Title
The role of balanced excitation and inhibition in cortical circuits

Permalink
https://escholarship.org/uc/item/3nm5m98r

Author
Murphy, Brendan Keith

Publication Date
2007-12-20

Peer reviewed|Thesis/dissertation

The role of balanced excitation and inhibition in cortical circuits

by

Brendan K. Murphy

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOPHYSICS

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

UMI Number: 3289349

UMI®

The role of balanced excitation and inhibition in cortical circuits

For my family.

Acknowledgments

When I decided to go to graduate school, I could not have hoped for a better environment than the one I found at UCSF. This place is so full of great people, and I have received so much help given so openly that I can't possibly thank everyone by name. Intellectually and scientifically, I have had some of the most exciting times of my life here, but also my fair share of frustration. I thank my wife Sarah for sharing my excitement about new ideas and for her support throughout. At times, I think she was just dragging me along, when I'd lost sight of the good parts of graduate school. She is also almost single-handedly responsible for the proper use of commas in everything I've written during my graduate career. I don't know what I'd do without her. I thank my family for always telling me that I was smart and that I could do the things I wanted to do. The interest and encouragement of my parents and siblings is one of the things that has motivated me the most at every stage of my life, and graduate school has been no different.

I'm very fortunate to have had Ken as an advisor. When I first joined his lab, I knew essentially nothing about neuroscience. His patient explanations of everything from the architecture of visual cortex to the Jordan normal form have been invaluable to me. His curiosity and desire to deeply understand the scientific problems he works on are a source of inspiration. I also thank him for being so accommodating when he moved to Columbia several years ago, and for allowing me to continue working at UCSF where my wife was also in graduate school. Even from the opposite coast, Ken was more responsive and available than I had any right to expect.

I thank the Miller lab. I've enjoyed many conversations both scientific and non-with Michael, Thomas, Gabor and Sergei over the years. The folks at Columbia, especially Aylin and George, have always made me feel welcome when I visit and have made Columbia a great place to work, positively brimming with theorists. I've also had a great time and many good discussions with members of the Abbott lab at Columbia, especially Kanaka and Tim. I thank Larry Abbott for many very creative suggestions and discussions about my work. His ability to immediately see the most fundamental aspects of a problem has astonished me at times. I thank Flip Sabes and Michael Stryker for serving on my thesis committee, and the Stryker lab for including me in their lab meetings after Ken moved to Columbia.

As an undergraduate I worked for several summers in the lab of Michael Prisant.

I thank him for his friendship and for convincing me that I should go to graduate school. My experience in his lab showed me just how fun and interesting science can be.

I thank the many Biophysics students who have mustered genuine interest in my neuroscience talks despite it being outside their normal field of expertise. In particular I thank my good friend and fellow Biophysics student Justin Bradford, without whom graduate school would have been a lot less fun.

Lastly, I thank Julie Ransom. Her ability to imperturbably ask me for the third time for my poster title can only be the result of a genuine concern for the welfare and success of all Biophysics students. I've always been impressed by her competence and intelligence, and I thank her for looking out for me all these years.

. . .

The content of Chapter 2 was published in the *Journal of Neuroscience.* The content of Chapter 3 has been submitted to *Nature.* Ken Miller co-authored these two papers and supervised all of the research in this dissertation.

Abstract

The role of balanced excitation and inhibition in cortical circuits

by

Brendan K. Murphy

Doctor of Philosophy in Biophysics

University of California, San Francisco

Professor Kenneth D. Miller, Chair

Cortical networks are thought to operate in a state of tightly balanced excitation and inhibition. A typical cortical neuron receives thousands of synaptic inputs from other cortical neurons. The majority of these inputs are excitatory and the network is stable only because strong recurrent excitation is balanced by similarly strong feedback inhibition. Theoretical models have shown that sparsely connected networks with strong, balanced excitation and inhibition exhibit chaotic activity that is consistent with the highly variable responses and large membrane potential fluctuations observed in cortex *in vivo*. In these models, and hypothetically in cortical networks, large noise is a consequence of large excitatory inputs being balanced by inhibitory inputs, keeping the mean membrane potential below threshold but leaving large fluctuations about this mean. Direct experimental evidence for this balanced state has been provided by measurements of synaptic conductances *in vivo* and in cortical slices. In chapter two we show that noisy fluctuations in membrane potential are an essential part of a mechanism by which neurons can modulate the gain of their responses. Gain modulation is common in cortical responses and is thought to play an important role in cortical computation.

Membrane potential fluctuations during spontaneous activity are not entirely random. In cat primary visual cortex spontaneous activity has been shown to exhibit spatial patterns similar to those evoked by a visual stimulus, in which neurons with similar preferred orientations are co-active. This suggests that, from unstructured input, cortical circuits selectively amplify activity patterns related to normal function. Current understanding of such amplification involves elongation of the lifetime of a neural pattern by mutual synaptic excitation among the neurons involved. In chapter three we describe a new mechanism

by which recurrent networks with strong recurrent excitation and balancing feedback inhibition can amplify neural activity patterns without elongation of lifetime, which we call transient amplification. By better understanding the implications of balanced excitation and inhibition in cortical networks, we gain insight into the role of recurrent circuitry in cortical computation.

Professor Kenneth D. Miller
Dissertation Committee Chair

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

An intricate network of neural circuits in the cerebral cortex carries out many of the computations that make mammalian brains so interesting and humans conscious and intelligent. The work in this dissertation aims to contribute, at least in a small way, to the theoretical understanding of the properties of cortical circuitry that make these computations possible. A substantial portion of cortex is devoted to the processing of sensory information and I will focus mostly on primary visual cortex (V1), which has been intensively studied for many years. Many of the features of V1 are common in other cortical areas and thus results obtained in V1 are likely to have implications for cortical computation in general. V1 is the first area of cortex to receive visual information from the retina, by way of the Lateral Geniculate Nucleus (LGN). In turn, V1 projects to many other cortical areas involved in processing visual information.

The properties of neurons in both V1 and the LGN are commonly characterized by a receptive field, which is a region of the visual field where an appropriate visual stimulus will cause a neuron to respond by firing action potentials. The primary distinction between the responses of neurons in the LGN versus those in V1 is that V1 neurons are highly selective for the orientation of visual stimuli. Neurons in V1 respond strongly to elongated bar-like stimuli of a particular orientation, e.g. vertical or horizontal (Hubel and Wiesel 1962) (Fig. 1.1). LGN neurons are selective for the contrast of visual stimuli, e.g. light or dark, but not for orientation.

In many animals (cats, ferrets, monkeys and others, but puzzlingly not rats, mice,
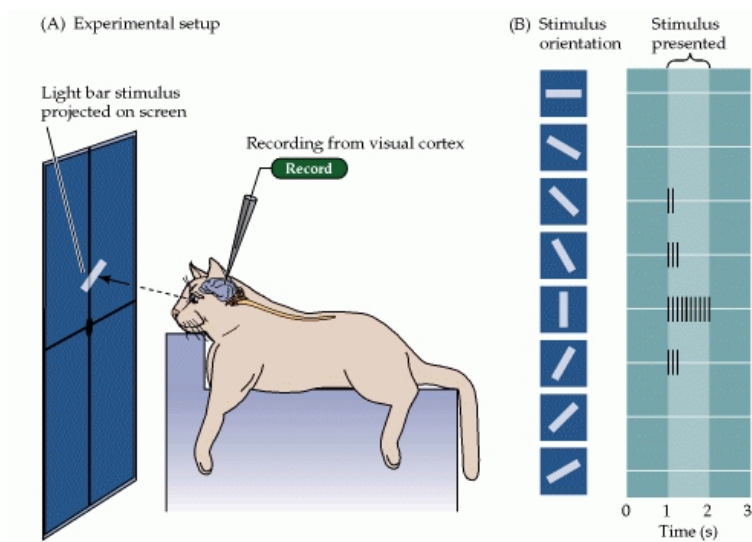
Figure 1.1: **Receptive fields in V1.** Panel A shows a typical setup for characterizing receptive fields in cat V1. The figure is taken from Purves (2001). Neurons in V1 respond strongly to a bar of the correct orientation (vertical in this case) but not to bars of other orientations, as shown in panel B.

or squirrels) cortex is organized into "columns" (Hubel and Wiesel 1962). Neurons that are nearby each other on the cortical surface tend to prefer similar orientations and this preference does not change through the depth of cortex, perpendicular to the surface. As you move "horizontally" across the surface, the orientation preferences of neurons generally change smoothly. At certain junctions, structures known as "pinwheels" occur where neurons of many different orientations can be found close together. Using optical imaging it is possible to record activity across large areas of the cortical surface and construct detailed orientation maps that label each part of the surface according to the preferred orientation of the neurons found there, as shown in Figure 1.2. As a general rule a roughly 1 mm^2 patch of cortical surface will contain neurons representing all orientations.

In depth, V1 is organized into six layers (Callaway 1998). Most inputs from the LGN are targeted to layer 4 where they make up a significant portion of the total input in terms of the influence on spike output, if not the total number of synapses (Ahmed et al. 1994, Chung and Ferster 1998, Ferster and Miller 2000, Ferster et al. 1996, LeVay and Gilbert 1976, Peters and Payne 1993, Reid and Alonso 1995, Tanaka 1983). Orientation selectivity arises from the feedforward arrangement of these LGN inputs onto V1 neurons
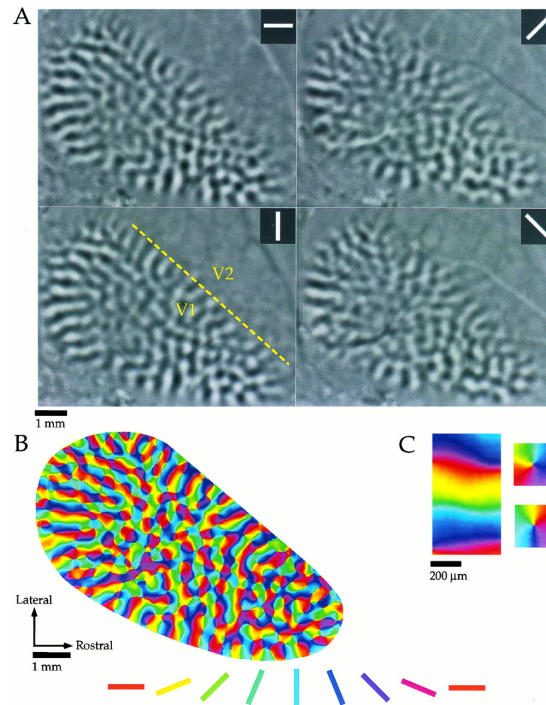
Figure 1.2: **Orientation map in V1.** An example orientation map from V1 in tree shrew obtained from optical imaging. The figure is taken from Bosking et al. (1997). The images in panel A are looking down at the surface of V1 during the presentation of four orientated visual stimuli (0°, 45°, 90° and 135°). This causes areas with neurons preferring each orientation to become active (dark areas). Panel B shows an orientation map obtained by combining the data in panel A. Panel C shows enlargements of both pinwheels and areas with smoothly changing orientation preferences.

in this layer (Ferster and Miller 2000, Hubel and Wiesel 1962, Reid and Alonso 1995), though selectivity may be enchanced by recurrent cortical interactions (Ben-Yishai et al. 1995, Somers et al. 1995, Sompolinsky and Shapley 1997). Neurons in other layers receive input primarily from other V1 neurons. Even in layer 4, intra-cortical synapses are a large source of input (Ahmed et al. 1994, Callaway 1998, Gilbert and Wiesel 1979, LeVay and Gilbert 1976, Peters and Payne 1993). Most of these synaptic connections are made over a fairly short horizontal distance of roughly 200 μm. Outside of layer 4, excitatory synapses can extend over longer distances making patchy connections between areas with similar functional reponse properites (Gilbert and Wiesel 1989).

