A Novel Circuit Model of Contextual Modulation and Normalization in Primary Visual Cortex

Daniel B. Rubin

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012
ABSTRACT

A Novel Circuit Model of Contextual Modulation and Normalization in Primary Visual Cortex

Daniel B. Rubin

The response of a neuron encoding information about a sensory stimulus is influenced by the context in which that information is presented. In the primary visual cortex (area V1), neurons respond selectively to stimuli presented to a relatively constrained region of visual space known as the classical receptive field (CRF). These responses are influenced by stimuli in a much larger region of visual space known as the extra-classical receptive field (eCRF). In that they cannot directly evoke a response from the neuron, surround stimuli in the eCRF provide the context for the input to the CRF. Though the past few decades of research have revealed many details of the complex and nuanced interactions between the CRF and eCRF, the circuit mechanisms underlying these interactions are still unknown. In this thesis, we present a simple, novel cortical circuit model that can account for a surprisingly diverse array of eCRF properties. This model relies on extensive recurrent interactions between excitatory and inhibitory neurons, connectivity that is strongest between neurons with similar stimulus preferences, and an expansive input-output neuronal nonlinearity. There is substantial evidence for all of these features in V1.

Through analytical and computational modeling techniques, we demonstrate how and why this circuit is able to account for such a comprehensive array of contextual modulations. In a linear network model, we demonstrate how surround suppression of both excitatory and
inhibitory neurons is achieved through the selective amplification of spatially-periodic patterns of activity. This amplification relies on the network operating as an inhibition-stabilized network, a dynamic regime previously shown to account for the paradoxical decrease in inhibition during surround suppression (Ozeki et al., 2009). With the addition of nonlinearity, effective connectivity strength scales with firing rate, and the network can transition between different dynamic regimes as a function of input strength. By moving into and out of the inhibition-stabilized state, the model can reproduce a number of contrast-dependent changes in the eCRF without requiring any asymmetry in the intrinsic contrast-response properties of the cells. This same model also provides a biologically plausible mechanism for cortical normalization, an operation that has been shown to be ubiquitous in V1. Through a winner-take-all population response, we demonstrate how this network undergoes a strong reduction in trial-to-trial variability at stimulus onset. We also propose a novel mechanism for attentional modulation in visual cortex. We then go on to test several of the critical predictions of the model using single unit electrophysiology. From these experiments, we find ample evidence for the spatially-periodic patterns of activity predicted by the model. Lastly, we show how this same circuit motif may underlie behavior in a higher cortical region, the lateral intraparietal area.
## Contents

List of Figures ................................................................................... v

Acknowledgments ................................................................................ ix

1 Introduction .................................................................................. 1

1.1 Primary visual cortex .............................................................. 2

1.2 Contextual modulation in V1 .................................................. 4

1.3 Underlying circuitry ............................................................... 8

1.4 Previous models ..................................................................... 9

1.5 Normalization ........................................................................ 14

1.6 Overview ................................................................................ 16

2 A Linear Model of Contextual Modulation in Visual Cortex ......... 19

2.1 Introduction and motivation .................................................. 19

2.2 Details of the linear model .................................................... 21

2.3 Simulations and parameters .................................................. 26

2.4 Results .................................................................................. 29

2.5 Analysis ................................................................................ 33

2.5.1 General features of an ISN ............................................. 33

2.5.2 Resonance, critical frequencies, and stability ................. 36

2.6 Discussion/conclusions .......................................................... 45
# Nonlinear Models of Contextual Modulation and Normalization

3.1 Introduction .................................................. 47

3.2 Nonlinear spatial model .................................. 49
   3.2.1 Introduction and motivation ......................... 49
   3.2.2 Model details and parameters ....................... 50
   3.2.3 Input dependence of dynamic regime ................. 52
   3.2.4 Results .................................................. 55
   3.2.5 Summary ................................................ 58

3.3 Normalization ............................................. 59
   3.3.1 Introduction and motivation ......................... 59
   3.3.2 Details of the model .................................. 59
   3.3.3 Results .................................................. 61
   3.3.4 Mechanisms ............................................. 63
   3.3.5 Analysis ................................................ 69
   3.3.6 Summary ................................................ 84

3.4 Large scale probabilistically connected topographic model ................. 85
   3.4.1 Details of the large scale model .................... 85
   3.4.2 Results .................................................. 87

3.5 Reduction of variability at stimulus onset ..................... 95
   3.5.1 Model details ............................................ 95
   3.5.2 Results .................................................. 99
   3.5.3 Summary ................................................ 103

3.6 Attentional modulation ................................... 103
   3.6.1 Introduction and motivation ......................... 103
   3.6.2 Results .................................................. 106
   3.6.3 Summary ................................................ 110

3.7 Discussion/conclusions .................................. 111
List of Figures

1.1 Stimulus selectivity in the CRF comes from the feedforward input. . . . . . . 3
1.2 Surround suppression is accompanied by a decrease in inhibitory conductance. 11
1.3 Schematic of the classic normalization model. ............................... 15

2.1 A cartoon of the linear firing rate model of V1 .............................. 21
2.2 The firing rate response of the model to stimuli of increasing length. . . . . 28
2.3 An ISN with spatially-periodic E and I firing rates resonates at its preferred spatial frequency. ................................................................. 31
2.4 The conductance changes in response to stimuli of increasing length. . . . . 32
2.5 Inhibition-stabilized networks respond paradoxically to increased input to inhibitory cells. ................................................................. 34
2.6 A 180° phase shift in the relative responses of E and I around a critical spatial frequency. ................................................................. 42

3.1 The input dependence of dynamic regime in the nonlinear network model. . 52
3.2 Length-tuning at multiple levels of contrast. ................................. 55
3.3 Contrast dependence of summation field size and surround valence. ....... 57
3.4 A cartoon of the nonlinear ring model. ........................................... 60
3.5 Sublinear addition of multiple stimuli. ............................................ 62
3.6 Non-orthogonal gratings are also normalized. .............................. 64
3.7 Winner-take-all responses. ............................................................ 65
3.8 Normalization versus stimulus contrast. ........................................... 66
3.9 Width-tuning in orientation space through nonlinear scaling of effective con-
nectivity. ......................................................................................... 67
3.10 Two neuron approximation of the full ring model. ......................... 72
3.11 The topographic map of preferred orientation. ............................. 86
3.12 Forty example length-tuning curves. .............................................. 87
3.13 Length- and orientation-tuning properties of the population. ........... 89
3.14 Length-tuning for different levels of stimulus strength. ................. 90
3.15 Contrast dependent contrast modulation spatial frequency tuning. ... 91
3.16 The orientation dependence of surround suppression. ................. 92
3.17 Surround suppression versus stimulus strength and coherence. ....... 94
3.18 A reduction in variability at stimulus onset. ................................. 99
3.19 Variability reduction in non-responsive cells. ............................... 101
3.20 The decrease in trial-to-trial variability depends on spatial correlations in the
   noisy input. ..................................................................................... 102
3.21 The reduction in variability in a much more complex model of cortex. .. 104
3.22 The qualitative effect of attention depends on the relative sizes of the atten-
tional and stimulus fields. ................................................................. 107
3.23 Modeling attention in the nonlinear ring model. ............................ 109
3.24 Attention modulates the strength of surround suppression in V4. ....... 111

4.1 Examples of slow changes in cortical excitability. ............................ 122
4.2 Orientation and direction tuning properties ..................................... 127
4.3 Luminance spatial frequency tuning properties .............................. 129
4.4 Six example size-tuning curves .................................................... 130
4.5 Statistics of size-tuning fits ......................................................... 133
4.6 Size-tuning at multiple levels of stimulus contrast ......................... 135
4.7 Four example position-tuning curves .......................................... 137
4.8 Statistics of position-tuning fits. .......................... 138
4.9 Luminance spatial frequency versus size and position tuning spatial frequency 140
4.10 Contrast modulation spatial frequency tuning ........................ 141

5.1 Choice task design and behavior. .......................... 151
5.2 Pairs of LIP neurons decorrelate following target onset. .......................... 153
5.3 Correlation encodes reward history. .......................... 158
5.4 Decorrelation by spatially specific saccade signal. .......................... 161
5.5 Decorrelation by spatial and non-spatial inputs in a nonlinear inhibition stabilized model. .......................... 163
5.6 Relationship between activity and correlation depends on the ratio of stimulated excitatory vs. inhibition neurons. .......................... 166

7.1 Resonant frequency increases with increasing stimulus strength. .......................... 186
7.2 Contrast-dependent changes in the eCRF are insensitive to stimulus shape. .......................... 187
7.3 The switch from facilitation to suppression depends on surround stimulus size. 188
7.4 A switch from supralinear to sublinear addition: comparison to MT data. .......................... 189
7.5 Preferred stimulus width shrinks with increasing stimulus strength in the ring model. .......................... 190
7.6 The switch from supra- to sublinear addition depends on stimulus width. .......................... 191
7.7 Twenty iterations of the iterative approximations of the contrast response curves. .......................... 192
7.8 Spatial frequency dependent paradoxical response. .......................... 193
7.9 Sublinear addition of multiple stimuli in the full model. .......................... 194
7.10 A switch from surround facilitation to surround suppression. .......................... 195
7.11 A reduction in variability at stimulus onset with a simpler noise model. .......................... 196
7.12 Variability reduction in non-responsive cells in the simpler noise model. .......................... 197
7.13 The decrease in variability is a decrease in shared noise in the simpler noise model. ................................................................. 198
7.14 The mean field model sublinearly adds multiple stimuli. ......................... 199
7.15 The decrease in variability is a decrease in shared noise in the complex model. 200
7.16 In the complex model, the reduction in variability at stimulus onset does not

  require spatially correlated external noise. ........................................ 201
7.17 In the complex model, the reduction in variability at stimulus onset does not

  require external noise. ................................................................. 202
7.18 Slow fluctuations in cortical excitability do not obscure stimulus selectivity. . 203
7.19 Boot strap analysis of size-tuning model fits. ..................................... 204
7.20 Boot strap analysis of position-tuning model fits. ............................... 205
7.21 Correlations between best fit frequencies in both the model and experiment. 206
7.22 Average correlation coefficient as a function of bin size in pre-target epoch. . 207
7.23 Correlations decrease with receptive field distance. ............................. 208
Acknowledgments

First and foremost, I wish to thank my advisor and mentor, Ken Miller. Your constant enthusiasm and support has been an inspiration to me, and your scientific rigor has demonstrated to me the right way to answer a scientific question. I am truly privileged to have had this opportunity to work alongside you these past few years.

I want to thank Larry Abbott and Stefano Fusi for their wisdom and encouragement as teachers, mentors, and role models. I also want to extend a collective thanks to Larry, Stefano, and Ken for making the Theory Center what it is. I cannot imagine a better environment in which to learn and to do science.

I am very grateful to Steve Van Hooser for graciously and patiently welcoming me into the world of experimental neuroscience, and for giving me the opportunity to put my model to the test. I also wish to thank Aniruddha Das for sharing his guidance, support, and wisdom along the way.

The work presented here would not be possible without the help of many people. It goes without saying that all of this work was completed in close collaboration with Ken Miller. The earlier modeling work was done alongside my fellow graduate student Jun Zhao. Yashar Ahmadian contributed greatly to the nonlinear analysis. Steve Van Hooser and Jared Clemens were instrumental in the design and implementation of the single unit electrophysiology experiments. Liam Paninski shared his wisdom on how to analyze the data from the single unit recordings. Annegret Falkner and Michael Goldberg designed and
implemented the experiments in LIP. Greg Wayne, Rebecca Levy, Paul Hanson, and Anne Schwieger provided helpful comments on earlier drafts of this thesis.

I am also grateful to the many teachers and friends I have had while working here in the Theory Center. In particular, I wish to thank Xaq Pitkow, Taro Toyoizumi, Omri Barak, and Srdjan Ostojic for the generosity of their time and knowledge. I am deeply grateful to Micky Vidne, David Sussillo, Evan Schaffer, Pablo Jercog, Baktash Babadi, Greg Wayne, Max Nikitchenko, Brian DePasquale, Jayant Kulkarni, Mattia Rigotti, Saul Kato, Ann Kennedy, David Pfau, Alex Ramirez, Anthony DeCostanzo, and Avi Ziskind for their friendship and support.

Most importantly, I want to thank all of my loving and supportive family. I owe an immeasurable thanks to Anne Schwieger, whose companionship, support, and encouragement has gotten me through the toughest times unscathed. To my older sister Rachel, I want to thank you for always setting the bar so high for me – your constant success is a source of inspiration. And of course, my parents, who I want to thank for everything. I know that you ultimately deserve the credit for this work, and for everything else I hope to accomplish.
Chapter 1

Introduction

One of the hallmarks of cortical computation is the parallelization and compartmentalization of information processing. At the macroscopic level, different regions of the brain are specialized for specific tasks, such as processing sensory input of a particular modality, planning the sequences of muscle contractions necessary for an action, or retrieving semantic information from a sequence of spoken words. At the microscopic level, this same pattern of parallelization and specialization is repeated. Within a particular cortical circuit, individual neurons are concerned with only a tiny piece of the world – a small patch of the visual scene, the deflection velocity of a single whisker, or the force applied across a single joint.

In spite of this extreme compartmentalization, the brain is exquisitely skilled at taking global-level information, that is, the context of a particular sensory or motor stimulus, and using it to tailor the activity of its specialized units appropriately. This too is true at both the macroscopic and microscopic levels. At the level of cortical circuits, the response of a neuron to a stimulus in its receptive field will depend strongly on the context in which it is presented, and on the stimuli presented to neurons with nearby or even distant receptive fields.

How does the brain perform this type of contextual modulation? How is information from individual neurons with relatively constrained glimpses of the sensory world pooled
1.1 Primary visual cortex

The visual cortex is made up of the regions of the brain that control and subserve the sense of vision. Visual processing occurs hierarchically; as signals are passed from one brain region to the next, processing becomes increasingly more specialized and complex. At the top of this hierarchy, brain regions are specialized for very specific visual tasks, such as face and object recognition (Fujita, 2002; Gross et al., 1972; Kobatake and Tanaka, 1994), color processing (Conway, 2009; Schein and Desimone, 1990), motion detection (Lisberger et al., 1987; Koch et al., 1989; Ilg, 2008), and behaviorally-relevant associative tasks (Bisley and Goldberg, 2010). And at the other end of the hierarchy, at the very first level of cortical processing, lies the primary visual cortex. Primary visual cortex, also called area V1 or striate cortex (named for the stria of Gennari, a band of myelinated axons that terminate in layer 4 (Kandel et al., 2000)), receives the bulk of its input directly from the lateral geniculate nucleus (LGN) of the thalamus. Primary visual cortex ultimately projects to all higher visual cortices (Van Essen, 1979; Felleman and Van Essen, 1991), and is necessary for the conscious perception of vision (Stoerig and Cowey, 1995; Cowey, 2010).
Neurons in V1 respond preferentially to oriented bars and gratings located within a small region of retinotopic space known as the classical receptive field (CRF). The CRFs in V1 are arranged topographically, such that the neurons form a map of visual space on the cortical surface and nearby neurons have neighboring CRFs. Within the CRF, cells in V1 are tuned along several axes of stimulus preference. Most notably, V1 neurons are selective for stimulus orientation (Hubel and Wiesel, 1959) (Figure 1.1A), a preference that persists at both low and high contrasts (Sclar and Freeman, 1982). Neurons are also tuned for the spatial frequency (Tolhurst and Thompson, 1981; De Valois et al., 1982; 1985), temporal frequency (Foster et al., 1985), and direction (Hubel and Wiesel, 1962) of a moving stimulus,
1.2. Contextual modulation in V1

Stimuli outside of the CRF, by definition, do not evoke a spiking response from a neuron. However, a stimulus outside the CRF can alter the response of a neuron to a stimulus within its CRF. This modulatory effect can extend a considerable distance in retinotopic space, often several times the diameter of the CRF itself. The area around the CRF in which stimuli can modulate the neural response is known as the non-classical, or extra-classical receptive field (nCRF or eCRF, used interchangeably). Stimuli in the eCRF are often suppressive, though sometimes they are facilitating. Furthermore, the magnitude of these modulatory effects can vary strongly with the specific characteristics and configuration of the stimuli both in the CRF and eCRF. In terms of computational function, the eCRF is thought to provide the context for the response of a single neuron, by providing it with information from the visual scene outside the extent of its receptive field. Contextual modulation is a ubiquitous feature of the early visual system, and is highly conserved across species (Van Hooser, 2007). However, the circuit mechanisms underlying the eCRF are still unknown.

Like the CRF, the eCRF is tuned for the orientation of bars and gratings. The presentation of a stimulus to the eCRF can suppress the response to a stimulus in the CRF, and this “surround suppression” is usually maximal for iso-oriented stimuli (Blakemore and
1.2. Contextual modulation in V1

Tobin, 1972; Nelson and Frost, 1978; Akasaki et al., 2002). Often, the orientation tuning of surround suppression is slightly broader than that of the CRF (Li and Li, 1994). Less commonly, suppression is observed for orthogonally oriented stimuli (Levitt and Lund, 1997). Surround suppression is rarely strongest for obliquely oriented surround stimuli, but, if the CRF is shown an obliquely oriented stimulus, suppression will tend to be maximal when the eCRF stimulus is of matching orientation, rather than the CRF’s preferred (Sillito et al., 1995; Cavanaugh et al., 2002b).

Sometimes, the eCRF plays a facilitatory role, and this facilitation is also orientation-dependent. In this case, the surrounding stimulus causes an increase in the center response, generally when paired with a low-contrast central stimulus (Li and Li, 1994; Sillito et al., 1995; Cavanaugh et al., 2002b; Sengpiel et al., 1997; Polat et al., 1998; Crook et al., 2002). Surround facilitation has been observed with both iso-oriented (Sengpiel et al., 1997; Sillito et al., 1995) and orthogonally-oriented surrounding stimuli (Levitt and Lund, 2002; Cavanaugh et al., 2002b; Sillito et al., 1995). The eCRF is tuned along other axes of stimulus quality as well. Surround suppression is strongest when the spatial frequency of the surrounding stimulus matches the preferred spatial frequency for the center (Li and Li, 1994). Distinct from this tuning is the preference of the eCRF for the spatial frequency of the arrangement of the surrounding stimuli (e.g. the contrast envelope of a contrast modulated surround stimulus), and this tuning is generally at much lower spatial frequencies than those observed for luminance gratings within the CRF (Tanaka and Ohzawa, 2009). The eCRF is also sensitive to the speed (Li and Li, 1994) and direction (Li and Li, 1994; Sengpiel et al., 1997; Cavanaugh et al., 2002b; Akasaki et al., 2002) of the surrounding stimuli, although these preferences seem to be less strictly enforced.

Whether or not the eCRF has a suppressive or facilitatory effect depends most strongly on the absolute and relative contrast of the stimuli in the center and surround. Though changing the contrast of the surround stimulus alters the magnitude of the modulatory effect in a fairly straightforward, monotonic fashion (Sadakane et al., 2006), a number of
eCRF properties undergo qualitative changes at different levels of center contrast. An iso-oriented, facilitatory surround stimulus to a low contrast central stimulus will often become suppressive for a high contrast central stimulus (Sengpiel et al., 1997; Polat et al., 1998). Firing rate and conductance length-tuning curves that have only a single peak at low contrast can have two peaks at high contrast (Sengpiel et al., 1997; Anderson et al., 2001). An orthogonally-oriented surround stimulus, which has on average no effect on a high contrast central stimulus, becomes suppressive for a low-contrast central stimulus (Cavanaugh et al., 2002b), and more generally, the orientation tuning of the eCRF weakens with low center contrast (Levitt and Lund, 1997). The size of the spatial summation field, measured as the region over which an increase in the diameter of a circular stimulus will cause an increase in firing rate, shrinks 2-10 fold at high contrast (Cavanaugh et al., 2002b; Sceniak et al., 1999; Song and Li, 2008).

Though the specifics of all of the eCRF properties discussed are quite variable from cell to cell, in general, the strength and frequency of these properties tend to be fairly unimodally distributed (Akasaki et al., 2002; Jones et al., 2000). That is to say, there does not appear to be a distinct class of neurons that uniformly do not undergo eCRF modulation in the same direction as the overall population. This is significant, because it implies that both the excitatory and inhibitory cells in the network are subject to the same forms of modulation. Though studies reporting cellular identities are surprisingly rare, in at least one such study, it was shown through morphological analysis that both excitatory and inhibitory neurons undergo surround suppression (Song and Li, 2008). Furthermore, there have been a number of studies showing that the inhibition received by cells is decreased when the center and surround are stimulated together (Anderson et al., 2001; Ozeki et al., 2004; 2009). Given that inhibitory connections in cortex have substantially shorter ranges than the excitatory neurons (Callaway, 2004), this would further imply that inhibitory neurons are themselves surround suppressed. This point becomes significant as we consider circuit mechanisms below.
Though the majority of work in this field has focused on primary visual cortex, contextual modulation has also been observed in higher visual areas. In visual area V2, a region with similar response properties to V1 but with more sensitivity to complex (non-luminance defined) borders and binocularity disparity (Boynton and Hegde, 2004; Orban, 2008), surround suppression is observed in neurons at about the same frequency as in V1, with a similar or perhaps slightly stronger average effect (Shushruth et al., 2009; Van den Bergh et al., 2010). Surround suppression is also observed visual area V4 (Desimone and Schein, 1987; Sundberg et al., 2009). V4 provides the major visual input to the inferior temporal cortex, a region known to be involved in object recognition (Fujita, 2002; Kobatake and Tanaka, 1994). Interestingly, in V4, the eCRF is tuned to the color of the surround stimulus, such that the response to a preferred center stimulus is maximally suppressed by a surround of the same color (Schein and Desimone, 1990). The strength of suppression even shows smooth tuning for the wavelength of the stimulus in the surround, in a manner not dissimilar from surround orientation tuning in V1. Surround suppression has also been extensively documented in area MT (Allman et al., 1985; Born and Tootell, 1992; Bradley and Andersen, 1998; Britten and Heuer, 1999; Anton-Erxleben et al., 2009), a region involved in motion detection. As in V1, the strength of surround suppression observed in MT decreases at low contrast (Pack et al., 2005), although it actually increases with decreasing stimulus coherence (Hunter and Born, 2011). In the lateral intraparietal area (LIP), a higher cortical region that codes for saccade priority in visual space, the response to a flashed stimulus is suppressed when monkeys plan a saccade to neighboring regions of visual space (Falkner et al., 2010).

Though the details of these studies, and the V1 literature in general, is extensive, it is worth noting that under natural viewing conditions, V1 neurons are rarely (if ever) exposed to stimuli that fill only the CRF (or the CRF plus some precisely defined region of the eCRF). Most of what we see fills the entire visual scene, and since the vast majority of a visual scene falls outside the CRF of any one neuron, most of the visual information
contextualizing a particular stimulus winds up in the eCRF. Of course, as V1 is comprised of a topographic map of visual space, the eCRF of one cell is simply a collection of CRFs from other cells, and similarly, the CRF of one cell may provide eCRF input for hundreds of other cells. Viewed in this light, the question of contextual modulation becomes much deeper than simply a question of who provides input to whom. Rather, there is likely a complex network of effects, and thus a question ideally suited for a computationally-driven analysis.

1.3 Underlying circuitry

Understanding the nature of stimulus selectivity in the CRF and eCRF gives insight to the structure of the underlying cortical circuitry. Within the CRF, there is considerable evidence that stimulus selectivity arises from the feedforward input from the LGN (Miller, 2003). The spatial arrangement of LGN input determines a neuron’s preferred orientation and spatial frequency (Reid and Alonso, 1995; Lampl et al., 2001) (Figure 1.1), and the width of orientation tuning can be understood from this along with feedforward inhibition (Miller, 2003; Palmer and Miller, 2007) and/or contrast-dependent changes in voltage noise (Finn et al., 2007). Temporal aspects of CRF tuning, such as temporal frequency and direction tuning, can be explained by temporally-dependent synaptic properties, such as synaptic depression (Chance et al., 1998; Kayser et al., 2001) and slow excitatory conductances (Krukowski and Miller, 2001). Stimulus phase preference in simple cells is most likely inherited from the LGN input, whereas its loss in complex cells may come from either a lack of phase specificity in the input from simple cells (Hubel and Wiesel, 1962; Adelson and Bergen, 1985) or strong local recurrent connections between complex cells (Chance et al., 1999). Consistent with a feedforward basis for the CRF selectivity, cells are more narrowly tuned for orientation when tested with higher spatial frequency gratings (Lampl et al., 2001). Additionally, after inactivation of cortex through cooling (Ferster et al., 1996) or electric shock (Chung and Ferster, 1998), cells showed the same orientation tuning as in the intact cortex, indicating
that this tuning is primarily a function of the feedforward input rather than the recurrent intracortical connectivity.

Unlike the CRF, there is substantial evidence that the eCRF emerges from either recurrent or intracortical feedback connectivity. Surround suppression in cortex is tuned for orientation (Blakemore and Tobin, 1972; Nelson and Frost, 1978; Akasaki et al., 2002), but surround suppression in LGN is relatively insensitive to orientation (Naito et al., 2007; Ozeki et al., 2009). Orientation selectivity first appears in V1, which is itself highly recurrently connected (Binzegger et al., 2004; Stepanyants et al., 2008). In particular, connectivity in V1 is selective for preferred stimulus orientation – the probability of two cells being synaptically connected is directly related to their similarity in orientation tuning, both for local (Ko et al., 2011) and especially long-range horizontal connections (Ts’o et al., 1986; Gilbert and Wiesel, 1989; Malach et al., 1993; Das and Gilbert, 1995; Bosking et al., 1997; Stettler et al., 2002). These connections can extend up to 6 to 8 mm (Gilbert, 1992), which could account for the spatial extent of the eCRF (Bair et al., 2003; Tusa et al., 1978). However, it has been argued that the speed at which suppression occurs may necessitate feedback from higher cortical areas (Angelucci and Bressloff, 2006; Angelucci and Bullier, 2003; Angelucci et al., 2002), since horizontal axons are unmyelinated and slow, whereas cortico-cortico feedback is transmitted through faster myelinated axons. Furthermore, the timing (Durand et al., 2007; Knierim and van Essen, 1992; Smith et al., 2006) and the susceptibility to adaptation (Durand et al., 2007; Webb et al., 2005) of the eCRF suggest a cortical, rather than feedforward, origin.

1.4 Previous models

There are many previously published models that explain some or many of the features of the eCRF detailed above. These models vary in their level of analysis. Some are purely phenomenological, and explain the nature of center and surround interactions in terms of the computational goals of the visual system. These include models operating to achieve
optimal decoding (Chen et al., 2006), predictive coding (Rao and Ballard, 1999; Spratling, 2010), normalization (Xing and Heeger, 2001), and saliency detection (Li, 2002; Koene and Zhaoping, 2007). Other models are more of a hybrid between phenomenological and circuit.

Often, experimental data is shown to fit well to a model or function of a particular form, which is then proposed as the basis of a potential circuit mechanism. The most popular of these would be the “Difference of Gaussians” (Sceniak et al., 1999; Levitt and Lund, 2002; DeAngelis et al., 1994; Wang et al., 2009) and “Ratio of Gaussians” (Cavanaugh et al., 2002a) models. Others, such as the one we will be presenting in subsequent chapters, are based at the circuit level, and aim to understand specifically how different classes of neurons are interacting to produce certain output behaviors.

One earlier circuit-level model that has been widely adopted specifically addresses the contrast-dependent changes in the eCRF. Along with some degree of spatially extensive recurrent connectivity, the key mechanism in this model is that the inhibitory cells in the network have both higher contrast gain and higher contrast threshold than the excitatory cells (Somers et al., 1998). At low contrast, only the excitatory neurons are above threshold in the center. When input is received from the surround (which is by definition subthreshold), only the activity of the excitatory cells increase, yielding facilitation. At higher center contrast, both excitatory and inhibitory cells are above threshold. Increased input to the local network will increase the firing rates of both excitatory and inhibitory neurons, but since inhibition has a greater gain, the net effect is suppression. This model makes the critical prediction that inhibitory cells should have detectably higher contrast gain and threshold than excitatory cells. However, intracellular recordings from putative excitatory and inhibitory cells (based on spike waveform analysis) in V1 reveal that they all have roughly the same gain and contrast threshold, and differ significantly only in their maximum firing rate (Contreras and Palmer, 2003).
1.4. Previous models

Surround suppression is accompanied by a decrease in inhibitory conductance. A. Average spike response over a cycle of stimulus presentation for a simple cell. Note surround suppression with the larger, iso-oriented stimulus, and the loss of suppression as the surround is rotated away from the orientation of the center. B. Average membrane potential across a cycle at three different levels of injected current. C–E. Cycle-averaged changes in total input conductance (C), excitatory conductance (D), and inhibitory conductance (E). Note that during surround suppression, there is a decrease in both excitatory and inhibitory conductance (From Ozeki et al. (2009)).

Another circuit-level model invoked strong recurrent inhibition (I → I connectivity) as the mechanistic basis for contrast-dependent changes in the orientation selectivity of surround modulation (Dragoi and Sur, 2000). This model requires that local I → I connections have broader orientation target specificity than E → E connections, such that an obliquely-
oriented surround (relative to the center’s preferred orientation) excites oblique-selective center I cells, which in turn inhibit preferred-selective center I cells. This releases preferred-selective center E cells from tonic levels of inhibition, establishing a tri-synaptic disinhibitory circuit that generates facilitation in the presence of an obliquely oriented surround. This model features long-range E $\rightarrow$ E connectivity with broader orientation tuning than local E $\rightarrow$ E connectivity. Though there is no real consensus on this last point, at least some studies indicate that if there is an asymmetry in the tuning of connectivity at different distances, it is likely the short range connections that are more broadly tuned (Bosking et al., 1997; Das and Gilbert, 1999; Nauhaus et al., 2008). This model also cannot account for any changes in the valence of surround effects with contrast.

Yet another circuit model proposes that the eCRF arises not locally in V1, but rather from fast-feedback loops from higher cortical areas (Schwabe et al., 2006). The motivation for this model is a set of experimental results concluding that horizontal, intra-areal connections are too slow to account for the dynamics of the far eCRF, and that feedback connections from higher cortical areas are fast enough to account for contextual-influence from the far surround (Angelucci et al., 2002; Angelucci and Bullier, 2003; Angelucci and Bressloff, 2006). This model proposes that the high contrast receptive field is determined by the spatial extent of the feed forward input, the low contrast summation field is determined by the spatial extent of the horizontal connections, and any surround modulation occurring at retinotopic distances further than the low contrast summation field must be carried by feedback input from higher cortical regions. It is not clear, however, whether feedback connections from higher cortical areas have sufficient functional specificity to account for V1 surround modulation (Stettler et al., 2002). Additionally, in this model, as in the model by Somers et al. (1998), the increase in the spatial extent of the summation field as well as the emergence of surround facilitation at low contrast are both caused by inhibitory neurons with higher contrast gain and threshold.

The last circuit model we consider, which forms the basis of most of the work presented
in this thesis, is the inhibition stabilized network (ISN) model (Ozeki et al., 2009). All of the previously presented circuit models operate under the intuitive assumption that during surround suppression, the amount of inhibition received by neurons should be increased compared to the non-surround suppressed state. By recording intracellularly and using voltage clamp to isolate the excitatory and inhibitory input conductances, Ozeki et al. (2009) found that during surround suppression, neurons actually received less inhibitory input (Figure 1.2). The ISN model was built to explain this paradoxical result. The novel mechanistic feature of this model is recurrent excitatory connectivity of sufficient strength to make the network unstable in the absence of dynamic feedback inhibition. With inhibition, though, the network is stable, and in response to increased excitatory input to inhibitory neurons, both the excitatory and inhibitory neurons in the network decrease their firing rate (Tsodyks et al., 1997). The model predicts a transient rise in inhibition immediately following surround stimulus onset, but at the steady state, a net decrease in the inhibitory input to the excitatory cells. The reason for this effect is that in the ISN model, input from the surround does not actually inhibit the cells, in the classical sense. Rather, by introducing a slight bias in the input towards the inhibitory neurons, it instead causes a de-amplification of both E and I responses (Murphy and Miller, 2009).

In the ISN model, the surround signal was considered to be simply an additional excitatory input directed towards the inhibitory neurons. In this thesis, we begin by expanding this general model and considering circuits in which center and surround are modeled self-consistently. As in V1, the eCRF of some cells is simply the CRF of other cells, and the network must self-generate the appropriate levels and spatial distributions of inputs. As we will show in the next chapter, this extension of the ISN model is able to reproduce and explain a number of eCRF properties. In the following chapter, we further extend this basic framework by introducing nonlinearity into our model. Here we are able to explore a much broader class of contextual modulations, and we find that the nonlinear ISN is able to recapitulate virtually all of the effects described above. Using a combination of numerical
and analytical modeling approaches, we can begin to understand the circuit mechanisms underlying contextual modulation in V1.

1.5 Normalization

Related to contextual modulation is the concept of normalization in cortical circuits. Proposed by some to be the “canonical operation” of cortex (Heeger et al., 1996; Reynolds and Heeger, 2009), normalization refers to a computation in which the responses of individual neurons are scaled by the total input into the network (Heeger, 1992) (Figure 1.3). This process is thought computationally to provide a way to keep neuronal firing rates within the most sensitive region of their dynamic regime, such that the most informative aspects of the population code, the relationships between individual units, remain precise. As a phenomenological model of sensory cortex, the normalization model has proven to be extraordinarily powerful and versatile. This class of models can account for the responses of single neurons to multiple preferred stimuli as well as pairs of orthogonal stimuli. The predictions of this model, originally tested on single cells, have also been shown to hold at the level of neural populations. Recently it has been shown that in V1, the response of the neuronal population to two simultaneously presented stimuli is very close to the average of the responses to the individual stimuli alone, whether measured with intrinsic optical imaging (MacEvoy et al., 2009), extracellular electrode array (Busse et al., 2009), or even fMRI (Brouwer and Heeger, 2011), and this has been cited as further evidence of normalization. Normalization models can also account for many of the neurophysiological effects of attention (Reynolds and Heeger, 2009).

Though a powerful phenomenological model of sensory cortex, mechanistically, there is still no consensus as to how cortical circuits actually perform this operation. Previously published models have proposed numerous different mechanisms by which the responses of individual neurons might be scaled by the total population activity. One of the earliest mechanistic models proposed feedback from a pool of “normalizing” inhibitory interneu-
1.5. Normalization

FIG. 1.3: **Schematic of the classic normalization model.** In the original normalization model, simple cell outputs ($S_i$) at multiple phases are pooled to generate one complex cell ($C_i$). The outputs from complex cells at all orientations and nearby spatial frequencies are pooled together to generate a feedback signal that divisively suppresses the simple cell responses (From Heeger (1992)).

rons (Heeger, 1993) to produce a divisive operation through shunting inhibition (Carandini et al., 1997). In this model, the inhibitory normalizing pool projects nonspecifically to all of the neurons in the local network in order to achieve normalization. However, it has been shown that in primary visual cortex, the inhibition received by cells actually closely matches the excitation received in orientation tuning (Ferster, 1986; Anderson et al., 2000), and there is little evidence for inhibitory cells with such nonspecific projections. Later models proposed different feedforward mechanisms to account for normalization, such as activity dependent depression in the thalamocortical synapses (Carandini et al., 2002). Experiments have shown that normalization is still observed when tested with flashed stimuli (as opposed to drifting gratings), and it has been argued that the duration of these stimuli is too brief to allow
synaptic depression to play a major role (MacEvoy et al., 2009). Input-dependent changes in the level of spontaneous background activity have also been shown to produce an effective gain modulation (Chance et al., 2002; Finn et al., 2007), which could account for some of the effects ascribed to normalization. In general, though, it is still an open question as to what mechanism or mechanisms produce normalization in cortex.

It is important here to clarify what might be a rather subtle point in this whole discussion. Though we speak of contextual modulation and normalization in separate sections, and refer to distinct bodies of literature in discussing them, we are by no means implying that the two are somehow distinct operations. They are rather just two different ways of looking at the response properties of V1 and trying to understand what it is doing. It’s quite simple to see how surround suppression can be viewed through the lens of normalization: the addition of more input to the network (i.e. a surround stimulus) causes a sublinear scaling of the response to the stimulus in CRF, precisely as the normalization model would predict. Historically, though, the normalization model has also been used to describe features of the V1 response not classically considered contextual modulation, such as contrast saturation or suppression by non-preferred stimuli within the CRF (Heeger, 1992). Ultimately, as we hope to show in this work, the circuit mechanism that we have uncovered is able to account for the results from both bodies of literature. Whether one wishes to think that this model “accomplishes contextual modulation and normalization” or rather “accomplishes contextual modulation through normalization” is perhaps a matter of personal taste. We would say neutrally that both contextual modulation and normalization can be accounted for by our model.

1.6 Overview

The goal of this work is to understand the circuit mechanisms underlying contextual modulation and normalization in visual cortex. In doing so, we will present a class of circuit models that we have developed, which share a common, unifying circuit motif. In their
ability to account for an unprecedented diversity of cortical response features, we feel that these models represent a parsimonious solution to many of the open questions in the fields of contextual modulation and normalization. In the next chapter, we will introduce our first extension to the original ISN model (Ozeki et al., 2009), a linear model of spatial interactions across retinotopic/cortical space. We study this model and use it to demonstrate how a network in the inhibition stabilized regime can self-consistently produce surround suppression of both excitatory and inhibitory neurons. In the subsequent chapter, we greatly enhance the basic linear model by introducing neuronal nonlinearity. In doing so, we produce not only a substantially more powerful and flexible model of contextual interactions, but also discover that our basic circuit motif is capable of producing all of the behaviors traditionally ascribed to models of normalization. In the following chapter, we discuss some experimental tests of our circuit model. In the final results chapter, we show one example of how the same circuit motif may underlie the behavior of a higher cortical region, the lateral intraparietal area. Though much of this work focuses specifically on primary visual cortex, there is little reason to believe that the circuit mechanisms underlying these effects should be unique to this one cortical region. Indeed, it is quite likely that other sensory cortices, including both higher visual areas and the cortical areas serving other sensory modalities, may rely on very similar circuit mechanisms, adapted over evolutionary timescales, to achieve computational goals selective to their particular functional needs.

