THE INTERPLAY OF NEURAL DYNAMICS AND CONNECTIVITY STRUCTURES IN THE VISUAL CORTEX

by

CALEB JAMES HOLT

A DISSERTATION

Presented to the Department of Physics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2020

DISSERTATION APPROVAL PAGE

Student: Caleb James Holt

Title: The Interplay of Neural Dynamics and Connectivity Structures in the Visual Cortex

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics by:

Tristan Ursell	Chair
Yashar Ahmadian	Advisor
Paul Csonka	Core Member
Daniel Steck	Core Member
Michael Wehr	Institutional Representative

and

Kate Mondloch	Interim Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2020

DISSERTATION ABSTRACT

Caleb James Holt

Doctor of Philosophy

Department of Physics

June 2020

Title: The Interplay of Neural Dynamics and Connectivity Structures in the Visual Cortex

Here, we theoretically study how cortical networks' synaptic connectivity shapes their spiking activity dynamics, and in turn, how dynamics shape the structure of synaptic connectivity.

In the first part of this work, we study rhythmic oscillations of neural activity in the primary visual cortex (V1). These oscillations are characterized by power-spectra with peaks frequencies anywhere from 30 to 80 Hz, the gamma band. Gamma peaks shift to higher frequencies when V1 is stimulated by increasing contrasts. Moreover, the peak frequency depends on the *local* contrast stimulating a V1 sub-network. The local nature of this contrast-dependence would be trivially explained if the long-range intra-cortical connections within V1 were weak. However, experimental observations, such as the suppression of spiking rates with increasing stimulus size, point to those same long-range connections being functionally strong. Here we show that a model of V1 can balance the strength

of short- and long-range connections to successfully reproduce observations of gamma peak's local contrast-dependence as well as "surround suppression" of firing rates. Thus, we demonstrate how cortical network dynamics can be shaped by connectivity structures.

Visual information is relayed to V1 from the thalamus. The patterning of thalamocortical connections determines the location of visual space which V1 neurons respond to, their receptive field (RF). Cortical neurons moreover develop particular RF filters from thalamocortical connections. in carnivores and primates, particular features of those filters, such as their orientation, develop smooth maps across the cortex. However, in those same species, the RF filters of neighboring neurons are diverse and heterogeneous. Previous theoretical models of RF filter development predicted that if a smooth map forms for one feature, then RF filters should be homogeneous; failing to predict the observed co-presence. We extend those models by properly considering cortical activity dynamics. We show that in the presence of multiple timescales in cortical dynamics and thalamic inputs, RF filters can develop qualitatively different feature organizations, from smooth maps to heterogeneity. Thus, we demonstrate how cortical network dynamics shape connectivity structures.

This dissertation contains previously published co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Caleb James Holt

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:
University of Oregon, Eugene, OR
Univeristy of Oklahoma, Norman, OK

DEGREES AWARDED:
Doctor of Philosophy, Physics, 2020, University of Oregon
Bachelor of Science, Physics, 2013, University of Oklahoma

AREAS OF SPECIAL INTEREST:
Computational Neuroscience

PROFESSIONAL EXPERIENCE:

Teaching Assistant, University of Oregon, 2013-2020
Research Assistant, Ahmadian lab, University of Oregon, 2017-2018

GRANTS, AWARDS AND HONORS:

3rd Place, Three Minute Thesis Competition, University of Oregon, 2018

PUBLICATIONS:

Holt, C., Miller, K., Ahmadian, Y. (2020). The stabilized supralinear network accounts for the stimulus dependence of visual cortical gamma oscillations. *In preparation*

ACKNOWLEDGEMENTS

I would like to first and foremost acknowledge Dr. Yashar Ahmadian, my adviser. Through his tutelage I have become a better scientist, mathematician, programmer, and finally a better writer. Moreover, through our interactions I have become a better person. In addition to Dr. Ahmadian, I would like to acknowledge Dr. Tristan Ursell for admirably handling his role as committee head. I would also be remiss to not acknowledge Drs. Paul Csonka, Daniel Steck, and Michael Wehr for being inspirational teachers during my graduate studies and serving on my committee.

I would like to also acknowledge our neurotheory group. Dr. Takafumi Arakaki is an incredible scientist and has substantially helped with the technical parts of this work. Even though our work schedules didn't quite overlap, it was always great knowing he was working behind me (both literally, his desk was behind mine, and metaphorically). Dr. Gregory Barello was a wonderful research mentor as well. In particular, I would like to acknowledge Gabe for constantly encouraging me in my work, and being a shoulder to lean on when I needed it. Elliott Abe has been a constant companion, and tirelessly problem-solved with me. He will likely be much more productive now that my desk no longer borders his. I know he will do amazing things in his academic career. David Wyrick is an amazing data scientist. He has been wonderful to have around and joke with, even when he's wrong about free-will. Nicu Istrate shows great promise as a mathematical mind, and puts in the work to delve deep in theoretical studies. Finally I would be remiss to not mention Dr. Luca Mazzucato who's infectious enthusiasm for science has lifted all our spirits.

My graduate career started in experimental optics working for Dr. Miriam Deutsch, and special acknowledgment is due Dr. Andrea Goering for being a wonderful mentor during that time. Andrea showed me how to be a proper scientist, and for that I am grateful.

On a personal note, I would not have made it through graduate school without the love and support of my friends and family. I would especially like to acknowledge my first year cohort, and in particular highlight Spencer Alexander and John Myers. Further thanks is due to The Rock House: Jordan and Haley Palamos, Amelia and Ben Rhodeland, Kentaro Hoeger, Ariel Steele, Andrew Blaikie, and Savannah Logan for being wonderful friends during my graduate career. And to Amy Turner, for being a light to me, and helping me get through this trying time.

But I wouldn't have made it very far in graduate school without the Farmhouse. To those who have ever stopped by and hung out, thank you. To those who ever lived there special thanks are due. Michał Jander, thank you for coming in with us and getting us to play games together. Anna Guenther, for making the house warm and being a great cat-mom. Emily Ziff-Levine, for singing in the kitchen and being a joy in the house. Hannah Jarvis, for all the support and care you gave me, and your kind and gentle spirit. Fehmi Yasin, for getting us all to cook together and being a brother to me all these years. Ian Petersen, for romping through the outdoors and bringing joy and frivolity to my life. David Miller, for showing you care in the little moments and an eagerness for adventures. And finally, Julian Smith, for laughter and discussions for days and days. In my life, I've found it isn't so much where you are, as who you're with. And I've been very fortunate to be with you all for so long.

This work is dedicated to my mom, for constantly asking me math questions in car rides; my father, for teaching the value of hard work; my brother, for pushing me to continue to ask questions. Thank you for the love and support.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

INTRODUCTION: THE BRAIN AS PHYSICS

There are three fundamental forces that physics currently knows of: the gravitational force, the strong force, and the electroweak force. This work can be entirely described by one half of one of the three fundamental forces, the forces of electricity and magnetism, because, at its core, it deals with the movement and transport of ions through atoms and molecules. Collectively these ions, atoms and molecules make up the neuron, a nerve cell, whose membrane is full of channels and pumps for ions to move through. However, describing that movement purely using Maxwell's and Schrödinger's equations is hopelessly complex and needlessly obfuscating. As such, in this work we develop and use phenomenological models describing the activity of single neurons and collections of neurons that form an interconnected network. These phenomenological models can be used to faithfully recreate aspects of experimentally observed dynamical phenomena in biological neural networks, and yield novel testable predictions about the mechanisms underlying them.

In the rest of this introduction we review some key aspects of neurobiology to give the broad relevant background for our phenomenological models in Secs. 1.1-1.4. The reader who is familiar with that background may consider skipping ahead to Sec. 1.5 in which we give an overview of the focus of this work and the problems that motivate it. In Sec. 1.1 we introduce the structures of neurons and the action potential, a key communication signal between neurons. In Sec. 1.2 we briefly discuss the gross anatomy of the central nervous system and the brain. In Sec. 1.3, we discuss how the synaptic connections between neurons change due to their

activity, a process known as synaptic plasticity. In Sec. 1.4, we introduce reduced models of the biological complexity of neural activity and synaptic connections. In Sec. 1.5, we outline how those reduced models are applied in the remaining chapters of this work to gain insight about two sets of neurobiological phenomena. Lastly, we end this chapter with the abstracts for the works covered in Chs. II and III.

1.1. Neurons and action potentials

First, let us consider a single neuron (Fig. 1.1A). Neurons are biological cells characterized by three key compartments. The large bulbous part of the neuron is the soma, the cell body, which contains much of the cellular machinery of the neuron, including its nucleus and genome (Fig. 1.1A, red arrows). In Fig. 1.1A, above the soma are extensive branching structures highlighted by blue braces. These are the dendrites of the neuron, and function as its input channels. The dark black branch jutting out of the bottom of the neuron in Fig. 1.1A, is the axon, which functions as the output channel of the neuron (highlighted by green braces). Neurons communicate via these input and output channels through synapses, the connections between the axons of one neuron and the dendrite of another. Before further discussing the synapses which form the communication media, we will first give an overview of the electrical properties of neurons which generate the communicated signals.

The lipid membrane of a neuron is largely insulating, allowing neurons to act as capacitors (with a capacitance per unit area on the order of $1 \frac{\mu \mathrm{F}}{\mathrm{cm}^2}$ [3]) that store electric charge. The membrane is however an imperfect and leaky insulator; it is punctuated by special kinds of proteins called ion channels, which allow the

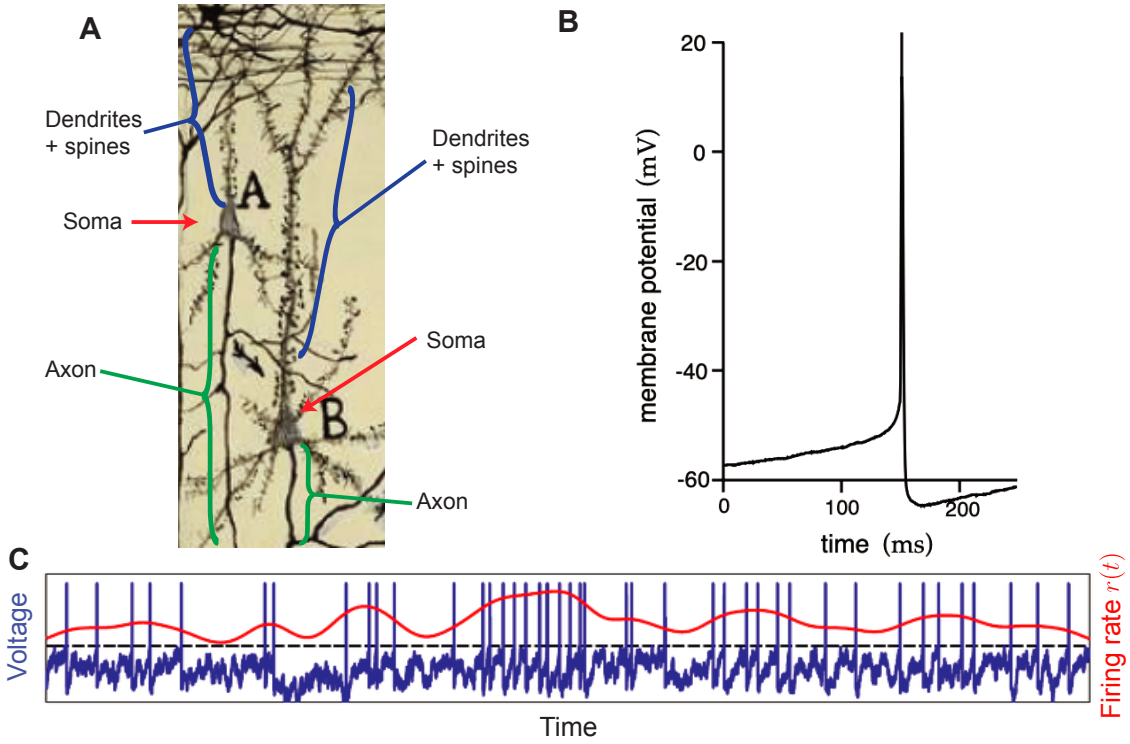

FIGURE 1.1. Introductory figure to neurons, spikes, and firing rates. **A:** Two examples of cortical excitatory neurons (A and B) drawn by Santiago Ramon y Cajal [1]. The bulbous structure of each neuron is its soma or cell body (red arrows). The branches with many small dots (little bumps called spines) around them are dendrites (blue brace), the input channels of the neurons. The dark black branches coming out the base of the neuron are the axons (green brace), the neurons' output channel. **B:** An example of an action potential (recorded by L. Rutherford in the laboratory of G. Turrigiano; image from [2]). **C:** A cartoon example of a voltage trace (blue) recorded from a neuron, containing many action potentials or spikes, and the associated firing rate (red). Every time the membrane voltage membrane crosses the spiking threshold (dashed black line) an action potential is emitted.

movement of charged ions in or out of the cell. If we neglect active biological processes for the moment, the membrane can thus be approximated electrically as a simple RC circuit, with an RC decay time (the time-scale over which stored charge in the cell leaks out) of around 10-20 milliseconds. At the same time, active biological devices called ion pumps (made up of other membrane proteins) spend metabolic energy, and act as batteries that constantly charge this leaky capacitor. Ion pumps maintain a generally negative electric potential across the cell membrane (referred to as the membrane potential), which at baseline sits at around -65 mV relative to the neuron's extracellular surround [4].

If the membrane potential of a neuron is raised above a specific threshold value, then an active positive feedback process initiates as follows. As membrane voltage rises beyond the threshold, certain voltage-gated ion channels in the cell membrane open up, allowing for an influx of positively charged sodium ions which raise the membrane potential even more. This positive feedback runaway process (with the voltage increase opening up more ion-channels and the latter causing a further rise in the membrane voltage, and so forth) ultimately results in a sharp and large peak in the membrane potential (with an amplitude on the order of $+100$ mV, and width on the order of 1-3 ms), known as an action potential or "spike" because of its shape (Fig. 1.1B). In general, only changes in membrane potential large enough to produce action potentials or spikes are communicated on to other neurons. For any change in membrane potential which does not exceed threshold, and in the absence of inputs, the membrane voltage simply returns to its resting potential. However, action potentials, once generated, travel down the axon of the neuron and arrive at its axonal terminals.

An axonal terminal is a bulge of neural membrane at the end of an axonal branch, and makes up the "pre-synaptic" part of a chemical synapse: a connection between two neurons. At the synapse, the axonal terminal meets with another bulge of neural membrane, called a spine, [1] on a dendrite of the "post-synaptic" neuron (Fig. 1.1A). The spine and the axonal terminal, together with the cleft between the two, make up the synapse (Fig. 1.3A). Once action potentials reach the axonal terminal of the pre-synaptic neuron, they cause the release of neurotransmitter molecules which diffuse through the cleft and bind to synaptic receptors on the dendritic membrane of the post-synaptic neuron. Those receptors in turn open up ion channels which allow in a flux of ions which changes the membrane potential of the post-synaptic cell. In this way, action potentials of a pre-synaptic neuron affect the membrane potential of its downstream post-synaptic partners.

As we said above, sub-threshold changes in the membrane voltage do not communicate to other neurons. Therefore, from a computational point of view, we can reduce the complex biological activity of a network of neurons to their spiking activity, or "activity" for short. One important measure of a neuron's spiking activity and its responses to stimuli is its firing rate, which is roughly the number of spikes fired by that neuron in a unit of time. More precisely, the instantaneous firing rate of a neuron (Fig. 1.1C, red curve) can be derived from its spike train (Fig. 1.1C, the spikes in the blue curve) by many means, *e.g.*, by taking a running average of spike counts in some sliding window. In the rest of this dissertation,

[1] As with everything in biology, there are many exceptions: not all chemical synapses are formed on post-synaptic spines. Many form on spineless dendrites and yet others can target the soma or the initial segment of axon of the post-synaptic cell.

when referring to a neuron's activity or response, we more precisely mean its firing rate or a change therein.

1.2. Neocortex and sensory cortical areas

Neurons are present throughout the body, from the muscles and gut to the central nervous system, *i.e.* the spinal cord and the brain. This work will focus on studying networks of neurons within the neocortex of the brain. The neocortex (hereafter cortex) is the outer wrinkled sheet of grey matter in the mammalian brain composed of a six-layered structure. It was the most recent area of the brain to evolve. Broadly speaking, different areas of the cortex have been specialized for different functions through evolution, with areas specializing in processing sensory stimuli lying towards the back of the head (posterior) and areas specializing in motor control and planning towards the front (anterior).

The areas which process sensory stimuli are known as sensory cortices. Each sense has an associated set of corresponding areas within the cortex, *e.g.* vision has the visual cortices, hearing the auditory cortices, and smell the olfactory cortices. Moreover, for each sensory modality the cortical area in which sensory information arrives first is referred to as the primary sensory cortex, *e.g.* primary auditory cortex, or primary visual cortex (also known as V1; Fig. 1.2).

Primary sensory cortices typically receive sensory information from their corresponding sensory organ via the thalamus (the exception being the primary olfactory cortex). The thalamus lies within a structure in the center middle part of the brain known as the diencephalon. For example, the eyes' retinas connect to the lateral geniculate nucleus (LGN) of the thalamus via the optic nerves and the optic tract; see Fig. 1.2. Through these connections, each thalamic neuron

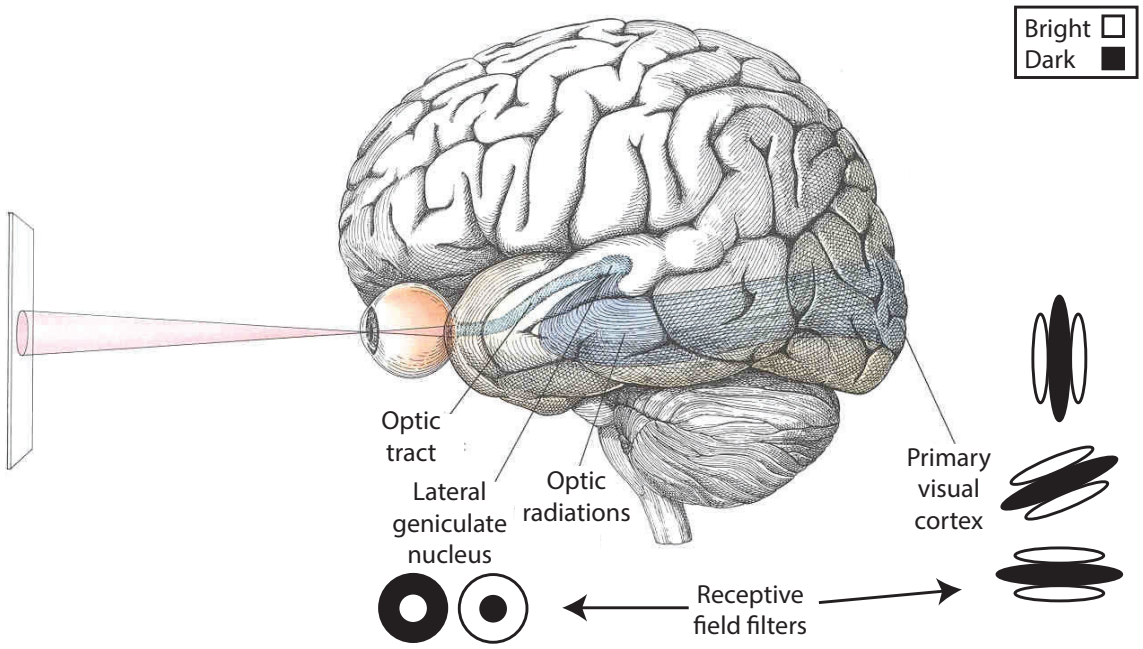

FIGURE 1.2. A sketch of the cerebral cortex (by David H. Hubel [5]) depicting the early visual pathway. Signals from the retina are communicated along the optic tact to the lateral geniculate nucleus (LGN) of the thalamus. The optic radiations form the feedforward thalamocortical connections to the primary visual cortex (V1). At the bottom and right margins, we show stylized examples of receptive field filters for ON- and OFF-center cells in the LGN, and for orientation selective cells in V1.

within the LGN receives information about a particular location in the visual field, known as its receptive field (RF). Moreover, LGN neurons respond strongly to certain visual patterns within their RF and not to others; the preferred pattern of a cell within its RF is therefore called its RF filter. To a first approximation, the thalamus functions as a relay center, sending signals from the sensory organs to the cortex largely unchanged (indeed, in the LGN the RF filters are mainly inherited from retinal cells, and thus signals from the retina are communicated to V1 with little transformation). These signals are communicated via feedforward thalamocortical connections (see *e.g.* the optic radiations in Fig. 1.2).

As we will see in Ch. III, these feedforward thalamocortical connections determine the RF filters of the neurons in the primary visual cortex (and similarly in other primary sensory cortices), and thus the stimulus characteristics or *features* that those neurons respond to most strongly [6, 7]. Cortical neurons thus acquire selectivities for particular features of stimuli, and in particular develop a RF. Moreover, nearby neurons prefer nearby regions of the visual field; their RF locations therefore form a smooth map of the visual field across the cortex, a property known as retinotopy. V1 neurons acquire RF filters from combinations of the RF filters of LGN neurons (Fig. 1.2). However, the RF filters of V1 neurons develop selectivities to features in the stimulus that are not present in the RF filters of LGN neurons. In particular, V1 neurons prefer (or respond most strongly to) oriented patterns of light and dark bands; as such orientation preference and selectivity fist emerges in the visual pathway in V1.

1.3. Synaptic plasticity

As we said above, visual cortical RF filters arise from the structure of feedforward thalamocortical connections. Feedforward thalamocortical connections undergo rapid changes in strength during late pre- or early post-natal development (depending on the species) [8, 9, 10]. The feature preferences which develop during this time depend on the activity of the pre- and post-synaptic neurons (by contrast, retinotopy in V1 seems to develop according to a genetically defined program and in this sense is "hard-wired") [11, 12]. The process of synapses changing their strength based on the activity of the connected neurons is known as synaptic plasticity. There have been many models for how synapses change with activity, but we will rely on so-called Hebbian plasticity or Hebbian learning (see

Ch. III for more details) to model development in this work. Hebbian learning is known by the aphorism that "cells that fire together, wire together" [13].

The synaptic connections between neurons come in two main types, excitatory or inhibitory. Either an action potential arriving at the axonal terminal causes the membrane potential of the post-synaptic neuron to increase, in which case the synapse is *excitatory* as the post-synaptic neuron is more likely to fire a spike, or an action potential causes the membrane potential of the post-syanptic neuron to decrease, in which case the synapse is *inhibitory* as the post-synaptic neuron is now less likely to fire a spike. Excitatory and inhibitory synapses release different neurotransmitter molecules, and cannot change into one another by synaptic plasticity. Interestingly, the synapses formed by a particular (pre-synaptic) neuron are generally either all excitatory or all inhibitory, a principle known as Dale's Law or Dale's Principle [14, 15, 16]. Moreover, long-range connections in the brain, *e.g.* feedfoward thalamocortical connections, are generally excitatory. It is generally the excitatory synapses that undergo Hebbian or Hebbian-like plasticity.

In addition to respecting the constraints imposed by the fixed excitatory/inhibitory type of a synapse, biologically plausible plasticity models cannot allow for unbounded growths in synaptic strengths. In Ch. III, we explore the development of orientation preference and other selectivities in V1, by activity-dependent Hebbian plasticity in thalamocortical connections, under the above constraints.

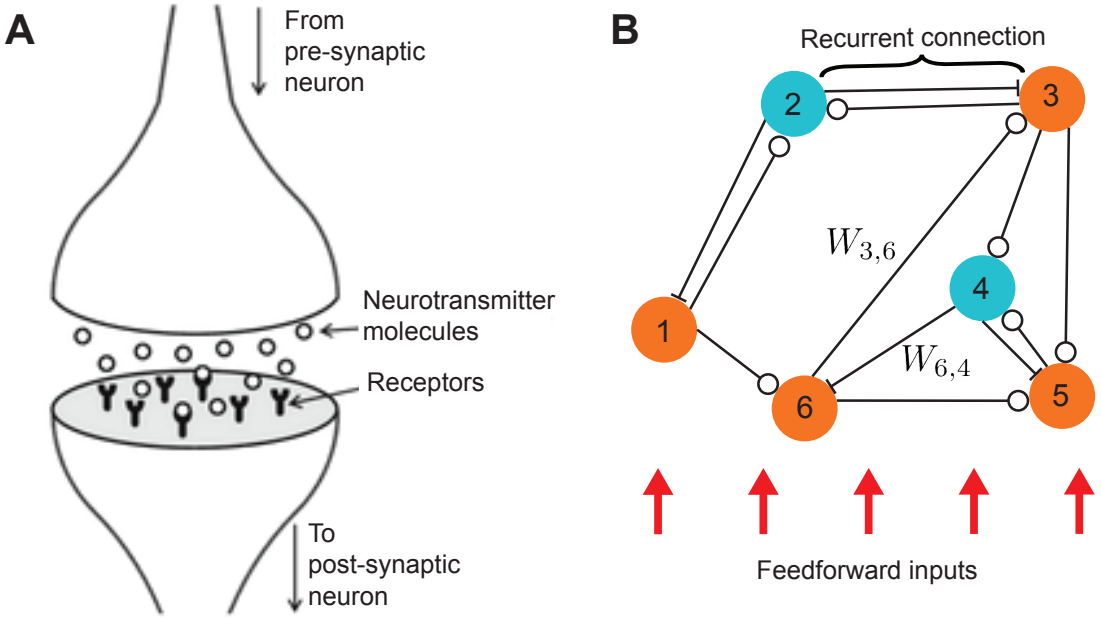

FIGURE 1.3. Example of a chemical synapse and networks of neurons **A:** Diagram of a chemical synapse (from [17]). The top part depicts the axonal terminal of the pre-synaptic neuron, while the bottom structure is a spine on a dendrite of the post-synaptic neurons. In the synaptic the cleft between the two, the pre-synaptic neuron releases neurotransmitter molecules whenever an action potential fired by the pre-synaptic neuron reaches the terminal. The neurotransmitters then diffuse in the extracellular space and attach to receptors on the post-synaptic cell's membrane. Receptors of the post-synaptic then open their ion channels resulting in a change in the membrane potential of the post-synaptic neuron. **B:** Example of a recurrent network with external input from a stimulus. Stimulus inputs are shown in red, and recurrent connections are shown in black. The network is composed of excitatory (orange) and inhibitory (cyan) neurons, with excitatory synapses ending in a circle and inhibitory synapses ending in a line. An example connection from neuron 1 to 2 is labelled as $W_{2,1}$.

1.4. Reduced models of neurons and synapses

Neurons fire spikes constantly, often in an irregular noisy-looking fashion (with an average firing rate around $5 - 10$ Hz, in brain regions of interest to this work [18]). While there is some debate as to whether the timing of spikes versus their rate is more important for neural communication [19] (see [2] Ch. 1 for a more detailed discussion), the firing rate, or number of spikes per unit time, is undoubtedly a key feature of neural responses. Recall that the instantaneous firing rate of a neuron can be approximated by taking a running average of spike counts in some sliding window (Fig. 1.1C). In this work, we adopt the commonly held "rate coding" point of view, and will study phenomena that can be modeled directly at the coarse-grained level of firing rates, without resolving or modeling individual spikes (similar to hydrodynamic modeling, which does not refer to the molecules composing a fluid). In short, we reduce the complex biological machinery and electrophysiology of the nerve cell to a model in which the entire neuron is represented by a single dynamical variable: its instantaneous firing rate $r(t)$.

Now that we have reduced the biological complexity of the neuron, we similarly seek to reduce the complexity of the synapse. A key characteristic of synapses is their synaptic efficacy or strength, which is measured by the change in the membrane potential of the post-synaptic neuron resulting from a spike by the pre-synaptic neuron. Biologically, the synaptic efficacy depends on a number of factors, including the amount of neurotransmitter released by the pre-synaptic terminal and the number of synaptic receptors expressed on the post-synaptic membrane (Fig. 1.3A). In this work, the synaptic strength or *weight*, is defined as the total change in the net input to the post-synaptic neuron for a unit change in

the firing rate of the pre-synaptic neuron. Thus, we again reduce all the biological complexity of the synapse to one number, its weight. In this work, we will denote the weight of synaptic connections between neurons in the form $W_{post,pre}$, *e.g.* $W_{3,6}$ is the connection from neuron 6 to neuron 3 (see Fig. 1.3B).

Fig. 1.3B schematically shows a network of reduced rate neurons connected by synaptic weights. Excitatory neurons are depicted in orange with connections ending in circles in Fig. 1.3B, while inhibitory neurons are depicted in cyan with connections ending in lines.

Broadly speaking, neurons receive two forms of connections, feedforward, *e.g.* thalamocortical connections, or recurrent, *e.g.* connections within the cortex. Feedforward connections go from one population to another; the connections only go one way, *e.g.* the feedforward inputs in Fig. 1.3B (red arrows). Recurrent connections, on the other hand, include reciprocal connections between neurons, as in the connections between 2 and 3 in Fig. 1.3B. More generally, recurrent connections can form loops, as formed, *e.g.*, by the connections between neurons $3, 4$ and 6 in Fig. 1.3B. Networks may have both feedforward and recurrent connectivity structures. In Fig. 1.3, recurrent connections which end in circles are excitatory, while recurrent connections which end in lines are inhibitory. In this work, feedforward connections correspond to thalamocortical connections, while recurrent connections correspond to intra-cortical connections within V1.

In a network of N neurons, we will arrange the firing rates of all neurons in an N-dimensional rate vector, $\boldsymbol{r}(t)$. Similarly, the connections between neurons can be represented as an $N \times N$ matrix, typically labelled as W, with matrix element W_{ij} denoting the weight of the synapse from neuron j to neuron i. In a network respecting Dale's law, introduced above, the weights of all synaptic projections

from an excitatory (inhibitory) neuron are positive (negative). For example, in Fig. 1.3, both $W_{3,6}$ and $W_{5,6}$ are greater than zero as both are connections from an excitatory neuron. Mathematically, Dale's law imposes the constraint that all elements in each column of the weight matrix, W, share the same sign.

The total net input to a neuron is the sum of its recurrent and feedforward inputs. We denote the latter (shown by the red arrows in Fig. 1.3B) as an N-dimensional vector, $\mathbf{I}$. The recurrent input to a neuron is the sum of the contributions of all the other neurons inside the network. The input contribution of neuron j to neuron i is proportional to the firing rate of the pre-synpatic neuron, r_j, and to the weight of synaptic connection it forms onto neuron i, W_{ij}. Thus the vector of recurrent inputs to all neurons can be written as the matrix-vector product $W\boldsymbol{r}$. At steady-state, the rate of a neuron is related to its total net input, v, by a nonlinear input/output (I/O) transfer function, $f(v)$, which is a biological property of the neuron. Therefore, assuming a steady-state equilibrium exists, the network's rate vector, $\boldsymbol{r}$, in that steady-state must satisfy

$$\boldsymbol{r} = f(W\boldsymbol{r} + \mathbf{I}). \tag{1.1}$$

To obtain an equation describing the dynamics outside the steady state, it is assumed that the rate of each neuron (in isolation) would exponentially relax towards its equilibrium value over a relaxation time τ_r. This leads to the following set of coupled ordinary differential equations (or dynamical system), known as rate equations

$$\tau_r \frac{d\boldsymbol{r}(t)}{dt} = -\boldsymbol{r}(t) + f(W\boldsymbol{r}(t) + \mathbf{I}(t)). \tag{1.2}$$

The neuronal I/O transfer function, $f(v)$, is a key ingredient of firing rate equations. Many different classes of I/O transfer functions have been used in theoretical neuroscience, many of which saturate at an upper bound, *e.g.* $f(x) = 1/(1 + e^x)$, while others do not. In Ch. II, we will study how a non-saturating I/O transfer function is particularly useful in reproducing certain features of the dynamics of neural activity in cortex.

1.5. The focus of this work

The uniting theme of this dissertation research has been the study of how synaptic connections and neural activity dynamics impact one another. In Ch. II, we use a previously successful model of cortical activity to study the recurrent connectivity structures underlying oscillations of neural activity, and thus study how recurrent connections impact the dynamics of neural activity at fast time-scales. Then in Ch. III, we study the development of neural selectivities via synaptic plasticity in feedforward thalamocortical connections. Specifically, we model how the intrinsic dynamics of the cortical network, which was not previously considered in theoretical models of cortical plasticity and development, impacts the development of feedforward selectivity structures through a novel mechanism. In short, we study how the dynamics of neural activity at multiple (but relatively fast) time-scales impact the development of connectivity structures. In the following paragraphs, we present the relevant background to each chapter.

As we have said, a group of neighboring neurons within V1 are responsive only to a certain location within the visual field, known as their receptive field (RF). In experimental studies, neural responses to stimuli within their RF's can be recorded using a variety of techniques. In particular, some studies record from

the local population of neurons by inserting an electrode into the extracellular fluid surrounding the neurons. These recordings show a noisy trace of the electrical activity of the population, and when low-pass filtered are known as the local field potential (LFP), a measure of the population's average activity (see [20, 21] for examples). Populations of neurons within primary visual cortex exhibit elevated rhythms in their LFP when stimuli are present in their RF. The rhythms with the greatest increases in power occur in the gamma band of frequencies ($30 - 80$ Hz).

In Ch. II we study under what spatial recurrent connectivity structures, a model of V1 can reproduce certain experimental observations regarding gamma oscillations. That chapter contains co-authored and published work by Caleb Holt, and Drs. Ken Miller and Yashar Ahmadian. We show that a previously successful model of cortical dynamics can produce gamma oscillations in its LFP. We then show that this cortical model robustly produces these oscillations and their dependence on stimulus strength across many parameter choices in a simple two-population (one excitatory, one inhibitory) network. We further demonstrate that an extended retinotopic version of this model successfully captures how gamma oscillations depend on stimulus strength in a spatially local manner. Moreover, we find that a particular cortical connectivity structure is key to producing the above phenomenon.

To show that the dynamics of cortical activity at multiple time-scales are important in shaping feedforward connectivity structure, in Ch. III we study RF feature development in V1. That chapter contains co-authored and un-published work by Caleb Holt, and Drs. Ken Miller and Yashar Ahmadian. During RF feature development, the feedforward thalamocortical connections change based on the activity of the pre- and post-synaptic neurons. Different feature selectivities

are known to develop particular organizations across the cortex, either a smooth map where nearby neurons prefer similar features or a disorderly (so-called salt-and-pepper) organization where the selectivity of a neuron to a feature has little to no correlation with its neighbors' selectivities to that feature. Previous models of feedforward thalamocortical development predicted that the selectivity to every feature would develop the same type of organization across the cortex [22]. Either all feature selectivities would develop a smooth map, or a salt-and-pepper organization. However, in experimental studies, a mixture of selectivity organization are observed for different features in the same visual cortex [23].

Previous theoretical models of feature selectivity development using Hebbian learning did not consider the dynamics of the post-synaptic cortical population. Rather, the activity of the cortical population was assumed to be always at the fixed-point corresponding to its instantaneous LGN input. Here we show that properly accounting for the intrinsic dynamics of V1 and the temporal variations in the LGN input, which can both feature multiple time-scales, gives rise to a more biologically plausible model of feedforward connectivity development. This new theoretical model of development leads to a certain coupling between multiple time-scales of LGN input features and cortical response dynamics. This coupling allows selectivities to features that vary at different time-scales to potentially develop qualitatively different organizations (*e.g.* smooth versus salt-and-pepper), as is observed biologically. We show results from simulations using our new model of feedforward development in network models with both one- and two-dimensional retinotopy. The cortical population in these models develops different organizations of selectivity to different input features. Additionally, we compare how these models perform against a previous model of feedforward

development which did not adequately account for cortical dynamics, and show that the previous model fails to develop different organizations for the features considered [22]. We finish this dissertation with discussions of generalizations and future directions for research in Ch. IV.

1.5.1. Abstract for Ch.II

When stimulated, visual cortical populations exhibit fast rhythmic activity with frequencies in the gamma band (30-80 Hz). The gamma rhythm manifests as a broad resonance peak in the power-spectrum of recorded local field potentials which exhibit several interesting stimulus dependencies. In particular, in macaque primary visual cortex (V1), the gamma peak frequency increases with increasing stimulus contrast. Moreover, this contrast dependence is local: when contrast varies smoothly over visual space, the gamma peak frequency in each cortical column is controlled by the local contrast in the column's receptive field. No parsimonious mechanistic explanation for these contrast dependencies of V1 gamma oscillations has been proposed. The stabilized supralinear network (SSN) is a mechanistic model of cortical networks that has accounted for a range of response nonlinearities and contextual modulations in the visual cortex and their contrast dependencies. We first show that a reduced SSN model lacking topography robustly captures the contrast dependence of gamma peak frequency, and provides a mechanistic explanation for this effect based on the observed non-saturating and supralinear input-output function of V1 neurons. Given this result, the local dependence on contrast can trivially be captured in a retinotopically extended SSN which lacks horizontal synaptic connections between cortical columns, as in that case the gamma peak in a cortical column would

be independently set by that column's feedforward input. However, contextual modulation phenomena such as surround suppression are known to be partly mediated by horizontal connections. We thus explored whether a retinotopically organized SSN model of V1 with strong excitatory horizontal connections can exhibit both surround suppression and the local contrast dependence of gamma peak frequency. We found that such a SSN can account for both effects, but only when the excitatory projections are composed of two components with different patterns of spatial fall-off with distance: a short-range component which only targets the source column, combined with a long-range component that targets columns neighboring the source column. We thus make a specific qualitative prediction for the spatial structure of horizontal connections in macaque V1, consistent with the columnar structure of cortex.

1.5.2. Abstract for Ch.III

In many mammalian species, the primary visual cortex (V1) develops smooth maps for several receptive field features, such as preferred orientation or spatial frequency. Such maps may be beneficial in minimizing wiring lengths between neurons selective to similar features. Nevertheless, even in visual cortices with smooth maps, the receptive fields of nearby neurons show considerable heterogeneity. Correspondingly, some receptive field features are largely uncorrelated between nearby cells, and average signal correlations between nearby cells are near zero. Such random, "salt-and-pepper" organizations may in turn be advantageous in reducing the response redundancy of local V1 populations and increasing their information content. Thus a combination of smooth maps for some features, and salt-and-pepper organizations for others, may provide both

the benefits of wiring length minimization and informational efficiency. Previous theoretical models have described the development of V1 feature selectivity and maps based on activity-dependent Hebbian plasticity. However, these models inevitably predict that V1 either develops smooth maps for all features (when long-range recurrent cortical excitation is strong) or salt-and-pepper organizations for all features (if cortical recurrent excitation is weak or short-range); they fail to account for the biologically observed co-presence. We propose that this failure is in part due to these models neglecting the intrinsic temporal dynamics of V1, and show that if properly considered, cortical interactions at slow and fast time scales will couple to the slow and fast features of inputs to V1, respectively. This can lead to the development of smooth maps for slow input features and salt-and-pepper organizations for fast input features. In particular, by simulating 1-and 2-dimensional topographic models of development of plastic feedforward thalamocortical connections, we show that our framework can sustain both smooth maps and salt-and-pepper organizations, providing a more biologically plausible mechanism for receptive field feature development in V1.

CHAPTER II

THE STABILIZED SUPRALINEAR NETWORK ACCOUNTS FOR THE STIMULUS DEPENDENCE OF VISUAL CORTICAL GAMMA OSCILLATIONS

The simulations and analyses done in this chapter were carried out by myself under the guidance of Dr. Yashar Ahmadian, the principal investigator who conceived the project. Dr. Ken Miller originated some of the questions motivating this work, and contributed to some of the conceptual solutions.

2.1. Introduction

When presented with a stimulus, populations of neurons within visual cortices exhibit elevated rhythmic activity with frequencies in the so-called gamma band (30-80 Hz) [20, 21]. These gamma oscillations can be observed in local field potential (LFP) or electroencephalogram (EEG) recordings and, when present, manifest as peaks in the LFP/EEG power-spectra. It has been proposed that gamma oscillations perform key functions in neural processing such as feature binding [24], dynamic communication between cortical areas [25, 26, 27], or as a timing or "clock" mechanism that can enable coding by spike timing [28, 29, 30, 31, 32]. These proposals, however, remain controversial.

While the functional role of gamma rhythms is not fully understood, much is known about their phenomenology. For example, defining characteristics of gamma oscillations, such as peak frequency, half-width, and height of the spectral gamma peak, exhibit systematic dependencies on various stimulus parameters [20, 21, 33, 34]. In particular, in the primary visual cortex (V1) of macaque monkeys, the power-spectrum gamma peak moves to higher frequencies as the

contrast of large and uniform grating stimuli is increased [20, 21]. This establishes a monotonic relationship between gamma peak frequency and the grating contrast. We will refer to this contrast-frequency relationship, obtained using a grating stimulus with uniform contrast, as the "contrast dependence" of gamma peak frequency.

Moreover, when animals are presented with a stimulus whose contrast smoothly varies over the visual field (and hence over nearby cortical columns in V1), it is the local stimulus contrast that determines the peak frequency of gamma oscillations [21]. Specifically, Ref. [21] used a Gabor stimulus (which has smoothly decaying contrast with increasing distance from the stimulus center), and found that the gamma peak frequency of different V1 recording sites match the predictions resulting from the frequency-contrast relationship obtained from the uniform grating experiment when using the local Gabor contrast in that site's receptive field. We refer to this second effect as the "local contrast dependence" of gamma peak frequency.

It is well-known that networks of excitatory and inhibitory neurons with biological neural and synaptic time-constants can exhibit oscillations with frequency in the gamma band [35, 36]. There exists a class of network models with gamma frequency tied to the single neuronal spiking [37, 38, 39]. These models come in two types: those that rely on recurrent connections between inhibitory units, known as inhibitory network gamma (ING), and those that rely on recurrent interactions between pyramidal (excitatory) and inhibitory neurons, known as pyramidal-inhibitory network gamma (PING). However, these ING and PING models generally produce gamma rhythms by synchronized regular spiking of different neurons firing at the same frequency across the network. In the visual

cortex (as well as other cortical areas), by contrast, neurons typically fire spikes irregularly, in a Poisson-like manner [40], with a wide distribution of firing rates [41]. Lastly these models produce very sharp gamma peaks with higher harmonics [37], which is not seen in cortical studies (see *e.g.* [21]). Here, by contrast, we develop a rate model of gamma rhythms, appropriate for modeling the asychronous irregular state of cortical activity.

No mechanistic circuit model of visual cortex has been proposed which can robustly account for the local contrast dependence of gamma oscillations. Ref. [20] developed a rate model which does account for the dependence of gamma peak frequency to changes in the global contrast. However, that model only explicitly accounts for the interactions between a local excitatory and inhibitory population, and thus cannot account for the local nature of the gamma peak frequency's contrast dependence. Moreover, even in the case of a stimulus with globally uniform contrast, this model could only produce very small contrast-dependent changes in peak frequency, which furthermore required the scaling of the intrinsic time constant of the excitatory neurons depending on contrast. As such, we seek to develop a self-contained and more parsimonious mechanistic model which can account for changes in global and local contrast dependencies of gamma with fixed parameters.