The role of the recurrent circuitry in cortex and, in particular, the balance between excitatory and inhibitory inputs, is the subject of much of this dissertation. Intracellular recordings of membrane potentials and conductances *in vivo* suggest that cortical neurons operate in a high conductance state in which large excitatory and inhibitory synaptic conductances are balanced (Anderson et al. 2000b, Destexhe et al. 2003, Haider et al. 2006, Paré et al. 1998, Steriade et al. 2001) (but, see Waters and Helmchen 2006). In addition, cortical activity *in vivo* is often highly variable, with irregular spike times and large subthreshold fluctuations in membrane potential (Anderson et al. 2000b, Shadlen and Newsome 1994). The irregular activity of cortical neurons is not consistent with the integration of many small excitatory inputs (Softky and Koch 1993). Rather, irregular spike times are the result of neurons having a subthreshold mean membrane potential and firing when flucutations in the membrane potential cross threshold (Shadlen and Newsome 1998, Troyer and Miller 1997). Theoretical results show that this kind of activity can be generated in sparsely connected networks with strong, balanced excitation and inhibition (Brunel 2000, van Vreeswijk and Sompolinsky 1996). Similar results can be obtained in models of single neurons with large, fluctuating excitatory and inhibitory background conductances (Destexhe and Paré 1999, Destexhe et al. 2001). In all of these models, excitation and inhibition are balanced on average, keeping the mean membrane potential below the threshold for spiking but leaving large fluctuations about this mean. Together with the direct experimental measurements of conductances, these theoretical results provide strong evidence that cortical networks operate in a state of tightly balanced excitation and inhibition. Further evidence to support this idea comes from the observation that supression of inhibition in cortex results in instability and runaway activity (Chagnac-Amitai and Connors 1989).

It is not obvious why cortical networks operate in such a noisy and carefully

balanced state but there are a number of theories about why it might be advantageous. Van Vreeswijk and Sompolinsky (1998) have shown that this state allows the network to respond more quickly to inputs and noise has been shown to be important for the propagation of firing rate signals through spiking neural networks (van Rossum et al. 2002, Vogels and Abbott 2005). In this dissertation, we discuss two additional implications of this feature of cortical networks. We show in Chapter 2 that noisy membrane potential fluctuations are important for generating multiplicative changes in neural gain (see also Chance et al. 2002). In Chapter 3 we introduce another consequence of the balanced state: balanced networks robustly support a mechanism of input amplification previously undescribed in neural systems, which we call transient amplification.

1.2   Neuron Models

The results in this dissertation are all theoretical and based on models of cortical activity. At one end of the spectrum of possible models are complicated multi-compartment models that involve detailed descriptions of both the geometry of the cell body and the neurites and the placement of ion channels on the cell membrane. Although these models can be very realistic, they involve a large number of parameters and are computationally expensive to simulate. At the other end of the spectrum are simple linear rate models that describe a neuron using a single number to represent its firing rate at any given time. The models used in this dissertation are largely in the less detailed end of the spectrum and are briefly described below.

1.2.1   Spiking models

A very common model of cortical neurons is the conductance based integrate-and-fire (IAF) model. It has the advantage of being computationally efficient, making it possible to easily simulate large networks. For example, Chapter 3 describes a simulation consisting of 50,000 IAF neurons. This model is relatively simple and has a single compartment, ignoring details of cell morphology such as dendrites or axons.

The membrane potential (V) of the IAF model neuron is governed by the equation:

$$\tau_m \frac{\mathrm{d}V}{\mathrm{d}t} = (V_{\text{leak}} - V) + \sum_i g_i(V_i - V) \tag{1.1}$$

Here τ_m is the resting membrane time constant, defined as $\frac{C}{g_\text{leak}}$ where C is the capacitance of the cell membrane and g_leak is the leak conductance. V_leak is the reversal potential of the leak conductance, the g_i are synaptic conductances in units of the leak conductance, and the V_i are the corresponding reversal potentials. We can define $V_\infty = \frac{V_\text{leak} + \sum_i g_i V_i}{1 + \sum_i g_i}$ and rearrange things to obtain:

$$\tau_\text{eff} \frac{\mathrm{d}V}{\mathrm{d}t} = V_\infty - V \tag{1.2}$$

where $\tau_\text{eff} = \frac{\tau_m}{1 + \sum_i g_i}$, the effective membrane constant when taking into account synaptic conductances. Now it's easy to see that the membrane potential of the neuron simply approaches the value of V_∞ with time constant τ_eff. Real neurons fire action potentials when the membrane potential reaches a certain threshold. IAF models do not simulate the voltage gated sodium and potassium channels responsible for action potential generation. Instead they simply record that an action potential has occurred when the membrane potential reaches threshold and then reset the membrane potential to a sub-threshold value, possibly then holding it there for a refractory period.

Eq. 1.1 must be integrated numerically. Throughout this dissertation, I use a second order algorithm described in Shelley and Tao (2001). This integration scheme corrects for numerical errors introduced when the membrane potential is reset after a spike. Simpler first order integration, such as Euler's method, will also work, but a smaller timestep is required to achieve similar accuracy.

More realistic spiking models simulate the generation of action potentials explicitly by incorporating voltage gated sodium and potassium conductances, as in the Hodgkin-Huxley model. For this type of model, and anything more complicated, the NEURON simulation environment (Hines and Carnevale 1997) is very useful. NEURON allows you to build very detailed models with multiple compartments, incorporating many different kinds of conductances and cell geometries.

1.2.2 Rate models

It can be difficult to tell why a complicated numerical model, like the network of IAF neurons in Chapter 3, produces the results that it does. Rate models are arguably less realistic, but they are much more amenable to mathematical analysis and depend on fewer assumptions about the detailed properties of real neurons.

In a rate model, the details of spiking are ignored and a neuron is described by its

firing rate as a function of time $r(t)$. In a short time interval Δt, the probability that the neuron will fire an action potential is given by $r\Delta t$. The firing rate of the neuron approaches some function of it's input, $F(I)$, with time constant τ:

$$\tau\frac{d\mathbf{r}}{dt} = -\mathbf{r} + \mathbf{F}(\mathbf{Wr} + \mathbf{h}) \tag{1.3}$$

Here, $\mathbf{r}$ is an N-dimensional vector representing the firing rates of a population of N neurons (the ith element r_i is the firing rate of the ith neuron). $\mathbf{W}$ is an $N \times N$ synaptic connectivity matrix (W_{ij} is the strength of connection from neuron j to neuron i). $\mathbf{Wr}$ represents input from other neurons within the network and $\mathbf{h}$ represents external inputs to the network. The vector $\mathbf{F}$ is just the firing-rate function $F(I)$ operating on each part of the input vector $\mathbf{I} = \mathbf{Wr} + \mathbf{h}$, such that $\mathbf{F}(\mathbf{I}) = (F(I_1), \ldots, F(I_N))^T$. The time constant τ is related to the synaptic and membrane time constants of the individual neurons.

The model can be simplified further by assuming that $F(I)$ is linear. Although firing-rate functions for real neurons are clearly not linear for all levels of input, this approximation is often acceptable for certain ranges of firing rate over which $F(I)$ is nearly linear. Alternatively, a linear approximation may be sufficient to describe the dynamics of relatively small fluctuations around a fixed point of the full non-linear system. In either case the activity of a network of N neurons can be described by a simple linear vector equation:

$$\tau\frac{d\mathbf{r}}{dt} = -\mathbf{r} + \mathbf{Wr} + \mathbf{h} \tag{1.4}$$

The fact that these equations are linear greatly simplifies mathematical analysis and makes available all the tools of linear algebra. For example, the rates seem to depend in a complicated way on the connectivity of the network, specified by $\mathbf{W}$. But because the equations are linear they can be solved by changing coordinate systems to a basis consisting of the eigenvectors of $\mathbf{W}$. In this basis Eq. 1.4 is reduced to a set of N independent, one dimensional equations that are then easily solved.

We show in Chapter 3 that the dynamics of a simple linear rate model are surprisingly similar to that of a much more complicated network of integrate-and-fire neurons when the network is responding to a steady, statistically stationary input. In general when the activities of neurons in a network are not highly correlated and spiking is asynchronous the dynamics of a spiking model can be approximated well by a rate model, though not necessarily a linear one (Brunel 2000, Lerchner et al. 2006, Shriki et al. 2003, Sompolinsky and White 2005).

Chapter 2

Multiplicative Gain Changes

The material in this chapter was published in the *Journal of Neuroscience* as:

Brendan K. Murphy[1] and Kenneth D. Miller[1,2]

[2]Depts. of Physiology and Otolaryngology

[1]Biophysics Graduate Program

[1]W.M. Keck Center for Integrative Neuroscience

[2]Sloan-Swartz Center for Theoretical Neurobiology at UCSF

University of California

San Francisco, CA 94143-0444

Email: murphy@phy.ucsf.edu, ken@phy.ucsf.edu

Abstract

We model the effects of excitation and inhibition on the gain of cortical neurons. Previous theoretical work has concluded that excitation or inhibition alone will not cause a multiplicative gain change in the curve of firing rate vs. input current. However, such gain changes *in vivo* are measured in the curve of firing rate vs. stimulus parameter. We find that when this curve is considered, and when the non-linear relationships between stimulus parameter and input current and between input current and firing rate *in vivo* are taken into account, then simple excitation or inhibition alone can induce a multiplicative gain change. In particular, the power-law relationship between voltage and firing rate that is

induced by neuronal noise is critical to this result. This suggests an unexpectedly simple mechanism that may underly the gain modulations commonly observed in cortex. More generally, it suggests that a smaller input will multiplicatively modulate the gain of a larger one when both converge on a common cortical target.