Though not at all a focus of this work, it is interesting to consider, as an aside, why visual cortex has received so much more attention by neuroscientists over the past sixty years than other brain regions. Certainly, from an engineering standpoint, visual cortex has the advantage of responding to a highly controllable input space. The ease with which one can deliver a precise visual stimulus using only a CRT computer monitor, as compared to the elaborate equipment necessary for delivering an olfactory or tactile stimulus, makes for a much more easily controlled experiment. Technically, too, visual cortex benefits from an unusually accessible anatomic location. Wrapped around the posterior pole of the cortex, V1 is
1.6. Overview

Amenable to both electrical and optical recordings with relatively simple surgical techniques. Perhaps most importantly, though, is the unique way in which vision, moreso than any other sense, is intimately intertwined with our experiential percept of consciousness and awareness. Were we an advanced society of bats, or mice, or carrier pigeons, it is not unlikely that we would be debating similarly our extensive focus on auditory, olfactory, or magnetosensory cortices. But we are people, and so vision it is.
Chapter 2

A Linear Model of Contextual Modulation in Visual Cortex

2.1 Introduction and motivation

The activity of neurons in visual cortex is under the influence of a large area of retinotopic space. In primary visual cortex (V1), this area is conventionally divided into two distinct regions — the classical and extra-classical receptive fields — which are distinguished by their response to stimuli within. Stimuli in the classical receptive field (CRF), or receptive field center, can elicit a response directly from the neuron, whereas stimuli in the extra-classical receptive field (eCRF), or receptive field surround, modulate the neural response to stimuli in the center (Cavanaugh et al., 2002a;b; Sceniak et al., 1999; Kapadia et al., 1999; Ozeki et al., 2004; 2009; Akasaki et al., 2002; Angelucci and Bressloff, 2006). Thus, the extra-classical receptive field is believed to play its part not in stimulus detection, but rather in tailoring the response of a neuron to varying contexts. Numerous experiments have illustrated the potential importance of this property to the computational function of the early cortical visual system, in terms of both stimulus detection and selectivity (Rao and Ballard, 1999; Vinje and Gallant, 2000; 2002; Chen et al., 2005; Haider et al., 2010). The mechanistic basis
of the eCRF, however, is still poorly understood.

Recently, it has been shown that during surround suppression, the amount of both excitation and inhibition received by surround suppressed V1 neurons paradoxically decreases (Ozeki et al., 2009). Despite being suppressed, the neurons were found to receive less inhibition during center and surround stimulation than during center stimulation alone. If the addition of a surround stimulus causes an increase in the input (either excitatory or inhibitory) to the region of cortex representing the center stimulus, this paradoxical response can only occur in a network in which the excitatory recurrent connections are sufficiently strong to make the network unstable in the absence of dynamic feedback inhibition (Tsodyks et al., 1997; Ozeki et al., 2009). Such a network is called an inhibition-stabilized network (ISN).

In the original ISN model (Ozeki et al., 2009), only the region of cortex representing the center stimulus was modeled explicitly; the surround was modeled as an additional external input directed preferentially to inhibitory cells. In this study, we wished to explore the requirements for surround suppression of both excitatory (E) and inhibitory (I) neurons in a circuit in which center and surround are both modeled self-consistently. To do this, we extended the original ISN model to include a spatial dimension, such that each E-I pair now represents a small cluster of neurons sharing the same receptive field center. Center and surround regions in this model are thus defined relative to the placement of stimuli (Figure 2.1), such that a stimulus in the surround receptive field of one group of cells is located in the center receptive field of neighboring cells. Through this simple extension, we are able to reproduce the main finding of Ozeki et al. (2009), which is that surround suppression of both excitation and inhibition requires an ISN. Furthermore, we show mechanistically why this must be the case in a spatially-extended network. Lastly, we describe how the mechanism underlying surround suppression in this model, the selective amplification of spatially periodic patterns of activity, may explain other previously unrelated forms of contextual modulation. In particular, this mechanism explains the tuning of neurons for the envelope
2.2 Details of the linear model

We model a section of V1 as a linear network positioned along a single axis of cortical space. At each position, \( x \), along this spatial axis there is a pair of neuronal elements, one excitatory and one inhibitory. Each element is represented by its firing rate, which evolves over time.
according to the equations (Wilson and Cowan, 1972)

\[
\tau_E \frac{d}{dt} r_E(x) = -r_E(x) + W_{EE} \ast r_E(x) - W_{EI} r_I(x) + c_E h(x) \tag{2.1}
\]

\[
\tau_I \frac{d}{dt} r_I(x) = -r_I(x) + W_{IE} \ast r_E(x) - W_{II} \ast r_I(x) + c_I h(x) \tag{2.2}
\]

Here, \(r_E(x)\) and \(r_I(x)\) represent the firing rates of the excitatory and inhibitory elements at position \(x\); \(c_E\) and \(c_I\) are the magnitudes and \(h(x)\) the shape of the feedforward input; \(\tau_E\) and \(\tau_I\) are the E and I cell time constants; \(W_{YZ}(x - x')\) represents the synaptic strength from a cell of type \(Z\) (E or I) at position \(x'\) to a cell of type \(Y\) at position \(x\), and depends only on the distance between them; and \(*\) signifies spatial convolution (e.g. \(W_{EE} \ast r_E = \sum_{x'} W_{EE}(x - x') r_E(x')\), where the sum is over all other grid positions \(x'\)).

These can also be expressed as a single system of equations as

\[
T \frac{d}{dt} r = -r + Wr + ch \tag{2.3}
\]

In this formalism \(r\) is a \(2N\)-dimensional vector of firing rates. The first \(N\) elements of \(r\) represent the firing rates of the excitatory units and the second \(N\) elements represent the firing rates of the inhibitory units. Each unit in the network receives both excitatory and inhibitory intracortical synaptic input from the other neurons in the network. The strength of this input is determined by the \(2N \times 2N\) synaptic connection matrix \(W\). By convention, the columns of \(W\) correspond to the presynaptic units, the rows identify the postsynaptic units, and the elements of the matrix indicate the strength of the synaptic connection between them. \(W\) can be expressed as a \(2 \times 2\) block matrix composed of four \(N \times N\) submatrices, each of which describes one of the four types of synaptic connections: E to E, I to E, E to I, and I to I. In these terms, \(W = \begin{pmatrix} W_{EE} & -W_{EI} \\ W_{IE} & -W_{II} \end{pmatrix}\), and if grid points are \(x_i\), with \(i = 1 \rightarrow N\), then \((W_{YZ})_{ij} = W_{YZ}(x_i - x_j)\).

Each neuron also receives feedforward input from outside of the network. The relative
2.2. Details of the linear model

Amount of feedforward input received by each neuron is given by the $2N$-dimensional vector $\mathbf{h}$. Again, the first $N$ elements represent the feedforward input to the excitatory units and the second $N$ elements represent the feedforward input to the inhibitory units. The shape of the input, $\mathbf{h}$, is scaled by two magnitude parameters, $c_E$ and $c_I$, which may or may not be equal. The scaled input is abbreviated in equation 2.3 as $\mathbf{c}\mathbf{h}$. $\mathbf{T}$ is a $2N \times 2N$ diagonal matrix of membrane time constants; the first $N$ elements of the diagonal equal $\tau_E$ and the second $N$ elements equal $\tau_I$. In equation 2.3, the firing-rate response is modeled as a linear function of the total feedforward and intracortical synaptic input; cells are neither thresholded nor do they saturate. Negative firing rates are allowed, and simply represent a level of firing below the baseline firing rate.

These neuronal elements are not meant to simulate the response of any one neuron, but rather represent the composite response of a small, homogenous group of neurons with receptive fields in the same region of retinotopic space. It is for this reason that we can at the present simply consider the firing rate response as representative of a local population average. In Ozeki et al. (2009), which employed a very similar model of the V1 cortical network, the validity of this assumption was tested explicitly. Here, the authors replaced a single pair of E-I elements with a multi-neuron representation, which used 100 excitatory and 100 inhibitory neurons, with connectivity varying probabilistically between the cells. They showed that this larger model reproduces the mean responses of the population firing-rate averages, albeit with more biologically realistic variability. Thus, for the phenomena of interest, our current level of modeling detail is appropriate.

Synaptic connection strength is a function of position and cell type (E or I). Excitatory elements make synaptic connections onto other populations in the network through long-range horizontal projections, the strength of which decay monotonically as a function of distance and depend on the cell type of the post-synaptic target:
2.2. Details of the linear model

\[ W_{EE}(x - x') = J_{EE} e^{-\frac{(x-x')^2}{2\sigma_{EE}^2}} \]

\[ W_{IE}(x - x') = J_{IE} e^{-\frac{(x-x')^2}{2\sigma_{IE}^2}} \]  

(2.4)

\[ J_{EE}, J_{IE}, \sigma_{IE}, \text{ and } \sigma_{IE} \] are parameters of the model. Synaptic connections originating from inhibitory populations are local, inhibiting only the cells at the same retinotopic position (Miller, 2003):

\[ W_{EI}(x - x') = J_{EI} \delta(x, x') \]

\[ W_{II}(x - x') = J_{II} \delta(x, x') \]  

(2.5)

\[ \delta(x, x') = \begin{cases} 1 & \text{if } x = x' \\ 0 & \text{if } x \neq x' \end{cases} \]

As above, \( J_{EI} \) and \( J_{II} \) are parameters of the model. Because the neuronal elements do not represent individual neurons, but rather small populations of neurons, local synaptic connections are allowed in both the E and I populations. All of the neuronal elements of a given type (E or I) are assumed to have the same membrane time constant which may or may not be the same in the two different cell types (Kelly and Van Essen, 1974).

Experiments using intracellular recording techniques often report the synaptic excitation and inhibition received by cells (Anderson et al., 2000; 2001). In terms of our model, we quantify the total excitation received by any element in the network as the sum of its feedforward input and its intracortical excitatory input, which is the sum of the firing rate of each excitatory element in the network scaled by the strength of its synapse onto the element in question. The total inhibition received by a cell is the firing rate of its local inhibitory neuronal element scaled by the strength of its synapse.
Feedforward input is represented in terms of the receptive fields of the cortical elements to which it projects. In spatial extent, the CRF of an individual neuron in V1 is well approximated by a Gaussian function of retinotopic space (Li and Li, 1994; Song and Li, 2008), that is, a small point stimulus located closer to the center of a neuron’s CRF will elicit a greater response than one located closer to the periphery. The stimuli we are interested in studying, however, are not points, but rather spatially extensive stimuli of various sizes. V1 neurons summate input spatially over their CRF (Henry et al., 1978), therefore we model the total input to any neuronal element in response to a stimulus of some length by integrating the CRF over the length of the stimulus. The integral of a Gaussian curve has a sigmoid shape (given by the error function), so we expect that as a stimulus is lengthened, the input to any one neuron in the network should rise and plateau sigmoidally. The midpoint of the rising phase will depend on the neuron’s position in the network, and the slope will depend on the width of the Gaussian curve describing the CRF. Assuming that the CRFs of all of the neurons have the same width, are evenly spaced, and that the CRFs of a given neuronal type (E or I) have the same maximum amplitude, the total feedforward input to any element in the network at a position $x$ can be phenomenologically described by a smoothed step function whose width is determined by the stimulus length, $l$, and whose sharpness is determined by the parameter $\sigma_{RF}$:

$$h(x) = \left( \frac{1}{1 + e^{-\frac{x + l/2}{\sigma_{RF}}}} \right) \left( 1 - \frac{1}{1 + e^{-\frac{x - l/2}{\sigma_{RF}}}} \right)$$  \hspace{1cm} (2.6)$$

We use this step-function input for the majority of simulations. To better understand the role of the input shape on the network response properties, we also run some simulations using a Gaussian input profile (which, in terms of our model, would represent a Gabor patch input):

$$h(x) = e^{-\frac{x^2}{2\sigma_t^2}}$$  \hspace{1cm} (2.7)$$

where $\sigma_t$, which determines the width of the Gaussian, is defined as: $\sigma_t = \frac{l}{2\sqrt{2\ln(2)}}$ such that the stimulus length $l$ is the full-width at half-height of the Gaussian curve. Given the
spatial blurring that may occur as stimuli are transmitted through the retina and LGN, as well as nonuniformities in the spatial structure of the receptive fields of neurons in these two regions (Cai et al., 1997; Reid and Shapley, 1992; Field et al., 2010), it may be most appropriate to consider inputs comprised of a weighted mixture of these two functions. To produce mixtures of these two inputs, we added the two stimuli with positive coefficients that sum to one (i.e. \( h(x) = c_1 \times \text{StepFunction} + c_2 \times \text{Gaussian} \), where \( c_1 + c_2 = 1 \)). For the plots in Figures 2.2 and 2.4 we used \( c_1 = 0, 0.2, 0.4, 0.6, 0.8, 1.0 \).

In this model, the stimulus presented to the neurons is assumed to be optimal for the CRF (i.e. in orientation, temporal frequency, spatial frequency, and direction). For all simulations, the stimulus is centered at the position \( x = 0 \). The relative magnitude of the feedforward inputs to the E and I cells are given by \( c_E \) and \( c_I \). The “sharpness” of the stimulus profile is given by \( \sigma_{RF} \), which is the same for both cell types (Anderson et al., 2000). As is the case biologically, feedforward input to V1 is always excitatory (Reid and Alonso, 1995).

### 2.3 Simulations and parameters

The distance between neighboring E-I pairs is \( \Delta x \). The stimulus length, \( l \), is specified in each simulation. Except where otherwise noted, simulations were run using a single set of model parameters. For the sake of generalizability, \( \sigma_{EE} \) in the model has been set equal to 1, and all other parameters of distance (\( \sigma_{IE}, \sigma_{RF}, \) and \( \Delta x \)) are in units normalized to this value. The parameters defining the feedforward input are: \( c_E = 1.2, c_I = 1.0 \), and (for the step function input defined in equation 2.6) \( \sigma_{RF} = 0.33 \). The parameters defining the cortical network are: \( N = 401, \Delta x = 0.5, \tau_E = 20 \text{ ms}, \tau_I = 10 \text{ ms}, J_{EE} = 0.385, J_{IE} = 1.0, J_{EI} = 0.55, J_{II} = 1.5, \sigma_{EE} = 1.0, \) and \( \sigma_{IE} = 2.0 \). The stimulus is always centered at \( x = 0 \). In the figures, stimulus lengths are divided by a cortical magnification factor to allow us to plot results in terms of retinotopic degrees (to match electrophysiological studies.) Here we use a magnification factor of 2.0 mm/deg, which corresponds to approximately 5 degrees
2.3. Simulations and parameters

eccentricity in macaque or 1-3 degrees eccentricity in cat (Van Essen et al., 1984; Albus, 1975). Cells receive a spatially uniform baseline input of strength 5.

These parameters were chosen to fulfill three main constraints. First, we wanted to ensure the model was operating in the ISN regime. This means that in the absence of dynamic inhibition, the recurrent excitatory connectivity is strong enough to make the network unstable, but the network is stable in the presence of inhibition. Mathematically, this occurs when at least one eigenvalue of $W_{EE}$ is greater than 1, but all of the eigenvalues of $W$ are less than 1 (Ozeki et al., 2009). In our system, the maximum eigenvalue of the matrix $W_{EE}$ is very closely approximated by $J_{EE} \sigma_{EE} \sqrt{2\pi}/\Delta x$. For our chosen parameters, $(0.385) \ast (1) \ast (\sqrt{2\pi})/(0.5) > 1$, so the network is an ISN. Second, as we determined in the course of our analysis, the E to I connections must extend further in space than the E to E connections (in model parameters, $\sigma_{IE} > \sigma_{EE}$). Third, we wanted the spatial scale of connectivity to roughly match the known anatomy of the system. With these constraints in place, we were still left with a large range within which to define the model’s parameters. The single set of parameters finally chosen represents just one point within a robust behavioral space that is relatively insensitive to changes in individual parameters. Fine tuning is not needed to produce the results presented in the main text.

For simulations in which only the steady-state response of the network is of interest (as in Figure 2.2), we can take advantage of the analytic solution to the linear system and calculate the steady-states directly. For instances in which the dynamics of the response is interesting (as in Figure 2.6), we simulate the network using a standard forward-Euler algorithm with a fixed 1 ms time step (which was also used many times to confirm that the analytic solutions reached the same steady state. In addition, all simulations were also tested with shorter time-steps to see if it made a difference in the steady-states or dynamics, which it did not). All simulations were run in Matlab (Mathworks Inc, Natick, MA).
2.3. Simulations and parameters

FIG. 2.2: The firing rate response of the model to stimuli of increasing length. The top row simply demonstrates gratings of increasing size. The second row shows a series of plots of the input to the network. The input is either a smoothed step function (in black), a Gaussian curve (in green), or some mixture of the two. There are a total of six curves on each plot: one for the pure step function, one for the pure Gaussian, and four mixtures. The third and fourth rows show the excitatory and inhibitory firing rates across the network. For excitatory firing rates, the curves transition from red (purely step-function input) to yellow (Gaussian input). For inhibitory cells, the transition is from blue for step-function input to cyan for Gaussian input. These plots show that large stimuli induce spatially-periodic patterns of activity in the cortical network. On the bottom are firing rate length-tuning curves of the cells at the center of the network showing strong surround suppression for both E and I, as well as a second-peak for long stimuli. Color scheme is the same as above.
2.4 Results

As a first test of surround suppression, we ran a series of simulations to find the steady-state responses to stimuli of fixed contrast and increasing diameter. Surprisingly, we found that for large stimuli, the network responded to these smoothed-step function inputs by producing standing-waves of activity over cortical space (Figure 2.2). These waves increased in spatial extent with stimulus width, and importantly, the edge of the stimulus representation in cortical space was always tied to a peak of the standing wave. Intuitively, these peaked edges occur because these regions of cortex receive less surround suppression than the regions representing the interior of the stimulus. An analogous result has been observed in psychophysical studies of contrast threshold in the presence of multiple flanking stimuli (Adini et al., 1997). Interestingly, different stimulus sizes induce the largest amplitude oscillations in the excitatory (fourth from the left in row three of Figure 2.2) and inhibitory (fifth from the left in row four of Figure 2.2) populations. As we will show below, this occurs because the excitatory and inhibitory subnetworks actually have different resonant spatial frequencies.

By plotting curves of the firing rate response (or conductance, Figure 2.4) versus stimulus length for the cells located at the center of the network, we can see how these standing waves produce surround suppression of both the excitatory and inhibitory populations. Initially, there is a rise in response magnitude with increasing stimulus size that ultimately peaks when the stimulus is at the preferred length for the cells. As stimulus size is increased past this preferred value, the responses decrease, and the cells show surround suppression. For stimuli larger still, there is a second rise in response magnitude. This model predicts that the length-tuning curves produced in response to stimuli with sharply-defined contrast edges should actually contain multiple peaks. This is precisely what was observed by Anderson et al. (2001) for inhibitory (and to a lesser extent excitatory) conductances. These authors proposed the existence of some secondary inhibitory input that was only activated by very large stimuli, but here we show that this second peak is in fact produced by the same
mechanism that produces the first peak and surround suppression, that is, a spatially periodic pattern of activity over cortical space.

For stimuli with more smoothly defined contrast profiles, length-tuning curves produced by the model may have only a single peak (Figure 2.2, yellow and cyan curves). This occurs because with more smoothly decreasing input contrast profiles, there are no clearly defined "edges" to the stimulus representation in the cortical network that receive less surround suppression than their neighbors, and thus no standing waves over cortical space. We therefore predict that the presence of multi-peaked length-tuning curves in experiments should depend strongly on the contrast profile of the stimuli used. Specifically, multi-peaked length tuning curves should be most visible in experiments using sharply defined, rather than Gabor, stimuli. We explain mathematically why this should be true in the analysis below.

If, as we hypothesize, our network produces surround suppression because of an intrinsic resonant spatial frequency, then we should be able to test for this resonance directly by probing the network with spatially-periodic stimuli. To do so, we first calculated the power spectra of the population responses of both the excitatory and inhibitory neurons during stimulation with stimuli of various sizes, and plotted these spectra versus spatial frequency (Figure 2.3 top, colored dashed lines). Next, we probed our network with sinusoidally contrast-modulated (CM) grating stimuli over a range of CM spatial frequencies, and plotted the maximum firing rate evoked in both the excitatory and inhibitory populations as a function of CM spatial frequency on the same set of axes (Figure 2.3 top, solid lines). We found that the maximum response of the cells in the network was evoked by a CM grating stimulus modulated at the same spatial frequency as the network oscillation; that is, the network resonates at its preferred spatial frequency. Furthermore, both of these peak frequencies matched exactly the analytical resonant frequencies (Figure 2.3 top, black dashed lines), whose derivation is described in the analysis below.
2.4. Results

An ISN with spatially-periodic E and I firing rates resonates at its preferred spatial frequency. A., B. Left y-axes: driving the network with a full-screen sinusoidally contrast modulated stimuli reveals strong CM spatial frequency tuning for both the excitatory (solid red) and inhibitory (solid blue) cells. The dashed vertical black lines are the analytically calculated resonant frequencies. Right y-axes: plotted in dashed colors are the normalized power spectra of the population response to small and large spatial stimuli (from the length-tuning simulations). The mean has been subtracted from the population response prior to Fourier analysis. The small stimulus for both the E and I figures is 0.5° long; the large stimulus is 4.5° for the E cells and 5.25° for the I cells. C., D. The resonant frequencies of the network depend most strongly on the spatial extent of synaptic connectivity, determined by the parameters $\sigma_{IE}$ and $\sigma_{EE}$. The preferred CM spatial frequencies from simulations as a function of the ratio of $\sigma_{IE}/\sigma_{EE}$ are shown in C; the analytically derived resonant frequencies are in D. Where the resonant frequency of I cells is greater than that of E cells, the network is unstable (shaded in red).
2.4. Results

FIG. 2.4: The conductances changes in response to stimuli of increasing length. This figure is analogous to Figure 2.2, showing the change in excitation and inhibition over both network position for all cells and stimulus length for the center cells. However, here we are plotting the excitatory and inhibitory conductances received by the excitatory cells, rather than the firing rate. Note the decrease in the relative amplitude of the periodic modulation in the excitatory conductance compared to the excitatory firing rates in Figure 2.2, which is a result of the Gaussian shape of the excitatory spatial connectivity functions.

Interestingly, an analogous spatial resonance was observed experimentally by Tanaka and Ozhawa (2009), who used large, contrast-modulated gratings to probe the structure of center-surround complexes of cells in V1. They found in their study that V1 cells do in fact have a preferred CM spatial frequency, and this frequency is 1-6 times lower than the preferred luminance spatial frequency (mean ± sd: 2.1 ± 0.9), reflecting the product of an interaction occurring over a spatial scale that extends beyond the CRF of individual neurons. Our model produces the same result, and offers a novel mechanistic explanation for this as-of-yet unexplained experimental observation: spatially-periodic patterns of activity. In the model, the optimal CM spatial frequency corresponds to the frequency of the spatial
activity, that is, the optimal stimulus drives the peaks but not the troughs of activity. A neuron’s summation field should fill about 1/2 a cycle of the spatial period, since a larger size would drive surround suppression. The summation field in V1 most commonly contains 1-3 cycles of the CRF preferred luminance spatial frequency (De Valois et al., 1985), suggesting a CM spatial frequency 2-6 times larger than the CRF preferred luminance spatial frequency, as observed experimentally (Tanaka and Ohzawa, 2009).

2.5 Analysis

In this section, we describe in detail an analysis of this network. Because this is a linear network, we are able to easily obtain an analytic solution for the fixed-point behavior of the model. By carefully choosing our basis set, we can use this solution to understand the mechanisms and requirements underlying the network behavior described above. We begin by briefly recapitulating the main finding from Ozeki et al. (2009), and then quickly move into the novel results and predictions from this extension of that work.

2.5.1 General features of an ISN

Ozeki et al. (2009) showed that in a network in the ISN regime, increased excitatory input to inhibitory cells will cause a paradoxical decrease in inhibitory firing rates. This can be seen by studying a simple two-population model, comprised of a single pair of recurrently connected excitatory and inhibitory neurons.

\[
\tau_E \frac{d}{dt} r_E = -r_E + J_{EE} r_E - J_{EI} r_I + c_E \tag{2.8}
\]

\[
\tau_I \frac{d}{dt} r_I = -r_I + J_{IE} r_E - J_{II} r_I + c_I \tag{2.9}
\]
Inhibition-stabilized networks respond paradoxically to increased input to I cells. This figure demonstrates this effect in a simple, 2-population model (1 E cell and 1 I cell). The top row shows the time course of the firing rate of the I cell in response to a small excitatory input (and no input to the E cell). The bottom row shows the changes in the steady state firing rate for both the E and I cell in response to inputs of increasing strength, again only to the I cell. The first column has parameters that place it in the ISN regime. The second column starts with the same parameter set, but $J_{EE}$ is scaled back by a factor of 4, so the network is not in the ISN regime. In the third column, all four synaptic weight parameters are scaled back by a factor 4. Observe in both rows that only in the ISN do the steady-state inhibitory firing rates decrease in response to increased excitatory input to the I cell.
The fixed-point firing rates are:

\[ r_E = \frac{1}{\det(1-W)} \left( (1 + J_{II}) c_E - J_{EI} c_I \right) \]

\[ r_I = \frac{1}{\det(1-W)} \left( (1 - J_{EE}) c_I + J_{IE} c_E \right) \]

In this simple 2-neuron example, \( W = \begin{pmatrix} J_{EE} & -J_{EI} \\ J_{IE} & -J_{II} \end{pmatrix} \). Since stability requires that the determinant of \( W \) be positive, from equation 2.11 it is clear that in the absence of any change in the input to the excitatory cells (i.e. \( c_E = 0 \)), an increase in excitatory input to inhibitory cells (\( c_I > 0 \)) will cause a decrease in the inhibitory firing rates if and only if \( J_{EE} > 1 \). In this simple two-population model, \( J_{EE} > 1 \) is precisely the requirement for the network to be in the ISN regime. This basic response property is illustrated in a simple 2-population linear model in Figure 2.5. For this toy model, we use the following parameters: \( \tau_E = 20, \tau_I = 10, J_{EE} = 1.5, J_{IE} = 1.5 J_{EI} = 2.5, \) and \( J_{II} = 2.0 \).

For the spatially-extended model, the fixed-point firing rates can be solved in an equivalent fashion.

\[ \mathbf{r} = (1 - \mathbf{W})^{-1} \mathbf{c} \]

The solution in this form is not particularly useful for gaining an intuition of the network’s behavior, because the firing rate at any one position \( x \) depends on the inputs and rates at all other positions. To decouple these equations, we can take advantage of the translational-invariance of the network connectivity functions. We say that connections are translationally invariant because the strength of the synaptic connection between any two neurons of a given type (E or I) depends only on their relative distance in the network, and not their absolute position. Translationally invariant matrices have the convenient property of being diagonalized by the Fourier transform, and so to take advantage of this property of
our network, we change our basis of analysis from one of position \((x)\) to spatial frequency \((k)\):

\[
\tilde{r}_E(k) = \frac{1}{\det(1 - \tilde{W}(k))} \left( (1 + \tilde{w}_{II}(k)) \tilde{h}_E(k) - \tilde{w}_{EI}(k) \tilde{h}_I(k) \right) \tag{2.13}
\]

\[
\tilde{r}_I(k) = \frac{1}{\det(1 - \tilde{W}(k))} \left( (1 - \tilde{w}_{EE}(k)) \tilde{h}_I(k) + \tilde{w}_{IE}(k) \tilde{h}_E(k) \right) \tag{2.14}
\]

where \(\tilde{r}, \tilde{w},\) and \(\tilde{h}\) refer to the spatial Fourier transforms of the firing rates, connectivity functions, and input, respectively.

In the frequency domain, each E-I pair at a given spatial frequency \((k)\) evolves independently of all other frequencies. So, by moving to a Fourier basis, we have essentially replaced our fully coupled system by an equivalent collection of \(N\) unconnected E-I pairs, each of which represents the activity of excitation and inhibition at a different spatial frequency. We can make this transformation because sinusoids of increasing frequency are an eigenbasis of the connection submatrices in this network. For this same reason, we can say that E and I sinusoids of frequency \(k\) will behave as an ISN when \(\tilde{w}_{EE}(k) > 1\). We will use this representation in presenting the rest of the analysis of the linear model. Additionally, for the sake of the analysis, we transition here from considering a discrete grid of neurons to considering a continuum of neurons along a continuous position \(x\) (and thus continuous in spatial frequency \(k\)).

### 2.5.2 Resonance, critical frequencies, and stability

#### 2.5.2.1 Inhibitory resonant frequency

One of the conclusions presented in the results section is that surround suppression of both excitatory and inhibitory firing rates in an ISN occurs through the selective amplification of spatially periodic patterns of activity. Furthermore, we can show analytically that with local inhibitory connectivity, only an ISN can generate spatially-periodic patterns of activity in the...
inhibitory subpopulation of neurons. To understand why this is so, consider the fixed-point solution of the inhibitory firing rates from equation 2.14. Letting \( h_E = \alpha \tilde{h}_I \) and expanding the determinant term to show its dependence on the individual synaptic weight functions, the inhibitory firing rates can be expressed with the following equation:

\[
\tilde{r}_I(k) = \left( \frac{(1 - \tilde{w}_{EE}(k) + \alpha \tilde{w}_{IE}(k))}{(1 + \tilde{w}_{II}(k)) (1 - \tilde{w}_{EE}(k)) + \tilde{w}_{EI}(k) \tilde{w}_{IE}(k)} \right) \tilde{h}_I(k) \tag{2.15}
\]

The large fractional term within the parentheses acts as a linear filter on the input, \( \tilde{h} \), and represents how the intracortical connectivity modulates the feedforward input to the network. In moving to the Fourier basis, firing rates and synaptic weights have been expressed in terms of spatial frequency, \( k \). Maxima at \( k \neq 0 \) in the network filter within the bandwidth of \( \tilde{h}_I(k) \) introduce non-DC peaks in the power-spectrum of the firing rates, which translate in linear space into activity with a defined spatial period. In other words, non-DC peaks in this linear filter allow the network to transform a non-periodic input into an output with spatial oscillations, which will have a frequency very close to the location of the maximum of this cortical-connectivity filter. By solving for these maxima we can find the conditions under which the network will demonstrate spatial periodicity. To find these points, we solve for the roots of the first derivative of the network filter with respect to \( k \). The denominator of this derivative is \( \det(1 - \tilde{W}(k))^2 \), and since stability requires that this determinant always be positive, this derivative is defined and continuous for all \( k \). The roots are given by zeros in the numerator, which equals:

\[
(\tilde{w}_{EI}(k) - \alpha (1 + \tilde{w}_{II}(k)))(\partial \tilde{w}_{IE}(k) (\tilde{w}_{EE}(k) - 1) - \partial \tilde{w}_{EE}(k) \tilde{w}_{IE}(k)) \tag{2.16}
\]

with all derivatives taken with respect to spatial frequency, \( k \). Because the inhibitory connectivity is local (i.e. \( W_{EI}(x) \) and \( W_{II}(x) \) are \( \delta \) functions in the spatial domain (eq. 2.5)), \( \tilde{w}_{EI}(k) \), and \( \tilde{w}_{II}(k) \) are constants in the Fourier domain, and \( \alpha \) is a constant as well. Because \( W_{EE}(x) \) and \( W_{IE}(x) \) are even functions of \( x \), both \( \tilde{w}_{EE}(k) \) and \( \tilde{w}_{IE}(k) \) are even functions of
2.5. Analysis

$k$, so $\partial \tilde{w}_{EE}(0)/\partial k$ and $\partial \tilde{w}_{IE}(0)/\partial k$ both equal 0. Assuming $\tilde{w}_{EI}(k) \neq \alpha (1 + \tilde{w}_{II}(k))$, the only zeros\(^1\) will occur at $k = 0$ and where:

$$\tilde{w}_{EE}(k) = \frac{\partial \tilde{w}_{EE}(k) \tilde{w}_{IE}(k)}{\partial \tilde{w}_{IE}(k)} + 1 \quad \text{(2.17)}$$

Because the functions describing the spatial extent of synaptic connections in the linear domain, $W_{EE}(x)$ and $W_{IE}(x)$, do not themselves contain any intrinsic periodic behavior, then the derivatives of the Fourier representation of these same functions, $\partial \tilde{w}_{EE}(k)$ and $\partial \tilde{w}_{IE}(k)$, will always have the same sign. Because $\tilde{w}_{IE}(k)$ is, by definition, a positive number (the Fourier transform of a Gaussian is a Gaussian), the fraction on the right hand side of equation 2.17 will always be positive, and equation 2.17 will only have a real solution, and the network filter a non-zero peak, when $\tilde{w}_{EE}(k) > 1$. Thus the network must be an ISN — $\tilde{w}_{EE}(k)$ must be greater than 1 for some value of $k$ — in order to have spatially periodic inhibitory firing rates. However, unlike the case of the paradoxical response illustrated above, being within the ISN regime is not necessarily sufficient to produce this behavior. Depending on the form and magnitude of all the spatially-extensive synaptic weight functions, the strength of recurrent excitation necessary to produce periodicity may be substantially larger than the strength required to put the network into the ISN regime.

With our particular choice of synaptic connectivity functions (equations 2.4 and 2.5), equation 2.17 can be used to predict the spatial frequency at which inhibitory firing rates will resonate. We call this frequency the resonant frequency of inhibition, $k_{rI}$.

$$k_{rI} = 2 \ln \left( J_{EE} \sigma_{EE} \sqrt{2 \pi \left( \frac{\sigma_{IE}^2 - \sigma_{EE}^2}{\sigma_{IE}^2} \right)} \right)$$

The roots given by equation 2.18 exist when the argument of the logarithm is both positive and greater than 1 in magnitude. Since the first of these two requirements tells us

\[^1\text{A sanity check: if } \tilde{w}_{EI}(k) = \alpha (1 + \tilde{w}_{II}(k)), \text{ then the derivative of the inhibitory firing rates is predicted to be zero at all frequencies. Inspecting equation 2.13 reveals that this exact same condition will result in zero excitatory firing rates at all frequencies.}\]
that \( \sigma_{IE} > \sigma_{EE} \), we know \( \frac{\sigma_{IE}^2 - \sigma_{EE}^2}{\sigma_{IE}^2} < 1 \), and thus the second requirement will be met when:

\[
J_{EE} \sigma_{EE} \sqrt{2 \pi} > \left( \frac{\sigma_{IE}^2}{\sigma_{IE}^2 - \sigma_{EE}^2} \right)
\]

which tells us that at a very minimum:

\[
J_{EE} \sigma_{EE} \sqrt{2 \pi} > 1
\]

The term on the left of inequality 2.20 is the maximum eigenvalue of \( W_{EE} \); and so this inequality is equivalent to requiring that the network be an ISN. Thus, by solving for the preferred spatial frequency of the inhibitory populations, we can see again that only a network in the inhibition-stabilized regime is able to preferentially amplify non-zero (non-DC) spatial frequencies in the inhibitory population.

### 2.5.2.2 Excitatory resonant frequency

In the length-tuning simulations, in addition to spatially periodic inhibition, we also observe spatially periodic excitation. Interestingly, the second peaks in inhibition and excitation occurred at slightly different stimulus lengths. To understand why this occurs, we perform an analogous analysis to the one presented above to find the resonant frequency of excitation, \( k_{rE} \).

\[
k_{rE} = \sqrt{\frac{2 \ln \left( \frac{J_{EI} J_{EE} \sigma_{IE}^3}{J_{EE} (1 + J_{II}) \sigma_{EE}^2} \right)}{(\sigma_{IE}^2 - \sigma_{EE}^2)}}
\]

Unlike inhibition, the existence of a real \( k_{rE} \) does not require the network to operate within the ISN regime. However, assuming that the network is already in a regime in which the inhibitory elements have a real resonant frequency (and consequently, the network is an ISN), then we can show that \( k_{rE} \) must be real and greater than 0 as well. As shown above, a real value for \( k_{rI} \) requires that \( \sigma_{IE} > \sigma_{EE} \). When this is true, a real value for \( k_{rE} \) from
equation 2.21 requires that the argument of the logarithm be greater than 1, thus:

\[ J_{EI}J_{IE} \sigma_{IE}^3 > J_{EE} (1 + J_{II}) \sigma_{EE}^3 \]  

(2.22)

Because this network is linear, we know that it will have stable fixed-points when all of the eigenvalues of \((1 - \tilde{\mathbf{W}}(k))\) have a negative real part, and for this to occur, the determinant of \((1 - \tilde{\mathbf{W}}(k))\) must be positive for all values of \(k\). By solving for the minima of the determinant over \(k\) and imposing the condition that these minima have a value greater than zero, we can get the following requirement for stability:

\[
\left( \frac{J_{EI}J_{IE} \sigma_{IE}^3}{J_{EE} (1 + J_{II}) \sigma_{EE}^3} \right) \left( \frac{\sigma_{EE}^2}{\sigma_{IE}^2 - \sigma_{EE}^2} \right) > J_{EE} \sigma_{EE} \sqrt{2\pi} \left( \frac{\sigma_{IE}^2 - \sigma_{EE}^2}{\sigma_{IE}^2} \right) 
\]

(2.23)

Since we have stipulated that the network is already in a regime in which the inhibitory population has a real resonant frequency, we have satisfied inequality 2.19. Therefore we know that:

\[ J_{EE} \sigma_{EE} \sqrt{2\pi} \left( \frac{\sigma_{IE}^2 - \sigma_{EE}^2}{\sigma_{IE}^2} \right) > 1 \]  

(2.24)

Since resonant inhibition also requires that \(\sigma_{IE} > \sigma_{EE}\), the fractional exponent must be positive, and so a necessary condition for stability is that:

\[ J_{EI}J_{IE} \sigma_{IE}^3 > J_{EE} (1 + J_{II}) \sigma_{EE}^3 \]  

(2.25)

This, of course, is the exact same condition that guarantees that the excitatory population has a real resonant frequency. Thus, inequalities 2.22 and 2.25 tell us that in a network in which the inhibitory elements have spatially periodic activity, stability obliges the excitatory activity be spatially periodic as well.

From equations 2.21 and 2.18, we can see that the resonant spatial frequencies depend most strongly on the spatial extent of the long-range synaptic connectivity. To go back and test this in our model, we varied the spatial extent of E to I connections \(\sigma_{IE}\) while keeping
all other parameters constant, and found the preferred spatial frequencies for both the E and I populations (the locations of the peaks of the solid curves in the top row of Figure 2.3). The simulations (Figure 2.3 bottom left) are an essentially perfect match for the analytically derived resonant frequencies (Figure 2.3 bottom right). The analytic solutions reveal that for $\sigma_{IE}$ sufficiently wide, the network will actually be unstable (red area on the plot). This is perhaps a counter-intuitive result (i.e. destabilization from too much inhibition), and occurs precisely where the resonant frequencies of inhibition becomes greater than that of excitation.