It is not clear how the local nature of the contrast-dependence of gamma oscillations can be reconciled with key features of cortical circuits. This locality would naturally emerge if cortical columns were non- or weakly interacting; in that case each column's oscillation properties would clearly be determined by its feedforward input (controlled by the local contrast). However, nearby cortical columns do interact strongly via the prominent horizontal connections connecting

them [42]. These interactions manifest, *e.g.*, in contextual modulations of V1 responses, such as in surround suppression [43], which are known to be mediated partly by horizontal connections [44].

Surround suppression is the phenomena wherein stimuli outside the classical receptive field (RF) of V1 neurons, which by themselves cannot drive the cell to respond, nevertheless modulate the cells' response, typically by suppressing it. Surround suppression results in a non-monotonic "size tuning curve", which is obtained by measuring a cell's response to circular gratings of varying sizes centered on that cell's RF: the response first increases with increasing stimulus size, but then decreases as the grating increasingly covers regions surrounding the RF. It is unclear whether a model of V1, featuring biologically plausible horizontal connections, can capture both surround suppression and the local contrast dependence of gamma oscillations.

A parsimonious, biologically plausible model of cortical circuitry which has successfully accounted for a range of cortical contextual modulations and their contrast dependence is the stabilized supralinear network (SSN) [45, 46]. In particular, the SSN robustly captures the contrast dependencies of surround suppression, such as the observation that size tuning curves peak at lower stimulus sizes with increasing stimulus contrast [45].

Being a rate model with excitatory and inhibitory neurons, we expect the SSN to be able to exhibit oscillations similar to gamma rhythms. Gamma oscillations do not behave like sustained oscillations, as sustained oscillations display a sharp peak in the power spectrum, typically with subsequent harmonics. By contrast, gamma peaks are broad, with no visible harmonics, consistent with transient (damped) and noise-driven oscillations [47, 48, 49]. We thus studied the

SSN in a regime that in the absence of time-dependent external inputs its firing rates reach a steady state, but can exhibit damped oscillations when perturbed (technically, this means the network is close to, but below a Hopf bifurcation, *i.e.*, a transition to a regime of sustained oscillations). When perturbed by structureless noise that is sufficiently fast (the biological network's irregular spiking can provide such a noise source), these noise-driven damped oscillations manifest as a resonance peak in the power-spectrum of network activity [49, 50]. As pointed out previously, to capture a resonance frequency in the gamma band, it is key to account for fast synaptic filtering as provided by the fast AMPA and GABA receptors [36, 51]. We will thus extend the SSN model to properly account for different synaptic currents through different receptor types, and their different filtering timescales.

The key feature of SSN is the supralinear rectified-power law input-output (I/O) function of its neurons. Based on intracellular recordings, this is a good fit to the non-saturating and expansive relationship between the firing rate and membrane voltage of V1 neurons [52]. This supralinear I/O transfer function results in increasing gains (the slope of I/O curve) with increasing net input to or output rate of neurons (see Fig. 2.1A), which typically go up with increasing external drive to the network which is controlled by stimulus contrast. This increase of gains with increasing drive, which in turns results in the strengthening of effective recurrent connections, is the mechanism underlying SSN's ability to capture the contrast-dependence of contextual modulations like surround suppression. We hypothesize that the same mechanism allows the SSN to robustly account for the contrast dependence of gamma peak frequency.

The aim of this work is to explore this hypothesis. We start Sec. 2.2 by developing an extended version of the SSN which accounts for synaptic inputs mediated by different synaptic receptor types. We then study a reduced SSN composed of two units representing excitatory (E) and inhibitory (I) sub-populations. We show that, for a wide range of biological parameters, this reduced SSN model generates gamma oscillations with peak frequency that robustly increases with increasing external drive to the network. We show that this robust contrast dependence is indeed a consequence of the expansive and non-saturating SSN neural nonlinearity. We next investigate the local contrast dependence using an expanded retinotopically organized SSN model of V1, with E and I units in different cortical columns. We show that this network is capable of reproducing the local contrast dependence of gamma oscillations while exhibiting realistic surround suppression, but only when excitatory projections have a certain spatial profile. Specifically, we show that when excitatory horizontal connection strengths have a smooth fall-off with distance between source and target cortical columns, the network fails to capture local contrast dependence.

Accounting for the latter requires a sharp immediate fall in the strength of excitatory projections as soon as the target column deviates from the source column. In a finer grained model, this would correspond to two spatial scales in the horizontal connection fall-off: a shorter length-scale on the order of the V1 mini-column size, and a longer ranged component stretching over a few mini-columns. This "local plus long-range" spatial structure of horizontal connections balances the trade-off between capturing local contrast dependence (requiring short-range or weak horizontal connections) and surround suppression (requiring the opposite). While achieving this balance requires a mild degree of parameter

tuning, we show that it does not require fine-tuning of parameters and is robust to parameter variations of up to $\pm 10\%$, if not larger. In Sec. 2.3, we present the details of the reduced and retinotopic models, the mathematical analyses, and numerical simulations. In Sec. 2.4, we conclude by discussing the implications of our findings for the structure of cortical horizontal connections and the shape of neural input/output nonlinearities.

2.2. Results

2.2.1. Stabilized supralinear network (SSN)

In this chapter we will use the Stabilized Supralinear Network (SSN) to model cortical dynamics, and gamma oscillations in particular. The SSN has been a successful model of cortical response dynamics and has successfully accounted for a range of cortical contextual modulation effects, such as surround suppression, and their contrast dependence [45, 46].

The SSN model belongs to the firing rate equations class of models introduced in Eq. (1.2). The key component of the SSN is its non-saturating and expansive neuronal Input/Output (I/O) transfer function. Neurons cannot fire action potentials beyond a certain rate (around 200 Hz, or perhaps higher) as they enter a so-called refractory period after they fire a spike, during which they cannot generate another action potential. However, average firing rates in cortex generally stay quite low compared to those bounds. In fact, in V1 no saturation is observed in the I/O function of single-neurons throughout their entire natural range of firing rates driven by visual stimuli, including optimal preferred stimuli. Moreover, the relationship between a V1 neuron's output firing rate and its mean membrane potential (measured relative to the rest potential, and used as a surrogate for the

neuron's mean input) is well approximated by a supralinear rectified power law [53, 54]. Thus, in the notation of Eq. (1.2), the SSN uses the I/O transfer function

$$f(v) = k[v]_+^n \tag{2.1}$$

where v is the net input to the neuron, k is a positive constant, $n > 1$ (corresponding to supralinearity), and we defined $[x]_+ \equiv \max(0, x)$ (denoting rectification). See Fig. 2.1A's inset for a plot of this I/O function for the case $n = 2$.

The dynamical rate equations governing the dynamics of an SSN composed of N neurons are thus given by [46]

$$\mathrm{T}\frac{d\boldsymbol{r}_t}{dt} = -\boldsymbol{r}_t + f\left(W\boldsymbol{r}_t + \mathbf{I}_t\right) \tag{2.2}$$

where $\boldsymbol{r}_t$ and $\mathbf{I}_t$ are the N-dimensional vectors of the neurons' instantaneous firing rates and external inputs, respectively, W is the $N \times N$ matrix of synaptic connection weights, and f acts element-wise. These equations are thus the same as the generic firing rate equations Eq. (1.2), with the I/O nonlinearity given by the supralinear power law Eq. (2.1), and the following additional modification. Here we allow for different neurons to have different relaxation times, by turning the scalar time-constant τ_r of Eq. (1.2) to a diagonal $N \times N$ matrix, T, with positive diagonal elements corresponding to the relaxation times of different neurons.

The SSN's connectivity matrix W respects Dale's principle (see Sec. 1.3), meaning each neuron forms either purely excitatory (E) or purely inhibitory (I) connections [16]. We will organize the elements of the rate vector, $\boldsymbol{r}_t$, such that the first N_E elements correspond to excitatory neurons, and the last $N_I = N - N_E$

inhibitory cells. Further, we assume an equal number of excitatory and inhibitory neurons, i.e. $N_E = N_I = N/2$. Thus W can be written as

$$W = \begin{pmatrix} W_{EE} & -W_{EI} \\ W_{IE} & -W_{II} \end{pmatrix} \tag{2.3}$$

with blocks W_{ab} ($a, b \in [E, I]$) by definition composed of non-negative elements (in our convention, W_{ab} denotes connections from the population of type b to the population of type a). We note that even though so far we have been treating the elements of the firing rate vector, $\boldsymbol{r}_t$, as representing the firing rates of single neurons, in many applications of firing rate equations, including ours below, the "neurons" should more accurately be interpreted as sub-populations of neurons (or equivalently as typical or mean-field neurons in a sub-population). Thus a component of $\boldsymbol{r}_t$ should be taken as representing the average firing rate of a neuronal sub-population. Similarly a single element, W_{ij}, of the matrix W represents the summed synaptic weight received by a neuron in sub-population i from all the single neurons in population j. We thus use the terms "unit" or "sub-population" below, instead of (or interchangeably with) "neuron". (This interpretation also justifies our choice $N_E = N_I$, despite the fact that in the cortex excitatory single neurons outnumber inhibitory ones by about a factor of four.)

As shown in [46], when the external input is a constant, the dynamical solutions of Eq. (2.2) generically converge to a stable steady state (or fixed point), provided that the inhibitory neurons are sufficiently fast (*i.e.* their time-constants, as given by the corresponding elements of T are fast relative to the time-constants

of excitatory neurons).[1] Steady-state solutions are obtained by setting the left hand side of Eq. (2.2) to zero.

Firing rate equations can also be formulated differently by treating the vector of net inputs to neurons, $\mathbf{v}_t$, as the dynamical variables, instead of rates, $\boldsymbol{r}_t$. This leads to the equations

$$\mathrm{T}\frac{d\mathbf{v}_t}{dt} = -\mathbf{v}_t + Wf(\mathbf{v}_t) + \mathbf{I}_t. \tag{2.4}$$

We also introduce this formulation here, as it is more closely related to the extended SSN model that we will introduce below in Sec. 2.2.3. When the external input is constant, it can be shown [55] that (under mild conditions on W) the steady-state solutions of Eq. (2.4) are equivalent to the steady-state solutions of Eq. (2.2) via the identification $\boldsymbol{r} = f(\mathbf{v})$ (or equivalently $\mathbf{v} = W\boldsymbol{r} + \mathbf{I}$).

2.2.2. Gamma rhythms as noise-driven oscillations

Empirical evidence is most consistent with visual cortical gamma oscillations resulting from noise-driven fluctuations, as opposed to resulting from sustained coherent oscillations [47, 48, 49, 50]. Sustained oscillations would have several defining features, such as a sharp peak in the power spectrum, likely trailed by subsequent harmonic peaks. Such oscillations would also be auto-coherent, *i.e.* have a consistent phase over several oscillation cycles, and occur with regularity both in timing and duration. Gamma oscillations, however, display none of the above features. They manifest as a single broad peak in power spectra [20, 21]. They are not auto-coherent either, and both their timing and their durations vary stochastically [47, 48]. All of which suggests that gamma oscillations are the result of filtered noise being amplified around resonance frequencies of the cortical circuit

[1] This is the reason behind the adjective "stabilized" in SSN.

[49]. In the language of dynamical systems, sustained oscillations occur above a so-called Hopf bifurcation. Indeed rate equation models of excitatory/inhibitory networks, such as the SSN, can generically undergo such a Hopf bifurcation when the feedback loop between the E and I neural populations is strong enough and inhibitory neurons are relatively slow [56]. But given the above evidence, we assume that (or rather seek parameter regimes in which) our SSN model sits below but close to a Hopf bifurcation, where it has a stable fixed point or steady state. In such a regime, the rate network can nevertheless exhibit damped or decaying oscillations when perturbed away from the stable fixed point, *e.g.* by input noise. These damped oscillations confer the network with a natural "resonance" frequency, which is determined by the network's fixed parameters and the background rates at the operating fixed point. This resonance manifests as a relatively broad peak in the power-spectrum of network activity fluctuations, even when the driving noise is temporally structureless and has no such peak in its power-spectrum.

To model such noise-driven oscillations using the SSN, we expand it, as in [57], to include noisy inputs, but with a faster source of noise to better study gamma oscillations than the fluctuations studied in [57]. As such, we model the external input as consisting of two terms $\mathbf{I}_t \to \mathbf{I}_{DC} + \boldsymbol{\eta}_t$, where $\mathbf{I}_{DC}$ represents the stimulus drive to the network (by a steady and time-independent stimulus), while $\boldsymbol{\eta}_t$ represents the noise in the network

$$\mathrm{T}\frac{d\mathbf{v}_t}{dt} = -\mathbf{v}_t + Wf(\mathbf{v}_t) + \mathbf{I}_{DC} + \boldsymbol{\eta}_t \tag{2.5}$$

We assume that the input noise, $\boldsymbol{\eta}_t$, has independent components which are identically distributed across different E and I units, and with zero mean. We

further assume $\boldsymbol{\eta}_t$ is temporally correlated as a pink noise with an exponentially decaying autocorrlation function with correlation times on the order of a few milliseconds (our main results are robust to changes in this parameter; in particular the noise can be assumed to be temporally white, which corresponds to the limit of zero correlation time). This noise could be attributed to several sources, *i.e.* feedforward noise from the thalamus, feedback from higher areas, or even spiking noise within our network, resulting from finite-size effects (the finite size of the implicitly-modeled neural sub-populations underlying the SSN's units).

2.2.3. SSN with multiple synaptic current types

With the aim of modeling gamma oscillations, we also extend the SSN model to properly account for different synaptic current types and their different receptor kinetics. The SSN model, as introduced in Sec. 2.2.1, has so far been used to model properties of steady-state network responses to constant inputs or stimuli [45, 46], or to model network dynamics in the presence of noise which results in relatively slow and shared variability across the network and the suppression of that variability by stimuli [57]. Gamma oscillations are by comparison faster dynamical phenomena, and as shown previously [36, 51, 58], networks of excitatory and inhibitory spiking neurons can generate oscillations with frequency in the gamma band when fast synaptic filtering by different synaptic receptor types is properly accounted for.

We thus extended the SSN model beyond its original form, presented in the previous subsections, in which each neuron is described by a single dynamical variable: either its firing rate or its total net input. In the extended model, by contrast, each neuron will have more than one dynamical variable, corresponding

to its inputs through different synaptic receptor types. Concretely, we will include the three main ionotropic synaptic receptors in the model: AMPA and NMDA receptors which mediate excitatory inputs, and GABA A receptors (we will shorten it to GABA) which mediate the inhibitory input.[2] To model the kinetics of different receptors, we will ignore the very fast rise-times of all receptor types (as their corresponding time-scales are considerably faster than the time-scales relevant to gamma oscillations), and only account for the receptor decay-times. This decay-time leads to a low-pass filtering of the corresponding input type. For the input of type α to some neuron ($\alpha \in \{\text{AMPA}, \text{NMDA}, \text{GABA}\}$), this low-pass filtering is described by the differential equation $\tau_\alpha \frac{dv_t^\alpha}{dt} + v_t^\alpha = u_t^\alpha$, where τ_α is the decay-time of the receptors of type α. Here u_t^α and v_t^α denote the input of type α at time t, before and after the low-pass filtering, respectively. If we arrange these inputs to different neurons into N-dimensional vectors, $\mathbf{v}_t^\alpha$ and $\mathbf{u}_t^\alpha$, we can rewrite the above equation in vector form as $\tau_\alpha \frac{d\mathbf{v}_t^\alpha}{dt} = -\mathbf{v}_t^\alpha + \mathbf{u}_t^\alpha$. Just as in the original SSN, $\mathbf{u}_t^\alpha$ is given by the sum of recurrent and feedforward (or external) input contributions. The only difference is that synaptic weights and external inputs should now be decomposed into the contributions of different synaptic receptor types. We thus have $\mathbf{u}_t^\alpha = W^\alpha \boldsymbol{r}_t + \mathbf{I}_t^\alpha$ where $\boldsymbol{r}_t$ is the N-dimensional vector of instantaneous firing rates, $\mathbf{I}_t^\alpha$ is the external input entering through receptors of type α, and W^α is a $N \times N$ matrix containing the contributions of receptors of type α to total synaptic weight matrix, such that $W = \sum_\alpha W^\alpha$. Finally, we still assume

[2]The three receptor types are named after their respective specific agonist molecules: α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA), N-methyl-D-aspartate (NMDA), and γ-aminobutyric acid (GABA). These ligand molecules are specific agonists of their corresponding receptor types, in the sense that they attach specifically to receptors of that type and activate them, which leads to the opening of associated ion-channels on the post-synaptic membrane. GABA is actually the endogenous inhibitory neurotransmitter of the central nervous system, while the endogenous excitatory neurotransmitter activating both NMDA and AMPA receptors is glutamate (not NMDA or AMPA).

that the firing rate is related instantaneously to the total net input, $\sum_\alpha v_t^\alpha$ by the rectified power law I/O nonlinearity of SSN, Eq. (2.1).[3] Thus $\boldsymbol{r}_t = f(\sum_\beta \mathbf{v}_t^\beta)$ where $f(\cdot)$ acts component-wise. The $3N$ dynamical variables $\mathbf{v}_t^\alpha$ are therefore governed by the equations

$$\tau_\alpha \frac{d\mathbf{v}_t^\alpha}{dt} = -\mathbf{v}_t^\alpha + W^\alpha \boldsymbol{r}_t + \mathbf{I}_t^\alpha \tag{2.6}$$

$$\boldsymbol{r}_t = f\big(\sum_\beta \mathbf{v}_t^\beta\big), \tag{2.7}$$

where $\alpha, \beta \in \{\text{AMPA}, \text{NMDA}, \text{GABA}\}$. As in the previous subsection, we assume that the external input is composed of a time-independent term, $\mathbf{I}_{DC}^\alpha$, which is the contribution of the stationary stimulus, as well as a stochastic input noise term, $\boldsymbol{\eta}_t^\alpha$. The dynamics of $\mathbf{v}_t^\alpha$ are thus governed by the following system of coupled stochastic differential equations

$$\tau_\alpha \frac{d\mathbf{v}_t^\alpha}{dt} = -\mathbf{v}_t^\alpha + W^\alpha f\big(\sum_\beta \mathbf{v}_t^\beta\big) + \mathbf{I}_{DC}^\alpha + \boldsymbol{\eta}_t^\alpha. \tag{2.8}$$

Since inhibition is only mediated by GABA, the elements of W^{AMPA} and W^{NMDA} on inhibitory columns (corresponding to pre-synaptic inhibitory neurons) are zero. The opposite is true for W^{GABA}: that matrix has zero excitatory columns. As for the break-up of excitatory weights between W^{AMPA} and W^{NMDA}, we made the simplifying assumption that in all excitatory connections the same fixed fraction, ρ_{NMDA}, of the total weight is mediated by NMDA, with the remainder (fraction $1 - \rho_{\text{NMDA}}$) mediated by AMPA.

[3] According to the analysis of [59], based on solutions of the Fokker-Planck equations governing voltage and firing statistics of spiking neurons, this instantaneous approximation is very accurate when the synaptic decay-times of the fast receptors are much faster than the RC membrane time-constant, as is true for AMPA and GABA decay-times.

As for the external inputs, we assumed that the stimulus input and input noise only enter through the AMPA receptors (as pertains the excitatory stimulus input, the mixture of NMDA or AMPA is inconsequential, as this input is not time-dependent; this assumption for the noise can be easily relaxed and will make no qualitative difference to our results on characteristics of LFP power spectra in the gamma band). We further assumed that input noise is independent and identically distributed (i.i.d.) across the units or neurons in the network, but temporally correlated as in so-called pink noise Eq. (2.43).

As we discuss below in Sec. 2.2.6, the scale of the gamma resonance frequency in this network is set by the timescale of the faster receptor types, AMPA and GABA. Here we used decay-times of 5 and 7 ms for AMPA and GABA receptors, respectively, compared to 100 ms for the slow NMDA.

2.2.4. Two-population model

We start by studying a simplified two-population model of V1 consisting of two units (or representative mean-field neurons): one excitatory and one inhibitory unit representing V1's excitatory and inhibitory neural sub-populations, respectively (Fig. 2.1A). This reduced model is appropriate for studying cases in which the spatial profile of the activity is not relevant, for example, when the grating stimulus is very large in size, and thus we can assume the entire V1 network is uniformly activated by the stimulus. We assume that both sub-populations have the same rectified power law I/O nonlinearity (Fig. 2.1A, inset) with the same parameters. These two units receive feedforward inputs, and make reciprocal synaptic connections with each other as well as themselves. We model increasing contrast by increasing the feedforward input, $\mathbf{I}_{DC}$ to our system.

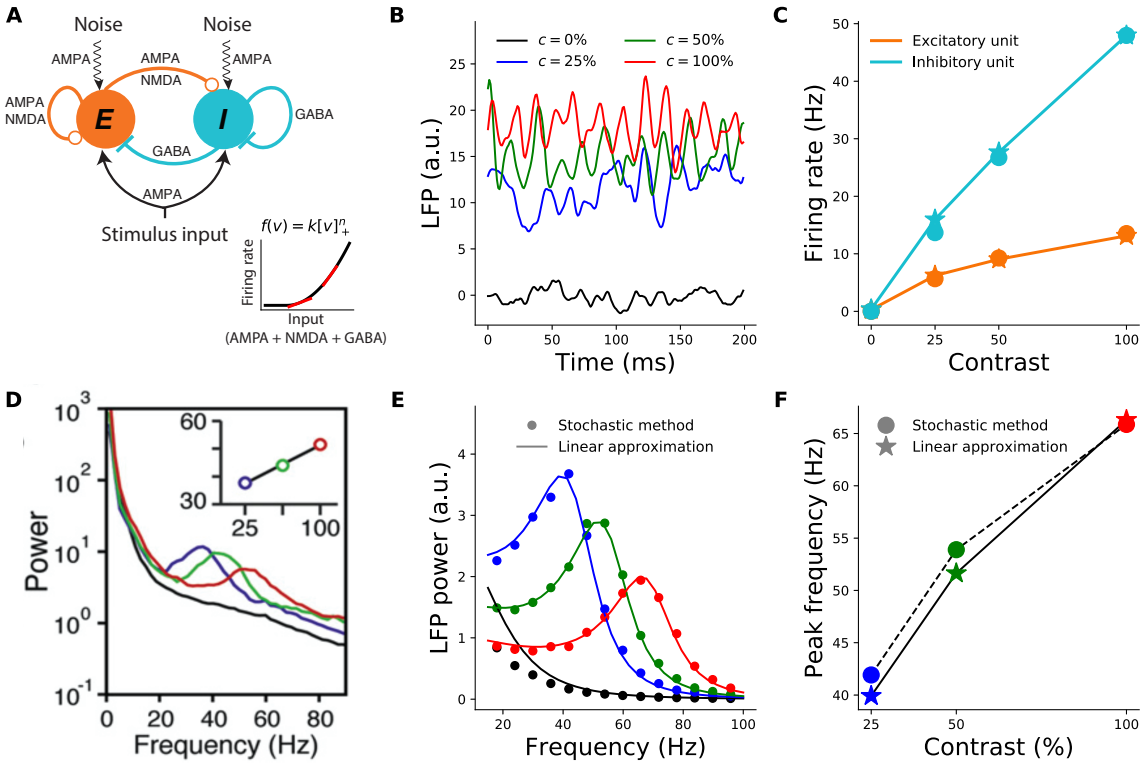

FIGURE 2.1. Contrast dependence of the gamma peak frequency in the 2-population model. **A:** Schematic of the 2-population Stabilized Supralinear Network (SSN). Excitatory (E) connections end in a circle; inhibitory (I) connections end in a line. Each unit represents a sub-population of V1 neurons of the corresponding E /I type. Both receive noisy input from the stimulus. **Inset:** The rectified power law Input/Output transfer function of SSN units (black). Red lines indicate the slope of the I/O function at particular locations. **B:** Local field potential (LFP) traces, modeled as total net input to the E unit, from the stochastic model simulations under four different stimulus contrasts (c): 0% (black) equivalent to no stimulus or spontaneous activity, 25% (blue), 50% (green), 100% (red). The same color scheme for stimulus contrast is used throughout the paper. **C:** Mean firing rates of the excitatory (orange) and inhibitory (cyan) units as a function of contrast, from the stochastic simulations (dots) and linearized approximation (stars). Note that the dots and stars closely overlap. **D:** Reproduction of figure 1I from [21] showing the average of experimentally measured LFP power-spectra in Macaque V1. The inset shows the dependence of gamma peak frequency on the contrast of the grating stimulus covering the recording site's receptive field. **E:** LFP power-spectra for $c = 0\%$, 25%, 50%, 100% (black, blue, green, and red curves, respectively) calculated from the noise-driven stochastic SSN simulations (dots), or using the linearized approximation (solid lines). **F:** Gamma peak frequency as a function of contrast, obtained from power-spectra calculated using stochastic simulations (dots and dashed line) or the linearized approximation (stars and solid line).

For the first results shown in Fig. 2.1, we simulated the stochastic Eq. (2.8) directly. Fig 2.1C (dots) shows the average firing rates found from the stochastic simulation. We modeled the LFP signal as being the net inputs to the E subpopulation. Figure 2.1B shows examples of raw simulated LFP traces for different stimulus contrasts. The LFP fluctuations exhibit oscillatory behavior, especially at high contrasts. These oscillations can be studied via their power-spectra which we calculated by taking the numerical Fourier transform of the simulated LFP traces (Fig. 2.1E, dots). As Fig. 2.1F shows, the peak frequencies (dots) of the simulated LFP power-spectra shift to higher frequencies with increasing contrast. Thus, a two-population SSN can reproduce the contrast dependence of the gamma peak frequency.

In Sec. 2.3.4 we develop a linearized approximation scheme for calculating the LFP power-spectra. Compared to simulations of the stochastic SSN model, the linear approximation allows us to numerically compute the LFP power-spectra much more efficiently. More importantly, the linear approximation allows for analytical approximations and insights, which (as we show, *e.g.*, in Sec. 2.2.6) elucidate the mechanism underlying the contrast dependence of the gamma peak.

To do so, we first find the steady state of our network, in the absence of noise, by solving $\mathbf{v}_*^\alpha = W^\alpha f(\sum_\beta \mathbf{v}_*^\beta) + \mathbf{I}_{DC}^\alpha$. This equation can be simplified if we define $\mathbf{v}_* \equiv \sum_\alpha \mathbf{v}_*^\alpha$ and $\mathbf{I}_{DC} \equiv \sum_\alpha \mathbf{I}_{DC}^\alpha$, and sum Eq. (2.9) over α. We then arrive at the same fixed equation for $\mathbf{v}_*$ as in the original SSN in Eqn. (2.2) [46]:

$$\mathbf{v}_* = W f(\mathbf{v}_*) + \mathbf{I}_{DC}. \tag{2.9}$$

Fig. 2.1C, shows the firing rates of E and I units at the steady-state solution Eq. (2.9) (the stars), for different values of the stimulus contrast, in excellent agreement with the mean rates obtained from the stochastic simulation (the dots).

After numerically solving Eq. (2.9), we then expand Eq. (2.8) to first order in the noise and noise-drive deviations around the fixed point to obtain

$$\tau_\alpha \frac{d\delta \mathbf{v}_t^\alpha}{dt} = -\delta \mathbf{v}_t^\alpha(t) + \tilde{W}^\alpha \sum_\beta \delta \mathbf{v}_t^\beta + \boldsymbol{\eta}_t^\alpha \tag{2.10}$$

where $\delta \mathbf{v}_t^\alpha$ denote the noise-driven fluctuations of net inputs around the fixed point, and we defined

$$\tilde{W}^\alpha \equiv W^\alpha \operatorname{diag}(f'(\mathbf{v}_*)) \,, \tag{2.11}$$

where $f'(\mathbf{v}_*)$ denotes the vector of slopes of the I/O functions of different neurons at the fixed point. These I/O slopes are commonly referred to as (neural) gains (see the red tangent lines in Fig. 2.1A inset). As we explain in the next subsection, the gains and their dependence on the fixed point rates (themselves dependent on the stimulus $\mathbf{I}_{DC}^\alpha$, via Eq. (2.9)) play a crucial role in the contrast dependence of gamma peak frequency in the SSN. Given the linear and time-homogeneous Eq. (2.10), we can use Fourier and linear systems analysis to obtain an analytical formula for the power-spectrum of LFP under the linear approximation (see Sec. 2.3.4).

As shown in Fig. 2.1E, the power-spectra obtained using the linear approximation agree very well with those estimated from the direct stochastic simulations of Eq. (2.8). Below, we thus calculate all power-spectra using the (computationally much faster) linear approximation, instead of stochastic simulations of Eq. (2.8).

As Fig. 2.1F (stars) shows, the peak frequency of our LFP power-spectra found using the linear approximation shift to higher frequencies with increasing contrast (and again closely match those found using the stochastic simulations). The 2-population SSN model thus captures the empirically observed contrast dependence of gamma peak frequency, as reported in from [21] (figure 1I therein, reproduced here as Fig. 2.1D).

The linearized approximation also allows for a qualitative understanding of the gamma peak contrast dependence, as we explain in a subsection below. We first show that the two-population SSN robustly reproduces the contrast dependence of gamma peak frequency across parameter choices.

2.2.5. Robustness of the two-population model

As the insight given in the previous subsection is quite general, only relying on the expansive and non-saturating nature of SSN's I/O nonlinearity, we expect that the model would be able to capture the gamma peaks's contrast dependence robustly, without any need for tuning of its connectivity or other parameters. To demonstrate this robustness, we simulated 1000 different instances of the 2-population network model with randomly drawn parameters. The sampled parameters were the connectivity strengths between the populations, i.e. $E \to E$, $E \to I$, $I \to E$, and $I \to I$, the strength of input to both the excitatory and inhibitory unit, and the NMDA fraction of excitatory synaptic weights (labelled as J_{EE}, J_{EI}, J_{EI}, J_{EI}, g_E, g_I, ρ_N respectively). The parameters were chosen uniformly from a biologically plausible range.

We sampled the parameters randomly and independently, but applied rejection criteria to ensure the existence of a stable fixed point. For instance, we

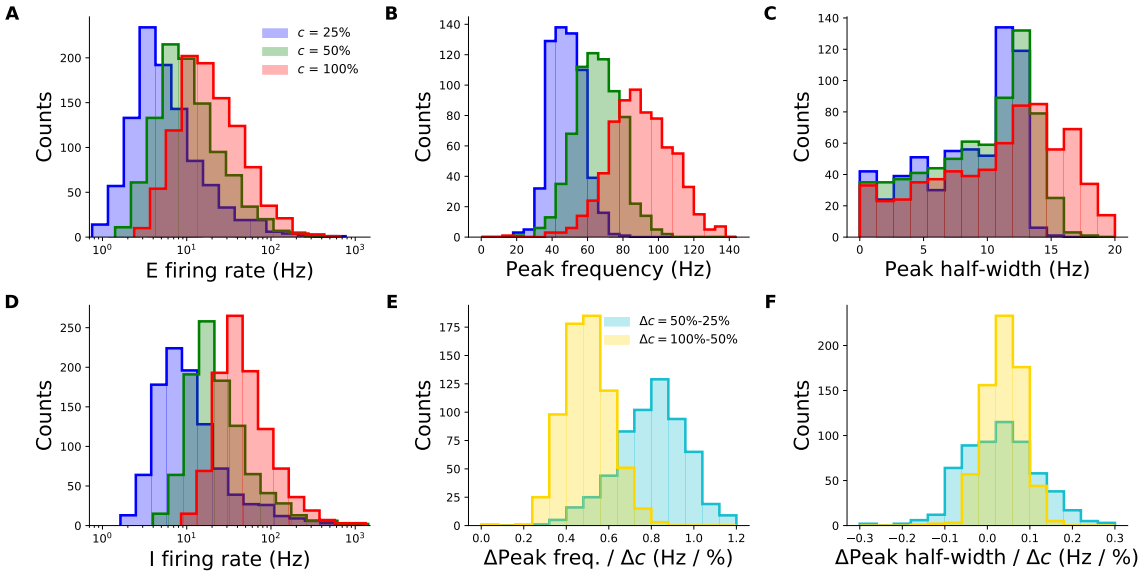

FIGURE 2.2. Robustness of the contrast-dependence of gamma peak frequency to network parameter variations. One thousand 2-population SSN's were simulated with randomly sampled parameters, and all histograms show simulation counts. **A:** Distributions of the excitatory unit's firing rate in response to 25%, 50%, and 100% contrast stimuli (blue, green, red), plotted on a logarithmic scale. 100% of networks shown across all contrasts. **B:** Distributions of the gamma peak frequencies at different stimulus contrasts. 60.9%, 50%, 50% of networks had gamma peaks at 25% (blue), 50% (green), and 100% (red) contrast respectively. **C:** Distributions of the gamma peak half-widths at different stimulus contrasts. **D:** Same as panel A, but for the inhibitory unit. 100% of networks shown across all contrasts. **E:** Distributions of the change in gamma peak-frequency normalized by the change in stimulus contrast, either 25% and 50% (cyan - 35.8% of networks) or 50% and 100% (yellow - 48.4%). **F:** Same as panel **E**, but for gamma peak half-width.

discarded parameter draws that would lead to unstable firing rates or to so-called supersaturation of excitatory firing rates based on criteria in [46] (see Sec. A.1 for details).

By and large, our simulations of randomly sampled 2-population SSN models produced fixed-point firing rates for the excitatory and inhibitory sub-populations that are within the biologically plausible window of values, across all contrasts. In Fig. 2.2A and D, we show the distribution of steady state firing rates for

contrasts at 25% (blue), 50% (green), and 100% (red). Furthermore, the two-population networks also typically produce peak frequencies that are within the gamma band (30 - 80 Hz) for all contrast conditions (Fig. 2.2B). The distributions also shift towards higher frequencies with increasing contrast, suggesting that the two-population SSN is indeed able to reproduce the contrast dependence of gamma peak frequency robustly. To demonstrate this more directly, we show the distributions of the changes in peak frequency normalized by the change in contrast in Fig. 2.2E. Note that none of the networks here simulated produced negative changes in peak frequency for positive changes in the stimulus contrast.

As a further corroboration of our model, we also studied how the half-width of the gamma peak changed with increasing contrast. While [20, 21] did not quantify changes in their gamma peak half-width with increasing contrast, their results suggest that no significant change in half-width was observed (Fig. 2.1D). Similarly, in our two-population network, changes in the gamma half-width were often relatively small, especially when normalized by the changes in contrast (Fig. 2.2F).

2.2.6. Intuition for the contrast dependence of gamma peak frequency

As we will show in this section, the SSN sheds light on the mechanism underlying the contrast dependence of the gamma peak, and further pins it to the expansive and non-saturating nature of the neuronal I/O transfer function.

Using the linearization around the fixed point, the power-spectra for the network can be found as Eq. (2.40) (making use of the Green's function, see Eqns. 2.39 and 2.41). The position of the resonance frequency in the power-spectrum is largely determined by the imaginary part of a corresponding

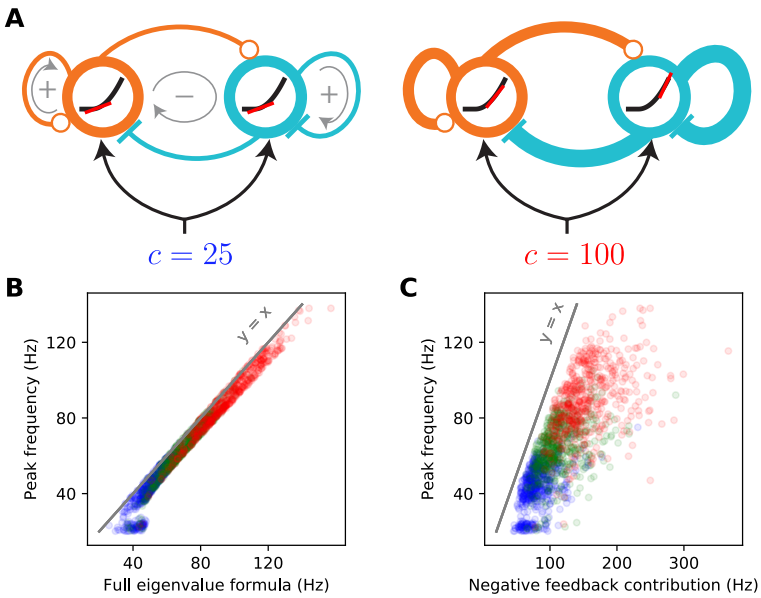

FIGURE 2.3. The supralinear nature of the neural transfer function can explain the contrast dependence of gamma frequency. **A:** Schematic diagrams of the 2-population SSN (see Fig. 2.1-A) receiving a low (left) or high (right) contrast stimulus. The thickness of connection lines represents the strength of the corresponding effective connection weight, which is the product of the anatomical weight and the input/ouput gain of the presynaptic neuron. The gain is the slope (red line) of the neural supralinear transfer function (black curve), shown inside the circles representing the E (orange) and I (cyan) units. A resonance frequency exists when the effective "negative feedback" (gray arrow enlcosing a minus sign) dominates the effective "positive feedback" (gray arrows enclosing positive signs), in the sense of the inequality Eq. (2.14). As the stimulus drive (c) increases (right panel), the neurons' firing rates in the network's operating point increase. As the transfer function is supralinear, this translates to a higher neural gains and stronger effective connections. When a resonance frequency already exists at the lower contrast, this strengthening of effective recurrent connections leads to an increase in the gamma peak frequency, approximately given by the imaginary part of the linearized SSN's complex eigenvalue, Eq. (2.13). **B:** The eigenvalue formula (2.15) provides an excellent approximation to the gamma peak frequency across sampled networks (see Fig. 2.2) and contrasts; correlation coefficient = 0.98 ($p < 10^{-6}$), for all data points combined across 25% (blue), 50% (green), and 100% (red) contrasts. **C:** The negative feedback loop contribution to the resonance frequency (Eq. (2.15) with the second term under the square root neglected) overestimates gamma peak frequency but is positively and significantly correlated with it; correlation coefficient = 0.75 ($p < 10^{-6}$).

eigenvalue of the Jacobian matrix (which also determines the singularities in the Green's function for Eq. (2.10)). The Jacobian matrix, $\mathcal{J}$ can be obtained by rewriting Eq. (2.10) in the standard form

$$\frac{d\mathbf{x}_t}{dt} = \mathcal{J}\mathbf{x}_t \tag{2.12}$$

where $\mathbf{x}_t$ is a $3N$-dimensional vector formed by concatenating the three $\delta\mathbf{v}^\alpha$ vectors together. In general, the Jacobian is thus a $3N \times 3N$ matrix; it is 6×6 for the 2-population model. (In Sec. 2.3.4, below, we give the explicit form of the Jacobian; see Eq. (2.37)). Previously, the eigenvalues of the Jacobian for an E-I firing rate network (similar to Eq. (2.2) for the case $N_E = N_I = 1$) were analyzed by [35], and conditions for emergence of (damped or sustained) oscillations were found. Such a model, however, does not account for different synaptic receptor types. In Sec. 2.3.6, we prove that in the absence of the NMDA receptors, two of the eigenvalues of the resulting 4×4 Jacobian correspond to the eigenvalues of a standard E-I rate model whose E and I neural time-constants are given, respectively, by the decay times of AMPA and GABA receptors (the other two eigenvalues are always real and thus cannot lead to oscillations). They are given by:

$$\begin{aligned} 2\lambda_\pm = &\gamma_E(\tilde{W}_{EE} - 1) - \gamma_I(\tilde{W}_{II} + 1) \\ &\pm \sqrt{\Big[\gamma_E(\tilde{W}_{EE} - 1) + \gamma_I(\tilde{W}_{II} + 1)\Big]^2 - 4\gamma_E\gamma_I\tilde{W}_{EI}\tilde{W}_{IE}} \end{aligned} \tag{2.13}$$

where $\gamma_E = \frac{1}{\tau_{\text{AMPA}}}$ and $\gamma_I = \frac{1}{\tau_{\text{GABA}}}$. Here, $a, b \in \{E, I\}$ denote the E/I units, and we defined $\tilde{W}_{ab} = W_{ab}f'(v_*^a)$, where $W_{ab} = \sum_{\alpha\in\{\text{AMPA,GABA}\}} W_{ab}^\alpha$ is the total synaptic weight from unit b to unit a, and $f'(v_*^a)$ is the gain (the slope of

the I/O function) of unit a at the background fixed point set by the stimulus. As we show further in Sec. 2.3.6, even in the presence of NMDA receptors (which are slow compared to the time-scales of gamma oscillations), Eq. (2.13) is still approximately correct. The caveat is that the excitatory effective weights in the formula only include the AMPA contribution (*i.e.*, $\tilde{W}_{ab}$ in Eq. (2.13) are still given by $\tilde{W}_{ab} \equiv \sum_{\alpha \in \{\text{AMPA,GABA}\}} \tilde{W}^{\alpha}_{ab}$ with NMDA not contributing). Indeed our simulations show that this approximation is very good (Fig. 2.3B)

We refer to $\tilde{W}_{ab}$ as the effective synaptic connection weights (or to their matrix as the effective connectivity matrix). Unlike the raw synaptic weights, the effective connection weights are thus modulated by the neural gains, and thereby the firing rates in the background fixed points and thus by the stimulus input which sets those. Note that the effective connection weights, $\tilde{W}_{ab}$ are dimensionless quantities: they give the coupling strengths between the net-input fluctuations of units a and b.

The condition for the emergence of (damped or sustained) oscillations is that the above eigenvalues are complex (in which case the two eigenvalues are complex conjugates). This happens when the expression under the radical in Eq. (2.13) is negative, or

$$4\gamma_E\gamma_I\tilde{W}_{EI}\tilde{W}_{IE} > \Big[\gamma_E(\tilde{W}_{EE} - 1) + \gamma_I(\tilde{W}_{II} + 1)\Big]^2 . \tag{2.14}$$

Qualitatively, the left hand side of the above inequality is a measure of the strength of the effective negative feedback between the E and I sub-populations, while the right hand size is a measure of the positive feedback in the network (self-excitation of E population, and self-inhibition of I population), see Fig. 2.3A, gray loops. Oscillations thus emerge when the negative feedback loop between E and I is sufficiently strong, in the precise sense of Eq. (2.14).

During spontaneous activity (when the external input is zero or very weak), the rates of both E and I populations are very small. This means that the spontaneous activity operating point sits near the rectification of the neuronal I/O transfer functions where the neural gains are very low. Thus in the spontaneous activity state, the dimensionless effective connections are relatively small. In the limit of $\tilde{W}_{ab} \to 0$, the left hand side of Eq. (2.14) goes to zero, while its right side goes to $(\gamma_I - \gamma_E)^2$ which is generically positive; hence the inequality is not satisfied. This shows that the spontaneous activity state generically does not exhibit oscillations, in agreement with lack of empirical observation of gamma oscillation during spontaneous activity.