2.1   Introduction

Gain modulation is a roughly multiplicative or divisive change in a neuron's tuning curve to one stimulus parameter as some other parameter or state is modified. Such gain changes are frequently observed in the responses of cortical neurons, and are thought to play an important role in neural computations (reviewed in Salinas and Thier 2000). A particularly well studied example exists in monkey posterior parietal cortex, where the responses of neurons to the retinal position of a visual stimulus are multiplicatively scaled by eye position (Andersen and Mountcastle 1983, Andersen et al. 1985). Similar modulation of responses by eye position is seen in a variety of visual areas (Boussaoud et al. 1993, Bremmer et al. 1997b, Galletti and Battaglini 1989, Trotter and Celebrini 1999). This gain modulation has been proposed to underlie coordinate transforms computed by these neurons, which are necessary for visually guided reaching (Pouget and Sejnowski 1997, Pouget and Snyder 2000, Salinas and Abbott 2001, 1995, Zipser and Andersen 1988). Gain modulations have also been observed in the enhancement of neural responses by attention. McAdams and Maunsell (1999a) showed that attention can increase the gain of the orientation tuning curves of neurons in areas V1 and V4 of macaque visual cortex, while Treue and Martinez-Trujillo (1999) found that attention increases the gain of direction tuning curves in macaque area MT. Modulation of responses in V1 by stimuli outside the classical receptive field appears to be divisive in character(Cavanaugh et al. 2002, Muller et al. 2003, Palmer and Nafziger 2002). Gain modulation can also be induced pharmacologically: Fox et al. (1990) found that by iontophoretically applying NMDA to neurons in cat V1, they could increase the gain of the neuron's contrast response curve.

Despite the apparent importance of multiplicative gain modulation in the cortex, the mechanisms responsible for producing such gain changes are not well understood (but, see Chance et al. 2002, Doiron et al. 2001, Fox and Daw 1992, Mel 1993, Salinas and Abbott 1996, Smith et al. 2002, Srinivasan and Bernard 1976). In particular, it has been concluded that simple excitation or inhibition alone cannot achieve a gain change (*e.g.* Chance et al.

2002, Holt and Koch 1997), except at low firing rates (Doiron et al. 2001). Instead, it has recently been shown that concurrent, balanced increases in background excitation and inhibition together, which cause an increase in current noise and in conductance with no net depolarization or hyperpolarization, can serve to divisively decrease gain (Chance et al. 2002). These conclusions were based on examining the gain of the relationship between injected current and firing rate. However, multiplicative gain changes in cortex *in vivo* are observed in the relationship between a *stimulus parameter* and firing rate. Here we consider the non-linear relationship between stimulus parameter and injected current, as well as the non-linear relationship between injected current and firing rate. We show that multiplicative gain changes arise robustly from the simple addition of excitation or inhibition alone, provided the modulating excitation or inhibition is small relative to the peak of the tuning curve of the driving excitation. That is, the observed cortical gain changes can be induced if the modulating influence simply adds or subtracts excitation or inhibition.

An important part of our model is the large background synaptic conductances to which neurons are subject *in vivo*, which give rise to a noisy sub-threshold membrane potential (Anderson et al. 2000b, Destexhe and Paré 1999). A noisy sub-threshold membrane potential in turn gives rise to an expansive power law relationship between the average membrane potential and the firing rate of a neuron (Hansel and van Vreeswijk 2002, Miller and Troyer 2002). This non-linear relationship between voltage, or input current, and firing rate, along with the non-linear relationship between stimulus parameter and input current, together cause excitation or inhibition alone to yield roughly multiplicative gain changes in neuronal responses. We demonstrate this using both numerical simulations and a simple analytical model.

2.2 Methods

We simulate a cortical neuron using two models: a conductance-base, integrate-and-fire model and a Hodgkin-Huxley type model.

2.2.1 Integrate-and-Fire Model

The integrate-and-fire model is described by the following equation:

$$C\frac{\mathrm{d}V}{\mathrm{d}t} = g_{\text{leak}}(E_{\text{leak}} - V) + \sum_i g_i(E_i - V) + I_{\text{inj}} \tag{2.1}$$

where C is the capacitance, g_{leak} is the leak conductance, E_{leak} is the resting membrane potential, I_{inj} is the injected current, and the g_i are conductances with corresponding reversal potentials E_i. When the voltage reaches the spike threshold V_{thresh}, it is reset to V_{reset} and held there for a refractory period t_{refract}.

The parameters for the integrate-and-fire model and its synaptic and noise currents were selected to match cortical properties, primarily in the course of previous work (Krukowski and Miller 2001, Palmer and Miller 2002, Troyer and Miller 1997, Troyer et al. 1998) but with minor adjustments in the present work. In particular the parameters were designed without reference to (and before obtaining) the results presented in this paper. The values for the following parameters are the same in each simulation: $g_{\text{leak}} = 10nS$, $C = 488pF$, $E_{\text{leak}} = -70mV$, $V_{\text{thresh}} = -54mV$, $V_{\text{reset}} = -60mV$, and $t_{\text{refract}} = 1.7ms$. The value of C was chosen so that, after taking into account additional background (noise) conductances described below, the resting membrane time constant is $20mS$. This is consistent with values of $15 - 24ms$ observed *in vivo* for cortical neurons (Hirsch et al. 1998).

There are two excitatory synaptic conductances, NMDA and AMPA, and two inhibitory synaptic conductances, GABA$_{\text{A}}$ and GABA$_{\text{B}}$. Their reversal potentials are $E_{\text{NMDA}} = 0mV$, $E_{\text{AMPA}} = 0mV$, $E_{\text{GABA}_{\text{A}}} = -70mV$, and $E_{\text{GABA}_{\text{B}}} = -90mV$. The NMDA conductance is voltage-dependent in accordance with the model described in Jahr and Stevens (1990), using $[\text{Mg}^{++}] = 1.2$ mM. The voltage we use to compute the NMDA conductance is the "shadow voltage", V_s. V_s is obtained by integrating the membrane potential continuously in time in the absence of a spike threshold, *e.g.* it is not reset when it reaches the spike threshold. This is meant to approximate the potential experienced by NMDA channels located on the dendrites of the neuron, and to eliminate discontinuities in the conductance at spike times. This model yields an NMDA conductance that is 3.7% of maximum at the resting potential of the neuron (where maximum is defined to be the conductance at a voltage of 100 mV) and 10.6% of maximum when $V_s = -51mV$ (the largest in this study).

The time courses of AMPA, GABA$_{\text{A}}$, and GABA$_{\text{B}}$ conductances following presynaptic action potentials are modeled as a difference of single exponentials:

$$g(t) = \sum_{\Delta t_j} \bar{g}\left(e^{-\Delta t_j/\tau^{\text{fall}}} - e^{-\Delta t_j/\text{Tr}}\right) \tag{2.2}$$

Here Δt_j is defined as $(t - t_j)$, where t_j is the time of the jth pre-synaptic action potential and $t_j < t$. Time constants are Tr$_{\text{AMPA}} = 0.25ms$, $\tau^{\text{fall}}_{\text{AMPA}} = 1.75ms$, Tr$_{\text{GABA}_{\text{A}}} = 0.75ms$,

$\tau_{\text{GABA}_\text{A}}^{\text{fall}} = 5.25ms$, Tr $_{\text{GABA}_\text{B}} = 40ms$, and $\tau_{\text{GABA}_\text{B}}^{\text{fall}} = 80ms$. Parameters for GABAergic synaptic conductances are set to roughly match experimental data (Benardo 1994, Connors et al. 1988). NMDA conductances decay as a double exponential with a fast and slow component:

$$g(t, V_s) = \sum_{\Delta t_j} \bar{g}(V_s)\Big(f_{\text{fast}}e^{-\Delta t_j/\tau_{\text{fast}}^{\text{fall}}} + f_{\text{slow}}e^{-\Delta t_j/\tau_{\text{slow}}^{\text{fall}}} - e^{-\Delta t_j/\text{Tr}}\Big) \tag{2.3}$$

Parameters for NMDA conductances are taken from experimental data for adult rats (Carmignoto and Vicini 1992): $\tau_{\text{fast}}^{\text{fall}} = 63ms$, $\tau_{\text{slow}}^{\text{fall}} = 200ms$, $f_{\text{fast}} = 0.88$, and $f_{\text{slow}} = (1 - f_{\text{fast}})$. We set Tr $_{\text{NMDA}} = 5.5ms$ to match the experimentally observed 10-90% rise time for NMDA receptor mediated post synaptic currents (Lester et al. 1990).

The size of synaptic conductances evoked by pre-synaptic action potentials are set in terms of their total conductance integrated over time in units of $nS \cdot ms$. The relative strengths of AMPA and NMDA conductances are set to match those observed in thalamocortical slices (Crair and Malenka 1995). This, along with $[\text{Mg}^{++}] = 1.2$ mM, yields the result that, at V_{thresh}, the time integrated conductance for NMDA is 2.57 times that of AMPA. The AMPA conductance evoked by a single excitatory pre-synaptic action potential is set to 2.8 $nS \cdot ms$. At V_{thresh}, the evoked NMDA conductance is 7.2 $nS \cdot ms$. The GABA$_\text{A}$ conductance resulting from a single inhibitory pre-synaptic action potential is set to 8 $nS \cdot ms$, and the GABA$_\text{B}$ conductance to 2 $nS \cdot ms$.

In addition to the above described synaptic conductances, the model includes two fluctuating background conductances, an inhibitory conductance $g_I(t)$ with $E_{\text{rev}} = -80mV$ and an excitatory conductance $g_E(t)$ with $E_{\text{rev}} = 0mV$. These conductances are meant to simulate the background synaptic input received by cortical neurons *in vivo*. They are produced by an Ornstein-Uhlenbeck process, as described in Destexhe et al. (2001):

$$\frac{dg(t)}{dt} = \frac{g_0 - g(t)}{\tau} + \chi(t)\sqrt{\frac{\sigma^2}{\tau}} \tag{2.4}$$

where g_0 is the mean conductance, τ is a noise time constant, σ^2 is the variance of the conductance, and $\chi(t)$ is a Gaussian random variable with zero mean and a standard deviation of one. Parameters were chosen by beginning with the parameters used in Palmer and Miller (2002), and adjusting these to produce membrane potential fluctuations of about 5mV at rest and a rest potential of about -70mV. This is in accordance with recordings from cortical cells *in vivo* (Anderson et al. 2000b). For the inhibitory background conductance,

$g_0 = 12.0nS$, $\tau = 34.1ms$, and $\sigma = 4.3nS$. For the excitatory background conductance, $g_0 = 2.4nS$, $\tau = 34.1ms$, and $\sigma = 2.4nS$. The noise time constant of 34.1 ms was chosen to cause the power spectrum of the voltage noise to match that seen in experimental data generously shared with us by Jeff Anderson and David Ferster (this matching was performed by S.E. Palmer in the lab of KDM, Palmer and Miller 2002). The mean background conductance, combined with g_{leak}, gives the cell a resting input resistance (R) of $41M\Omega$. The capacitance (C) is set to give the cell a resting membrane time constant ($\tau_m = RC$) of 20ms.