This analysis also reveals why stimuli with edged contrast profiles produce multi-peaked length-tuning curves but those with Gaussian contrast profiles usually do not. In the Fourier domain, a step-function is transformed into a sinc function ($\frac{\sin(kl)}{k}$, where $l$ is the width of the step function), and in general edged stimulus profiles will be characterized by varying degrees of “ringing” in their Fourier transforms. This ringing will induce sign changes in the Fourier representation whose positions depend on stimulus length, resulting in a periodic flip in the phase of the resonant frequency with increasing stimulus size. These phase changes give rise to the periodic rise and fall of the length-tuning curves. Gaussian stimuli, on the other hand, are simply transformed into Gaussians curves over frequency with inversely proportional width (narrow stimuli will have broad Fourier representations, and vice-versa). For sufficiently wide Gaussian stimuli, the power spectrum goes to zero at a frequency below that of the peaks in the network response filters ($k_{r\,E}$ and $k_{r\,I}$). Thus the network’s resonant frequencies receive no input, and the firing rates show no spatial periodicity. For narrower Gaussian inputs, the Fourier representation may be wide enough to have non-zero power at the network’s preferred frequencies (and in response the network does have spatially-periodic activity – see the yellow curves in the leftmost plots of Figure 2.2), but in our model this “sufficiently narrow” range ends before the stimulus length that would produce the second peak in the length-tuning curves.
FIG. 2.6: The ISN model predicts a 180° phase shift in the relative responses of E and I to direct input to inhibitory cells around a critical spatial frequency. We stimulate only the inhibitory cells in the network with a photostimulus with a defined spatial frequency. The stimulus drifts at 2 Hz. The low spatial frequency stimulus on the left has a spatial frequency of 0.2 cycles/degree. The higher spatial frequency stimulus on the right is at 0.5 cycles/degree. In the top right a summary plot shows the relative phases of excitatory and inhibitory firing rates as a function of photostimulus spatial frequency. The vertical dashed line is the analytically calculated critical frequency.
2.5. Analysis

As the length $l$ of a step-function stimulus is increased, there can be stimulus lengths where $\frac{\sin(kl)}{k} = 0$ for one of the peak frequencies ($k_{rE}$ and $k_{ri}$). This will be observed across the population as a relative weakening in the amplitude of the spatial oscillation (though not a complete absence of periodicity, as spatial frequencies near the peak are also relatively amplified).

2.5.2.3 Critical frequency

Previous theoretical work on the ISN regime has focused on the paradoxical decrease in inhibitory firing rates in response to increased excitatory input (Tsodyks et al., 1997; Ozeki et al., 2009). With the addition of a spatial dimension in our current model, this paradoxical response will depend on the stimulus spatial frequency. The fixed-point firing-rate solutions in equations 2.13 and 2.14 can be used to understand this effect. As these fixed-point solutions are written in terms of the eigenvectors of the connection submatrices, the requirement for being an ISN is simply that $\tilde{w}_{EE}(k) > 1$ for some value of $k$. When feedforward input is biased completely to inhibitory populations, such that $\tilde{h}_E(k) = 0$, the fixed-point firing rate equations reduce to:

\begin{align}
\tilde{r}_E(k) &= -\frac{1}{\det(1 - \mathbf{W}(k))} \tilde{w}_{EI}(k) \tilde{h}_I(k) \quad (2.26) \\
\tilde{r}_I(k) &= \frac{1}{\det(1 - \mathbf{W}(k))} (1 - \tilde{w}_{EE}(k)) \tilde{h}_I(k) \quad (2.27)
\end{align}

Since $\det(1 - \mathbf{W}(k))$ must be positive for the network to be stable, only when $\tilde{w}_{EE}(k) > 1$ will an increase in $\tilde{h}_I(k)$ cause a decrease in $\tilde{r}_I(k)$, and thus the network must be operating within the ISN regime to yield the paradoxical decrease in inhibitory firing rates. Because the strength of connectivity decreases monotonically to zero as a function of distance, in the Fourier domain the strength of connectivity also decreases as a function of increasing spatial frequency. Each spatial frequency behaves as an independently evolving E-I pair, and so there should be a particular spatial frequency at which the strength of recurrent
excitation, $\tilde{w}_{EE}(k)$, drops below the level sufficient to support the ISN regime, and the behavior of the network switches from ISN to non-ISN. This critical frequency, $k_c$, is where $J_{EE}\sigma_{EE}\sqrt{2\pi e}^{-\left(k_c\sigma_{EE}/\sqrt{2}\right)^2} = 1$. Re-arranging to solve for $k_c$:

$$k_c = \sqrt{2 \ln \left( \frac{J_{EE}\sigma_{EE}\sqrt{2\pi}}{\sigma_{EE}} \right)}$$  \hspace{1cm} (2.28)

Equation 2.28 shows that the critical frequency, $k_c$, above which the network will shift from ISN to non-ISN behavior, decreases nearly linearly with increasingly extensive excitatory-to-excitatory synaptic connections, as described by the decay term, $\sigma_{EE}$. Additionally, if the product $J_{EE}\sigma_{EE}\sqrt{2\pi} < 1$ then the value of $k_c$ will be imaginary, demonstrating that there will be no frequency for which the network will be operating as an ISN.

The existence of this cut-off frequency yields an interesting prediction with regard to the behavior of the network in response to inhibitory inputs of differing spatial power spectra — all stimulus frequencies below $k_c$ should respond with ISN-like behavior, and all frequencies above $k_c$ should act as non-ISNs. Thus if one were to stimulate only I cells (for example with channelrhodopsin) at increasing spatial frequencies, there would be a cutoff frequency at which the responses of E and I cells undergo a $180^\circ$ transition in their relative phases. At low spatial frequencies the network operates as an ISN, and E and I cells should be spatially and temporally in phase with each other (and out of phase with the input), but above this cutoff frequency, the E and I populations should be out of phase with each other (and inhibitory cells would be in phase with the input). We tested this prediction in our model by stimulating only the inhibitory cells in the network with a “photostimulus” with a defined spatial frequency. Stimuli with a spatial frequency below the critical frequency fall within the ISN regime, and so E and I firing rates are modulated in phase with each other, and out of phase with the input (Figure 2.6 left). This is the spatial analogue of the paradoxical response described by Ozeki et al. (2009). Stimuli with a spatial frequency above the critical frequency are outside of the ISN regime, and so E and I move out of phase with each other, and I moves in phase with the input (Figure 2.6 right).
By simply examining equations 2.28 and 2.18, we can see that not only must the network be within the ISN regime to have inhibitory resonance, but further, that the resonant frequency of the inhibitory elements, $k_{ri}$, must be less than the critical frequency, $k_c$. Thus the resonant frequency of the inhibitory elements must itself fall within the range of frequencies that are inhibition-stabilized.

2.6 Discussion/conclusions

We have shown here that in a circuit model of V1 in which the center and surround receptive fields are modeled self-consistently, surround suppression of both E and I cells requires that the network operate within the ISN regime. This finding reaffirms the results of Ozeki et al. (2009), and extends their results by demonstrating a possible mechanism by which this suppression is achieved: the selective amplification of spatially-periodic patterns of activity over cortical space.

In addition to explaining surround suppression of both E and I neurons, this mechanism also offers a novel explanation for the experimental findings of Tanaka and Ohzawa (2009), who observed tuning in V1 to the modulation spatial frequency of a CM grating. Such a tuning may simply be epiphenomenal, or may instead play a role in establishing the well-known tuning of neurons in V2 to “second-order” spatial frequencies (Zhou and Baker, 1994; 1996; Leventhal et al., 1998; Mareschal and Baker, 1998). This tuning has been implicated as a key step in distinguishing texture boundaries, and may be important in other early visual computations (Baker and Mareschal, 2001). This simple model makes a number of testable experimental predictions. One of the most straightforward of these is that one should be able to observe spatial periodicity in the responses of V1 neurons. There is actually considerable evidence of this in the literature. A number of length- and size-tuning experiments in V1 actually report neurons whose responses vary nonmonotonically in the presence of progressively larger stimuli (Anderson et al., 2001; Sengpiel et al., 1997; Li and Li, 1994; De Valois et al., 1985; Wang et al., 2009). More
often, however, such periodicity is not reported. The chief reason for this is likely that the majority of length-tuning studies probe the stimulus space with only 8-12 stimuli, whose sizes are typically logarithmically spaced (Ichida et al., 2007; Sceniak et al., 1999; Schwabe et al., 2010; Shushruth et al., 2009; Ozeki et al., 2004). Though such a protocol may be useful for accurately estimating the size of the summation field (the first peak in a length-tuning curve), they will invariably miss the fine structure of the far surround. We address this potential pitfall directly in two experimental tests presented in a later chapter.

Another prediction of this model is the existence of a critical spatial frequency, above which connectivity strength decays to a level unable to support the ISN regime. Probing for such a frequency experimentally would require a way to selectively stimulate a population of inhibitory cells without also driving excitatory cells. Though virtually impossible only a few years ago, with the advent of sophisticated optogenetic techniques (Deisseroth, 2011; Yizhar et al., 2011), such a test has now become technically possible, and we are excited to attempt this soon.

Importantly, all of the results and analyses presented here have assumed linear neurons. In reality, we know that neurons are highly nonlinear units, and this nonlinearity invariably has a major impact on the properties of the cortical network (Miller and Troyer, 2002; Kayser et al., 2001). In the subsequent chapter, we explore the effect of including a simple, static nonlinearity into our neural model. The effects we observe are exciting, as the simple circuit motif we have described above is transformed into an exceptionally powerful model of visual cortex.
Chapter 3

Nonlinear Models of Contextual Modulation and Normalization

3.1 Introduction

Though a simple linear model is capable of explaining several interesting forms of contextual modulation, these behaviors are really only a minute subset of the full repertoire of V1 responses. In particular, it has been shown in numerous experiments that contextual modulation depends strongly on the relative and absolute contrasts of the center and surround stimuli (Sengpiel et al., 1997; Polat et al., 1998; Anderson et al., 2001; Cavanaugh et al., 2002b; Sceniak et al., 1999; Song and Li, 2008; Wang et al., 2009). In a linear network, responses can only scale linearly with input strength. To explain contrast-dependent changes in network behavior, we must consider the role of nonlinearity in our model.

Previous models of V1 have invoked the nonlinearity inherent in the neural response as a mechanistic explanation for the contrast-dependency of a number of extra-classical receptive field (eCRF) properties (Somers et al., 1998; Schwabe et al., 2006). However, these models all depend on an asymmetry in the intrinsic, biophysical response properties of excitatory and inhibitory neurons, a feature which may or may not exist (Contreras and
3.1. Introduction

Palmer, 2003). As an alternative hypothesis, we propose that V1 could use nonlinearity to transition between effective dynamic regimes, behaving sometimes as an ISN and sometimes as a non-ISN. If such a transition were possible, it could allow the network to respond in qualitatively different ways to stimuli of different strength, without requiring an asymmetry in intrinsic cellular parameters. In the first portion of this chapter, we test this hypothesis in our model of spatial contextual modulation by introducing a symmetric (between E and I), static nonlinearity and studying how this new feature alters the response properties of the network. We show that this slight modification to the model is sufficient to account for a surprisingly diverse array eCRF of properties.

By including nonlinearity in our network, we are now able to explore a more diverse array of cortical behaviors. One such behavior is normalization. Normalization refers to an operation in which the responses of individual neurons are scaled, most typically through division, by some metric of the overall level of input to the network. Phenomenological models of normalization have been shown repeatedly to accurately predict the responses of individual cells (Heeger, 1992; Carandini and Heeger, 1994) and neuronal populations (MacEvoy et al., 2009; Busse et al., 2009), however, there has to date been no consensus on the underlying circuit mechanisms. Previously it has been proposed that normalization may be accounted for by shunting inhibition (Carandini et al., 1997), thalamocortical synaptic depression (Carandini et al., 2002), or through an increase in the level of noisy background activity, which causes a divisive reduction in response gain (Chance et al., 2002; Finn et al., 2007). In the second portion of this chapter, we show how the nonlinear version of our model offers a novel circuit mechanism for cortical normalization. We then show how this same model can account for the decrease in trial-to-trial variability observed at stimulus onset (Churchland et al., 2010).

In the third section of this chapter, we combine our models of spatial contextual modulation and normalization into a unified, large-scale model of primary visual cortex, and use it to explore some of the more nuanced features of contextual modulation. Lastly, we propose
3.2 Nonlinear spatial model

3.2.1 Introduction and motivation

In addition to position and size, one of the most important factors affecting the modulatory role played by the eCRF is the absolute and relative contrast of the stimuli in the center and surround. A number of eCRF properties undergo qualitative changes with changes in center stimulus contrast. The response to a low contrast center stimulus can be facilitated by an iso-oriented, high contrast surround, whereas the response to a high contrast center stimulus will almost always be suppressed by a high contrast surround (Sengpiel et al., 1997; Polat et al., 1998). Firing rate and conductance length-tuning curves that have only a single peak at low contrast can have two peaks at high contrast (Sengpiel et al., 1997; Anderson et al., 2001). An orthogonally oriented surround stimulus, which has on average no effect on a high contrast central stimulus, becomes suppressive for a low contrast central stimulus (Cavanaugh et al., 2002b), and more generally, the orientation tuning of the eCRF weakens with low center contrast (Levitt and Lund, 1997). The size of a neuron’s summation field, measured as the region of space over which an increase in the diameter of a circular stimulus will increase firing rate, shrinks by up to 2-10 fold at high contrast (Cavanaugh et al., 2002a; Sceniak et al., 1999; Song and Li, 2008).

Previous models have proposed that these contrast-dependent changes result from an intrinsic asymmetry between the contrast response properties of excitatory and inhibitory neurons (Somers et al., 1998; Schwabe et al., 2006). In these models, it is assumed that the inhibitory neurons have both a higher contrast threshold and contrast gain. At low contrast, only excitatory neurons in the center are active, so additional input from the surround (which is by definition subthreshold) yields a purely facilitatory effect. At high contrast, both E
and I cells are active in the center. Since the inhibitory cells have a higher gain, additional input will drive these cells more, causing a net suppression. The experimental prediction from these models is that if one measures the contrast response properties of identified E and I cells, this asymmetry should be apparent. A recent attempt to do this, however, has shown that E and I cells have statistically indistinguishable contrast threshold and gain, and vary only in their maximum firing rate (Contreras and Palmer, 2003).

As an alternative hypothesis, we propose that V1 could use nonlinearity to transition between effective regimes, behaving as either an ISN or non-ISN depending on the shape and magnitude of the network input. Such a transition would allow the network to respond in qualitatively different ways to stimuli of different strength, without requiring any asymmetry between E and I cells in intrinsic cellular parameters. We test this hypothesis in our model of spatial contextual modulation by introducing a static nonlinearity into the network that is identical for E and I cells and studying how this addition alters the response properties of the network. Here we show that this slight modification to the model is sufficient to account for a surprisingly diverse array of eCRF properties.

### 3.2.2 Model details and parameters

In this section, we introduce a rectifying, nonlinear input-output function into the network, and use this model to offer a novel mechanistic explanation for qualitative, contrast-dependent changes in certain V1 eCRF response properties. We model the input-output relationship as a power-law function:

\[ F(x) = k([x]^+)^n \]  

(3.1)

where \([x]^+ = \max(x, 0)\) and \(n > 1\). This description has been shown to be a good approximation of the true input-output relationship of neurons in V1 (Miller and Troyer, 2002; Priebe et al., 2004). In implementing this nonlinearity into our model, we replace the linear state equations with the following:
3.2. Nonlinear spatial model

\[
\tau_E \frac{d}{dt} r_E(x) = -r_E(x) + k \left( [W_{EE} \ast r_E(x) - W_{EI} \ast r_I(x) + c_E h(x)]_+ \right)^n
\]  (3.2)

\[
\tau_I \frac{d}{dt} r_I(x) = -r_I(x) + k \left( [W_{IE} \ast r_E(x) - W_{II} \ast r_I(x) + c_I h(x)]_+ \right)^n
\]  (3.3)

In all simulations, we use the same values of \( k \) and \( n \) for both the excitatory and inhibitory populations. This parameter choice is inspired by the experimental work of Contreras and Palmer (2003), who recorded intracellularly from putative excitatory and inhibitory neurons in V1 and characterized their contrast-response functions. They showed that, for both the firing rate and membrane potential response of V1 neurons to stimuli of increasing contrast, the parameters determining the gain of the nonlinearity in E and I cells were statistically indistinguishable, and that these populations differed significantly only in their maximum firing rate.

Connectivity follows the same rules as in the linear model, and as before, the \( \ast \) denotes the spatial convolution of firing rates over the grid with the connectivity functions. For simulations of the nonlinear model, we used the following parameters: \( N = 101, \Delta x = 0.5, \tau_E = 20 \text{ ms}, \tau_I = 10 \text{ ms}, J_{EE} = 1.00, J_{IE} = 1.25, J_{EI} = 1.0, J_{II} = 0.75, \sigma_{EE} = 1.0, \sigma_{IE} = 2.0, k = 0.01, \) and \( n = 2.2 \). Stimulus strength \( (c \text{ in equations 3.2 and 3.3}) \) was equal for both E and I, and varied from 0 to 100. The parameter defining the sharpness of the step-function input was \( \sigma_{RF} = 0.125 \). Baseline input (except where otherwise reported) was set to 0. To best match the electrophysiological recordings we use a cortical magnification factor of 1.5 mm/deg. Because of the nonlinearity, steady-state responses can no longer be calculated analytically; both steady-state and dynamic responses for this and the other nonlinear models are calculated through simulation. As above, we used a forward-Euler algorithm with a fixed 1 ms time-step.
3.2. Nonlinear spatial model

The input dependence of dynamic regime in the nonlinear network model. A. The responses of both E and I cells to increased input only to inhibitory cells while both populations are driven by a constant input. For weak constant input, I firing rates increase and E firing rates decrease with increased input to inhibitory cells. At higher levels of constant input, the network moves into the ISN regime, so both E and I firing rates decrease. B. The net change in E and I firing rates versus tonic input strength. C. The maximum real part of the eigenvalues of both $\hat{W}$ and $\hat{W}_{EE}$, derived from the instantaneous linearization of the nonlinear system. The stimulus strength where the maximum real eigenvalue of $\hat{W}_{EE}$ goes unstable (> 0) matches almost exactly the stimulus strength where inhibitory cells begin to respond paradoxically to increased excitatory input. The network actually enters the ISN regime at a slightly (< 0.5 units) lower stimulus strength, but there is a small region of stimulus strength where the input to the inhibitory cells used to test for the paradoxical response decreases the E firing rates enough to push the network out the ISN regime.

3.2.3 Input dependence of dynamic regime

With the addition of an expansive nonlinearity, we predicted that our network would be able to transition between the non-ISN and ISN regimes as a function of input strength. At low input strength (and consequently low firing rates), neuronal gain is very shallow, such that
a small change in the input to any one cell causes relatively little change in the firing rate of that cell, and consequently little change in the input to other cells in the network. Thus for weak inputs, the effective connectivity will be too weak to support the ISN regime. As the strength of the input is increased and neurons move into the steeper portion of their input-output curves, small changes in the input to one cell can cause large changes in the firing rate and the input to other cells, so that effective connectivity is much stronger – strong enough to support ISN dynamics. Depending on the shape of the stimulus, there should be a particular threshold contrast at which the network enters the ISN regime.

To understand how the addition of nonlinearity to the network model influences dynamic regime, we derive an analytic expression for the linearization of the dynamics around the fixed point firing rates by finding the Jacobian matrix \( \hat{W} \) of the dynamical system:

\[
\hat{W} = T^{-1} \left( \Phi \left( \begin{array}{cc}
knW_{EE} & -knW_{EI} \\
knW_{IE} & -knW_{II}
\end{array} \right) - I \right)
\]

(3.4)

Where \( \Phi \) is a diagonal matrix defined as:

\[
\Phi = \left( \begin{array}{cc}
\text{diag}(\Phi_E) & 0 \\
0 & \text{diag}(\Phi_I)
\end{array} \right)
\]

(3.5)

Here \( \text{diag}(v) \) denotes a diagonal matrix with the elements of the vector \( v \) along the diagonal. \( \Phi_E \) and \( \Phi_I \) are \( N \)-dimensional vectors that depend on the excitatory and inhibitory firing rates, defined as:

\[
\Phi_E = (W_{EE}r_E - W_{EI}r_I + c_c h)^{n-1}
\]

(3.6)

\[
\Phi_I = (W_{IE}r_E - W_{II}r_I + c_I h)^{n-1}
\]

(3.7)

Exponentiation is performed element-by-element. \( T \) is a diagonal matrix of time constants; the first \( N \) entries of the diagonal equal \( \tau_E \) and the second \( N \) entries equal \( \tau_I \). In the
linearization, the dynamics around the fixed-point \( r_{fp} \) evolve as:

\[
\frac{d\hat{r}}{dt} = \hat{W}\hat{r}
\]

where \( \hat{r} \) is the vector of firing rates relative to the fixed point, such that \( \hat{r} = r - r_{fp} \).

Using this approximation, we can numerically calculate the eigenvalues and eigenvectors of the effective linear weight matrix, \( \hat{W} \), at a given steady-state and use them to predict the effective dynamic regime of the nonlinear network. Just as in the linear model, we say that the nonlinear model is an effective ISN when the excitatory subnetwork is unstable, but the overall network is stable. In terms of the linearization, this occurs when at least one eigenvalue of \( \hat{W}_{EE} \) (the upper right quadrant of \( \hat{W} \)) is positive, but all of the eigenvalues of \( \hat{W} \) are negative. Because the linearization depends on the fixed-point firing rates of the network, which in turn depend on the shape \( h \) and magnitude \( c \) of the stimulus, it is possible that the network may be able to transition between effective dynamic regimes depending on the specifics of the stimulus.

Using this formulation, we can explore how well the behavior of the nonlinear network can be predicted by the linearized approximation. As a simple test, we stimulated E and I cells with a small, centrally-located stimulus of increasing strength and calculated the eigenvalues of \( \hat{W} \) and \( \hat{W}_{EE} \). We then perturbed only the inhibitory cells with a small additional excitatory input, and recorded the responses of both E and I cells. We observe that when the linearization predicts a switch into an effective ISN regime (the largest eigenvalue of \( \hat{W}_{EE} \) has real part > 1), the inhibitory cells begin to respond to the additional excitatory input with a decrease in firing rate (Figure 3.1).
3.2. Nonlinear spatial model

3.2.4 Results

To test our hypothesis that the nonlinear transition between the non-ISN and ISN regimes can account for the contrast-dependent differences in the effect of surround modulation, we reran the length-tuning simulations on the nonlinear spatial model using stimuli of increasing contrast. For low-contrast stimuli, the response of the cells at the center of the network...
3.2. Nonlinear spatial model

network simply grows with the expanding input, eventually plateauing for very large stimuli. With increasing stimulus contrast, though, a clear transition can be seen. There first appears only weak surround suppression, which is then followed by progressively stronger surround suppression and as well as the emergence of a second peak in the length-tuning curve (Figure 3.2). These results indicate the gradual emergence and strengthening of the spatially-periodic activity explored in the linear model, exactly as predicted by a network that smoothly transitions into the ISN regime. This same contrast-dependence was observed experimentally by Anderson et al. (2001), who reported a second peak in inhibitory length-tuning at high contrast, but not at low contrast, as well as Sengpiel et al. (1997), who observed the same effect for firing rate in a V1 neuron.

In addition to an increase in the strength of spatially-periodic activity, these length-tuning curves also show an increase in the spatial frequency at which the activity is oscillating. Using the linear approximation derived above, we can solve for the resonant frequencies of the network as a function of stimulus strength. To do this, we drive the network with spatially and temporally constant inputs of different magnitudes. Because of the spatial-uniformity of the resulting activity pattern, the effective connectivity submatrices (\(\hat{W}_{EE}, \hat{W}_{IE}, \hat{W}_{EI}, \) and \(\hat{W}_{II}\)) will be translationally invariant. As in the previous chapter, we can use the eigenvalues of these connection submatrices to numerically calculate the maximally amplified spatial frequency at each level of input strength. We find that the fixed-point solution predicts a monotonic increase in spatial frequency with input magnitude (Supplementary Figure 7.1). This results in a leftward shift in the first peak of the length-tuning curves with increasing contrast, causing a contrast-dependent reduction in summation field size (Figure 3.3A). The magnitude of this shift varies from approximately 1-12 fold for E cells and 1-8 fold I cells, depending on the specific values chosen to represent “low” and “high” contrast. Contrast-dependent changes in summation field size have been observed several times experimentally (Sceniak et al., 1999; Cavanaugh et al., 2002a), though the mechanism underlying this shift is still unknown. We propose that an increase in the amplitude and
3.2. Nonlinear spatial model

FIG. 3.3: **Contrast dependence of summation field size and surround valence.** **A.** Summation field size is calculated as the stimulus size that yields the first maximum response on the length-tuning curve. Values are normalized to the summation field size at the maximum stimulus strength (dashed line) to show the range of changes observed. **B.** The dashed red line shows the contrast response curve of an excitatory cell to a small centrally located stimulus. With a surround stimulus held constant at a strength of 50, the strength of the center stimulus is systematically varied (solid curve). For a weak center stimulus, the high-contrast surround is facilitative. For a stronger center stimulus, the high-contrast surround is suppressive. The center stimulus exactly fills the high-contrast summation field for the neuron, which for these parameters is 0.55°. The surround in this case consists of two flanking stimuli adjacent and each equal in length to the center stimulus, such that the center-surround complex is 3 times the length of the original center stimulus. In Supplementary Figure 7.3 we vary the size of the surround stimulus parametrically.

frequency of spatially periodic activity, as predicted by our model, may underlie this effect.

It has been observed that a high contrast surround stimulus can facilitate the response to a low contrast center, yet suppress the response to a high contrast center stimulus (Sengpiel et al., 1997; Polat et al., 1998). This switch from facilitation to suppression occurs naturally in our model as a consequence of the shrinking summation field. As the contrast of the center stimulus is increased, a surround stimulus that was originally located within the summation field will now be in the suppressive surround, and thus its relative effect on the center will be reversed (Figure 3.3B).

To confirm that these results were not dependent on the specific shape of the input, we
repeated these two simulations using a stimulus with a Gaussian spatial profile, rather than the smoothed-step function profile. Both the tests of summation field size and the transition from facilitation to suppression show no qualitative difference with Gaussian rather than step-function inputs (Supplementary Figure 7.2). However, the transition from a facilitatory to suppressive surround does depend strongly on the relative size of the surround stimulus used. This is illustrated in Supplementary Figure 7.3. In Figure 3.3B, we used a surround stimulus with a diameter three times the size the center stimulus. In the supplementary materials, we vary the surround size parametrically while probing over the range of center stimulus strengths. For relatively small surrounds (2-4x), there is a substantial region of facilitation. For larger surrounds, however, the effect is always suppressive. Later in this chapter, we show that in a more realistic large-scale network model, this same trend exists, however there is substantial cell-to-cell variability. Some cells show regions of facilitation even for fairly large surrounds, while others have only suppression even with small surround stimuli.

3.2.5 Summary

This simple nonlinear spatial model demonstrates a novel mechanism that can explain the contrast-dependence of a number of eCRF properties. Through a transition in effective dynamic regime, activity in the model undergoes a contrast-dependent increase in both the amplitude and frequency of spatial periodicity. These changes cause a decrease in summation field size that can cause a switch from surround facilitation to surround suppression. Importantly, none of these effects rely on an asymmetry in the input-output functions of the E and I cells, a previously proposed mechanism (Somers et al., 1998; Schwabe et al., 2010) that has proven experimentally elusive (Contreras and Palmer, 2003; Nowak et al., 2010).
3.3 Normalization

3.3.1 Introduction and motivation

Recently it has been shown that the V1 population response to multiple stimuli is a sublinear sum (roughly, the average) of the responses to the individual stimuli alone (MacEvoy et al., 2009; Busse et al., 2009). Similarly, in V4, the response to two stimuli in different portions of the receptive field is roughly the average of the individual responses (Reynolds et al., 1999). These and other nonlinear response properties have been phenomenologically described as “normalization”, but their mechanistic origins are unknown. Though some features traditionally ascribed to normalization are present in the feedforward inputs to cortex and/or can arise through nonlinearities of cells and synapses, rather than circuit properties (Kayser et al., 2001; Lauritzen et al., 2001; Carandini et al., 2002; Priebe and Ferster, 2006; Li et al., 2006a), there is still considerable evidence that cortical circuits show normalizing behavior. In this section we explore to what extent our nonlinear circuit model can account for normalizing features of the V1 response, focusing in particular on the sublinear addition of multiple stimuli.

3.3.2 Details of the model

To explore whether our circuit model could offer some mechanistic insight into this form of normalization, we converted the nonlinear spatial model described above into a ring model of orientation specificity (Figure 3.4). Positions on the ring correspond to preferred orientation $\theta$, which ranges from $0^\circ$ to $180^\circ$ with movement around the ring (and $0^\circ$ and $180^\circ$ represent the same orientation). Rather than a spatially extended patch of cortex, as modeled above, we think of this model as describing a single orientation hypercolumn in V1. We use the same nonlinear rate equations as in the nonlinear spatial model (eqs. 3.2 and 3.3), and again we use identical nonlinearity parameters ($k$ and $n$) for both E and I cells.
A cartoon of the nonlinear ring model. In this model, the location of each neuron on the ring is defined by its stimulus orientation preference. All four types of synaptic connections (E \Rightarrow E, E \Rightarrow I, I \Rightarrow E, and I \Rightarrow I) have the same extent in orientation space, and differ only in magnitude. An input to this network is modeled as a Gaussian curve centered on the neuron with the matching orientation preference.

\[ \tau_E \frac{d}{dt} r_E(\theta) = -r_E(\theta) + k \left( [W_{EE} \ast r_E(\theta) - W_{EI} \ast r_I(\theta) + c_E h(\theta)]_+ \right)^n \] (3.9)

\[ \tau_I \frac{d}{dt} r_I(\theta) = -r_I(\theta) + k \left( [W_{IE} \ast r_E(\theta) - W_{II} \ast r_I(\theta) + c_I h(\theta)]_+ \right)^n \] (3.10)

As we are modeling the interactions between neurons with different preferred orientations, the connectivity rules have been altered to reflect the pattern of synaptic connectivity believed to operate in the orientation domain. A number of experiments have shown that neurons in V1 connect preferentially to other neurons with the same preferred orientation (Malach et al., 1993; Bosking et al., 1997), and that the orientation tuning for excitatory and inhibitory inputs is approximately equal in width (Ferster, 1986; Anderson et al., 2000). We again model connectivity as decreasing in strength as a Gaussian function of distance (in
preferred orientation, $\theta$), but used the same width parameter, $\sigma_{ori}$, for all four connection types ($E \Rightarrow E$, $E \Rightarrow I$, $I \Rightarrow E$, and $I \Rightarrow I$). With $a, b \in \{E, I\}$:

$$W_{ab}(\theta - \theta') = J_{ab} e^{-\frac{(\theta - \theta')^2}{2\sigma_{ori}^2}} \quad (3.11)$$

For simulations of the nonlinear ring model, we used the following parameters: $N = 180$, $\Delta \theta = 1^\circ$, $\tau_E = 20$ ms, $\tau_I = 10$ ms, $J_{EE} = 0.0441$, $J_{IE} = 0.04158$, $J_{EI} = 0.0231$, $J_{II} = 0.01827$, $\sigma_{ori} = 32^\circ$, $k = 0.04$, and $n = 2.0$. Feedforward stimuli are modeled as Gaussian curves defined in orientation space by the neuron on which they are centered. Input width was set to $\sigma_{FF} = 30^\circ$. Stimulus strength was equal for $E$ and $I$ and varied from 0 to 100.

### 3.3.3 Results

To test if our model could reproduce the experimental finding that V1 sublinearly adds multiple orthogonal inputs (MacEvoy et al., 2009; Busse et al., 2009), we ran a series of simulations using high contrast stimuli oriented at either $45^\circ$, $135^\circ$, or both $45^\circ$ and $135^\circ$ (Figure 3.5). The model shows clear sublinear addition of the responses, and as was the case with surround suppression, there was no increase in inhibition. To quantify this effect, we found the best-fit weights for the equation:

$$R_{1+2} = w_1 R_1 + w_2 R_2 \quad (3.12)$$

where $R_1$ is the response to one stimulus, $R_2$ is the response to the other stimulus, $R_{1+2}$ is the response to both presented simultaneously, and $w_1$ and $w_2$ are the two weights we fit. When $w_1$ and $w_2$ are less than one, the model demonstrates sublinear addition.

Normalization was also observed with equal contrast gratings presented at non-orthogonal orientations (MacEvoy et al., 2009). To test our model for this response feature, we reran our test of sublinear addition while parametrically varying the orientation difference between the
3.3. Normalization

Results

Mean

Sum

Stimulus 1
Stimulus 2
Stimulus 1+2

W1 = 0.68
W2 = 0.68

Preferred Orientation

0 20 40 60 80 100 120 140 160 180

Excitatory Firing Rate

Inhibitory Firing Rate

FIG. 3.5:

Sublinear addition of multiple stimuli. The top two rows show the firing rate responses of E and I cells to a 45° stimulus and a 135° stimulus. The responses to both stimuli shown simultaneously are plotted in the third row. All inputs used in this simulation are strength = 50. The magnitudes of the sublinear weights are indicated on the third row plot. A composite plot showing the actual response (red), as well as the mean (blue) and linear sum (green) of the responses to the two individual stimuli is shown in the fourth row.

two stimuli. In Figure 3.6 we show the responses of the network to pairs of stimuli separated by 90°, 40°, and 20° as well as the additive weights across all orientation differences tested. As in the experimental data, responses were strongly sublinear for all orientation pairs. The model also reproduces the “winner-take-all” effect described by Busse et al. (2009), who found that as the contrast of one grating was increased relative to its orthogonal partner, the response of the population would quickly resemble the response to the higher contrast grating alone (Figure 3.7) As we will show later, this effect actually follows directly from the property of sublinear summation (see section 3.3.5.6).

At least one previous experiment demonstrated that cross-orientation suppression could be accounted for by contrast saturation in the LGN (Li et al., 2006a). Arguing against this mechanism, MacEvoy et al. (2009) repeated their experiments at multiple levels of
stimulus contrast. Sublinear addition was observed at all contrast levels tested (12.5%, 25%, and 50%). To test our model for this property, we repeated our test of normalization across the full range of physiological contrasts. Consistent with this experiment, we observed sublinear weights between 0.6 – 0.7 for almost the entire range of stimulus strengths tested. However, for very weak stimuli (< 10), we actually observed the opposite effect: the network adds inputs supralinearly (Figure 3.8). Interestingly, an analogous effect has been observed multiple times in brain areas responsible for multisensory integration, such as the superior colliculus (Meredith and Stein, 1986; Perrault et al., 2003; Stanford et al., 2005; Stein and Stanford, 2008; Ohshiro et al., 2011), and at least once in area MT (Heuer and Britten, 2002) (Supplemental Figure 7.4). This prediction has not, to our knowledge, been tested explicitly in V1.

3.3.4 Mechanisms

Previous models have posited that normalization occurs through the divisive action of an inhibitory pool of “normalizing interneurons”, which are typically modeled as being less stimulus specific (or completely nonspecific) than the excitatory projection neurons (Heeger, 1993; Carandini et al., 1997). In these models, the activity of this normalizing population reflects the overall level of input to the network. By providing a divisive feedback signal to all of the excitatory neurons, it effectively normalizes the network’s activity. In our model, we have no such pool of nonselective, globally projecting interneurons. Nor does our model produce an increase in either the inhibitory firing rates or inhibitory conductance. Thus, we sought to understand what alternative mechanism was accounting for normalization.
3.3. Normalization

FIG. 3.6: **Non-orthogonal gratings are also normalized.** A. The three firing rate plots show the individual responses, the combined response, and the sum and the mean of the individual responses for orientation differences of 90°, 40°, and 20°. B. The sublinear weights for E and I are also plotted across stimulus orientation difference. Stimulus strength is set to 50 for all of these simulations.

Interestingly, we noted a number of parallels between this model and our model of surround suppression. In both cases, the firing rates of both the E and I cells are reduced (both populations are surround suppressed and normalized). Similarly, we observed in both
3.3. Normalization

FIG. 3.7: **Winner-take-all responses.** A. The four response profiles show the excitatory cell firing rates across the network as the stimulus strength difference is increased. The four stimulus strength pairs shown here are (40,40), (50,30), (60,20), and (70,10). B. A summary plot shows the best-fit weights as a function of increasing stimulus strength disparity.

cases a transition between a facilitating low contrast regime and a sublinear, suppressive high contrast regime. Furthermore, if we perform a test of “width-tuning” in the ring model, we find that at a given level of input strength, the network responds maximally to feedforward input of a particular width in orientation space. Stimuli that are wider or narrower than this preferred width elicit a sub-maximal response (Figure 3.9A). Similarly to the nonlinear spatial model above, the preferred orientation width (analogous to the summation field size in the spatial domain) of the network shrinks with increasing stimulus strength (Supplementary Figure 7.5).

Of course, in the real biological system, the width of a stimulus in orientation space is not an independently variable parameter, but rather dependent on both the stimulus and the feedforward connectivity (Reid and Alonso, 1995; Troyer et al., 1998). Though orientation width can be altered by varying the stimulus size and spatial frequency, these manipulations would have direct effects on the firing rate (e.g. the cell’s response is tuned for spatial frequency) that would make it impossible to isolate the effect of changing orientation width.
3.3. **Normalization**

FIG. 3.8: **Normalization versus stimulus strength.** The additive weights for equal strength stimuli are calculated while varying stimulus strength. For most of the physiological range, weights are strongly sublinear for both E and I cells. For very weak stimuli, the model predicts that addition will instead be supralinear.

But just because the stimulus orientation width cannot itself be manipulated, it doesn’t mean that the network might not have some preferred width that changes with input strength. If normalization were occurring through the same contrast-dependent surround suppression we observed in the spatial model, this would also explain the contrast-dependent switch from supra- to sublinear addition. At low contrast the preferred stimulus width is actually larger than the fixed width of a single feedforward stimulus, and so by adding a second stimulus, the neuron receives an effectively wider input. This produces an effective surround facilitation, which will boost the response. At high contrast, the preferred width is equal to or narrower than the width of a single stimulus, and so adding a second stimulus simply produces surround suppression. This mechanism makes the prediction that the contrast at which the network switches from supra- to sublinear addition should depend on the feedforward stimulus width (Supplementary Figure 7.6).

This contrast-dependent width-tuning in the orientation domain should be distinguished
3.3. **Normalization**

![Diagram](image)

**FIG. 3.9:**

**Width-tuning in orientation space through nonlinear scaling of effective connectivity.** **A.** Plotted here are the firing rates elicited by inputs of different width at five different levels of stimulus strength. Rates are normalized to the maximum rate for each stimulus strength level. With increasing stimulus strength, the width of the stimulus eliciting the maximum response from the network decreases. For all but the weakest stimulus, the preferred width is less than the width of the feedforward input (30 degrees), so multiple stimuli cause a decrease in firing rate. Only for very low-contrast stimuli will the second stimulus actually cause an increase in firing rate, because the preferred width of the network at this contrast level is larger than the feedforward input **B.** With increasing contrast, the asymmetry between the width of the firing rate profiles of the excitatory and inhibitory cells grows, causing an increase in the difference between the effective $\sigma_{EE}$ and $\sigma_{IE}$.

from the width of orientation tuning within the CRF, which is known to be contrast invariant (Sclar and Freeman, 1982). The contrast-dependent tuning that we describe refers to the stimulus width (in orientation space) that yields the maximum response. Contrast-invariant orientation tuning is measured with a stimulus of constant width in orientation space whose center is rotated around the orientation domain. The spatial analogues of these two tuning properties are the spatial summation field size, which shrinks with increasing contrast (Sce- niak et al., 1999), and the minimum response field (MRF), which does not (Song and Li, 2008).