On the other hand, when the condition Eq. (2.14) holds, the frequency of the eigenmode oscillations are given by the imaginary part of the eigenvalues, *i.e.*, by

$$\text{resonance frequency} = \frac{1}{2\pi}\sqrt{\gamma_E\gamma_I\tilde{W}_{EI}\tilde{W}_{IE} - \left[\gamma_E(\tilde{W}_{EE}-1)/2 + \gamma_I(\tilde{W}_{II}+1)/2\right]^2} \tag{2.15}$$

(the division by 2π is because the eigenvalue imaginary parts give the angular frequency). As we will discuss further below the resonance frequency (or approximately the gamma peak frequency [Fig. 2.3B]) thus depends on the effective connections weights and is thus modulated by the neural gains. However, the scale (or order of magnitude) of this frequency is set by γ_E and γ_I, *i.e.*, by the decay times of AMPA and GABA. To see this, recall that the effective connection weights are dimensionless and we therefore expect them to be generically order one. Thus to determine the scale of the resonance frequency, we set the recurrent weights to one, and moreover ignore the contribution of the "positive feedback" term under the square root in Eq. (2.15). This yields: resonance frequency

$\sim \sqrt{\gamma_E \gamma_I}/(2\pi)$. For $\tau_{AMPA} \sim \tau_{GABA} \sim 4-6$ ms, this yields a resonance frequency on the order of 30-40 Hz, matching the observed gamma frequency for intermediate stimulus contrasts.

We now use Eq. (2.15) to provide insight into the contrast dependence of the gamma peak frequency, see Fig. 2.3. In our two-population network, as contrasts increase the fixed-point firing rates increase (Fig. 2.1C). Because the SSN I/O transfer function is non-saturating and supralinear, as the rates increase the gains (*i.e.*, the slope of the I/O transfer functions) of the E and I cells are also guaranteed to increase (Fig. 2.3A). The increase in the gains leads in turn to the strengthening of the effective connection weights. When Eq. (2.14) is satisfied, a rough approximation (Fig. 2.3C) is obtained by ignoring the positive feedback contribution (*i.e.*, the second term under the square root in Eq. (2.13)). With only the negative contribution retained, it is clear that an increase in neural gains leads to an increase in the resonance frequency (the precise conditions for this to occur are given in Sec. 2.3.7). Thus as contrasts increase, we expect the location of the gamma peak in the LFP power-spectrum to increase as well.

2.2.7. Retinotopic SSN

We next investigated whether the SSN can account for the local contrast dependence of gamma peak frequency by expanding our network from two units representing global E and I populations to many units that are retinotopically organized. We thus model the cortex as a two-dimensional grid which has an E and I sub-population at each grid location corresponding to a cortical column (Fig. 2.4A). In the retinotopic SSN, the stimulus input can vary across the network: each column can receive a different input proportional to the contrast

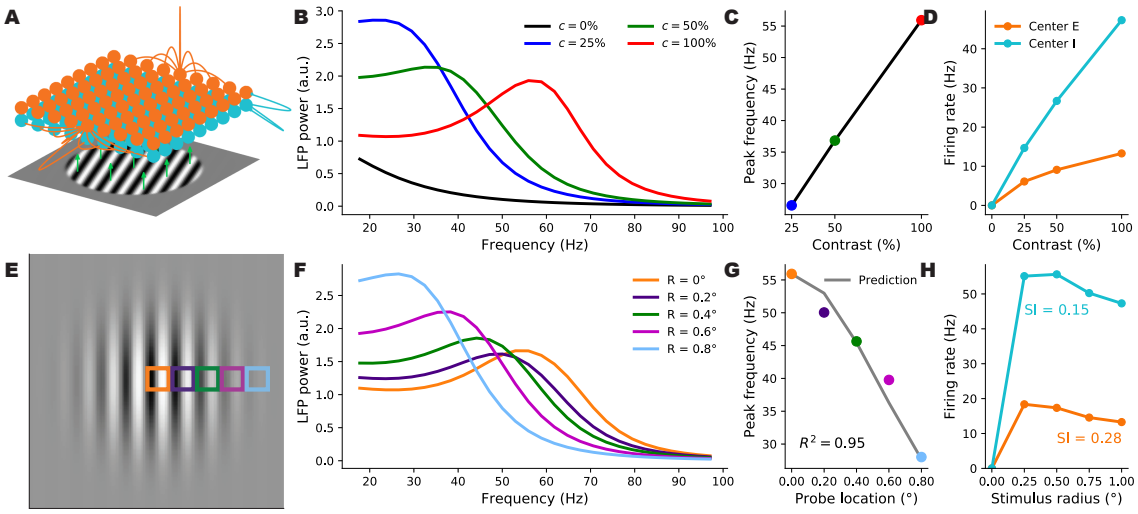

FIGURE 2.4. A retinotopically-structured SSN model captures the local contrast dependence of gamma peak-frequency, as well as the surround suppression of excitatory and inhibitory firing rates. **A:** Schematic of the retinotopic grid and inputs. At each location in retinotopic space, there exists an excitatory (orange) and inhibitory (cyan) sub-population which receives input (green arrows) from the stimulus (grating). Conenctions from excitatory units are shown in orange, while connections from inhibitory units are shown in cyan. **B:** LFP power-spectra of the excitatory center sub-population during different flat grating stimuli contrast, 0% (black), 25% (blue), 50% (green), and 100%. **C:** Peak frequency of the power-spectra gamma band under various flat grating contrast conditions. **D:** Fixed-point firing rates for the excitatory (orange) and inhibitory (cyan) sub-populations at the center of the retinotopic grid for increasing stimulus contrast. **E:** Representation of the Gabor input to the full retinotopic grid of units. Each square is colored according to the retinotopic receptive field of the sub-populations represented in the grid, 0^o displaced from the center (orange), 0.2^o (navy), 0.4^o (green), 0.6^o (magenta), and 0.8^o (sky blue). **F:** Power-spectra of the excitatory sub-population colored according to distance displaced from the center of the grid. **G:** Gamma peak frequency of the power-spectra at increasing distance from the center of the grid. The gray curve shows a prediction of the displaced units' peak frequency given their local contrast based on the shift in gamma peak frequency of the center E sub-population as the contrast of a flat grating changes. **H:** Suppression curve of the firing rates. The firing rates of the center excitatory and inhibitory subpopulations vary non-monotonically as a flat grating stimulus is grown in size, and their suppression indices are 0.27 for the center E sub-population (orange), and 0.21 for the center I sub-population (cyan).

within its receptive field. In the current study we used flat grating stimuli of various sizes and contrasts (the stimulus in Fig. 2.4A), or a Gabor stimulus similar to the one used in [21] (Fig. 2.4E). Flat grating stimuli are disks of oriented light and dark bands, within which the contrast is uniform and sharply falls to zero at the disk edge. The Gabor stimulus is a grating with a contrast profile that decays smoothly with increasing distance from the stimulus center, according to a Gaussian profile (see Sec. 2.3 for details). Gamma oscillations show weak dependence on stimulus orientation [20], possibly due to the averaging of LFP over an area larger than the size of orientation minicolumns. To keep our model parsimonious and computationally more tractable, we thus chose the size of our cortical columns to be roughly half the hypercolumn size in Macaque, and neglected the orientation map structure and the dependence of external inputs and horizontal connections on preferred orientation.

We wish to study the trade-off in the model between capturing surround suppression of firing rates and capturing the local dependence of gamma peak frequency, and whether parameter choices exist for which the model can capture both of these effects.

One extreme of connectivity structure is one where the local intra-columnar connectivity is not emphasized. In such a case, the connectivity between all units smoothly decays with increasing distance between the units. Such a network though, often has a gamma peak frequency shared across the grid (see Sec. 2.2.9 for more details). On the other hand, the other extreme of connectivity structure, weak long-range connectivity, should produce the local contrast dependence of gamma peak frequencies, but fail to produce surround suppression of firing rates. In the limit of weak long-range horizontal connectivity, different columns in our

grid will approximate independent two-population networks, and as we showed above (Figs. 2.1–2.2), the two-population SSN robustly produces the contrast dependence of gamma peak frequency. Therefore in this regime we expect the network to trivially capture the local contrast dependence of gamma (with each column's gamma peak solely controlled by its direct feedforward stimulus input), but struggle to produce surround suppression in firing rates [44].

To study the balance between the two regimes, we include parameters which govern the spatial extent of connections within our network. Specifically, we include a term which linearly interpolates between the two extremes of connectivity structure $\lambda_{a,b}$, where $\lambda_{a,b} = 0$ signifies a strong horizontal connectivity regime while $\lambda_{a,b} = 1$ signifies purely local (intra-columnar) connectivity. Here we adopt the notation that a unit or neural sub-population can be specified by its type a or b, which are either E or I, and location $\mathbf{x}$ where $\|\mathbf{x}\|$ is the distance from the retinotopic center. We denote synaptic connections in the form $post, pre$, e.g. $\lambda_{a,b}$ controls the locality of connections from units of type b to units of type a, and that $W_{\mathbf{x},\mathbf{y}|a,b}$ represents the connection strength from the unit of type b to the unit of type a at locations $\mathbf{y}$ and $\mathbf{x}$ respectively.

We make a simplifying assumption consistent with the principles underlying the SSN and assume that connections from inhibitory sub-populations to either excitatory or inhibitory sub-populations are extremely local (2.4A, cyan lines), i.e. that they "synapse" only onto themselves and the associated E sub-population at their retinotopic location. To capture the trade-off between highly local connections among the excitatory sub-populations, which should mediate the local contrast dependence and surround suppression, we model the scale of connection between E to E and E to I sub-populations as a linear interpolation between an

exponential decay and purely local connections.

$$W_{\mathbf{x},\mathbf{y}|a,b} = \lambda_{a,E}\delta_{\mathbf{x},\mathbf{y}} + (1 - \lambda_{a,E})e^{\frac{-|\mathbf{x}-\mathbf{y}|}{\sigma_{a,E}}} \tag{2.16}$$

where δ is the Kronecker δ, and $\lambda_{a,E}$ interpolates between purely local ($\lambda_{a,E} = 1$) and purely exponentially decaying connections ($\lambda_{a,E} = 0$). Further, we allow $\lambda_{a,E}$ and $\sigma_{a,E}$ to vary depending on whether the unit at $\mathbf{x}$ is of type $a = E$ or $a = I$. Recall that these are connections from E sub-populations.

In (Fig. 2.4A, orange lines) we show an example of how the strength of excitatory connections scale across the retinotopic grid. We show E to E connections on the far edge of the grid, while E to I connections are shown on the near edge of the grid. Note the sharp peaks for self connection on the far side, and the sharp local connection with the associated retinotopic inhibitory sub-population. Additionally we show the highly local connections from I to E or I to I on the right corner of the grid (Fig. 2.4A, cyan).

We found the power-spectra in our retinotopic SSN using the linear approximation to the network's dynamics, and found the fixed-point firing rates via Eq. (2.9). We "present" large flat gratings with varying contrasts to our network, which ensures that the firing rates converge, and verify that the firing rates of the center sub-populations increase with increasing contrast (Fig. 2.4D). Based on our previously built intuition, increasing firing rates means that the center E sub-population's power-spectra should demonstrate the contrast dependence of gamma peak frequency, and it does indeed (Fig. 2.4B-C).

To study surround suppression, we "present" flat gratings at max contrast ($c = 100\%$) and vary their spatial size. The retinotopic SSN has steady-state firing rates that behave non-monotonically, and suppress with increasing stimulus size

(Fig. 2.4H). The network's center E sub-population has a suppression index (SI) of 0.28 consistent with biologically reported values of suppression indices [34], where the suppression index is defined as $SI = 1 - \frac{r_\infty}{max(r)}$ with r_∞ being the firing rate at largest contrast.

To study the local contrast dependence, the network was "presented" with a Gabor stimulus, which has spatially decaying contrast. The receptive field of the center E sub-population is represented by the orange square on the Gabor stimulus (Fig. 2.4E). The RFs of studied non-center retinotopic E sub-populations are also shown with a square colored according to the of their RF displacement from the center of the grid (navy: 0.2^o, green: 0.4^o, magenta: 0.6^o, cyan: 0.8^o) (Fig. 2.4E). The power-spectra for all E sub-populations were calculated again using the linear approximation, and the peak frequency from the power-spectra were extracted (Fig. 2.4F-G, colored dots).

Reference [21] found that the LFP power's peak frequency at each retinotopic location was well predicted by a linear fit to the contrast dependence of the center E sub-population. Using a similar linear fit (Fig. 2.4C, black line), the prediction of what each retinotopic sub-population's peak frequency should be based on the local contrast in their RFs is shown as a gray line in Fig. 2.4. Our linear interpolation excellently predicts the observed power spectra gamma peaks found in Fig. 2.4F. Here we use the $R^2 = 1 - \frac{SSE}{var}$ metric to classify how well the gamma peak frequency in various RFs are predicted by the linear fit, where SSE means the sum of squared errors (*e.g.*the sum of the square of the difference between the linear fit and the found gamma peak frequencies). The denominator, var, refers to the variance of our simulated data (*e.g.*the variance of our gamma peak frequencies). Note that R^2 is bounded above by 1 when the linear prediction

perfectly matches the data, but unbounded below. In the example shown in Fig. 2.4F, we find an $R^2 = 0.95$, indicating a very good fit.

Thus the retinotopic SSN can produce both the surround suppression of firing rates as well as the local contrast dependence of gamma peak frequency.

2.2.8. Robustness of the Retinotopic SSN

To show that the parameters for the retinotopic SSN were not fine-tuned, we simulated 1000 networks with perturbed parameters. The perturbations were drawn from uniform distributions between -10% to 10% of the parameters previous value.

These networks produced biologically plausible center excitatory and inhibitory firing rates for changing contrasts of the large flat grating stimuli (Fig. 2.5A and D). The distributions correspond to contrasts at $25\%, 50\%$, and 100% (blue, green, red). The perturbed networks also robustly produce peak frequencies in the gamma band (Fig. 2.5B), and have peak frequencies that generally change positively with increasing contrast (Fig. 2.5E).

If these networks also display the local contrast dependence, then the power-spectra of the spatially displaced E sub-populations should be well predicted by a linear interpolation of the center E sub-population's contrast dependence on large flat gratings. We made a similar fit as the gray line in Fig. 2.4G for each perturbed network, and found how well it predicted the observed peak frequencies at the displaced spatial locations. The vast majority of our networks which produced gamma peaks show a predictable ($R^2 > 0.8$) local contrast dependence of gamma peak frequency (Fig. 2.5F, green bar). Relatively fewer networks which produced gamma peaks were not predictable by the contrast dependence of their center E

sub-population (Fig. 2.5F, red bar). The inset shows the full distribution of R^2 for all the perturbed networks. All of which shows that the SSN used in Fig. 2.4 is not fine-tuned, but robust to pertubations of $\pm 10\%$.

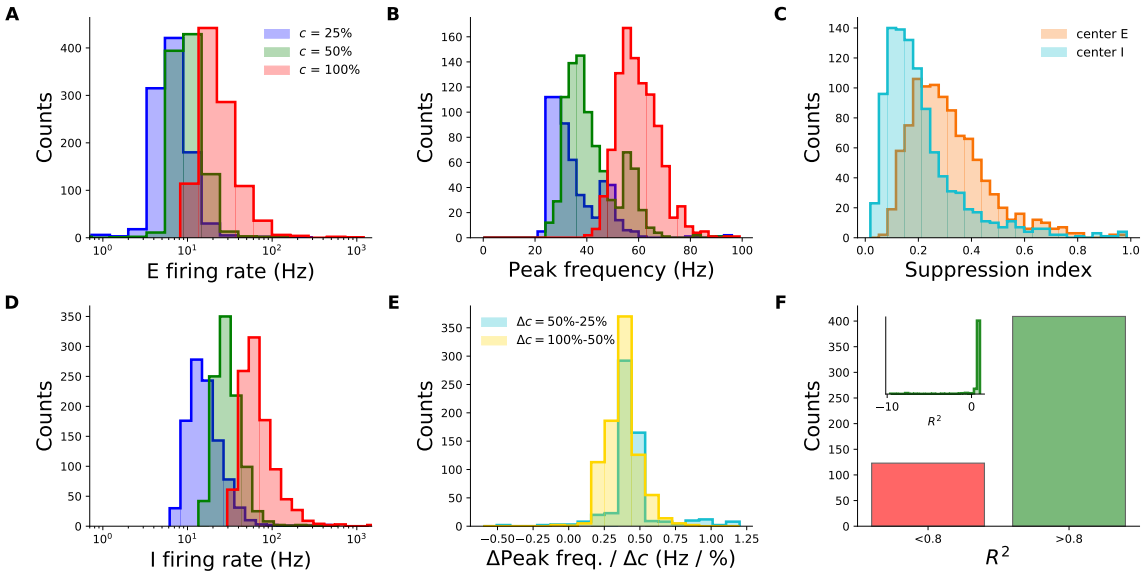

FIGURE 2.5. Robustness of the retinotopic SSN to parameter perturbations. We simulated 1000 different retinotopic SSN networks with parameters randomly and independently perturbed up to 10% of their value in the network of Fig. 2.4. All histograms show counts of these sampled networks. For gamma-related histograms only the fraction of sampled networks exhibiting a gamma peak in the relevant stimulus conditions are included. **A & D:** Distributions of the center E and I unit's firing rate (logarithmic scale) in response to uniform gratings with contrasts 25% (blue), 50 % (green), and 100% (red). **B:** Distributions of gamma peak frequency at stimulus center, for different contrasts of the largest uniform grating (same color code as in **A**). (The histograms for 25%, 50%, and 100% contrasts include 57.5%, 88%, and 99.8% of sampled networks, respectively.) **C:** Distributions of the suppression index for the center E (orange) and I (cyan) sub-populations. **E:** Distributions of the change in gamma peak-frequency normalized by the change in stimulus contrast, changing from 25% to 50% (cyan) or from 50% to 100% (yellow). The histograms include 57.5% and 88% of the sampled networks, respectively. The vast majority of perturbed networks produce positive changes in peak frequency as the stimulus contrast grows. **F:** The distribution of the goodness-of-fit, as measured by R^2, of the local prediction for gamma peak frequency (the gray line in Fig. 2.4G, but computed independently for different sampled networks). The goodness-of-fit, R^2, of the gray line in (Fig. 2.4G) computed for many different simulations. The contrast dependence of gamma peak frequency of the excitatory center sub-population robustly predicts ($R^2 > 0.8$) the peak frequency found in the spatially extended excitatory sub-populations when presenting a Gabor stimulus to the network (green bar). There were fewer where the contrast contrast dependence of the center E sub-population did not predict ($R^2 > 0.8$) the local contrast dependence (red bar). **Inset:** The full distribution of R^2. 57.3% of networks shown.

2.2.9. Retinotopic SSN with a smooth fall-off of excitatory horizontal connectivity

Here we present results from a retinotopic SSN without any excess in local connectivity, *e.g.* $\lambda_{ab} = 0$.

An SSN with a connectivity structure similar to [45] can robustly produce surround suppression of firing rates, but fails to produce the local contrast dependence of gamma peak frequency (Figs. 2.6). For networks with a smooth fall-off of long-range horizontal connectivity, it is possible to produce gamma peaks across the across retinotopic locations (Fig. 2.6A and C). However, the gamma peak frequency across retinotopic locations is shared (Fig 2.6D). As such, the contrast dependence of the center E sub-population (Fig 2.6A) fails to predict the gamma peak frequency of retinotopically displaced units when a Gabor stimulus is presented (Fig 2.6D, gray line). For Fig. 2.6D, the $R^2 \ll 1$, and the prediction fails. The firing rates of the center E sub-population were, however, suppressed by increasing stimulus size (Fig 2.6B).

To test how representative these results were, we simulated 1000 networks with non-local connectivity structures. The vast majority of networks simulated with this connectivity structure showed some non-zero suppression index $SI = 1 - \frac{r_\infty}{max(r)}$ (Fig 2.7B). The power-spectra of these networks, though, often failed to produce gamma peaks (Fig 2.7A) or had gamma peak frequencies which decreased with increasing contrast . These networks thereby failed to reproduce the contrast dependence of gamma peak frequency (Fig 2.7D). Moreover, the contrast dependence of the center E sub-population failed to predict the peak frequency at spatially extended RFs when the network was presented with a Gabor stimulus (Fig 2.7C and E).

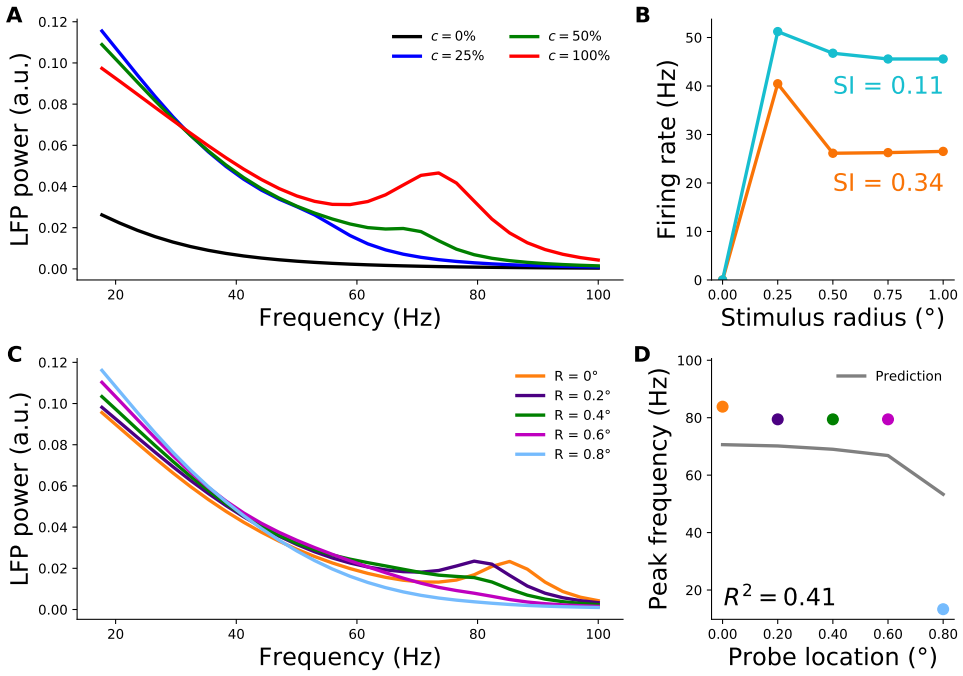

FIGURE 2.6. Retinotopic SSN without an excess in intra-columnar E-connections fails to capture the local contrast dependence of gamma peak frequency. **A:** LFP power-spectra at the center cortical column, for different contrasts of the uniform grating stimulus: 0% (black), 25% (blue), 50% (green), 100% (red). **B:** Firing rate size-tuning curves of the center E (orange) and I (cyan) sub-populations, based on responses to 100% contrast gratings. **C:** LFP power-spectra in the Gabor stimulus condition at different retinotopic distances from the stimulus center: 0^o (orange), 0.2^o (navy), 0.4^o (green), 0.6^o (magenta), and 0.8^o (sky blue). **D:** Gamma peak frequency as a function of distance from Gabor stimulus center, for power-spectra in **C**. The gray curve shows the local prediction of the peak frequency given the local contrast in the Gabor stimulus, and the relationship (derived from the power-spectra in **A**) between peak frequency at the center of the uniform grating and its contrast.

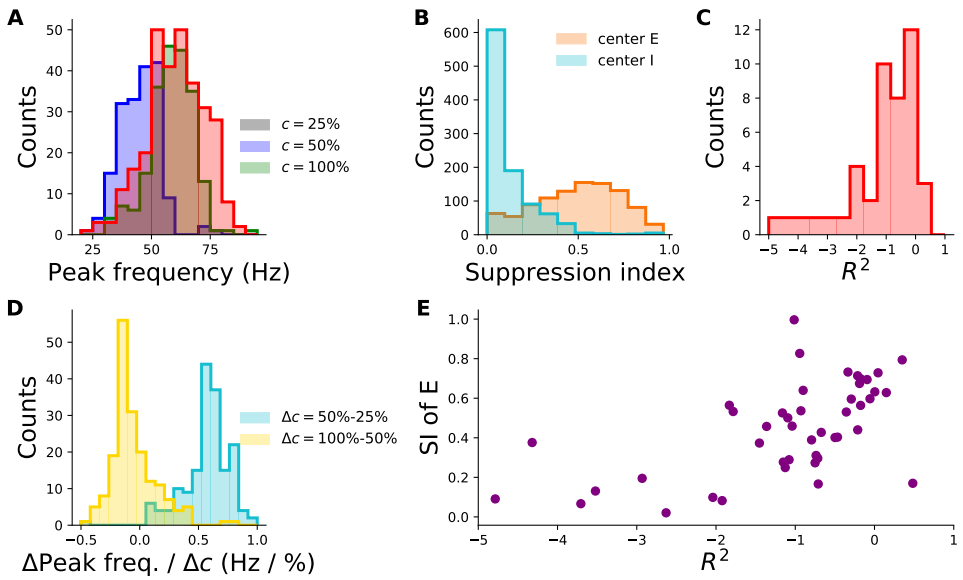

FIGURE 2.7. Across many parameter choices the retinotopic SSN without an excess in intra-columnar E-connections fails to capture the local contrast dependence of gamma peak frequency. We simulated a thousand such retinotopic SSN's with randomly sampled parameter sets. The histogram in panel includes **B** all sampled networks, while histograms in **A** and **C-E** contain only the minority of sampled networks which exhibited a gamma peak in the relevant stimulus conditions. **A:** Distributions of gamma peak frequency at stimulus center, for different uniform grating contrasts at 25% (blue), 50 % (green), and 100% (red). **B:** Distributions of the suppression index for the center E (orange) and I (cyan) sub-populations, extracted from size-tuning curves. **C:** Distributions of the change in gamma peak-frequency normalized by the change in stimulus contrast, from 25% to 50% (cyan) or from 50% to 100% (yellow); the respective histograms include 17.3% and 18.2% of sampled networks. **D:** The distribution of the goodness-of-fit, as measured by R^2, of the local prediction for gamma peak frequency (the gray line in Fig. 2.6D, but computed independently for different sampled networks). **E:** The joint distribution of the suppression index of the center E population and the local gamma contrast dependence (as measured by the R^2 histogrammed in **H**). 4.5% of sampled networks are shown. No sampled network achieved an R^2 above 0.5.

2.3. Methods

2.3.1. Synaptic SSN

To better model synaptic filtering, we must further expand the SSN to include various receptor types within our sub-populations. We consider inhibitory synapses to be mediated by GABA receptors, but we include both AMPA and NMDA receptor types for excitatory units. In addition to better modelling biology, splitting the excitatory sub-population into two receptor types also helps us dynamically.

By including NMDA as well as AMPA and GABA, the resonant frequencies of the network shifts to nearer the gamma band (30-80 Hz), due to NMDA weakening the effective connectivity of the AMPA receptors. Importantly, including these different receptor types does not change the numerical value of the fixed-point or steady-state solution of our network. However, the stability of that fixed point may be affected, and so we ensure our expanded networks were stable by studying the real part of the eigenvalues of the network's Jacobian.

In addition to weakening the effective connectivity of AMPA, including NMDA also sufficiently slows down excitation compared to inhibition. *E-I* networks fail to produce oscillations when excitation is faster than inhibition. As AMPA receptors have a faster time scale than GABA receptors, if we only considered those synaptic types our networks would fail to oscillate. By choosing some percentage of our excitatory sub-populations' synapses to express NMDA instead of AMPA, the time-scale of excitation becomes effectively slower than inhibtion, allowing our network to produce damped oscillations.

Mathematically, the SSN can be expanded to include multiple synaptic types in the following way

$$\tau_\alpha \frac{d\mathbf{v}_t^\alpha}{dt} = -\mathbf{v}_t^\alpha + W^\alpha f\big(\sum_\beta \mathbf{v}_t^\beta\big) + \mathbf{I}_{DC}^\alpha + \boldsymbol{\eta}_t^\alpha \tag{2.17}$$

where $\alpha, \beta \in [A, G, N]$ for AMPA, GABA, NMDA respectively. Note that $\mathbf{v}^\alpha$ is an N dimensional vector describing the input voltage of receptor type α to every unit in our simulation.

The SSN expands from an N-dimensional rate equation to a $3N$-dimensional rate equation, and importantly we decompose the total connectivity matrix W into a connectivity W^α for the three receptor types:

$$W^A = \begin{pmatrix} W_{EE}^A & 0 \\ W_{IE}^A & 0 \end{pmatrix} \tag{2.18}$$

$$W^N = \begin{pmatrix} W_{EE}^N & 0 \\ W_{IE}^N & 0 \end{pmatrix} \tag{2.19}$$

$$W^G = \begin{pmatrix} 0 & -W_{EI}^G \\ 0 & -W_{II}^G \end{pmatrix} \tag{2.20}$$

2.3.2. Modelling LFP

We assume that local field potential (LFP) recordings predominantly measure net input fluctuations to excitatory (pyramidal) neurons, and thus in our model LFP we study the net input to our excitatory sub-populations. In principle

the instantaneous LFP signal at location $\mathbf{x}$ is

$$LFP_{\mathbf{x}}(t) \propto \mathbf{e}_E \cdot \sum_{\beta} \mathbf{v}_t^{\beta} = \sum_{i \in E} \sum_{\beta} \mathbf{v}_t^{\beta} \tag{2.21}$$

where $\mathbf{e}_E$ is the projection operator to the excitatory sub-populations, i.e. $e_{E,i} = 1$ if $i \in N_E$, else $e_{E,i} = 0$. Note that $\sum_{\beta} \mathbf{v}_t^{\beta}$ is the net input voltage to all N of our units.

If we further assume that the spatial extent of the recording probe is within a cortical column, meaning that within our network the recording is small relative to the size of our grid, we can impose that $\mathbf{e}_{E,i \in N_E} = \delta_{i,\mathbf{x}}$. Thus the LFP at $\mathbf{x}$ when the recording is small relative to the grid is given by

$$LFP_{\mathbf{x}}(t) \propto \sum_{\beta} \mathbf{v}_{\mathbf{x},t}^{\beta} \tag{2.22}$$

2.3.3. Two-population and retinotopic model

For the two-population model, we model V1 as two units: one for the E sub-population, and one for the I sub-population. The LFP in this case is simply the net input to the excitatory sub-population. We assume that each sub-population makes connections with itself and the other. Thus there are 4 connection types in our network: $E \to E$, $E \to I$, $I \to E$, and $I \to I$. Further, we impose that some fraction of E synapses express NMDA as their receptor ($\rho_N = \frac{\text{\# of NMDA}}{\text{total connections}}$).

Therefore,

$$W^A = \begin{pmatrix} (1-\rho_N)J_{EE} & 0 \\ (1-\rho_N)J_{IE} & 0 \end{pmatrix}$$
$$W^N = \begin{pmatrix} \rho_N J_{EE} & 0 \\ \rho_N J_{IE} & 0 \end{pmatrix} \tag{2.23}$$
$$W^G = \begin{pmatrix} 0 & -J_{EI} \\ 0 & -J_{II} \end{pmatrix}$$

where $J_{a,b}$ is the strength of connection from sub-population of type b to type a.

The DC input is modelled as feedforward inputs from the thalamus acting through AMPA receptors to both sub-populations, but with varying strengths g_E and g_I. The input scales linearly with increasing contrast, c.

$$\mathbf{I}^A_{DC} = c \begin{pmatrix} g_E \\ g_I \end{pmatrix} \tag{2.24}$$

$$\mathbf{I}^{\beta \neq A}_{DC} = 0 \tag{2.25}$$

When we expand the network to study the effect of spatially varying contrasts on gamma peak frequency, we model V1 as a two-dimensional grid of many units with receptive fields that obey retinotopy. Hence we named this model the retinotopic model, and each retinotopic location may be thought of as a cortical column. At each retinotopic location there are two sub-populations, E and I. Each unit x in our network therefore has a type ($a = E$ or I), and a location ($\mathbf{x}$) measured from the center of the network . The location in the network corresponds to its receptive field center.

We keep the same connection types in the retinotopic model as the two-population model. In addition to the overall strengths, $J^{\alpha}_{a,b}$, we scale the strength of connections with distance in receptive field centers by $W_{\mathbf{x},\mathbf{y}|a,b}$. The spatial extent depends on the types of the pre-synaptic and post-synaptic sub-population.

In particular, we scale connections from I sub-populations as a Gaussian in the distance between the two units, $\mathbf{x}$ and $\mathbf{y}$.

$$W_{\mathbf{x},\mathbf{y}|a,I} = e^{-\frac{(\mathbf{x}-\mathbf{y})^2}{2\sigma^2_{a,I}}} \tag{2.26}$$

where $a \in [E, I]$ denotes the type of the unit at $\mathbf{x}$. For simplicity, we assume $\sigma_{II} = \sigma_{EI}$, and that both are small with respect to the spacing between receptive fields, dx. Hence, in agreement with experimental observations, inhibitory connections are highly local and form connections with themselves and their retinotopic excitatory sub-population partner only.

One simple way to get the local contrast dependence of gamma peak frequency would be to model excitatory connections as also being extremely local. In such a case, each retinotopic location is distinct and separate from all the others in the network, and each location effectively is a two-population network. Two-population networks robustly display the contrast dependence of gamma peak frequency (Fig. 2.2E). Therefore the local contrast from the Gabor stimulus in each retinotopic location will determine the gamma peak frequency of that effective two-population system. Such a network should thus easily display the local contrast dependence of gamma peak frequency, as units receiving lower contrast will have lower frequency gamma peaks. However, such a network would fail to show surround suppression in the firing rates [44].

On the other hand, a network with smoothly decaying connection strength should display surround suppression of the firing rates with increasing stimulus size (Fig. 2.6B), but we would expect the gamma peak frequency to be shared by all simulated units, meaning that the local contrast dependence would not be observed (Fig. 2.6C).

To interpolate between the two regimes, we scale the strength of connections from E sub-populations by a linear mixture of complete locality (a Kronecker-δ), and exponential decay.

$$W_{\mathbf{x},\mathbf{y}|a,E} = \lambda_{a,E}\delta(\mathbf{x}-\mathbf{y}) + (1-\lambda_{a,E})e^{-\frac{|\mathbf{x}-\mathbf{y}|}{\sigma_{a,E}}} \tag{2.27}$$

where $a \in [E, I]$ denote the type of the post-synaptic unit. Importantly, $\lambda_{E,E} \neq \lambda_{I,E}$ and $\sigma_{E,E} \neq \sigma_{I,E}$.

The connectivity matrix W therefore has the following form:

$$W^{\alpha}_{\mathbf{x},\mathbf{y}|a,b} = J^{\alpha}_{a,b} W_{\mathbf{x},\mathbf{y}|a,b} \tag{2.28}$$

Similarly, we assume that the feedforward input has the same relative strengths in our retinotopic model as the two-population (g_E and g_I are unchanged), but we must now model the spatial extent of the stimulus.

We model several different presentations of the flat grating stimulus. We vary the contrast while holding the spatial size (r_{stim}) fixed to study the contrast dependence of gamma peak frequency. We vary the spatial sizes at max contrast to study surround suppression of firing rates in our networks. We model the strength

of spatial inputs to our networks for the flat grating as

$$I_{\mathbf{x}} = 1 - \frac{1}{1 + e^{-\frac{|\mathbf{x}| - r_{stim}}{w_{RF}}}} \tag{2.29}$$

where w_{RF} is the decay of receptive fields in our network.

The Gabor stimulus has a contrast envelope which decays across space with some σ_{Gabor} given by

$$c(\mathbf{x}) = e^{-\frac{\mathbf{x}^2}{2\sigma^2_{Gabor}}} 100\% \tag{2.30}$$

Therefore, when we model a Gabor stimulus being presented to our network, we scale the strength of the spatial input to match the spatially decaying contrast of the Gabor (we assume that it is at max contrast at the center of the retinotopic grid).

$$I_{\mathbf{x}} = e^{-\frac{|\mathbf{x}|^2}{2\sigma^2_{Gabor}}} \tag{2.31}$$

where σ_{Gabor} is the Gabor's spatial decay constant. We used $\sigma_{Gabor} = 0.5°$, as in [21].

Again we model input as being feedforward from the thalamus through the AMPA channel, and therefore the input to unit $\mathbf{x}$ of type a is

$$I^A_{DC,a} = c g_a I_{\mathbf{x}} \tag{2.32}$$

$$I^{\beta \neq A}_{DC,X} = 0 \tag{2.33}$$

2.3.4. Linear Approximation

In order to study the power-spectra, we prefer to use the linear approximation to the network dynamics around its fixed point and study the

frequencies of noise-driven fluctuations directly. Such an approximation is computationally more efficient than simulating a stochastic network, and yields greater intuition than stochastic simulation does. With small, but finite, noise we model the system as receiving some perturbations around its fixed point, such that $\mathbf{v}_t^\alpha \rightarrow \mathbf{v}_*^\alpha + \delta\mathbf{v}_t^\alpha$ when $\mathbf{I}_t^\alpha \rightarrow \mathbf{I}_{DC}^\alpha + \boldsymbol{\eta}_t$, where $\boldsymbol{\eta}$ is some small additive noise that is private to each sub-population, and $\mathbf{v}_*^\alpha$ is the N-dimensional fixed-point of the input voltage of receptor type α satisfying

$$\mathbf{v}_*^\alpha = W^\alpha f(\sum_\beta \mathbf{v}_*^\beta) + \mathbf{I}_{DC}^\alpha \tag{2.34}$$

Therefore to first order the dynamical equation describing the fluctuations, $\delta\mathbf{v}_t^\alpha$, around the fixed point is

$$\tau^\alpha \frac{d\delta\mathbf{v}_t^\alpha}{dt} = -\delta\mathbf{v}^\alpha + W^\alpha \Phi \sum_\beta \delta\mathbf{v}_t^\beta + \boldsymbol{\eta}_t^\alpha \tag{2.35}$$

where we've defined τ^α as the α receptor time constant, and the gain matrix Φ as a diagonal matrix whose ii entries are

$$\Phi_{ii} \equiv f'(\sum_\beta v_{i*}^\beta) = nk[\sum_\beta v_{i*}^\beta]_+^{n-1} = nk^{\frac{1}{n}} \boldsymbol{r}_*^{1-\frac{1}{n}} \tag{2.36}$$

Note that the fixed-point firing rates, $\boldsymbol{r}_*$, can be solved using the simple E-I network (no different synaptic types) to save on computational cost, since they are numerically equivalent in both networks, provided stability of $\boldsymbol{r}_*$ when using the expanded synaptic network is checked.

To check the stability in the synaptic model we used the Jacobian of our $3N$-dimensional system, $\mathcal{J}$. Recall that the Jacobian is a $3N \times 3N$ matrix and can

be found by concatenating Eq. (2.35) across α. As such the Jacobian written in $N \times N$ blocks is

$$\mathcal{J} = \mathbf{T}^{-1} \begin{pmatrix} -I + W^A\Phi & W^A\Phi & W^A\Phi \\ W^N\Phi & -I + W^N\Phi & W^N\Phi \\ W^G\Phi & W^G\Phi & -I + W^G\Phi \end{pmatrix} \tag{2.37}$$

where $\mathbf{T} = \text{diag}(\tau_A, \tau_N, \tau_G) \otimes I_{N\times N}$, a $3N \times 3N$ diagonal matrix of the synaptic time-constants. By ensuring the eigenvalues of $\mathcal{J}$ have negative real part, we check the stability of the fixed point of Eq. (2.9).

Taking the Fourier transform of (2.35), and solving for $\widetilde{\delta\mathbf{v}}^\alpha$

$$\widetilde{\delta\mathbf{v}}^\alpha(f) = \sum_\beta \frac{1}{-i2\pi f\tau^\alpha\delta_{\alpha,\beta} + \delta_{\alpha,\beta} - W^\alpha\Phi c^{\alpha,\beta}} \widetilde{\boldsymbol{\eta}}^\beta(f) \tag{2.38}$$

$$\equiv \sum_\beta G_v^{\alpha,\beta}(f)\widetilde{\boldsymbol{\eta}}^\beta(f) \tag{2.39}$$

yields the fluctuation strength across frequencies, where have defined the coupling matrix $c^{\alpha,\beta} \equiv \mathbf{1}^\alpha\mathbf{1}^\beta$ as the matrix of all ones.

In order to solve for the power-spectra of these fluctuations, we find the covariance of the network, $C_v^{\alpha,\beta}(f)$, because, in general, the power-spectra $p(f)$ at a particular location is given by the product of the covariance matrix with some projection matrix $P = \mathbf{e}\mathbf{e}^T$. When dealing with various synaptic types, we must also consider coupling across populations and hence $P \to P^{\alpha,\beta} = P \otimes c^{\alpha,\beta}$

$$p(f) = Tr(P^{\alpha,\beta}C_v^{\alpha,\beta}(f)) \tag{2.40}$$

Since we have found the fluctuation strength across frequencies directly using the linear approximation, we solve for the covariance across frequencies

$$\begin{aligned} C_v^{\alpha,\beta}(f) &\propto \langle \widetilde{\delta \mathbf{v}}^{\alpha}(f) \widetilde{\delta \mathbf{v}}^{\beta}(f)^{\dagger} \rangle \\ &= \sum_{\gamma} \sum_{\epsilon} G_v^{\alpha,\gamma}(f) C_\eta^{\gamma,\epsilon}(f) G_v^{\epsilon,\beta\dagger}(f) \end{aligned} \tag{2.41}$$

where $C_\eta^{\gamma,\epsilon}(f)$ is the covariance of the input noise between synaptic populations γ, ϵ. We assume that noise enters only through the AMPA channel, and hence

$$C_v^{\alpha,\beta}(f) = G_v^{\alpha,A} C_\eta^{A,A} G_v^{A,\beta\dagger} \tag{2.42}$$

Further, we assume that the noise is colored in time, but i.i.d. across units in our network, i.e. $C_\eta^{A,A}(f) = I c_{\boldsymbol{\eta}}(f)$ where $c_\eta(f)$ is the scalar power-spectra of the external noise. In the case of pink noise, as used here,

$$c_{\boldsymbol{\eta}}(f) = I \frac{2\tau_{corr}}{|-2\pi i f \tau_{corr} + 1|^2} \tag{2.43}$$

where τ_{corr} is the correlation time of our noise.

In this case, instead of solving Eq. (2.40) directly, we will obtain a formula for the LFP power spectrum that is mathematically equivalent but is computationally more efficient. We define the N-dimensional vector

$$\mathbf{u}(\omega) = \sum_{\beta} G^{\beta,A}(f)^{\dagger} \mathbf{e}_E \tag{2.44}$$

Then the LFP power spectrum will be given by

$$p(\omega) = c_{\boldsymbol{\eta}}(f) \mathbf{u}(f)^{\dagger} \mathbf{u}(f) \tag{2.45}$$

This is how we evaluated the power spectra for our networks that were not explicitly found using a stochastic simulation.