2.2.2   Hodgkin-Huxley Model

Simulations of a more biophysically detailed single compartment model were produced using the NEURON simulation environment(Hines and Carnevale 1997). This model includes fluctuating background conductances produced by an Ornstein-Uhlenbeck process, voltage dependent sodium and potassium conductances to model action potentials, and a non-inactivating potassium conductance responsible for spike frequency adaptation as described in Destexhe et al. (2001) for a single compartment neuron model. Parameters for the model were taken from Destexhe et al. (2001): $g_{\text{leak}} = 0.045mS/cm^2$, $E_{\text{leak}} = -80mV$, $C = 1\mu F/cm^2$. For the inhibitory background conductance, $g_0 = 57.0nS$, $\tau = 10.5ms$, and $\sigma = 15.84nS$. For the excitatory background conductance, $g_0 = 12nS$, $\tau = 2.7ms$, and $\sigma = 7.2nS$. The densities of voltage dependent sodium and potassium channels are $480pS/\mu m^2$ and $100pS/\mu m^2$ respectively. The density of spike adaptation potassium channels is $3pS/\mu m^2$. The surface area of the simulated neuron is $34636\mu m^2$.

The model also includes a hyperpolarization-activated conductance (I_h) with kinetics modeled as in Migliore (2003). This conductance has a reversal potential of $-43mV$ (Stuart and Spruston 1998) and a density of $0.05mS/cm^2$.

2.3 Results

Our results are divided into three sections. First we will present the results of a series of numerical simulations of a single model neuron of the visual cortex. The model neuron is a conductance-based integrate-and-fire neuron. Its input includes noise conductances designed to match voltage noise observed *in vivo* (Anderson et al. 2000b), which make the RMS voltage noise about 5 mV. In the second section we will provide a simplified, but more intuitive, analytical model of the neuron which illustrates the generality and robustness of the results obtained with the integrate-and-fire model. Finally, to further illustrate the robustness of the results, we will show that the assumptions of the analytical model and the results of the integrate-and-fire model all hold for a more biophysically detailed Hodgkin-Huxley type model neuron that also receives noise conductances. This model includes conductances responsible for spike generation and spike frequency adaptation, as well as a sub-threshold hyperpolarization-activated conductance (I_h). We will refer to the two numerical models as either the integrate-and-fire model or the Hodgkin-Huxley model.

Gain is defined here as the slope of a plot of a stimulus parameter, such as contrast or orientation, versus the response of the neuron. A pure gain change is one in which the curve of response vs. stimulus parameter is multiplicatively scaled, so that the gain is scaled by a constant factor for all values of the stimulus parameter.

2.3.1 Integrate-and-Fire Simulations

Contrast Response Curve Gain

Neurons in the visual cortex respond to stimuli of increasing contrast with an increasing firing rate. Plots of contrast versus firing rate are often well fit by a hyperbolic ratio function (Albrecht and Hamilton 1982, Sclar et al. 1990):

$$R = R_{max}\left(\frac{C^n}{C^n + C_{50}{}^n}\right) + S \tag{2.5}$$

where C is the contrast, R is the firing rate, S is the spontaneous activity, and C_{50} is the contrast that gives a half maximal firing rate. We will refer to a plot of contrast versus firing rate as a Contrast Response (CR) curve. The gain of the neuron is the slope of the CR curve.

We assume that our model neuron receives a stimulus-driven excitatory Poisson input. The rate of this input is a hyperbolic ratio function of stimulus contrast, with $R_{max} = 2000hz$, $C_{50} = 0.133$, $n = 1.2$, and $S = 0$. This is designed to model synaptic input from an earlier stage of visual processing. We then study how the gain of the model neuron's response to the stimulus is altered by additional excitation or inhibition in the form of glutamate or GABA receptor binding drugs or direct injected current.

Iontophoresis We model iontophoretic application of drugs binding to NMDA, AMPA, GABA$_A$ or GABA$_B$ receptors by opening a constant conductance of the appropriate type, and study the effects on the neuron's CR curve (Fig. 2.1A).
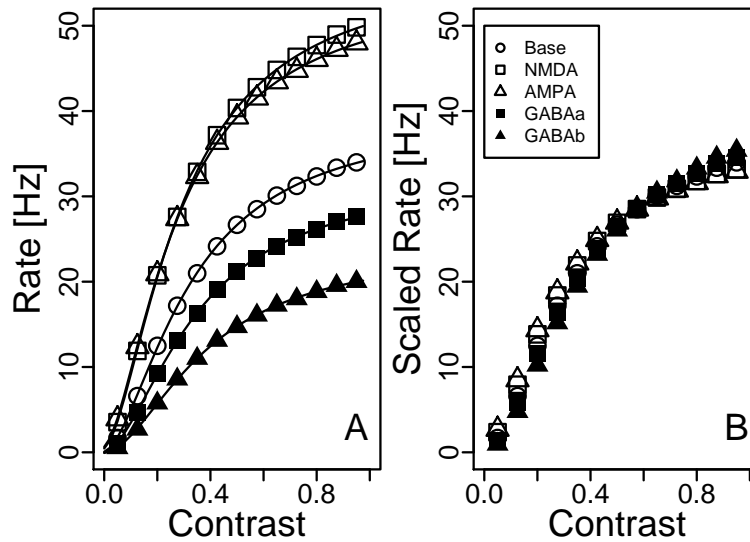
Figure 2.1: **Pharmacologically induced changes of contrast response curve gain.** (A) Plot of contrast versus firing rate for the model neuron without drugs (open circles), and with iontophoretic NMDA (open squares), AMPA (open triangles), GABA$_A$ (closed squares), or GABA$_B$ (closed triangles) applied. Each data point represents an average of twenty 60 second trials. Solid lines are fits of the data to a hyperbolic ratio function (Eq. 2.5). Parameters for the fits were obtained using a non-linear least squares algorithm and are given in Table 2.1. (B) Curves from A scaled to optimally (least squares) fit the Base curve.

We first simulated the iontophoresis of NMDA onto the neuron by opening a constant NMDA conductance equivalent to 10 nS if the neuron were held at +100 mV. Because of the voltage dependence of NMDA channels, the mean iontophoretic NMDA

conductance is 0.48 nS at zero contrast and increases to 1.1nS at maximum contrast. This conductance increased the firing rate of the model neuron at all contrasts. Maximum firing rate (at $C = 1.0$) increased to 50Hz from 34Hz in the baseline curve. The firing rate at $C = 0.0$ increased to 0.96Hz from 0.26Hz. This corresponds to a 46% increase in the average slope of the CR curve. One might imagine that this slope increase is caused by the voltage dependent increase of the NMDA conductance. However, when we simulated the iontophoresis of AMPA by opening a constant 1.0nS AMPA conductance, we observed a similar increase in firing rates. Maximum firing rate increased to 48Hz, and the firing rate at $C = 0.0$ increased to 0.81Hz. The average slope of the CR curve increased by 40%. Since the AMPA conductance is not voltage dependent, this implies another mechanism of gain change.

We next simulated the iontophoresis of inhibitory drugs binding to $GABA_A$ or $GABA_B$ receptors. In both cases the firing rate of the model neuron was reduced at all contrasts. A constant 2.0nS $GABA_A$ conductance decreased the maximum firing rate from 34Hz to 28Hz, and the firing rate at $C = 0.0$ from 0.26Hz to 0.15Hz. The average slope of the CR curve decreased by 17%. A constant 2.0nS $GABA_B$ conductance had a larger effect, decreasing the maximum firing rate to 20Hz and the firing at $C = 0.0$ to 0.06Hz. The average slope decreased by 41%. The different effects of $GABA_A$ and $GABA_B$ can be attributed to their different reversal potentials.

We scaled the iontophoretic CR curves to optimally fit the baseline curve (Fig. 2.1B). The overlap of the scaled CR curves indicates that changes in firing rate caused by iontophoretic conductances are very close to pure, multiplicative gain changes. Nonetheless, there are clearly systematic deviations from a purely multiplicative scaling. These deviations are made more clear by analyzing the fits of the CR curves to hyperbolic ratio functions (shown as solid lines in Fig. 2.1A). Fit parameters are given in Table 2.1. The most significant of these deviations are left and right shifts of the CR curve, indicated by changes in C_{50} (from Eq. 2.5). NMDA and AMPA conductances shifted the baseline curve left, decreasing C_{50} by 11% and 13% respectively. $GABA_A$ and $GABA_B$ shifted the curve right, increasing the baseline C_{50} by 6% and 17% respectively. NMDA and AMPA also increased spontaneous activity ($C = 0.0$) somewhat, while $GABA_A$ and $GABA_B$ reduced it.

	R_{max}	C_{50}	n	S	Scale
Base	39.5	0.325	1.66	0.0600	1.00
NMDA	56.5	0.290	1.61	0.452	1.50
AMPA	54.1	0.280	1.58	0.632	1.46
$GABA_A$	32.5	0.346	1.71	-0.0294	0.800
$GABA_B$	24.0	0.381	1.78	-0.0760	0.565

Table 2.1: Values of the fit parameters for the solid lines in Fig. 2.1A, and the scale factors used in Fig. 2.1B. Parameters are defined in Eq. 2.5.

Injected Current In our model, iontophoresis causes both polarizing ionic current and changes in conductance. The conductance changes are small relative to the neuron's mean resting conductance of about 24nS, rendering it unlikely that the conductance changes are a significant cause of the gain change. However, to directly separate the effects of injected current and conductance change, we next simulated the direct injection of current into the model neuron.

The effects of current injection on CR curves are very much like the effects of iontophoretic drugs (Fig. 2.2). Injection of depolarizing current into the neuron has effects similar to an iontophoretic AMPA or NMDA conductance. 50 pA of current increased the maximum firing rate from 34Hz to 47Hz, and the firing rate at $C = 0.0$ from 0.26Hz to 0.73Hz. The average slope of the CR curve increased by 37%. Conversely, injection of hyperpolarizing current has effects similar to an iontophoretic $GABA_A$ or $GABA_B$ conductance. -50 pA of current decreased the maximum firing rate to 24Hz, and the firing rate at $C = 0.0$ to 0.09Hz. The average slope of the CR curve decreased by 29%.