Though we found considerable evidence that normalization in this model was produced by surround suppression in the nonlinear ISN regime, we were at first puzzled about how this could occur. We have previously shown analytically, in the linear model, that
surround suppression depends critically on an asymmetry in the widths of $E \rightarrow E$ and $E \rightarrow I$ connections. In this current model, all four connection types have the same width, and so it would seem that surround suppression should be impossible. However, we then realized that in the same way that the effective connectivity strength in the nonlinear models scales with increasing firing rate, so does the the effective connectivity width scale with the width of the firing rate profile. If there is a slight difference in the width of activity patterns for $E$ and $I$ cells, then the effective connectivity functions will have the asymmetry necessary for surround suppression. And indeed we observe just such an asymmetry – in our model, once the network crosses the threshold into the effective ISN regime, the $I$ cell activity profiles are always slightly wider than the $E$ cell activity profiles. This makes intuitive sense, because while $I$ cells can suppress $E$ cell firing all the way to 0 at the edge of the activity bump, they cannot do the same to their own firing rates. This asymmetry in firing profiles causes the $E \rightarrow I$ to connections to be effectively wider than the $E \rightarrow E$ connections, providing the asymmetry in connectivity necessary for surround suppression (Figure 3.9B).

More generally, we find that networks with an expansive (e.g. power-law) input-output nonlinearity should always undergo a switch from a supralinearly to sublinearly summing regime. Given sufficiently strong recurrent excitatory connectivity, an expansive nonlinearity creates a generic switch in network behavior from a low contrast, facilitating, supralinear regime to a high contrast, suppressive, sublinearly adding regime. This occurs because at low contrast, when the gain on individual cells is still quite low, effective connectivity is weak and the dynamics of both the excitatory subnetwork and the overall network are stable. When a neuron receives an additional excitatory input, its response is increased supralinearly by its power-law input-output function; the connectivity of the network plays little role in the response. However, once a certain threshold input level is reached, gain becomes strong enough to make the excitatory subnetwork unstable (i.e. the network enters the ISN regime). In this state, if neurons responded supralinearly to their input, an unstable positive feedback loop would ensue and firing rates would go to infinity (or saturation). Because inhibition
is present in sufficient strength and speed, this potential for explosive activity instead shifts
the balance of E to I slightly towards I (for example with slightly wider activity profiles),
which de-amplifies the overall network (Murphy and Miller, 2009). In the end, the only
stable response is a sublinear increase (or even a net decrease) in firing rate. We detail this
argument in the analysis below.

3.3.5 Analysis

3.3.5.1 Two population nonlinear approximation

In the results section, we describe how normalization can be accomplished by the nonlinear
ring model; asymmetric firing rate profiles between E and I cells create an effective “Mex-
ican hat” in the connectivity, which allows the network to suppress stimuli that are wider
in orientation space than the contrast-dependent preferred stimulus width. This effective
asymmetric connectivity depends on the nonlinearity of the neurons – a linear model built
with equal connectivity widths cannot surround suppress.

Though this scheme describes how this particular network achieves normalization, what
we have found in our research is that more generally, all nonlinear networks of this basic
design (i.e. with extensive recurrence, an expansive nonlinearity, and sufficiently strong feed-
back inhibition to prevent instability) should all undergo an input-dependent transition in
stimulus response, from a supralinear regime to a sublinear (normalizing) regime.

To show more explicitly how this transition from supra- to sublinear addition can occur
in a general nonlinear model, we turn now to a simplified approximation of the model.
Though not an exact solution, the approximation detailed below is able to reproduce all of
the major important behaviors of the full model, and yet is analytically tractable enough to
yield interesting insight.

Recall that we define our network as \( N \) excitatory and \( N \) inhibitory neurons positioned
on a ring of preferred orientation whose firing rates evolve according the differential equations:
3.3. Normalization

\[
\tau_E \frac{d r_E(\theta)}{d t} = -r_E(\theta) + k([W_{EE} * r_E(\theta) - W_{EI} * r_I(\theta) + c_E h(\theta)]_+)^n \tag{3.13}
\]

\[
\tau_I \frac{d r_I(\theta)}{d t} = -r_I(\theta) + k([W_{IE} * r_E(\theta) - W_{II} * r_I(\theta) + c_I h(\theta)]_+)^n \tag{3.14}
\]

For now, let us focus our attention on the activity of a single pair of E and I cells located at \( \theta = 0 \). Define \( \vec{w} \) as the \( N \)-dimensional row vector giving the relative synaptic weights from all other positions in the network to the neurons at \( \theta = 0 \), normalized so the relative weight from \( \theta = 0 \) is 1 (recall that all four connectivity functions in this model have the same shape (Gaussians with the same width parameter)). The steady-state firing rates of these two neurons are:

\[
r_E(0) = k([J_{EE} (\vec{w} \cdot r_E) - J_{EI} (\vec{w} \cdot r_I) + c_E h(0)]_+)^n \tag{3.15}
\]

\[
r_I(0) = k([J_{IE} (\vec{w} \cdot r_E) - J_{II} (\vec{w} \cdot r_I) + c_I h(0)]_+)^n \tag{3.16}
\]

To get at a tractable solution, we now make the assumption that the excitatory and inhibitory firing rates across the network have the same basic shape, and only differ by a scalar, so that \( r_E = r_E(0) r \) and \( r_I = r_I(0) r \) (though we know that this assumption is faulty, as we will show below, the results are still fairly accurate), and \( r \) is the shape of the response normalized to have a peak of 1. We define the normalized activity profile in response to a single stimulus as \( r_1 \) and to two stimuli as \( r_2 \), and let \( \Psi_1 = \vec{w} \cdot r_1 \) and \( \Psi_2 = \vec{w} \cdot r_2 \) (and more generally \( \Psi = \vec{w} \cdot r \)). Note that \( \Psi_2 > \Psi_1 \), that is, the effect of adding a second stimulus is to increase \( \Psi \). Thus, normalization of E or I responses will occur if \( r_E(0) \) or \( r_I(0) \), respectively, decrease with increasing \( \Psi \). Since the input is simply scaled by the contrast, we let \( h(0) = 1 \), \( c_E = c_E h(0) \), and \( c_I = c_I h(0) \).

The excitatory and inhibitory firing rates at \( \theta = 0 \) are given by the following scalar equations:
3.3. Normalization

\[
    r_E(0) = k([J_{EE}r_E(0)\Psi - J_{EI}r_I(0)\Psi + c_E]_+)^n \tag{3.17}
\]

\[
    r_I(0) = k([J_{IE}r_E(0)\Psi - J_{II}r_I(0)\Psi + c_I]_+)^n \tag{3.18}
\]

Ignoring the rectification term for the time being, these two equations can be rearranged:

\[
    r_I(0) = \frac{J_{EE}r_E(0)\Psi + c_E - \left(\frac{r_E(0)}{k}\right)^{\frac{1}{n}}}{J_{EI}\Psi} \tag{3.19}
\]

\[
    r_E(0) = \frac{J_{II}r_I(0)\Psi - c_I + \left(\frac{r_I(0)}{k}\right)^{\frac{1}{n}}}{J_{IE}\Psi} \tag{3.20}
\]

Equations 3.19 and 3.20 define the nullclines of the excitatory and inhibitory firing rates, respectively, in a 2D space spanned by \( r_E(0) \) and \( r_I(0) \).

We can use these equations to look at a few features of the network. First, we can see where the network enters the inhibition-stabilized regime. If the excitatory nullcline has positive slope, the network must be an ISN (Ozeki et al., 2009).

\[
    \frac{\partial r_I(0)}{\partial r_E(0)} = \frac{J_{EE} - \frac{1}{n} \left(\frac{r_E(0)}{k}\right)^{\frac{1}{n}}}{J_{EI}\Psi nr_E(0)} \tag{3.21}
\]

So the network is in the ISN regime when:

\[
    J_{EE} > \frac{\left(\frac{r_E(0)}{k}\right)^{\frac{1}{n}}}{\Psi nr_E(0)} \tag{3.22}
\]

The denominator grows faster than the numerator (because \( n > 1 \)), demonstrating that there should be some minimum excitatory firing rate necessary to place the network in the ISN regime.
The second, and perhaps more interesting, way to use these nullclines is to approximate the steady state firing rates of the network. The network is at equilibrium where the two nullclines intersect. By plugging equation 3.19 into equation 3.20, we can derive an implicit function to define these equilibria:
3.3. Normalization

\[ r_E(0) = \frac{J_{II} \Psi \left( \frac{J_{EE} E(0)\Psi + c_E - \left( \frac{r_E(0)}{k} \right)^{\frac{1}{n}}}{J_{EI} \Psi \left( \frac{J_{EI} \Psi k}{J_{EI} \Psi} \right)^{\frac{1}{n}}} \right) + \left( \frac{J_{EE} E(0)\Psi + c_E - \left( \frac{r_E(0)}{k} \right)^{\frac{1}{n}}}{J_{EI} \Psi k} \right)^{\frac{1}{n}} - c_I}{J_{IE} \Psi} \] (3.23)

or equivalently:

\[ r_I(0) = \frac{J_{EE} \Psi \left( \frac{J_{II} \tau_1(0) \Psi - c_I + \left( \frac{\tau_1(0)}{k} \right)^{\frac{1}{n}}}{J_{IE} \Psi \left( \frac{J_{EI} \Psi k}{J_{EI} \Psi} \right)^{\frac{1}{n}}} \right) - \left( \frac{J_{II} \tau_1(0) \Psi - c_I + \left( \frac{\tau_1(0)}{k} \right)^{\frac{1}{n}}}{J_{IE} \Psi k} \right)^{\frac{1}{n}} + c_E}{J_{IE} \Psi} \] (3.24)

When evaluating this network numerically, one need only solve one of these two implicit equations; the other firing rate can be found just by plugging into equation 3.19 or 3.20.

3.3.5.2 Explicit representation

The implicit solutions above (equations 3.23 and 3.24) can be numerically evaluated quickly and efficiently, and are able to reproduce qualitatively the results from the full model (Figure 3.10). However, we would like to find an explicit representation of the equilibrium firing rates so that we can better understand the dependency of the fixed points on the model parameters.

One way to do this is to search our model for extrema. If we differentiate equation 3.23 with respect to contrast, we can get an explicit equation for \( \frac{\partial r_E}{\partial c} \). If, for the time being, we let \( n = 2 \) and \( c_E = c_I = c \), by setting this equation equal to zero, we can solve for \( c \) to find the contrast at which the maximum firing rate occurs (assuming the network “supersaturates”, i.e. at some contrast firing rates begin to decrease with increasing contrast).

\[ c_{\text{max}} = \frac{J_{EI}}{4 \left( J_{EI} - J_{II} \right)^2 k \Psi} - J_{EE} \Psi r_E(0) + \sqrt{\frac{r_E(0)}{k}} \] (3.25)

If we go back to equation 3.23 and substitute in \( c_{\text{max}} \) for \( c \), we can now solve for \( r_E(0) \)
exactly. Define:

\[ \beta = 2J_{EI} - 2J_{II} + J_{IE} - J_{EE} + 2\sqrt{(J_{EI} - J_{II})(J_{EI} + J_{IE} - J_{EE} - J_{II})} \] (3.26)

The maximum firing rate equals:

\[ r_{E}(0)_{\text{max}} = \frac{1}{4k\Psi^{2}\beta (J_{EI} - J_{II})} \] (3.27)

We see already that the maximum firing rate that can be achieved by the network decreases with increasing \( \Psi \). If the stimulus is widened or a second stimulus is added, the maximum firing rate will go down. However, the contrast at which that maximum occurs will also change. The previous expression for \( c_{\text{max}} \), in equation 3.25, is in terms of \( r_{E}(0) \). We can now take the expression for \( r_{E}(0)_{\text{max}} \) from 3.27, and plug it back into 3.25 to get an expression for the contrast at which the maximum firing rate will occur, in terms of model parameters.

\[ c_{\text{max}} = \frac{1}{k\Psi} \left( \frac{J_{EI}}{4(J_{EI} - J_{II})^{2}} - \frac{J_{EE}}{4\beta (J_{EI} - J_{II})} + \frac{1}{\sqrt{4\beta (J_{EI} - J_{II})}} \right) \] (3.28)

So we see that not only will the maximum firing rate decrease with increasing \( \Psi \), but furthermore, that the contrast at which the maximum will occur also decreases. When we add a second stimulus (increase \( \Psi \)), firing rates will both plateau and decrease sooner (i.e. at lower contrast) and have lower maximum values. These two effects combined effectively guarantee normalization in the high contrast regime.

From equation 3.22 we can find the firing rate at which the network will operate within the ISN regime:

\[ r_{E}(0) > \frac{1}{k(J_{EE}\Psi n)^{n}} \] (3.29)
3.3. Normalization

Assuming equality, we can plug this value of $r_E(0)$ back into equation 2.13 to solve for the value of $c$ at which the network enters the ISN regime:

$$c_{ISN} = \frac{2J_{EE}^2 J_{EI} - (J_{EI} - J_{II})(J_{EI} J_{IE} + J_{II} J_{EE}) - 2J_{EE} J_{EI} \sqrt{J_{EE}^2 - (J_{EE} + J_{IE})(J_{EI} - J_{II})}}{4k\Psi J_{EE}^2 (J_{EI} - J_{II})^2}$$

(3.30)

Like the contrast at maximum firing rate, the contrast at which the network enters the ISN regime also decreases with increasing $\Psi$.

3.3.5.3 Iterative approximation

By using implicit differentiation, we found an explicit expression for a few key points along the contrast response curve. However, a more general explicit expression remains elusive. Another approach we can take is to approximate an explicit expression for the firing rate by solving for $r_E$ of a particular order, and studying when the different order terms dominate the overall firing rate\(^1\). Our original expression for $r_E$ contains three separate orders: $r_E$, $r_E^{1/n}$, and $r_E^{1/n^2}$. At higher activity levels, the firing rate should be dominated by the highest order term, $r_E$. Solving for this term we have:

$$r_{EI} = \frac{1}{\Psi (J_{IE} J_{EI} - J_{EE} J_{II})} \left( J_{II} c_E - J_{EICI} - J_{II} \left( \frac{r_E J_{II}}{k} \right)^{\frac{1}{n}} + J_{EI} \left( \frac{J_{EE} r_E \Psi + c_E - \left( \frac{r_E}{k} \right)^{\frac{1}{n}}} {J_{EI} \Psi k} \right)^{\frac{1}{n}} \right)$$

(3.31)

As a first-order approximation, we get:

$$r_{E1} = \frac{1}{\Psi (J_{IE} J_{EI} - J_{EE} J_{II})} \left( J_{II} c_E - J_{EICI} + J_{EI} \left( \frac{c_E}{J_{EI} \Psi k} \right)^{\frac{1}{n}} \right)$$

(3.32)

This first-order approximation reveals that when the network is in the high firing rate

---

\(^1\)At this point, we will drop the $(0)$ from $r_E(0)$. All of our solutions are for the firing rates located at $\theta = 0$.\end{document}
3.3. Normalization

regime, the firing rate will scale as roughly \( \frac{1}{\Psi} \) (ignoring the \(( \frac{1}{\Psi} )^{\frac{1}{n}} \) inside the parentheses). In words, this tells us that in the high firing rate regime, firing rate will decrease with increasing overlap between the stimuli and the connectivity, whether it be by widening the single stimulus or adding a second stimulus. More precise approximations can be obtained iteratively by plugging \( r_{E1} \) back into \( r_{Ei} \) to get \( r_{E2} \), which we can again plug into \( r_{E1} \) to get \( r_{E3} \), and so on. This yields a closer and closer approximation of the actual solution we are interested in. Importantly, this solution is only in terms of the network parameters, and so is explicit for \( r_E \).

However, this iterative process fails for small values of \( r_E \) because the actual activity in the network is dominated by the low order terms (Supplementary Figure 7.7). Solving the full approximation now for \( r_E^{1/n^2} \) gives us:

\[
r_{Ei} = k \left( -c_E - J_{EE} r_{Ei} \Psi + J_{EI} k \Psi \left( \frac{J_{EIC1} - J_{II} c E + (J_{IE} J_{EI} - J_{EE} J_{II}) r_{Ei} \Psi + J_{ equal} \left( \frac{r_E}{k} \right)^{\Psi}}{J_{EI}} \right) \right)^n \]

(3.33)

Again, taking a first-order approximation:

\[
r_{E1} = k \left( -c_E + J_{EI} k \Psi \left( \frac{J_{EIC1} - J_{II} c E}{J_{EI}} \right)^n \right)^n \]

(3.34)

And here we see the opposite: when the activity in the network is dominated by the low-order terms in this description, the firing rate is expected to increase with \( \Psi \). This is precisely what we observe in the full model – an increase in firing rate with increasing stimulus width and supralinear addition of two stimuli at low contrast.

3.3.5.4 Contrast scaling and the balanced regime

The mechanism underlying normalization can be further elucidated by using this two-population approximation of the fixed-points to examine the effect of the expansive nonlinear-
ity on the scaling of responses to changes in stimulus width. Below, let $J = \begin{bmatrix} J_{EE} & -J_{EI} \\ J_{IE} & -J_{II} \end{bmatrix}$, $g = \begin{bmatrix} 1 \\ c_I/c_E \end{bmatrix}$, and $c = c_E$ (such that $g$ describes the relative strength of input to E and I cells and $c$ is the absolute magnitude of the input). The fixed point of the 2-D approximation can be written as:

$$r = k(\Psi Jr + cg)^n$$

(3.35)

To make this equation dimensionless, we define two additional variables: $y = \Psi r^k$ and $\alpha = kc^{n-1}\Psi$. Rewriting equation 3.35, we get

$$y = \alpha (Jy + g)^n$$

(3.36)

We can use this formulation to find the conditions under which firing rates will either increase or decrease with increases in stimulus width ($\Psi$):

$$\frac{dr}{d\Psi} = \frac{d\left(\frac{y}{\Psi}\right)}{d\Psi} = \frac{dy}{d\Psi} \frac{1}{\Psi} + \frac{dy}{d\Psi} \frac{c}{\Psi} + \frac{d(\Psi^{-1})}{d\Psi} c y = \frac{dy}{d\Psi} c - \frac{cy}{\Psi^2} = \frac{c}{\Psi} \left(\frac{dy}{d\Psi} - \frac{y}{\Psi}\right)$$

(3.37)

$$\frac{dy}{d\alpha} = \frac{dy}{d\alpha} \frac{d\alpha}{d\Psi} = \left(\frac{dy}{d\alpha} \frac{k c^{n-1}}{d\Psi}\right) = \alpha \frac{dy}{\Psi d\alpha}$$

(3.38)

Combine these two to get:

$$\frac{dr}{d\Psi} = \frac{c}{\Psi} \left(\frac{\alpha}{\Psi} \frac{dy}{d\alpha} - \frac{\alpha}{\Psi} \frac{y}{\alpha}\right) = \frac{c\alpha}{\Psi^2} \left(\frac{dy}{d\alpha} - \frac{y}{\alpha}\right)$$

(3.39)

And so we get that $\frac{dr}{d\Psi} > 0$ when $\frac{dy}{d\alpha} > \frac{y}{\alpha}$. This is equivalent to $\frac{d\ln y}{d\ln \alpha} > 1$. Intuitively, this is saying that an $\epsilon$ perturbation to $\ln \alpha$ will cause a perturbation to $\ln y > \ln \alpha$, so to
3.3. Normalization

first approximation \( \ln y \sim p \ln \alpha \) for \( p > 1 \) and thus:

\[
y \sim \alpha^p \quad \text{for} \quad p > 1
\]  

(3.40)

We can now use the same iterative approach described above (in terms of the dimensionless parameters as in equation 3.36) to find when \( y \) will scale as \( \alpha^p \) for \( p > 1 \). For low contrast \( (\alpha \ll 1) \):

\[
y = \alpha (Jy + g)^n
\]  

(3.41)

Expanding iteratively:

\[
y = \alpha \left( g + J\alpha \left( g + J\alpha (\ldots^n)_+ \right)_+ \right)_+^n
\]  

(3.42)

For low contrast, equation 3.42 converges, and we see that \( y \) scales with \( x_1\alpha^1 + x_2\alpha^2 + x_3\alpha^3 + \ldots \) (where \( x \) is some nonlinear combination of \( J \) and \( g \)). Thus the condition for \( \frac{\partial y}{\partial \Psi} > 0 \) is satisfied, indicating supralinear summation. At high contrast \( (\alpha \gg 1) \) however, this expansion explodes. This intuitively occurs in a regime in which excitation is unstable, i.e. in the ISN regime. If the dynamics are nonetheless stable, we need to re-examine:

\[
y = \alpha (Jy + g)^n
\]  

(3.43)

Regardless of the contrast regime, we know that if \( (Jy + g) \) has any component that scales as \( \alpha^p \) for \( p > -1/n \), then \( y \) will have to scale as \( \alpha^q \) for \( q = 1 + pn > 0 \). But then \( (Jy + g) \) has a component that goes as \( \alpha^q \), and so \( y \) will have arbitrarily high powers of \( \alpha \), as was the case in equation 3.42. We know that with \( \alpha \) large enough, equation 3.42 explodes\(^2\) because the coefficients \( J \) and \( g \) do not get small sufficiently quickly (they are both order 1). Since \( (Jy + g) \) already contains \( g \) and this goes as \( \alpha^0 \) and \( 0 > -1/n \), then we know

\(^2\)unless \( \det J = 0 \), but this solution requires fine tuning and so we discard this in search of a more general solution (and furthermore, our own simulation parameters do not meet this criterion, and yet we still observe sublinear addition at high contrast).
3.3. Normalization

this component must be canceled to leave only powers \( \leq -1/n \). So the leading order of \( y \) must go as \( \alpha^0 \) to cancel \( g \) and leave only powers of \( \alpha^{-1/n} \) and lower. If the leading order of \( y \) scales as \( \alpha^p \) for \( p = 0 \), from equations 3.39 and 3.40 we expect sublinear addition. If we say that \( y_0 \) is this leading order, then \( Jy_0 + g = 0 \), thus \( y_0 = -J^{-1}g \). This requirement is the same condition for stability derived by van Vreeswijk and Sompolinsky (1996; 1998) in their balanced model, and similarly we expect that our networks will dynamically reach this stable state whenever they respond stably in this high-contrast, high-input regime.

3.3.5.5 Stability analysis

In the previous sections of the analysis, we have shown that if the firing rates in the network don’t explode, then the network must have been dynamically stabilized, which implies normalization. But when will the network dynamically stabilize rather than explode? Empirically, we find that this essentially always happens so long as feedback inhibition (\( J_{IE} \) and \( J_{EI} \)) is sufficiently strong. But for the case \( n = 2 \), we can also analytically find a somewhat stronger condition that guarantees that any fixed point that exists is stable (though it doesn’t guarantee that a fixed point exists).

To begin, recall that previously we derived an expression for the Jacobian of this dynamical system (equation 3.4). For sake of clarity, let \( \tau_E = \tau_I \), \( n = 2 \), and let us write \( J_{xy} \) to mean \( J_{xy}\Psi \). Then, for this simple two-dimensional system, the linearized weight matrix at the fixed-point is given by:

\[
\hat{W} = \begin{pmatrix} 2 \left( \sqrt{kT_E} & 0 \\ 0 & \sqrt{kT_I} \right) & \begin{pmatrix} J_{EE} & -J_{EI} \\ J_{IE} & -J_{II} \end{pmatrix} - I \right)
\]

To simplify further, let us also use the following shorthand:

\[
\Phi \equiv \begin{pmatrix} \sqrt{kT_E} & 0 \\ 0 & \sqrt{kT_I} \end{pmatrix}
\]

(3.45)
and

\[ J \equiv \begin{pmatrix} J_{EE} & -J_{EI} \\ J_{IE} & -J_{II} \end{pmatrix} \] (3.46)

Such that we can also write \( \hat{W} = (2\Theta J - I) \). The linear system around the fixed-point will be stable if both of the eigenvalues of this matrix are negative. This is true iff the trace of \( \hat{W} < 0 \) and the determinant of \( \hat{W} > 0 \). For the requirement that the trace of \( \hat{W} < 0 \) we get that:

\[ 2J_{EE}\sqrt{k_{RE}} - 2J_{II}\sqrt{k_{RI}} < 2 \] (3.47)

which is equivalent to requiring that

\[ \text{tr} (2\Theta J) < 2 \] (3.48)

For the determinant requirement, note that for an arbitrary 2-dimensional system:

\[ \det (X - I) = \det (X) - \text{tr} (X) + 1 \] (3.49)

Letting \( X = (2\Theta J) \), we can see that the requirement \( \det (\hat{W}) > 0 \) will be fulfilled if:

\[ \det (2\Theta J) > 0 \] (3.50)

and

\[ \text{tr} (2\Theta J) < 1 \] (3.51)

Obviously, if inequality 3.51 is fulfilled, then the requirement that the trace of \( \hat{W} < 0 \) will be met, and so we can simply concern ourselves with these two inequalities (3.50 and 3.51). For the first requirement, we can use \( \det (2\Theta J) = \det (2\Phi) \det (J) \), and the knowledge that \( \det (2\Phi) > 0 \) (because firing rates and \( k \) are always positive and \( \Phi \) is diagonal), so that all we need to consider is the sign of \( \det (J) \).
3.3. Normalization

$$\det (\mathbf{J}) > 0 \Rightarrow J_{EE} < \frac{J_{IE} J_{EI}}{J_{II}}$$  \hspace{1cm} (3.52)

For the second requirement (inequality 3.51), we can use a slight modification to the trace requirement above:

$$2J_{EE}\sqrt{k r_E} - 2J_{II}\sqrt{k r_I} < 1$$  \hspace{1cm} (3.53)

And after rearranging:

$$J_{EE}\sqrt{r_E} - J_{II}\sqrt{r_I} < \frac{1}{2\sqrt{k}}$$  \hspace{1cm} (3.54)

Which will be guaranteed to be fulfilled if:

$$J_{EE}\sqrt{r_E} - J_{II}\sqrt{r_I} < 0$$  \hspace{1cm} (3.55)

To get inequality 3.55 in terms of the network parameters, we can use the fixed-point equation for $r_I$ to solve for $\sqrt{r_I}$ in terms of $\sqrt{r_E}$:

$$\sqrt{r_I} = \sqrt{k} \left( J_{IE} \sqrt{r_E}^2 - J_{II} \sqrt{r_I}^2 + c_I \right)$$  \hspace{1cm} (3.56)

$$\sqrt{r_I} = -1 \pm \frac{\sqrt{1 + 4c_I J_{II} k + 4J_{IE} J_{II} k \sqrt{r_E}^2}}{J_{II} \sqrt{k}}$$  \hspace{1cm} (3.57)

Because $r_I$ must be positive, we only need to consider the positive root. Plugging this into inequality 3.55 and rearranging, we get:

$$J_{EE}^2 - J_{IE} J_{II} < \frac{(1 + 4c_I J_{II} k)}{4r_E k}$$  \hspace{1cm} (3.58)

Since the right hand side of 3.58 is always positive, this bound will always be met as long as the left hand side is negative. Along with the requirement derived above from the determinant of $\mathbf{J}$, we have two inequalities that, when met, will ensure that the fixed-point
is stable:

\[ J_{EE}^2 < J_{IE} J_{II} \]  \hspace{1cm} (3.59)

\[ J_{EE} < \frac{J_{IE} J_{EI}}{J_{II}} \]  \hspace{1cm} (3.60)

These bounds are relatively simple to satisfy, requiring only sufficiently strong feedback inhibition \((J_{IE} \text{ and } J_{EI})\) to maintain stability. Though this analysis does not reveal whether or not a given parameter set will generate a system with a fixed-point, if a fixed-point exists and these bounds are met, stability will be guaranteed.

### 3.3.5.6 Winner-take-all responses

In the simulations, we saw that a network that sublinearly sums its inputs shows winner-take-all behavior for unequal strength stimuli. Here we show that (at least in the scalar case) sublinear addition always implies winner-take-all responses. A sublinear summation of responses can be expressed as

\[ r(a + b) = (r(a) + r(b))^n \]  \hspace{1cm} (3.61)

with \( n < 1 \). What we desire, then, are the relative weights on the responses to stimuli \( a \) and \( b \), or

\[ (r(a) + r(b))^n = w_1 r(a)^n + w_2 r(b)^n \]  \hspace{1cm} (3.62)

In the simulation results, \( r(a) \) and \( r(b) \) were \( N \) dimensional vectors, and so this problem had \( N \) equations and two unknowns. Thus we found the weights \( w_1 \) and \( w_2 \) using simply least-squares fitting. Here we are considering the scalar \( r \), and so have one equation with two unknowns. To address this problem, let us find a suitable constraint on \( w_1 \) and \( w_2 \). We know that when \( r(a) = r(b) \), we want that \( w_1 = w_2 = 2^{n-1} \), and that when \( r(b) = 0 \) that \( w_1 = 1 \) and \( w_2 = 0 \). Let us then constrain some \( L^p \) norm of the weight vector such that \((w_1^p + w_2^p)^{1/p} = 1\). The second condition satisfies the \( L^p \) norm constraint for any \( p \), so let us
use the first condition to solve for $p$:

$$(w_1^p + w_2^p)^{1/p} = 1$$

(3.63)

Raise both sides to the $p$ power, and solve for $p$:

$$w_1 = \sqrt[p]{1 - w_2^p}$$
$$2^{n-1} = \sqrt[p]{1 - 2^{p(n-1)}}$$
$$2^{p(n-1)} = 1 - 2^{p(n-1)}$$
$$2^{p(n-1)+1} = 1$$
$$p(n - 1) + 1 = 0$$
$$p = \frac{1}{1 - n}$$

(3.64)

For arbitrary responses to stimuli $a$ and $b$, $r(a)$ and $r(b)$, one solution to equation 3.62 is that $w_1$ and $w_2$ equal $\left(\frac{r(a)}{r(a) + r(b)}\right)^{1-n}$ and $\left(\frac{r(b)}{r(a) + r(b)}\right)^{1-n}$.

$$(r(a) + r(b))^n = \left(\frac{r(a)}{r(a) + r(b)}\right)^{1-n} r(a)^n + \left(\frac{r(b)}{r(a) + r(b)}\right)^{1-n} r(b)^n$$

(3.65)

$$(r(a) + r(b))^n = \left(\frac{1}{r(a) + r(b)}\right)^{1-n} r(a) + \left(\frac{1}{r(a) + r(b)}\right)^{1-n} r(b)$$

$$(r(a) + r(b))^n = \left(\frac{1}{r(a) + r(b)}\right)^{1-n} (r(a) + r(b))$$

$$r(a) + r(b) = r(a) + r(b)$$

(3.66)

And these two weights, $\left(\frac{r(a)}{r(a) + r(b)}\right)^{1-n}$ and $\left(\frac{r(b)}{r(a) + r(b)}\right)^{1-n}$, exactly fulfill the constraint $(w_1^p + w_2^p)^{1/p} = 1$ for $p = \frac{1}{1 - n}$. So as $r(a)$ grows relative to $r(b)$, their weights will go towards 1 and 0, yielding the winner-take-all response.
3.3. Normalization

3.3.6 Summary

We have described a circuit model that adds multiple inputs sublinearly, consistent with a number of experiments in V1 (MacEvoy et al., 2009; Busse et al., 2009). This operation, which has been described previously as normalization, occurs without the need for a pool of stimulus-nonspecific “normalizing” inhibitory interneurons, large shunting conductances, or an increase in inhibitory firing rates (Carandini and Heeger, 1994; Carandini et al., 1997). In fact, the entire network, both E and I cells, normalizes; this is a network level mechanism, and does require that some I cells do not normalize their responses in order to normalize others. The only necessary components are strong recurrent excitation balanced by inhibition and an accelerating neuronal input-output relationship, both of which are strongly supported by experiments (Ozeki et al., 2009; Nowak et al., 2010; Miller and Troyer, 2002). This finding is significant, because it demonstrates that normalization, at times referred to as “the canonical operation of cortex” (Heeger et al., 1996; Reynolds and Heeger, 2009), may be an intrinsic quality of cortical networks, rather than an operation that needs its own dedicated class of neurons.

Our model makes the prediction that for weak stimuli, cortex may add multiple stimuli supralinearly rather than sublinearly. Though this behavior has been observed in area MT (Heuer and Britten, 2002), this prediction has not, to our knowledge, been tested in V1. Busse et al. (2009) did test with relatively low contrast pairs (6%) without seeing supralinear addition, but this may not be low enough to enter the supralinear regime. Furthermore, because of the known role of feedforward input in processing cross-orientated stimuli (Kayser et al., 2001; Lauritzen et al., 2001; Carandini et al., 2002; Priebe and Ferster, 2006; Li et al., 2006a), the cross-orientation suppression paradigm may not be the best way to test this prediction. We would instead suggest a more direct experiment, for example using channelrhodopsin and overlapping patterns of photostimulation to directly activate populations of V1 neurons. By titrating the intensity of the photostimulus, one could use these direct cortical inputs to approximate weak and strong stimuli, and then probe systematically for a
transition between supralinear and sublinear addition.

3.4 Large scale probabilistically connected topographic model

3.4.1 Details of the large scale model

As a final proof of concept, we consider a large nonlinear network of E and I cells, in which each E/I pair is defined by both a two-dimensional position as well as a preferred orientation. The dynamics of this network are governed by the same nonlinear state equation as above (equations 3.2 and 3.3, with $W * r(x)$ redefined to mean $\sum_{x'} W(x, x')r(x')$). Connections are sparse and random. The probability of a connection from the cell of type $b$ at position $x'$ with preferred orientation $\theta'$ to the cell of type $a$ at position $x$ with preferred orientation $\theta$ decreases over distance in both the spatial and orientation domains:

$$p(W_{ab}|x - x', \theta - \theta') = \kappa_b e^{-\frac{(x-x')^2}{2\sigma_{ab}^2}} e^{-\frac{(\theta-\theta')^2}{2\sigma_{ori}^2}}$$ (3.67)

where $\kappa_b$ is used to set the overall sparseness of connections. These probabilities are used to stochastically generate connection matrices filled with only 1’s and 0’s. For each 1 in the weight matrix, a synaptic weight is chosen randomly from a Gaussian distribution with a mean equal to the magnitude parameter for the connection type (i.e. $J_{ab}$), and a standard deviation equal to 0.25 times the magnitude parameter. Weights below zero are rectified. To correct for the heterogeneity of orientation representation in the map and to maintain stability, synaptic weights onto each cell are normalized such that each cell receives the same total inhibitory and total excitatory synaptic weight from the network, and this total synaptic weight equals what the cells would have received if all connections of a given type were of uniform strength.

The cellular biophysics parameters ($\tau$, $n$, and $k$) are also noisily drawn from a Gaus-
The topographic map of preferred orientation. This map was generated by summing 30 complex plane waves equally spaced over 0 to $\pi$, with random phases and directions, and identical spatial frequencies. The phase of the complex number at each position in this $75 \times 75$ map is used to define the orientation preference of the cells at that position, where the phase angle corresponds to twice the preferred orientation. Kaschube et al. (2010) showed that this technique produces orientation maps with biologically realistic statistics.

The preferred orientation of each E/I pair is defined by a $75 \times 75$ orientation map generated using the model of Kaschube et al. (2010) (described in detail in the legend for Figure 3.11). The parameters used are: $N = 5625$, $\tau_E = 20$ ms, $\tau_I = 10$ ms, $\kappa_E = 0.1$, $\kappa_I = 0.5$ (such that local excitatory connection probability is 10% and local inhibitory connection probability is 50%), $J_{EE} = 0.033168$, $J_{IE} = 0.38831$, $J_{EI} = 0.09021$, $J_{II} = 0.077201$, $\sigma_{ori} = 45^\circ$, $k = 0.012$, $n = 2.2$, $\sigma_{EE} = 12.0$, $\sigma_{IE} = 12.0$, $\sigma_{EI} = 4.0$, and $\sigma_{II} = 4.0$. These distances are in grid spacings. To convert these to biological units, the full map is assumed to represent a cortical area corresponding to a $16^\circ \times 16^\circ$ patch of retinotopic space, which means that one grid spacing corresponds to $16/75$ degrees in retinotopic space. Assuming
a cortical magnification factor of 0.75 mm/deg, the standard deviation of the probability of the longest synaptic connections is: \(12 \times 16/75 \times 0.75 \approx 2\) mm cortex, such that 95% of connections will lie in a circle roughly 4 mm in radius, which is not unreasonable biologically. Boundaries in both orientation and retinotopic space are periodic.

![Stimulus Diameter vs Firing Rate](image1.png)

**FIG. 3.12:** Forty example length-tuning curves. We length-tuned 500 randomly selected E and I cells with a high-contrast (stimulus strength 40) grating. The orientation of the grating was set to be ideal for the randomly selected E-I cell pair. Shown on these two plots are 40 randomly selected length-tuning curves from the 500 cell data set, as well as the mean length-tuning curve from the full data set (inset).

### 3.4.2 Results

Using this more complex model, we were then able to study not only some idealized mean behavior of our circuits, but also the cell-to-cell variability of responses. Previous studies have
shown that in V1, not all cells experience the same amount of surround suppression; rather, there is a nearly uniform distribution of surround suppression strengths, with some cells being strongly suppressed and others not being surround suppressed at all (and some even surround facilitated) (Akasaki et al., 2002; Jones et al., 2000). To test for this property in our full model, we ran length-tuning tests on 500 randomly selected neurons. The orientation of the grating was set to be ideal for the randomly selected E-I cell pair. 40 randomly selected length-tuning curves for both E and I cells are shown in Figure 3.12. For each cell, we calculate a surround suppression index, which we define as: 
\[ SSI = \frac{r_{\text{max}} - r_{\text{full}}}{r_{\text{max}}} \]
where \( r_{\text{max}} \) is the maximum firing rate within the first two-thirds of the length-tuning curve, and \( r_{\text{full}} \) is the firing rate in response to the largest stimuli. As has been observed previously, there is a nearly uniform distribution of surround suppression indices when E and I cells are considered together (Figure 3.13A). When considered separately there is a clear distinction in the average strength of surround suppression between E and I, which might be interesting to test for experimentally. From the length-tuning curves, we also calculated distributions of summations field sizes (Figure 3.13B). We also studied the orientation tuning on the same set of E and I cells (Figure 3.13C). As has been observed experimentally, the orientation tuning of inhibitory cells is slightly but significantly wider than those of excitatory cells (Cardin et al., 2007).