2.3.5. Definition of gamma peak

We developed our own method of defining gamma peak frequency and width for the linear approximation power-spectra. For the stochastic simulations we relied on the previous method which found the peak frequency as the argmax of the difference between the stimulus presented power-spectra and the background or spontaneous power-spectra $f_0 = argmax(p(f)|_{c>0} - p(f)|_{c=0})$ [21]. In our exploration of parameter space using the linear approximation we found power-spectra that were "shoulders" rather than bumps or peaks, and this previous method would misidentify the peak frequency. It identified the start of the shoulder, rather than the frequency with the greatest increase of power, as being the peak frequency, and so we developed a new method of finding peak frequency which relies on the curvature of the power-spectra.

Generically, a peak occurs between two inflection points of a line when the curvature of that line is downward. As such, we defined our power-spectra peak frequency using the linear approximation method as the middle frequency of two inflection where the power-spectra has negative curvature. The half-width we defined as half the difference between inflection points with negative curvature. For the stochastic simulations, which were noisier than our linear approximation data, the inflection point method of finding the peak frequency did not work as well as the previous method. So for the stochastic simulation we used the previous method.

2.3.6. Proof the spectrum of a synaptic model with only AMPA and GABA is the same as the spectrum of an E/I rate model

Here we prove that the linearization spectrum of a synaptic model without NMDA is the same as the spectrum of an E/I rate model, with the exchange $\tau_{AMPA} \to \tau_E$ and $\tau_{GABA} \to \tau_I$. This means that, in particular, the formulae of [35] for eigenvalues in a 2-neuron/population model still hold for this model with the above replacements.

We start by rewriting the inverse Green's function, using the Green's function defined implicitly in Eq. (2.38), and also start general, allowing for q different receptor types:

$$G^{-1}(\omega) = A - \mathbf{W}\Phi P \tag{2.46}$$

where we define

$$A := -i\omega\mathbf{T} + I \qquad \in R^{qN\times qN} \tag{2.47}$$

$$\mathbf{W} := \begin{pmatrix} W^A \\ W^G \\ W^N \\ \vdots \end{pmatrix} \qquad \in R^{qN\times N} \tag{2.48}$$

$$P := \mathcal{I}_{N\times N} \otimes \mathbf{1}_q^{\mathrm{T}} = (\mathcal{I}_{N\times N}, \mathcal{I}_{N\times N}, \mathcal{I}_{N\times N}, \ldots) \qquad \in R^{N\times qN} \tag{2.49}$$

where $\mathbf{T} = \mathrm{diag}(\boldsymbol{\tau}_s) \otimes \mathcal{I}_{N\times N}$ and $\boldsymbol{\tau}_s \in R$ (in our case $\boldsymbol{\tau}_s = (\tau_A, \tau_G, \tau_N)$ or (τ_A, τ_G)), and $\mathbf{1}_q^{\mathrm{T}} = (1, \ldots, 1) \in R^q$.

The eigenvalue spectrum correspond to values of $z = -i\omega$ which make the determinant of $G^{-1}(\omega)$ vanish. Noting that the second term in Eq. (2.46) is

rank-deficient (has at most rank N, instead of full-rank qN), we make use of the "matrix determinant lemma" to write:

$$\det(G^{-1}(\omega)) = \det(A)\det(\mathcal{I}_{N\times N} - PA^{-1}\mathbf{W}\Phi) \tag{2.50}$$

It is not hard to see that

$$PA^{-1} = (\frac{\mathcal{I}_{N\times N}}{-i\omega\tau_A + 1}, \frac{\mathcal{I}_{N\times N}}{-i\omega\tau_G + 1}, \frac{\mathcal{I}_{N\times N}}{-i\omega\tau_N + 1}, \ldots) \tag{2.51}$$

and therefore

$$PA^{-1}\mathbf{W} = \sum_{\alpha=1}^{q} \frac{1}{-i\omega\tau_\alpha + 1} W^\alpha \tag{2.52}$$

We now limit to $q = 2$ with only AMPA and GABA,

$$PA^{-1}\mathbf{W} = \frac{1}{-i\omega\tau_A + 1} W^A + \frac{1}{-i\omega\tau_G + 1} W^G = W\widetilde{A}^{-1} \tag{2.53}$$

where we have made use of the specific forms of W^A and W^G (namely, that they have zero columns for inhibitory and excitatory neurons, respectively) from Eq. (2.23), and where we have defined

$$W = \sum_\alpha W^\alpha \qquad \in R^{N\times N} \tag{2.54}$$

$$\widetilde{A} := z\widetilde{T} + \mathcal{I}_{N\times N} \tag{2.55}$$

with $\widetilde{T} = \text{diag}(\tilde{\boldsymbol{\tau}}) \in R^{N\times N}$ where $\tilde{\boldsymbol{\tau}} = (\tau_A, \ldots, \tau_G, \ldots) \in R^N$ is the N-dimensional vector with first N_E components equal to τ_A and the last N_I components equal to τ_G. After identifying $\tau_{A/G}$ with $\tau_{E/I}$, we thus see that $\widetilde{T}$ is the same as the T matrix of the $\boldsymbol{r}$-model (which is N-dimensional), as is W its connectivity matrix.

Also noting that $(z\widetilde{T} + \mathcal{I}_{N\times N})^{-1}$ and Φ are both diagonal, we can commute them in Eq. (2.53) to obtain:

$$\det(G^{-1}(\omega)) = \det(A)\det(\mathcal{I}_{N\times N} - W\Phi\widetilde{A}^{-1}) \tag{2.56}$$

$$= \frac{\det(A)}{\det(\widetilde{A})}\det(z\widetilde{T} + \mathcal{I}_{N\times N} - W\Phi) \tag{2.57}$$

$$= \frac{\det(A)}{\det(\widetilde{A})}\det(z\widetilde{T} + \mathcal{I}_{N\times N} - \Phi W) \tag{2.58}$$

(to get the last line, do a similarity transform with Φ, of the matrix in the last determinant).

Now it is explicit that the zeros of the last determinant factor are the eigenvalues of the N-dimensional $\boldsymbol{r}$-system (after $\tau_{A/G} \leftrightarrow \tau_{E/I}$ identification).

The first factor, on the other hand, can be written as:

$$\frac{\det(A)}{\det(\tilde{A})} = \frac{(z\tau_A + 1)^N (z\tau_G + 1)^N}{(z\tau_A + 1)^{N_E}(z\tau_G + 1)^{N_I}} = (z\tau_A + 1)^{N_I}(z\tau_G + 1)^{N_E} \tag{2.59}$$

So the spectrum also has N additional *real* eigenvalues (in addition to those of the $\boldsymbol{r}$-model) with values $-\tau_A^{-1}$ and $-\tau_G^{-1}$, and multiplicities, N_I and N_E, respectively. (Thus in total we have $2N$ eigenvalues as we should.)

In particular, all oscillatory/complex eigenvalues are exactly those of the $\boldsymbol{r}$-model in the no-NMDA case, which in the 2-neuron case are given by the formulae in Tsodyks et al. 1997.

2.3.6.1. Approximate statement about role of NMDA:

We consider two regimes for the effect of NMDA:

1. when $|z|$ or ω are very small compared to the NMDA time-constant: $\omega \ll \tau_N^{-1}$.

2. when $|z|$ or ω are very large compared to the NMDA time-constant: $\omega \gg \tau_N^{-1}$.

The first regime is relevant for DC response and DC properties (such as surround suppression of steady-state rates). The second regime is approximately valid for gamma oscillations, thanks to the relatively high frequency of those.

In regime 1, it is obvious that the breakdown of E weights into the two types doesn't have any effects, simply because (setting ω to 0) time-scales don't play any role here. So the parameter ρ_N makes no difference to fixed point response properties.

In regime 2, looking at Eq. (2.52), we note that the prefactor $\frac{1}{-i\omega\tau_\alpha+1}$ for NMDA is very small and can be ignored. This means that for high frequencies (*e.g.*, approximately frequencies around gamma) we can simply kill all NMDA weights, and only consider the AMPA weight matrix, W^A. In particular, the model where $W^A \propto W^N$, then the effect of NMDA on the gamma peak is approximately equivalent to reducing total excitatory weights (which all affect DC properties) by a scalar factor (which in our formalism is $1 - \rho_N$) when it comes to gamma properties.

2.3.7. Theorems for the two-population model

We consider now the case of a two-neuron model with one excitatory and one inhibitory neuron (the neurons in all models of this article should be thought of as "mean field" neurons, each representing a statistically homogenous population of neurons of excitatory or inhibitory type), what we have here called the two-

population model. We will also assume no NMDA contribution (or equivalently work in the very slow NMDA regime, and replace all excitatory weights with their AMPA part, as explained at the end of Sec. 2.3.6).

In this case the gamma peak frequency is **closely approximated** by the imaginary part of the eigenvalues of the Jacobian matrix:

$$\mathcal{J} = -\mathbf{T}^{-1} + \mathbf{T}^{-1}\mathbf{W}\Phi \tag{2.60}$$

$$= \begin{pmatrix} \gamma_E(-1 + W_{EE}\Phi_E) & -\gamma_E W_{EI}\Phi_I \\ \gamma_I W_{IE}\Phi_E & \gamma_I(-1 - W_{II}\Phi_I) \end{pmatrix} \tag{2.61}$$

where we defined $\gamma_E \equiv \tau_{AMPA}^{-1}$ and $\gamma_I \equiv \tau_{GABA}^{-1}$. Noting that the trace and determinant of $\mathcal{J}$ yield the sum and product of the eigenvalues, respectively, we obtain the expression (see [35])

$$\begin{aligned} 2\lambda_{1,2} = {} & \gamma_E(W_{EE}\Phi_E - 1) - \gamma_I(W_{II}\Phi_I + 1) \\ & \pm \sqrt{\left[\gamma_E(W_{EE}\Phi_E - 1) + \gamma_I(W_{II}\Phi_I + 1)\right]^2 - 4\gamma_E\gamma_I W_{EI}W_{IE}\Phi_E\Phi_I} \end{aligned} \tag{2.62}$$

A gamma peak exists only if the expression under the square root is negative, *i.e.*

$$4\gamma_E\gamma_I W_{EI}W_{IE}\Phi_E\Phi_I > \left[\gamma_E(W_{EE}\Phi_E - 1) + \gamma_I(W_{II}\Phi_I + 1)\right]^2, \tag{2.63}$$

in which case, for the gamma peak angular frequency ω_0, we (**approximately**) have

$$4\omega_0^2 = 4\beta_E\beta_I W_{EI}W_{IE} - (\beta_E W_{EE} + \beta_I W_{II} + \gamma_I - \gamma_E)^2 \tag{2.64}$$

where we defined $\beta_X := \gamma_X\Phi_X$ for $X \in \{E, I\}$.

We will now obtain a simplified expression for the derivative of ω_0^2 with respect to the contrast c, using the rectified supralinear nonlinearity of the SSN. Using $\Phi_* = nk^{\frac{1}{n}} r_*^{1-\frac{1}{n}}$ (where r_* is the firing rate at fixed point) we obtain

$$\frac{d\beta_*}{dc} = \frac{n-1}{n}\beta_* \frac{d\ln r_*}{dc} \tag{2.65}$$

Then using Eq. (2.64), and defining

$$A := (\beta_E W_{EE} + \beta_I W_{II} + \gamma_I - \gamma_E) \tag{2.66}$$

and $(\cdots)' := \frac{d(\cdots)}{dc}$, we find:

$$\begin{aligned}\frac{n}{n-1}\frac{d\omega_0^2}{dc} &= \beta_E\beta_I W_{EI}W_{IE}(\ln r_E + \ln r_I)' \\ &\quad - \frac{2A}{4}\left(\beta_E W_{EE}(\ln r_E)' + \beta_I W_{II}(\ln r_I)'\right) \qquad (2.67)\\ &= \omega_0^2(\ln r_E + \ln r_I)' \\ &\quad + \frac{1}{2}A^2\left[\frac{(\ln r_E)' + (\ln r_I)'}{2} - \frac{\sum_a w_a(\ln r_a)'}{\gamma_I - \gamma_E + \sum_a w_a}\right] \qquad (2.68)\end{aligned}$$

where the sums are over $a \in \{E, I\}$ and we defined

$$w_a := \beta_a W_{aa} \qquad\qquad a \in \{E, I\} \tag{2.69}$$

Analysis of the sign of $\frac{d\omega_0^2}{dc}$:

Assuming that we are in the gamma oscillatory regime (*i.e.*, ω_0 is real) and that the fixed point rates increase with contrast, then from Eq. (2.68) we find that sufficient conditoin for $\frac{d\omega_0^2}{dc} > 0$ is that the factor in the square brackets in Eq. (2.68) is positive. In the solutions of SSN most relevant to cortical biology,

$(\ln r_I)'$ tends to be larger than $(\ln r_E)'$ (because excitatory rates tend to saturate or supersaturate earlier). We thus consider two extreme cases: $(\ln r_E)' = (\ln r_I)'$ and $(\ln r_E)' = 0$.

In the first case, the bracket becomes $(\ln r_I)' \left[1 - \frac{\sum_a w_a}{\gamma_I - \gamma_E + \sum_a w_a}\right] = (\ln r_I)' \frac{\gamma_I - \gamma_E}{\gamma_I - \gamma_E + \sum_a w_a}$, which is positive as long as $\gamma_I > \gamma_E$ (which is unfortunately not the case for GABA and AMPA).

In the second case, the bracket factor becomes $(\ln r_I)' \frac{\gamma_I - \gamma_E + \sum_a w_a - 2w_I}{2(\gamma_I - \gamma_E + \sum_a w_a)} = (\ln r_I)' \frac{\gamma_I - \gamma_E + w_E - w_I}{2(\gamma_I - \gamma_E + \sum_a w_a)}$. This is positive (as long as the denominator is positive, which is true as long as $\gamma_I > \gamma_E$) if

$$w_E - \gamma_E + \gamma_I > w_I \tag{2.70}$$

But the stability of the fixed point dictates that the expression on first line of Eq. (2.62) (the real part of the eigenvalues) has to be negative and thus

$$w_I > w_E - \gamma_E - \gamma_I \tag{2.71}$$

2.4. Discussion

In this work we have shown that the expanded SSN is able to robustly display the contrast dependence of gamma peak frequency in both a two-population and a retinotopic network. The retinotopic model successfully balances the trade-off in horizontal connection strength such that both the local contrast dependence of the gamma peak frequency and the surround suppression of firing rates are observed robustly. In order to capture gamma oscillations using the SSN, we expanded the model beyond an E-I network to a varied synaptic network

model. Crucially, the SSN account sheds light on the mechanism underlying the contrast dependence of gamma peak frequency and points to the key role of the non-saturating and expansive neural transfer function, observed empirically [52, 60], in giving rise to this effect.

Finding the power-spectra using the linearization to Eq. (2.8), helped us make analytic simplifications. From these simplifications, we gained insights about how the SSN may produce gamma (*e.g.* Secs. 2.2.6, 2.3.6, and 2.3.7). As gains and effective connectivity increase, the imaginary part of the corresponding eigenvalue of the Jacobian also increases. Moreover, by finding the power-spectra via linearization, we were able to rapidly compute power-spectra which allowed for extensive explorations of the model's parameter space. Throughout this work, we developed several techniques for finding parameters which robustly display the local contrast dependence of gamma peak frequency, which could aid in future searches looking to understand the functions underlying gamma oscillations.

In this work, partly for simplicity, we assumed an instantaneous I/O function between net synaptic input ($\sum_\beta \mathbf{v}_\beta$) and the output rate. This is based on the approximations discussed in [59], which is valid when the fast synaptic filtering time-constants (τ_{AMPA} and τ_{GABA}) are much smaller than the neuronal membrane time-constants. Note that our framework, however, can easily be generalized beyond this approximation by using the full neuronal linear response filter obtained from the Fokker-Planck treatment of [59]. This biggest impact caused by this treatment would be to render the gains used here frequency dependent (*i.e.* $\Phi \to \Phi(f)$). We expect this dependence to be weak because we are in the regime of fast synaptic filtering as compared to the neuronal membrane time constant, and so we expect the transfer function to be approximately instantaneous. Therefore

we do not expect that including the full neuronal linear response filter would change our qualitative results.

As we have shown, the SSN produces the local contrast dependence of gamma peak frequency and the surround suppression by balancing long-range and local horizontal connections. The local component of the connections used represents some elevated strength within a (micro)column relative to the long-range connections. This predicts that there exists an elevated probability of connection within a (micro)column in the structure of horizontal connectivity in macaque V1. Indeed, it has been noted that anatomical findings on the spatial profile of horizontal connections in the macaque cortex point to such a mixture of short-range or local and long-range connections, with the local component not extending beyond 0.4 mm (the size of our model's columns) [61].

Rodents, however, have a very different V1 organization than macaques. Typically rodents are thought to not have strong columnar organization across the cortex, based mainly on the salt-and-pepper organization of orientation preference in their V1 [7]. More recently, experimental evidence from mouse V1 suggests that there is no elevated strength in local connections relative to long-range connections [62]. As such, we predict that a similar experiment to [21] *in mouse* would find that there is no local contrast dependence of gamma peak frequency, unlike what is observed in macaque.

CHAPTER III

CORTICAL DYNAMICS AT MULTIPLE TIME-SCALES DRIVE THE JOINT GROWTH OF SMOOTH MAPS AND LOCAL CODING HETEROGENEITIES

The simulations and analyses done in this chapter were carried out by myself under the guidance of Dr. Yashar Ahmadian, the principal investigator who conceived the project. Dr. Ken Miller pointed out the question and the theoretical conundrum motivating this work, and has consulted us throughout the project.

3.1. Introduction

Sensory cortices are specialized regions of the brain which process sensory stimuli. Primary sensory cortices are cortical areas which first receive input from their corresponding sensory organs, typically via the thalamus. Neurons within sensory cortices of mature mammals respond selectively to certain features of the input. Such selectivities and preferences mean that sensory cortical neurons have the greatest response when the feature is present in the stimulus. Most response selectivities first emerge during embryonic or early post-natal development. For example, neurons in the early visual system only respond to stimuli in a certain small region of the visual field, called the neuron's receptive field (Fig. 3.1A). Receptive fields are different for different neurons, and all together they tile the entire visual field. Moreover, neurons within the primary visual cortex (V1) develop selectivity to oriented patterns of light and dark bands within their receptive field, with different neurons responding preferentially to different angles (referred to as the neuron's preferred orientation, see Fig. 1.2 for examples). Similarly, neurons within primary auditory cortex develop selectivities for a

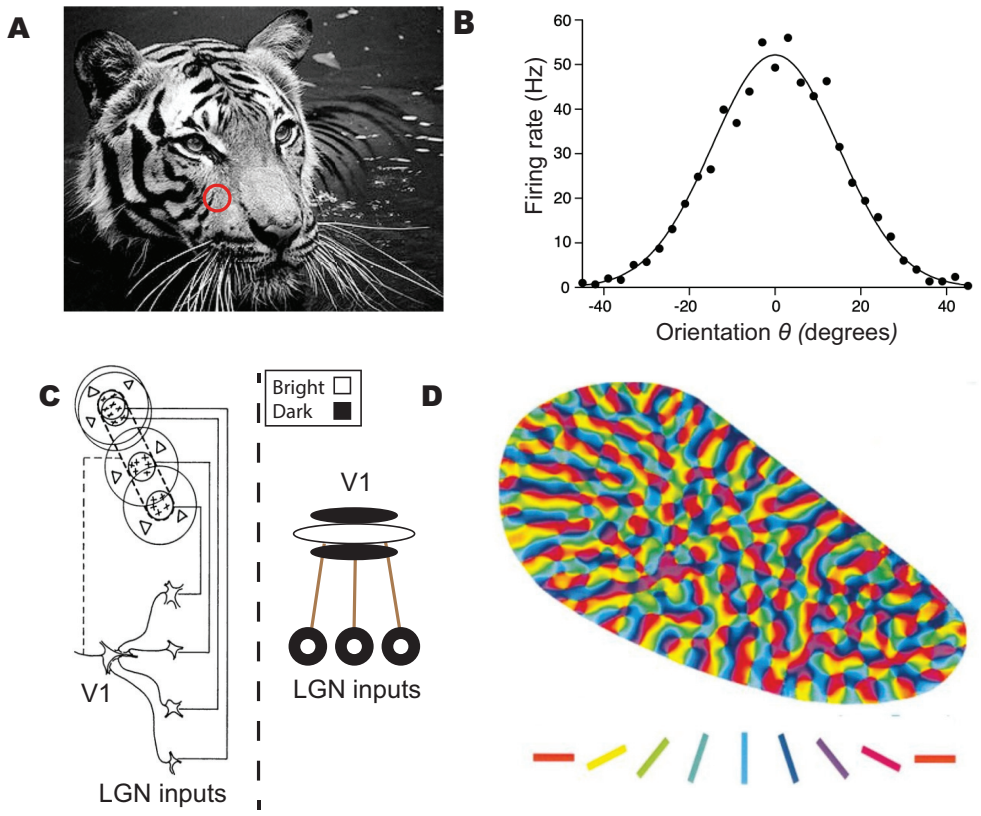

FIGURE 3.1. Example of receptive fields and orientation selectivity. Also how orientation selectivity develops from feedforward thalamocortical connections plus its organization across the cortex. **A:** Example receptive field (red circle) of a neuron within V1 **B:** Orientation selectivity example from [2] (data from [63]). **C:** Example of a V1 RF in cat. Left image shows a realistic example, while the right is a cartoon example for clarity. Left: Example of a V1 RF filter in cat. Shown is a V1 cell making connections with 4 LGN inputs (explosion symbols bottom right). Solid lines connect each LGN input to its RF. The RF filter of each LGN cell is shown as two circles with solid lines (plus signs indicate ON regions, triangles indicate OFF regions, hence these are ON-center, OFF-surround responsive cells). The resulting RF of the V1 cell is shown as a dashed outline on top of the circles, and connected to the V1 cell with the dashed line on the left. Taken from [6]. Right: we simplified the diagram on the left to a cartoon where the RF filter of the LGN neurons and V1 neuron are shown. Black filled shapes correspond to dark patches of visual input, while white filled shapes represent light patches of visual input. **D:** Orientation preference map from [64]. Shown is the orientation preference map from a tree shrew, a primate.

specific sound frequency, while neurons within the early olfactory pathway develop selectivity to certain odorants. This work will focus on the development of neural

selectivities and their spatial organization in V1. In principle though, the findings of this work could be applied to any sensory cortex as long as certain parts of the model were adjusted to match the new modality.

Moreover this work will focus on the development and organization across cortical populations of feature selectivities through activity-dependent plasticity, where the synaptic connections between neurons change depending on their activity. As such, we will not study the development of "hard-coded" or genetically determined features and organization. We will nevertheless summarize some key points about those features here. The receptive field (Fig. 3.1A) of neurons within V1 develops due to feedforward connections from the lateral geniculate nucleus (LGN) of the thalamus, and encodes the feature of visual solid angle or RF location. Feedforward connections mean that LGN cells form synapses onto V1 cells. The feature of RF location is encoded in a smooth map across the cortex, meaning that nearby neurons selectively respond to nearby regions of visual space; this smooth encoding of RF location across V1 is known as retinotopy. However, the development of retinotopy does not depend on the activity of V1 or the LGN. Rather the patterning of feedfoward connections from the LGN to V1 are genetically determined such that RF location selectivity develops and forms a smooth map in V1 as it develops in an early embryonic period [8, 65]. Thus while the development of retinotopy is outside the scope of this work, our models of cortex and LGN-to-V1 (thalamocortical) connections will have retinotopic organization.

Instead, this work will focus on late pre-natal (in primates) or early post-natal periods (in cats, ferrets), when feedforward LGN-to-V1 connections undergo rapid activity-dependent plasticity. During this period, neurons within these

areas develop selectivity to certain features of their input [8]. One such feature is stimulus orientation. When presented with an oriented stimulus, like a long bar or a stripe pattern, neurons within V1 respond with the greatest firing rate at one particular orientation (Fig. 3.1B), known as their preferred orientation. In the example shown in Fig. 3.1B, the cortical cell responds most strongly when the grating is oriented at 0°, thus 0° is called the preferred orientation of this cell. Thus V1 neurons develop and apply a spatial filter to incoming visual sensory information on top of their RF. It is known that orientation preference and selectivity arise primarily from the patterning of feedforward LGN-to-V1 connections [6, 54, 66, 67]. Hence, LGN neurons are the pre-synaptic population while V1 cells are the post-synaptic population. Importantly, orientation preference is not "hard-coded" or genetically determined, but rather develops by strengthening certain LGN-to-V1 synapses and weakening others due to the neural activity in the LGN and V1 [8].

In contrast to V1 cells, LGN cells are not orientation selective, but do have a RF and, moreover, an RF filter. More precisely, cells within LGN respond strongly to differing center surround combinations. For instance, an ON-center, OFF-surround cell responds strongly to a bright spot, surrounded by a dark spot, while the opposite is true for an OFF-center, ON-surround cell (for brevity we will refer to these an ON- or OFF-center cells (Fig. 3.1C). Neurons within V1 receive synaptic connections from combinations of these ON- and OFF-center cells, which gives rise to the oriented structure within their RF and thus their orientation preference and RF filter [6] (Fig. 3.1C).

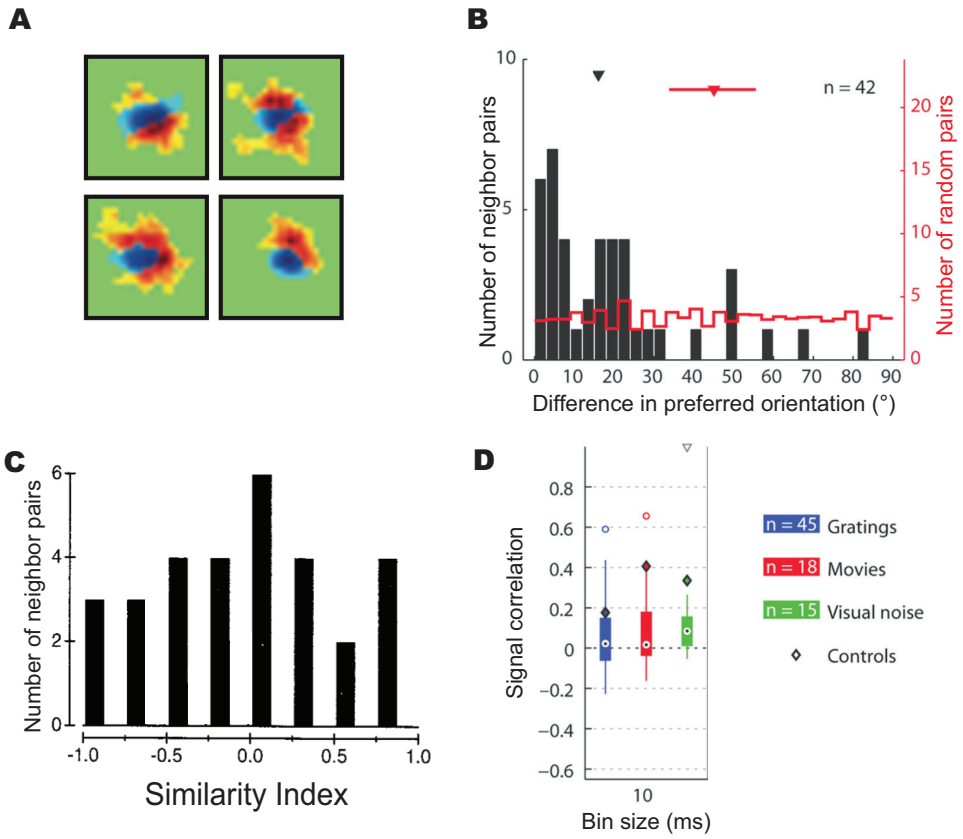

FIGURE 3.2. Local diversity of receptive field structure in cat V1. **A:** Examples of spatial receptive field filters for a group of neighboring V1 neurons. Red regions are ON-responsive (meaning they respond positively to increases in local brightness); blue regions are OFF-responsive (meaning they respond positively to increases in local darkness). The receptive fields are arranged from left to right, top to bottom as the corresponding cells are located in the cortex [68]. **B:** Distribution of preferred orientation differences across neighboring pairs of neurons (black bars) and across randomly selected pairs of neurons (red) [23]. **C:** Distribution of RF similarity indices across pairs of neighboring cells within V1 from [69]. **D:** Signal correlations across pairs of neighboring V1 cells for three different stimulus classes: gratings (blue), natural movies (red), and visual noise (green). Signal correlation is defined as the correlation coefficient of the trial-averaged responses of two neurons across different stimulus conditions or time bins. Shown are box-and-whisker plots of signal correlation, the boxes extend from the 25^{th} to the 75^{th} percentiles; the median is indicated by the black dot. Whiskers show the whole range of signal correlations, with outliers shown as circles. Diamonds represent expected correlations from identical, but noisy neurons, which serve as controls. More precisely, a single neuron's responses in different trials were partitioned into two sets, and the signal correlation of those two sets of trials was calculated as if they belonged to two distinct neurons [23].

In some species, the development of orientation selectivity occurs within a "critical period" of development. It is known that the development of many features or even learning of certain skills in animals have so-called critical periods. Orientation preference finishes development during this critical period, and moreover is stably represented in the cortex afterwards [8, 70]. If visual input is denied or altered for the organism during this period, *e.g.* by over-representing a particular orientation by putting cylindrical-lens-fitted goggles on the animal, then development of orientation preference is irreversibly changed [71].

In mature cats and higher mammals (*e.g.*, primates and humans), orientation preference is encoded in a smooth map across the cortex (Fig. 3.1D and Fig. 3.2B), such that nearby neurons have similar preferred orientations [23, 64, 72, 73, 74]. By contrast, in rodents (*e.g.* mice and rats) orientation preference develops a random, or salt-and-pepper (S&P) organization, meaning that nearby cells have an orientation preference which is uncorrelated with their neighbors [7, 75, 76]. In this work, we will not focus on comparing the development of qualitatively different organizations of orientation selectivity across different species. Rather, *this work focuses on the development of qualitatively different organizations of different feature selectivities in the same visual cortex, in species with smooth orientation maps.*

Thus in V1 of higher mammals, both orientation and RF location have smooth maps (despite their different mechanisms of development). These smooth maps are thought to help minimize wiring lengths within the brain between similarly tuned neurons. A possible explanation for why cats and higher mammals have developed smooth maps of orientation preference, while rodents have not, is

that the cortices of cats and higher mammals would be too large without them. By minimizing the wiring lengths, cortical volume can be kept to a minimum [77].

The fact that these key feature selectivities, *e.g.* orientation and RF location, develop smooth maps, despite their different mechanisms of development, may suggest that RF filters of neighboring V1 cells are very similar. However, even in these animals, not every V1 feature is organized in a smooth map. If that were the case, the RF filters of neighboring neurons would indeed have a high similarity with one another, and the response of neighboring neurons would be highly correlated and therefore redundant. Rather, the responses of cortical neurons to natural or artificial stimuli tend to be minimally correlated with their neighboring neurons [23] (Fig. 3.2D), implying that certain features must be encoded in salt-and-pepper (S&P) organization, a qualitatively different organization than a smooth map. In S&P organizations, rather than nearby cells preferring similar features, there is no correlation between the preferences of neighboring neurons. This low response correlation (also known as signal correlation) minimizes the redundancy of local cortical responses, and hence maximizes their information content.

To a first approximation, responses of neurons in the early visual system (retina, LGN, and the thalamorecipient layer of V1) can be approximated by linear functions of the visual stimulus. If for the moment we ignore temporal dynamics, the stimulus can be thought of as an image, which we denote by $s(\mathbf{a})$. Here, $\mathbf{a}$ denotes the location in the 2D visual field, and $s(\mathbf{a})$ is the light intensity (for this work we will ignore color) at that location. The response of the cell i, r_i, can then be approximated (up to a constant) by the inner product of the cell's receptive

field filter, $F_i(\mathbf{x})$, with the image stimulus as follows

$$r_i = \sum_{\mathbf{a}} F_i(\mathbf{a}) s(\mathbf{a}). \equiv \mathbf{F}_i \cdot \mathbf{s}. \tag{3.1}$$

Recall that the receptive field filters of V1 cells are oriented patterns (see *e.g.*, Fig. 3.1C) and emerge from the patterning of LGN-to-V1 connections.

Thus the low response (signal) correlation in V1 suggests that RF filters vary from neighbor to neighbor despite having similar orientation preferences, and indeed that is the case. Consider, for example, the work of [68] which studied RF filters across V1. We have reproduced a set of neighboring RF filters from their work in Fig. 3.2A. Note that the neurons prefer similar orientations but have varying spatial shapes of their RF filters, *e.g.* how big the light and dark lobes are and the location of those lobes within the RF, are very different (as opposed to the cartoon examples shown in Fig. 1.2). As such, we will now introduce two key measures for the similarity of RFs and neural coding, relative and absolute spatial correlations of RF filters.

To measure relative spatial correlation we will compare the Pearson correlation coefficient between neighboring RF filters centered at a standard location. Previous experimental studies labeled relative spatial correlation as the similarity index between RF filters [69]. They found that the distribution of similarity indices between neighbors was centered on zero. This suggests that while selectivities to certain features are organized in a smooth map, *e.g.* orientation preference (Fig. 3.2B), there exist selectivities to other features which develop a S&P organization across the cortex. These S&P organizations are thought to keep information between neural responses and stimuli high by keeping the correlations between responses low. As a corroborating measure, we will also study

the absolute spatial correlation. For absolute spatial correlations we will compute the Pearson correlation coefficient between feedforward connections of neighboring neurons. Mathematically, then, the absolute spatial correlation is the correlation between $\mathbf{F}_i$ and $\mathbf{F}_j$ when i and j refer to neighboring neurons. When the absolute spatial correlation is low, we expect neighboring neurons within the primary visual cortex to be driven by different populations of thalamic neurons. The signal correlations between those neighboring cortical neurons should be low as they receive different inputs. In summary, signal correlations and similarity indices near zero imply the existence of S&P organizations, even though orientation preference develops a smooth map across the cortex.

Therefore, a mixture of smooth maps and S&P organizations develop within the same cortex for different features. This co-presence of different organizations for different features may allow the cortex to "get the best of both worlds", so to speak, and minimize both wiring lengths and output activity correlations. Previous theoretical models, however, predicted that the same feature organization, either a smooth map or S&P organization, would develop for all features within the same cortex [22, 78, 79, 80].These studies showed that the feature selectivity organization which develops depends not on the features themselves, but on the linear response kernel of the cortical network (which will be introduced more mathematically in the next section). We will refer to the cortical linear response kernel as the cortical response generally. The cortical response kernel of the post-synaptic network describes how external input to one cortical neuron impacts the responses of other cells in its extended neighborhood (due to intra-cortical recurrent interactions). Furthermore, in previous works and in this work we assume that all cortical neurons have the same cortical response kernel. When the

spatial response kernel of the cortex is spatially broad and excitatory (positive), then the feature preferences of nearby cells align and become more similar over the course of development. Therefore the cortex develops a smooth selectivity map for all of the stimulus feature selectivities. If, however, the response kernel of the cortex was spatially narrow, meaning that external input to a cortical neuron had little to no effect on the response of others in its neighborhood, then the feature selectivity of nearby cortical cells would develop independently from each other. In this case, the cortex develops S&P selectivity organizations for all features of the input. Importantly, as a consequence of assumptions made by these models, the same cortical response kernel was shared by all features.Hence, once the spatial form of the response kernel was chosen for one feature, then it was set for all the other features as well. In short, previous models predicted the development of either a smooth map or S&P organization for all feature selectivities within the post-synaptic population.

Here we extend and generalize those previous studies. We show that when cortical dynamics at multiple time-scales are properly taken into account, the response kernel of the cortex can have different spatial profiles at different time-scales. Moreover, we show that a novel coupling mechanism emerges, wherein features of the cortical input that temporally vary at different time-scales couple most strongly to the cortical spatial response kernel at similar time-scales. This, in principle, allows the development of qualitatively different spatial organizations for selectivities to features that vary at different time-scales, within the same cortex. We present simulations of cortical development of both one-dimensional and two-dimensional models of cortex showing the development of qualitatively different feature organizations under our novel coupling mechanism, and compare

our results to biologically observed feature organizations. Lastly, we give a detailed account of the mathematical and computational methods used.

3.2. Theoretical results

3.2.1. Background and motivation

Previous theoretical models relied on modelling activity dependent plasticity or development under Hebbian learning and made four (I - IV) key assumptions [22, 78, 79, 80].

I Bilinear Hebbian learning

II Linear cortical response

III Instantaneous cortical response

IV Slow plasticity relative to input

Hebbian learning is most commonly identified by the aphorism that "cells that fire together, wire together", meaning that synaptic connections are strengthened when two neurons are co-active together and weakened when they are not [13, 81]. In particular, these previous models utilized a bilinear Hebbian learning rule (assumption I), meaning that the strength of connection between two neurons should grow proportionally with the product of the activity of the pre- and post-synaptic neurons. Consider a single excitatory synapse, whose weight we will denote by w. Let us further denote the activity of pre- and post-synaptic cells for that synapse by u and v, respectively. Mathematically, the Hebb rule is then

equivalent to the following differential equation governing the evolution of w:

$$\frac{dw(t)}{dt} \propto v(t)u(t). \tag{3.2}$$

In a modification of the Hebb rule known as the covariance rule [81], the pre- and post-synaptic activities are replaced with the deviations of those activities from their long-term temporal average. We will implicitly assume such a covariance rule below, by letting $v(t)$ and $u(t)$ denote the deviations of pre- and post-synaptic activities relative to their temporal means.

In this chapter we would like to model not just the evolution of a single synapse, but of the entire collection of thalamocortical synpases connecting the LGN to the (thalamorecipient layer of) V1. Our models will generally feature N_{pre} pre-synaptic LGN cells and N_{post} post-synaptic cortical cells. We will use W to denote the full $N_{pre} \times N_{post}$ matrix of synaptic weights connecting the LGN cells to V1 cells, such that the matrix element $W_{x,a}$ denotes the weight of the synapse connecting the a-th LGN cell to the x-th cortical cell. We similarly adopt a vector notation for the pre- and post-synaptic activities: we let $\mathbf{u}$ denote the N_{pre}-dimensional vector representing the activity of the pre-synaptic LGN population, and let $\mathbf{v}$ denote the N_{post}-dimensional vector representing the activity of the post-synaptic cortical population. We will index neurons from the pre-synaptic population with letters from the beginning of the alphabet, *e.g.* a, and index neurons from the post-synaptic population with letters from the end, *e.g.* x. Under the bilinear Hebbian learning then, the feedforward connection, $W_{x,a}$, from the pre-synaptic neuron at location a to the post-synaptic neuron at location x,

evolves according to $\frac{dW_{x,a}}{dt} \propto v_x(t)u_a(t)$, or in matrix-vector notation [1]:

$$\tau_W \frac{dW}{dt} = \mathbf{v}(t)\mathbf{u}^T(t). \tag{3.3}$$

Here τ_W denotes the time-scale of synaptic development. Note that we will use capitalized letters to represent matrices in our formalism, *e.g.* W, and bold-face for vectors, *e.g.* $\mathbf{v}$.

Previous models also assumed that the Input/Output transfer function of cortical cells was linear (assumption II). Consider an illustrative example with only one post-synaptic neurons such that $\mathbf{v} \rightarrow v$ and $W \rightarrow \mathbf{w}$ an N_{pre}-dimensional vector. In that case and under the assumption of a linear I/O transfer function, the activity of the post-synaptic neuron is determined by a linear weighted sum of its inputs plus a decay to its resting state

$$\tau_v \frac{dv(t)}{dt} = -v(t) + \mathbf{wu}(t) \tag{3.4}$$

These models further assume that the cortex is in a fixed-point of its dynamics, or that the response of the cortex is much faster than the dynamics of the thalamus $\tau_u \gg \tau_v$ and so it is "slaved" to the dynamics of the thalamus (assumption III). Under that assumption, the left-hand side of Eq. (3.4) is effectively zero, and the activity of the post-synaptic cell may be written as

$$v(t) = \mathbf{wu}(t) \tag{3.5}$$

[1]Recall that we are defining $\mathbf{u}$ and $\mathbf{v}$ to be the deviations of the pre- and post-synaptic population respectively from their long-term temporal average. As such, Eqn 3.3 can describe weakening of synapses as well as strengthening.

Substituting Eq. (3.5) into Eq. (3.3) yields

$$\tau_W \frac{d\mathbf{w}}{dt} = \mathbf{w}\mathbf{u}(t)\mathbf{u}(t)^T \tag{3.6}$$

The last assumption these previous studies made was that synaptic plasticity was slow relative to the dynamics of the activity $\tau_W \gg \tau_u \gg \tau_v$ (assumption IV). This assumption prevents synapses from constantly being "re-written" by changing activity, and is experimentally justified given that the thalamus has activity with a time-scale on the order of milliseconds, whereas plasticity changes on the order of seconds [82]. Hence we may average Eq. (3.6) over some time T such that

$$\tau_W \frac{d\mathbf{w}}{dt} = \mathbf{w}\frac{1}{T}\int_0^T dt\mathbf{u}\mathbf{u}^T \equiv \mathbf{w}C \tag{3.7}$$

where C is the $N_{pre} \times N_{pre}$ matrix of zero time-delay correlations of thalamic activity. Under these simplifying assumptions, $\mathbf{w}$ will grow along the eigenvectors (principal components) of C, and the post-synaptic neuron will therefore develop a selectivity to the eigenvectors of C. Hence, the post-synaptic neuron will respond greatest when the input is aligned to the principal eigenvector of C, and therefore the post-synaptic neuron will become selective to inputs that match that principal eigenvector.

Now, consider many post-synaptic neurons. In this case, the activity of post-synaptic populations $\mathbf{v}$ also has some dependence on the recurrent connectivity within the post-synaptic population and is given by

$$\tau_v \frac{d\mathbf{v}}{dt} = -\mathbf{v} + W\mathbf{u} + R\mathbf{v} \tag{3.8}$$

where R is the recurrent connectivity in the cortex. Imposing the same condition, that $\tau_u \gg \tau_v$, and borrowing from linear response theory whereby we assume the cortex is a "black-box" which linearly transforms its inputs, leads to

$$\mathbf{v}(t) = KW\mathbf{u}(t) \tag{3.9}$$

where K is $N_{post} \times N_{post}$ matrix describing the linear response kernel at zero time-lag of the cortex across the cortical population. Note that K is related to the recurrence in the network R, but from here on we will model K directly and remain agnostic as to the spatial structure of R.