The changes in firing rate caused by injected current are close to purely multiplicative gain changes (Fig. 2.2B). Deviations from purely multiplicative changes are indicated by the fit parameters in Table 2.2. These deviations are in the same direction and of similar magnitude to those seen above for iontophoresis. Positive current shifted the baseline CR curve left, decreasing C_{50} by 12%. Negative current shifted the curve right, increasing the baseline C_{50} by 12%. These results imply that polarizing current, and not a change in conductance, is the primary factor responsible for changing the gain of the model neuron.

Our model of the effects of injecting current can be directly compared to existing *in vivo* data. Sanchez-Vives et al. (2000) examined the effect on visual cortical contrast response curves of injecting a hyperpolarizing current into the cell (Fig. 2.3A). Effects in
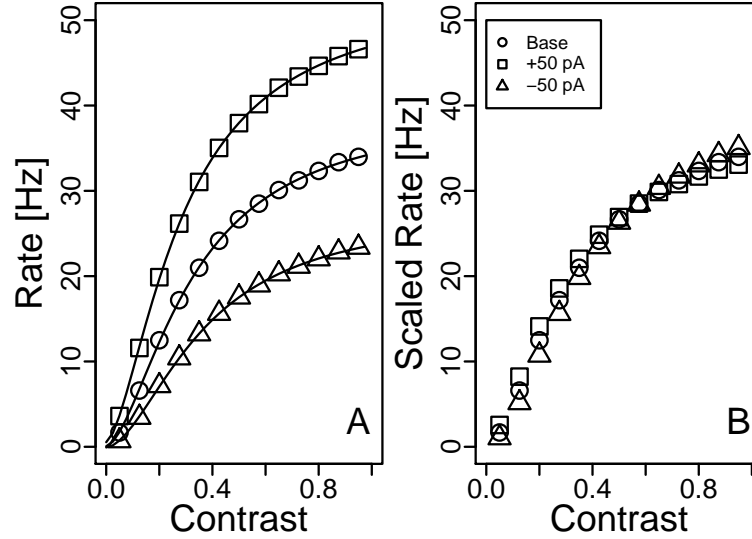
Figure 2.2: **Current induced changes of contrast response curve gain.** (A) Plot of contrast versus firing rate for the model neuron without injected current (circles) and with +50 pA (squares) or -50 pA (triangles) of current injected. Each data point represents an average of twenty 60 second trials. Solid lines are fits of the data to a hyperbolic ratio function (Eq. 2.5). Parameters for the fits are given in table 2.2. (B) Curves from A scaled to optimally (least squares) fit the Base curve.

	R_{max}	C_{50}	n	S	Scale
Base	39.5	0.325	1.66	0.0600	1.00
+50 pA	52.7	0.285	1.59	0.536	1.41
-50 pA	27.7	0.365	1.76	-0.0751	0.667

Table 2.2: Values of the fit parameters for the solid lines in Fig. 2.2A, and the scale factors used in Fig. 2.2B. Parameters are defined in Eq. 2.5.

the model are strikingly similar (Fig. 2.3B). The simulation produces a more purely multiplicative effect than is seen in the average experimental data (Fig. 2.3,D), but the deviations are in the same directions in simulation as in the data. In particular, the hyperpolarization induces a statistically significant (Sanchez-Vives et al. 2000) increase in C_{50} in the experimental data, as predicted by the model. It should be noted that the experimental data is averaged over many cells; if each cell were modulated multiplicatively but different cells had different contrast response curves and were modulated by different factors, then the modulation of the average would not be purely multiplicative. Thus it is possible that individual cells show a more purely multiplicative effect than the average data. In any case, the overall resemblance suggests that our simple model gives a reasonable representation of neurons *in vivo*.

Tuning Curve Gain

A number of influences, including attention and eye position, have been shown to change the gain of tuning curves in cortex. Although the mechanisms by which these gain changes occur are not clear, we will now demonstrate that, in our model neuron, one synaptic input can modulate the gain of the response to another. To do so we introduce two Poisson inputs to the neuron. The first is an excitatory driving input whose rate R is a Gaussian function of an arbitrary stimulus parameter θ:

$$R(\theta) = R_{max} \exp\left(-\frac{\theta^2}{2\sigma^2}\right) + S \tag{2.6}$$

The second is a modulatory input, either excitatory or inhibitory, whose rate is independent of θ. For the driving input, $\sigma = 1.0$, $R_{max} = 2000Hz$, and $S = 0Hz$.

We constructed tuning curves for the model neuron by plotting average firing rate versus θ, with and without modulatory inputs (Fig. 2.4A). The driving input alone produced a maximum firing rate at $\theta = 0.0$ of 41Hz, and a minimum firing rate at $\theta = \pm 3$ of 0.29Hz. We then added modulatory inputs to the neuron. A 250Hz excitatory input increased the firing rate of the neuron for all values of θ. The maximum response increased to 55Hz, and the firing rate at $\theta = \pm 3$ increased to 0.86Hz. A 250Hz inhibitory input decreased the firing rate of the model neuron for all values of θ. The maximum firing rate decreased to 31Hz, and the firing rate at $\theta = \pm 3$ decreased to 0.12Hz.

We scaled the tuning curves with modulatory input to best fit the baseline tuning curve (Fig. 2.4B). As in Figs. 2.1 and 2.2, the tuning curves with modulatory input can be
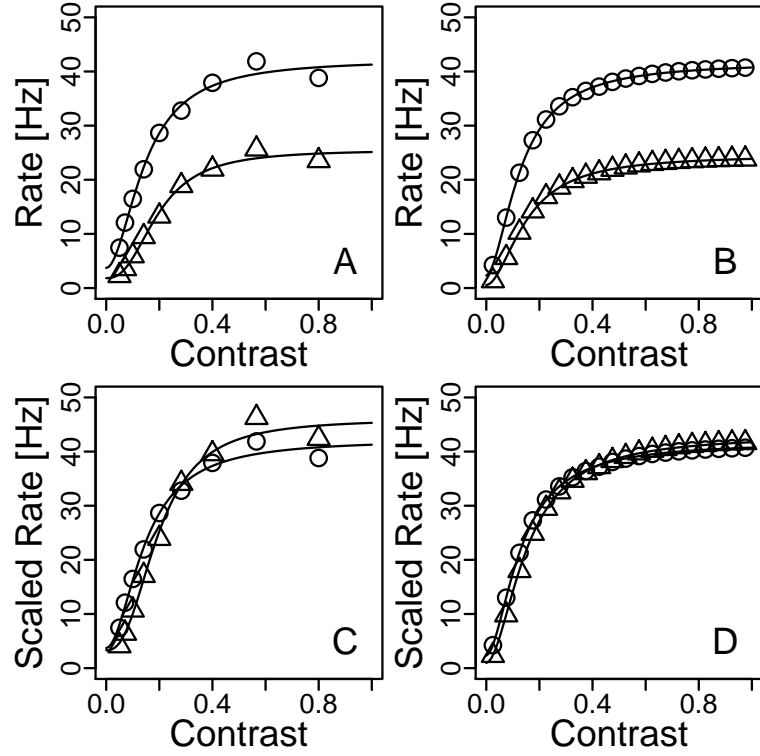
Figure 2.3: **Model matches the effects of hyperpolarizing current on contrast reponse curve gain *in vivo*.** (A) Replotted data from Sanchez-Vives et al. (2000), Fig. 12A, demonstrating the effects of hyperpolarizing injected current on the contrast response functions of neurons in primary visual cortex. Circles: curves with no injected current. Triangles: curves with hyperpolarizing current injected. (B) Our simulated CR curves. Control curve (circles) uses input parameters $R_{max} = 1400Hz$, $C_{50} = 0.08$, $n = 1.6$, and $S = 600Hz$, designed to reproduce control curve in A. Curve with hyperpolarizing current (triangles) used current of -90 pA, designed to produce a similar reduction of $R_{\max}$ as in A. Each data point represents an average of twenty 60 second trials. (C,D) Plots from A and B with the hyperpolarized curve scaled to optimally (least squares) match the control curve. Data in (A) are averages over multiple cells studied over four octaves of contrast; we have plotted the four octaves as 5% to 80%. Solid lines are fits of the data to a hyperbolic ratio function (Eq. 2.5). Fit parameters for the model are $R_{max} = 39Hz$, $C_{50} = 0.13$, $n = 1.8$, and $S = 2.3Hz$ without and $R_{max} = 24Hz$, $C_{50} = 0.15$, $n = 1.9$, and $S = 0.6Hz$ with hyperpolarizing current. Fit parameters for the experimental data are $R_{max} = 38Hz$, $C_{50} = 0.15$, $n = 1.9$, and $S = 3.7Hz$ without and $R_{max} = 24Hz$, $C_{50} = 0.2$, $n = 2.5$, and $S = 1.8Hz$ with hyperpolarizing current.
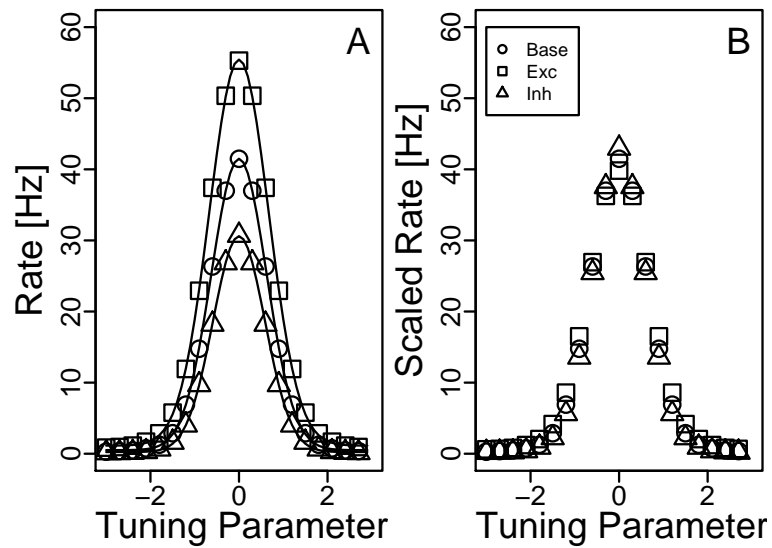
Figure 2.4: **Tuning curve gain is multiplicatively modulated by excitatory or inhibitory synaptic input.** (A) Plot of the tuning parameter in Eq. 2.6 versus firing rate for the model neuron without modulatory input (circles) and with a modulatory 250 Hz excitatory (squares) or 250 Hz inhibitory (triangles) Poisson input. Each data point represents an average of twenty 60 second trials. Solid lines are fits of the data to a Gaussian function (Eq. 2.6). Parameter values for the fits are given in table 2.3. (B) Curves from A scaled to optimally (least squares) fit the Base curve.

scaled to nearly fit the baseline curve. Parameters for Gaussian fits of the data in Fig. 2.4A are given in Table 2.3. These fits show that in addition to the multiplicative scaling, there are systematic changes in the width of the tuning curve caused by modulatory input. Excitatory modulation caused an 8% increase in the width of the tuning curve. Inhibitory modulation caused a 6% decrease in the width of the tuning curve.