We then tested this model on all the experimental paradigms explored above, and found that it is able to reproduce the previous analyzed behaviors. As before, strong surround suppression and spatially-periodic responses emerge only in the high contrast regime (Figure 3.14). When stimulated with full-field, sinusoidally contrast-modulated stimuli, the neurons show contrast-dependent tuning for contrast-modulation spatial frequency (Figure 3.15). As we observed in the linear spatial model, the relative phase of E and I cells in response to an input to inhibitory cells undergoes a 180° shift as the stimulus spatial frequency is raised above a critical value (Supplementary Figure 7.8). However, because this model has an expansive input-output nonlinearity, the paradoxical movement of excitatory and inhibitory
3.4. Large scale probabilistically connected topographic model

FIG. 3.13: 
Length- and orientation-tuning properties of the population. A. The distribution of surround suppression indices for E (red) and I (cells). For each of the 500 randomly selected cells, we calculated a surround suppression index using the maximum firing rate elicited by stimuli shorter than two-thirds of the largest stimulus shown (between 0 and 10.6 degrees) and the firing rate for the largest stimulus (16 degrees). The index is calculated as $SSI = \frac{r_{max} - r_{Full}}{r_{max}}$. Some cells that showed net facilitation have negative surround indices. B. A histogram of the summation field size for each cell. This is the stimulus size that elicited the maximum firing rate from the cell during the length-tuning test. The inhibitory cells clearly show a larger summation field size than the excitatory cells. C. The distribution of orientation tuning widths (measured as half-width at half-height). The mean tuning width for E and I cells is indicated on the plot, as well as the p-value of a two-sided t-test demonstrating that the two populations have significantly different means.

firing depends not only on driving the network at a spatial frequency below the critical frequency, but also on a baseline input strong enough to push the network into the effective ISN regime. As in the nonlinear ring model, the network demonstrates cross-orientation normalization at high contrast (Supplementary Figure 7.9). As was demonstrated in the one-dimensional nonlinear spatial model, a high contrast surround stimulus can facilitate a low contrast center stimulus but suppress a high contrast center stimulus (Supplementary Figure 7.10). Because we have included a more realistic degree of variability into our network, we observe more diverse behavior. Some neurons are facilitated by nearly all surround stimuli when the center is at low contrast, some are only facilitated by relatively small surrounds, and some are suppressed by all surround stimuli. The diversity of responses we observe here is representative of the assortment of responses observed in the literature and,
we would argue, at least partly explains some of the contention surrounding this particular experimental finding.

FIG. 3.14: Length-tuning for different levels of stimulus strength. Responses shown here are from 11 randomly selected E and I neurons. As with the 1-dimensional spatial model, periodic activity increased in magnitude with increasing stimulus strength.

Using this more detailed model, we can also explore the interactions between the spatial and orientation domains. For example, it has been shown that surround suppression in V1 is orientation dependent, such that a preferred center stimulus is maximally suppressed by
FIG. 3.15:  
Contrast dependent contrast modulation spatial frequency tuning. The network was stimulated by a full-field stimulus at the preferred orientation of the center E and I cells. The stimulus was overlaid with a sinusoidal contrast modulation, with a spatial frequency indicated on the x-axis of each plot. The maximum response of the E and I cells at the center of the network were recorded. A. The excitatory cell responses are plotted versus CM spatial frequency from yellow (for the weakest input) to red (for the strongest). Inhibitory responses are similarly denoted from cyan to blue. With increasing strength of the underlying luminance grating (here varied from 2 to 40), there is a steady rise in the preferred CM spatial frequency for both E and I cells. B. For both E and I, the preferred contrast modulation orientation was completely independent from CRF orientation tuning for 200 randomly selected neurons (E cells: \( r = -0.070, P = 0.489 \); I cells, \( r = 0.055, P = 0.584 \)). Shown here are histograms of the difference between preferred luminance and contrast modulation orientation.

an iso-oriented surround and minimally suppressed by an orthogonal surround (Cavanaugh et al., 2002b; Levitt and Lund, 1997). Our model is able to reproduce this effect (Figure 3.16A). Furthermore, it has been shown that when the center stimulus is at low contrast, the orientation specificity of surround suppression weakens. When we stimulate the center in our model with a much weaker stimulus than the surround, we similarly observe a broadening of the orientation tuning curves for surround suppression (Figure 3.16B).
3.4. Large scale probabilistically connected topographic model

FIG. 3.16: The orientation dependence of surround suppression. A. With the center stimulus at the center cells’ preferred orientation, the surround stimulus orientation was rotated in a full circle. The response of the center cell, normalized to the response to a center stimulus alone, is plotted versus the orientation difference between the center and surround. The response to the center stimulus alone is indicated by the dashed line at 1. Results here are the average of 50 randomly selected neurons, with the mean plotted on the thick line and standard deviation around the mean in the shaded region. B. With the center stimulus at much lower contrast than the surround, the orientation specificity of surround suppression decreases. Here we plot a histogram of the circular variances of the orientation tuning curves of surround suppression for 100 randomly selected E neurons and 100 randomly selected I neurons. The mean circular variance increases significantly ($p < 0.0001$) at low contrast, indicating that the tuning curves are less orientation specific at low contrast. When we split the total population into E and I cells, we find that only the E cells have a significant change in circular variance.

Primary visual cortex is not the only part of visual cortex in which cells experience surround suppression. In area MT, a cortical region involved in processing of visual motion,
there is also significant surround suppression (Allman et al., 1985; Born and Tootell, 1992; Bradley and Andersen, 1998; Britten and Heuer, 1999). As in V1, the strength of surround suppression decreases with decreasing stimulus contrast (Pack et al., 2005; Hunter and Born, 2011). To determine if a decrease in surround suppression was a general response of cortex to weak stimuli, these investigators tested the strength of surround suppression at different levels of stimulus coherence (a measure of what percentage of random dots are moving in a specified direction). Interestingly, they found that while suppression strength decreased with decreasing contrast, it actually increased with decreasing coherence (Hunter and Born, 2011). In our nonlinear ring model, we claim that normalization occurs as the result of an effective “width-tuning” in orientation space; making a stimulus wider in orientation space, like making it wider in real space, induces more surround suppression. While there is no easy way to manipulate stimulus width in orientation space in V1, these investigators had found a way to perform an analogous manipulation in MT, by parametrically varying stimulus coherence. By adding a direction-untuned component to the input while decreasing the direction tuned component, decreasing coherence effectively makes the stimulus wider in direction space, so we would predict that such a manipulation in our model should lead to increased surround suppression. To determine if our model would replicate these experimental findings, we adopted our model to area MT, and repeated the experiments of Hunter and Born. We are able to reproduce their results, finding that surround suppression strength decreases with decreasing stimulus contrast but increases with decreasing stimulus coherence (Figure 3.17).
3.4. Large scale probabilistically connected topographic model

Surround suppression versus stimulus strength and coherence. To model the recent experimental results of Hunter and Born (2011), who found that in area MT surround suppression decreased with decreasing stimulus contrast but increased with decreasing stimulus coherence, we performed length-tuning tests on the network while varying either stimulus strength or stimulus coherence. To lower stimulus coherence, we raised the baseline magnitude of the input to all orientations while simultaneously lowering the amplitude of the orientation-tuned component of the input, such that an input with 0% coherence was just an equal input to all orientations. This kept the maximum input constant while adjusting the level of the baseline input. For both the stimulus strength and stimulus coherence tests, we used the direction-tuning of MT cells, which have been reported to have a tuning width (full-width at half-height) of 83° on a 360° ring of directions (Albright, 1984). For a Gaussian stimulus in our model this gives \( \sigma_{FF} = 17.623^\circ \). The plots show the strength of surround suppression versus stimulus strength on the left and bottom axes, and surround suppression versus stimulus coherence on the right and top axes. Excitatory cells are shown in the top plot, and inhibitory cells in the bottom.
3.5  **Reduction of variability at stimulus onset**

Across many cortical regions, stimulus onset is accompanied by a marked reduction in the trial-to-trial variability of neural responses (Churchland et al., 2006; 2010). This reduction correlates with behavioral performance and is believed to play a part in cortical computation. The underlying circuit mechanism is unknown. We hypothesize that a circuit that can perform a “winner-take-all” operation on its inputs should be able to suppress the relative contribution of noise to the population response in the presence of a strong stimulus, and this suppression may be sufficient to reduce the trial-to-trial variability.

To test this hypothesis, we modified our nonlinear ring model to include one or more sources of variability, and studied the effects of stimulus onset in the presence of an ongoing noisy input. We built three different modified forms of the nonlinear ring model, with increasingly detailed noise-generating mechanisms, to explore this question. All three will be briefly presented here; the figures in this chapter were generated using the second form of the model, which is of intermediate complexity. The figures from the other two models (which are qualitatively similar) are presented in the Supplementary Figures.

### 3.5.1 Model details

In the first, and most basic modification, we simply added a noisy input $\eta$ into each neuron, such that the state equations of the network are now

\[
\tau_E \frac{d}{dt} r_E(\theta) = -r_E(\theta) + k \left( [W_{EE} \ast r_E(\theta) - W_{EI} \ast r_I(\theta) + c_E h(\theta) + \eta]_+ \right)^n \tag{3.68}
\]

\[
\tau_I \frac{d}{dt} r_I(\theta) = -r_I(\theta) + k \left( [W_{IE} \ast r_E(\theta) - W_{II} \ast r_I(\theta) + c_I h(\theta) + \eta]_+ \right)^n \tag{3.69}
\]
3.5. Reduction of variability at stimulus onset

The noise $\eta$ injected into each neuron is generated as follows. For both the E and I populations, we take an $N \times T$ matrix (where $N$ is the number of cells in the population and $T$ is the duration of the simulation in time steps) of random values drawn from a Gaussian distribution with a pre-specified mean and standard deviation. This white noise is then smoothed with Gaussian filters first over time for each neuron and then over space at each time step. For this first model, we use a temporal filter with a width of 25 ms and a spatial filter with a width of $1^\circ$. Both filters are normalized to have an integral of 1. All of the network parameters are the same as in the non-noisy ring model presented earlier. The white noise was generated from a Gaussian distribution with $\mu_\eta = 10$ and $\sigma_\eta = 40$. After being passed through both the temporal and spatial filters, the standard deviation of the noise ultimately injected in the network was approximately 2.5.

In the second noisy model, we modified our basic ring model even further, to account not only for the noisy input, but also for the variability inherent in the “spiking” of each neuron. We replaced each term in the input representing a firing rate by a normally distributed random process with a mean and variance given by what had previously been the firing rate. At each time step, this random process is used to produce vectors of excitatory and inhibitory firing rates, which are then used to calculate the input to every cell in the network. All of the parameters in this model are otherwise unchanged.

$$
\tau_E \frac{d}{dt} r_E(\theta) = -r_E(\theta) + k \left( [W_{EE} * \mathcal{N}(r_E(\theta), r_E(\theta)) - W_{EI} * \mathcal{N}(r_I(\theta), r_I(\theta))] + \mathcal{N}(c_E h(\theta), c_E h(\theta)) + \mathcal{N}(\mu_\eta, \sigma_\eta^2) ]_+ \right)^n
$$

$$
\tau_I \frac{d}{dt} r_I(\theta) = -r_I(\theta) + k \left( [W_{IE} * \mathcal{N}(r_I(\theta), r_I(\theta)) - W_{II} * \mathcal{N}(r_I(\theta), r_I(\theta))] + \mathcal{N}(c_I h(\theta), c_I h(\theta)) + \mathcal{N}(\mu_\eta, \sigma_\eta^2) ]_+ \right)^n
$$

where $\mathcal{N}(\mu, \sigma^2)$ is a random variable drawn from a normal distribution with mean $\mu$ and variance $\sigma^2$. 

As a last test, we implemented a markedly more complex model, derived from the mean-field solution to the diffusion approximation for a population of constant-leak linear integrate and fire neurons (Fusi and Mattia, 1999). This model allows us to generate realistic variability in the absence of any external noise, as would be observed in a population of integrate and fire neurons.

The model is based on the assumption that at each grid point, the E unit and I unit each represent a homogenous pool of \( N_{IF} \) integrate and fire neurons with identical statistics. Within a node, the population of neurons receives synaptic input from both itself and other nodes in the network. This input has some mean, \( \mu \) and variance \( \sigma^2 \) across the population. We model three different types of synaptic input, AMPA, NMDA, and GABA with the following dynamical equations:

\[
\tau_{AMPA} \frac{d}{dt} \mu_{EAMPA}(\theta) = -\mu_{EAMPA}(\theta) + W_{EEAMPA} r_{E}(\theta) N_{IF} + \mu_{EXTAMPA}(\theta)
\]

\[
\tau_{AMPA} \frac{d}{dt} \mu_{IAMPA}(\theta) = -\mu_{IAMPA}(\theta) + W_{IEAMPA} r_{E}(\theta) N_{IF} + \mu_{EXTAMPA}(\theta)
\]

\[
\left(\frac{\tau_{AMPA}}{2}\right) \frac{d}{dt} \sigma_{EAMPA}^2(\theta) = -\sigma_{EAMPA}^2(\theta) + W_{EEAMPA}^2 r_{E}(\theta) N_{IF} + \sigma_{EXTAMPA}^2(\theta)
\]

\[
\left(\frac{\tau_{AMPA}}{2}\right) \frac{d}{dt} \sigma_{IAMPA}^2(\theta) = -\sigma_{IAMPA}^2(\theta) + W_{IEAMPA}^2 r_{E}(\theta) N_{IF} + \sigma_{EXTAMPA}^2(\theta)
\]

\[
\tau_{NMDA} \frac{d}{dt} \mu_{ENMDA}(\theta) = -\mu_{ENMDA}(\theta) + W_{EENMDA} r_{E}(\theta) N_{IF} + \mu_{EXTNMDA}(\theta)
\]

\[
\tau_{NMDA} \frac{d}{dt} \mu_{INMDA}(\theta) = -\mu_{INMDA}(\theta) + W_{IENMDA} r_{E}(\theta) N_{IF} + \mu_{EXTNMDA}(\theta)
\]

\[
\left(\frac{\tau_{NMDA}}{2}\right) \frac{d}{dt} \sigma_{ENMDA}^2(\theta) = -\sigma_{ENMDA}^2(\theta) + W_{EENMDA}^2 r_{E}(\theta) N_{IF} + \sigma_{EXTNMDA}^2(\theta)
\]

\[
\left(\frac{\tau_{NMDA}}{2}\right) \frac{d}{dt} \sigma_{INMDA}^2(\theta) = -\sigma_{INMDA}^2(\theta) + W_{IENMDA}^2 r_{E}(\theta) N_{IF} + \sigma_{EXTNMDA}^2(\theta)
\]

\[
\tau_{GABA} \frac{d}{dt} \mu_{EGABA}(\theta) = -\mu_{EGABA}(\theta) + W_{IEGABA} r_{I}(\theta) N_{IF}
\]

\[
\tau_{GABA} \frac{d}{dt} \mu_{IGABA}(\theta) = -\mu_{IGABA}(\theta) + W_{IIGABA} r_{I}(\theta) N_{IF}
\]
3.5. Reduction of variability at stimulus onset

\[
\left(\frac{\tau_{\text{GABA}}}{2}\right) \frac{d}{dt} \sigma^2_{\text{GABA}}(\theta) = -\sigma^2_{\text{GABA}}(\theta) + W^2_{\text{IE}} r_I(\theta) N_{IF} \tag{3.77}
\]

\[
\left(\frac{\tau_{\text{GABA}}}{2}\right) \frac{d}{dt} \sigma^2_{\text{I}}(\theta) = -\sigma^2_{\text{I}}(\theta) + W^2_{\text{II}} r_I(\theta) N_{IF}
\]

At each time step, we compute the mean and the variance of the input to each node as:

\[
\mu_E(\theta) = \mu_{E_{\text{AMPA}}} + \mu_{E_{\text{NMDA}}} - \mu_{E_{\text{GABA}}}
\]

\[
\mu_I(\theta) = \mu_{I_{\text{AMPA}}} + \mu_{I_{\text{NMDA}}} - \mu_{I_{\text{GABA}}}
\]

\[
\sigma^2_E(\theta) = \sigma^2_{E_{\text{AMPA}}} + \sigma^2_{E_{\text{NMDA}}} + \sigma^2_{E_{\text{GABA}}}
\]

\[
\sigma^2_I(\theta) = \sigma^2_{I_{\text{AMPA}}} + \sigma^2_{I_{\text{NMDA}}} + \sigma^2_{I_{\text{GABA}}}
\]

For each population, the mean and average input across the network is then used to calculate the mean firing rate:

\[
\nu_x(\theta) = \left[\left(\frac{\sigma^2_x(\theta)}{2\mu_x^2(\theta)}\right) \left(e^{-\frac{2\mu_x(\theta)\Theta}{\sigma_x^2(\theta)}} - e^{-\frac{2\mu_x(\theta)H}{\sigma_x^2(\theta)}}\right) + \frac{\Theta - H}{\mu_x(\theta)}\right]^{-1} + \tau_{\text{ref}}
\]

This equation gives the firing rate of an integrate and fire neuron with noisy input of mean \(\mu\) and variance \(\sigma^2\), spike threshold \(\Theta\), reset potential \(H\), and refractory period \(\tau_{\text{ref}}\); the \(x\) subscript denotes E or I. To account for the finite size of individual neuronal pools, the actual firing rates used in the following time step are drawn randomly from a Gaussian distribution with mean and variance equal to the calculated mean rate (Renart et al., 2004):

\[
r(\theta) = \mathcal{N}(\nu(\theta), \nu(\theta)) \tag{3.81}
\]

In this model, we used the following parameters: \(N = 180\), \(\Delta \theta = 1^\circ\), \(\tau_{\text{AMPA}} = 5\text{ms}\), \(\tau_{\text{GABA}} = 5\text{ms}\), \(\tau_{\text{NMDA}} = 250\text{ms}\), \(J_{E_{\text{AMPA}}} = 0.0003583125\), \(J_{I_{\text{AMPA}}} = 0.00135135\), \(J_{E_{\text{NMDA}}} = 0.0003583125\), \(J_{I_{\text{NMDA}}} = 0.00135135\), \(J_{E_{\text{GABA}}} = 0.000675675\), \(J_{I_{\text{GABA}}} = 0.00118755\), \(N_{IF} = 1000\), \(\sigma_{\text{ori}} = 45^\circ\), and \(\sigma_{\text{FF}} = 30^\circ\). The feedforward input \((\mu_{\text{EXT}_{\text{AMPA}}}, \mu_{\text{EXT}_{\text{NMDA}}}, \sigma^2_{\text{EXT}_{\text{AMPA}}}, \sigma^2_{\text{EXT}_{\text{NMDA}}})\) was determined by the strength of the input, and was made up of equal parts AMPA and NMDA currents, such that for stimulus strength \(c\),
3.5. Reduction of variability at stimulus onset

The top two plots show the responses of the E and I cells centered at the stimulus orientation from 1,000 trials. The stimulus is turned on at 2,000 ms. The mean at each time step is plotted as the solid red and blue lines, and the standard deviation about the mean is shown in the shaded red and blue areas. The bottom plots show the Fano factor \( \frac{\sigma^2}{\mu} \) calculated at each time step.

\[
\mu_{\text{EXTAMP}} = \mu_{\text{EXTNMDA}} = \sigma_{\text{EXTAMP}}^2 = \sigma_{\text{EXTNMDA}}^2 = c/2.
\]

For the input-output function (Equation 3.80) we used \( \Theta = 1, H = 0, \) and \( \tau_{\text{ref}} = 0. \) For filtering the noisy input, we use a temporal filter with a width of 10 ms and a spatial filter with a width of 5°.

### 3.5.2 Results

To test the effect of stimulus onset on trial-to-trial variability in these networks, we ran 1,000 simulations of a simple stimulus presentation trial. On each trial, we begin by injecting the noisy input \( \eta \) alone for 2,000 milliseconds. We then turn on an oriented stimulus with strength 50, and continue the simulation for an additional 3,000 ms. On each trial, we
3.5. Reduction of variability at stimulus onset

record the firing rate of each neuron in the network over time, and use this to calculate the mean and the variance across trials for each cell at every time step. From these measures, we then calculate the Fano factor of each cell as function of time.

\[
F(t) = \frac{\sigma^2(t)}{\mu(t)} \tag{3.82}
\]

The Fano factor is a commonly used metric of variability, and the one used by Churchland et al. (2010) to quantify the reduction in variability they observe.

In the top row of Figure 3.18, we show the mean and standard deviation of the firing rate for both the excitatory and inhibitory neuron located at the center of the stimulus, and on the plot below each, the Fano factor over time. For both cells, the Fano factor clearly decreases at stimulus onset. It is possible, of course, that this effect is due solely to the fact that the mean firing rate of these neurons is increased by the stimulus. However, when we look at other neurons on the ring, we see the same effect. Even for neurons that are suppressed by the stimulus, and thus show a decrease in mean firing rate, there is a decrease in the Fano factor at stimulus onset (Figure 3.19). This matches exactly the experimental findings.

From recordings with multi-electrode arrays, it has been shown the stimulus evoked reduction in variability results primarily from a reduction in the “shared noise” in the network. The “private noise”, that is, the variability intrinsic in individual neurons, is relatively unaffected by stimulus onset. In our model, we can manipulate the degree to which the noisy input is shared in the network by varying the width of the Gaussian spatial noise filter. When we repeat our basic stimulus presentation paradigm over a range of filter widths, we observe a dependence in the magnitude of variability reduction on the spatial extent of the correlations introduced by the spatial filter (Figure 3.20A). This was true again for both the neurons located at the stimulus center as well as for those further away on the ring.

In the work of Churchland et al. (2010), the reduction in shared noise was demonstrated by using factor analysis to separate the observed variability into two components, a “shared”
variability and a “private” variability. Factor analysis is a dimensionality reduction technique used to explain the source of variance within a data set. Essentially, after pre-specifying a number of common factors, it fits the model $\text{Cov}(X) = \Lambda\Lambda^T + \Psi$, where $\Lambda$ is an $n \times k$ factor loading matrix and $\Psi$ is a diagonal matrix of the private variances. Factor analysis is related to principal component analysis (PCA), but unlike PCA, which finds the directions of maximum variance within a system, factor analysis instead finds the directions of maximum correlation (Roweis and Ghahramani, 1999).

To compare our results with the experimental data, we performed factor analysis on a subset of the neurons in our network before and after the stimulus onset. Like Churchland et al. (2010), we specifically chose a “mean-matched” sample of neurons to study, in order
The decrease in trial-to-trial variability depends on spatial correlations in the noisy input. A. The same experiments as above are now repeated for 21 different values of the spatial filter width (0° to 2° in 0.1° steps). The Fano Factor from the pre-stimulus and post-stimulus epochs are averaged over time, and the percent change is plotted as a function of spatial filter width. For both the cells located at the preferred (left y-axis) and non-preferred (right y-axis) locations, the effect is same: with little or no spatial correlation in the input, the reduction in variability is small. With increasing spatial correlation in the input, the drop in variability grows much larger. B. Factor analysis reveals a reduction in shared noise. As in Churchland et al. (2010), we performed our factor on a mean-matched sample of cells. In terms of our model, we selected the 28 cells whose mean firing rate changed the least (less than 5) in the pre- and post-stimulus epochs. We then time-averaged the firing rates of each of these cells over equal-sized time segments from the pre- and post-stimulus epochs for each of the 1,000 trials. Using these two 28 by 1000 matrices, we performed factor analysis (assuming 4 common factors) to split the total variance from each epoch into “shared” and “private” components.
leak linear integrate and fire neurons reproduces all of the previously observed behaviors (Figure 3.21, Supplementary Figures 7.14 and 7.15). The only major difference in this model is that because the noise is predominantly generated intrinsically, rather than being externally added, the model no longer relies on spatial correlations in the input noise for the reduction of trial-to-trial variability (Supplementary Figure 7.16). With the spatial filter set to zero, there is still a strong reduction in variability. In fact, there’s no need for externally added noise at all (Supplementary Figure 7.17). Even with just a constant input to all of the cells the network, the network still undergoes a reduction in variability at stimulus onset.

3.5.3 Summary

Here we have demonstrated that the same circuit model that intrinsically normalizes its inputs also produces a reduction in trial-to-trial variability at stimulus onset. This reduction occurs because of the transition from sublinear addition to winner-take-all response summation with increasing disparity in input strength. The strong stimulus suppresses the contribution of the underlying noise process to the neuron’s response, reducing the trial-to-trial variability. As in experiments, this reduction in variability occurs both in neurons that are activated and those that are not activated or are suppressed by the stimulus. The suppressed variability is that “shared” between the neurons – each neuron’s “private” variability is not suppressed. This points to a potential computational function – to enhance signal to noise ratio by suppressing noise from inputs.

3.6 Attentional modulation

3.6.1 Introduction and motivation

Attention has been shown to have a powerful modulatory effect on both task performance and neuronal response, and changes in the latter can often be powerful predictors of the former. For example, when a preferred and non-preferred stimulus are both presented to
FIG. 3.21: The reduction in variability in a much more complex model of cortex. As in Figures 3.18 and 3.19, stimulus onset causes a reduction in trial-to-trial variability in all the neurons in the network, regardless of whether they are excited, suppressed, or ambivalent to the stimulus presented.
3.6. Attentional modulation

the receptive field of a V4 neuron, the cell’s response is roughly the average of the responses evoked by each stimulus alone. By attending to either the preferred or non-preferred stimulus, the response will be shifted towards the response evoked by the attended stimulus alone (Reynolds and Desimone, 2003). Similarly, attention to a stimulus in the suppressive surround of a V4 neuron increases the suppression induced, whereas attention to the center reduces the suppression (Sundberg et al., 2009). Attention causes a similar percentage increase in the firing rates of excitatory and inhibitory neurons (Mitchell et al., 2007), and also causes a decrease in neuronal variability and correlation (Cohen and Maunsell, 2009; Mitchell et al., 2007).

In previous work, it was shown that non-normal networks (such as we model) with strong recurrent excitation and inhibition exhibit “balanced amplification” (Murphy and Miller, 2009); small inputs biased toward either excitatory (or inhibitory) cells drive large increases (or decreases) in both excitatory and inhibitory firing rates. The surround suppression we have been studying can be understood as a contextually-driven decrease in balanced amplification (Ozeki et al., 2009). Input from the surround is slightly biased towards inhibitory cells, which lowers the gain, and consequently leads to suppression, of both the excitatory and inhibitory cells. As much work on attention has concluded that the chief effect of attention on the local cortical circuit is an increase in neuronal gain (Mitchell et al., 2007; Reynolds and Heeger, 2009), we wondered if perhaps attentional modulation could be taking advantage of the same balanced amplification mechanism. Instead of biasing input towards inhibitory cells to get suppression, perhaps there is a bias towards excitatory cells that causes a global boost in the gain of both E and I cells.

To explore this mechanism, we used our various models of visual cortex to replicate some well known models and experiments on visual attention. In all cases, attention was modeled as simply an additional excitatory input directed preferentially towards the excitatory neurons within a spatiotopically defined locus of attention. We imagine this excitatory input represents some sort of top-down influence on the cortical circuit, either through fast-acting
ionotropic transmitters or some slower metabotropic neuromodulatory input.

### 3.6.2 Results

As a first test, we attempted to reproduce the modeling results of Reynolds and Heeger (2009), who found in their “normalization model of attention” that the expected effect of attention is strongly influenced by the relative sizes of the stimulus and the attention field. Specifically, they predicted that when attention is directed to a relatively large area, the effect on the response to a small stimulus should be predominantly a change in “contrast-gain”, such that cells respond to stimuli as if they were effectively at higher contrast. This would be appreciated as a leftward shift in a contrast-response curve for a stimulus, with relatively little change in the maximum firing rate. For a large stimulus and a small attentional field, they instead predict a change in “response-gain”, such that all responses are boosted multiplicatively.

Here we use the spatial, nonlinear ISN model to study the two different effects of attention described by Reynolds and Heeger. Attention is modeled as a small additional input only to excitatory cells over a defined spatial area, and we calculate contrast response curves with and without attention. To quantify changes in the contrast response properties, we fit each curve to a standard Naka-Rushton equation (Naka and Rushton, 1966):

$$R(c) = R_{\max} \left( \frac{c^n}{c_{50}^n + c^n} \right)$$

(3.83)

where $R_{\max}$ is the plateau firing rate, $n$ describes the steepness of the contrast response curve, and $c_{50}$ is the strength of the stimulus at which the response is 50% of its maximum.

With a large attentional field and small stimulus, the effect of attention is predominantly a leftward shift in the contrast-response function, as predicted by Reynolds and Heeger (2009). We quantify this change in “contrast gain” as the difference in the $c_{50}$ parameters of the contrast response curves produced with and without attention (Figure 3.22A).
The qualitative effect of attention depends on the relative sizes of the attentional and stimulus fields. Here we use the spatial, nonlinear ISN model to study the two different effects of attention, as described by Reynolds and Heeger (2009). Attention is modeled as a small additional input (strength 4) only to excitatory cells over a defined spatial area, and we calculate contrast response curves with (red curves) and without (cyan curves) attention. **A.** With a large attention field and small stimulus, the contrast gain increases by 8.37 units of stimulus strength when attending, whereas the response gain increases by only a factor 1.05. **B.** In the “small attention field large stimulus” condition, we see the opposite. There is almost no change in contrast gain (an decrease of 0.47 units of stimulus strength), but a 3.02-fold increase in response gain when attending.

We compare this to the “response gain”, which we quantify as the ratio of $R_{\text{max}}$ parameters with and without attention. With a large stimulus and small attention field, the effect of attention is reversed: there is virtually no change in the contrast gain, and a much larger change in the response gain (Figure 3.22B).

We next reproduce some well known experimental results on attention. To begin, we modeled the experiment of Reynolds and Desimone (2003), who probed the responses of V4 neurons with preferred and non-preferred stimuli, presented either alone or together to the receptive field of a single neuron. They found that in the simultaneous presentation
3.6. Attentional modulation

condition, attending to a non-preferred stimulus caused a relative suppression compared to an attend-away condition, whereas attending to the preferred stimulus boosted the response. To simulate this experiment (Figure 3.23A), we recorded the response of a cell to a strong stimulus of preferred orientation in the nonlinear ring model. We then add a non-preferred stimulus at the orthogonal orientation to the ring and varied the strength of this probe stimulus from 0 to 80. The addition of the non-preferred probe is always suppressive (especially at high contrast), and with increasing probe strength suppression is increased. We then repeat the same test with attention (as an extra input to excitatory cells of strength 5) directed either towards the preferred stimulus or the probe stimulus. When attention is directed towards the preferred stimulus, the amount of suppression is decreased. When attention is directed to the probe stimulus, suppression is enhanced.

In a related experiment, Treue and Martinez-Trujillo (1999) recorded from a neuron in area MT while presenting two stimuli to the neuron’s receptive field. One of the stimuli was always moving in a non-preferred direction, while the direction of the other stimulus was systematically varied. Compared to an attend-away condition, responses of MT neurons were relatively suppressed at all stimulus directions when attention was directed towards the non-preferred stimulus, but relatively enhanced when attending towards the varying stimulus. We find the same result if we repeat this test in our nonlinear ring model (Figure 3.23B). Like Treue and Martinez-Trujillo (1999), the change we observe occurs without a substantial change in the width of orientation tuning, indicating a mostly multiplicative scaling.

In another experiment on attention in area MT, Martinez-Trujillo and Treue (2002) again recorded from neurons while presenting two stimuli within the receptive field. One stimulus was moving in a preferred direction, and the other in a non-preferred direction. They then varied the contrast of the preferred stimulus while holding the contrast of the non-preferred stimulus fixed, and directed a monkey to attend either to the non-preferred stimulus or outside of the receptive field. They found that attending to the non-preferred stimulus caused predominantly a change in contrast-gain. Reynolds and Heeger were able to
3.6. Attentional modulation

FIG. 3.23: **Modeling attention in the nonlinear ring model.**  
A. Attention enhances the suppressive effect of non-preferred stimuli in V4. A strong stimulus (strength 40) of preferred orientation is shown to a cell in the nonlinear ring model. We then add a non-preferred stimulus with orthogonal orientation to the ring and vary the strength of this probe from 0 to 80, and record the firing rate. The addition of the non-preferred probe is suppressive, and with increasing probe contrast the suppression is increased. We then repeat the same test, with attention (as an extra input to excitatory cells of strength 5) directed either towards the preferred stimulus or the probe stimulus. When attention is directed towards the preferred stimulus, suppression is decreased. When attention is directed to the probe stimulus, suppression is enhanced.  
B. In the presence of a non-preferred probe stimulus (orientation 135° and strength 40), we vary the orientation of a test stimulus between 0° and 180°, while recording from the cell at 45° and attending either to the non-preferred probe (cyan), the varying stimulus (red), or away (orange). Attention produces an almost exclusively multiplicative change in response. Normalized responses are shown in the inset. There is virtually no change in tuning width, as observed experimentally (Treue and Martinez Trujillo, 1999).  
C. In the presence of a fixed-strength non-preferred stimulus, the contrast of a preferred stimulus is systematically varied while attention is directed either away (cyan) or towards the non-preferred stimulus (red). Attention to the non-preferred stimulus produces mainly a reduction in contrast gain (Martinez-Trujillo and Treue, 2002)  
D. Showing preferred and non-preferred stimuli of equal but varying contrast while attending to one or the other produces a much larger change in response gain, as in (Reynolds and Heeger, 2009) The response gain and contrast gain values indicated correspond to attending to the non-preferred probe relative to the preferred stimulus.
reproduce this effect in their model (2009), and furthermore predicted that if the contrast of both the preferred and non-preferred stimulus were varied together, attending to one or the other stimulus would produce a much larger change in response gain. Using the nonlinear ring model again, we modeled both of these stimulus conditions, and find analogous results (Figure 3.23C, D).

In yet another experiment, Sundberg et al. (2009) found that in V4, the strength of surround suppression could be either enhanced or decreased by attending specifically to the surround or center stimulus. For this simulation (Figure 3.24), we used the large topographic model, and randomly selected 25 E and I cells. For each cell, we measured the response to a strong stimulus of preferred orientation centered on the receptive field, and then added a strong surround stimulus of the same orientation to the surround. The response to the cell was measured again in the absence of an attentional input (the “Attend Away” condition), as well as with an attentional input directed towards the center or surround stimulus. As was observed experimentally, attending to the surround boosted the amount of surround suppression, whereas attending to the center greatly weakened the surround suppression (and in some cases gave a response greater than the unattended center alone).

3.6.3 Summary

With a simple addition to our nonlinear circuit models, we are able to reproduce a number of experimental results on attention in visual cortex (Martinez-Trujillo and Treue, 2002; Treue and Martinez Trujillo, 1999; Sundberg et al., 2009; Reynolds and Desimone, 2003). Through balanced amplification (Murphy and Miller, 2009), a small additional excitatory input to excitatory cells causes a nonlinear scaling of firing rates in a manner consistent with a number of experimental observations. Importantly, these simple models are able to account for the two different forms of gain changes reported in the literature, contrast-gain and response-gain (Martinez-Trujillo and Treue, 2002; Reynolds and Heeger, 2009). This model provides a simple, plausible mechanism through which higher cortical feedback can
3.7 Discussion/conclusions

We have presented here a simple, unified model of V1 that is able to reproduce a number of classical and extra-classical receptive field properties, including length-tuning curves with multiple peaks (Anderson et al., 2001), contrast-modulation spatial frequency tuning (Tanaka and Ohzawa, 2009), cross-orientation normalization (MacEvoy et al., 2009; Busse et al., 2009), and a number of contrast-dependent extra-classical receptive field properties, in particular shrinking summation fields (Sceniak et al., 1999; Cavanaugh et al., 2002a) and a switch in the valence of a high-contrast surround with increasing center stimulus contrast (Sengpiel et al., 1997; Polat et al., 1998). Using a series of firing-rate models of increasing complexity, culminating in a nonlinear, topographically specific model of cortex, we have shown how all boost responses within a topographically-defined locus of attention.
of these properties can be produced by a network operating within the inhibition-stabilized regime (Ozeki et al., 2009), and how with an expansive neuronal input-output function, a network with strong recurrent excitatory connectivity can transition into this regime as a function of stimulus strength.

This model makes a number of testable experimental predictions. It predicts that V1 should respond to high-contrast, sharply-defined, spatially uniform stimuli with a spatially-periodic pattern of activity, and that these periodic activity patterns should produce length-tuning curves with multiple peaks. The spatial period of activity measured in length-tuning curves or across the population should increase with increasing stimulus contrast. Furthermore, because inhibitory cells resonate at a slightly lower spatial frequency, measurement of input conductance around the second peak in the length-tuning curve should reveal that the second peak for excitatory conductance is found at slightly smaller stimulus lengths than the second peak for inhibitory conductance. Both the spatially periodic patterns of activity and the second peaks in firing rate and conductance may disappear when tested with a stimulus overlaid by a Gaussian (or similarly smoothed) contrast window, or when tested with low contrast stimuli that do not drive the network into an effective ISN regime.

This model also predicts that the phase difference of the responses of excitatory and inhibitory cells to a spatially periodic input directed only to inhibitory cells (e.g. with channelrhodopsin) should depend on the spatial frequency of the stimulus. There should be a 180° transition in the relative phase of E and I cells about some critical spatial frequency, as illustrated in Figure 2.6 and Supplementary Figure 7.8. Because of the accelerating input-output nonlinearity, we predict that for this transition to occur, the network would need to be driven with a visual stimulus to ensure that it was operating in an effective ISN regime. The critical spatial frequency around which the 180° phase transition will occur should depend on the luminance contrast of this visual stimulus. Yet another prediction is that the preferred contrast-modulation spatial frequency, as measured by Tanaka and Ohzawa (2009), should increase with increasing luminance contrast. Lastly, we predict that at very low contrast,
multiple inputs to cortex should sum supralinearly, rather than sublinearly.

This network demonstrates that certain previously posited mechanisms for contrast-dependent changes in the eCRF may not be necessary. In particular, it has been proposed that contrast-dependent shrinking of summation field size and the switch from a facilitative to suppressive surround with center contrast relies on an asymmetry in the nonlinearity between E and I cells (Somers et al., 1998; Schwabe et al., 2010) (in terms of our model, the parameters \( n \) and \( k \) in equations 3.2, 3.3). Though recent electrophysiology has shown that these values may be different between E and I cells (Haider et al., 2010), other work, fitting the input-output function to a more complex powerlaw relationship, has shown that the transfer function of E and I cells may be quite similar (Nowak et al., 2010). Importantly, the nonlinearity of the contrast response function of regular spiking and fast spiking cells (putative E and I cells, respectively) has been shown to be nearly identical, up to an overall scale factor (Contreras and Palmer, 2003). Given the equivocality of experimental results, our goal here was simply to show that, though this asymmetry may be present to some degree, it is not necessary to produce these contrast-dependent changes. Rather, these contrast-dependent changes emerge as the network transitions into an effective ISN regime with increasing stimulus strength. This transition shifts the network's resonant spatial frequency into a non-DC mode, which then produces spatially periodic patterns of activity and surround suppression.

Underlying many of the response properties we have presented here is a novel mechanism for context-dependent modulation: spatially periodic response patterns. Spatial resonance enables these networks to respond to patterns of feedforward input in a highly flexible fashion. Through a resonant spatial frequency that increases with stimulus strength, this network can sum low-contrast stimuli over larger regions of cortex while at the same time providing a simple way to isolate high-contrast stimuli over smaller cortical areas. Such a network is intrinsically tuned to transition smoothly from a stimulus detection mode to a stimulus discrimination mode. In the next chapter, we will present some experimental tests of this
model by probing cortex for evidence of these spatially periodic responses.