Substituting Eq. (3.9) into Eq. (3.3), yields

$$\tau_W \frac{dW}{dt} = KW\mathbf{u}(t)\mathbf{u}^T(t) \tag{3.10}$$

Again, we assume that plasticity is slow relative to the dynamics of the thalamus such that $\tau_W \gg \tau_u \gg \tau_v$ (assumption IV), and average with respect to time

$$\tau_W \frac{dW}{dt} = KW\frac{1}{T}\int_0^T dt\mathbf{u}(t)\mathbf{u}^T(t) \equiv KWC \tag{3.11}$$

Fig. 3.3 shows as schematic of these models [2]. The post-synaptic population is shown as green circles, while the pre-synaptic population is shown in red. The spatial form of the cortical response kernel K is shown above the post-synaptic population, while the spatial shape of the input correlations,C, are shown below

[2]Note that we are suppressing the convolution symbol in our formalism, opting to denote convolutions across space by

$$KWC \quad = \quad K * W * C \equiv \sum_{x'}\sum_{a'} K(x,x')W(x',a')C(a',a) \tag{3.12}$$

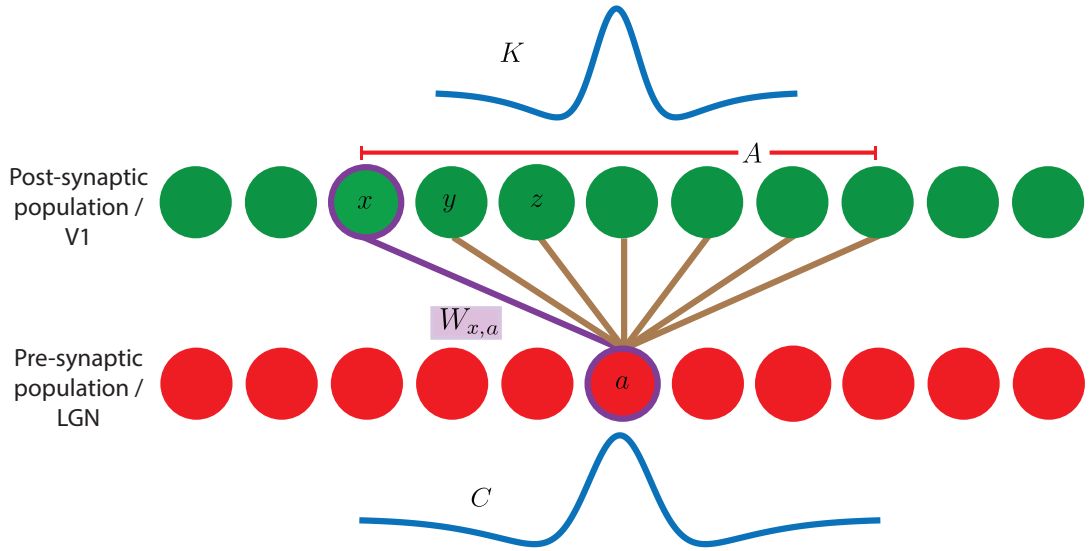

FIGURE 3.3. Schematic of previous theoretical models [22]. Here we show two one-dimensional representations of networks of neurons. Green circles represent the post-synaptic or V1 neurons, while red represent the pre-synaptic or LGN neurons. The spatial structure of the input correlations C is shown below the LGN. The spatial structure of the cortical response kernel, K, of the cortex is shown above V1. Examples of feedforward connections are shown in brown lines, except for the connection between a and x which we have highlighted in purple. Addtionally the spatial extent of the arbor is shown above the cortex.

the post-synaptic population. Feedforward connections are depicted as brown lines from the pre-synaptic population to the post-synaptic population, and the connection from a to x, $W_{x,a}$, is highlighted in purple. Importantly, W grows along the principal eigenvalue of C, and therefore the post-synaptic population again develops selectivity to that feature of the input. Note that only one spatial form of the response kernel is allowed in this formalism. Now, that we have considered many post-synaptic neurons though, the spatial form of K will determine how that selectivity is patterned across the post-synaptic population.

If K has very short-ranged excitation, *i.e.* K approximates a Kronecker-δ (Fig. 3.4-bottom left), then the post-synaptic populations can be thought

of as many independent neurons, each of which will receive some feedforward input governed by Eq. (3.7). Hence, each neuron in the post-synaptic population will develop a random selectivity to the features in C determined by the initial conditions of each individual effective $\mathbf{w}$, and therefore the post-synaptic population develops a S&P organization for all features within C (Fig. 3.4-bottom middle). On the other hand, if K is long-range and positive (which in analogy with recurrent connections we will call positive K excitatory), then nearby cells positively drive each other, such that nearby cells are more likely to respond to the same input, causing those neurons to develop similar feature preferences [22]. Hence for broad excitatory K, the post-synaptic population will develop a smooth map for all features within C (Fig. 3.4-top). We demonstrate that a smooth or S&P organization have developed for a given feature by studying the correlation of preferences to that feature across the post-syanptic population. S&P organizations will have correlation that quickly decays to zero with increasing distance, indicating that neurons develop preferences independent of the preferences of their neighbors, while smooth maps have long-range correlations indicating that clusters of preferences have developed (Fig. 3.4 Right - orange line).

Critically though, these models predict that every feature within C is governed by the same cortical response kernel K, and so all features within C develop the same organization across the post-synaptic population. However, biologically, a mixture of organizations is observed within the same cortex, and so we wish to expand upon these previous models and develop a mechanism under which a mixture of organizations for different features may occur. In particular, we will examine the assumptions these models were based on, and

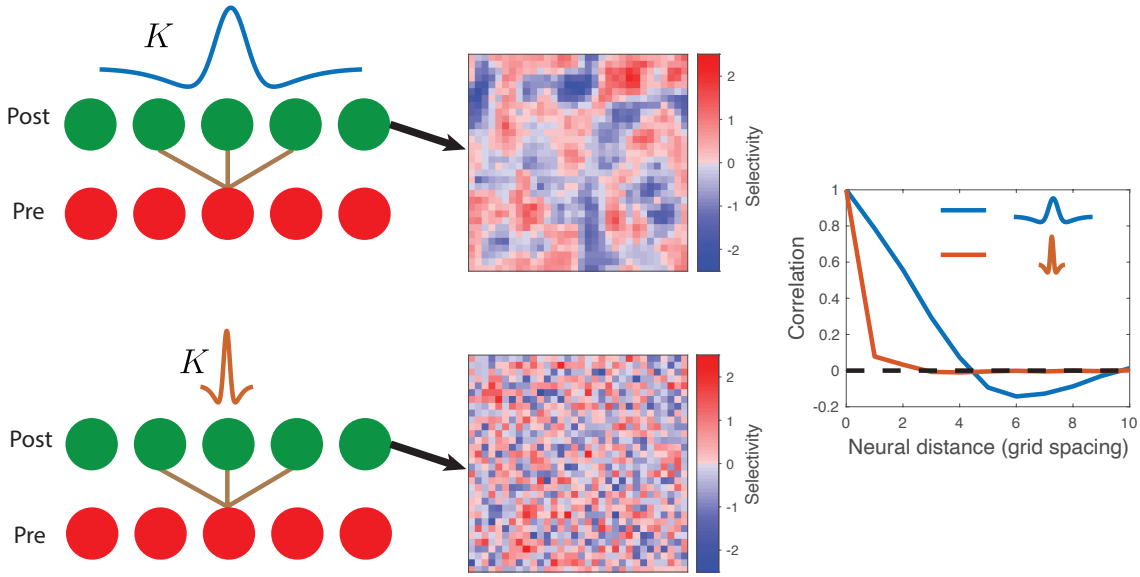

FIGURE 3.4. Previous theoretical models predict either smooth maps for all features of S&P organizations for all features. Shown here are two different feedforward models with different cortical response kernel K. **Left:** For each model, the pre-synaptic population is shown as red circles, and the post-synaptic population is shown as green green cirlces. An example set of feedforward connections, W, between the two are shown in brown. The cortical response kernel,K, of the post-synaptic population is shown above the models. The top model has long-range excitatory response kernel K, and the model on bottom has a very narrow response kernel. **Middle:** After development, both post-synaptic populations develop selectivity to the input feature (heat maps- each pixel shows the selectivity of a post-synaptic neuron), but the organization of that selectivity is very different between the two models. The top model develops a smooth map of the feature selectivity, while the organization of the bottom model develops a S&P organization. **Right:** We demonstrate the difference in the two organizations by studying the correlation of the feature preference across the network. The model with the broadly excitatory kernel (blue line) has long-range correlations across the post-synaptic population indicating that nearby neurons develop similar preferences and hence that feature forms a smooth map across that population. While the model with the narrow response kernel (orange line) has practically no correlations with its neighbors in the post-synaptic population indicating the development of a S&P organization.

relax the assumption that the cortex is "slaved" to the dynamics of the thalamus (assumption III), which is not observed experimentally.

3.2.2. Considering cortical dynamics

Previous models of selectivity or preference organization development relied upon four main assumptions

I Bilinear Hebbian learning: $\tau_W \frac{dW}{dt} = \mathbf{v}(t)\mathbf{u}(t)$

II Linear cortical response: $\tau_v \frac{d\mathbf{v}}{dt} = -\mathbf{v} + W\mathbf{u} + R\mathbf{v}$

III Instantaneous cortical response: $\mathbf{v}(t) = KW\mathbf{u}(t)$

IV Slow plasticity relative to input: $\tau_W \frac{dW}{dt} = KWC$

Here we relax assumption III, which relies on cortical dynamics being much faster than the dynamics of the thalamus $\tau_u \gg \tau_v$. In reality, this assumption is not particularly well-justified, as the dynamics of the cortex happen on the same order of time-scales as the thalamus, $\tau_u \approx \tau_v$. By relaxing this assumption we gain a new "degree of freedom", and here show how features at multiple time-scales couple strongly to response kernels with similar time-scales. This coupling allows preferences to features at one time-scale to develop an organization that most strongly depends on the spatial form of the response kernel at that or a similar time-scale. Thus selectivity to a certain feature may develop a qualitatively different organization across the post-synaptic population from a selectivity to another feature, provided the response kernels at their respective time-scales have different spatial profiles. Thus the post-synaptic population is no longer forced to develop the same organization of feature preferences for all features present in the feedforward input.

Again we will keep most of the assumptions from previous models (I, II, IV) and assume the cortex responds linearly, but we now fully treat its dynamics,

which allows us to make use of multiple time-scales. Assumption II allows us to again use linear response theory, from which we know that the response of the cortex must be a convolution of its weighted inputs. Previously we used this theory to arrive at Eq. (3.9), however, when the dynamics of the cortex at multiple times-scales are considered, the convolution of the weighted inputs is no longer purely spatial, but temporal as well, leading to

$$\mathbf{v}(t) = \int_{-\infty}^{t} d\tau K(\tau) W \mathbf{u}(t-\tau) \tag{3.13}$$

where now $K(\tau)$ is an $N_{post} \times N_{post}$ matrix that describes the linear response kernel at some time-lag τ. As stated previously, K is related to the recurrence in the network, but from here on we model K directly and remain agnostic about the structure of the recurrence in the network.

Given that we still wish to use a bilinear Hebbian learning rule (assumption I), we substitute Eq. (3.13) in Eq. (3.3) and properly treat the temporal components to find

$$\tau_W \frac{dW}{dt} = \int_{-\infty}^{t} d\tau K(\tau) W \mathbf{u}(t-\tau) \mathbf{u}^T(t) \tag{3.14}$$

Again we consider time-scales of plasticity that are much longer than the time-scales of dynamics, $\tau_W \gg \tau_u \approx \tau_v$ (IV). As such, we average Eq. (3.14) over time to find

$$\tau_W \frac{dW}{dt} = \left\langle \int_{-\infty}^{t} d\tau K(\tau) W \mathbf{u}(t-\tau) \mathbf{u}^T(t) \right\rangle_t \equiv \int_{-\infty}^{\infty} d\tau K(\tau) W C(-\tau) \tag{3.15}$$

where $C(\tau)$ is an $N_{pre} \times N_{pre}$ matrix which describes the input correlations at time-delay τ, thus the input correlations also have multiple time-scales. Equivalently, applying Parseval's theorem to get a better sense of how $C(\tau)$ and $K(\tau)$ couple across time-scales, we find

$$\tau_W \frac{dW}{dt} = \int_{-\infty}^{\infty} df \widetilde{K}(f) W \widetilde{C}(f) \tag{3.16}$$

where $\widetilde{K}(f)$ and $\widetilde{C}(f)$ are the (element-wise) Fourier transforms of $K(\tau)$ and $C(\tau)$, respectively. Here we can see that again the post-synaptic population will develop a selectivity to the eigenvectors (principal components) of $\widetilde{C}(f)$, and even though $\widetilde{C}$ will have different features across multiple times-scales, the post-synaptic population will become selective to all of them via Eq. (3.16). Those different features (eigenvectors) at multiple time-scales will couple to the response kernel which has a time-scale most similar to its own. Hence at one time-scale, the spatial form of the response kernel, $\widetilde{K}(f)$, may be broad and excitatory, causing preference to the feature (eigenvector) which shares its time-scale to be encoded in a smooth map across the post-synaptic population, while at a different time-scale, the spatial form of the response kernel may be narrow, leading to the development of a S&P organization for the preference of the feature which matches that time-scale.

Eq. (3.16) is quite general, and may be used for arbitrary response kernels and input correlations. Here, we will limit ourselves to the following forms of the spatiotemporal correlation matrices and response kernels to gain more insight into

the contribution of different correlation and dynamic times-scales.

$$C_{a,a'}(t) = \sum_{j=1}^{M} C^j_{a,a'} e^{-\frac{|t|}{\tau_j}} \tag{3.17}$$

$$K_{x,x'}(t) = \Theta(t) \sum_{i=1}^{S} K^i_{x,x'} e^{-\gamma_i t} \tag{3.18}$$

where $C^i_{a,a'}$ and $K^i_{x,x'}$ are now purely spatial correlations matrices and response kernels for the i^{th} time-scale τ_i or $\frac{1}{\gamma_i}$, respectively. Note that $K(t)$ is restricted to be causal (hence the Heaviside Θ), as it should be. In general, the cortical response kernel $K(t)$ and input correlation $C(t)$ may have different numbers of time-scales, $S \neq M$. Note that by considering the dynamics of the cortex itself, we essentially gain a "degree of freedom" in our Hebbian learning rule, such that the response kernel $K(t)$ may take many different spatial forms across multiple time-scales. Whereas previously in Eq. (3.11), when the spatial form of K had been set for one feature within the input correlation C, it was set for all other features as well. Importantly though, in this new framework the post-synaptic populations will still develop selectivity to the same features in $C(t)$, but now each feature may develop different organizations across the post-synaptic population depending on the spatial form of $K(t)$ at that time-scale.

We can now use eqns 3.17 in 3.15, or equivalently, take their Fourier Transforms

$$\widetilde{C}(f) = \textstyle\sum_j^M \frac{2\tau_j}{|1+2\pi i f \tau_j|^2} C^j \tag{3.19}$$

$$\widetilde{K}(f) = \textstyle\sum_i^S \frac{1}{1+\frac{2\pi i f}{\gamma_i}} K^i \tag{3.20}$$

and use them in Eq. (3.16).

$$\tau_W \frac{dW}{dt} = \sum_i^S \sum_j^M \int_{-\infty}^{\infty} df \frac{1}{1+\frac{2\pi i f}{\gamma_i}} \frac{2\tau_j}{|1+2\pi i f \tau_j|^2} K^i W C^j \tag{3.21}$$

Note that K^i, C^j and W have no frequency dependence, and so can be moved outside the integral. As such, we need only consider the frequency factors under the integral.

Let us consider the integral with respect to frequency in more detail, suppressing the sums and spatial factors. First, we find denominators that are real

$$\int_{-\infty}^{\infty} df \quad \frac{1}{1+\frac{2\pi i f}{\gamma_i}} \frac{1-\frac{2\pi i f}{\gamma_i}}{1-\frac{2\pi i f}{\gamma_i}} \frac{2\tau_j}{|1+2\pi f i \tau_j|^2} \tag{3.22}$$

$$\int_{-\infty}^{\infty} df \quad \frac{1-\frac{2\pi i f}{\gamma_i}}{|1+\frac{2\pi i f}{\gamma_i}|^2} \frac{2\tau_j}{|1+2\pi f i \tau_j|^2} \tag{3.23}$$

The imaginary term in Eq. (3.23) is an odd function with respect to frequency, and since the integral is over all frequencies (from $-\infty$ to ∞), the imaginary contribution to the integral is zero, and we are left with purely real terms (as desired since synaptic strengths are real). Hence to find the growth of W we need to evaluate the following integral

$$\int_{-\infty}^{\infty} df \frac{2\tau_j}{|1+\frac{2\pi i f}{\gamma_i}|^2 |1+2\pi i f \tau_j|^2} = 2\int_0^{\infty} df \frac{2\tau_j}{|1+\frac{2\pi i f}{\gamma_i}|^2 |1+2\pi i f \tau_j|^2} \tag{3.24}$$

As the denominators are the only part of the integrand that depends on frequency, we employ the method of partial fractions to solve this integral

analytically

$$\frac{1}{|1+\frac{2\pi f}{\gamma_i}|^2|1+2\pi f\tau_j|^2} = \frac{A}{1+(\frac{2\pi f}{\gamma_i})^2} - \frac{B}{1+(2\pi f\tau_j)^2} \tag{3.25}$$

and find $A = \frac{1}{1-\tau_j^2\gamma_i^2}$ and $B = \frac{\tau_j^2\gamma_i^2}{1-\tau_j^2\gamma_i^2}$. Recall that $\int_0^\infty dx \frac{1}{1+(ax)^2} = \frac{1}{a}\arctan(ax)|_0^\infty = \frac{\pi}{2a}$, and use that identity in Eq. (3.25)

$$4\tau_j \int_0^\infty df \frac{1}{|1+\frac{2\pi f}{\gamma_i}|^2|1+2\pi f\tau_j|^2} = \int_0^\infty df \frac{A}{1+(\frac{2\pi f}{\gamma_i})^2} - \int_0^\infty df \frac{B}{1+(2\pi f\tau_j)^2} \tag{3.26}$$

$$\rightarrow 4\tau_j \frac{\pi}{2}\frac{1}{2\pi}[\gamma_i A - \frac{1}{\tau_j}B] = \tau_j\gamma_i\big(\frac{1}{1-\tau_j^2\gamma_i^2} - \frac{\tau_j\gamma_i}{1-\tau_j^2\gamma_i^2}\big) \tag{3.27}$$

and simplifying we arrive at

$$\tau_W \frac{dW}{dt} = \sum_i \sum_j \int_{-\infty}^\infty df \frac{1}{1+\frac{2\pi i f}{\gamma_i}} \frac{2\tau_j}{|1+2\pi i f\tau_j|^2} K^i W C^j = \sum_i \sum_j \frac{\tau_j\gamma_i}{1+\tau_j\gamma_i} K^i W C^j \tag{3.28}$$

Eqn 3.28 is quite general and valid for arbitrary spatial shapes of C^j and K^i as well as arbitrary numbers of time-scales, M and S. To demonstrate that our framework enables the development of qualitatively different organizations for different features, we consider the simplest non-trivial case and study the development of selectivities when there exist only two time-scales in both the input correlations, C, and response kernels, K.

3.2.3. Two time-scale model

From here on, we restrict our framework to correlations with two time-scales, τ_{slow} and τ_{fast}, and response kernels with two time-scales $\frac{1}{\gamma_{slow}}$ and $\frac{1}{\gamma_{fast}}$. As such we will assume that the input correlation is now given as

$$C(t) = C^{slow} e^{-\frac{|t|}{\tau_{slow}}} + C^{fast} e^{-\frac{|t|}{\tau_{fast}}} \tag{3.29}$$

and the principal eigenvectors of C^{slow} and C^{fast} will be the features that the post-synaptic population becomes selective for (note C^{slow} and C^{fast} are purely spatial). Likewise, the cortical response kernel of the post-synaptic population also contains only two time-scales

$$K(t) = \Theta(t)K^{slow}e^{-\gamma_{slow}t} + \Theta(t)K^{fast}e^{-\gamma_{fast}t} \tag{3.30}$$

and we further simplify by assuming that $\frac{1}{\gamma_i} = \tau_i$.

Under these simplifying assumptions, Eq. (3.28) reduces to

$$\begin{aligned}\tau_W \frac{dW}{dt} = \Big[\frac{1}{2}K^{slow} &+ \frac{\tau_{slow}}{\tau_{fast}+\tau_{slow}}K^{fast}\Big]WC^{slow} \\ &+ \Big[\frac{\tau_{fast}}{\tau_{slow}+\tau_{fast}}K^{slow} + \frac{1}{2}K^{fast}\Big]WC^{fast}\end{aligned} \tag{3.31}$$

The post-synaptic population will develop selectivity to the principal eigenvectors of C^{fast} and C^{slow}. If the time-scales of those features are widely separated, *i.e.* $\tau_{slow} \gg \tau_{fast}$, then those feature selectivities in the post-synaptic population are able to develop qualitatively different organizations, depending on the spatial form of K^{fast} and K^{slow} as

$$\tau_W \frac{dW}{dt} \approx \Big[\frac{1}{2}K^{slow} + \frac{\tau_{slow}}{\tau_{fast}+\tau_{slow}}K^{fast}\Big]WC^{slow} + \frac{1}{2}K^{fast}WC^{fast} \tag{3.32}$$

Hence the spatial form K^{fast} will govern the organization of selectivity to C^{fast}, while the organization to the selectivity of C^{slow} will be governed by a mix of K^{slow} and K^{fast}.

To control the relative strength of selectivity to the two features, we normalize the C's and the K's according to some norm (see Sec. B), and factor

out the overall norm or strength of the two terms involving the slow and fast features, as two scalars, λ^{slow} and λ^{fast}. That way, the relative strengths of growth along $\mathbf{e}^{slow}$ and $\mathbf{e}^{fast}$ were more easily controlled via λ_{slow} and λ_{fast}. With this convention

, Eq. (3.31) becomes

$$\begin{aligned}\tau_W \frac{dW}{dt} = \quad & \lambda^{slow} \quad \Big[\frac{1}{2}K^{slow} + \frac{\tau_{slow}}{\tau_{fast}+\tau_{slow}}K^{fast}\Big]WC^{slow} \\ & + \quad \lambda^{fast}\Big[\frac{\tau_{fast}}{\tau_{slow}+\tau_{fast}}K^{slow} + \frac{1}{2}K^{fast}\Big]WC^{fast}\end{aligned} \tag{3.33}$$

Note that in the one-dimensional networks as well λ^{slow} and λ^{fast} will also control the relative strengths of the *slow* and *fast* feature, but our reported λ^{slow} and λ^{fast} numerical values should be weighted by a combination of the eigenvalues of K^{slow} and K^{fast} (see Sec. 3.4 for more details).

Previous theoretical models only considered input correlations with one feature

$$\tau_W \frac{dW}{dt} = \lambda KWC \tag{3.34}$$

Therefore to make more appropriate comparisons between our framework determined by Eqn 3.33 and previous theoretical models, we adapt Eqn 3.34 such that there are two features present within the input correlations in this framework as well to

$$\tau_W \frac{dW}{dt} = \lambda^{slow} K^{slow} WC^{slow} + \ell^{fast} K^{slow} WC^{fast} \tag{3.35}$$

where $\ell^{fast} \neq \lambda^{fast}$. However, we do enforce the same spatial form of K^{slow} across both frameworks. We choose the numerical values of λ^{slow}, λ^{fast}, and ℓ^{fast} such that the post-syanptic network develops the same relative strength of selectivity

to their respective features. We refer to Eqn 3.35 as the **static framework** as cortical activity is entirely set by the input it receives [22] and our framework given by Eqn 3.33 as the **dynamic framework**.

Here we make a further assumption that the spatial form of K^{slow} will be such that it promotes the development of smooth maps and as such is long-range and excitatory, while we impose that the spatial form of K^{fast} will be such that it promotes the development of S&P organizations. This is a modelling choice motivated by biology. We expect that at the slower time-scale input signals at one cell will propagate out to and affect the activity of its neighbors, and hence the slow response kernel should be long-range. By the same logic, at the fast time-scale we expect there to be relatively little time for input signals to affect the activity of their neighbors, and hence we expect the fast response kernel to be spatially narrow. However, this a modeling assumption as there exists relatively little data about the response kernels of cortical populations.

Given that we are interested in understanding the development of orientation selectivity, in our model each "cell" or "neuron" more accurately describes a cortical column. As such, it is natural to assume that the fast response kernel might describe some self-excitation of the cortical column, and so we assume it is described by a very local excitatory spatial profile. Whereas, for the slow response kernel, we assume that it is determined by some longer-range connections in our network, and spatially we model it as a Mexican-Hat or difference of Gaussians (DoG) profile, similar to [22]. As long as the "excitatory" component of our DoG response is broad enough, then [22] showed that DoG response kernels will promote the development of smooth maps. However, these are all merely intuitions, as the actual response kernels within the cortex remain unknown. Were

it actually to be the reverse, *e.g.* such that the fast response kernel promotes the development of smooth maps while the slow response kernel promotes the development of S&P organizations, our model would still apply, but the spatial forms of the response kernels should be changed to match experiment.

The only constraints we place on the spatial profiles of the feedforward input correlations are such that their principal eigenvectors of C^{fast} and C^{slow} are orthogonal to one another, which ensures the development of one does not directly interfere with the development of the other. For simplicity we will also consider their spatial forms to also be DoG, and assume that the spatial form of C^{slow} is longer-range than C^{fast}. We assume that the correlations at the slower time-scale are longer-range because we expect nearby LGN neurons to respond similarly on average, while correlations on the faster time-scale should be shorter ranged due to the noise within neural activity (*e.g.* spiking activity). However, if experimental evidence showed that faster time-scale correlations are broader and longer-range, then those spatial forms for the correlation structure could easily be used in our model rather than what we have proposed here. We would expect the cortex to develop selectivity to both features regardless of the actual spatial form of the input correlations.

If W was permitted to grow without bounds, then we expect the development of selectivity to C^{slow} to have no impact on the development of selectivity to C^{fast} as their principal eigenvectors are orthogonal to one another. However, W is not observed to grow without bounds, and so we include an upper limit and a constraint on the strength of synaptic connections. Through this mechanism, the growth of selectivity to C^{slow} and C^{fast} are no longer independent,

but we will show that qualitatively different organizations still develop through our simulations.

3.2.4. Normalization and Arbors

Under the proposed formulation, W would grow without bound, which is not experimentally observed. Rather there are limits as to how strong or weak any particular synapse can become. As such, we include in our model of W an upper and lower bound, such that any element of W can not exceed a maximum value, and impose that every element of W has zero as a lower bound. We chose zero as the lower bound because long-range connections within the brain are typically excitatory. Therefore, the feedforward connection strength W is positive semi-definite.

Once a synapses has hit either the upper or lower bound, that synapse is "frozen", or unable to develop further. Experimentally orientation preference is a relatively stable feature preference across the cortex [70], and by including a freezing mechanism we model the critical period during which orientation preference is set in the cortex [8, 71]. Importantly, we impose that a synapse from one pre-synaptic cell onto a particular post-synaptic cell can only strengthen at the expense of weakening a synapse from a different pre-synaptic cell onto the same post-synaptic cell [22]. Through this mechanism the development of one feature organization can impact the development of the other (see Sec. 3.4.4 for more details).

To better model biology, we include an arbor function to enforce retinotopy on our network as not every LGN neuron communicates with every V1 neuron, and further modeling the development of retinotopy remains outside this work as

retinotopy develops through genetic, rather than activity dependent, plasticity. The arbor function, A imposes a spatial limit on the range of feedforward connections (Fig. 3.5-red line). The arbor is such that feedforward connections from an LGN population to a V1 column are set to zero when they are beyond the spatial limit, and finite within it. As such, Eq. (3.28) becomes

$$\tau_W \frac{dW}{dt} = A \odot \sum_i \sum_j \frac{\gamma_i \tau_j}{1 + \gamma_i \tau_j} K^i W C^j \tag{3.36}$$

where we have signified element-wise multiplication of the arbor A with the summed over time-scale changed in W as $\odot$. Note that the arbor function is a $N_{post} \times N_{pre}$ matrix which only depends on the relative difference in location between the pre- and the post-synaptic neuron.

3.3. Simulation results

Here we simulate the development of selectivity organizations under the formalism developed above considering only two-time scales within the input correlations, $\widetilde{C}(f)$, and cortical response kernels, $\widetilde{K}(f)$. Unless otherwise noted, we impose an arbor function, A, limiting connections from the pre-synaptic to the post-synaptic population which enforces retinotopy in these networks. Moreover, we impose that feedforward connections from the pre- to the post-synaptic population, W, are bounded and normalized as above. We present results from simulations of one-dimensional networks and two-dimensional which show that by treating the dynamics of the post-synaptic (or cortical) population on an equal footing with the pre-synaptic (or thalamic) population, two different features can develop qualitatively different organizations. We first show that our framework

predicts qualitatively different selectivity organizations to develop for the features within C^{slow} and C^{fast} when one-dimensional networks are used to represent both the pre- and the post-synaptic populations. Then we expand our framework to two-dimensional grids for the pre- and post-synaptic populations and show the selectivity to the features within C^{slow} and C^{fast} again develop qualitatively different organizations in this expanded network simulation as well.

Further, we demonstrate our framework is capable of predicting the development of a smooth map for orientation preference when the pre-synaptic population is modeled as consisting of two populations, either ON-center or OFF-center responsive neurons in analogy with the LGN. Our framework improves upon previous models of orientation preference development by predicting more heterogeneous receptive fields for cortical neurons. Moreover, our framework succeeds in predicting biologically plausible receptive fields for the cortical population when input correlations have many time-scales. Whereas the formalism developed previously, when adapted to include a correlation structure with multiple features, predicts receptive fields that are very weakly selective to orientation. Importantly though, our new framework succeeds in predicting the development of qualitatively different organizations for different features within the same cortex, which had not been done before.

3.3.1. One-dimensional network results

We have simulated the development of feedforward plasticity between two one-dimensional rings of neurons (Fig. 3.5 - circles), representing a pre-synaptic population and a post-synaptic population. In analogy with the visual system,

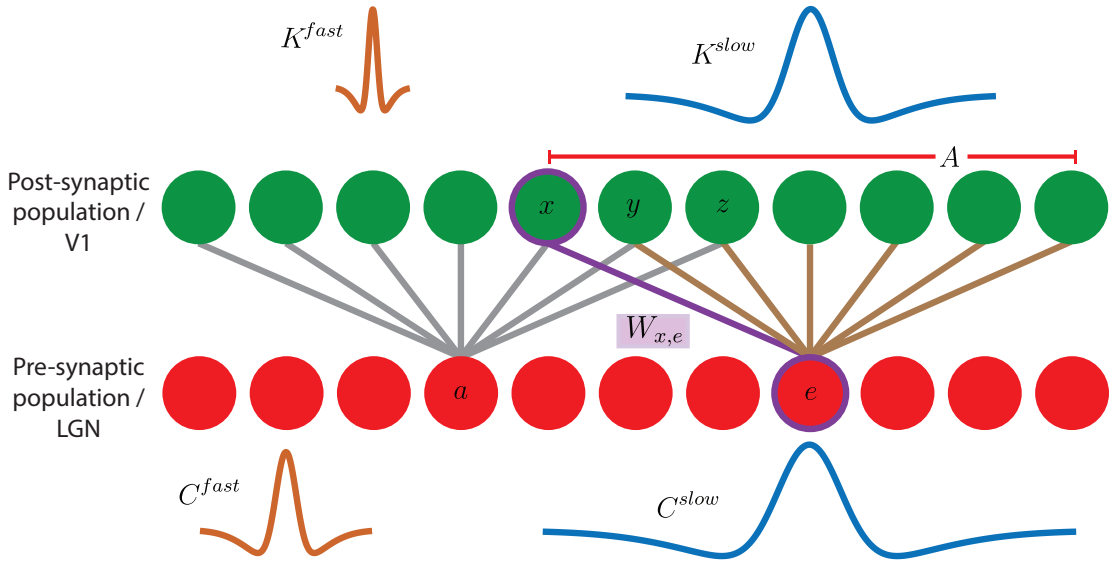

FIGURE 3.5. Schematic of one-dimensional networks, representing the LGN or pre-synaptic population (red circles, early alphabet), and V1 or post-synaptic population (green circles, end of alphabet). Shown below the LGN are the spatial forms of the correlation at the slow (blue) and fast (orange) time-scale, similarly, the response kernels are shown arbor V1. The arbor, the spatial extent of connections from the LGN to V1 is shown as a red line above a set representative connections (brown lines). Two sets of representative connections are shown from pre-synaptic neurons a (gray) and e (brown). The connection,$W_{x,e}$ from LGN cell e to V1 cell x is shown in purple as an example of an element of W.

we consider the pre-synaptic population to be the LGN and the post-synaptic population to be V1.

Note that we only modeled the development of feedforward connections from the LGN to V1, and did not model the development of recurrence within V1. Rather, we assumed that the spatial profiles of the response kernels within V1 remain fixed over time, and dominated by two time-scales, one fast and slow. We further chose that the spatial extent of the slow kernel, K^{slow}, was long range and had a DoG form (Fig. 3.5 - blue curve above V1), while the spatial extent of the

fast kernel, K^{fast}, was very narrow and approximated as a Kronecker-δ function (Fig. 3.5 - orange curve above V1).

Likewise, we assumed the input correlation structure $\widetilde{C}(f)$ only contained the same two time-scales as the response kernel $\widetilde{K}(f)$, however, we considered two cases of spatial correlation structure. The simple case had input correlations with only one eigenvector each (not pictured in Fig 3.5), while for the other we assume the spatial correlations were DoGs at both time-scales (Fig. 3.5 - curves below LGN). We imposed no sense of retinotopy in the simple case for the input correlations, rather we use this case to check our intuition about how selectivity should grow under normalization. To that end, we removed possible confounds, such as the arbor function or the possibility of non-principal eigenvectors of C^{slow} being aligned with the principal eigenvector of C^{fast} or vice versa.

For the simple single eigenvector case, we assumed that C^{slow} and C^{fast} were projection matrices, *i.e.* $C^j = \mathbf{e}^j(\mathbf{e}^j)^T$ with $j \in [slow, fast]$. In this simple case, C^{slow} and C^{fast} are single-rank and composed of only one linearly independent eigenvector. Moreover, $\mathbf{e}^{slow}$ and $\mathbf{e}^{fast}$ were randomly chosen N_{pre}-dimensional vectors, conditioned such that $\mathbf{e}^{slow}$ and $\mathbf{e}^{fast}$ are orthogonal (*i.e.* $\mathbf{e}^{fast} \cdot \mathbf{e}^{slow} = 0$). For this simple case then, we ensure that the only features our post-synaptic population becomes selective for are those eigenvectors. When we consider more realistic DoG C^j, then both C^{slow} and C^{fast} contain many eigenvectors. We choose their spatial shapes such that the principal eigenvectors are orthogonal to one another, however, there exist many other eigenvectors within C^{slow} and C^{fast} when realistic spatial shapes are used which may or may not be orthogonal. Hence for this simple case, we choose $C^j = \mathbf{e}^j(\mathbf{e}^j)^T$ to ensure the growth of C^{fast} in no way impacts the growth of C^{slow}, until the bounds of W are reached. Moreover, we

also *do not* impose an arbor function on this condition. This is the only model we present results for without an arbor. However, We do still impose constraints and normalization on W. In this simple model, we ensure that the only impact to the growth of selectivity to both the $fast$ and $slow$ feature, outside of Eq. (3.31), is the normalization and constraints on W. Therefore, it is a good control to test our intuition about how selectivity ought to develop to these features under normalization.

Then, we model C^{slow} and C^{fast} as depending on the relative distance between pre-synaptic neurons, and to be multi-rank. We modeled the spatial form of the slow time-scale correlations, C^{slow}, as being much broader than the spatial form of the fast time-scale correlations C^{fast} (Fig. 3.5A, curves below LGN). To enforce retinotopy in this network, we restricted the spatial range of possible connection from the LGN population by imposing an arbor, A (Fig. 3.5 - red line), on feedforwad connections, W (Fig. 3.5-brown and gray lines).

We then simulated Eq. (3.31) according to the algorithm outlined in [22] for these two conditions on the input correlation structure (see Sec. 3.4.5). In particular, we simulated the growth of the feedforward connection strength W until $\geq$ 90% of the synapses within the arbor saturated, meaning 90% of the synapses had values at either the upper or lower bound. To study how selective the post-synaptic population became after development, we modeled the selectivity of a post-synaptic cell as the inner product of its individual $\mathbf{w}_x$ (which can be thought of as its RF filter $\mathbf{F}_x$) with the principal eigenvector of the fast or slow input correlation, $\mathbf{e}^i$ where $i \in [slow, fast]$. Thus the selectivity to the i^{th} feature at position x, S^i_x, can be found across the whole post-synaptic population, *i.e.* for

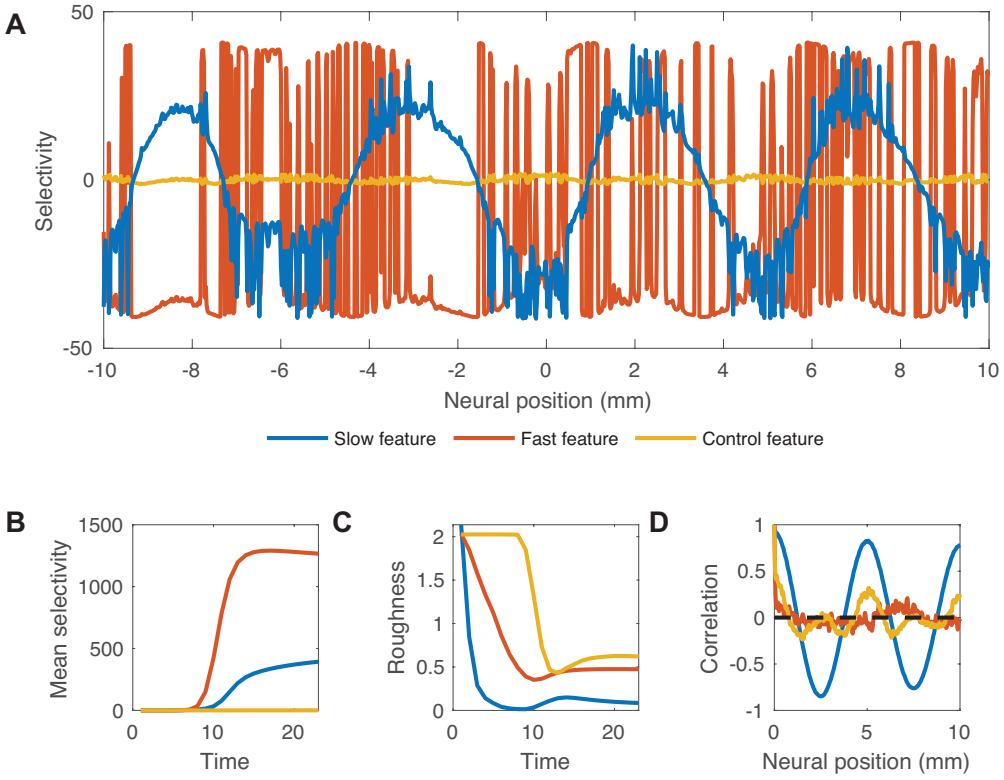

FIGURE 3.6. One-dimensional network with two time-scale feature organization and development for simplest input correlations, *i.e.* projection operators. Note that an arbor was not used, and LGN neurons communicated with all V1 neurons. **A:** The selectivity of every post-synaptic neuron is shown to the slow feature (blue), fast feature (orange), and the control(yellow). Selectivity is defined as $W \cdot \mathbf{e}^i$ with $i \in [slow, fast, control]$. Note that periodic boundary conditions are used so the cell at -10 is next to the cell at 10. **B:** The time evolution of the mean selectivity to the slow (blue), fast (orange), and control (yellow). **C:** The time evolution of the nearest-neighbor difference (Roughness) of the selectivity to the slow (blue), fast (orange), control (yellow). A higher value indicates a less smooth map. **D:** The spatial correlation of the feature selectivity to the slow (blue), fast (orange), control (yellow) features across the network.

all x, by

$$\mathbf{S}^i = W \cdot \mathbf{e}^i \tag{3.37}$$

This defines selectivity because it represents how strong the input would be to a post-synaptic neuron when the feature $\mathbf{e}^i$ is present. If selectivity is a large and positive value, that means the post-synaptic neuron would receive a high input whenever the input activity matches $\mathbf{e}^i$, and thus respond very strongly. Hence that post-synaptic neuron would prefer activity aligned with $\mathbf{e}^i$. However, if the selectivity is a large and negative value, the post-synaptic neuron would respond most strongly and therefore prefer when input activity is anti-aligned to $\mathbf{e}^i$. The post-synaptic neuron would be un-selective for the feature, meaning its response would not change much depending on how input activity aligns or anti-aligns with $\mathbf{e}^i$, if the selectivity was zero.

Recall that we define orientation preference by the stimulus orientation which elicits the greatest response in the cortex. In a similar way, after development, we expect the post-synaptic population to develop a preference for the input either aligning or being anti-aligned with $\mathbf{e}^i$, regardless of the form used for C^i. The post-synaptic neuron at x either prefers the feature when $\mathbf{w}_x \cdot \mathbf{e}^i > 0$ or the negative of the feature when $\mathbf{w}_x \cdot \mathbf{e}^i < 0$, or is unselective for it when $\mathbf{w}_x \cdot \mathbf{e}_i \approx 0$.