	R_{max}	σ	S	Scale Factor
Base	41.0	0.622	0.508	1.00
250hz Exc	54.3	0.669	1.14	1.39
250hz Inh	30.4	0.588	0.235	0.715

Table 2.3: Values of the fit parameters for the solid lines in Fig. 2.4A, and the scale factors used in Fig. 2.4B. Parameters are defined in Eq. 2.6.

Results obtained by injecting constant current, instead of adding a modulatory Poisson input, are very similar (data not shown). None of the modulatory inputs used in any of these simulations had an affect on the response variability of the model neuron as measured by the coefficient of variation of inter-spike intervals.

Although we have chosen the parameters for the integrate-and-fire model carefully in order to match the experimentally measured properties of cortical neurons, our results do not depend on the detailed parameters of the simulations. For example, nothing qualitatively changes if the membrane time constant is doubled or halved (by correspondingly changing the capacitance), or if the stimulus-dependent Poisson input is replaced by an injected current (so long as the current's amplitude remains the same non-linear function of the stimulus parameter). The reason for this robustness is illustrated by the following simple analytical model, which shows the more general conditions required for these results to hold.

2.3.2 Analytical Model

We express the neuron's state in terms of the *shadow voltage*, defined to be the voltage the neuron would have if it did not spike or undergo post-spike voltage resets. The effect of the voltage noise in the model neuron is to make the neuron's firing rate f depend on its mean shadow voltage V as a power law (Hansel and van Vreeswijk 2002, Miller and Troyer 2002) (Fig. 2.5):

$$f = kV^{\alpha} \tag{2.7}$$
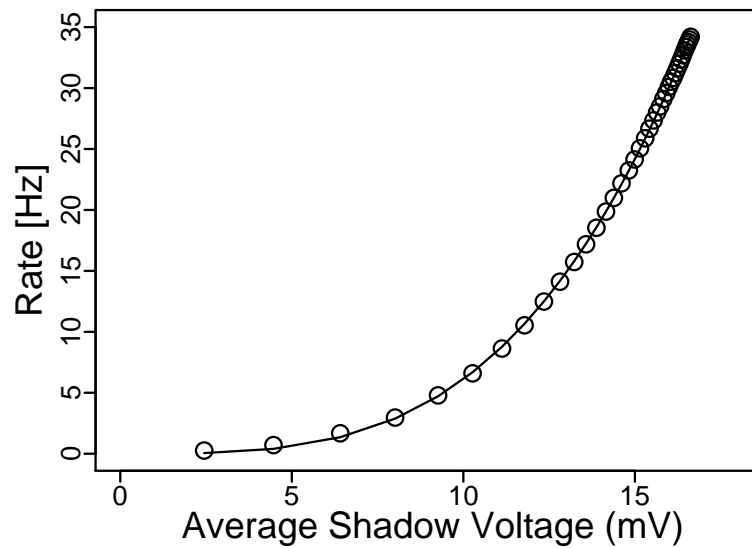
Figure 2.5: **Firing rate is a power law function of mean voltage** A plot of the average shadow voltage versus firing rate (circles) for the model neuron driven by an excitatory Poisson input (the same input as the Base curve in Fig. 2.1). Each data point represents an average of twenty 60 second trials. The voltage axis is shifted such that V_{rest} is 0mV. The solid line is a fit of $f = kV^\alpha$ to the data, with $k = 0.0025$ Hz/[mV]$^\alpha$ and $\alpha = 3.4$. The shadow voltage is the voltage the neuron would have if spiking and post-spiking reset were ignored, see text.

Here, α is about 3.4. Furthermore the shadow voltage is linear in the mean input, so it can be thought of as a function of a driving input and a modulatory input:

$$V = d(p) + m \tag{2.8}$$

The modulatory input m is a constant, while the driving input $d(p)$ is a function of an input parameter p. In our numerical simulations p corresponds to either contrast or the tuning parameter θ. Likewise m corresponds to iontophoretic application of drugs, injected current, or modulatory synaptic inputs. Using these two equations we obtain an expression for firing rate with respect to $d(p)$ and m:

$$f = k\Big(d(p) + m\Big)^{\alpha} \tag{2.9}$$

This equation is already sufficient to largely explain the results of our simulations. Letting $d(p)$ be a sigmoid (Fig. 2.6A) or a Gaussian (Fig. 2.6B), we can compare $k\Big(d(p)\Big)^{\alpha}$ (solid lines in Fig. 2.6) to $k\Big(d(p) + m\Big)^{\alpha}$ for positive m, renormalized to best match $k\Big(d(p)\Big)^{\alpha}$ (dotted lines in Fig. 2.6). Here, positive m represents an excitatory modulatory input. We find that the effect of excitatory modulation is to approximately multiply a sigmoidal contrast response curve, but with a slight left-shift of the curve (Fig. 2.6A); and to approximately multiply a Gaussian tuning curve, but with a slight widening of the curve (Fig. 2.6B). In both cases, the result of the simple model given by Eq. 2.9 is essentially identical to the results of the integrate-and-fire simulations.

To further understand why Eq. 2.9 leads to a gain change, we define the gain as the derivative of the firing rate with respect to the input parameter. Taking the derivative of Eq. 2.9 with respect to p we obtain an expression for the gain:

$$\frac{\partial f}{\partial p} = k\alpha\Big(d(p) + m\Big)^{\alpha-1}\frac{\partial d(p)}{\partial p} \tag{2.10}$$

Because $\alpha > 1$ the gain is directly related to the modulatory input, m.

However, Chance et al. (2002) point out that a change in $\frac{\partial f}{\partial p}$ for a given p can arise from either a "true" gain change, or a simple left or right shift of the non-linear f versus p curve. To distinguish between these two cases they plot $\frac{\partial f}{\partial p}$ versus f. Using Eq. 2.9 we can rewrite Eq. 2.10 as:

$$\frac{\partial f}{\partial p} = k\alpha \left(\frac{f}{k}\right)^{\frac{\alpha-1}{\alpha}} \frac{\partial d(p)}{\partial p} \tag{2.11}$$
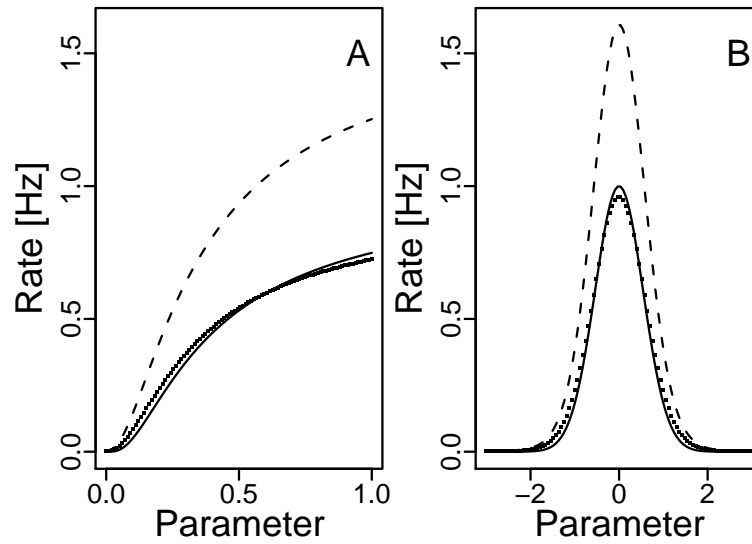
Figure 2.6: **A simple analytical explanation of gain changes.** Plots of Eq. 2.9 with $\alpha = 3.4$ and $k = 1.0$ Hz/[mV]$^\alpha$. The solid lines correspond to a plot of $d(p)$ with $m = 0$mV, and the dashed lines correspond to $m = 0.15$mV. The dotted lines are the $m = 0.15$mV curves optimally (least squares) scaled to fit the $m = 0$mV curves. (A) $d(p)$ is a hyperbolic ratio function with $C_{50} = 0.133$, $n = 1.2$, and $R_{max} = 1.0$mV. (B) $d(p)$ is a Gaussian with $\sigma = 1.0$, and an amplitude of 1.0mV.

When d(p) is linear in p, $\frac{\partial d(p)}{\partial p}$ is a constant. Therefore $\frac{\partial f}{\partial p}$ has no dependence on m, and changes in modulatory inputs (m) do not affect the gain. This can be seen more directly from Eq. 2.9: if $d(p) = ap+b$, then $f = k\,(ap + b + m)^\alpha$, so that m simply left- or right-shifts the curve of f vs. p without changing its shape. In Chance et al. (2002) the driving input is an injected current, and $d(p)$ is a linear function of p. In our results, and in many biologically relevant situations, $d(p)$ is not linear. In this case $\frac{\partial d(p)}{\partial p}$ in Eq. 2.11 is not constant, but depends on p; when reexpressed in terms of f, there will also be a dependence on m. That is, the value of $\frac{\partial f}{\partial p}$ as a function of f depends on m. Thus changes in modulatory inputs do not simply shift the f versus p curve, they change its gain – its slope for a given value of f.

The gain changes observed in our numerical simulations are very nearly multiplicative – the slope at each point is changed by roughly the same factor. The above arguments explain why there should be a gain change, but not why it should be nearly multiplicative. To make this more clear it is useful to look at the expansion of Eq. 2.9 to first order in $\frac{m}{d(p)}$:

$$k\Big(d(p) + m\Big)^\alpha \approx k\Big(d(p)^\alpha + \alpha m d(p)^{\alpha-1}\Big) \tag{2.12}$$

With no modulatory input, $m = 0$, the firing rate is simply $kd(p)^\alpha$. If a non-zero modulatory input had a purely multiplicative effect on firing rate we would expect it to add to the firing rate an amount proportional to $d(p)^\alpha$. Equation 2.12 shows that a non-zero m actually adds an amount proportional to $d(p)^{\alpha-1}$. Thus the multiplicative effects of modulatory inputs in our model depend on $d(p)^{\alpha-1}$ being similar in shape to $d(p)^\alpha$, which in turn depends on α being substantially larger than 1. Of course, $d(p)^\alpha$ and $d(p)^{\alpha-1}$ cannot be perfectly identical in shape unless $d(p)$ is a constant, so small discrepancies from a perfectly multiplicative scaling are predicted; similar discrepancies are seen in some experimental results, as addressed in the Discussion.