Another important mechanistic implication of these modeling results is that normalization, the sublinear addition and relative scaling of responses to multiple stimuli, can occur without utilizing a non-specific “normalizing pool” of inhibitory interneurons (Carandini et al., 1997). All that is required in our model is an expansive neuronal nonlinearity (which can be identical for both E and I cells), recurrent connectivity sufficiently strong to generate an intrinsically unstable regime, and fast, strong feedback inhibition capable of maintaining stability. Though this model does not rule out the possibility of non-specific inhibitory input, we feel this mechanism is less likely to be playing a major role in normalization for two reasons. First, the size of the conductance change necessary to produce the shunting inhibition predicted to come from the normalizing pool is much larger than has been observed experimentally (Carandini et al., 1997). Second, there are no reports confirming the existence of these non-normalizing inhibitory neurons that are normalizing everyone else in the network (though certainly they may have been missed).

Is normalization simply a form of surround suppression? We showed that in our ring model, the sublinear addition of stimuli occurs because the network is able to induce an effective asymmetry in connectivity, producing a contrast-dependent width-tuning in orientation space. However, we have shown analytically that even without this asymmetry and width tuning, in a network built from neurons with expansive input-output curves and strong recurrent excitation, there should always be a regime in which stimuli are sublinearly added. So is it rather that surround suppression is simply a form of normalization? Or is there perhaps something more general going on?

We would argue that what we have described here is a general circuit motif that, by virtue of the potential for unstable feedback excitation, is driven into an intrinsically normalizing regime. In essence, in order to maintain stability, the network has established a feedback mechanism whereby each additional excitatory input a reduction in the gain of the excitatory elements. When additional excitation is directed at a single target (i.e. a sin-
gle stimulus is increased in contrast), the result is a steadily decelerating contrast response curve, culminating in contrast saturation and eventually super-saturation. When the additional excitation is instead directed at multiple targets (i.e. a second stimulus is added to the network), the recurrent connectivity causes a decrease in gain at both locations, leading to the type of normalization we study above. When the recurrent connectivity and stimulus space happen to both live along the same continuous dimension (e.g. orientation or retinotopic space), the result will be surround suppression. However, with completely random connectivity, normalization is still observed. It just doesn’t look like surround suppression anymore. We feel this mechanism may represent a novel way to think about the cortical computation of normalization.
Chapter 4

Experimental Tests of the Model

4.1 Introduction

It is well known that the response of neurons in primary visual cortex (V1) can be strongly suppressed by a surrounding stimulus (Cavanaugh et al., 2002a;b; Sceniak et al., 1999; Kapadia et al., 1999; Ozeki et al., 2004; 2009; Akasaki et al., 2002; Angelucci and Bressloff, 2006). Evidence suggests that this effect originates intracortically, through horizontal and feedback connections, rather than in the feedforward input (Angelucci and Bressloff, 2006; Ozeki et al., 2009). Recent experiments have revealed that the firing rates of both excitatory (E) and inhibitory (I) neurons are suppressed by high contrast stimuli (Song and Li, 2008), as are the excitatory and inhibitory conductances received by surround suppressed neurons (Ozeki et al., 2009). Our previously presented computational models detail the circuit mechanisms and requirements for suppression of both E and I (Rubin and Miller, 2009; 2010a;b; Miller and Rubin, 2010). Key amongst these is the requirement for strong recurrent excitatory connectivity balanced by strong feedback inhibition, a network configuration which we have called the “inhibition stabilized network” (ISN)(Ozeki et al., 2009; Tsodyks et al., 1997). Networks in this regime also selectively amplify spatially-periodic patterns of activity, a feature which we have offered as an mechanistic explanation for certain experi-
mental observations, such as the non-monotonic changes in inhibitory conductance observed during length-tuning (Anderson et al., 2001) and contrast modulation spatial frequency tuning (Tanaka and Ohzawa, 2009).

These circuit models make a number of straightforward, testable predictions. We predict that large, high contrast stimuli should evoke spatially-periodic patterns of activity over cortical space. In our model, these patterns produce firing rate size-tuning curves with multiple peaks as well as tuning for the envelope spatial frequency of stimuli modulated by a sinusoidal contrast modulation envelope. Based on our analysis and simulation of nonlinear models, we further predict that the presence or absence of this spatially periodic activity will depend on the strength of the stimulus. Spatial-periodicity will be more prevalent at high rather than low contrast, and the dominant spatial frequency observed should increase with increasing stimulus strength.

To test these predictions, we recorded single units from the primary visual cortex (V1) of adult ferrets (*Mustela putorius furo*) and conducted three experiments to explore the periodic nature of signals from the extra-classical receptive field (eCRF). In the first experiment, we performed a detailed size-tuning study designed to resolve the fine spatial structure of the receptive field surround. To do this, we measured the responses to 30 different stimuli with linearly spaced size steps to produce size-tuning curves with high spatial resolution. In the second study, we probed the cortex for spatially-periodic activity by translating a large grating stimulus across an axis of retinotopic space, a test of what we call “position tuning”. In the third study, we measured the response of neurons to contrast modulated drifting gratings while varying both the envelope spatial frequency and the contrast of the underlying luminance grating. From the responses, we constructed spatial frequency tuning curves for the envelope spatial frequency at multiple levels of luminance contrast. We then used these curves to test the prediction that the preferred envelope spatial frequency increases with increasing luminance contrast.

Overall, we find that spatial periodicity is a common and dominant component of V1
neuronal responses. Size- and position-tuning curves are described significantly better by a model that takes this periodicity into account, and the majority of neurons tested show an increase in the preferred contrast modulation spatial frequency with increasing luminance contrast. These results provide strong experimental evidence that surround suppression and contextual modulation in V1 is accomplished in part by the amplification of spatially-periodic patterns of activity.

4.2 Methods

4.2.1 Approvals

Animal care protocols conformed to the guidelines established by the National Institutes of Health and were approved by the Brandeis University Institutional Animal Care and Use Committee.

4.2.2 Experimental preparation and recording

Eight adult ferrets were prepared for extracellular recording. Anesthesia was induced with a mixture of ketamine hydrochloride (30 mg/kg) and xylazine hydrochloride (3.0 mg/kg), and maintained with 0.5–1.5% isoflurane. After initial anesthesia, 0.1 mg of atropine sulfate (0.5 mg/ml in 0.2 ml) was administered. A rectal temperature probe was inserted and EKG electrodes were attached. Incision sites were treated with a small subcutaneous injection of bupivacaine (0.25%). A nylon tracheal tube was inserted by tracheotomy, and animals were respirated on a mixture of 40% oxygen and 60% nitrous oxide during the subsequent surgeries. These values were adjusted to 50% oxygen and 50% nitrous oxide prior to the presentation of stimuli. Respiration rate and stroke volumes were adjusted to maintain an end-tidal CO$_2$ between 3.2% - 5.0%. An intraperitoneal (IP) line was then inserted, and was used during experiments to maintain paralysis of the extra-ocular muscles through an infusion of 21 µg/mL gallamine triethiodide in lactated Ringer’s solution with 5% glucose.
The animal was secured in a stereotaxic frame, a hole approximately 2 – 4 mm wide was made in the skull over primary visual cortex, and the overlying dura was dissected away. Heart rate, end-tidal CO$_2$, and temperature were monitored continuously throughout surgeries and recordings.

Extracellular signals were recorded using a carbon fiber microelectrode (Kation Scientific), amplified through a differential AC amplifier (A-M Systems Model 1700), and acquired for online analysis with a Micro 1401-3 Analog to Digital Converter (Cambridge Electronic Design). Spike sorting was done by hand in Spike2 (Cambridge Electronic Design). Stimuli were generated by a Mac Pro computer using the Psychophysics Toolbox (Brainard, 1997) and displayed on a Sharp Triton CRT monitor.

4.2.3 Visual stimulation

Stimuli were presented as follows. The receptive field properties of isolated single units were initially characterized qualitatively with a manually controlled computer generated stimulus, which was used to roughly identify the spatial coordinates and stimulus preferences of the receptive field. After an initial coarse mapping, the monitor was repositioned to be centered on the neuron’s receptive field, and the spatial coordinates were again mapped. To more precisely identify the receptive field position, we monitored the spiking response of the isolated cell while switching between a circular stimulus and an annular stimulus whose inner radius equaled the radius of the circular stimulus. By progressively decreasing this radius while searching for a position that yielded a response to the circular stimulus but no response to the annular stimulus, we were able to quickly locate the center of the classical receptive field (Song and Li, 2008).

Once the receptive field was qualitatively tuned, we then ran a series of quantitative tuning stimuli. We first found the cells’ preferred direction of motion using 16 drifting grating stimuli (in 22.5 degree steps). The motion direction that yielded the maximum response was set as the preferred direction for that cell, and was used in all subsequent tests.
We then found the cell’s preferred luminance spatial frequency using seven different stimuli, and the spatial frequency that yielded the maximum response was similarly used for the remainder of the tests.

After recalibrating the stimulus to the preferred orientation and spatial frequency of the cell, we ran a high spatial resolution size-tuning test. 30 different circular drifting gratings, varying in size from 1 to 30 degrees in radius, were presented in random order to the cell. For a subset of cells, we ran the size-tuning test at more than one level of stimulus contrast.

We next ran what we call a “position tuning” test. A large (15 degree radius) stimulus was presented at one of twenty-one different positions on the screen. Each position was generated by translating the center of the stimulus in 1.5 degree steps along the axis parallel to the orientation of the grating (perpendicular to its direction of motion), covering a distance of up to +/- fifteen degrees from the center of the receptive field.

In the last series of tests, we presented a large full-screen grating stimulus at the preferred orientation and luminance spatial frequency that was overlaid with a slowly drifting low frequency sinusoidal contrast modulation (CM). The amplitude of the modulation envelope equaled the luminance contrast, such that the overall stimulus contrast varied from the chosen luminance contrast to 0% contrast. We first performed a quantitative tuning for the preferred orientation of the CM grating, using full-screen stimuli with 8 different CM orientations (from 0° to 180° in 22.5° steps) at 70% luminance contrast. Once the preferred orientation was determined (measured as the peak of the F1 at the temporal frequency of the modulating grating), this orientation was used to perform contrast modulation spatial frequency tuning at four levels of luminance contrast (4%, 8%, 16%, and 64%).

For most experiments, individual stimuli were presented for 2 seconds, with a drift rate of 4 Hz and an interstimulus interval of 4 seconds. For the two experiments using contrast modulated gratings, stimuli were presented for 4 seconds, and the overlying modulation grating was drifted at 1 Hz. In all experiments, stimuli were presented 4 or 5 times in a pseudorandom order.
4.2. Methods

4.2.4 Data analysis

4.2.4.1 Maximum Likelihood Estimation (MLE) for Denoising

Because we wished to explore a large region of stimulus parameter space with each cell, we were constrained in the number of times we could present each stimulus by the amount of time we could reasonably expect to record from a single neuron. In general, each stimulus was presented only four or five times. During both the experiments and the subsequent data analysis, we observed that there was often a very slow ($< 0.001$ Hz) fluctuation in the magnitude of neural responses, as if there was some slow change in overall cortical excitability. Over the course of minutes, activity would seem to slowly flow from periods of relatively high activity to periods marked by weak or no responses to stimuli (Figure 4.1). Because these fluctuations occurred over very long time scales, they did not seem to greatly obscure the selectivity of responses (see Supplementary Figure 7.18). However, because we were limited to only four or five presentations per stimulus, they did greatly increase the apparent fluctuations in the various response curves. To remove this source of stimulus-nonspecific variability, we used our raw spiking data to find the maximum likelihood estimate (MLE) of the parameters in the following general model:

$$r(t) = g(t)s_i(t)$$  \hspace{1cm} (4.1)

where $r(t)$ is the measured firing rate at time $t$, $g(t)$ is a time varying process representing the fluctuating change in overall cortical variability, and $s_i(t)$ is the true response of the neuron to stimulus $i$. To set the temporal bin size, we use the actual stimulus onset and offset times recorded during the experiments. To fit $g(t)$, we first generated a set of basis functions of time; for simplicity we chose a sum of four sine waves. To generate these functions, we fit the raw firing rate curve (generated by averaging the spike train with a 100ms sliding temporal
4.2. Methods

FIG. 4.1: Examples of slow changes in cortical excitability. The firing rate versus time for six example cells is shown in blue (calculated with a 100ms sliding window). The initial estimate of the nonspecific scaling function, $g(t)$, is in green. The adjusted $g(t)$ that maximizes the likelihood of both $g(t)$ and the stimulus specific response vector $s_i(t)$ is shown in red.

window) to a 12-parameter function of the form:

$$\sum_{i=1}^{4} A_i \sin(\omega_i t + \phi_i)$$  \hspace{1cm} (4.2)

using a nonlinear least squares curve fitting algorithm. To ensure that these basis functions interfered minimally with our estimate of the stimulus-selective response, the frequency of the sine waves used for the basis functions was bounded from above at 10 cycles per experiment duration (typically on the order of 5-45 minutes). Thus we were reasonably sure that the choice of basis functions would not cause the true stimulus selective responses to be classified as fluctuations.

Having generated a 12-dimensional initial basis function (with an amplitude, frequency, and phase for each of the four sinusoids), we then used an MLE approach to find the best
parameters describing the neural response to the data. For the F0 component of the response, we used a Poisson model of neural spiking:

$$f(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (4.3)$$

Here, $k$ is the actual number of observed spikes in a given time bin, $\lambda = r(t)dt$ is the expected number of spikes in a given bin of duration $dt$, and $r(t)$ is as described in equation 4.1. For practical reasons, it is simpler to maximize the log of the likelihood, rather than the likelihood directly. The likelihood function that we maximized was:

$$\text{Log Likelihood} = \sum_t k \log (g(t)s_i(t)dt) - g(t)s_i(t)dt - k! \quad (4.4)$$

Because we postulated a multiplicative interaction between “global excitability” ($g(t)$) and the stimulus specific response ($s_i(t)$), in order to fit the parameters of each component we used an alternating approach, wherein we first found the MLE parameters for $s_i(t)$ while holding the parameters of $g(t)$ constant, then found the best parameters for $g(t)$ while holding $s_i(t)$ constant. During the $g(t)$ stage, we fit the four dimensional vector of amplitude parameters for the four sinusoids chosen as the basis function. During the $s_i(t)$ stage, we fit the $N$-dimensional vector of responses to the $N$ different stimuli used for a given experiment. For example, during the size-tuning experiments we presented 30 different stimulus sizes, and during the MLE fit we found the 30-dimensional vector of responses that best described the neural responses. This alternating approach was used because for a given parameter vector, the relationship between the parameters and the likelihood estimate was always concave. This assures that likelihood maximum found during an individual subroutine is the global maximum over the parameters. We alternated between the two optimization subroutines until the likelihood estimate converged.

For the F1 component, the measured output is no longer an integer number of spikes, but rather the power of the F1 component of the response, so a Poisson model does not
4.2. Methods

really make sense. Here we instead assumed a Gaussian distribution of neural responses:

\[ f(k, \lambda) = e^{-\frac{(k-\lambda)^2}{2\sigma^2}} \] (4.5)

The free parameter \( \sigma \) ends up being arbitrary, since the likelihood function we maximize looks like:

\[ \text{Log Likelihood} = \sum_t \frac{(k - g(t)s_i(t)dt)^2}{2\sigma^2} \] (4.6)

so \( \sigma \) affects the magnitude but not the position of the maximum.

In addition to removing unwanted fluctuations from the data, this approach has an additional advantage. For the experiments in which we probed stimulus space with a high resolution set of stimuli (e.g. the size-tuning experiments), we can add smoothness constraints to the MLE process. To do this, we simply add a constraint on the maximum allowable second derivative magnitude across the response vector. This allows us to take advantage of the fact that we expect stimulus responses to be smooth over the stimulus parameter space (e.g. size-tuning curves should be a smooth function of size), and we are probing on a fine enough grid of stimulus quality that we don’t expect large changes in response between two adjacent stimuli. In essence, we recoup some of the averaging power lost by using fewer presentations per stimulus by effectively averaging over adjacent lengths.

To find error bounds for our estimates, we derive an expression for the matrix of second derivatives (i.e. the Hessian matrix) of responses over parameter space. By assuming that the parameters we found are normally distributed, the diagonal of the negative inverse of the Hessian matrix evaluated at the MLE will give the estimate of the variance of the each parameter (because the second derivative of a Gaussian curve evaluated at its maximum is simply \(-1/\sigma^2\)). Unless otherwise noted, the responses and error bars presented throughout the results section are derived from this MLE process. Because the MLE process fits a purely multiplicative model, the magnitudes of the various tuning curves are arbitrary up to
4.2. Methods

4.2.4.2 Model fitting

To use the experimental data to test different hypotheses of surround structure, we fit the data from the size- and position-tuning experiments to two different conceptual models. The first is meant to represent a classic model of surround suppression, in which the response to an excitatory central region is suppressed by an inhibitory surround. Modeling the center receptive field as one Gaussian curve of position and the surround as a second, wider Gaussian curve:

\[
r(x) = a_1 e^{-\frac{(x-a_2)^2}{2a_3^2}} - a_4 e^{-\frac{(x-a_5)^2}{2a_6^2}} - a_7
\]  

Here, \( x \) is retinotopic position. We fit the parameters \( a_1 - a_7 \), which define the spatial extent and magnitude of the center and surround receptive fields, to our position-tuning data. This equation models the difference of two Gaussian curves, and so has been called the Difference of Gaussians (DoG) model (Sceniak et al., 1999; Levitt and Lund, 2002). We compare fits of this model to one inspired by our circuit model, in which the excitatory center is sinusoidally modulated by the surround:

\[
r(x) = b_1 e^{-\frac{(x-b_3)^2}{2b_5^2}} - b_4 e^{-\frac{(x-b_6)^2}{2b_9^2}} \left( \sin (b_7 x + b_8) + b_9 \right) - b_{10}
\]  

Here we fit the parameters \( b_1 - b_{10} \). The surround term is modulated by a sinusoid with a frequency given by the parameter \( b_7 \). We call this model the Sinusoidal Surround Modulation model (or SSM for short).

For the size-tuning data, we are interested in the response of a neuron as it integrates a stimulus over its center and surround receptive fields. For the Difference of Gaussians model, we used the following function:

\[
r(x) = a_1 \left( \left( \frac{1}{\sqrt{\pi}} \int_0^{a_2(x-a_3)} e^{-z^2} dz \right) + \frac{1}{2} \right) - a_4 \left( \left( \frac{1}{\sqrt{\pi}} \int_0^{a_5(x-a_6)} e^{-z^2} dz \right) + a_7 \right)
\]
where $x$ now refers to stimulus length. For the Sinusoidal Surround Modulation model, we fit the data to:

$$r(x) = b_1 \left( \left( \frac{1}{\sqrt{\pi}} \int_0^{b_2(x-b_3)} e^{-z^2} dz \right) + \frac{1}{2} \right) - b_4 \left( \left( \frac{1}{\sqrt{\pi}} \int_0^{b_5(x-b_6)} e^{-z^2} dz \right) + \frac{1}{2} \right) e^{(b_7+b_8 i)x+b_9 i} + b_{10}$$

(4.10)

The surround term is modulated by a complex exponential with a frequency given by the parameter $b_8$ (we consider only the real part of this function).

### 4.3 Results

#### 4.3.1 Basic response properties

We recorded from a total of 90 neurons in the ferret primary visual cortex. For each isolated cell, we first performed a series of quantitative tests of orientation and spatial frequency tuning. The stimulus that evoked the maximum response from the cell across these two dimensions was then used in the subsequent experiments.

First we determined the preferred orientation and direction of each cell using a high-contrast drifting sinusoidal stimulus (Figure 4.2). All direction-tuning stimuli were at 70% contrast, with a spatial frequency of 0.1 cycles/degree and a temporal frequency of 4 Hz. We used 16 different stimuli oriented between $0^\circ$ to $337.5^\circ$ (in $22.5^\circ$ steps), presented in a pseudorandom order.

In addition to ascertaining the preferred direction of motion, we further characterized the tuning properties of the cells. Several different metrics exist to quantify the orientation and direction selectivity. One measure of orientation selectivity commonly used is the bandwidth of the response in orientation space, which we define as the half-width at half the maximum response. Another quantification commonly used is the circular variance (Palmer and Miller, 2007), defined as:
4.3. Results

FIG. 4.2: Orientation and direction tuning properties. A. The distribution of circular variances across cells, which is relatively uniform. B. The distribution of orientation bandwidths, defined as the half-width at half the maximum response. C. As has been previously observed, cells with narrow bandwidth have a highly variable circular variance, but cells with wide bandwidth have uniformly large circular variance. D. The mean direction tuning for all cells, aligned at the preferred direction (shaded area indicates SEM). E., F. The frequency histogram and cumulative histogram of direction selectivity, defined as $DI = \frac{r(\theta_p) - r(\theta_o)}{r(\theta_p) + r(\theta_o)}$.

$$CV = 1 - \left| \frac{\sum_j r(\theta_j) e^{j2\theta_j}}{\sum_j r(\theta_j)} \right|$$  \hspace{1cm} (4.11)

A cell with a circular variance equal to 1 responds equally to all orientations, whereas a cell with a circular variance equal to 0 responds only to one orientation (in either direction in this case). This metric has the advantage of being a more global measure of orientation tuning, whereas bandwidth is insensitive to features of the neuron’s response under the half-maximum. The distributions of bandwidths and circular variances, as well as the relationship between the two, is very similar to previous studies (Ringach et al., 2002; Gur et al., 2005); in particular, as was observed by Ringach et al.(2002), cells with narrow bandwidths have a
very highly variable circular variance, whereas cells with wider bandwidths tend to have a uniformly large circular variance (Figure 4.2C).

In addition, we quantified the direction selectivity of each cell using the following index (Jagadeesh et al., 1997):

\[ DI = \frac{r(\theta_p) - r(\theta_o)}{r(\theta_p) + r(\theta_o)} \] (4.12)

with \( \theta_p \) is the preferred direction and \( \theta_o \) is the opposite direction. The distribution of direction indices is consistent with previously published reports from ferret V1 (Li et al., 2008; Usrey et al., 2003).

Next, we tested each cell’s spatial frequency tuning to high contrast luminance gratings drifting in the preferred direction. For each cell, we determined both the preferred spatial frequency and the spatial frequency bandwidth. Bandwidth was defined as:

\[ B = \log_2 (SF_{high}) - \log_2 (SF_{low}) \] (4.13)

with \( SF_{high} \) and \( SF_{low} \) defined as the spatial frequencies at half the maximal response above and below the preferred spatial frequency. For cells whose response did not drop to less than half the maximal response above or below the preferred spatial frequency, bandwidth was not calculated; these were classified simply as high-pass or low-pass, respectively.

Spatial frequency tuning curves were recorded for 88 cells (Figure 4.3). Of these, 23 were classified as low-pass, 5 as high-pass, and 2 were so broadly tuned that they were neither low- nor high-pass. The remainder were classified as band-pass, and their bandwidths were calculated. The mean preferred spatial frequency over the whole population of neurons was 0.16 ± 0.01 cycles/degree (± standard error), and the mean bandwidth (of the band-pass neurons) was 2.1 ± 0.14 octaves. These numbers are also consistent with previously published data from ferret V1 (Li et al., 2006b).
4.3. Results

4.3.2 Size tuning

One of the most straightforward predictions from our circuit model is that stimulation of the center and surround should induce a spatially-periodic pattern of activity over cortical space, and this pattern should translate into length- and size-tuning curves with multiple peaks. Though there has been some hint of this in the literature (Anderson et al., 2001; Sengpiel et al., 1997), in general this effect has not been widely discussed. One reason why this feature of the neural response may have been largely overlooked is that previous experiments have not probed the stimulus space in a way that would properly reveal this effect. Most published experiments on size-tuning search for the size of the stimulus that yields the maximum response from the neuron, which in general is relatively small (DeAngelis et al., 1994; Anderson et al., 2001; Song and Li, 2008; Ozeki et al., 2004; Akasaki et al., 2002; Jones et al., 2001). These previous studies have all typically used somewhere between 8-10 different size stimuli, with the majority clustered around the summation field size of the neuron. These studies tend to use only one or two very large stimuli, with the simple goal

FIG. 4.3: Luminance spatial frequency tuning properties. A. The distribution of preferred spatial frequencies. B. The distribution of spatial frequency bandwidth, defined as $B = \log_2(SF_{high}) - \log_2(SF_{low})$, with $SF_{high}$ and $SF_{low}$ defined as the spatial frequencies at half the maximal response above and below the preferred spatial frequency.
FIG. 4.4: **Six example size-tuning curves.** The response of the cell versus stimulus radius is indicated in black on each plot. Error bars indicate the standard deviation of the estimate from the MLE algorithm. The red curve is the best fit Difference of Gaussians (DoG) model, and the blue curve is the best fit Sinusoidal Surround Modulation (SSM) model. Firing rates on each curve are normalized to set peaks to 1.

of coarsely demonstrating surround suppression. With such low spatial resolution in the surround, it is nearly impossible to infer or explore the finer spatial structure.

Thus, to test our hypothesis, we performed size-tuning experiments on V1 neurons using stimuli at 30 different sizes (varying from 1° to 30° in 1° steps). Stimuli are set to be of preferred orientation and spatial frequency for each cell, and are presented in a pseudorandom order. In total, we recorded size-tuning curves from 76 neurons. From this data, we are able to describe the spatial structure of the surround response to grating stimuli with high spatial resolution. The results from several example cells are shown in figure 4.4.

The most striking aspect of these results is the clear presence of spatial periodicity in the responses. In our data set, almost all cells display some degree of spatial periodicity, with most demonstrating two or even three peaks in their size-tuning curves. To quantify
the degree to which this periodicity was present in the data, we fit each of the size-tuning
curves to two conceptual models. The first of these is a Difference of Gaussians (DoG)
model, in which an excitatory center is modulated by a spatially uniformly surround. This
general model has been used many times to describe surround suppression (Sceniak et al.,
1999; Levitt and Lund, 2002). We compare the fit to this conceptual model to one inspired
by our circuit model, which we call a Sinusoidal Surround Modulation (SSM) model. In
this conceptual model, an excitatory center is modulated by a surround whose magnitude
oscillates with stimulus size.

For all 76 neurons, the SSM model provided a better fit to the data than the DoG
model, as quantified as $1/(\sum \text{error}^2)$. This is not particularly meaningful on its own, as
the SSM model has three additional parameters, and the DoG model is simply a con-
strained version of the SSM model. Because these two models are nested (in that the
SSM model can be turned into the DoG model by setting parameters $b_7$, $b_8$, and $b_9$ equal
to 0), we can test for the statistical significance of this improvement by using a like-
lihood ratio test (Knight, 1999). The likelihood ratio test defines a test statistic
$$D = -2 \ln \left( \frac{\text{likelihood for null model (DoG)}}{\text{likelihood for alternative model (SSM)}} \right),$$
which corresponds to how much more likely the results were to be observed given the full (alternative) model than the constrained (null) model. For the purpose of hypothesis testing, this statistic is often approximated to be $\chi^2$
distributed with degrees of freedom equal to the difference in the number of free param-
ters between the alternative and null models. Using this test, we find that the SSM model
provides a significantly better ($p < 0.01$) fit for 97% (74/76) of cells (Figure 4.5A). This indi-
cates that for the vast majority of the cells, the periodicity in the surround was a significant
component of the response.

To ensure that we were not over-fitting our size-tuning curves to the SSM model, we
performed a boot-strap analysis (Supplementary Figure 7.19). For each cell, we reran the
fitting algorithm to the SSM model 100 times, each time leaving out a random 20% of the
data points. This allowed us to generate a range of estimates for each of the 10 parameters
in the SSM model, and to assess how much the best-fit curves differed when generated using only part of the data. In general, the boot-strap fit curves very closely agreed to the fit obtained with the full data set for each cell. For each of the 100 curves generated for each cell, we quantified the normalized difference between the boot-strap fit curves (\(X\)) and the full 30 data point curve (\(Y\)) as: 
\[ \Delta = \frac{\sum (Y_i - X_i)^2}{\sum Y_i^2}. \]
Plotting a histogram of these normalized differences on a log-log plot reveals a clear power law distribution, indicating that in general the difference between fits generated with 20% of the data missing and curves generated with the full data set were all clustered very close to 0. This reassures us that we did not just arbitrarily fit a curve through some noise in the data.

From our computational models, we predict that the dominant spatial frequency amplified by cortex should increase with increasing stimulus strength. To test this prediction, on a subset of the neurons in our study, we performed size-tuning studies at one or more additional levels of luminance contrast (Figure 4.6). For each cell at each contrast level, we fit the size-tuning curves with the SSM model to obtain the best-fit spatial frequency. Indeed, we found that for the majority of neurons, the dominant spatial frequency of size-tuning increased with increasing stimulus contrast (Figure 4.6B).

This data set also allows us to rule out a potential alternative hypothesis. If the spatially periodic responses we observed were not caused by the emergence of a spatial resonance, as we have modeled, but instead due to some static property of the functional architecture (for example the periodic spacing of orientation columns), we would expect that the relative phase of periodicity in the size-tuning curve should be roughly constant at different levels of contrast. If, instead, this periodicity was due solely to the emergent resonance of the network, the frequency and phase of which depend on strength and spatial distribution of the input, we would expect no particular relationship between relative phases from size tuning curves recorded at different levels of contrast.
4.3. Results

FIG. 4.5: Statistics of size-tuning fits. A. The reciprocal of the sum squared error for the DoG and SSM models. Not surprisingly, all 76 cells are better described by the SSM model (all points lie above the unity line). Cells that are described significantly better ($p < 0.01$) by the SSM model are colored in blue. Those that are not are in red. B. The best-fit spatial frequencies of size tuning for all 76 cells tested. The best-fit spatial frequency for the full data fit is indicated by the black circle, and the mean and standard deviation of the best fit frequencies from the boot strap analysis are in blue.

To address this question, for each cell on which we measured size-tuning at more than one level of contrast, we calculated the cross-correlation between the size-tuning curves measured at different levels of contrast (after discarding the first 40% of the curve, which in general
4.3. Results

for all cells simply captures spatial summation). We plot the cross-correlation of each pair of size-tuning curves from each cell against the ratio of the stimulus contrasts in Figure 4.6C. For all contrast ratios, we find that the mean correlation coefficients are not significantly different from 0, indicating that size-tuning curves measured at different levels of contrast are not significantly correlated. This argues against the hypothesis that periodicity emerges from a static property of the functional architecture.

In one previous study of size-tuning and stimulus contrast, it was noted that there is often a re-emergence of activity at large stimulus sizes after surround suppression, an effect these authors called “counter-suppression” (Wang et al., 2009). Interestingly, these investigators found the appearance of counter-suppression to be more predominant with low contrast stimuli. This result initially seems at odds with our prediction, which is that surround suppression and re-emergence are the result of spatially-periodic activity that emerges at higher stimulus contrasts. However, this discrepancy may be caused by the particular way in which these investigators quantified the observed counter-suppression. For each cell, they calculated a counter-suppression index (CSI), defined as: \[ \text{CSI} = \frac{R_{cs} - R_{min}}{R_{max}}. \] In this index, \( R_{max} \) is the firing rate at the summation field peak, \( R_{min} \) is the minimum firing rate in the surround suppressed region, and \( R_{cs} \) is the counter-suppression firing rate (i.e. the maximum rate of the second peak). It is worth noting, though, that differentiating between counter suppression and simple summation in a size-tuning curve is really quite a subjective endeavor. In studying more thoroughly their data as well as our own, we observe that this index may falsely assign high values of counter-suppression to cells that, in reality, show only summation and facilitation. If a size-tuning curve is generally showing facilitation, but has one or two small dips that are falsely assigned as troughs, the difference between \( R_{cs} \) and \( R_{min} \) is then going to be quite large, whereas \( R_{max} \) is relatively small.
4.3. Results

A. Using the definition of counter-suppression index (CSI) from Wang et al. (Wang et al., 2009), we observe a decrease in CSI with increasing contrast. B. From our fit to the SSM model, we see an increase in dominant spatial frequency with increasing contrast. C. The cross-correlations of all pairs of size-tuning curves measured for a given cell plotted against the base-2 log of the ratio of the stimulus contrasts. For all contrast ratios (except for 3.125 (log2(3.125) = 1.6439), for which there was only 1 data point), the mean (indicated by a red x) was not significantly different from 0.

To test this supposition, we used the definition of CSI from Wang et al. (2009) and applied it to our own data to calculate a counter suppression index for each cell. To remove any subjectivity from the process, we simply used all extrema in each size-tuning curve, and calculated a CSI for each consecutive triplet of peaks in a given curve. For each cell, we then averaged all of the individual indices to get the mean CSI for the cell. At low contrast, because neurons tend to have much larger summation fields and generally facilitatory surrounds, we also observed much larger CSI's at low compared to high contrast (Figure 4.6A). We would argue, though, that this is not capturing a decrease in the re-emergence of firing rates after surround suppression at high contrast, but rather the fact that at low contrast more cells are likely to be exclusively summing, albeit with some bumps along the way, and this index will inadvertently assign these cells high values of counter-suppression.
4.3.3 Position tuning

From our circuit model, we predict that the spatial periodicity that we observe in the size-tuning data results from spatially periodic patterns of activity over cortical space. Ideally, to further test this prediction, we would measure a spatial profile of responses from several neurons over an axis of cortical space. However, to obtain a sufficient number of distinct spatial measurements, one would conceivably need 10 or more electrodes simultaneously recording from a subset of the population all sharing a common set of preferred stimulus characteristics (e.g. stimulus orientation and spatial frequency). In an in vivo preparation such as ours, this experiment would be technically impossible. As a surrogate experiment, then, we recorded from isolated single units while translating the stimulus position over an axis of retinotopic space, in the hope that the evoked pattern of activity would translate with the stimulus. In essence, we attempted to replace the movement of the recording electrode by $x^\circ$ in one direction of retinotopic space by an equivalent movement of the stimulus in $x^\circ$ of the opposite direction. We call this a test of “position tuning”.

We presented a large (15$^\circ$) grating stimulus at one of 21 equally-spaced positions along the axis of the preferred orientation. Position was varied over a total of 30$^\circ$ of cortical space, from 15$^\circ$ to one side of the CRF center to 15$^\circ$ on the other side, in 1.5$^\circ$ steps. As always, stimuli were set to be of preferred orientation and spatial frequency for each cell, and were presented in pseudorandom order. We recorded position-tuning curves from a total of 74 neurons (Figure 4.7).

We again fit these curves to the two conceptual models, the DoG model and the SSM model, altered here to reflect the constant stimulus size but varying position on the screen. Because of the difference in the cells’ preferred orientations, the asymmetric dimensions of the monitor, and the variability in receptive field shape and size, for some cells the position tuning stimulus was moved completely out of the cell’s receptive field, whereas for others it was not. Overall, though, the cells again show a marked degree of spatial periodicity, and the SSM provided a significantly ($p < 0.01$) better fit for 89% (66/74) of cells. Boot-strap
4.3. Results

FIG. 4.7: Four example position-tuning curves. Responses are plotted in black versus stimulus position relative to the receptive field center. Error bars indicate the standard deviation of the estimate from the MLE algorithm. The red curve is the best fit Difference of Gaussians (DoG) model, and the blue curve is the best fit Sinusoidal Surround Modulation (SSM) model. Firing rates on each curve are normalized to 1.

analysis again reveals that these fits are robust to leaving out a substantial portion (20%) of the data (Supplementary Figure 7.20).

For both the size and position tuning studies, our models predict that the resonant spatial frequencies driving the network should be lower than those preferred by the classical receptive field. This relationship was observed by Tanaka and Ohzawa in their test of contrast modulation spatial frequency tuning (2009). To examine if this held true in our novel tests of spatial periodicity, we examined the relationship between the preferred luminance spatial frequency and the best-fit spatial frequencies from both the size- and position-tuning tests
(Figure 4.9). For 74/76 cells, the preferred luminance spatial frequency was greater than
the best-fit spatial frequency from the size-tuning test, and for 66/74 cells, the preferred
luminance spatial frequency was greater than the best-fit spatial frequency from the position-
tuning test.

![Graph A](image)

**FIG. 4.8:**

**Statistics of position-tuning fits.** **A.** The reciprocal of the sum squared error for the DoG and
SSM models. As with the size-tuning data, cells with significantly better fits ($p < 0.01$) are colored
in blue, and those that are not are in red. **B.** The best-fit spatial frequencies of position tuning for
all 74 cells tested. The best fit frequency for the full data fit is indicated by the black circle, the
mean and standard deviations of the best fit frequency from the boot strap analyses are in blue.
Contrast modulation spatial frequency tuning

In our last experiment, we tested cells for tuning to the envelope spatial frequency of a sinusoidally contrast modulated luminance grating (Figure 4.10). In particular, we wished to determine if and how this tuning changes with the underlying luminance grating contrast. Tanaka and Ohzawa explored the response of V1 neurons to large contrast modulated gratings covering both the receptive field center and surround, and showed that they are well tuned to the spatial frequency of contrast modulation (2009). They found that this tuning typically peaks at spatial frequencies between 2 to 3 octaves lower than the preferred luminance spatial frequency, is orientation sensitive but uncorrelated with the orientation preference to the underlying luminance grating, and is insensitive to the contrast modulation drift direction. Our circuit models of V1 reproduce all of these findings, and make a further prediction. In our models, because of the accelerating neuronal input-output function, the effective strength of recurrent connectivity increases with increasing stimulus strength. Increasing connectivity strength causes an increase in the resonant spatial frequencies of the excitatory and inhibitory populations (Supplementary Figure 7.1). As we interpret the contrast modulation tuning observed by Tanaka and Ohzawa as a “pinging” of the network’s resonant frequencies, we predict that the preferred contrast modulation spatial frequency should increase with increasing luminance contrast.

To test this prediction, we first determined the preferred contrast modulation orientation of each cell (Figure 4.10C). As has been observed previously, there was no relationship between the preferred luminance grating orientation and preferred contrast modulation orientation (Pearson’s $r = 0.12, p = 0.33$). We then drove the cells with full-screen contrast modulated drifting gratings over a range of contrast modulation spatial frequencies and luminance contrasts (luminance grating orientation and spatial frequency and contrast modulation orientation were fixed at the preferred values for the cell). The contrast modulation was drifted at lower temporal frequency (1 Hz) than the underlying luminance grating (4 Hz), so that we could isolate the F1 response at the frequency of the modulation (Tanaka
4.3. Results

FIG. 4.9: Luminance spatial frequency versus size and position tuning spatial frequency. A. For each cell, the ratio between the preferred luminance spatial frequency in the classical receptive field and the best-fit spatial frequency from the size-tuning curve. For all but two cells the ratio is greater than 1. B. The distribution of the ratios between the preferred luminance spatial frequency and the best-fit spatial frequency from the position-tuning curve.