For our one-dimensional networks, we studied 1024 neurons in both the pre-synaptic population and the post-synaptic population, ranging across positions -10 μm to 10 μm. We are simulating the development of feedforward thalamocortical connections, as such our initial conditions were such that all elements of $W \in [0.8, 1.2]$ randomly and uniformly. Under this initial condition, we are imposing that the feedforward thalamocortical connections be excitatory as observed. Imposing this excitatory condition though, introduces an artefact to our data, as W has some nonzero mean. For the results shown in this subsection,

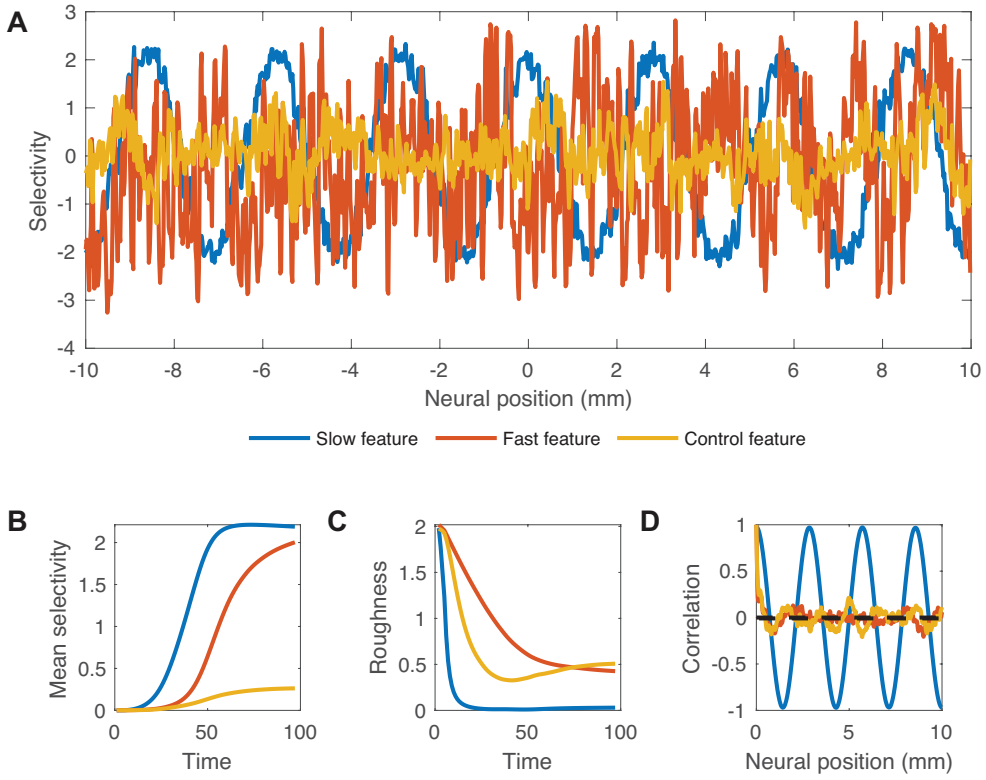

FIGURE 3.7. One-dimensional network with an arbor function (retinotopy), two time-scale feature organization and development. **A:** The selectivity of every post-synaptic neuron is shown to the slow feature (blue), fast feature (orange), and the control(yellow). Selectivity is defined as $W \cdot \mathbf{e}^i$ with $i \in [slow, fast, control]$. Note that periodic boundary conditions are used so the cell at -10 is next to the cell at 10. **B:** The time evolution of the mean selectivity to the slow (blue), fast (orange), and control (yellow). **C:** The time evolution of the nearest-neighbor difference (Roughness) of the selectivity to the slow (blue), fast (orange), control (yellow). A higher value indicates a less smooth map. **D:** The spatial correlation of the feature selectivity to the slow (blue), fast (orange), control (yellow) features across the network.

we removed the mean from W before calculating any related quantity, *e.g.* the selectivity to certain features.

In Figs 3.6 we present results for the single-rank input correlation structure. Recall that the input correlation used in Fig. 3.6 was chosen to ensure no

confounds in the development of W. Therefore, we chose C^j to be single-rank and imposed no arbor to ensure that selectivity to both $\mathbf{e}^{slow}$ and $\mathbf{e}^{fast}$ develop under normalization. Furthermore, we expect both to develop an organization which depend on the spatial form of K^i that they couple most strongly to, *e.g.* we expect $\mathbf{e}^{slow}$ to develop a smooth map while $\mathbf{e}^{fast}$ develops a S&P organization. In panel A of Fig 3.6, we show the selectivity for to the slow feature, $\mathbf{e}^{slow}$, in blue and the selectivity to the fast feature, $\mathbf{e}^{fast}$, in orange across the post-synaptic population. We further studied the selectivity to a feature which was not present in our input correlation, $\widetilde{C}(f)$, which we used as a control. The feature $\mathbf{e}^{control}$ was chosen to be orthogonal to be $\mathbf{e}^{slow}$ and $\mathbf{e}^{fast}$. As long as selectivity to the control feature stays low relative to the selectivity of the fast and slow features, then we know that the network did not just become selective to all features, instead it becomes selective for just the features in the input. Recall that in both the projection and the spatially dependent correlation cases the only constraint we made on the forms of C^{slow} and C^{fast} was that their principal eigenvectors are orthogonal, thus $\mathbf{e}^{slow}$ is orthogonal to $\mathbf{e}^{fast}$ and $\mathbf{e}^{control}$. Qualitatively, the post-synaptic population did become selective to both the slow feature and the fast feature, but stayed relatively un-selective to the control feature for both of the considered input correlation conditions.

The selectivity to the control feature did not substantially grow over time for either input correlation condition (Figs. 3.6B), but did show some change due to the normalization condition on W (not shown are results from unbounded simulations of W when the control feature did not change over time). Importantly, though, the selectivity to both the slow and the fast feature did grow over time, and, moreover, developed qualitatively different

organizations across the post-synaptic population regardless of the choice of input correlation structure. Preference for the smooth feature becomes encoded across the post-synaptic population in a regular repeating pattern with a relatively small amount of variation. Clear clusters form for preferences where the input activity pattern matches the eigenvector, $\mathbf{e}^{slow}$, versus the anti-aligned pattern, $-\mathbf{e}^{slow}$. Importantly, preference for the fast feature has no discernible pattern and neighbors are just as likely to prefer input activity aligned to the positive eigenvector $\mathbf{e}_{fast}$ as its negative $-\mathbf{e}_{fast}$.

We quantified how different the two organizations were using the nearest-neighbor metric, meaning we studied differences in preference for the same feature normalized by the mean selectivity to that feature between neighboring units in our post-synaptic population which we labeled as "Roughness" (Fig. 3.6C).Furthermore, we studied how correlated feature preferences were across the one-dimensional network (Fig. 3.6D).We expect smooth maps to have a low value of roughness as nearest neighbors prefer similarly valued features and so their difference will be low, while in a S&P organization the roughness value will be large because nearest neighbors will have no correlation in their preferred feature values and so the difference in their preferred values will be high. Similarly, for a smooth map we expect to see the feature correlation to be long-range across the post-synaptic population, while for a S&P organization, we expect the correlation to quickly fall-off to zero and be uncorrelated across the cortex.

We found that while all were maximally rough at the start of development, the fast feature had clearly developed a much "rougher" organization than the slow feature post-development. Further, we found preferences for the slow feature to be highly correlated even at large spatial distances, while the preferences for the fast

feature sharply dropped off and stayed near zero across the network. Neighboring units therefore were equally likely to prefer anti-aligned fast features as aligned fast features, emblematic of a S&P organization, while neighboring post-synaptic neurons are highly likely to prefer similarly aligned slow features indicating that a smooth map developed. Thus the preference for the slow feature within the correlation, $\mathbf{e}^{slow}$, organizes as a smooth map, while the preference for the fast feature within the correlation, $\mathbf{e}^{fast}$, develops a S&P organization across the post-synaptic population.

When we consider more realistic input correlations structures (Fig. 3.5 - curves below LGN), the above results hold. In Fig. 3.7A, we show the selectivity which develops across our post-synaptic population when an arbor and DoG spatial forms for C^{slow} and C^{fast} were used. Recall that we choose their spatial shapes such that the principal eigenvectors $\mathbf{e}^{slow}$ and $\mathbf{e}^{fast}$ are orthongoal to one another. However, in this DoG case, the input correlations are multi-rank, and so not every eigenvector within C^{slow} may be orthogonal to C^{fast}. Moreover, in this case, we also impose some notion of retinotopy through an arbor function, A (Fig. 3.5 - red line), wherein we restrict the range of connections from LGN to V1. Even with multi-rank input correlations an arbor, the post-synaptic population develops a selectivity to $\mathbf{e}^{slow}$ that is much more clustered then the selectivity to $\mathbf{e}^{fast}$, and moreover becomes more selective to $\mathbf{e}^{fast}$ and $\mathbf{e}^{slow}$ than the control feature, $\mathbf{e}^{control}$.

We quantify the development and organization of these selectivity structures in the same way as before. Importantly, the selectivity to the *slow* and $fast$ feature become higher on average than the *control* feature (Fig 3.7B). The preference to the *slow* feature develops a smooth map as shown by having a low

roughness, and long-range correlation, while the preference to the $fast$ feature develops a S&P organization as shown by its high roughness and correlation which quickly drops to zero (Fig 3.7C-D). Thus even with more realistic DoG input correlations and an arbor function, qualitatively different feature preference organizations develop across the post-syanptic population.

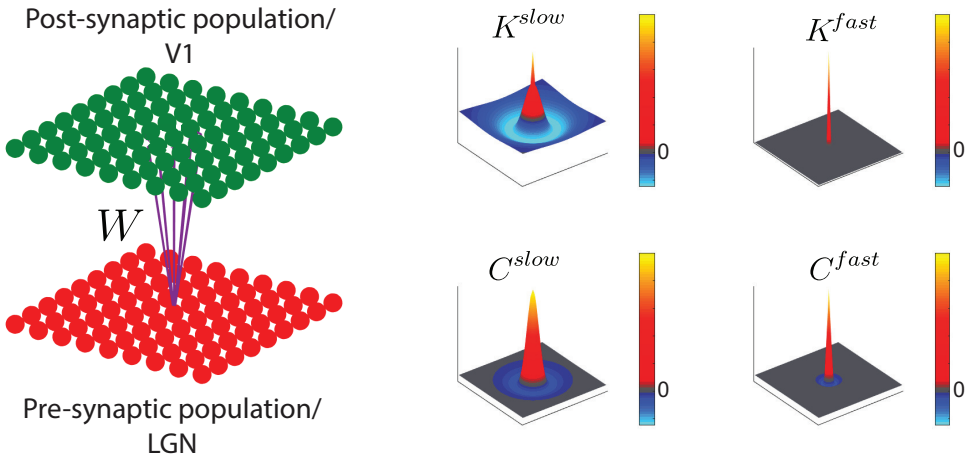

FIGURE 3.8. Schematic of two-dimensional grid networks with spatial forms of the correlation and response kernels. Shapes on the left represent the LGN or pre-synaptic population (red spheres), and V1 or post-synaptic population (green spheres). Some representative connections from the LGN to V1 are shown in purple, W. The spatial forms of K^{slow}, K^{fast}, C^{slow}, and C^{fast} are shown on the right as surface plots with blue-to-cyan indicating negative values and red-to-yellow indicting positive values.

3.3.2. Two-dimensional network results

Here we show that this formalism can be expanded to two-dimensional grids of neurons and the same result holds (Fig. 3.8 - left); the post-synaptic population develops qualitatively different selectivity organizations to the slow and fast feature when expanded. Again, we impose an arbor function such that retinotopy is enforced within the network. We maintain our simplifying assumption that there

exist only two time-scales in the input correlation structure and response kernel, but now we expand their spatial forms to be two-dimensional (Fig. 3.8 - right).

The spatial forms of C^{fast} and C^{slow} are again chosen to be DoG structures with orthogonal principal eigenvectors, where we have assumed that the spatial extent of C^{slow} is broader than C^{fast}. Similarly, we model the response kernels within the post-synaptic population as being two-dimensional DoG structures, where K^{slow} is assumed to be much broader than K^{fast}. We again choose the spatial form of K^{fast} to approximate a Kronecker-δ. We also assume that the slow time-scale is much much longer than the fast time-scale such that the the development of the organizations to each feature is dominated by the similar time-scale response kernel.

Again we simulate the development of the feedforward connection strengths W until 90% of the synapses saturate, or have values that are either the upper or lower bound of W. Then we study how selective the neurons across the post-synaptic population became to the slow and fast feature. The network of post-synaptic units develops selectivity to both the fast and the slow feature (Fig. 3.9A-B), but does not become selective for the control feature (Fig. 3.9C). Each heat map within Fig. 3.9 shows the full network of the post-synaptic population with each pixel representing the selectivity of the unit at that location. The coloring of the pixels denotes the selectivity to input activity either aligned with $\mathbf{e}^i$ (red) or anti-aligned with $\mathbf{e}^i$ (blue). If a neuron is unselective for input aligning with that feature, then its selectivity is zero (white). We again choose the control feature to be orthogonal to both the slow and the fast feature (Fig. 3.9C), and the post-synaptic network remains unselective for that feature. Thus the post-synaptic

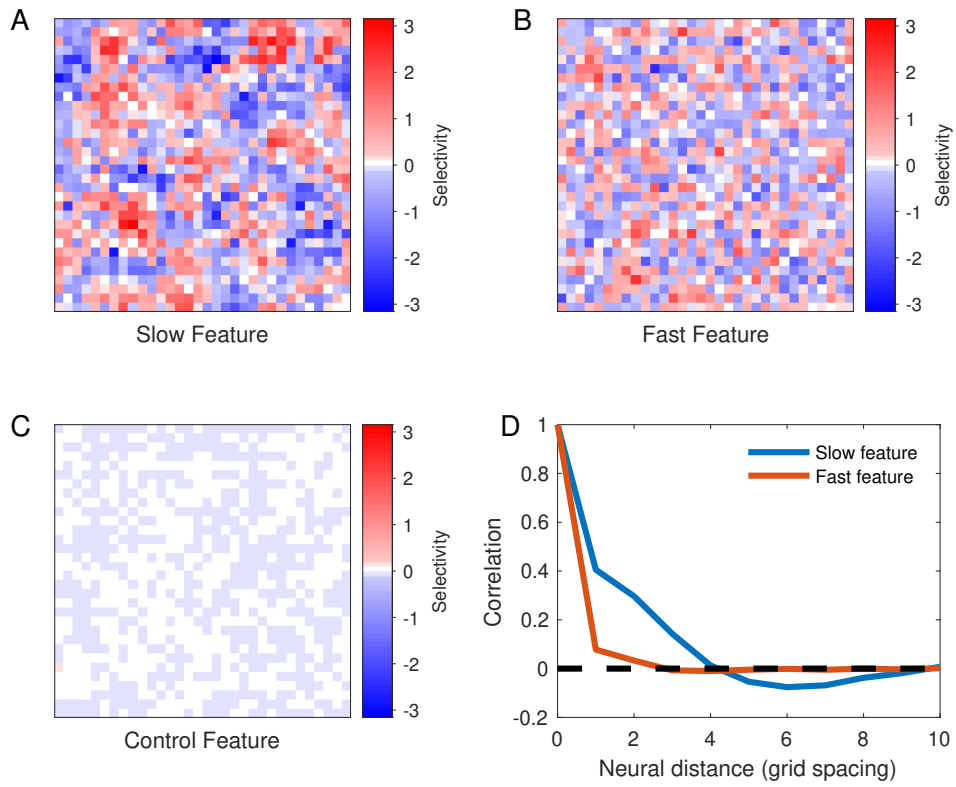

FIGURE 3.9. Two-dimensional, two time-scale cortical dynamic model develops qualitatively different feature selectivity organizations to the slow and fast input feature. Here each pixel represents one neuron in our post-synaptic population. This network was simulated with the cortical dynamics being treated equally with the thalamic dynamics. **A:** Heat map of the selectivity to the slow feature. **B:**Heat map of the selectivity to the fast feature. **C:** Heat map of the selectivity to the control feature. **D:** The average correlation across space for the selectivity to the slow feature (blue), and the fast feature (red). Not shown is the average correlation across space for the selectivity to the control feature since the selectivity did not significantly develop

population does not become selective to all features, but only the features within the input correlation structure.

To demonstrate that the slow feature and fast feature develop different organizations across the post-synaptic population, we study the spatial correlations of the different feature selectivities across the post-synaptic population. We

have neglected studying the organization of the control feature since the post-synaptic population remains unselective for it after development. The slow feature selectivity is positively correlated across several post-synaptic cells, in any direction. Moreover the selectivity to the slow feature becomes slightly anti-correlated with neurons that are located in the negative lobe of the DoG form for K^{slow}. Hence the slow feature selectivity of one unit is correlated to its neighbors indicating that a smooth map for the selectivity to the slow feature has developed. Whereas, the correlation of fast feature preferences quickly drops to zero beyond its auto-correlation (Neural distance = 0). Hence the selectivity to the fast feature has low correlations with its neighbors, and thus the fast feature preference of each post-synaptic cell is independent of its neighbors. Therefore the fast feature is encoded in a S&P organization.

Here again our model develops qualitatively different selectivity organizations to different features within the same network. Showing that when cortical dynamics are treated properly the post-synaptic population develops different organizations to different features. Whereas, if we adapt the previous formalism to include multiple time-scales in the correlation input structures, while still treating the cortical neurons as being "slaved" to their inputs, *e.g.* Eqn 3.35, the feature selectivities for both time-scales are encoded in the same organization (Fig. 3.10A-B). Similar to Fig. 3.9, the whole network is represented as pixels with coloring corresponding to selectivity of a particular feature. Here, again, the post-synaptic population remains un-selective to the control feature. However, the slow and the fast feature have very similar correlations across the cortex. Moreover, the feature preference to both features is highly correlated with its neighbors, indicating that smooth maps develop for both features (Fig. 3.10D). Hence the

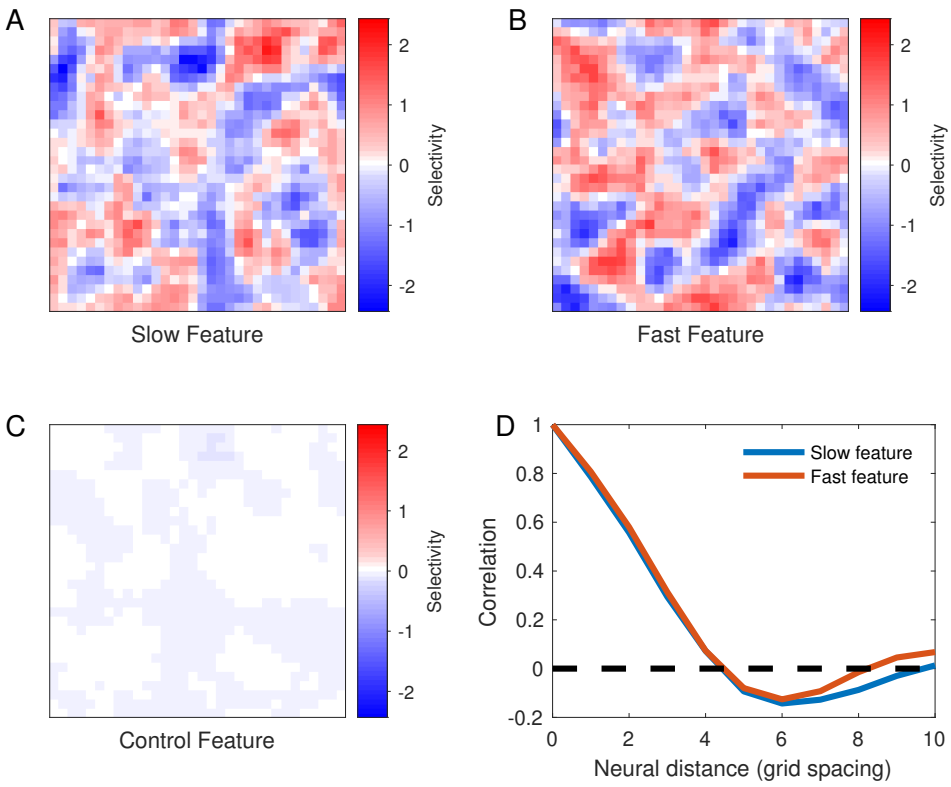

FIGURE 3.10. Two time-scale two-dimensional model without cortical dynamics develops smooth maps of selectivity to both the slow and fast feature. Here each pixel represents one neuron in our post-synaptic population. This network was simulated using the formalism outlined in [22], and so cortical dyanmics are "slaved" to the thalamic inputs. **A:** Heat map of the selectivity to the slow feature. **B:**Heat map of the selectivity to the fast feature. **C:** Heat map of the selectivity to the control feature. **D:** The average correlation across space for the selectivity to the slow feature (blue), and the fast feature (organge). Not shown is the average correlation across space for the selectivity to the control feature since the selectivity did not significantly develop.

static framework which does not fully consider the cortical dynamics fails to predict the development of qualitatively different spatial organizations across the cortex, while our framework successfully develops two qualitatively different organizations across the cortex.

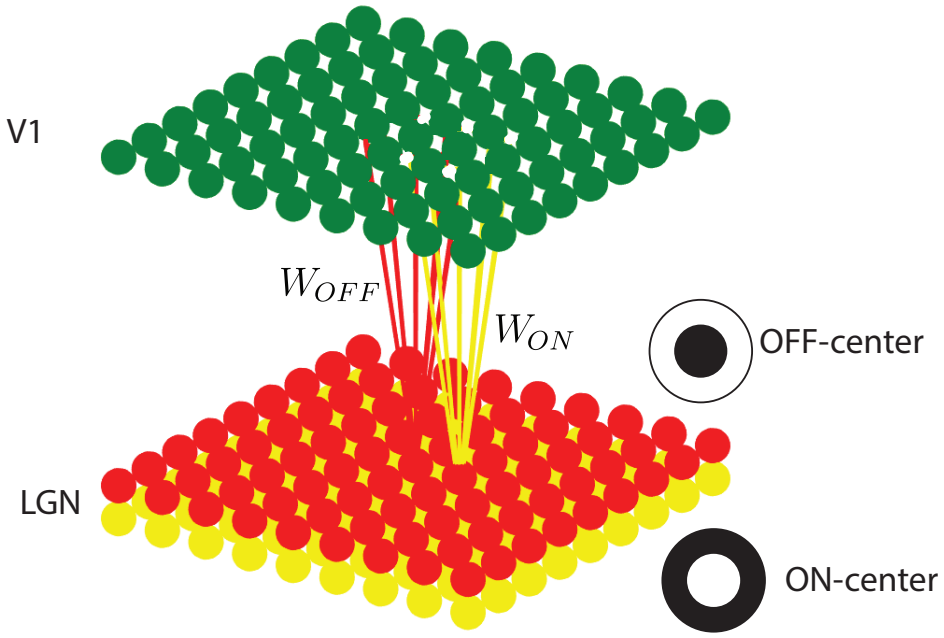

FIGURE 3.11. Here we show a schematic of the expanded LGN population to two networks consisting of ON-center (yellow spheres) and OFF-center neurons (red spheres). A representative RF filter is shown for each population type. We also show some representative connections from the ON-center W_{ON} (yellow lines) and OFF-center W_{OFF} (red lines), to V1 (green spheres).

3.3.3. V1 feature development

So far we have considered rather abstract and mathematical features being encoded in the post-synaptic population, but this work was motivated to explain the co-presence of heterogeneity in receptive fields with smooth feature preference maps, *e.g.* orientation preference develops a smooth map in the primary visual cortex, while the RFs themselves show broad heterogeneity. In order to study the development of receptive field features directly, we expand our model to include ON-center and OFF-center populations in the LGN. We now consider the LGN to consist of two networks, one consisting of neurons that respond to ON-center stimuli and another population which responds to OFF-center stimuli (Fig. 3.11).

We model the LGN populations such that at each retinotopic location there exist corresponding complementary pairs of neurons, one for each population, *e.g.* there exists one ON-center and one OFF-center neuron at each retinotopic location. We thus treat the LGN as having a shared position index between the populations, corresponding to retinotopic location. By expanding to multiple LGN populations, we must consider correlations within population and between them, similarly feedforward connections in this expansion originate from one specific population but terminate at the same post-synaptic population (V1).

Hence we now consider feedforward connections which originate from one specific population, and going forward we denote this population as a subscript *e.g.* W_{ON}. Consider the development of feedforward connectivity from the ON-center population, W_{ON}.

$$\tau_W \frac{dW_{ON}}{dt} = \mathbf{v}\mathbf{u}_{ON}^T \tag{3.38}$$

We assume that both LGN populations have the same number of neurons, N_{pre}, and hence $\mathbf{u}_{ON}$ and $\mathbf{u}_{OFF}$ are both N_{pre}-dimensional. Note that $\mathbf{v}$ remains an N_{post}-dimensional vector without a subscript since it does not belong to either population within the LGN; it does, however, receive input from both populations.

Retaining assumption II, a linear response cortex, the activity of the cortex is still given as a decay to a resting state plus a sum of its weighted inputs, but now those inputs originate from two different populations as

$$\tau_v \frac{d\mathbf{v}}{dt} = -\mathbf{v} + W_{ON}\mathbf{u}_{ON} + W_{OFF}\mathbf{u}_{OFF} + R\mathbf{v} \tag{3.39}$$

Again we rely on linear response theory where we know that the activity of the cortex is a convolution across time and space of its weighted inputs with its linear

response kernel at time-lag τ

$$\mathbf{v}(t) = \int_{-\infty}^{t} d\tau K(\tau)\big[W_{ON}\mathbf{u}_{ON}(t-\tau) + W_{OFF}\mathbf{u}_{OFF}(t-\tau)\big] \tag{3.40}$$

Substituting back into Eq. (3.38),

$$\tau_W \frac{dW_{ON}}{dt} = \int_{-\infty}^{t} d\tau K(\tau)\big[W_{ON}\mathbf{u}_{ON}(t-\tau)\mathbf{u}_{ON}^T(t) + W_{OFF}\mathbf{u}_{OFF}(t-\tau)\mathbf{u}_{ON}^T(t)\big] \tag{3.41}$$

and again applying assumption IV that synaptic plasticity is slow with respect to the dynamics of activity, we find

$$\tau_W \frac{dW_{ON}}{dt} = \int_{-\infty}^{\infty} d\tau K(\tau)\big[W_{ON}C_{ON,ON}(-\tau) + W_{OFF}C_{OFF,ON}(-\tau)\big] \tag{3.42}$$

where now $C_{p,q}(\tau)$ represents the correlation between populations $p, q \in [ON, OFF]$ at time-delay τ. Similarly the development for feedforward OFF-center population connections is given by Eq. (3.42) with $ON \rightarrow OFF$ and vice versa.

Following the formalism outlined in Sec 3.2.2, we find that feedforward connectivity from a population p (where $p \in [ON, OFF]$) is generally governed by

$$\tau_W \frac{dW_p}{dt} = \sum_i \sum_j \frac{\tau_j \gamma_i}{1 + \tau_j \gamma_i} K^i\big[W_p C_{p,p}^j + W_q C_{q,p}^j\big] \tag{3.43}$$

where $q \neq p$, and i, j again correspond to different time-scales within our model. Note that $C_{p,p}^j$ or $C_{p,q}^j$ are again the purely spatial forms of the correlations inputs at time-scale j, and similarly K^i is the purely spatial response kernel at time-scale i. We simplify Eq. (3.43) by assuming that $C_{p,q} \propto -C_{p,p}$ when $q \neq p$ such that the same feature (eigenvector) is shared between correlations within and between populations. The post-synaptic population will therefore again develop selectivity

to the features encoded in the input correlation structures, and moreover the organization of those selectivities will be determined by the spatial structure of the response kernel of the cortex at that time-scale.

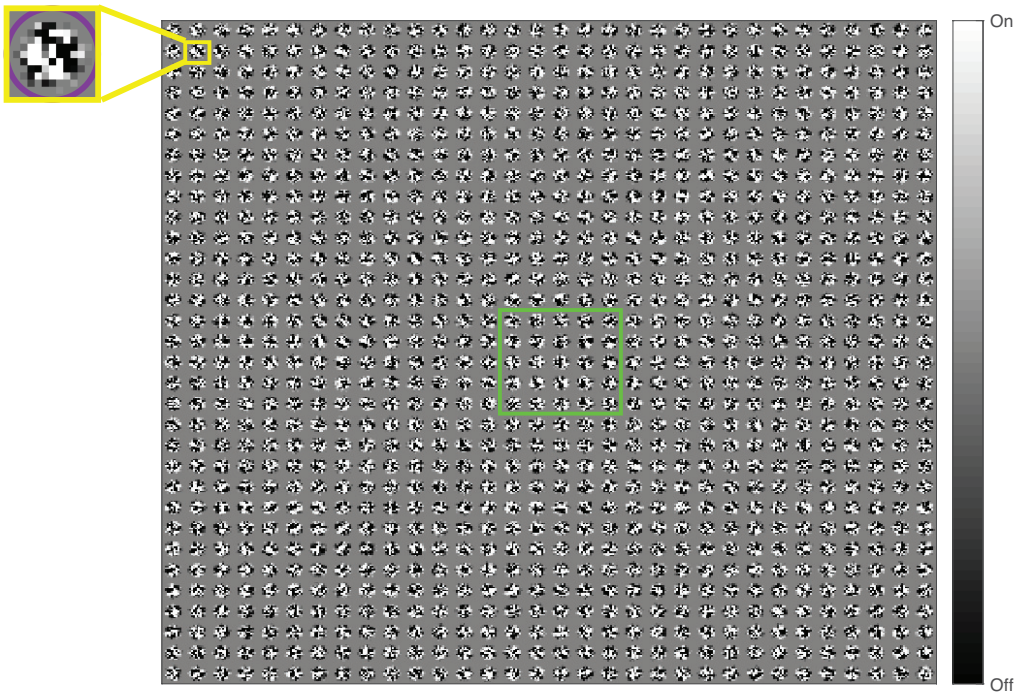

FIGURE 3.12. Receptive fields of 2D V1 population. The receptive field (RF) of each cortical neuron are shown on a grid. Shown for each cortical neuron are all connections within an arbor. An example is shown in the yellow square, each post-synaptic neuron has a corresponding square which has dimensions of $D_A \times D_A$ pixels (D_A is the arbor diameter). The arbor is shown as a purple circle. Each pixel shows the connection from the LGN populations to the corresponding cortical neuron dominated by ON-center (white) or OFF-center (black) responsive population. The green box highlights the example region used in later plots.

In addition to selectivity to the eigenvector features of the input, our model will also develop biologically realistic RFs (Fig. 3.12). We study these RFs directly through $W \equiv W_{ON} - W_{OFF}$. In Fig. 3.12, we show the values of W for each post-synaptic neuron. For each post-synaptic neuron, we show a $D_A \times D_A$

square (Fig. 3.12-yellow square) of connections from the LGN, with D_A the arbor diameter. Each pixel shows feedforward synaptic strengths that the corresponding post-synaptic neuron receives, that are either dominated by the ON-center (white) or OFF-center (black) population. Note that gray pixels represent zero feedforward strength, which means that the connections at that pixel is either outside the arbor, or receives equally strong inputs from both populations.

Recall that previously we imposed a normalization condition on the development of W such that a synapse from the pre-synaptic population to one cortical cell could only grow at the expense of weakening a different synapse, *i.e.* the total synaptic strength onto a post-synaptic cell remains fixed across time. Now, however, that means that connections from the LGN populations may grow at the expense of synapses within their population but also by weakening connections from the other population as well. The most common outcome of this competition is that one population dominates over a patch of retinotopic locations, resulting in the post-synaptic population becoming selective for either a light (ON-center dominates) or dark (OFF-center dominates) patch at that retinotopic location (Fig. 3.12), and indeed that is what we see. As a result, feedforward inputs matching those RF patches dominated by connections from ON- or OFF-center responsive neurons will drive cortical neurons to respond strongly; hence the post-synaptic population develops selectivity to particular patches of light and dark bands (Fig. 3.13). Thus the cells within our post-synaptic population now develop selectivity or preference to particular orientations of the stimulus. We first show that the networks developed under our dynamic framework and the static framework develop selectivity to the particular orientations (Fig. 3.13), and then study the organization of that selectivity across the cortex (Fig. 3.14).

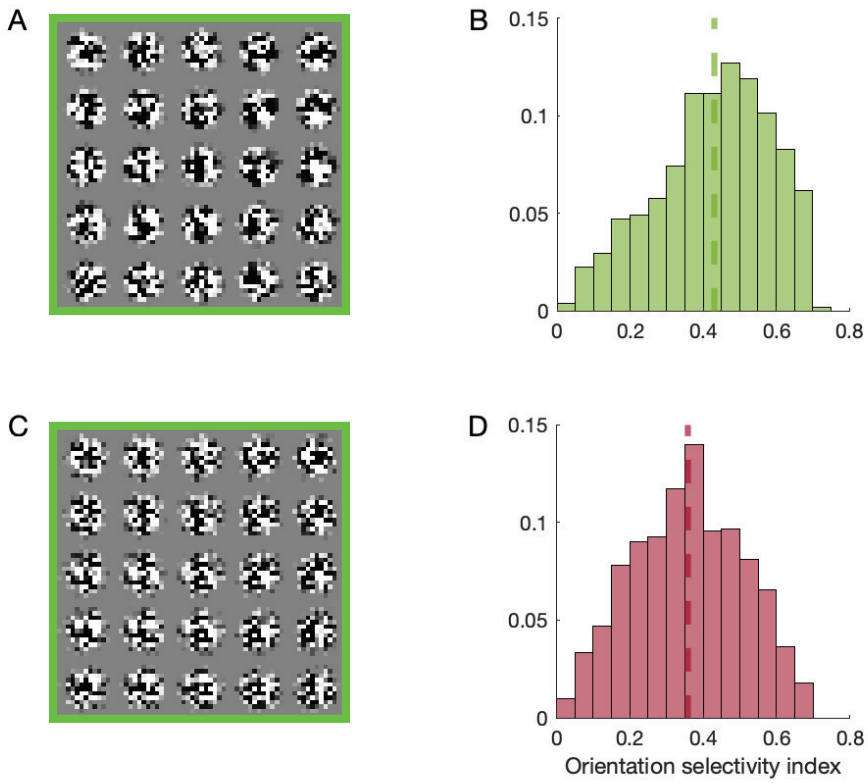

FIGURE 3.13. Orientation selectivity in a network developed under our dynamic model versus a model with a static cortex [22], plus examples of each models' receptive fields (RFs). **A: (C:)** Shown are the RFs located within the green square from Fig. 3.12 developed under the dynamic (static) framework. Each pixel represents a feedforward connection from the pre-synaptic population dominated by the ON-center responsive population (white), OFF-center responsive population (black), or neither (gray). **B: (D:)** Distributions or orientation selectivity index developed under the dyanmic (static) framework. Orientation selectivity index measures how strongly a neuron in the post-synaptic population responds to its preferred orientation over other orientations. The dashed line shows the mean of the distribution.

Orientation selectivity is a measure of how strongly the cortical neuron responds to its preferred orientation over others (see Sec. 3.4.6). At one end, weak orientation selectivity means that the neuron would respond over its baseline activity as long as there is a stimulus inside its receptive field regardless of the

orientation. At the other end, strong orientation selectivity means that the neuron only responds over its baseline activity when the stimulus closely matches its preferred orientation. Our dynamic framework does develop better orientation selectivity than the static framework. In Fig. 3.13, we show two sets of example cells (the cells from Fig. 3.12- green square) developed under our dynamic framework and the static framework in A and C respectively. Qualitatively, the example cells developed under the static framework have noisier RF filters than the example cells developed under our dynamic framework. We would thus expect that the cells developed under our framework to have higher orientation selectivity indices than the cells developed under the static framework, and indeed that is what we see (Fig. 3.13B and D). Orientation selectivity indices are a measure of how strongly a neuron responds to its preferred orientation over other orientations (see Sec. 3.4.6 for more details). The distributions show the orientation selectivity index for all neurons in the post-synaptic populations developed under our dynamic framework (Fig. 3.13B) or the static framework (Fig. 3.13D). Our framework becomes substantially more selective than the static framework; the mean selectivity for our framework is 0.43 while the mean selectivity for the static framework is 0.36 (Fig. 3.13B and D, dashed lines). Thus our framework better approximates biology than the static framework, as neurons within V1 are typically selective to their preferred orientation [73].

However, even though our framework develops selectivity which better approximates biology, the static framework produces a smoother organization of its selectivity than our dynamic framework. In Fig. 3.14A (C), we show the preferred orientation across the post-synaptic population developed under the dynamic (static) framework. Each pixel again corresponds to a neuron within

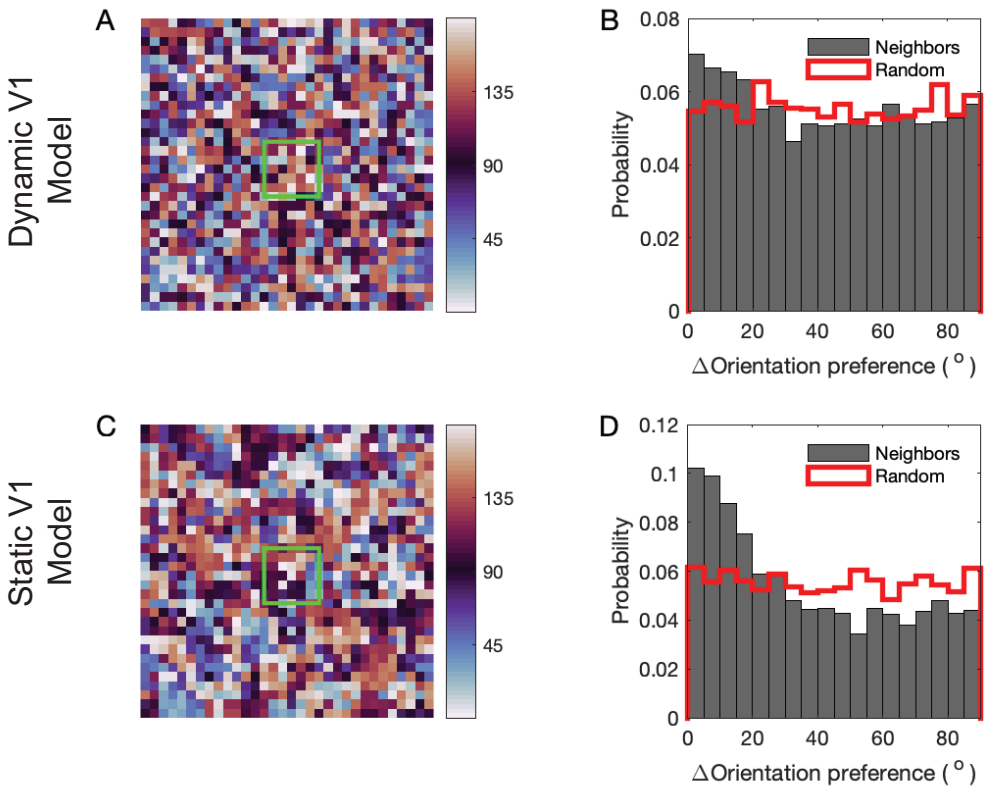

FIGURE 3.14. Orientation preference maps for a network developed under our dynamic model versus a model with a static cortex [22]. **A: (C:)** Each pixel corresponds to a post-synaptic neuron. The coloring of that pixel shows the preferred orientation at that location developed under our dynamic (static) framework. Green square is around the same example cells from Fig. 3.13. **B: (D:)** Distributions of differences in orientation preference between neighbors (black bars), and random post-synaptic neurons (red line). Differences between neighbors being more clustered around zero indicates a smoother mapping of orientation preference.

our post-synaptic population at that location. The coloring of that pixel shows the preferred orientation of that neuron. We have chosen a cyclic color map as orientations of one hundred and eighty degrees are equivalent to orientations at zero degrees. By eye, there appear to be more uniform clusters of similar colors in the static framework (C) than our dynamic framework, implying that smoother

organization of orientation preference has developed in the static framework. However in these heat maps, there is no notion of orientation selectivity.

The post-synaptic population develops selectivity to all orientations equally in both frameworks, as the difference in preferred orientation between any two randomly selected neurons approaches a uniform distribution (Fig. 3.14B and D, red line). For a smooth map, we expect differences in preferred orientation between neighboring neurons to be highly clustered around zero, meaning that preferred orientation has just slightly changed from neighbor to neighbor. A smooth map, then, should have a peak in the distribution near zero that is much taller than the red line. Hence, both frameworks develop a smooth map for orientation preference as the difference in preferred orientation between neighbors is clustered around zero (Fig. 3.14B and D, black bars), matching experimental works [23]. However, the static framework is more clustered around zero, and hence develops a smoother map for orientation preference across the cortex.

By including multiple input correlation features, though, both the static framework and our dynamic framework develop less smooth orientation preference maps than previous theoretical models [22] for the parameter choices shown.We have chosen these parameters, *i.e.* λ^{slow}, λ^{fast} and ℓ^{fast}, such that the same relative selectivity to the fast and slow features of the input correlation is observed (Figs 3.9 and 3.10). There exist parameter choices, though, for which our framework develops significantly smoother orientation preference maps than the static framework, but often those choices come at the expense of weakening selectivity to the fast input correlation feature. As this work focuses on developing qualitatively different maps for two features, we have used the parameter

choices that highlight this effect, at the expense of weakening the smoothness of orientation preference.

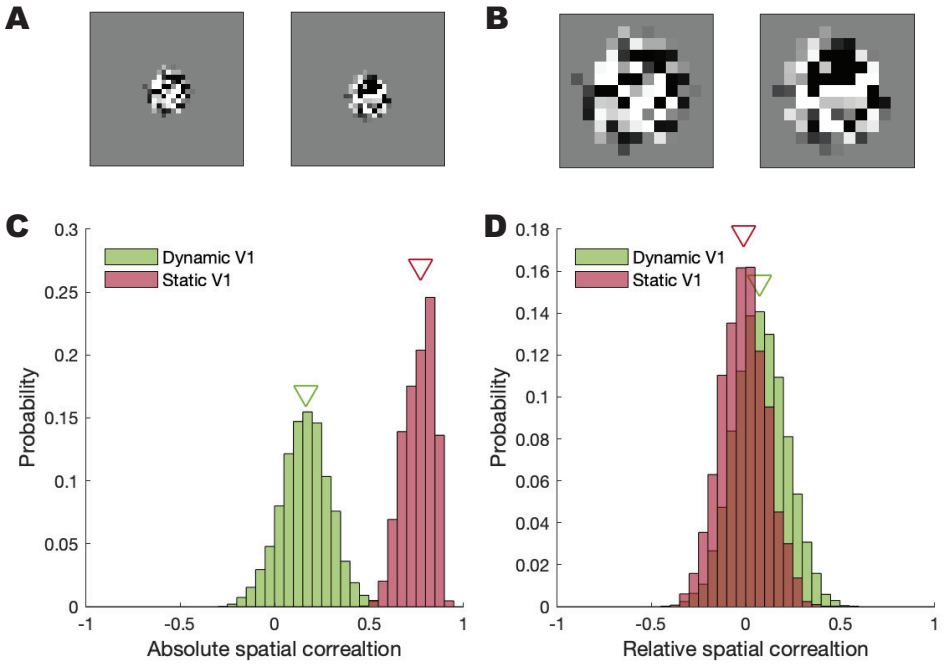

FIGURE 3.15. Examples of absolute and relative connectivities plus comparisons of Relative and absolute spatial correlation between a model with cortical dynamics and one without [22]. **A**: Two example absolute connectivities developed under our dynamic model. The absolute spatial correlation compares the connections across the LGN network. The absolute spatial correlation between the two examples shown is 0.028. **B**:Two example relative connectivities developed under our dynamic model (same post-synaptic neurons as **A**). The relative spatial correlation compares connections within the arbor to look at relative patches. The relative spatial correlation between the two examples shown is 0.21. **C**: Distributions of absolute spatial correlation between feedforward connectivity of neighboring units in our dynamic framework (green), and the static framework (red). **B:** Distributions of relative spatial correlation between neighboring RFs, or the spatial correlation of feedforward connections within an arbor between neighboring units developed under the dyanmic framework (green) and the static framework (red).