From the inputs represented in our simulations the relationship between $d(p)^\alpha$ and $d(p)^{\alpha-1}$ is most easily seen when $d(p)$ is a Gaussian function. In this case $d(p)^\alpha = \exp(-\alpha\frac{p^2}{2\sigma^2})$ and $d(p)^{\alpha-1} = \exp(-(\alpha-1)\frac{p^2}{2\sigma^2})$. So when $m = 0$ the firing rate of the neuron is a Gaussian, which is narrower than the Gaussian input $d(p)$ by a factor of $\sqrt{\alpha}$. A positive m adds an amount proportional to a slightly wider Gaussian (narrower than the Gaussian input by a factor of $\sqrt{\alpha - 1}$). This produces a new tuning curve that is multiplicatively scaled and wider than the curve with $m = 0$ by an amount that depends on m and α. The width should increase by an amount no less than zero, and no more than a factor of $\frac{\sqrt{\alpha}}{\sqrt{\alpha-1}}$. For $\alpha = 3.4$ the width should increase no more than 19%, which is consistent with results

from the integrate-and-fire simulation in which modulatory excitation increased the width of the tuning curve by 8%, and consistent with the slight widening observed in Fig. 2.6B.

The relationship between $d(p)^\alpha$ and $d(p)^{\alpha-1}$ is less clear when $d(p)$ is a hyperbolic ratio function. However, it is easy to show that if $d(p) \propto \frac{C^n}{C^n + C_{50}{}^n}$, then $[d(p)]^\alpha$ reaches its half-maximum at $C = \frac{C_{50}{}^n}{[2^{\frac{1}{\alpha}}-1]^{\frac{1}{n}}}$, which increases with increasing α. That is, $d(p)^{\alpha-1}$ should reach its half-maximum at a slightly lower value of C than $d(p)^\alpha$, so that adding $d(p)^{\alpha-1}$ to $d(p)^\alpha$ should cause a slight left-shifting of the sigmoid curve. This is the result of excitatory modulation as seen in simulations and in Fig. 2.6A.

The approximation in Eq. 2.12 is justified in cases when $\frac{m}{d(p)}$ is small. Clearly this is not always the case in our simulations; for instance, $d(p)$ goes to zero when contrast is zero. In this case the value of the function goes to km^α, and the relative error of the approximation is very large. However, as long as m is sufficiently small, the absolute error is small, and the approximation is useful. The usefulness of this approximation in describing our numerical results is demonstrated in Fig. 2.7, which plots the right hand side of Eq. 2.12 using values for $d(p)$ and m taken from our integrate-and-fire simulations. $d(p)$ is simply the mean shadow voltage (without modulatory inputs) of the model neuron in the simulated CR and tuning curves. m is the mean difference between the shadow voltage of the neuron with a modulatory input and the baseline. The firing rates reconstituted using Eq. 2.12 (Fig. 2.7, solid lines) correspond well with the actual firing rates in the integrate-and-fire simulations (Fig. 2.7, symbols). This indicates that this approximation is a reasonable one for describing these simulations.

The analytical model assumes that inputs cause additive changes in the shadow voltage. This is true for input currents. However, input conductances need not translate additively into input currents. In order to completely describe the effects of an input conductance, one has to consider reversal potential effects by which the current flowing through a conductance depends not only on the size of the conductance but also on the driving force. As such, the analytical model does not completely account for the effects of conductances, particularly shunting inhibitory conductances with reversal potentials close to rest. Nonetheless, we have found that shunting (GABA-A) conductances behave similarly to injected currents in our numerical model, causing multiplicative changes in the curve of stimulus parameter vs. firing rate. This is probably because, over the range of shadow voltages for which the neuron's firing rate is significantly different from zero, the changes
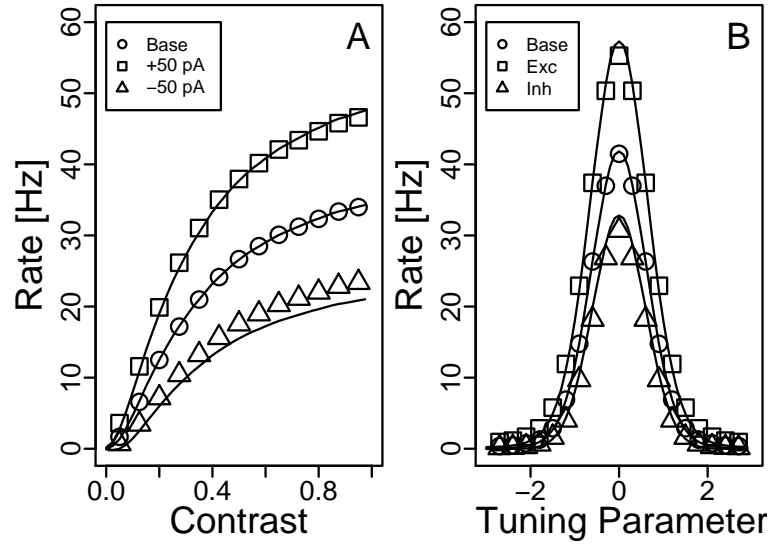
Figure 2.7: **Comparision of the analytical model to simulations.** Comparison of predictions of Eq. 2.12 (solid lines) to results of simulations (symbols). Solid lines are plots of the right hand side of Eq. 2.12 using values for $d(p)$ and m taken from the numerical simulations. $d(p)$ is the mean shadow voltage of the baseline curve without modulation. m is the mean difference between the shadow voltage of the curve with modulation and the baseline curve. Symbols are repeated from Figs. 2.2A and 2.4A. (A) $d(p)$ and m taken from CR curves generated by numerical simulation (Fig. 2.2). (B) $d(p)$ and m taken from tuning curves generated by numerical simulation (Fig. 2.4). Circles: tuning curves without modulatory input. Squares: tuning curves with 50 pA injected current (A) or with 250 Hz excitatory Poisson input (B). Triangles: tuning curves with -50 pA injected current (A) or 250 Hz inhibitory Poisson input (B).

in the driving force are relatively small. Similar arguments (Holt and Koch 1997) lead to the result that shunting conductances, like injected currents, have an additive effect on the curve of input current vs. firing rate (Chance et al. 2002, Holt and Koch 1997).

We are arguing that addition of two inputs, followed by raising to a power, gives an approximate multiplication. If the input-output function were an exponential (*e.g.*, Gabbiani et al. 2002) rather than a power law, this relationship would be exact: $e^{a+b} = e^a e^b$. This raises the question whether a better analytic approximation to our results might be given by an exponential rather than a power law. However, an exponential input-output relationship for the stimulus-induced firing rate must have the form $f = k(e^V - 1)$ in order that $f = 0$ when $V = 0$ (where 0 represents rest). We tried fitting a function of this form to the neuron's input-output relationship but the fit is visibly considerably worse than that shown in Fig. 2.5 for a power law. Furthermore the fits to the simulation data using this equation (equivalents of Figs. 2.6 and 2.7) are quite poor, particularly for excitatory modulatory input. We conclude that the power law gives the better description of our simulations.

The success of this simple analytical model in describing the more complex biophysical model used in simulations demonstrates the robustness of our results. The analytical model shows that achievement of multiplicative gain modulation depends on only two features of the biophysical model: the shadow voltage should be a roughly linear function of the mean input, and the output rate should be a power law of the shadow voltage with an exponent significantly larger than 1. These are both attributes of a wide variety of biophysical models with a wide variety of parameters. In particular, the robustness with which noise induces a power law in a series of models has been demonstrated elsewhere (Hansel and van Vreeswijk 2002, Miller and Troyer 2002), and we have also verified this: the power $\alpha = 3.39$ found in our integrate-and-fire model becomes $\alpha = 3.33$ if only AMPA and no NMDA is used for excitatory currents, $\alpha = 4.06$ if the time constant is doubled and $\alpha = 3.16$ if the time constant is halved (the time constant was manipulated by changing the capacitance), with excellent fits of a power law in all cases. The change in α with a change in time constant is as expected theoretically: it has been shown that α should primarily depend on the distance from rest to threshold in units of the standard deviation of the noise (Miller and Troyer 2002); doubling the time constant decreases the noise and hence increases this distance, which increases α, while halving the time constant increases the noise and hence decreases this distance, which decreases α.

Do these two features hold for real cortical neurons? The assumption that the voltage of a neuron is approximately a linear sum of its inputs is often made, and indeed it is under this assumption that a mechanism of gain modulation by inhibition or excitation alone has proven elusive – if inputs can multiplicatively influence one another, multiplicative gain modulations are likely to be easier to attain. Nonetheless our model provides such a mechanism under the assumption of linear input summation. Much evidence exists that summation in cortical or hippocampal pyramidal neurons can be linear (Cash and Yuste 1998, 1999, Jagadeesh et al. 1993, 1997). However, pyramidal neurons contain voltage-dependent conductances that can affect the summation of inputs (reviewed in Reyes 2001) and can cause inputs to summate in a non-linear manner (Nettleton and Spain 2000, Schwindt and Crill 1998, Wei et al. 2001). In addition, dendritic integration can be non-linear, although nonlinear conductances can correct this and linearize dendritic integration (Bernander et al. 1994, Cash and Yuste 1998, 1999). A recent modeling study suggests that integration of multiple inputs on a single thin apical dendrite may be nonlinear, but that integration between dendrites is remarkably linear (Poirazi et al. 2003a,b).

The second key assumption, of a power law relationship between voltage and firing rate, seems likely to hold in many cortical neurons. This is a general outcome of the presence of neural noise in a variety of neural models (Hansel and van Vreeswijk 2002, Miller and Troyer 2002), and many studies indicate the presence of substantial voltage noise in cortex (Anderson et al. 2000b, Arieli et al. 1996, Azouz and Gray 1999, Hô and Destexhe 2000, Paré et al. 1998, Tsodyks et al. 1999). More specific evidence is provided by the finding of Anderson et al. (2000b) that voltage noise can transform contrast-invariant voltage tuning into contrast-invariant spiking tuning in visual cortical neurons. As shown by Miller and Troyer (2002), a power law transformation from voltage to spiking rate is the only such transformation that can achieve this, thus indicating that such a transformation is found in visual cortical cells.

The above arguments suggest that the two assumptions of our model may hold in many cortical cells. To provide further evidence that our proposed mechanism of gain modulation can apply to real cortical neurons, we now present numerical simulations of a model that includes some of the more detailed biophysical properties of cortical pyramidal neurons.

2.3.3   Hodgkin-Huxley Simulations

The Hodgkin-Huxley neuron model includes three potentially important conductances not present in our integrate-and-fire simulations: voltage activated spiking conductances, a spike frequency adaptation conductance, and a hyperpolarization-activated mixed cation conductance I_h. To demonstrate that the mechanism of gain change we are proposing is valid in this neuron model we will first show that the mean voltage of the neuron is approximately a linear function of the input it receives, and that the firing rate of the neuron is related to the mean voltage by a power law. We will then generate CR and Gaussian tuning curves for the model neuron with and without modulatory inputs. These curves are generated in the same way as the curves in the integrate-and-fire simulations presented above, but for simplicity the synaptic inputs (both modulatory and driving inputs) have been replaced with constant excitatory ($E_{rev} = 0mV$) or inhibitory ($E_{rev} = -80mV$) conductances. Replacing these inputs with injected currents yields very similar results (not shown).