We used the magnitude of this F1 response to construct, for each cell, a contrast modulation spatial frequency tuning curve at each of the four luminance contrasts (4%, 8%, 16%, and 64%) tested.

Across the population, we found that the majority of cells (36/50) showed an increase in their preferred contrast modulation spatial frequency at high contrast, when compared to the response at the lowest contrast to which the cell responded. Of the remainder of cells, 8/50 showed a decrease, and 6/50 preferred the same CM spatial frequency at both low and high contrast (however, of these 6, in 2 cases the preferred CM spatial frequency increased for the two intermediate contrasts, and then decreased at the highest contrast, perhaps suggesting an influence of contrast super-saturation on some cells (Li and Creutzfeldt, 1984; Tyler and Apkarian, 1985; Ledgeway et al., 2005; Peirce, 2007)). Additionally, the mean preferred CM spatial frequency increases, but decelerates, with increasing contrast, an effect also observed...
4.3. Results

in our circuit models (Supplementary Figure 7.1).

FIG. 4.10: Contrast modulation spatial frequency tuning. A. Seven example cells are shown. Normalized tuning curves are shown at four levels of luminance contrast: 4% (blue), 8% (green), 16% (red), and 64% (cyan). Response is measured as the magnitude of the F1 component at the modulation (rather than luminance) drift rate. B. An example of a contrast modulated grating, the stimulus used in this experiment. C. A histogram of the distribution of differences between preferred carrier and contrast modulation orientation, both measured at high contrast. D. The distribution of the ratio of preferred luminance spatial frequency (at 64% contrast) to preferred contrast modulation spatial frequency. E. The mean preferred CM spatial frequency increases with stimulus contrast. Error bars indicate SEM. F. A pie chart summarizing the population data set. 72% of cells preferred a higher spatial frequency at high contrast, whereas only 16% preferred a lower spatial frequency, and 12% preferred the same spatial frequency. Of the 36/50 cells that preferred higher spatial frequency at high contrast, 25 of them (69%) were low-pass (i.e. preferred 0.01 cycles per degree) at the lowest contrast at which they responded.
4.4 Discussion/conclusions

From these three experiments, we find substantial evidence supporting our hypothesis that surround stimulation is accompanied by the amplification of spatially-periodic patterns of neural activity. For all but two cells, size-tuning curves were better described by a conceptual model with a sinusoidally modulated surround, and nearly all cells were also better described as having a spatially-periodic position-tuning profile. We observed tuning for the spatial frequency of a sinusoidal contrast modulation envelope for large drifting gratings, and the preferred frequency of this tuning increased with increasing luminance contrast.

Though the position-tuning and contrast modulation spatial frequency tuning components of our study are relatively novel, it is worth pointing out that over the past few decades, there have been numerous studies characterizing the size-tuning properties of neurons in primary visual cortex (Li and Li, 1994; DeAngelis et al., 1994; Anderson et al., 2001; Song and Li, 2008; Wang et al., 2009; Ozeki et al., 2004; Akasaki et al., 2002; Jones et al., 2001). Why should we have expected to find something new in a subject area that has been so extensively studied? As it turns out, virtually none of the previously published experiments have focussed in detail on the spatial structure of the surround. The majority of these studies probed the stimulus space with between 8-12 different stimuli, and the stimuli used tend to be clustered within the bounds of the classical receptive field (CRF). One very common approach has been to use logarithmically-spaced step sizes, such that of the 8-12 stimuli, only 2-4 are large enough to drive both center and surround. Of course, this sort of experimental approach is entirely appropriate given the goal of most of these studies, which is to precisely measure the size of the summation field. But this approach will certainly fail to reveal either the presence of absence of periodic behavior in the surround. There have been some studies in which the stimulus space is more finely sampled (Li and Li, 1994; Wang et al., 2009), and encouragingly, these studies report the frequent presence of a “disinhibitory” or “counter-suppressive” surround, which looks a lot like the spatial-periodicity we study here. Though the experimental results of Wang et al. (2009) show the opposite contrast-dependency we
have predicted, as we discussed above, this contradiction may simply be an effect of the particular index used to quantify the results.

In the test of position-tuning, we frequently observed cells whose responses changed non-monotonically with changes in stimulus position. Aside from being better fit by the SSM model, this result is more-or-less completely at odds with the traditional notion of the standard Gaussian-shaped receptive field. But it is exactly what is predicted by our model of spatial contextual interactions. There is no other model, to our knowledge, that predicts these types of spatially-periodic responses to translations of a uniform stimulus across the receptive field.

In implementing and analyzing the results of these studies, there were some technical challenges to overcome. Traditionally, when one wishes to assess the response of a neuron to a class of stimuli (such as gratings of different sizes), it is preferable to present the individual stimuli as many times as possible to obtain a confident measure of the response. With only 8-12 stimuli, it is reasonable to expect to be able to record the response to each stimulus upwards of 10-12 times. With this relatively large number of samples, one can be fairly confident that fluctuations and other non-stimulus relevant changes in the cortical state will be effectively averaged out of the data. However, to record the response to 20-30 different stimuli for each of 3 different experiments, as we have done here, we can only reasonably expect to show each stimulus 4 or 5 times before the experiment duration becomes so long that we can’t reliably record from a single neuron. As a result, simple averages of responses tend to be substantially nosier, and easily corrupted by stimulus-independent changes in the cortical state. To address this constraint, we have employed more advanced statistical tools to better approximate the response of the neurons to the stimuli. These data analysis techniques may be of general interest to neuroscientists working with electrophysiological data from anesthetized animals, where global changes in cortical excitability may be a fairly general phenomenon (Roughan and Laming, 1998; Kiviniemi et al., 2000; Lorincz et al., 2009).
4.4. Discussion/conclusions

Though for many cells the spatial-periodicity was clearly a powerful component of the spatial structure of the surround, in general we did not observe any strong correlations between the best-fit spatial frequencies recorded between different tests. This may be because of the different neuronal subpopulations recruited by the different stimulation paradigms. In the size-tuning tests, additional neuronal populations are recruited concentrically with increasing stimulus size, whereas the periodicity observed in position-tuning tests is likely recruited from increasingly distinct subpopulations at the larger displacement values. And in the contrast modulation spatial frequency tuning paradigm, the neural population recruited is made up of the entirety of cortex driven by our monitor. This is important to consider, because the spatial resonance we are studying here is not a fixed property of individual cells, like orientation preference or receptive field location, but rather an emergent property of the interacting network. In theory, because of the intrinsic nonlinearity of real neurons, each unique combination of active neurons in a particular population response could give rise to a completely unique pattern of spatial resonance. Because we assume cellular properties change relatively slowly over cortical space, we may reasonably assume that this resonance similarly changes relatively slowly with changes in the visual stimulus. Even still, the spatially-periodic response recorded from a given size- or position-tuning study likely results from the sum of some unknown number of different frequencies, each amplified by slightly different regions of the cortical space, and each interacting in some unique way. Thus, the periodicity we have measured, though itself a network level effect, is likely to be the result of multiple interacting resonant subnetworks, and each unique pattern of cortical activation may give rise to a similarly unique pattern of spatial resonance.

To further reassure ourselves that we were not overlooking an important relationship, we ran simulations of size-tuning, position-tuning, and contrast modulation spatial frequency tuning in 100 randomly selected neurons in our probabilistically-connected topographic model of V1, which we have designed to display a realistic degree of cell-to-cell variation. We then used the SSM models presented above to find the best-fit spatial fre-
quencies for simulated size-tuning and position-tuning data. Indeed, across this sample of cells, we found a similar lack of correlation between any of the three tests of spatial resonance (Supplementary Figure 7.21).

What computational function might this spatially-periodic activity serve? Certainly, it is possible that everything we’ve observed here is simply epiphenomenal. In wiring itself in such a way that it effectively normalizes its inputs, cortex may simply have inadvertently placed itself into a parameter regime that produces spatial periodicity. In our simple model, the requirements for achieving this activity are quite general. As long as the network has strong recurrent excitation, the need to maintain stability (by either anatomic or effective asymmetry in $E \rightarrow I$ and $E \rightarrow E$ connectivity) will almost guarantee the necessary conditions for periodicity on some spatial scale. Whether or not this spatial scale is of the right order of magnitude to influence responses may simply be a matter of chance.

Assuming this periodicity does serve a real function, though, it is interesting to consider what it might be. In area V2, cells are known to respond to “second-order” stimuli (such as contrast modulated gratings), which are defined by variations in properties other than luminance (Baker and Mareschal, 2001). Because of this feature of V2 responses, it has been proposed that V2 plays a role in texture segregation, boundary perception, and higher-order motion detection (Leventhal et al., 1998; Mareschal and Baker, 1999). However, unlike the response properties we have explored here in V1, neurons in V2 share a common set of tuning properties for both luminance and second-order stimuli (Mareschal and Baker, 1998; 1999). It is known that V2 receives the largest portion of its input from V1 (Van Essen, 2005), and so perhaps periodicity in V1 helps shape the receptive field structure of V2. Alternatively, it could be that V1 actually plays a larger role in detecting and discriminating second-order boundaries than has previously been thought. Previous work has characterized V1 neurons as generally unresponsive to second-order stimuli, but did so by driving cells with a carrier grating tuned specifically to be outside of the cells’ spatial frequency response band (Zhou and Baker, 1994; 1996). Here, we find that cells do show tuning for second-order stimulus
features, but only when being driven by an underlying carrier grating to which they can respond. The presence for this tuning, however, argues against the idea that V1 cells are generally unconcerned with second-order stimuli.

Another potential computational benefit of this sort of periodic activity may be a sort of “spatial binding”. In the same way that temporal oscillations across brain regions have been proposed to “bind” disparate computational processes (Tononi and Koch, 2008), perhaps spatial oscillations like those we have observed help to cluster the responses to discrete objects and texture groups within the visual scene. Such a spatial cluster may be useful for downstream areas. One could conceive of a readout system that quickly detects object edges by simply finding local maxima in the response field, and then unifies the two or more edges into a single object or texture field by looking for the standing wave that connects them.

In future work, we plan to use more sophisticated experimental techniques to test further predictions of the model. In particular, the ISN model makes the prediction that under the right conditions, the relative phase difference between excitatory and inhibitory activity can be modulated by changing the spatial frequency of the input to inhibitory neurons. Stimulating inhibitory cells at a low spatial frequency will cause them to drift in phase with the excitatory population, but stimulating them at a higher spatial frequency will generate out-of-phase firing. Of course, testing this prediction requires having some way to stimulate only the inhibitory population. Fortunately, with the advent of optogenetic techniques (Deisseroth, 2011), we are working to develop a way to do exactly this.
Chapter 5

Decorrelation of Parietal Neurons During Saccadic Choice

The work described in this chapter is a collaboration with the lab of Michael Goldberg. The theoretical and computational work was done by myself and Ken Miller. The experiments and analysis were designed and conducted by Annegret Falkner and Michael Goldberg. The introduction and experimental portions of this chapter were written by Annegret Falkner; the theoretical components were written by myself.

5.1 Abstract

Though it has long been suggested that correlations represent an independent channel of information in the brain separate from the firing rates, the relationship between correlated “noise” in the brain and saccadic behavior remains unclear (Cohen and Kohn, 2011). While neural activity in the monkey lateral intraparietal area (LIP) encodes the priority of spatial locations in the visual field (Bisley and Goldberg, 2003) and can be used to select the targets of upcoming saccadic eye movements (Ipata et al., 2006), it is unknown whether correlations between neurons in LIP represent behaviorally relevant information beyond what
is encoded by the spike rate alone. We simultaneously recorded from multiple macaque LIP neurons that encode the locations of visual targets during a saccadic choice task and examined whether changes in the “noise” correlation were related to both changes in neural activity and the monkeys’ behavior. We found that pairs of neurons that were positively correlated in the spontaneous activity prior to the target onset were decorrelated following target onset as firing rates increased. Additionally we found that the strength of the positive correlation prior to the appearance of the targets reliably encoded a measure of the monkey’s history of acquired rewards such that higher levels of correlation were associated with poorer saccadic performance. However, unlike after target onset, changes in the correlation during the spontaneous activity were not accompanied by changes in firing rate. These two seemingly paradoxical decorrelation effects can be accounted for by a single inhibition-stabilized network model of normalization that suggests that they result from the confluence of both spatially specific visual signals and non-spatial reward signals to LIP.

5.2 Introduction

To generate appropriate behaviors, the brain must combine information about the external world with a measure of the animal’s internal state (i.e. reward and motivation). Some of these variables, for example spatially specific signals about the locations of visual stimuli in the world, may be encoded in the firing rates of individual neurons, while other information that is shared across neurons may be undetectable at the level of the single neuron. Though in many cases, particularly in the oculomotor system, the firing rates of individual neurons have a direct relationship with the upcoming behaviors, it is unclear the extent to which shared variability, which manifests itself as correlated “noise” between simultaneously recorded neurons, can encode relevant information and influence saccadic behavior.

Many visual cortical areas exhibit correlations between neurons during the ongoing spontaneous activity (Kohn and Smith, 2005; Cohen and Maunsell, 2009) and it is now well established that decreases both in correlation (the shared variability) and in the indepen-
dent variability of neurons can lead to potential improvements in visual discriminability and encoding capacity (Bair et al., 2001; Mitchell et al., 2009). Decorrelations have been shown to result from both cognitive (Cohen and Newsome, 2008) and direct visual (Oram, 2011) stimulation and these changes have been shown to be more important in improving population sensitivity than either changes in the activity or individual variability (Cohen and Maunsell, 2009), since correlated noise can never be averaged out of a population, regardless of its size. It stands to reason then, that trials in which correlations are effectively reduced even prior to the start of the trial could provide a behavioral advantage over those in which the firing rates are swamped by correlated noise, though this has not been explicitly tested.

The lateral intraparietal area (LIP) converts visual information into a priority map for saccade selection and has an important role in decision-making (Roitman and Shadlen, 2002) and reward processing (Platt and Glimcher, 1999; Sugrue et al., 2004). LIP neurons receive spatially specific inputs from visual areas such as V1 and the superior colliculus, and also receive spatially specific top-down signals from the frontal eye fields in the frontal cortex. Additionally, LIP neurons are known to receive neuromodulatory signals from the brainstem including the locus ceruleus (Baizer et al., 1993), which are largely presumed to be non-spatial and associated with motivation and arousal.

Models of saccadic decision-making have routinely modeled this process as an independent “rise-to-threshold” with no role for correlations (Hanes and Schall, 1996; Ratcliff and McKoon, 2008), or with mutual inhibition between competing choice options (Constantinidis and Wang, 2004; Wong et al., 2007), however previous reports of correlations in the parietal cortex (Lee et al., 1998) suggest that these views are overly simplified. LIP neurons also exhibit strong surround suppressive effects (Falkner et al., 2010), which mediate competition between multiple visual stimuli prior to eye movements. Increases in surround suppression have been linked to both improvements in saccade latency and performance, though the precise role of inhibition in coordinating information between neurons is unknown.

Since saccade behavior is strongly modulated by both external and internal information
and the latencies of saccades are sensitive to the task related information (Lauwereyns et al., 2002), saccade latency provides us with a direct trial-by-trial readout of the monkey’s overall level of motivation. To examine whether saccade behavior is influenced by levels of correlated noise, we examined pairwise correlations between neurons in LIP in both the spontaneous activity and prior to the saccadic choice, and established links to both the firing rates and saccade behavior for these epochs.

The inhibition stabilized network model has been used previously to model aspects of visual cortical processing such as surround suppression, normalization, and “winner-take-all” stimulus selection (Ozeki et al., 2009; Rubin and Miller, 2010b). Since the primary features of this network share anatomical and physiological similarity with LIP, we used this model to shed some light on the relationship between firing rates and correlation and to suggest a potential decorrelation mechanism in cortex.

5.3 Results

5.3.1 Choice and saccade behavior

We used a dynamic foraging task (Sugrue et al., 2004; Lau and Glimcher, 2008) that required the monkey to make a saccadic choice on each trial between 2 simultaneously presented visual targets (Figure 5.1A). During this task, monkeys initiated the start of the trial by fixating on a central red cue. After a fixation delay, 2 colored targets (green and blue) appeared and monkeys waited for a variable delay, after which the fixation point was extinguished and monkeys had up to 400ms to foveate their chosen target. Targets were pseudo-randomly associated with relative reward probabilities ranging from 0.25 to 0.75 that were adjusted every 200-300 trials between the colored targets, though absolute reward magnitudes remained fixed over the duration of the session.
5.3. Results

FIG. 5.1:  

Choice task design and behavior. A. Monkeys performed a visual foraging task during which they made a free choice between 2 targets presented at the centers of the response fields of 2 simultaneously recorded LIP neurons with widely separated RFs. Monkeys had up to 400 ms to foveate their chosen target. B. Across sessions, monkeys’ average choice reflected the relative reward ratio of the chosen target. Error bars are +/- SE. Monkey D (red, n=71 behavioral sessions), slope=0.650, p=0.004, r-square=0.956. Monkey I (blue, n=33 behavioral sessions), slope=0.617, p=0.0002, r-square=0.993. C. Saccade latencies do not reflect relative reward of chosen target. Monkey D slope=0.019, p=0.993, r-square=0.000. Monkey I slope=-0.0003, p=0.993, r-square=0.000. D. Saccade latencies are well predicted by a single value measure of the monkey’s acquired rewards. Monkey D slope=-29.889, p=0.004, r-square=0.953. Monkey I slope=-19.668, p=0.001, r-square=0.983. E. Histogram of regression slopes from individual sessions regressing saccade latency with reward history. Average regression slope=-0.122, 63/104 (61%) behavioral sessions had significant regression slopes.
As observed in previous reports of foraging tasks (Sugrue et al., 2004), we found that both monkeys (Monkey D & Monkey I) adjusted their choice strategies to match the changing relative reward probabilities (Figure 5.1B), indicating that they were well aware of the changing parameters of the task (Monkey D, n=71 sessions, slope=0.650, p=0.004, average slope across sessions=0.579 SD=0.364; Monkey I, n=33 sessions, slope=0.61, p=0.0002, average slope across sessions=0.627 SD=0.275). However, the average saccade latencies for these choices were uncorrelated with the relative reward value of the acquired target (Figure 5.1C) (Monkey D, slope=0.019, p=0.993, average slope across sessions=0.011 SD=0.063; Monkey I, slope=-0.0003, p=0.993, average slope across sessions=0.016 SD=0.070).

Instead, we found that a much stronger predictor of the monkeys’ trial-to-trial saccade latency was the monkeys’ history of rewards (Figure 5.1D). When sorted within each session by whether the monkey had successfully harvested a reward on the previous trial or not, we found that the average saccade latency for when the monkey had failed to receive a reward on the previous trial was significantly longer than when he had successfully harvested a reward (p=0.0004, student’s t-test, comparison of rewarded and unrewarded trials, n=104). This effect was remarkably consistent irrespective of properties related to the choice, including the target choice on the current trial or the previous trial, the color of the target or the RF location of the previous or current choice, and also remained significant when analysis was restricted to only trials with equivalent relative reward probabilities. To determine whether saccade latencies were sensitive not only to a single missed reward, but an accumulation of missed or successful rewards, we fit saccadic latencies within each session with a vector that represents the trial-by-trial value of monkey’s history of reward (Reward history). To generate this term, we filtered the monkey’s total history of rewards with an exponential decay constant so that recent rewards would be more influential than distant rewards (see Methods). For each session we determined the best-fit linear regression between the saccade latency and the reward history term and extracted the session-by-session regression slope and best-fit decay constant. Regression slopes were mostly negative for each
session (Figure 5.1E, average slope=-0.12, SD=0.02, n=104) and 63/104 sessions (61%) were individually significant (regression p-value<0.05) indicating that during individual sessions successful harvesting of rewards (a high reward history) was associated with faster saccade latencies.

FIG. 5.2: Pairs of LIP neurons decorrelate following target onset. A. Spike counts of pair of example neurons during choices to targets in RF1 aligned to the onset of the targets (left panel) and the initiation of the saccade (right panel). Solid line shows response to cell with RF1 responding to the chosen target. Dotted line shows simultaneously recorded response to cell with RF2 to rejected target. Spike counts are for 300 ms bins stepped every 25 ms aligned on the end of the bin. B. Correlation coefficient of de-meaned spike counts shown in A. Error bars shown computed using jackknife methods. Red line shows correlations computed after shuffling trials. C. Population average spike counts for pairs of neurons with non-overlapping RFs. Conventions as in A. Shaded area is +/-SE. (n=114 choices from 67 pairs of neurons). D. Population average correlation coefficient (r) of spike counts for pairs of neurons shown in C. Insets show histograms of correlation coefficients at the time bins indicated by the black triangles. Pre-target epoch (300 ms prior to target onset, left panel), mean r=0.186, pre-saccade epoch (300 ms prior to saccade, right panel) mean r=0.058.
Across all behavioral sessions, when trials were grouped by reward history and averaged across sessions, average saccade latencies were well fit by reward history (Figure 5.1D, Monkey D slope=-29.889, p=0.004, Monkey I slope=-19.668, p=0.001).

Expected reward, which is determined not by the total reward stream but by the individual reward stream to a particular target has been shown previously to predict both choice and the firing rates of LIP neurons (Sugrue et al., 2004). We also fit this trial-to-trial measure with saccade latencies for each session and found that expected reward systematically accounted for less of the total variance of saccade latencies than total reward history (p=0.001, t-test of r-square values of expected reward compared to total reward, n=104) and fewer significant sessions than for the total reward stream (52/104, 50.0%).

### 5.3.2 Decorrelation of neural signals following target onset

We assessed the contributions of neural activity and “noise” correlations to the monkey’s saccadic behavior by recording from pairs of LIP neurons during the choice task. To avoid cross contamination, pairs of neurons were recorded from widely spaced (>2mm) independently moveable electrodes and recorded cells were selected to eliminate any signal correlation due to visual RF overlap (see inclusion criteria in methods). This was necessary in order to isolate the correlations due to shared variability (so-called neural noise) that are not associated with responses to the targets themselves, since neurons with RF overlap will both be stimulated by the same target and could trivially have correlations produced by the stimulus itself. We recorded from 208 neurons (104 pairs of neurons), 67 pairs of which were determined to have non-overlapping RF centers using a single target saccade task.

For each saccadic choice, activity consisted of the simultaneously recorded responses to the chosen target and the rejected target. As expected, activity increased after target onset for both neurons, and activity corresponding to the chosen target was greater than the activity evoked by the rejected target (Figure 5.2A) since the response to the chosen target is modulated by both a visual signal and a saccade preparation/attentional signal. For each
5.3. Results

Within a single pair of cells for choices to the target appearing in RF1, both the spike...
counts and the Pearson correlation changed dynamically over the behavioral time course, even though the monkey’s eye position had not changed. In the period prior to the target onset where the monkey is fixating waiting for the trial to begin (the pre-target epoch) the correlation is significantly positive and maximal (Figure 5.2B left panel), but after the targets appear in the RF of each cell the magnitude of the correlation steeply declines until in this pair it eventually reaches a minimum at the time of the initiation of the saccade (Figure 5.2B right panel). The correlation coefficient returns to positive after the eye moves.

We calculated the noise correlation for each choice in each pair of cells and averaged across the population. Across the population of cell pairs we observed similar changes in spike count (Figure 5.2C) and correlation trajectory (Figure 5.2D) over the course of the decision. In this population average, the correlation was significantly positive in the pre-target epoch (students t-test p<<0.001), and taken individually 58% of cell pairs had significantly positive correlations in this epoch.

Across the population correlations were significantly reduced prior to the saccade. For almost every individual choice, correlations in the 300 ms before the target appearance were significantly higher than they were 300ms before the saccade (Figure 5.2D). In alignment with previous reports, we found a small but systematic negative relationship between the change in correlation and the change in firing rate between the pre-target epoch and the pre-saccade epoch for the response to the chosen target (p=0.0063, least squares regression, comparison of percent change in firing rate with percent change in correlation between pre-target and pre-stimulus epochs, n=67 pairs).

Correlations can increase trivially as spike rates increases (Zeitler et al., 2006; de la Rocha et al., 2007). To ensure that the decrease in Pearson correlation observed over the course of the saccadic decision is not due to simply a reduction in the raw spike rate, we calculated the average spike count from the baseline vs. the spike count from the decision period. As expected, the spike counts for the decisions where saccades were executed into the RF were significantly enhanced in the pre-saccadic decision bin in comparison with the
5.3. Results

pre-target bin (p=0.011, paired t-test, n=67). The average spike counts for the cells where saccades were made opposite the cells’ RF were on average static (p=0.949, paired t-test). Since the spike counts were actually increased for one set of choices and were unchanged for the other, these changes cannot account for the change in correlation observed between the 2 epochs.

Though we did not systematically vary this parameter in this study and accepted cells only when they fit our inclusion criteria, we found that both correlations in the pre-target epoch and the pre-saccade epoch decreased as a function of the distance between the RF centers of the recorded cells (Supplementary Figure 7.23A).

5.3.3 Correlations and synchrony

In concordance with several other cortical areas, we found that correlations increase for increasing bin sizes (Supplementary Figure 7.22), suggesting that these correlations emerge on a slow timescale. This does not, however, preclude a contribution from synchronously occurring spikes. We calculated the incidence of synchronously occurring spikes by computing the cross correlation for each pair of cells during the pre-target and pre-saccadic epoch. For each epoch, we recalculated the cross correlation after shuffling trials and iterated this calculation 1000 times to extract estimates of the 95% confidence bounds for each pair of cells. Individually, 34/104 pairs of cells (25%) had significantly higher incidence of synchronous spikes than would be expected by chance in the pre-target epoch. During the pre-saccadic epoch, despite the fact that there are more spikes evoked, only one pair of cells (1%) had significantly higher incidence of coincident spikes.
5.3. Results

FIG. 5.3: **Correlation encodes reward history.** 

A. Average spike counts of trials divided by whether the monkey had successfully harvested a reward on the previous trial (blue) and those where the monkey had been unsuccessful the previous trial (red) show very little change. Solid lines are the responses to the chosen target and dotted lines are the responses to the rejected targets. Other conventions as in Figure 5.2. Average spike counts are not significantly different between rewarded and unrewarded trials for the pre-target epoch (red triangle, p=0.576 and p=0.489 for chosen and rejected target spike counts respectively, student’s t-test) or during the pre-saccade epoch (black triangle, p=0.831 and p=0.936 for chosen and rejected targets). 

B. Average correlation coefficients for the spike counts of previously rewarded (blue) and unrewarded trials (red). Trials where the monkey was previously rewarded are more decorrelated in the pre-target epoch (red triangle, p=0.004, student’s t-test) relative to trials where the monkey was unrewarded. Correlations are no longer significantly different in the pre-saccade epoch (black triangle, p=0.690, student’s t-test). 

C. Correlations are strongly modulated by the monkeys’ reward history. Population average of correlation coefficients computed in each session after sorting by reward history. Pre-target epoch (gray) slope=-0.99, p=0.004, r-square=0.995. Pre-saccade epoch (black) slope=-0.087, p=0.048, r-square=0.776. 

D,E. Spike counts for the responses to the chosen targets (D) and the rejected targets (E) are weakly modulated by reward history. For chosen targets, pre-target epoch (gray) slope=-0.559, p=0.010, r-square=0.918, and pre-saccade epoch (black) slope=0.432, p=0.021, r-square=0.867. For rejected targets, pre-target epoch slope=-0.489, p=0.144, r-square=0.562, pre-saccade epoch slope=0.007, p=0.971, r-square=0.001.
5.3.4 Correlation encodes reward history

Since the monkeys’ trial-to-trial saccade latency is strongly modulated by their previous history of rewards (Figure 5.1D,E), we investigated whether this variable was encoded in either the spiking activity or the correlations between LIP neurons during this task. We divided the trials in each session according to whether a reward was successfully harvested on the previous trial and plotted the neural activity and noise correlations for these trials (Figure 5.3A,B). Neural spiking activity showed no significant difference between trials where the monkey received a previous reward and those that did not (p=0.567 for chosen target response, p=0.489 for rejected target response, paired t-test). In contrast, the average correlation coefficient across sessions was strongly modulated by previous reward: trials where the monkey failed to receive a previous reward were more strongly positively correlated than the trials where monkeys had been successful (p=0.004, paired t-test, n=67 pairs). This effect begins in the spontaneous activity prior to the target onset and is exhibited several hundred ms into the trial. The average values of the correlation converge when aligned to the saccade onset, both exhibiting the stereotyped decrease over the trial described above (p=0.690, paired t-test between rewarded and unrewarded trials in pre-saccade bin). Trials where the monkey had previously received a reward exhibited less change than trials in which the monkey failed to receive a reward (mean percent change for unrewarded =175.3% SE=70.4, mean percent change for rewarded trials = 90.0% SE=52.6). Since trials where the monkey was unsuccessful at harvesting a reward in the previous trial are also associated with increased saccade latencies, this indicates that increased correlation is associated with poorer saccadic performance on this task.

Since correlations between neurons cannot be computed for individual trials and because it requires a sufficient number of trials to achieve an accurate estimate of the coefficient, we determined the relationship between correlation and the monkey’s reward history by dividing the trials during each session according to the reward history computed by regressing with saccade latency (see methods). The reward history term is bounded between 0 and 1, so trials
5.3. Results

were binned into 5 equally spaced bins. We extracted the spike counts associated with both
the pre-target epoch and the pre-saccadic epoch and computed the correlations across trials
for each pair in each bin and averaged across sessions. Similar to the saccade latency, we
found that the average correlation coefficients were strongly negatively modulated by reward
history, with high reward history values being associated with the lowest correlations. This
was true for both the pre-target epoch (Figure 5.3C-E, gray) and the pre-saccadic epoch
(Figure 5.3C-E, black).

Surprisingly we found that the average spike counts varied little across differing values
of reward history for both epochs (Figure 5.3D,E), and average spike counts between the
highest and lowest reward history bins differed by less than half a spike for the duration of the
bin (0.359 spikes/300ms SE=0.020 for pre-target epoch, 0.315 spikes/300ms for pre-saccade
epoch, n=67 pairs).

5.3.5 Correlations separately encode visual and cognitive signals

In a subset of LIP neuron pairs (n=33 pairs of neurons), we varied the spatial configuration of
the task such that only one RF encoded the location of a saccade target. The other saccade
target was placed diametrically opposed, in the opposite hemifield such that the second RF
was excited by neither saccade target (Figure 5.4A). Now monkeys could choose the target
in RF1, or could choose the target that was opposite RF1, and had no incentive to plan a
saccade to the second RF. We divided the trials by whether the monkey made a saccade to
the target in RF1 or to the target in neither RF and plotted both the neural activity and
the noise correlation associated with these trials (Figure 5.4B,C). As expected, the neural
activity in RF1 when the monkey planned saccades to that location was greater than the
activity of those same cells when the monkey planned a saccade to the target in neither RF.
During both of these choices, the neural activity encoding the empty space that was neither
target were not positively modulated by the onset of the target during either choice.
5.3. Results

**FIG. 5.4:**

**Decorrelation by spatially specific saccade signal.** 

**A.** Modified task design. Monkeys could choose the target in RF1 (left, responses shown in green) or target opposite RF1 (right, responses shown in blue).  

**B.** Average spike count (per 300ms bin) for choices to target in RF1 (green) and to target opposite RF1 (blue). Solid lines show response of neuron with RF1 and dotted line show response of empty RF. (n=33 pairs of neurons)  

**C.** Noise correlation of neurons during modified choice task. Correlations are decreased for saccadic choices into RF1 relative to choices to target opposite RF1. Conventions as in Figure 5.2.
This spatial configuration allowed us to ask two separate questions. First, do neurons still exhibit correlated activity when they do not both encode the locations of choice targets? We found that pairs of neurons had significant positive correlations in the pre-target epoch, even when one of the neurons did not encode a saccade target location. This suggests that this positive correlation emerges from non-spatial shared variability in the network or a shared input across many spatial locations, rather than a spatially specific input to the saccade target locations only. This activity was also modulated by the monkey’s history of reward in the pre-target epoch (p=0.040, paired t-test, comparison of rewarded and unrewarded trials during pre-target epoch).

A second question is whether correlations are modulated separately by visual and saccadic signals. The firing rates of LIP neurons have a spatially specific signal that represents both the visual location of the target and a cognitive signal that reflects the priority of that location (Colby et al., 1996). In this version of the task, the initial visual input to both RFs is identical (with one RF containing no target), while the priority signal differs depending on what the upcoming choice will be. We found that there is a significant difference in the correlation depending on whether the monkey plans a saccade to the RF compared to when the monkey plans a saccade to the target in neither RF: pairs that have higher firing rates that reflect higher priority are associated with more decorrelation from the initial pre-target epoch. This effect is maximal in the bins preceding the pre-saccadic epoch when firing rates are maximally separated and suggests that visual and cognitive factors can make separate contributions to decorrelating the noise in LIP over the course of the trial.

5.3.6 Decorrelation by spatial and non-spatial inputs

Previous reports suggest that changes in correlation can have a tight relationship with changes in firing (Cohen and Maunsell, 2009) rate or can be evoked without concurrent changes in activity (Churchland et al., 2010). In LIP, we find that both relationships are simultaneously valid: in the spontaneous activity, correlations are reduced with little change
5.3. Results

to the firing rate (Figure 5.3), while during the task epoch, increased signal is associated with decreased correlation (Figure 5.4). Can a single model of LIP account for both of these relationships? To test this, we modified a nonlinear inhibition stabilized network (ISN) model to elucidate the relationship between activity and correlation.

FIG. 5.5: 
Decorrelation by spatial and non-spatial inputs in a nonlinear inhibition stabilized model. A. Schematic of the modified ISN model. Cortex is modeled as a 180 degree ring of excitatory neurons (red) and inhibitory neurons (blue) with mutual suppression and recurrent connectivity. Simulations consisted of “recording” from neurons on the ring while presenting spatial and non-spatial inputs. B. Spatial inputs increase firing rates and decorrelate neurons. Both spatial inputs are presented simultaneously. Top panel shows neural activity in response to the chosen stimulus (blue) and rejected stimulus (green). Correlation is shown aligned to stimulus onset. C. Nonspatial input decorrelates spontaneous activity without changing activity. “High reward” (dotted line) corresponds to increased non-spatial signal relative to “low reward” (solid line). Nonspatial input is equal to all cells in the network. D. Increasing spatial input strength leads to increased firing rates and decreased correlation. Shown is the activity and correlation in the pre-stimulus epoch (red) and the post-stimulus epoch (blue) for a range of input strengths. E. Increasing non-spatial input strength leads to decreased correlation but no change in firing rate. Conventions as in D.
The ISN is a relatively simple rate model consisting of recurrently connected excitatory and inhibitory neurons arranged around a 180° ring (Figure 5.5A) such that neurons along this ring receive varied levels of surround suppression from their neighbors. LIP anatomy and physiology supports both of these modest characterizations (surround suppression and recurrent connectivity) since it has been shown to exhibit strong spatially tuned suppressive interactions, and is classically defined by its robust persistent activity during a memory guided saccade task, a known characteristic of neural networks that can be generated through strong recurrent connectivity (Wang, 2001; Brunel, 2003).

In this model, each neuron is represented by its firing rate that evolves as a function of time according to a pair of dynamical equations governing the relationship between excitation and inhibition (Methods). The strength of the input to each neuron is determined by the matrices of synaptic weights which remain unchanged over the course of the trial. Neurons in this model can also receive feed forward input (that can vary across the “spatiotopic field” of the network) and also receive a stationary Gaussian noise input (see Methods for details of simulation parameters). Inputs that have different spatiotopic response profiles represent visual stimuli to neurons with non-overlapping RFs and we can select neurons along the ring to “record” from to calculate the changes in activity and correlation in time.

We simulated the neural activity during the choice task by presenting orthogonal neurons in the network with 2 simultaneous choice inputs, allowing one input to be greater than the other to represent the confluence of the visual and saccade related signals in the inputs to the chosen neuron (Figure 5.5B, top panel, blue). We simulated 1000 trials using these inputs and using this resultant vector of firing rates for the two recorded neurons, computed the correlation across trials over time (Figure 5.5B, bottom panel). Similar to our pairwise neural recordings in the macaque, following the onset of the targets there is an abrupt decrease in the correlation. Additionally, there is a direct relationship between the strength of the spatial input and the change in the correlation (Figure 5.5D). We simulated the spatial inputs using a range of input strengths and averaged across simulations during both the pre-target epoch
(Figure 5.5D, red) and the post-target epoch (Figure 5.5D, blue). As expected, there is no change in either firing rate or in correlation in the pre-target epoch, since the input has not yet occurred. However, increased inputs to the simultaneously recorded cells result in an increased decorrelation, similar to that observed in our data.

The absolute magnitude of the correlation depends on the relative positions of the recorded neurons in the ring. Similar to our observed neural data, this magnitude of the correlation decreases as a function of the distance between the recorded neurons in both the pre-target and post-target epoch (Supplementary Figure 7.23B).

We next modeled the effects of reward history on the spontaneous activity in the model. We modeled these effects with the assumption that there is no spatial specificity to the input, consistent with the role of a neuromodulator. Instead of spatially selective inputs representing visual signals, our simulated reward input was spatially non-selective over all excitatory and inhibitory neurons in the network. During these simulations, we found that when this non-spatial input is strong (i.e. reward signal is high), correlations are decreased, though firing rates are unchanged (Figure 5.5C). Similar to the response to the spatial input, increased non-spatial input increases the amount of correlation decrease from its unstimulated state (Figure 5.5E, bottom panel). However, in accordance with our observed data, increasing the strength of the non-spatial input produces no appreciable modulation of firing rate (Figure 5.5E, top panel).

Though LIP receives feed forward signals that represent visual and saccade related information that are known to project to excitatory neurons, it is unknown the extent to which inhibitory cells (which also have spatiotopic receptive fields) receive this information. Additionally, non-spatial neuromodulators are known to interact with many receptor subtypes that exist in various combinations and configurations across cells types. Thus we wished to explore if there was any dependence in the relationship between input and correlation on the configuration of the spatial and non-spatial inputs.

Using the ISN model, we tested whether the negative relationship observed between
firing rates and correlation observed following spatial inputs was affected by the relative contributions of excitatory and inhibitory neurons. We repeated the above simulations at varying ratios of input to inhibitory vs. excitatory neurons, and extracted the relative changes in firing rate and correlation for each I/E ratio. For the spatial input, we found that the negative relationship between firing rate and correlation did not depend on the relative contributions of the E and I cells and that the change in firing rate accounted for a high proportion of the variance of the change in correlation (Figure 5.6A). Changes in I/E ratio did constrict the amount by which the firing rate changes after stimulus onset (high levels of inhibition are linked with less change in firing rate following stimulus onset), which changes the slope of this relationship.

![Graph A](image1)

**FIG. 5.6:** Relationship between activity and correlation depends on the ratio of input to excitatory vs. inhibition neurons. A. Change in firing rate vs. change in correlation between pre and post-stimulus epochs shown for a range of I/E ratios for spatial stimulus. Different points correspond to different strength spatial inputs. B. Change in activity compared with change in correlation between pre and post-stimulus epochs shown for a range of I/E ratios for non-spatial stimulus. Different points correspond to different strength non-spatial inputs. See plot for individual statistics.