Moreover, we study the spatial correlation of feedforward connections between neighboring post-synaptic neurons as well. We compare the spatial

correlation in the post-synaptic neurons in two ways, relative and absolute. Within our cortical population, the absolute correlation should be clustered around zero, which would keep the correlation in the output between cortical neurons low, as is observed in [8]. Additionally, we would like the relative correlation to be clustered around zero as well, as that means that each receptive field has developed a selectivity to some spatial feature independent of its neighbors, similar to [69]. For the absolute spatial correlation, we compute the Pearson correlation coefficient between neighboring post-synaptic neurons for all of their connections across the LGN. Essentially, for the absolute spatial correlation, we compute the correlation across rows of W which correspond to neighboring locations in the grid of post-synaptic neurons. For the relative spatial correlation, we overlay the receptive fields of two neighboring neurons and compute the Pearson correlation coefficient between their RFs. Essentially, we shift the arbors of all post-synaptic neurons to the same location, and compute the correlation between the arbor-connections of neighboring post-synaptic location.

For both relative and absolute spatial correlation our dynamic framework is clustered near zero (Fig. 3.15), performing much better than previous models [22]. Whereas the static framework is only clustered around zero for relative spatial correlation, but remains near one for the absolute spaital correlation. Hence our dynamic framework improves upon previous model by moving the distributions of spatial correlation, both relative and absolute, closer to zero.

While our cortex still develops qualitatively different maps to our previously considered features, the eigenvectors within C^{slow} and C^{fast} (Fig. 3.9) and develops low spatial correlations as desired, it develops more of a S&P organization than is observed in biology [8]. There remain many parameter choices in our model for

which the biological value is unknown. Moreover, we use a relatively simplistic model of input correlations and cortical response kernels with only two time-scales. It is likely that biological input correlations and cortical response kernels are much more complicated and involve many more time-scales. By considering the "full-rank" time structure of the input correlations and response kernel, the cortex may develop a smooth orientation preference and keep the correlation between RFs low. The prediction of this model, though, is that features at those different time-scales would couple strongly to the cortical response kernel with a similar time-scale, and therefore that feature selectivity would develop an organization across the cortex governed by the spatial shape of the cortical response kernel at that time-scale. Hence selectivity to features which couple strongly into narrow response kernels would develop S&P organizations while selectivity to features which couple strongly to broadly excitatory response kernels would develop smooth maps (Fig. 3.9). Moreover, by including the dynamics of the cortex, distributions of spatial correlation become more closely centered on zero for both relative RF correlations and absolute.

3.4. Methods

In order to study the development of feedforward connectivity under a Hebbian framework, we built on previous theoretical frameworks [22], but relaxed a key assumption, namely that the cortex responds infinitely fast to its inputs. Recall that by relaxing that assumption, the cortical response kernel and input correlation couple across time. WE restricted ourselves to cortical response kernels and input correlations which had M and S time-scales respectively. Under this new framework (see Sec. 3.2.2), we arrived at the following unconstrained equation

for the development of feedforward connectivity

$$\tau_W \frac{d}{dt} W_p = A \odot \sum_{j=1}^{M} \sum_{i=1}^{S} \frac{\gamma_i \tau_j}{1 + \gamma_i \tau_j} K^i \left[W_p C^j_{p,p} + W_q C^j_{q,p} \right] \tag{3.44}$$

which contains only spatial forms for the response kernels K^i and $C^j_{q,p}$, and the arbor A acts element-wise on the summed terms. Here, in this section, we present our methods for simulating networks whose feedforward connections obey Eqn 3.44. Specifically we discuss designing spatial structure of the networks themselves, the input correlations and the cortical linear response kernels, and lastly the functional form of the arbor. We also present how Eqn 3.44 is constrained such that the total synaptic strength on a post-synaptic population neuron does not change over time, and impose that W_p does not grow without bounds. Moreover, we outline our integration method for solving the differential Eqn 3.44, and present how receptive field properties are calculated.

3.4.1. A note on the forms of the networks

For both the one- and two-dimensional network simulations, we imposed periodic boundary conditions. For the one-dimensional networks, this means that both the LGN and V1 networks existed on a ring, and in Fig 3.5 the far right and far left neurons of both populations would be next to each other. Similarly, we imposed periodic boundary conditions on the two-dimensional networks such that the LGN and V1 formed a torus. This was done to avoid effects from the boundary.

Obviously the cortex does not form a torus or a ring. The cortex does however contain many more neurons than we were able to efficiently simulate.

Since we were unable to simulate as many neurons as the cortex contains, we imposed periodic boundary conditions to avoid finite-size effects. Periodic boundary conditions are not essential to our model, and the results would hold if a larger simulation was run without periodic boundary conditions.

3.4.2. Defining the structure of K^i and $C^j_{p,q}$

Here we have chosen the spatial form of K^i to be long-range and difference of Gaussians (DoG) in spatial form at slower time-scales, but the response kernels approximates a Kronecker-δ at faster time-scales. Specifically in the two time-scale we choose K^{slow} to be DoG, and K^{fast} to be a Kronecker-δ. Previously it had been shown that a spatially broad and excitatory cortical response kernel would cause the organization of selectivity to features in the input correlations to be a smooth map, while a short-range or broadly inhibitory response kernel would cause the feature selectivity to be encoded in a S&P organization. We assume that the cortical response kernel governing slower time-scales is broader than the response kernel at fast time-scales because it matches our intuition. We believe this to be a valid assumption because it seems reasonable that at longer time-scales the input signal at one cortical cell can propagate outwards and affect the activity at other cells, but at shorter and faster time-scales we expect self-interaction to be the biggest effect on the post-synaptic population dynamics. Further there is a lack of experimental evidence as to the actual shape of the cortical response kernel of the cortex. Importantly though, our model makes predictions about the organization of selectivity features given the spatial structure of the response kernel at that time-scale, but does not depend on the structure at certain time-scales having certain structures. If, for instance, the response kernel at fast time-scales should be

found to be broad as well, then our theory predicts that the feature of fast time-scale input correlations would also be encoded in a smooth map.

We have chosen the spatial form of $C^j_{q,p}$ to be Mexican-hat, or difference of Gaussians (DoG) at all time-scales. As such, K^{slow} and all $C^j_{p,q}$ are defined by their own σ^E and σ^I which set the length scale of the excitatory and inhibitory lobe for the DoG respectively. Moreover in the two time-scale model, we construct $C^{slow}_{p,q}$ such that its principal eigenvalue is orthogonal to the principal eigenvalue $C^{fast}_{p,q}$ for all p, q. When the LGN has many sub-populations, we impose that $C^j_{p,p} = C^j_{q,q}$ for $q \neq p$, and that $C^j_{q,p} \propto C^j_{p,p}$ when $p \neq q$ for $j \in [slow, fast]$. Thus we can define $C^j_D = C^j_{p,p} - C^j_{p,q}$ which has the same eigenvectors of the original $C^j_{p,p}$, and thus we maintain that $\mathbf{e}^{slow}$ is orthogonal to $\mathbf{e}^{fast}$. Lastly, we normalize the principal eigenvalues of C^{slow}_D and C^{fast}_D to be 1, as well as the principal eigenvalues of K^{slow} and K^{fast}. To control the relative strength of selectivity to the slow and the fast feature, we therefore introduce a scalar factor for both time-scales, λ_{slow} and λ_{fast}, as follows.

$$\begin{aligned}\tau_W \frac{d}{dt} W_p = A \odot \sum_j \Bigg[& \lambda_{slow} \quad \frac{\gamma_{slow}\tau_j}{1+\gamma_{slow}\tau_j} K^{slow}\left[W_p C^j_{p,p} + W_q C^j_{q,p}\right] \\ & + \lambda_{fast} \quad \frac{\gamma_{fast}\tau_j}{1+\gamma_{fast}\tau_j} K^{fast}\left[W_p C^j_{p,p} + W_q C^j_{q,p}\right]\Bigg] \end{aligned} \tag{3.45}$$

These scalar factors ensure that the feature selectivity at both time-scales develop the similarly strong selectivities.

3.4.3. Arbor function

The arbor function, A, is imposed upon the feedforward connectivity such that the pre-synaptic population only forms connections with a subset of the

post-synaptic population. In general, $A = A(\mathbf{x}, \mathbf{a})$ a function of the presynaptic location, $\mathbf{a}$, and the post-synaptic location, $\mathbf{x}$. Practically, $A(\mathbf{x}, \mathbf{a})$ is a matrix with the $\mathbf{a}, \mathbf{x}$ element given by its functional form. In order to enforce retinotopy on our networks, though, A only depends on the difference in location of the pre- and post-synaptic units, $A(\mathbf{x}, \mathbf{a}) = A(\mathbf{x} - \mathbf{a})$ (Fig. 3.5). The simplest form of an arbor function of this type would be the "pill-box" arbor with radius R_A, where

$$A(\mathbf{x} - \mathbf{a}) \quad = 1 \text{ if } \|\mathbf{x} - \mathbf{a}\| < R_A \tag{3.46}$$

$$A(\mathbf{x} - \mathbf{a}) \qquad = 0 \text{ otherwise} \tag{3.47}$$

Such a form allows all synapses to saturate at the maximum upper bound, regardless of how far separated they are. This arbor was used in our one-dimensional simulations.

However, such an arbor does not take into account the physical limitations that underlie the synaptic connections from axons to dendrites. Recall that axons are the output channels of the pre-synaptic neuron, and dendrites are the input channel of the post-synaptic neuron. Here we are not concerned about the axons reaching the post-synaptic population, but we do wish to study how the arbor forms connections with the dendrites. Both axons and dendrites are defined by a branching pattern. If we think about a cone containing all the branches of both the axon and the dendrite, we expect the cone for the axon to be narrower as it makes a more directed branching pattern (see Fig. 1.1A). Since we are not concerned about how the axons reach the post-synaptic population, we will think about the once as being just its base. Hence we can model the extents of the axon and the dendrite as being circular.

Furthermore, we assume that the chances of connection from a pre-synaptic neuron to a post-synaptic neuron should scale as the overlap of those two circles. To that end, then we define our arbor as the overlap of two circles centered on the pre- and post-synaptic neurons. The circle centered on the pre-synaptic neuron represents the branching extent of its axon, and similarly the circle centered on the post-synaptic neuron represents the branching extent of its dendrites. We normalize the arbor such that when one circles is completely contained in the other the arbor has value one, and when the distance between the neurons is greater than the arbor radius, R_A, the arbor is zero. Specifically we choose a circle of radius R_A for the post-synaptic neuron, and a circle of radius $r_A = c_A R_A$ to describe the spatial extent of the axon, with $c_a < 1$. The spatial overlap of two circles is given by

$$\begin{aligned} s(\boldsymbol{r}, R_A, r_a) &= R_A^2 \frac{\arccos\left(\boldsymbol{r}^2 + R_A^2 - r_A^2\right)}{2\boldsymbol{r}R_A} \\ &+ \frac{\arccos\left(\boldsymbol{r}^2 - R_A^2 + r_A^2\right)}{2\boldsymbol{r}r_A} - \frac{1}{2}\sqrt{(4\boldsymbol{r}^2 R_A^2 - (R_A^2 + \boldsymbol{r}^2 - r_A^2)^2)} \end{aligned} \quad (3.48)$$

The strength of the arbor at separation $\boldsymbol{r} = \mathbf{x} - \mathbf{a}$ is given as

$$\begin{aligned} A(\boldsymbol{r}) &= 1 && \text{if} && \boldsymbol{r} < R_A - r_A \\ A(\boldsymbol{r}) &= \frac{1}{\pi r_a^2} s(\boldsymbol{r}, R_A, r_a) && \text{if} && R_A - r_a < \boldsymbol{r} < R_A \\ A(\boldsymbol{r}) &= 0 && \text{otherwise} \end{aligned} \quad (3.49)$$

The effect of including a spatially decaying arbor is to limit the maximum strength of synaptic connection of widely separated neurons, but a tapered arbor versus a pill-box arbor does not substantially change the development of the

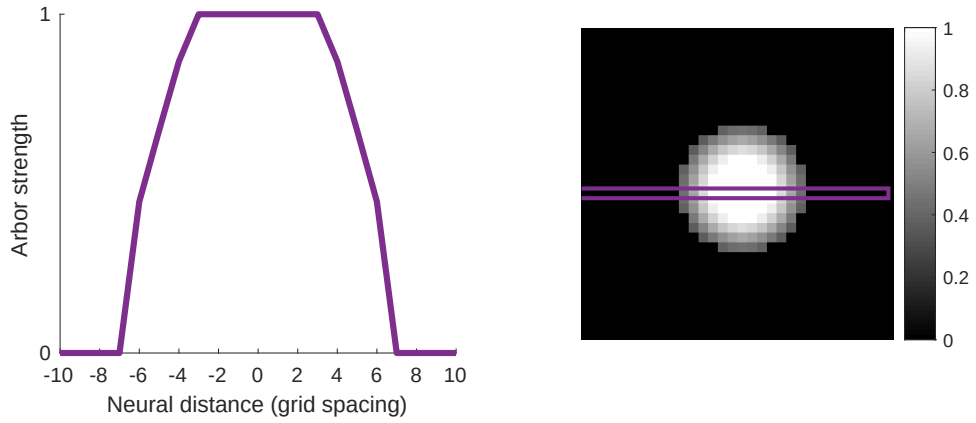

FIGURE 3.16. Cross section and two-dimensional representation of the arbor function used. On the left is shown the cross section of the arbor strength across a subset of neural distances. On the right is shown the two dimensional form of the arbor, and highlighted with a purple box is the line through which the cross section is taken.

receptive fields [22]. Synapses have an upper bound given by the product of the upper bound at no separation, W_{max}, with the arbor strength at the separation between their pre- and post-syanptic neurons, *i.e.* $W_{max,\mathbf{x},\mathbf{a}} = A(\mathbf{x} - \mathbf{a})W_{max}$ (see Sec. 3.4.4). Previous work suggests that the role of the arbor function is primarily determined by the cutoff of its Fourier transform, or in other words the spatial frequency at which the its Fourier transform becomes negligibly small compared to the value at zero spatial frequency [22]. The corresponding wavelength of the cutoff frequency gives an "effective" arbor diameter, which for the values used here, is about 11.3.

3.4.4. Constraints and bounding

In addition to including an arbor, we also constrain equation Eqn 3.44 such that the total synaptic strength onto a single post-synaptic unit does not change over time. Thus we modify Eqn 3.44 to include a subtractive constraint via

$$\left.\frac{d}{dt}\right|_C W_p(\mathbf{x},\mathbf{a}) = \frac{d}{dt}W_p(\mathbf{x},\mathbf{a}) - \epsilon(\mathbf{x})A(\mathbf{x}-\mathbf{a}) \tag{3.50}$$

where $\frac{d}{dt}W_p(\mathbf{x},\mathbf{a})$ represents the unconstrained derivative term, *e.g.* Eqn 3.44, and $\epsilon(\mathbf{x})$ is a function which keeps the total synaptic strength at post-synaptic location $\mathbf{x}$ constant and is given by

$$\epsilon(\mathbf{x}) = \frac{1}{2\sum_{\mathbf{a}} A(\mathbf{x}-\mathbf{a})} \sum_{\mathbf{a}} [\frac{d}{dt}W_{ON}(x,a) + \frac{d}{dt}W_{OFF}] \tag{3.51}$$

Here we have adopted the notation that the unconstrained equations are given as $\frac{d}{dt}W_p$, while we will always denote the constrained equations with $\left.\frac{d}{dt}\right|_C W_p$, like Eqn 3.50. Note also that we will use letters in the early part of the alphabet to denote locations within the pre-synaptic population, and letters from later in the alphabet to denote the locations of the post-synaptic population, *e.g.* a for the pre-synaptic population and x for the post-synaptic population. Eqn 3.51 has been chosen such that the total change in synaptic connection strength for a post-synaptic cell is zero

$$\sum_{\mathbf{a}} [\left.\frac{d}{dt}\right|_C W_{ON}(\mathbf{x},\mathbf{a}) + \left.\frac{d}{dt}\right|_C W_{OFF}(\mathbf{x},\mathbf{a})] = 0 \tag{3.52}$$

for all $\mathbf{x}$, *i.e.* for all cortical cells.

Further, we impose that $W_p(\mathbf{x},\mathbf{a})$ is bounded above by some value $W_0 A(\mathbf{x}-\mathbf{a})$ (where W_0 is some scalar value setting the max possible value for W), and below by zero. In order to do so, whenever an element of W_p grows beyond one of those bounds, we reset the value to the bound, and further "freeze" the development of that synapse. "Freezing" the value of that synapse aids in

computational efficiency, but otherwise has little impact on the development of the other feedforward connections. To ensure that Eqn 3.52 remains satisfied whenever a particular connection $W_p(\mathbf{x}, \mathbf{a})$ is reset and frozen, we adopt the algorithm outlined in [22].

3.4.5. Simulation algorithm

Here we present the algorithm for solving Eqn 3.50 for ON-center responsive LGN populations which can easily be applied to the OFF-center population by exchanging OFF for ON in all equations, and vice versa. Moreover, it is more obvious how the expanded model can be compressed to the single LGN population model than expanding the single population to many.

Initially all values for W_{ON} are randomly assigned a strength uniformly sampled from $(1 \pm w)A(\mathbf{x} - \mathbf{a})$, with $w = 0.2$ the noise in the initial connectivity strength. The initial synaptic strengths are multiplicatively normalized via step 5 in the algorithm described below. Thereafter, each synapse was updated each time-step via the algorith described below until the upper or lower bounds were reached, at which point the synapse was "frozen" at the bound it reached. The simulations continue until $\geq 90\%$ of all the synapses with nonzero arbor reach one of the bounds.

Step 1: Compute the right-hand-side of the unconstrained derivative in Eqn 3.44

$$\Delta_{ON} \equiv \frac{1}{\tau_W}\big[A \odot \sum_i \sum_j \frac{\gamma_i \tau_j}{1 + \gamma_i \tau_j} K^i \Big[W_{ON} C^j_{ON,ON} + W_{OFF} C^j_{OFF,ON} \Big] \big] \qquad (3.53)$$

Note that we have suppressed the convolution symbol in our formalism, opting instead to denote convolutions across space by writing matrices next to each other.

$$\begin{aligned} K^i W_{ON} C^j_{ON,ON} &= K^i * W_{ON} * C^j_{ON,ON} \\ &\equiv \sum_{\mathbf{x}'} \sum_{\mathbf{a}'} K^i(\mathbf{x}, \mathbf{x}') W_{ON}(\mathbf{x}', \mathbf{a}') C^j_{ON,ON}(\mathbf{a}', \mathbf{a}) \end{aligned} \tag{3.54}$$

Convolutions can be computationally costly to find, but we can significantly expedite the simulation by utilizing a property inherent to the structure of our cortical response kernels and correlations matrices. Given that we are assuming each cortical or post-synaptic neuron has the same spatial form of the cortical response which only depends on the relative spacing between cortical neurons, K^i is a Toeplitz matrix. Similarly, since we have defined the correlations by the relative spacing between LGN or pre-synaptic neurons, then $C^j_{p,q}$ is also Toeplitz. Toeplitz matrices have a particular property that significantly speeds up the simulation, namely the eigenvalues of a Toeplitz matrix can found via their Fourier transform (to be clear this Fourier transform is spatial) [83]. Hence we can take the Fourier transform of Eqn 3.53, so that the convolutions over neural space become simple element-by-element products in Fourier space which are relatively easy to solve, then take the inverse Fourier transform of those products to find Δ_{ON}.

During the first evaluation of Δ_{ON}, we set the time step of our integration, dt. We choose dt such that the standard deviation of change in synaptic weights is $\sigma_\Delta = 0.01$ To ensure that our time-step is not too large, if $dt > dt_0$, then we set $dt = max(dt/2, dt_0)$, where dt_0 is chosen based on the size of the arbor.

Step 2: Modify Δ_{ON} with subtractive constraints that conserve the total synaptic strength over each cortical cell.

$$\Delta_{ON}(\mathbf{x},\mathbf{a})\Big|_C = \Delta_{ON}(\mathbf{x},\mathbf{a}) - \epsilon(\mathbf{x})A(\mathbf{x}-\mathbf{a}) \tag{3.55}$$

Step 3: Use this constrained Δ_{ON} along with the constrained Δ_{ON} from previous time-steps to compute the total change in each synapse using a three-step integration method. The three-step integration method allows larger time-steps, dt, and speeds computation. To find the change in W_{ON}, we used

$$W_{ON}(t+1) = W_{ON}(t) + dt(23F_t - 16F_{t-1} + 5F_{t-2})/12 \tag{3.56}$$

where $F_t = \Delta_{ON}\Big|_C$ at that corresponding time step. However, this formalism is only valid after the third time-step. The first two updates are given by

$$W_{ON}(t=1) = W_{ON}(t=0) + dtF_0 \tag{3.57}$$

$$W_{ON}(t=2) = W_{ON}(t=1) + dt(2F_1 - F_0) \tag{3.58}$$

After $t = 4$, the step size was doubled, so that only even-numbered time-steps were computed afterwards, *i.e.* $W_{ON}(t+2) = W_{ON}(t) + dt(23F_t - 16F_{t-2} + 5F_{t-4})/12$.

Step 4: If an element of W_{ON} reached or exceeded the upper or lower bound, reset the element to the bound, *i.e.* if $W_{ON}(\mathbf{x},\mathbf{a};t) < 0$ or $W_{ON}(\mathbf{x},\mathbf{a};t) > W_o A(\mathbf{x}-\mathbf{a})$ reset it to $W_{ON}(\mathbf{x},\mathbf{a};t) = 0$ or $W_{ON}(\mathbf{x},\mathbf{a};t) = W_o A(\mathbf{x}-\mathbf{a})$ respectively. Further we "freeze" synapses at the bounds so that they no longer change over time.

Step 5: If any synapses are cut off or reset in the previous step, then the total synaptic strength onto a cortical neuron will have changed. To correct for this, we multiplicatively renormalize the synaptic strength on those cortical cells. All synaptic strengths on those cells, excepting the synapses at the bounds, are multiplied by

$$\gamma(\mathbf{x}) = \frac{2\sum_{\mathbf{b}} A(\mathbf{x}-\mathbf{b})}{\sum_{\mathbf{b}}\left[W_{ON}(\mathbf{x},\mathbf{b}) + W_{OFF}(\mathbf{x},\mathbf{b})\right]} \tag{3.59}$$

which returns the total synaptic strength onto that cell to $2\sum_{\mathbf{b}} A(\mathbf{x}-\mathbf{b})$, as previous.

Steps 2 and **5** must be updated once synapses have started saturating at either the upper or lower bound. Let $W_{active}(\mathbf{x})$ be defined as the sum of all unfrozen feedforward synaptic strengths onto the post-synaptic cell at location $\mathbf{x}$, including both ON- and OFF-center populations, and similarly define W_{frozen} as the sum of frozen feedforward strengths. Further let $A_{ON,active}(\mathbf{x},\mathbf{a}) = A(\mathbf{x},\mathbf{a})$ if that synapse is active (or unfrozen), but $A_{ON}(\mathbf{x},\mathbf{a}) = 0$ if the synapse is frozen. Then for **step 2**, $\epsilon(\mathbf{x})$ is updated in the following way

$$\epsilon(\mathbf{x}) = \frac{\sum_{\mathbf{b}} \Delta_{ON}(\mathbf{x},\mathbf{b}) + \Delta_{OFF}(\mathbf{x},\mathbf{b})}{\sum_{\mathbf{b}}\left[A_{ON,active}(\mathbf{x}-\mathbf{b}) + A_{OFF,active}(\mathbf{x}-\mathbf{b})\right]} \tag{3.60}$$

and for **step 5**, $\gamma(\mathbf{x})$ is similarly updated via

$$\gamma(\mathbf{x}) = \left[- W_{frozen}(\mathbf{x}) + 2\frac{\sum_{\mathbf{b}} A(\mathbf{x}-\mathbf{b})}{W_{active}(\mathbf{x})}\right] \tag{3.61}$$

Lastly we impose that $0.8 \leq \gamma(\mathbf{x}) \leq 1.2$.

3.4.6. Determining RF features

After simulating the development of feedforward synapses, we study the selectivity to certain features. First we define selectivity to the principal component of C^j as being the projection of the feedforward connectivity strengths of W onto $\mathbf{e}^j$,

$$\mathbf{S}^j \equiv W\mathbf{e}^j \tag{3.62}$$

where $\mathbf{S}^j$ is the N_{post}-dimensional vector of the selectivity. Given our construction of the input correlations when multiple LGN sub-populations are modeled, particularly that C_D^j has the same principal eigenvector as $C_{p,q}^j$, the same selectivity organizations to the principal eigenvector develop when multiple LGN sub-populations are considered as when the LGN has only one population.

Moreover when the LGN contains both ON- and OFF-center response populations, then the post-synaptic neurons will develop selectivity to features which correspond to experimentally observed RF features. We have defined the RF of our post-synaptic population to be $W = W_{ON} - W_{OFF}$, and mainly consider connections with the pre-synaptic populations which are within the arbor. To find the preferred orientation or phase of our cortical neurons, we find the response, $R_{\mathbf{x}}(\theta, \phi)$, to a variety of sinusoidal gratings which have varying orientations (θ) and phases (ϕ). We model the response of a post-synaptic cell as being the overlap of the RF with a particular grating

$$R_{\mathbf{x}}(\theta, \phi) = \sum_{\mathbf{a}} W(\mathbf{x}, \mathbf{a}) \sin\left([2\pi \mathbf{k} \cdot (\mathbf{x} - \mathbf{a})] + \phi\right) \tag{3.63}$$

where the orientation θ is the angle of the vector $\mathbf{k}$, *i.e.* $\mathbf{k} = k_x \cos\theta\hat{x} + k_y \sin\theta\hat{y}$. The preferred orientation and phase can then be found as the argument maximum of $R_{\mathbf{x}}(\theta, \phi)$.

However, this measure gives no sense of the degree of selectivity to the preferred orientation, known as the orientation selectivity index. Therefore, to make a more experimentally consistent measurement of the preferred orientation, we define $\mathcal{R}_{\mathbf{x}}(\theta)$ which is the N_θ-dimensional vector of responses of the post-synaptic cell at location $\mathbf{x}$ to a grating with the above defined preferred phase. We can find the the vector sum, $v_{\mathbf{x}}$ of $\mathcal{R}_{\mathbf{x}}(\theta)$, by $v_{\mathbf{x}} = \sum_\theta e^{2i\theta}\mathcal{R}_{\mathbf{x}}(\theta)$; the orientation selectivity is thus $\|v_{\mathbf{x}}\|/N_\theta$, and the preferred orientation is the $\arg(v_{\mathbf{x}})/2$ because $\theta \in [0, 180]$ for grating orientations. The preferred phase of the post-synaptic population can be found in a similar manner, but constructing a N_ϕ-dimensional vector of responses $\mathcal{R}_{\mathbf{x}}(\phi)$ to the previously defined preferred orientation.

We found the spatial correlation of RFs across the post-synaptic population in two different ways, one relative and on absolute. To find the relative correlation between neighboring RFs, we shifted the RF of all post-synaptic neurons such they all centered on the same retinotopic location, and then found the Pearson correlation coefficient of their spatial forms. So we only considered the overlaps between an $D_A \times D_A$ grid for each neuron, rather than the correlation between their connections to the full grid which was done for the absolute RF comparison. More concretely, when comparing the absolute spatial correlations across the post-synaptic population, we again found the Pearson correlation coefficient but without centering, *i.e.* we found the Pearson correlation coefficient between the rows of W which correspond to neighboring locations in the post-synaptic network.

3.5. Discussion

Here we have presented results from a theoretical model of how feedforward thalamacortical connections develop under Hebbian plasticity. We have expanded upon previous theoretical models [78, 79, 80] and in particular [22] by including a more realistic treatment of cortical activity dynamics. We showed that when properly considered, the cortical linear response kernel may have different spatial shapes at different time-scales, essentially gaining a "degree of freedom" for our model. Moreover, we showed that a novel coupling mechanisms emerges wherein features of the cortical input that temporally vary at different time-scales couple most strongly to the spatial cortical linear response kernel at similar time-scales. The spatial shape of the cortical linear response kernel at a particular time-scale determines the organization of the selectivity to the feature that shares that time-scale. When the cortical linear response kernel is broad and excitatory, it promotes the growth of smooth maps for features, but when the cortical linear response kernel is narrow, it promotes the growth of S&P organizations. Different spatial cortical linear response kernels may have different spatial shapes at different time-scales. Therefore, different feature organizations may develop across the same cortex which was not possible in previous models. We presented simulations which bore out this mathematical result in both one- and two-dimensional network simulations for simple input correlations and cortical linear response kernels.

In this work, we have largely restricted ourselves to thinking about the primary visual cortex of cats and higher mammals. In principle though, Eq. (3.15) (or equivalently Eq. (3.16)) is quite general and applicable to many species and systems not considered here. Moreover, in this work we have considered stylized and simplistic input correlations and cortical response kernels as a proof of

principal. In the future, it would be interesting to see the results of this model using input correlations and cortical response kernels from experimental studies. We expect that there exist many more time-scales than two in the real cortex and thalamus. At each time-scale though, we expect the organization of that feature across the cortex to be determined by the spatial cortical response which shares its time-scale.

Note that in this work we are not interested in studying the differences in feedforward thalamocortical development across species. Rather, we sought to provide an expanded theoretical framework that could describe how the cortex in cats and primates develops a smooth map for orientation preference while simultaneously developing heterogeneous receptive field (RF) filters. However our work does make predictions about the spatial structures of the rodent cortical linear response kernels. Within rodents, a S&P has been observed for orientation preference [7, 75]. As such, we expect that their cortical response kernels are highly local across time-scales.

In this work we also did not include recurrent plasticity within the cortex. Such plasticity would change our cortical linear response kernels over time. We did not include cortical plasticity partly for modeling simplicity and transparency, but also due to experimental observations as well. There are correlated structures within the brain of ferrets which exist before and are predictive of orientation preference maps [84]. Such structures could likely be explained by cortical linear response kernels. As these structures exist before orientation preference maps, then it seems as though the cortical linear response kernels do not change much during the developmental period studied here. As such, including small changes to the

linear response kernels would be unlikely to change our results in a qualitative way, and the results presented here should hold.

Our framework as described here, though, represents a significant step forward in understanding the development of receptive fields as we have demonstrated that by considering the dynamics of the cortex there exists a coupling mechanism which can produce qualitatively different feature organizations across the cortex.

CHAPTER IV

CONCLUSIONS AND FUTURE WORK

Within this dissertation we have studied the impact of synaptic connections on neural activity dynamics, and vice versa. We have shown that certain connectivity structures are key when simulating oscillating neural activity, and that the activity of the cortex could possibly play a key role in shaping feedforward thalamocortical connectoinis. These two major works (Ch. II and III) demonstrate that studying neural activity dynamics without properly considering the connectivity structure or vice versa will leave an incomplete impression of what is necessary for the brain to work the way it does. In Ch. II, we show that the SSN is able to accurately capture the dynamics of the cortex only when a proper horizontal connectivity structure is considered. In Ch. III we show that the development of receptive field filters depends upon the dynamics of the cortex to simultaneously develop a smooth map for orientation preference and heterogeneous spatial RF filters. Thus the brain relies on this beautiful dance where the dynamics shape the connectivity while being simultaneously being governed by the established connectivity.

4.1. Gamma peak frequencies and surround suppression

In Ch. II, we showed that the stabilized supralinear network (SSN) provides a mechanistic circuit model which accurately depicts the stimulus dependence of gamma oscillations. By modelling horizontal connections as including some strong local (intra-columnar) dependence in their connectivity profile, the SSN can produce both the locality of contrast dependence as well as display the

surround suppression of firing rates. Without including some elevated strength of connections to the local population, though, the network is unable to produce both the locality of contrast dependence but robustly produces the surround suppression of firing rates. On the other extreme, when the network either only has local connections or local connections are much stronger than long range connections, we showed that the network robustly displays the local contrast dependence of gamma peak frequency but fails to produce the surround suppression of firing rates. However, when the proper balance is found between the two, the SSN reproduces the locality of contrast dependence and the surround suppression of firing rates robustly up to perturbations of $\pm 10\%$. Moreover, we demonstrated how the SSN is able to capture these effects through its non-saturating non-linearity.

Interestingly, this model was not able to capture a related effect on the gamma peak frequency with surround suppression. [20] reports gamma peak frequency shifts to lower frequencies as the stimulus size increases. Further, the peak itself gets sharper and has higher power with increasing stimulus size. Despite our numerous searches, we were never able to reproduce this effect in our simulations of gamma peak frequency and surround suppression. What we found instead was there were models where one effect was seen, but not the other. For instance, the gamma peak frequency decreased with increasing stimulus size, but the peak width got wider, or the gamma peak width got narrower and sharper, but the peak moved to higher frequencies. This effect though, may simply be beyond the model that we have used here. Experimental evidence has shown that the effects on the gamma power-spectrum with increasing stimulus size are due to feedback connections to V1 from higher visual areas [50]. As such, in order to

include those effects, we suggest to build on the model used in Fig. 2.4, but include a higher visual area also modelled as an SSN.

Moreover, [20] report a related effect on gamma peak frequency when different orientations are considered. Our network does not model populations as having an orientation preference, and rather treats each populations as a hypercolumn within V1. Thus a finer model than ours would have to be used in order to study the effects of changing orientation on gamma peak frequency. Moreover such a model would need to include a connection strength dependence on the relative difference in orientation preference, as seen in [42, 64].

Lastly our model does not provide any intuition as to how gamma peak is used by higher areas within the brain, rather it provides a mechanistic model for producing such oscillations. More studies would need to be conducted to see what sort of signalling mechanism, if any, gamma peak frequency provides to higher areas.

To summarize, there many different phenomena that models of the dynamics of cortex may explain. Here we have provided a parsimonious model which explains several observed phenomena of cortical activity dynamics, namely the local contrast dependence of gamma peak frequency with the surround suppression of firing rates.

4.2. Receptive field development

In Ch. III, we showed how considering the dynamics of the cortex is essential to developing qualitatively different organizations of selectivities or preferences to input features. We simulated the development of feedforward connections under two different frameworks, the dynamic framework which properly considers the

dynamics of the cortex and the static framework which treats the cortex as being constantly "slaved" to the input dynamics ($\tau_u \gg \tau_v$). Within our dynamic framework, the post-synaptic population becomes selective to two different features, and the selectivity to those features are organized in a qualitatively different way across the post-synaptic (cortical) population. We showed that the selectivity to one feature is encoded in a smooth map, while the selectivity to the other is encoded in salt-and-pepper (S&P) organization for a simple two-time scale model under our dynamic framework. While the static framework has a post-synaptic population which becomes selective to both features, nevertheless that same post-syanptic population develops the same organization of selectivity to both features, *e.g.* a smooth map for both.

Biologically a mixture of different organizations of feature selectivity are observed. Some features, such as orientation within the primary visual cortex (V1) of cats and higher mammals, are encoded in a smooth map. Within those same animals though, other features, such as the relative location of the light and dark patches within the RF filter, have a S&P organization as evidenced by the low correlations between the spatial shapes of the receptive fields (RFs) of neighboring cells [69]. Moreover, output correlations remain low among neighboring neurons within V1 [23], one of the main benefits of S&P organizations.

However, while our framework can predict the co-presence of qualitatively different organizations across the cortex to abstract and mathematical features such as the eigenvectors of the input correlations C^{fast} and C^{slow}, the static framework predicts a smoother organization of orientation preference in our post-synaptic neurons. One possible expansion to our model which may help our framework to develop a smoother orientation would be to use a non-linear

cortex (relax Assumption II). Using a nonlinear cortex, such as the SSN power law, may provide a level of flexibility to our network that allows the development of heterogeneous RF filters while keeping orientation preference as smooth as biological observations, [42, 64]. This increase in flexibility, though, comes at the cost of losing analytic tractability.

Another possible way to improve on our results would be to use more biologically plausible input correlation structures and linear response kernels. Here, we used difference of Gaussian spatial forms for our linear response kernel and input correlations, but the actual structures within the brain be much more complicated. Moreover, The eigenvectors of the actual input correlations may not be orthogonal, which would cause a blending of the two organizations, which may help smooth the orientation preference map. There is also many more time-scales than two within the brain, and considering more time-scales in our network may provide more realistic RF structures. As such, another next step should be a systematic study of input correlations within the lateral geniculate nucleus (LGN) of the thalamus, and a study of the response kernels of the cortex. The framework developed here though, shows how those many time-scales within the input should couple to the response kernels of the post-synaptic population.

One particularly interesting avenue of exploration to this author's eyes remains the study of synaptic strengths beyond their weights. The weight of a synapse is due to a host of factors, among them the number of nreceptor sites in the membrane of the post-synaptic cell, the amount of neuro-transmitter released by the pre-synaptic cell and more. All of those factors comprise what we have modelled as the feedforward weight, W. However, each of those factors may have their own dynamics happening on different time-scales which may help with

receptive field map formation. Recent work by [85] showed that when synapses have more factors than just a synaptic strength, networks are able to store more memories in the synapses. Given that we have shown here how important temporal dynamics of neural activity are to receptive field development, imagine what possibilities may open up when the dynamics of factors beyond synaptic strength are considered.

4.3. Parting words

These have been enjoyable projects, and the author is very glad he made the switch to neuroscience many years ago. Partly that is because he gets to do way more and way better math in this current work than he did previously, but it is also because he still gets to model and study a physical system. Moreover, it has allowed him to study and think about his own brain, which is fun. Additionally, neuroscience seems to have a lot of growth coming its way as a field.

Even though the author remains unsure of his own future directions, he has been offered a post-doctoral position previously (it didn't work out due to timing and the author being unsure how excited he was about the research). He has also applied to Lewis and Clark as a physics professor, but has yet to hear back. To be fair though, no liberal arts college is probably hiring in the midst of the COVID-19 pandemic so he's holding that plan loosely.

Thank you so much for reading this far. I hope you have learned something.

-CJH

APPENDIX A

MATHEMATICAL DETAILS FOR CH. II

Within our simulations for gamma peak frequency, we use biologically plausible values whenever possible for our parameters. In order to study the effect of horizontal connections, we did several different searches to try to find networks which best displayed the contrast dependence of gamma peak frequency, the local contrast dependence of gamma peak frequency, and the surround suppression of firing rates.

We searched over the following parameters for the two population networks:

$J_{E,E}$ strength of E to E connections

$J_{E,I}$ strength of I to E connections

$J_{I,E}$ strength of E to I connections

$J_{I,I}$ strength of I to I connections

g_E strength of feedforward E inputs

g_I strength of feedforward E inputs

ρ_N NMDA fraction of E weights

And expanded the search to include the following parameters when modelling the retinotopic network

$\lambda_{E,E}$ the locality of E to E connections

$\lambda_{E,I}$ the locality of E to I connections

$\sigma_{E,E}$ the spatial decay of E to E connections

$\sigma_{E,I}$ the spatial decay of E to I connections

Initially, we searched a grid of plausible parameter values. But searching the whole grid became computationally prohibitive.

We also performed gradient descent on our parameters using the JAX package of python since JAX can evaluate a gradient on an arbitrary function given a loss. We used a mean-square-error (MSE) loss on the shapes of our power spectra by comparing our simulated power spectra with idealized versions of the power-spectra found in [21], *i.e.* $loss_{PS} = \langle (PS_{ideal}(f) - PS_{simulated}(f))^2 \rangle_f$, where PS_{ideal} was based on [21] and $PS_{simulated}$ was found during our simulation for some parameter choices . We also include a regularizing term in our loss for our firing rates to be within biological regimes. The loss was zero if $r_E < 80$ Hz and $r_I < 100$ Hz for all E and I units in our network, but grew approximately linearly as the rates exceeded those limits, *i.e.* $loss_r = \langle \ln(1 + \exp(\boldsymbol{r} - \mathbf{UB})) \rangle_r$ where $\mathbf{UB}$ are the upper bounds for E and I units. For our retinotopic networks, we included a term which kept our suppression index above some minimum value. The loss for the suppression index was 0 when the network had a suppression index that was above or equal to some minimum value, but the loss increased linearly when the simulated suppression index was below that value. Typical experimental

observations of surround suppression report suppression indices in the range of $0.3 - 0.6$, and so we used 0.2 was used as our lower bound, SI_{min}. For the surround suppression loss, we used $loss_{SS} = \frac{A}{2}(|SI_{min} - SI_{simulated}| + SI_{min} - SI_{simulated})$, where A controlled the weight of this loss term ($A = 10$ was used here). Our total loss during these simulations was $L = loss_{PS} + loss_r + loss_{SS}$.

Using JAX helped narrow down the parameter ranges that we should search in, but ultimately the search that yielded the results used in this work was a random parameter search. Wherein, we set the ranges of the parameters according to some biological considerations guided by results from the JAX searches, and each parameter was drawn uniformly from that range. The shared parameters for the model producing the power-spectra in Figs. 2.1 and 2.4 were:

$$J_{EE} = 4.43$$

$$J_{EI} = 1.65$$

$$J_{IE} = 5.03$$

$$J_{II} = 1.24$$

$$g_E = 0.37$$

$$g_I = 0.26$$

$$\rho_N = 0.5$$

The remaining parameters used to specify the spatial connectivity in Fig. 2.4 were:

$$\lambda_{EE} = 0.4$$

$$\lambda_{EI} = 0.7$$

$$\sigma_{EE} = 0.20 \text{ mm}$$

$$\sigma_{EI} = 0.40 \text{ mm}$$

The example power-spectra shown in Fig. 2.6, had the following values for its parameters:

$$J_{EE} = 4.20$$

$$J_{EI} = 3.15$$

$$J_{IE} = 3.61$$

$$J_{II} = 1.86$$

$$g_E = 0.58$$

$$g_I = 0.23$$

$$\rho_N = 0.42$$

$$\lambda_{EE} = 0$$

$$\lambda_{EI} = 0$$

$$\sigma_{EE} = 0.22 \text{ mm}$$

$$\sigma_{EI} = 0.24 \text{ mm}$$

Lastly there are several fixed parameters for the networks that did not vary throughout our simulations. For both the two-population and the retinotopic

model, we chose

$$n = 2 \qquad \text{(exponent of SSN's nonlinearity)} \tag{A.1}$$

$$k = 0.04 \qquad \text{(prefactor of SSN's nonlinearity)} \tag{A.2}$$

$$\tau_A = 5\,\text{ms} \qquad \text{(decay time of AMPA)} \tag{A.3}$$

$$\tau_G = 7\,\text{ms} \qquad \text{(decay time of GABA)} \tag{A.4}$$

$$\tau_N = 100\,\text{ms} \qquad \text{(decay time of NMDA)} \tag{A.5}$$

$$\tau_{corr} = 5\,\text{ms} \qquad \text{(correlation time of noise)} \tag{A.6}$$

To find the fixed-point firing rates of the network without considering different synaptic types, i.e. in a simple E-I network, we use the following time scales for excitation and inhibition

$$\tau_E = 30\ \text{ms} \tag{A.7}$$

$$\tau_I = 10\ \text{ms} \tag{A.8}$$

Recall that such simple E-I networks have the same numerical value of fixed-point firing rates as the expanded synaptic version. To ensure that the fixed point remains stable going from the simple E-I network to the network with various synaptic types, we checked that the real part of the eigenvalues of the Jacobian of our network were negative.

To study the retinotopic model, we expanded our simulation to a 17×17 grid with 2 units (one E and one I) at each location. Thus we expanded the number of units of our simulation from 2 to $17^2 * 2 = 578$. We kept the following spatial

parameters fixed as well for our retinotopic simulations

$$w_{RF} = 0.04^\circ \qquad \text{(grating input margin width)} \tag{A.9}$$

$$\sigma_{Gabor} = 0.5^\circ \qquad \text{(Gabor stimulus sigma)} \tag{A.10}$$

$$\Delta x = 0.4\,\text{mm} \qquad \text{(column size)} \tag{A.11}$$

$$N_{\text{grid}} = 17^2 \qquad \text{(number of columns)} \tag{A.12}$$

$$L = 6.4\,\text{mm} \qquad \text{(grid side length)} \tag{A.13}$$

$$M = 2\,\frac{\text{mm}}{\text{deg.}} \qquad \text{(cortical magnification factor)} \tag{A.14}$$

And finally we fixed the following parameters for the scaling of connections within our retinotopic model:

$$\sigma_{aI} = 0.09 \text{ mm for } a \in [E, I] \tag{A.15}$$

A.1. Random explorations of model parameter spaces

For studying the robustness of the contrast dependence of gamma frequency in the two-population model (results shown in Fig. 2.2), we allowed the parameters (*e.g.* J_{EE}, J_{EI}, J_{IE}, J_{II}, g_E, g_I, ρ_{NMDA}) to vary widely within biologically plausible ranges. To determine the ranges for the recurrent connection strength we first made rough biological point estimates for the recurrent E and I weights (*i.e.*, J_{aE} and J_{aI}, respectively, for $a \in \{E, I\}$), as well as the (excitatory) feedforward weights (g_E and g_I); we denote these estimates by J_E^*, J_I^* and g^*, respectively (see below for a description of how we estimated these estimates). Finally, we assumed that recurrent V1 excitatory synapses are dominated by AMPA, rather than NMDA, and therefore sampled ρ_N uniformly at random in the interval $[0, 0.5]$.

The ranges of all varied model parameters were thus given by

$$J_{EE} \in [0.5\, J_E^*, 1.5\, J_E^*] = [2.2, 6.6] \tag{A.16}$$

$$J_{EI} \in [0.5\, J_I^*, 1.5\, J_I^*] = [1.1, 3.3] \tag{A.17}$$

$$J_{IE} \in [0.5\, J_E^*, 1.5\, J_E^*] = [2.2, 6.6] \tag{A.18}$$

$$J_{II} \in [0.5\, J_E^*, 1.5\, J_E^*] = [1.1, 3.3] \tag{A.19}$$

$$g_E \in [0.5\, g^*, 1.5\, g^*] = [0.22, 0.66] \tag{A.20}$$

$$g_I \in [0.5\, g^*, 1.5\, g^*] = [0.22, 0.66] \tag{A.21}$$

$$\rho_N \in [0, 0.5] \tag{A.22}$$

We then independently sampled the different parameters, uniformly at random, from the above ranges. However, we rejected some sampled parameter sets, to ensure stability of the fixed points and lack of excitatory supersaturation based on criteria found in [46]. The first criteria was to reject parameter sets for which $J_{EE}J_{II} > J_{EI}J_{IE}$, or where the product of strengths of intra-population connections was greater than the product of inter-population connections. [46] found the opposite inequality, *i.e.*, $J_{EE}J_{II} < J_{EI}J_{IE}$, to be a necessary condition for the existence of a stable fixed point. We moreover enforced another inequality involving the recurrent and feedforward connection weights, in order to prevent (strong) supersaturation of excitatory rates and guarantee sufficient growth of those rates with contrast [46].It is worth noting that Within the biological regime there is a marginal and small volume of parameter space that allows for sufficiently strong excitatory rates despite their supersaturation at very high contrasts; here, mostly in order to speed up sampling, we did not explore this region of the parameter space. We thus rejected parameter choices where $J_{II}g_E > J_{EI}g_I$, to

avoid the so-called supersaturation of excitatory firing rates in [46]. Parameter choices which violated the above conditions were then re-drawn from their distributions until the above criteria were satisfied.

To estimate the mid-range estimates, J_E^*, J_I^* and g^*, we relied on estimates of the effect of recurrent and feedforward inputs on the membrane voltage of a post-synaptic neuron. To connect these latter estimates to our model parameters, however, we first need to express SSN's powerlaw input-output function, Eq. (2.1), more biologically, in terms of the relationship between membrane voltage and output firing rate: $r = k_{\text{bio}}(\Delta V)^n$, where ΔV denotes the raise in membrane potential from baseline (at which the mean firing rate is assumed to be very low, and hence neglected). From [52, 60], a $\Delta V \sim 20$ mV leads to a firing rate of ~ 80 Hz. For a power-law exponent of $n = 2$, this translates to a power-law prefactor $k_{\text{bio}} = (80\,\text{Hz})/(20\,\text{mV})^2 \sim 0.2\,\text{Hz/mV}^2$. In our model, we have instead used the value $k = 0.04$, Eq. (A.2), for the power-law pre-factor, for consistency with previous SSN publications [45, 46]). Thanks to the scaling property of the power law, however, we can convert from the biological units to our arbitrary units, by rescaling estimated changes in the voltage (due to feedforward or recurrent inputs) by $\gamma \equiv (k_{\text{bio}}/k)^{\frac{1}{n}} = \sqrt{0.2/0.04} = \sqrt{5}$.

We estimated the effect of feedforward inputs on membrane voltage using measurements in cats and mice [52, 67, 86, 87, 88] (see [89] for a review and discussion of these measurements). Based on these measurements, we estimate the maximum feedforward input, achieved for 100% contrast to be on the order of the rest to threshold distance, which is around 20 mV [**?**]. This yields $g^* = \gamma(20\text{mV})/(100\%) = 0.2\sqrt{5} \simeq 0.447$, used in Eqs. (A.20)–(A.21).

To estimate the total excitatory recurrent weight, J_{aE} ($a \in \{E, I\}$), suppose that a typical V1 neuron of type a receives K_{aE} excitatory inputs from within V1. Suppose these inputs produce unitary post-synaptic potentials (PSP) that have an exponential time course, with a mean amplitude j_{aE} and time-constant $\tau_{\rm PSP}$, and have mean rate r_E. Then the mean depolarization produced by these excitatory inputs is $\Delta V \sim j_{aE} K_{aE} \tau_{\rm PSP} r_E$. After conversion to our units, we thus obtain the estimate $J_{aE} = \gamma \Delta V / \Delta r_E \sim \gamma\, j_{aE} K_{aE} \tau_{\rm PSP}$. Based on anatomical measurements for sensory cortex (reviewed in [89]) we estimate $K_{aE} \sim 400$ (with uncertainty around ± 200). And based on electrophysiological measurements we assume the median EPSP amplitude to be ~ 0.5 mV. For fast synapses the PSP time-course is mostly set by the membrane time constant, so we take $\tau_{\rm PSP} \sim \tau_{\rm mem} \sim 10$ ms $= 0.01$ Hz^{-1}. We thus obtain $J_E^* = \sqrt{5} \times 0.5 \times 400 \times 0.01 \simeq 4.47$ Hz^{-1}, which then lead to the ranges Eq. (A.16) and (A.18). The obvious modification of the above yields an estimate for total inhibitory weights, J_{aI}. For inhibitory PSP amplitudes, we used the value of 0.5 mV, but we assumed half as many inhibitory pre-synaptic inputs, K_{aI}, due to the smaller number of inhibitory cells in the circuit. We thus took $J_I^* = 0.5 J_E^*$, which lead to the ranges Eq. (A.17) and (A.19).

In Fig. 2.5, we studied the robustness of results obtained in the retinotopic model of Fig. 2.4 to local parameter perturbations. We limited the relative strength of these perturbations to within $\pm 10\%$ of each parameter's unperturbed value. More explicitly, we independently sampled the model parameters, uniformly

at random, from the following ranges:

$$J_{EE} \in [3.99, 4.89] \tag{A.23}$$

$$J_{EI} \in [1.48, 1.82] \tag{A.24}$$

$$J_{IE} \in [4.52, 6.76] \tag{A.25}$$

$$J_{II} \in [1.12, 1.24] \tag{A.26}$$

$$g_E \in [0.33, 0.41] \tag{A.27}$$

$$g_I \in [0.24, 0.29] \tag{A.28}$$

$$\rho_N \in [0.45, 0.55] \tag{A.29}$$

$$\lambda_{EE} \in [0.36, 0.44] \tag{A.30}$$

$$\lambda_{EI} \in [0.63, 0.77] \tag{A.31}$$

$$\sigma_{EE} \in [0.18, 0.22] \text{ mm} \tag{A.32}$$

$$\sigma_{EI} \in [0.36, 0.44] \text{ mm} \tag{A.33}$$

When studying the retinotopic SSN with a smooth fall-off of excitatory horizontal connectivity (*i.e.*, with $\lambda_{EE} = \lambda_{IE} = 0$) we again varied the parameters over the full range of their biologically plausible values. Thus the shared parameters between the two-population networks and this retinotopic model were sampled as described above, from the ranges Eqs. (A.16)–(A.22). We also chose a biologically plausible range for the length-scales of long-range excitatory horizontal connections, $\sigma_{E,E}$ and $\sigma_{I,E}$:

$$\sigma_{EE}, \sigma_{IE} \in [0.15, 0.5] \text{ mm} \tag{A.34}$$

We sampled these randomly and uniformly from the above range, subject to the condition $\sigma_{I,E} > \sigma_{E,E}$, which is necessary for obtaining surround suppression [45].

APPENDIX B

MATHEMATICAL DETAILS FOR CH. III

In this appendix we give the value of the parameter in the various 1D and 2D models of thalamocortical connectivity development studied in Ch. III. There were several parameters which remained fixed across our simulations (*i.e.* one-dimensional ring vs two-dimensional grid networks).

$$\tau_W = 1000\,\text{ms} \quad \text{(plasticity time-scale)} \tag{B.1}$$

$$\tau_{slow} = 100\,\text{ms} \quad \text{(slow feature time-scale)} \tag{B.2}$$

$$\tau_{fast} = 1\,\text{ms} \quad \text{(fast feature time-scale)} \tag{B.3}$$

Recall from Sec. 3.2.3 that we assumed $\gamma_i = \frac{1}{\tau_i}$, or that the relaxation time-scales of the linear response kernels ($\frac{1}{\gamma_i}$) were the same as the correlation times of the inputs (τ_i), at the same time-scale. The time-step of our integration method is set during the first time-step such that the standard deviation of the change in synaptic weights is 0.01, see Sec. 3.4.5. Moreover we kept the same bounds on all elements of W in all models:

$$W_{max} = 4 \quad \text{(upper bound } W\text{)} \tag{B.4}$$

$$W_{min} = 0 \quad \text{(lower bound } W\text{)} \tag{B.5}$$

In the simulations of all one-dimensional ring models, Sec 3.3.1, we used

$$
\begin{aligned}
L &= 10\,\text{mm} && \text{(half the circumference of the ring)} && \text{(B.6)}\\
\mathbf{x} &\in && [-L, L] && \text{(B.7)}\\
\mathbf{a} &\in && [-L, L] && \text{(B.8)}\\
N_{post} &= 1024 && \text{(number of post-synaptic neurons)} && \text{(B.9)}\\
N_{pre} &= 1024 && \text{(number of pre-synaptic neurons)} && \text{(B.10)}\\
& && && \text{(B.11)}
\end{aligned}
$$

for the spatial grid parameters. We simulated two different types of one-dimensional ring networks which differed in the type of their input correlation. In the simpler version of the model (with results shown in Fig. 3.6), we used single-rank input correlations, $C^j = \mathbf{e}^j(\mathbf{e}^j)^T$, at both the slow and the fast time-scales. We randomly generated the principal eigenvectors $\mathbf{e}^{slow}$ and $\mathbf{e}^{fast}$, and then made $\mathbf{e}^{fast}$ orthogonal to $\mathbf{e}^{slow}$ using the so-called Graham-Schmidt orthonormaliztion procedure. As such, C^{slow} and C^{fast} were single-rank projection operators for this first choice of input correlation structure (see Fig. 3.6).

In a more complex version of the 1D ring models, we used a Mexican hat or difference of Gaussian (DoG) spatial profile for the input correlations at both the slow and the fast time-scales. The DoG profile is in general given by

$$DoG(\mathbf{x}, \mathbf{y}) = \alpha_+ \mathcal{N}_+ e^{-\frac{|\mathbf{x}-\mathbf{y}|^2}{2\sigma_+}} - \alpha_- \mathcal{N}_- e^{-\frac{|\mathbf{x}-\mathbf{y}|^2}{2\sigma_-}} \qquad \text{(B.12)}$$

where $\mathcal{N}_+$ and $\mathcal{N}_-$ were chosen such that with α_E and α_I set to one, the absolute value of each term in Eq. (B.12) would yield one if summed over $\mathbf{y}$. We call this

the row-normalization of the lobes. For the ring networks, the spatial length scales (σ) and strengths (α) of the DoG lobes underlying the input correlation structure we used were

$$\sigma_{+,C}^{slow} = 0.3\,\mathrm{mm} \tag{B.13}$$

$$\sigma_{-,C}^{slow} = 1\,\mathrm{mm} \tag{B.14}$$

$$\sigma_{+,C}^{fast} = 0.033\,\mathrm{mm} \tag{B.15}$$

$$\sigma_{-,C}^{fast} = 0.1\,\mathrm{mm} \tag{B.16}$$

$$\alpha_{+} = 10 \tag{B.17}$$

$$\alpha_{-} = 10 \tag{B.18}$$

We used the same α_+ and α_- in the slow and fast input correlation matrices. Because we row-normalized each Gaussian lobe, $C^j(0,0)$ is nonzero for $j \in [slow, fast]$.

While the input correlations had a DoG spatial profile at both time-scales, only the cortical linear response kernel at the slow time-scale had a DoG profile. By contrast, the cortical linear response kernel at the fast time-scale was chosen as a very narrow Gaussian function to approximate a delta-function. In the notation introduced above, in the DoG for the slow linear response kernel we used

$$\sigma_{+,K}^{slow} = 0.5\,\mathrm{mm} \tag{B.19}$$

$$\sigma_{-,K}^{slow} = 1.5\,\mathrm{mm} \tag{B.20}$$

$$\alpha_{+,K}^{slow} = 10 \tag{B.21}$$

$$\alpha_{-,K}^{slow} = 10 \tag{B.22}$$

On the other hand, we modelled the fast linear response kernel as approximating a Kronecker-δ. For the fast linear response kernel we used

$$K^{fast}(\mathbf{x},\mathbf{y}) = \alpha^{fast} e^{-\frac{|\mathbf{x}-\mathbf{y}|^2}{2\sigma^{fast}}} \tag{B.23}$$

$$\sigma^{fast} = 0.01\,\text{mm} \tag{B.24}$$

$$\alpha^{fast} = 0.67 \tag{B.25}$$

where σ^{fast} is smaller than the grid spacing $dx = 0.0195$ mm, and $alpha^{fast}$ was chosen such that the growth of selectivity to the $fast$ feature was approximately as strong as the growth of selectivity to the *slow* feature. Here expand on the discussion before Eq. (3.33) about normalzation. After constructing the input correlations, we normalized their respective principal eigenvalues to one, such that the relative selectivity to slow and fast features could be controlled with λ_{slow} and λ_{fast}.

$$\lambda_{slow} = 1 \tag{B.26}$$

$$\lambda_{fast} = 10 \tag{B.27}$$

For the one-dimensional ring networks, we did not normalize the principal eigenvalue of the cortical response kernels to one.

For the one-dimensional ring networks with DoG input correlation structures, we used a "pill-box" arbor (see Sec. 3.4.3) with radius

$$R_A = 1\,\text{mm}\,(\text{arbor radius}) \tag{B.28}$$

We now describe the parameters of the two-dimensional networks. In order to keep the same number of pre- and post-syanptic neurons between the one-dimensional ring networks and the two-dimensional grid networks, we choose the number of neurons per dimension $n_i = 32$ where $i \in [pre, post]$. The number of neurons is then $N_i = n_i^d$ where d is the dimensionality of the network ($d = 1$ for ring networks, $d = 2$ for grid networks). For the two-dimensional networks, we give all length-scales in units of the grid spacing dx; these values can be converted to millimeters by multiplying them by $dx = 0.0195$ mm. The grid specifications were:

$$L = 16 \quad \text{(half of the grid side length)} \tag{B.29}$$

$$\mathbf{x} \in ([-L, L], [-L, L]) \tag{B.30}$$

$$\mathbf{a} \in ([-L, L], [-L, L]) \tag{B.31}$$

$$\tag{B.32}$$

For the grid networks, we used the circular overlap arbor function as described in Sec. 3.4.3 (see Eqns. 3.48 and 3.49) with parameters

$$R_A = 6.5 \quad \text{(arbor radius)} \tag{B.33}$$

$$c_A = 0.5 \quad \text{(scalar of the axon circle radius).} \tag{B.34}$$

Recall that for a circular overlap arbor, we model the dendrites of the post-synaptic neuron and the axon of the pre-synaptic neuron as being circles. In analogy with biology where dendrites have a greater bounding area than the axons, we scale the arbor circle by c_A such that it is smaller than the dendritic circle. The value of the arbor is then determined by the amount of overlap between the two

circles, with complete overlap being one. Fig 3.16 shows a representative arbor made using this circular overlap. The result is an arbor which tapers at the edges.

We again used DoG structures for the spatial profiles of the input correlations and the cortical linear response kernels, but now $\mathbf{x}$ and $\mathbf{y}$ are two-dimensional vectors rather than one-dimensional. For these two-dimensional simulations, we set $\mathcal{N}_+$ and $\mathcal{N}_-$ to one to be more consistent with previous work [22]. For the length scales of the input correlation structures we used (where again the units are in terms of grid spacing)

$$\sigma_{+,C}^{slow} = 0.92 \tag{B.35}$$

$$\sigma_{-,C}^{slow} = 2.76 \tag{B.36}$$

$$\sigma_{+,C}^{fast} = 0.31 \tag{B.37}$$

$$\sigma_{-,C}^{fast} = 0.92 \tag{B.38}$$

$$\alpha_{+,C} = 1 \tag{B.39}$$

$$\alpha_{-,C} = 1/9 \tag{B.40}$$

For our linear response kernels, we used

$$\sigma_{+,K}^{slow} = 1.38 \tag{B.41}$$

$$\sigma_{-,K}^{slow} = 4.14 \tag{B.42}$$

$$\alpha_{+,K}^{slow} = 1 \tag{B.43}$$

$$\alpha_{-,K}^{slow} = 1/9 \tag{B.44}$$

for the slow kernel, and for the fast kernel we used

$$\sigma^{fast} = 0.01 \tag{B.45}$$

$$\alpha^{fast} = 1 \tag{B.46}$$

where again we assumed the linear response kernel at the fast time-scale approximated a Kronecker-δ. Moreover, we normalized the principal eigenvalues of the input correlation structures and the cortical linear response kernels to one when we simulated two-dimensional networks. As discussed before Eq. (3.33), by normalizing our eigenvalues we could control the relative strengths of the $fast$ and *slow* feature using

$$\lambda^{slow} = 1 \tag{B.47}$$

$$\lambda^{fast} = 5.5 \tag{B.48}$$

$$\ell^{fast} = 2.5 \tag{B.49}$$

REFERENCES CITED

[1] Pablo Garcia-Lopez, Virginia Garcia-Marin, and Miguel Freire. The histological slides and drawings of Cajal. *Frontiers in Neuroanatomy*, 4(MARCH):1–16, 2010. ISSN 16625129. doi: 10.3389/neuro.05.009.2010.

[2] Peter Dayan and L.F. Abbott. *Theoretical Neuroscience*. 2001.

[3] Luc J. Gentet, Greg J. Stuart, and John D. Clements. Direct measurement of specific membrane capacitance in neurons. *Biophysical Journal*, 79(1): 314–320, 2000. ISSN 00063495. doi: 10.1016/S0006-3495(00)76293-X. URL `http://dx.doi.org/10.1016/S0006-3495(00)76293-X`.

[4] Rebecca Lewis, Katie E. Asplin, Gareth Bruce, Caroline Dart, Ali Mobasheri, and Richard Barrett-Jolley. The role of the membrane potential in chondrocyte volume regulation. *Journal of Cellular Physiology*, 226(11): 2979–2986, 2011. ISSN 00219541. doi: 10.1002/jcp.22646.

[5] Orientation Columns, Maps Of, and T H E Cortex. Hubel D.H. - Eye, Brain, and Vision (1995).

[6] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962. ISSN 14697793. doi: 10.1113/jphysiol.1962.sp006837.

[7] Matthias Kaschube. Neural maps versus salt-and-pepper organization in visual cortex. *Current Opinion in Neurobiology*, 24(1):95–102, 2014. ISSN 09594388. doi: 10.1016/j.conb.2013.08.017. URL `http://dx.doi.org/10.1016/j.conb.2013.08.017`.

[8] J. Sebastian Espinosa and Michael P. Stryker. Development and Plasticity of the Primary Visual Cortex. *Neuron*, 75(2):230–249, 2012. ISSN 08966273. doi: 10.1016/j.neuron.2012.06.009.

[9] Antonella Antonini and M. Stryker. Rapid remodeling of axonal arbors in the visual cortex. *Science*, 260(5115):1819–1821, jun 1993. ISSN 0036-8075. doi: 10.1126/science.8511592. URL `https://www.sciencemag.org/lookup/doi/10.1126/science.8511592`.

[10] M. F. Bear, A. Kleinschmidt, Q. Gu, and W. Singer. Disruption of experience-dependent synaptic modifications in striate cortex by infusion of an NMDA receptor antagonist. *Journal of Neuroscience*, 10(3):909–925, 1990. ISSN 02706474. doi: 10.1523/jneurosci.10-03-00909.1990.

[11] Jianhua Cang, Megumi Kaneko, Jena Yamada, Georgia Woods, Michael P. Stryker, and David A. Feldheim. Ephrin-As guide the formation of functional maps in the visual cortex. *Neuron*, 48(4):577–589, 2005. ISSN 08966273. doi: 10.1016/j.neuron.2005.10.026.

[12] Audrey Dufour, Julie Seibt, Lara Passante, Vanessa Depaepe, Thomas Ciossek, Jonas Frisén, Klas Kullander, John G. Flanagan, Franck Polleux, and Pierre Vanderhaeghen. Area specificity and topography of thalamocortical projections are controlled by ephrin/Eph genes. *Neuron*, 39(3):453–465, 2003. ISSN 08966273. doi: 10.1016/S0896-6273(03)00440-9.

[13] Donald Hebb. *The Organization of Behavior*. 1949. ISBN 0805843000.

[14] Henry Dale. Pharmacology and Nerve-Endings. *Proceedings of the Royal Society of Medicine*, 28(3):319–332, jan 1935. ISSN 0035-9157. doi: 10.1177/003591573502800330. URL `http://journals.sagepub.com/doi/10.1177/003591573502800330`.

[15] J. C. Eccles, P Fatt, and K Koketsu. Cholinergic and inhibitory synapses in a pathway from motor-axon collaterals to motoneurones. *The Journal of Physiology*, 126(3):524–562, dec 1954. ISSN 00223751. doi: 10.1113/jphysiol.1954.sp005226. URL `http://doi.wiley.com/10.1113/jphysiol.1954.sp005226`.

[16] Piergiorgio Strata and Robin Harvey. Dale's principle. *Brain Research Bulletin*, 50(5-6):349–350, 1999. ISSN 03619230. doi: 10.1016/S0361-9230(99)00100-8.

[17] Bernard Widrow, Youngsik Kim, Dookun Park, and Jose Krause Perin. Nature's Learning Rule. In *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, pages 1–30. Elsevier, 2019. doi: 10.1016/B978-0-12-815480-9.00001-3. URL `https://linkinghub.elsevier.com/retrieve/pii/B9780128154809000013`.

[18] Alain Destexhe and Denis Paré. Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *Journal of Neurophysiology*, 81(4):1531–1547, 1999. ISSN 00223077. doi: 10.1152/jn.1999.81.4.1531.

[19] Romain Brette. Philosophy of the spike: Rate-based vs. Spike-based theories of the brain. *Frontiers in Systems Neuroscience*, 9(November):1–14, 2015. ISSN 16625137. doi: 10.3389/fnsys.2015.00151.

[20] Xiaoxuan Jia, Dajun Xing, and Adam Kohn. No consistent relationship between gamma power and peak frequency in macaque primary visual cortex. *Journal of Neuroscience*, 33(1):17–25, 2013. ISSN 02706474. doi: 10.1523/JNEUROSCI.1687-12.2013.

[21] Supratim Ray and John H R Maunsell. Differences in Gamma Frequencies across Visual Cortex Restrict Their Possible Use in Computation. *Neuron*, 67 (5):885–896, 2010. ISSN 08966273. doi: 10.1016/j.neuron.2010.08.004. URL http://dx.doi.org/10.1016/j.neuron.2010.08.004.

[22] Kenneth D. Miller. A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. *Journal of Neuroscience*, 14(1):409–441, 1994. ISSN 02706474. doi: 10.1523/jneurosci.14-01-00409.1994.

[23] Kevan A.C. Martin and Sylvia Schröder. Functional heterogeneity in neighboring neurons of cat primary visual cortex in response to both artificial and natural stimuli. *Journal of Neuroscience*, 33(17):7325–7344, 2013. ISSN 02706474. doi: 10.1523/JNEUROSCI.4071-12.2013.

[24] Wolf Singer. Neuronal Synchrony: A Versatile Code for the Definition of Relations? *Neuron*, 24(1):49–65, sep 1999. ISSN 08966273. doi: 10.1016/S0896-6273(00)80821-1. URL https://linkinghub.elsevier.com/retrieve/pii/S0896627300808211.

[25] Pascal Fries. A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10):474–480, 2005. ISSN 13646613. doi: 10.1016/j.tics.2005.08.011.

[26] Pascal Fries. Rhythms for Cognition: Communication through Coherence. *Neuron*, 88(1):220–235, 2015. ISSN 10974199. doi: 10.1016/j.neuron.2015.09.034. URL http://dx.doi.org/10.1016/j.neuron.2015.09.034.

[27] Jianguang Ni, Thomas Wunderle, Christopher Murphy Lewis, Robert Desimone, Ilka Diester, and Pascal Fries. Gamma-Rhythmic Gain Modulation. *Neuron*, 92(1):240–251, oct 2016. ISSN 08966273. doi: 10.1016/j.neuron.2016.09.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0896627316305621.

[28] György Buzsáki and James J. Chrobak. Temporal structure in spatially organized neuronal ensembles: a role for interneuronal networks. *Current Opinion in Neurobiology*, 5(4):504–510, 1995. ISSN 09594388. doi: 10.1016/0959-4388(95)80012-3.

[29] John G.R. Jefferys, Roger D. Traub, and Miles A. Whittington. Neuronal networks for induced '40 Hz' rhythms. *Trends in Neurosciences*, 19(5): 202–208, 1996. ISSN 01662236. doi: 10.1016/S0166-2236(96)10023-0.

[30] Andreas Draguhn and György Buzsáki. Neuronal Oscillations in Cortical Networks. *Science*, 304(June):1926–1930, 2004.

[31] J. J. Hopfield. Encoding for computation: Recognizing brief dynamical patterns by exploiting effects of weak rhythms on action-potential timing. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):6255–6260, 2004. ISSN 00278424. doi: 10.1073/pnas.0401125101.

[32] Pascal Fries, Danko Nikolić, and Wolf Singer. The gamma cycle. *Trends in Neurosciences*, 30(7):309–316, 2007. ISSN 01662236. doi: 10.1016/j.tins.2007.05.005.

[33] J. Andrew Henrie and Robert Shapley. LFP power spectra in V1 cortex: The graded effect of stimulus contrast. *Journal of Neurophysiology*, 94(1):479–490, 2005. ISSN 00223077. doi: 10.1152/jn.00919.2004.

[34] M. A. Gieselmann and A. Thiele. Comparison of spatial integration and surround suppression characteristics in spiking activity and the local field potential in macaque V1. *European Journal of Neuroscience*, 28(3):447–459, 2008. ISSN 0953816X. doi: 10.1111/j.1460-9568.2008.06358.x.

[35] Misha V Tsodyks, William E Skaggs, Terrence J Sejnowski, and Bruce L McNaughton. Paradoxical effects of external modulation of inhibitory interneurons. *Journal of Neuroscience*, 17(11):4382–4388, 1997. ISSN 02706474.

[36] Nicolas Brunel and Xiao Jing Wang. What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. *Journal of Neurophysiology*, 90(1):415–430, 2003. ISSN 00223077. doi: 10.1152/jn.01095.2002.

[37] Roger D. Traue, John G.R. Jefferys, and Miles A. Whittington. Simulation of gamma rhythms in networks of interneurons and pyramidal cells. *Journal of Computational Neuroscience*, 4(2):141–150, 1997. ISSN 09295313. doi: 10.1023/A:1008839312043.

[38] Christoph Börgers and Nancy Kopell. Synchronization in Networks of Excitatory and Inhibitory Neurons with Sparse, Random Connectivity. *Neural Computation*, 15(3):509–538, mar 2003. ISSN 0899-7667. doi: 10.1162/089976603321192059. URL `papers2://publication/uuid/1E3B8B77-5F7D-47AD-B522-7F0164611E9B` `http://www.mitpressjournals.org/doi/10.1162/089976603321192059`.

[39] Miles A. Whittington, Mark O. Cunningham, Fiona E.N. LeBeau, Claudia Racca, and Roger D. Traub. Multiple origins of the cortical gamma rhythm. *Developmental Neurobiology*, 71(1):92–106, 2011. ISSN 19328451. doi: 10.1002/dneu.20814.

[40] Sami El Boustani, Martin Pospischil, Michelle Rudolph-Lilith, and Alain Destexhe. Activated cortical states: Experiments, analyses and models. *Journal of Physiology Paris*, 101(1-3):99–109, 2007. ISSN 09284257. doi: 10.1016/j.jphysparis.2007.10.001.

[41] György Buzsáki and Kenji Mizuseki. The log-dynamic brain: How skewed distributions affect network operations. *Nature Reviews Neuroscience*, 15(4): 264–278, 2014. ISSN 14710048. doi: 10.1038/nrn3687.

[42] C. D. Gilbert and T. N. Wiesel. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, 9 (7):2432–2422, 1989. ISSN 02706474. doi: 10.1523/jneurosci.09-07-02432.1989.

[43] James R. Cavanaugh, Wyeth Bair, and J. Anthony Movshon. Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *Journal of Neurophysiology*, 88(5):2547–2556, 2002. ISSN 00223077. doi: 10.1152/jn.00693.2001.

[44] Lars Schwabe, Jennifer M. Ichida, S. Shushruth, Pradeep Mangapathy, and Alessandra Angelucci. Contrast-dependence of surround suppression in Macaque V1: Experimental testing of a recurrent network model. *NeuroImage*, 52(3):777–792, 2010. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.01.032. URL `http://dx.doi.org/10.1016/j.neuroimage.2010.01.032`.

[45] Daniel B Rubin, Stephen D VanHooser, and Kenneth D Miller. The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 85(2):402–417, 2015. ISSN 10974199. doi: 10.1016/j.neuroN.2014.12.026.

[46] Yashar Ahmadian, Daniel B. Rubin, and Kenneth D. Miller. Analysis of the stabilized supralinear network. *Neural Computation*, 25(8):1994–2037, 2013. ISSN 08997667. doi: 10.1162/NECO_a_00472.

[47] Samuel P. Burns, Dajun Xing, Michael J. Shelley, and Robert M. Shapley. Searching for autocoherence in the cortical network with a time-frequency analysis of the local field potential. *Journal of Neuroscience*, 30(11): 4033–4047, 2010. ISSN 02706474. doi: 10.1523/JNEUROSCI.5319-09.2010.

[48] Samuel P. Burns, Dajun Xing, and Robert M. Shapley. Is gamma-band activity in the local field potential of v1 cortex a "clock" or filtered noise? *Journal of Neuroscience*, 31(26):9658–9664, 2011. ISSN 02706474. doi: 10.1523/JNEUROSCI.0660-11.2011.

[49] Dajun Xing, Yutai Shen, Samuel Burns, Chun I Yeh, Robert Shapley, and Wu Li. Stochastic generation of gamma-band activity in primary visual cortex of awake and anesthetized monkeys. *Journal of Neuroscience*, 32(40): 13873–13880, 2012. ISSN 02706474. doi: 10.1523/JNEUROSCI.5644-11.2012.

[50] Kukjin Kang, Michael Shelley, James Andrew Henrie, and Robert Shapley. LFP spectral peaks in V1 cortex: network resonance and cortico-cortical feedback. *Journal of Computational Neuroscience*, 29(3):495–507, dec 2010. ISSN 0929-5313. doi: 10.1007/s10827-009-0190-2. URL `http://link.springer.com/10.1007/s10827-009-0190-2`.

[51] Erwan Ledoux and Nicolas Brunel. Dynamics of networks of excitatory and inhibitory neurons in response to time-dependent inputs. *Frontiers in Computational Neuroscience*, 5(May):1–17, 2011. ISSN 16625188. doi: 10.3389/fncom.2011.00025.

[52] Ian M. Finn, Nicholas J. Priebe, and David Ferster. The Emergence of Contrast-Invariant Orientation Tuning in Simple Cells of Cat Visual Cortex. *Neuron*, 54(1):137–152, 2007. ISSN 08966273. doi: 10.1016/j.neuron.2007.02.029.

[53] JS Anderson, I Lampl, DC Gillespie, and D Ferster. The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science (New York, NY)*, 290(5498):1968–1972, January 2000.

[54] Nicholas J. Priebe and David Ferster. Inhibition, Spike Threshold, and Stimulus Selectivity in Primary Visual Cortex. *Neuron*, 57(4):482–497, 2008. ISSN 08966273. doi: 10.1016/j.neuron.2008.02.005.

[55] KD Miller and F Fumarola. Mathematical equivalence of two common forms of firing rate models of neural networks. *Neural Computation*, 24(1):25–31, January 2012.

[56] Misha Tsodyks, William E Skaggs, Terrence J Sejnowski, and Bruce L Mcnaughton. Paradoxical effects of inhibitory interneurons. *The Journal of Neuroscience*, 17(11):4382–4388, 1997. ISSN 0270-6474.

[57] Guillaume Hennequin, Yashar Ahmadian, Daniel B. Rubin, Máté Lengyel, and Kenneth D. Miller. The Dynamical Regime of Sensory Cortex: Stable Dynamics around a Single Stimulus-Tuned Attractor Account for Patterns of Noise Variability. *Neuron*, 98(4):846–860.e5, 2018. ISSN 10974199. doi: 10.1016/j.neuron.2018.04.017.

[58] Francesca Barbieri, Alberto Mazzoni, Nikos K. Logothetis, Stefano Panzeri, and Nicolas Brunel. Stimulus dependence of local field potential spectra: Experiment versus theory. *Journal of Neuroscience*, 34(44):14589–14605, 2014. ISSN 15292401. doi: 10.1523/JNEUROSCI.5365-13.2014.

[59] Nicolas Fourcaud. Fourcaud-Brunel. 2110:54, 2002. ISSN 10069291. doi: 10.1007/BF02879529. URL papers://a6591aa3-238d-4eda-a3d9-0181a4361186/Paper/p158.

[60] Nicholas J. Priebe and David Ferster. Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. *Neuron*, 45(1): 133–145, 2005. ISSN 08966273. doi: 10.1016/j.neuron.2004.12.024.

[61] Nicole Voges, Almut Schüz, Ad Aertsen, and Stefan Rotter. A modeler's view on the spatial structure of intrinsic horizontal connectivity in the neocortex. *Progress in Neurobiology*, 92(3):277–292, 2010. ISSN 03010082. doi: 10.1016/j.pneurobio.2010.05.001.

[62] Luigi Federico Rossi, Kenneth Harris, and Matteo Carandini. Excitatory and inhibitory intracortical circuits for orientation and direction selectivity. *bioRxiv*, page 556795, 2019. doi: 10.1101/556795. URL https://www.biorxiv.org/content/10.1101/556795v1.

[63] G H Henry, B Dreher, and P O Bishop. Orientation specificity of cells in cat striate cortex. *Journal of Neurophysiology*, 37(6):1394–1409, nov 1974. ISSN 0022-3077. doi: 10.1152/jn.1974.37.6.1394. URL https://www.physiology.org/doi/10.1152/jn.1974.37.6.1394.

[64] William H. Bosking, Ying Zhang, Brett Schofield, and David Fitzpatrick. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of Neuroscience*, 17(6):2112–2127, 1997. ISSN 02706474. doi: 10.1523/jneurosci.17-06-02112.1997.

[65] Andrew D. Huberman, Marla B. Feller, and Barbara Chapman. Mechanisms Underlying Development of Visual Maps and Receptive Fields. *Annual Review of Neuroscience*, (1):479–509, jul . ISSN 0147-006X. doi: 10.1146/annurev.neuro.31.060407.125533.

[66] J. M Alonso and R. C Reid. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378(November):16, 1995.

[67] Anthony D. Lien and Massimo Scanziani. Tuned thalamic excitation is amplified by visual cortical circuits. *Nature Neuroscience*, 16(9):1315–1323, 2013. ISSN 15461726. doi: 10.1038/nn.3488.

[68] Jens Kremkow, Jianzhong Jin, Yushi Wang, and Jose M. Alonso. Principles underlying sensory map topography in primary visual cortex. *Nature*, 533 (7601):52–57, 2016. ISSN 14764687. doi: 10.1038/nature17936.

[69] Gregory C. DeAngelis, Geoffrey M. Ghose, Izumi Ohzawa, and Ralph D. Freeman. Functional micro-organization of primary visual cortex: Receptive field analysis of nearby neurons. *Journal of Neuroscience*, 19(10):4046–4064, 1999. ISSN 02706474. doi: 10.1523/jneurosci.19-10-04046.1999.

[70] Yann Sweeney and Claudia Clopath. Population coupling predicts the plasticity of stimulus responses in cortical circuits. *bioRxiv*, page 265041, 2019. doi: 10.1101/265041. URL `https://www.biorxiv.org/content/10.1101/265041v2`.

[71] Shigeru Tanaka, Toshiki Tani, Jérôme Ribot, Kazunori O'Hashi, and Kazuyuki Imamura. A postnatal critical period for orientation plasticity in the cat visual cortex. *PLoS ONE*, 4(4), 2009. ISSN 19326203. doi: 10.1371/journal.pone.0005380.

[72] Kenichi Ohki, Sooyoung Chung, Prakash Kara, Mark Hübener, Tobias Bonhoeffer, and R. Clay Reid. Highly ordered arrangement of single neurons in orientation pinwheels. *Nature*, 442(7105):925–928, 2006. ISSN 14764687. doi: 10.1038/nature05019.

[73] G. G. Blasdel. Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, 12(8):3139–3161, 1992. ISSN 02706474. doi: 10.1523/jneurosci.12-08-03139.1992.

[74] Tobias Bonhoeffer and Amiram Grinvald. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, 353(6343):429–431, 1991. ISSN 00280836. doi: 10.1038/353429a0.

[75] Kenichi Ohki, Sooyoung Chung, Yeang H. Ch'ng, Prakash Kara, and R. Clay Reid. Functional imaging with cellular resolution reveals precise microarchitecture in visual cortex. *Nature*, 433(7026):597–603, 2005. ISSN 00280836. doi: 10.1038/nature03274.

[76] Sergej V. Girman, Yves Sauvé, and Raymond D. Lund. Receptive field properties of single neurons in rat primary visual cortex. *Journal of Neurophysiology*, 82(1):301–311, 1999. ISSN 00223077. doi: 10.1152/jn.1999.82.1.301.

[77] Alexei A. Koulakov and Dmitri B. Chklovskii. Orientation preference patterns in mammalian visual cortex: A wire length minimization approach. *Neuron*, 29(2):519–527, 2001. ISSN 08966273. doi: 10.1016/S0896-6273(01)00223-9.

[78] D. J.C. Mackay and K. D. Miller. Analysis of linsker's application of hebbian rules to linear networks. *Network: Computation in Neural Systems*, 1(3): 257–297, 1990. ISSN 0954898X. doi: 10.1088/0954-898X_1_3_001.

[79] K. Miller, J. Keller, and M. Stryker. Ocular dominance column development: analysis and simulation. *Science*, 245(4918):605–615, aug 1989. ISSN 0036-8075. doi: 10.1126/science.2762813. URL `https://www.sciencemag.org/lookup/doi/10.1126/science.2762813`.

[80] Kenneth D. Miller and David J. C. MacKay. The Role of Constraints in Hebbian Learning. *Neural Computation*, 6(1):100–126, jan 1994. ISSN 0899-7667. doi: 10.1162/neco.1994.6.1.100. URL `http://www.mitpressjournals.org/doi/10.1162/neco.1994.6.1.100`.

[81] T. J. Sejnowski. Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology*, 4(4):303–321, 1977. ISSN 0303-6812. doi: 10.1007/BF00275079. URL `http://link.springer.com/10.1007/BF00275079`.

[82] Stefano Fusi, Patrick J. Drew, and L. F. Abbott. Cascade models of synaptically stored memories. *Neuron*, 45(4):599–611, 2005. ISSN 08966273. doi: 10.1016/j.neuron.2005.02.001.

[83] Robert M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239, 2006. ISSN 15672190. doi: 10.1561/0100000006.

[84] Gordon B. Smith, Bettina Hein, David E. Whitney, David Fitzpatrick, and Matthias Kaschube. Distributed network interactions and their emergence in developing neocortex. *Nature Neuroscience*, 21(11):1600–1608, 2018. ISSN 15461726. doi: 10.1038/s41593-018-0247-5. URL `http://dx.doi.org/10.1038/s41593-018-0247-5`.

[85] Marcus K. Benna and Stefano Fusi. Computational principles of synaptic memory consolidation. *Nature Neuroscience*, 19(12):1697–1706, 2016. ISSN 15461726. doi: 10.1038/nn.4401.

[86] David Ferster, Sooyoung Chung, and Heidi Wheat. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature*, 380(6571): 249–252, 1996. ISSN 00280836. doi: 10.1038/380249a0.

[87] Sooyoung Chung and David Ferster. Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. *Neuron*, 20(6):1177–1189, 1998. ISSN 08966273. doi: 10.1016/S0896-6273(00)80498-5.

[88] Ling Yun Li, Ya Tang Li, Mu Zhou, Huizhong W. Tao, and Li I. Zhang. Intracortical multiplication of thalamocortical signals in mouse auditory cortex. *Nature Neuroscience*, 16(9):1179–1181, 2013. ISSN 10976256. doi: 10.1038/nn.3493.

[89] Yashar Ahmadian and Kenneth D. Miller. What is the dynamical regime of cerebral cortex? (March), 2019. URL `http://arxiv.org/abs/1908.10101`.