In contrast, for non-spatial inputs, the change in firing rate is a poor predictor of the
change in correlation (Figure 5.6B). This was true across all simulated I/E ratios, though only in the “balanced” state (I/E ratio of 1), did the non-spatial input have a negligible effect on firing rate.

5.4 Discussion

Neurons move from a state of correlation to state of relative decorrelation prior to saccadic choice and both visual and saccade-related signals provide separable effects on decorrelation prior to saccade. Depending on the magnitude, the level of correlation provides saccadic advantage or disadvantage; at the start of the trial the most positively correlated activity is associated with the longest saccade latencies. The magnitude of this correlation is associated with changes in motivational factors including a measure of the monkey’s acquired reward history. Paradoxically, the decorrelation prior to the saccade is associated with an increase in the firing rate at the chosen target while the decorrelation during the spontaneous activity is associated with little or no change in firing rate. Simulations using a nonlinear ISN allow us to simultaneously model both of these processes, showing that spatial inputs produce a consistent negative relationship between activity and correlation, while non-spatial inputs have a decoupled relationship with firing rate. In a “balanced” network where non-spatial inputs are received equally by both excitatory and inhibitory cells, correlations can change substantially while activity changes little or remains unchanged, such that correlations can signal behaviorally relevant information that is not encoded by the firing rates of individual neurons.

The strength of the initial correlations we observe is on the order of the correlations seen in other visual cortical areas, including V1 (Kohn and Smith, 2005), V4 (Cohen and Maunsell, 2009; Mitchell et al., 2009), and MT (Bair et al., 2001), and similar to previous studies (Smith and Kohn, 2008), significant correlations were observed even between neurons with widely separated RFs and anatomical distances (>2 mm). Since significant correlations in the spontaneous activity are seen in anesthetized preparations as well as in the behaving
animal, it is likely that are the result of a network property and not from any feed forward sensory inputs.

Decorrelation, however, can result from input to the network and the ISN provides an intuition of how this can occur. This model was originally developed to study the well-known cortical phenomenon of normalization (Rubin and Miller, 2010b), a property often ascribed to cortical circuits (Heeger, 1992; Carandini and Heeger, 1994; Carandini et al., 1997; Reynolds and Heeger, 2009). Circuits that are known to produce normalization in vivo also perform a winner-take-all operation on their inputs (Busse et al., 2009). When the circuit receives two inputs of roughly equal strength, the response of the network is a sublinear sum of the responses to the two stimuli presented alone. However if the network receives simultaneous inputs that differ substantially in strength, the response of the network on the whole looks like the response to the stronger stimuli; the weaker stimulus is effectively ignored. Similar to the magnitude of the neural response, the response variability is also normalized. In our simulations, the correlation in the spontaneous activity is produced by a relatively weak, ongoing noisy processes. When a second input (or inputs) is introduced to the network, the winner-take-all mechanisms suppress the noisy weak input, resulting in less trial-to-trial variability. Importantly, as has been shown across cortical areas, the reduction in variability is the result of reduced shared variability between neurons (Churchland et al., 2010), and the “private” variability that by definition is unique to each neuron is unchanged by this process.

In our neural data, we propose that the drop in correlation seen in both the spontaneous activity and after the target onset is the result of a normalizing process, during which the addition of a non-spatial (which encodes a measure of the monkeys’ internal motivational state) and spatial input (which encodes the location of choice stimuli) reduces the shared noise through the network’s suppressive winner-take-all dynamics. In the case of the spatial signals, these inputs could correspond directly to feed-forward inputs from other visual areas and spatially specific signals from the frontal cortex that carry information about spatial
attention and saccade selection. Non-spatial signals, on the other hand, which may act to stabilize highly variable spontaneous activity (Rajan et al., 2010), work on a longer timescale and are likely to be neuromodulatory in nature. One possible mechanism is that an ascending brainstem structure releases a neuromodulator every time a reward is obtained. If the time to decay for this signal is long (as is strongly suggested by our saccade latency analysis), these signals may accumulate across trials such that a string of acquired rewards will result in a stronger non-spatial signal. Though in our data, we found little change in firing rate during the spontaneous activity across trials with different reward history values, this need not be the case for all non-spatial signals. By changing the ratio of the targets of the signal to prefer either excitatory cells or inhibitory cells, the firing rates of individual neurons can be positively or negatively modulated, while always resulting in a decorrelation. This mechanism is biologically plausible, since different neuromodulators would be expected to activate different constellations of receptor subtypes across classes of cells. In the model, this mechanism also allows for correlations to change on-the-fly” without changes to the synaptic weights between neurons.

5.4.1 Correlations and decoding

What is the role of decorrelation in cortical circuits? As has been previously observed, the presence of correlations and their relationship to the information conveyed in the neural signal depends critically on the type of decoding algorithm (Abbott and Dayan, 1999; Averbeck et al., 2006). Observing noise correlations in LIP simplifies this problem somewhat since there is strong evidence that saccade selection in LIP can be effectively decoded using a winner-take-all process in which the winning “peak” on LIP’s priority map will signal the moment-to-moment winner. A positive correlation in this decoding scheme would be potentially confounding to downstream decoders since distant peaks on LIP’s map would not be carrying independent information about the stimuli contained within their response fields.

Decorrelation has been associated with periods of wakefulness in rodents (Poulet and
Petersen, 2008) and this has been associated with higher signal to noise ratio during these epochs. As our saccade latency data similarly suggests, the magnitude of the correlation in the spontaneous activity is associated with a disadvantage for saccadic choice. It is unclear from these data whether correlations are causally involved in determining saccade latency, and the monkey’s reward history is likely only one factor of many that accounts for the actual trial-to-trial saccade latency. Rather, the saccade latency is an indicator of the monkey’s arousal level that reflects a pertinent internal state. It is the internal state that determines the correlation and reduced correlation allows for several potential decoding benefits, such as increased information in the first spike (Shamir, 2009).

These results are also highly consistent with several studies demonstrating that decorrelations accompany (and can primarily account for) an increase in spatial attention. An increase in spatial attention in our model provides the increased spatially specific input that results in decorrelation. However, our experiments and model go further in demonstrating that this is not the only mechanisms by which decorrelation may provide a decoding advantage. Not only does our model demonstrate that increases in firing rate are associated with increased decorrelations, it provides the framework for a more general mechanism that can be accessed by non-spatial motivational inputs as well as attentional or saccade related signals.

5.5 Materials and methods

We used two male rhesus monkeys (Macaca mulatta) weighing 8.12 kg in this experiment. All experimental protocols were approved by the Animal Care and Use Committees at Columbia University and the New York State Psychiatric Institute, and complied with the guidelines established by the Public Health Service Guide for the Care and Use of Laboratory Animals. We located the intraparietal sulcus in each monkey using a T1 volume scan obtained on a GE Signa 1.5 T magnet. Using standard sterile surgical techniques and endotracheal isoflurane general anesthesia we made a 2 cm trephine hole over the intraparietal sulcus and implanted
5.5. Materials and methods

12-16 titanium screws in the monkey’s skull and used them to anchor an acrylic cap in which we placed a head holding device, the recording chamber, and the plug for subconjunctival search coils for eye position recording.

5.5.1 Data collection and exclusion criteria

We used the REX/MEX/VEX system developed at the National Eye Institute’s Laboratory for Sensorimotor Research for behavioral control, visual stimulus display and data collection using Dell Optiplex PC’s running QNX (REX and MEX) and Windows 2000 (VEX). The monkeys sat in a dimly illuminated room with their head fixed and viewed a screen that stood 75 cm away. Visual stimuli were back-projected onto the screen using a LCD projector (Hitachi CP-X275) with a refresh rate of 75 Hz. We used a photodiode to register the actual times for stimulus onsets and offsets. Fixation point and saccade target stimuli were 0.3 degree wide colored squares. Fixation points were red and saccade targets were blue and green. We introduced the 2 separate electrodes per recording session into the same grid separated by a minimum of 2mm through separate guide tubes positioned in a 1 mm grid (Crist Instruments). We recorded single units from each electrode from area LIP with glass-insulated tungsten electrodes (Alpha Omega Engineering, Nazareth, Israel) while the monkeys performed a passive fixation task as white spots flashed sequentially at different locations in the visual field. We amplified, filtered and discriminated action potentials using an amplitude window discriminator (MEX software). Only well-isolated single neurons with highly discriminable waveforms were studied.

5.5.2 Neuron inclusion criteria

We considered neurons to be in LIP if they showed consistent visual, delay-period and saccade related responses during the memory-guided saccade task. For each neuron we isolated, we identified the center of the RF using flashed spots at 400ms intervals (4 per trial, located on a 40 x 40 degree grid with 5 degree spacing, less than 50 ms duration) during passive
5.5. Materials and methods

fixation. We defined the center of the RF as the spatial location of the flashed spot that elicited the maximum activity. Once 2 neurons were independently isolated, we tested each the response of each neuron using memory guided saccade task. We recorded the response to both neurons simultaneously while monkeys made memory guided saccades to a single saccade target. In one block of trials (50 trials), the target was placed in the response field of the first neuron (RF1) and in the second block of trials in RF2. For each block of trials we compared the activity during the delay period to the activity during prior to target onset. Neurons were considered to have sufficient delay period activity if activity was greater during the delay period (t-test, one-tailed p<0.05). Cells pairs were included in the choice task analysis and determined not to share stimulus evoked activity if during the memory guided saccade task, one cell had significant activity 30-300ms after target onset compared to an equivalent bin during the pre-target fixation period, and the other cell did not have an increase (one tailed t-test p>0.05). Un-stimulated cells could have a significant decrease in activity that would not be considered for exclusion by this test.

5.5.3 Task details

Once LIP cells were isolated on each electrode, the monkey was required to perform the free choice foraging task. For each trial in this task, the monkey fixated central red spot for 500 ms, at which point 2 saccade targets appeared simultaneously, one in the RF of each isolated cell. Either target (green or blue) could appear randomly in either RF. The targets were present for 750-1050 ms, at which point the fixation spot disappeared which was the cue for the monkey to choose one target. Monkeys had 400 ms after the go-cue to make a saccade to a 4.5×4.5 degree window around the saccade target. If the monkey’s eye was in the window for 100 ms from 400 ms to 500 ms after the go-cue, a beep indicated whether the monkey would receive a reward: a long beep signaled reward while a short beep indicated no reward. Rewards were determined using a changing relative probability schedule that was changed pseudo-randomly approximately every 200 trials. Reward magnitudes were fixed for the
duration of each session. The range of reward probabilities tested included 3:1,2:1,1:1,1:2,1:3, though not all relative probabilities were tested each session, depending on the number of trials and the monkeys’ satiety. Each target was re-baited (using a random flip of an independent coin for each target) each time that color target was chosen but uncollected rewards carried over across trials so that monkeys could harvest rewards maximally by visiting each target color with the same proportion at which it is rewarded relative to the other color. We did not use a changeover delay. Though monkeys did not always perform this task optimally, they did change their choice strategy when reward probabilities changed during a session, indicating that they had learned that the target reward probabilities had changed. The monkeys’ behavior was quantified by comparing the relative reward ratio (RewardGreen / (RewardGreen+RewardBlue)) to the monkeys’ choice ratio (ChoiceGreen / (ChoiceBlue+ChoiceGreen)). For the monkeys’ instantaneous choice ratio, the choice ratio was averaged over blocks of 10 trials. Saccade latencies for variable reward trials were normalized by the average saccade latency across all trial types for each saccade direction.

For a subset of cell pairs (n=33), we also recorded data for the “empty” RF task. The empty RF task is identical to the free choice task in every respect, except that the locations of the choice targets were changed so that a target appeared in RF1 and the second target appeared diametrically opposite RF1 such that it did not excite either RF1 or RF2. RF2 thus became the “empty” RF and was then no longer a choice option associated with a reward.

5.5.4 Data analysis

All data analysis programs were written in MATLAB (Mathworks Inc, Natick, MA). For the foraging task we examined the relationship between the spike count of each neuron during the choice separately for each saccade direction (saccade into RF1 and saccade into RF2) and reward probability. We used a sliding bin of 300 ms stepped every 25 ms and calculated the spike counts from each cell across the trial. Spike counts were normalized by
5.5. Materials and methods

subtracting the mean spike count from each trial type from the absolute spike count of each trial. We used a sliding average from the surrounding 10 trials of the same saccade trial type to eliminate slower fluctuations in rate that could be caused by slow changes in the monkeys’ alertness. Calculating the Pearson coefficient without this sliding average made no qualitative difference in the results. We also calculated the Pearson coefficient on the z-score of the spike counts rather than the raw spike counts themselves and again this made no qualitative difference in our findings.

We calculated the Pearson correlation of each pair of spike counts separately for each bin across the duration of the trial. The Pearson correlation was computed separately for each saccade direction within a given pair of cells, NOT pooled across saccade directions, which can produce spurious negative correlations. For single cells, error bars were calculated using Jackknife methods leaving out individual trials (iterated 1000 times). For populations of cells, error bars were calculated using standard error of the mean for population averages at each time step. Significant correlations were assessed using a t-test on the distribution of correlation coefficients for each bin independently, while pairwise significance testing in the pre-target and pre-saccadic epochs was done using paired t-test ($p < 0.05$) after testing for normality at each time-step. In order to achieve an appropriate estimate of the correlation, cell pairs were excluded from analysis if individual decisions did not have a minimum of 10 trials for each behavioral condition.

We validated that 300 ms was an appropriate bin to use (Bair et al., 2001; Kohn and Smith, 2005; Smith and Kohn, 2008; Mitchell et al., 2009), by calculating the Pearson correlation using different sized bins slid along the pre-target epoch (500 ms prior to target onset) in the choice task. Bin sizes used were 5, 10, 25, 50, 100, 150, 250, 300, 400, and 500 ms stepped through the duration of the baseline period at 50ms increments. For example, a bin size of 400ms would be calculated 3 times in during the 500ms (starting at -500, -450, and -400).

We calculated synchrony between neurons by taking the spike trains from the pre-target
5.5. Materials and methods

and pre-saccadic epoch and computed a cross average cross-correlogram across trials for each pair. For each pair of neurons, we then recomputed the cross correlogram after shuffling trial order and iterated this process 1000 times. From these shuffled correlograms we extracted 95% confidence bounds. We considered a neuron to have significant synchronous spiking if the value of the spike coincidence was greater than the upper 95% bound for a 3ms window across the 0 time lag.

For the reward history model, we extracted a measure of the reward history by computing a vector of previous rewards where positively rewarded trials were labeled with a 1 and negatively rewarded trials with correct saccades were labeled with a 0. Trials where the monkeys’ saccades were overly inaccurate or exceeded the time limit were excluded. This reward vector was then convolved with an exponential with a variable time constant tau. For each behavioral session, we determined the best-fit tau by regressing the saccade latency for each trial with the reward history term and minimizing the squared residuals minus a tau2 term which penalizes the model from over fitting. We performed this analysis for the total reward stream over all choices and also for choices only to each particular color target as controls. The coefficient of regression between latency and reward history was computed using the best-fit tau for each session. Distributions of coefficients were compared to a predicted mean of 0 using a t-test.

To determine the relationship between correlation and reward history, we binned the trials from each session into 5 equal bins using the best-fit tau for each session and recalculated the noise correlations and average spike counts across trials in each of the bins. Regressions were computed using standard techniques and linear fits were done using least-squares.

5.5.5 Model

We model cortex as a network of $N$ recurrently connected excitatory (E) and inhibitory (I) elements arranged around a 180° ring. Each element is represented by its firing rate, which evolves over time according to the equations (Wilson and Cowan, 1972):
\[
\tau_E \frac{d}{dt} r_E(\theta) = -r_E(\theta) + k \left( [W_{EE} * r_E(\theta) - W_{EI} * r_I(\theta) + c_E h(\theta) + \eta]_+ \right)^n \tag{5.1}
\]

\[
\tau_I \frac{d}{dt} r_I(\theta) = -r_I(\theta) + k \left( [W_{IE} * r_E(\theta) - W_{II} * r_I(\theta) + c_I h(\theta) + \eta]_+ \right)^n \tag{5.2}
\]

In the above equations, \( r_E(\theta) \) and \( r_I(\theta) \) are the firing rates for the E and I neurons at position \( \theta \). Each unit in the network receives both excitatory and inhibitory intracortical synaptic input from the other neurons in the network. The strength of this input is determined by the synaptic connectivity functions \( W_{EE}, W_{IE}, W_{EI}, \) and \( W_{II} \). \( W_{YZ}(\theta - \theta') \) represents the synaptic strength from a cell of type Z (E or I) at position \( \theta' \) to a cell of type Y at position \( \theta \), and depends only on the distance between them. The * signifies spatial convolution (e.g. \( W_{EE} * r_E = \sum_{\theta'} W_{EE}(\theta - \theta') r_E(\theta') \)), where the sum is over all other grid positions \( \theta' \).

Each neuron also receives feedforward input from outside of the network. The feedforward input received by neurons is given by \( h(\theta) \), which represents the shape of the input. It is scaled by two magnitude parameters, \( c_E \) and \( c_I \), which may or may not be equal. Each neuron also receive a constant noisy input, \( \eta \), which we describe in more detail below. The time constants of the E and I cells are \( \tau_E \) and \( \tau_I \). We model the nonlinear input-output function of the cells as a rectified powerlaw (Miller and Troyer, 2002; Priebe et al., 2004), with identical parameters \( n \) and \( k \) for both the E and I cells (Contreras and Palmer, 2003). For generality, we model connectivity as decreasing in strength monotonically with distance on the ring. All four connection types (E ⇒ E, E ⇒ I, I ⇒ E, and I ⇒ I) decay in strength at the same rate around the ring, and vary only in their magnitude. With \( a, b \in \{E, I\} \):

\[
W_{ab}(\theta - \theta') = J_{ab} e^{-\frac{(\theta - \theta')^2}{2\sigma^2}} \tag{5.3}
\]

For all of the simulations presented here, we used the following parameters: \( N = 180, \tau_E = 20 \, \text{ms}, \tau_I = 10 \, \text{ms}, J_{EE} = 0.0441, J_{IE} = 0.04158, J_{EI} = 0.0231, J_{II} = 0.01827, \)
\( \sigma = 32^\circ, \ k = 0.04, \) and \( n = 2.0. \) In this model, spatially localized feedforward stimuli are modeled as Gaussian curves defined in position by the neuron on which they are centered. The width of this Gaussian was set as \( \sigma_{FF} = 15^\circ \) Unless otherwise noted, stimulus strength was equal for E and I.

The noise \( \eta \) injected into each neuron is generated as follows. For both the E and I populations, we take an \( N \times T \) matrix (where \( N \) is the number of cells in the population and \( T \) is the duration of the simulation in time steps) of random values drawn from a Gaussian distribution with a pre-specified mean (\( \mu_\eta = 30 \)) and standard deviation (\( \sigma_\eta = 40 \)). This white noise is then smoothed with Gaussian filters first over time for each neuron and then over space at each time step. For all of the simulations presented here, we use a temporal filter with a width of 25 ms and a spatial filter with a width of 5\(^\circ\). Both filters are normalized to have an integral of 1.

All simulations are run in Matlab, using a forward-Euler integration algorithm with a fixed 1 ms time-step.

In all simulations presented here, a given stimulus presentation paradigm is specified (for example: start with a non-spatial input of value \( x \), and at time = 1,200 turn on a spatial input of magnitude \( y \)), and then 1,000 trials are run. To calculate the correlation between a pair of cells, at each time step we take the two 1,000 element vectors of firing rates (one from each cell), which contains the firing rates across the trials at the given time step, and simply calculate the correlation coefficient between them. When presenting these correlations on plots versus times, the correlation time-series curve is smoothed with a 300 ms sliding average window.
Chapter 6

Conclusions

6.1 Summary

We have presented a simple, yet surprisingly powerful, model that can explain the circuit mechanisms underlying contextual modulation in primary visual cortex. This work began by building off of the original ISN model (Ozeki et al., 2009), which showed that a simple two-neuron linear circuit model in the inhibition-stabilized regime will respond to an external input to I cells with a decrease in the firing rates of both the E and I cells. Using a slightly more elaborate linear model, we then demonstrated how a one-dimensional spatial network comprised of repeating recurrently-connected E-I units can self-consistently generate surround suppression of both the E and I cells through the selective amplification of spatially periodic patterns of activity. The key requirements for this activity are strong, recurrent excitation balanced by feedback inhibition (the ISN regime) as well as long-range excitatory to inhibitory connections that extend further in cortical space than excitatory to excitatory connections. This network, in addition to demonstrating surround suppression, also provides a mechanistic basis for the low frequency tuning of neurons to the envelope spatial frequency of a contrast modulated drifting grating.

By expanding this work and introducing nonlinearity into our model, we then showed
how this same circuit configuration could account for a number of the known contrast-dependent changes in the eCRF. The shrinking of summation field size, the switch from surround facilitation to suppression, and the emergence of spatial periodicity at high contrast all result from a nonlinear transition between dynamic regimes. Because of the expansive input-output nonlinearity, the effective connectivity strength scales with input magnitude. At low contrast, the network operates in an effective non-ISN regime, and contextual interactions are largely facilitatory. Above a certain threshold input strength, the network transitions into the effective ISN regime, and interactions are largely suppressive. This mechanism allows for contrast dependent changes in the modulatory role of the eCRF without requiring an asymmetry in the input-output functions of the excitatory and inhibitory neurons.

Upon further investigation, we found that this same nonlinear ISN model could account for a broad class of circuit behaviors generally described as normalization. Even with completely symmetric connectivity functions, this same circuit motif is able to generate the activity patterns necessary to produce surround suppression. Through a more thorough analysis, we find that all nonlinear circuits of this general prototype, with strong recurrent interactions and an accelerating input-output nonlinearity, should normalize their inputs. This model makes the prediction that for very weak inputs, the circuit ought to actually add inputs supralinearly, rather than sublinearly.

By combining our understanding of contextual interactions across the spatial and orientation domains, we constructed a topographic, probabilistically-connected model of V1, which we then used to explore a variety of eCRF effects observed in vivo. With this model, we are able to reproduce myriad contextual interactions. In addition to recapitulating the effects observed in the simpler models, we find that this model has orientation specific surround suppression with a tuning width that broadens at low center contrast. The model also produces a nearly uniform distribution of surround suppression strengths across neurons, such as has been observed experimentally (Akasaki et al., 2002; Jones et al., 2000). Furthermore, we have been able to use this model to help understand confusing experimental
results. For example, when we explored spatial periodicity experimentally, we were surprised to find that the best-fit spatial frequencies from size- and position-tuning curves were uncorrelated across neurons. When we simulated the same experiments in this more realistically variable network model and fit the simulated size- and position-tuning curves with the same conceptual models, we found the same lack of correlation.

We then introduced various sources of temporally-fluctuating “noisy” input to the network, and find that through winner-take-all normalization, the onset of a feedforward stimulus strongly reduces trial-to-trial variability. This reduction is specific to the component of the variability shared across neurons in the network; as has been observed experimentally, “private” variability is not affected by stimulus onset (Churchland et al., 2010). Using multiple circuit models at different levels of complexity, we confirm that this reduction in variability is a circuit mechanism, and not an artifact of our choice of noise model.

To briefly explore some of the well known effects of attention on neural responses, we developed a framework for studying this additional form of modulation in our network. We modeled the influence of attention on the local circuit as a small additional input to excitatory cells. By tilting the balance of input slightly in favor of E, attention essentially operates through the same mechanism as surround suppression, but in the opposite direction. The small difference in the input to E and I increases the gain of the local circuit, resulting in an increase in the firing rates of both cell types. This simple mechanism can account for a surprising number of previously reported attentional effects in visual cortex.

We then performed multiple experiments to test some of the key assumptions of our circuit model. We probed neurons in V1 for spatially periodic activity by measuring their responses to stimuli of many different sizes, and observed significant oscillations in their size-tuning curves. By translating a large stimulus across retinotopic space, we measured “position-tuning” curves and saw that these as well demonstrated significant spatial periodicity. We measured the tuning for the contrast envelope of a contrast-modulated drifting grating, and found that this tuning increased with increasing levels of luminance grating.
6.2 Perspectives and future directions

All three of these results provide strong evidence for periodicity in spatial contextual interactions, as predicted by our circuit model.

Lastly, we showed how the normalization-induced decrease in shared variability could account for the decorrelation of activity in the lateral intraparietal area (LIP). We demonstrated two distinct pathways for decorrelation. A spatially-specific signal caused a strong decorrelation and was accompanied by a significant change in neural firing, as is observed in LIP after stimulus onset. A spatially-nonspecific input to all the neurons in the network caused only a change in correlation with no change in firing rate, as is observed in LIP during epochs of high recent reward history. We propose that such a spatially-nonspecific input could be carried by a slow, diffuse neuromodulatory input that varies with the monkey’s motivational/satietal state.

6.2 Perspectives and future directions

One major line of research that remains to be addressed concerns the computational function of contextual modulation and normalization. What specific operations are enabled by the various forms of modulation we have modeled, and how is cortex using these operations to process information? One of the major challenges to addressing this issue is the diversity of projection targets of V1. The pyramidal neurons of layer 2/3 project directly and indirectly to several different higher visual areas, each with its own distinct computational function. Before we can assert the computational implications of contextual modulation, it will be necessary to understand what information V1 is providing to these higher regions. Do neurons with different projection targets have distinctly different response properties, and if so, how do they differ? For example, it may be that neurons with strong tuning for the envelope spatial frequency of a contrast-modulated stimulus provide input mainly to neurons in V2. We would then argue that this form of contextual modulation in V1 gives rise to the higher order boundary selectivity characteristic of V2 neurons. Given the diversity of V1 projection targets, it will ultimately require careful experiments linking the physiological
function of individual neurons and populations of cells with anatomy to fully address these questions.

To gain inroads towards understanding the computational functions of contextual modulation and normalization, we will undoubtedly need to consider how this network, and more generally V1, processes natural scenes. Most of the literature on V1 considers the responses to stereotyped, unnatural stimuli like drifting bars and gratings, yet the visual system we study has evolved to respond to the types of images we encounter in our daily lives (or at least use to encounter 10,000 years ago). Understanding the computations enabled by this model will undoubtedly require considering the statistics of natural scenes, and understanding what information from them needs to be passed to higher cortical regions.

In terms of more direct extensions to this work, one major missing piece is a realistic model of neural dynamics. Throughout this work, we mainly consider the steady state of the network, and pay relatively little attention to the temporal structure of responses. In the real visual system, sensory processing is a dynamic, on-going process, and the time-course of neural responses may play a major part in processing sensory information. To this end, it would be worthwhile to expand and complexify this model to include more realistic temporal properties. This would likely require including more realistic neural rate models, such as the mean field model presented in Chapter 3, as well as temporal nonlinearities, such as spike-rate adaptation and short term synaptic plasticity.

Experimentally, there are several interesting ways that we may further validate and elucidate this model. In Chapter 3, we used our nonlinear ISN model to propose a mechanism for normalization in the cortex. We demonstrated that our circuit model, like the phenomenological normalization models (MacEvoy et al., 2009; Busse et al., 2009), sublinearly adds multiple orthogonal and oblique inputs. It is still unknown, however, whether cortex actually adds this type stimuli sublinearly. These experiments (MacEvoy et al., 2009; Busse et al., 2009) make the assumption that stimuli are processed more or less linearly through the retina and LGN, but it is entirely possible (and even probable (Kayser et al., 2001;
6.3. Concluding remarks

Lauritzen et al., 2001; Priebe and Ferster, 2006; Li et al., 2006a)) that cross-oriented stimuli are at least partially suppressed prior to ever reaching cortex. We propose to test for sublinear addition of inputs injected directly into cortex through the use of neurons expressing channelrhodopsin. By projecting distinct patterns of photostimulation directly onto cortex, we could then test definitively for the presence or absence of sublinear addition of multiple stimuli. Additionally, if sublinear addition is observed, we could further test the validity of our circuit model by probing cortex with pairs of weak stimuli to search for a supralinear regime.

Another experiment to directly test the ISN model requires using a channelrhodopsin that is restricted to inhibitory neurons. With this, we could selectively perturb different spatial patterns of inhibitory neurons and observe responses of both the E and I cells in the network. The ISN model predicts that the relative phase of E and I cell firing rates should depend on the spatial frequency of the I cell input pattern, because the strength of recurrent connectivity in the spatial Fourier basis falls off with increasing spatial frequency. Low spatial frequency photo-input should drive the network paradoxically, such that both E and I cell firing rates move in phase. With high spatial frequency photo-input, the I cells should respond nonparadoxically, and show increased firing rates where the E cells decrease their firing rate (Supplementary Figure 7.8).

6.3 Concluding remarks

We have presented a simple circuit model of primary visual cortex that can account for a substantial breadth of experimental data. The key features of this model, extensive recurrent connectivity and expansive input-output nonlinearity, are well-established features of the cortical circuit. We believe that this model represents a general solution to the questions of circuit mechanisms underlying contextual modulation and normalization. As subsequent models build upon this general framework with additional detail and complexity, we believe it may soon be possible to understand more completely the information processing and
6.3. Concluding remarks

computational strategies of visual cortex.
Chapter 7

Supplementary Figures

FIG. 7.1: **Resonant frequency increases with increasing stimulus strength.** The resonant spatial frequencies of both the excitatory and inhibitory populations are calculated from the Jacobian matrix at the fixed point in the one-dimensional nonlinear model of spatial contextual interactions. With increasing stimulus strength, the resonant frequencies of both populations increase. Note that as in the linear model, the resonant frequency of inhibition is always less than the critical frequency.
FIG. 7.2: **Contrast-dependent changes in eCRF valence are insensitive to stimulus shape.** The same two simulations from Figure 3.3 are repeated here, but with Gaussian rather than step-function inputs. The results, a shrinking of summation fields with increasing contrast and a switch from surround facilitation to surround suppression with increasing center contrast, are essentially the same.
FIG. 7.3: 

The switch from facilitation to suppression depends on surround stimulus size. This figure demonstrates the same effect as in Figure 3.3B, but here the size of the surround is varied parametrically while the center size is held constant as the size that exactly fills the summation field of the neuron. Smaller surround stimuli evoke substantial facilitation at low center contrast, but larger surround stimuli are suppressive at all contrasts. The size in the legend indicates the diameter of the center + surround relative to the center alone.
FIG. 7.4: A switch from supralinear to sublinear addition: comparison to MT data. A. In area MT, it has been shown that responses to multiple low-contrast stimuli add supralinearly, but responses to higher-contrast stimuli add sublinearly. Convex contours on this plot indicate supralinear addition, while concave contours indicate sublinear addition. From Heuer and Britten (2002). B. The data from Figure 3.8, replotted to match the MT data. We also observe supralinear addition at low contrast and sublinear addition at high contrast.
FIG. 7.5: Preferred stimulus width shrinks with increasing stimulus strength in the ring model. For very weak inputs, the preferred width of the network is infinite, in that it responds more strongly to a constant input across orientation space than to any Gaussian. Once a threshold strength is reached, the preferred stimulus width decreases monotonically with increasing input strength. This plot expands on Figure 3.9, and shows that the contrast-dependent contraction of orientation tuning width is analogous to the contrast-dependence of summation field size (Figures 3.3 and 7.2).
FIG. 7.6: The switch from supra- to sublinear addition depends on stimulus width. The relative additive weights in response to two equal strength stimuli are plotted as a function of both stimulus width and contrast. Because the network’s preferred stimulus width shrinks with increasing stimulus strength, if normalization is due to surround suppression, the contrast at which the network transitions from a supra-linear to sublinear additive regime should decrease with increasing stimulus width.
FIG. 7.7: Twenty iterations of the iterative approximations of the contrast response curves. For both the 1 stimulus and 2 stimuli conditions, the iterative approximations of the reduced normalization model accurately predict behavior within either the high contrast or low contrast regime. The numerically calculated contrast response curves are plotted in black. The 1 stimulus condition is shown on the left. The first order approximation of the contrast response curve is in red, and subsequent iterations are in progressively lighter shades. The 2 stimuli condition is shown on the right, and iterations go from blue to cyan. The top row uses the high contrast iterative approximation, which goes unstable for low firing rates. The bottom row uses the low contrast iterative approximation, which goes unstable at high firing rates.
Spatial frequency dependent paradoxical response. In the nonlinear model, the spatial frequency paradoxical response as illustrated in Figure 2.6 also depends on the strength of the tonic input to both the E and the I cells. Here we repeat the test from Figure 2.6 in the full-size topographic nonlinear model with both weak and strong tonic stimuli, and both high and low spatial frequency photostimulation. For both of the weak tonic input conditions, the cells respond non-paradoxically to the photostimulus directed to the inhibitory cells, because the network is not in the ISN regime. For the high strength input conditions, we regain the effect of Figure 2.6: low spatial frequency photostimulation drives the cells paradoxically, because the input falls in the ISN regime of spatial frequencies, but high spatial frequency photostimulation does not. The input, excitatory firing rates, and inhibitory firing rates are all normalized to both their minimum and maximum values. The photostimulus was drifted at 3 Hz. For the low input strength conditions, the network was stimulated with a tonic input of strength 1 and the photostimulus was a sinusoid with amplitude 1. For the high input strength conditions, the network was stimulated with a tonic input of strength 40 and the photostimulus was a sinusoid with amplitude 10. The low spatial frequency photostimulus had a frequency of 0.03 cycles/degree, whereas the high spatial frequency stimulus was 0.5 cycles/degree.
FIG. 7.9: **Sublinear addition of multiple stimuli in the full model.** As in Figure 3.8, the network was driven with equal contrast, full-field stimuli with a 90° difference in orientation. Neurons are grouped into 18 equal-width bins based on their preferred orientations, and the firing rate responses within the bins are averaged. Shown here is the mean of the weights calculated for both E and I populations for 25 different pairs of orthogonal stimuli at a variety of stimulus strengths.
FIG. 7.10: A switch from surround facilitation to surround suppression. As in Figure 3.3B, a high contrast surround (here surround stimulus strength = 40) switches from having a facilitatory to suppressive effect on a center stimulus with increasing center stimulus strength. Four example cells are shown. For each, the center stimulus was fit to exactly fill the summation field of the cell, and the center stimulus strength was varied from 0 to 40 in the presence or absence of strong surround stimuli of varying sizes. On average, the smaller surrounds (e.g. 1.5 and 2.0 times the center size) were more likely to be facilitatory at low center strength, while larger stimuli were more generally suppressive. From these examples, one can see the considerable variability amongst the cells, with some cells showing considerable facilitation at low center strength for smaller surrounds, and others showing suppression for all surround sizes.
FIG. 7.11:
A reduction in variability at stimulus onset with a simpler noise model. This figure demonstrates the same effect as in Figure 3.18, but here we use a simpler noise model. The only source of variability in the network is the externally injected, stationary noise process $\eta$. The noise is smoothed with Gaussian filters over space and time.
FIG. 7.12: Variability reduction in non-responsive cells in the simpler noise model. As in Figure 3.19, the reduction in variability is observed in all cells in the network, even those that are relatively suppressed by the stimulus.
FIG. 7.13:
The decrease in variability is a decrease in shared noise in the simpler noise model. A. As in Figure 3.20A, the decrease in variability at stimulus onset depends on spatial correlations in the noise. Spatially white noise is relatively unaffected by stimulus onset. B. As in Figure 3.20B, factor analysis of a subset of cells that experience little to no change in firing rate at stimulus onset reveals that the reduction in variability is primarily a reduction in the shared network noise, and the level of private noise is unchanged by stimulus onset.
FIG. 7.14: The mean field model sublinearly adds multiple stimuli. As in the simple ring model presented in Chapter 3, the more complex model based on the finite-size mean-field approximation of a network of constant-leak integrate and fire neurons also produces our benchmark of normalization, the sublinear addition of two orthogonal stimuli.
FIG. 7.15: The decrease in variability is a decrease in shared noise in the complex model. As in Figures 3.20 and 7.13, factor analysis of a subset of cells that experience little to no change in firing rate at stimulus onset reveals that the reduction in variability is primarily a reduction in the shared network noise, and the level of private noise is relatively unchanged by stimulus onset.
FIG. 7.16:
In the complex model, the reduction in variability at stimulus onset does not require spatially correlated external noise. The basic stimulus onset paradigm is repeated, but here the external noise injected into the network is not passed through a spatial filter. The model nonetheless shows a reduction in variability for both stimulus-preferring and stimulus-non-preferring cells.
FIG. 7.17:
In the complex model, the reduction in variability at stimulus onset does not require external noise. The basic stimulus onset paradigm is repeated, but here the external input into the network is constant over space and time. All of the variability is generated within the network, and this internally generated variability is still suppressed at stimulus onset.
FIG. 7.18: Slow fluctuations in cortical excitability do not obscure stimulus selectivity. Here we plot a firing rate curve from Figure 4.1 while highlighting four 100 second time bins. Firing rates in each subplot are binned by stimulus presentation epoch. During periods of both high and low cortical excitability, the cell can be seen to still respond selectivity to stimuli. Note the changing scale on the y-axes of the subplots.
FIG. 7.19: Boot strap analysis of size-tuning model fits. A. Three examples of the boot-strap fitting process are shown on the right. The original SSM curve, fit with all 30 data points, is in green. The 100 boot-strap fits, each fit with 24 randomly selected points, are shown in blue. B. A log-log histogram of the normalized difference between the boot-strap fits and the full data fit shows a powerlaw distribution, indicating that most boot-strap fits differ very little from their respective full-data fit.
FIG. 7.20:
**Boot strap analysis of position-tuning model fits.** A. Three examples of the boot-strap fitting process are shown on the right. B. A log-log histogram shows the normalized difference between the boot-strap fits and the full data fit.
FIG. 7.21:
Correlations between best fit frequencies in both the model and experiment. A. The best fit frequencies from size-tuning, position-tuning, and contrast modulation spatial frequency tuning in the experimental data. There is no correlation between any pair. The calculated correlation coefficient and the p-value of the correlation are indicated on each plot. B. The same three tests were run on the large-scale probabilistic model. As in the data, there is no significant correlation between any pair of the three spatial spatial frequencies.
Average correlation coefficient as a function of bin size in pre-target epoch. Average correlation coefficient as a function of bin size in pre-target epoch. Bin sizes used were 5, 10, 25, 50, 100, 150, 250, 300, 400, and 500 ms stepped through the duration of the baseline period at 50ms increments. Error bars show SE across pairs of neurons (n=67 pairs of neurons).
Correlations decrease with receptive field distance. A. Correlations decrease as a function of the distance between the response field centers of neurons in the monkey for both the pre-target epoch (blue) and the pre-saccadic epoch (red). Distances shown are the distances in visual degrees between the choice targets. B. Correlations decrease as a function of the distance along the ring between “recorded” neurons. Conventions as in A. Distance units here are arbitrary, with 90 representing the furthest distance on the ring. Close distances on this plot represent neurons with overlapping RFs (the RFs in the model have a half-width at half height of 12°).
Bibliography