Input-Output Relationship

A power law relationship between mean voltage and firing rate has been reported previously for a Hodgkin-Huxley type model neuron (Hansel and van Vreeswijk 2002) with a noisy membrane potential. We confirm that this relationship is also present in the model neuron studied here (Fig. 2.8B). The mean voltage of the model neuron is related to the output firing rate by a power law with $\alpha = 2.91$ and $k = 0.033$. In addition we find that the relationship between excitatory input conductance and mean voltage is roughly linear over this range of output firing rate (Fig. 2.8A), although some deviation from linearity is seen as voltage approaches threshold.

Contrast Response Curve

We construct a CR curve by introducing a stimulus driven excitatory conductance whose magnitude is a hyperbolic ratio function of contrast with parameters $R_{max} = 20nS$, $C_{50} = 0.133$, $n = 1.2$, and $S = 0$ (see Eq. 2.5). The modulatory inputs do not vary with contrast and are either a 3.5 nS excitatory conductance, or a 8.5 nS inhibitory conductance.

Modulatory inputs have an effect on the CR curve very similar to that in the integrate-and-fire model (Fig. 2.9, compare to Fig. 2.2). A 3.5 nS excitatory conductance
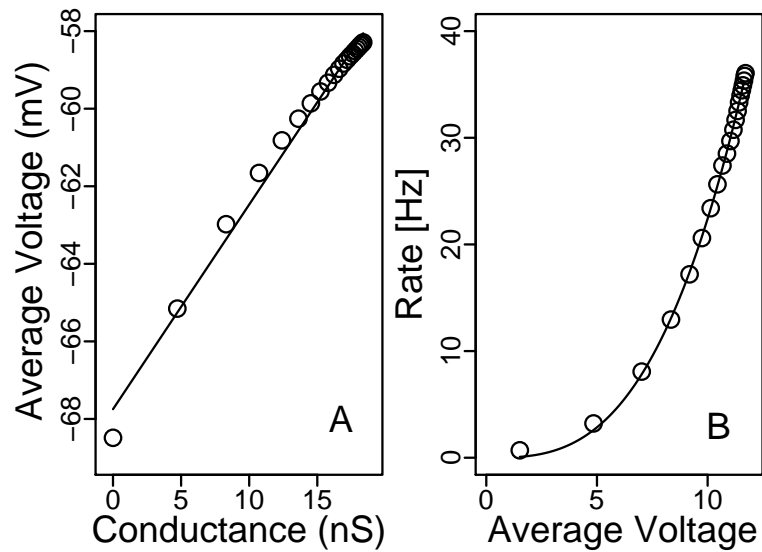
Figure 2.8: **Relationships between input conductance, mean voltage, and firing rate in the Hodgkin-Huxley model.** (A) Plot of the input conductance versus average voltage (circles) for the Hodgkin-Huxley model neuron. The solid line is a linear fit to the data, with a slope of 0.528 mV/nS and an intercept of -67.7 mV. (B) Plot of the average voltage versus firing rate (circles) for the Hodgkin-Huxley model neuron. The voltage axis is shifted such that V_{rest} is 0mV. The solid line is a fit of $f = kV^\alpha$ to the data, with $k = 0.024$ Hz/[mV]$^\alpha$ and $\alpha = 2.98$. In both A and B the input is the same as the Base curve in Fig. 2.9. Each data point represents an average of ten trials, five seconds each.

increased the maximum firing rate from 36 Hz to 49 Hz, and the firing rate at C = 0.0 from 0.7 Hz to 2.3 Hz. The average slope of the CR curve increased by 32%. A 8.5 nS inhibitory conductance decreased the maximum firing rate to 25 Hz, and the firing rate at C = 0.0 to 0.18 Hz. The average slope of the CR curve decreased by 30%.
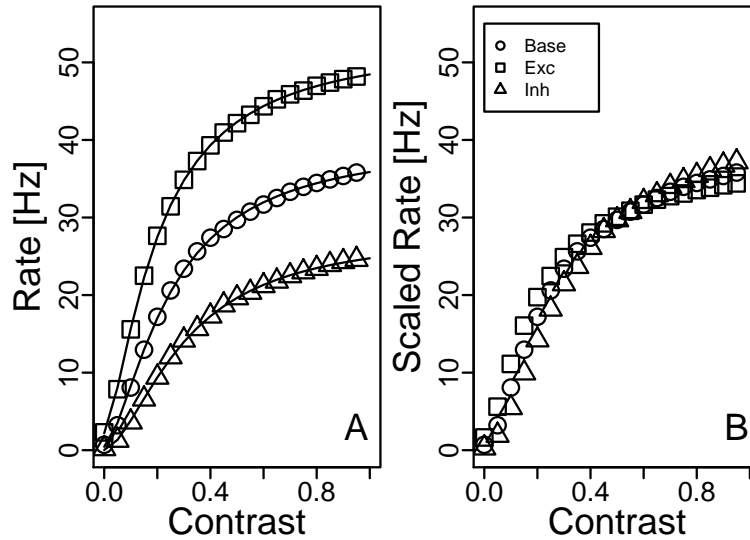
Figure 2.9: **Contrast response curve gain changes in the Hodgkin-Huxley model.** (A) Plot of contrast versus firing rate for the Hodgkin-Huxley model neuron without modulatory input (circles) and with 3.5 nS excitatory (squares) or 8.5 nS inhibitory (triangles) modulatory conductances. Each data point represents an average of ten trials, five seconds each. Solid lines are fits of the data to a hyperbolic ratio function (Eq. 2.5). Parameters for the fits are given in table 2.4. (B) Curves from A scaled to optimally (least squares) fit the Base curve.

The changes in firing rate caused by the modulatory inputs are close to purely multiplicative gain changes (Fig. 2.9B). Deviations from purely multiplicative changes are indicated by the fit parameters in Table 2.4. Excitatory modulatory conductance shifted the baseline CR curve left, decreasing C_{50} by 19%. Inhibitory modulatory conductance shifted the curve right, increasing the baseline C_{50} by 23%.

Tuning Curve

We construct a tuning curve by introducing a stimulus driven excitatory conductance whose magnitude is a Gaussian function of a stimulus parameter θ with parameters

	R_{max}	C_{50}	n	S	Scale
Base	39.48	0.2463	1.546	0.4358	1
Exc	50.71	0.2005	1.459	2.179	1.401
Inh	28.04	0.3022	1.677	0.02561	0.6615

Table 2.4: Values of the fit parameters for the solid lines in Fig. 2.9A, and the scale factors used in Fig. 2.9B. Parameters are defined in Eq. 2.5.

$R_{max} = 17.5nS$, $\sigma = 1.0$, and $S = 0Hz$ (see Eq. 2.6). The modulatory inputs were the same as for the CR curve, a 3.5 nS excitatory conductance or an 8.5 nS inhibitory conductance.

Modulatory inputs have an effect on the tuning curve very similar to that in the integrate-and-fire model (Fig. 2.10, compare to Fig. 2.4). The driving input alone produced a maximum firing rate at $\theta = 0.0$ of 33Hz, and a minimum firing rate at $\theta = \pm 2.5$ of 0.7Hz. We then added modulatory inputs to the neuron. The excitatory modulatory input increased the firing rate of the neuron for all values of θ. The maximum response increased to 46Hz, and the firing rate at $\theta = \pm 2.5$ increased to 2.3Hz. The inhibitory modulatory input decreased the firing rate of the model neuron for all values of θ. The maximum firing rate decreased to 22Hz, and the firing rate at $\theta = \pm 2.5$ decreased to 0.18Hz.

The changes in firing rate caused by modulatory inputs are close to purely multiplicative gain changes (Fig. 2.10B). Deviations from purely multiplicative changes are indicated by the fit parameters in Table 2.5. An excitatory modulatory conductance caused a 12% increase in tuning curve width, while an inhibitory modulatory conductance caused a 11% decrease in width.

	R_{max}	σ	S	Scale Factor
Base	32.19	0.5044	0.7883	1
Exc	43.21	0.5644	2.44	1.478
Inh	21.91	0.4474	0.2701	0.6271

Table 2.5: Values of the fit parameters for the solid lines in Fig. 2.10A, and the scale factors used in Fig. 2.10B. Parameters are defined in Eq. 2.6.

For both the CR curve and the tuning curve, deviations from purely multiplicative changes are in the same direction as the integrate-and-fire model and are similar in magnitude (though somewhat larger), which suggests that our proposed mechanism of gain change also operates in this more detailed biophysical model.
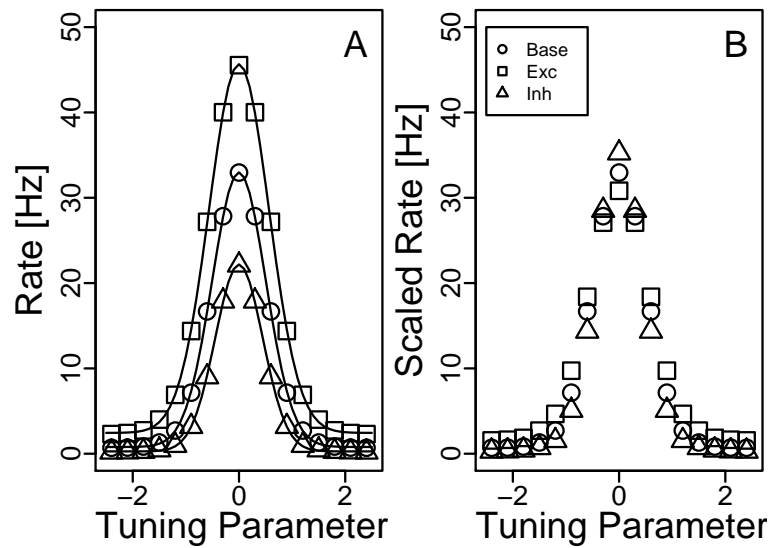
Figure 2.10: **Tuning curve gain changes in the Hodgkin-Huxley model.** (A) Plot of the tuning parameter in Eq. 2.6 versus firing rate for the Hodgkin-Huxley model neuron without modulatory input (circles) and with 3.5 nS excitatory (squares) or 8.5 nS (triangles) inhibitory nodulatory conductances. Each data point represents an average of ten trials, five seconds each. Parameter values for the fits are given in table 2.5. (B) Curves from A scaled to optimally (least squares) fit the Base curve.

2.4 Discussion

The results demonstrate that changes in excitatory or inhibitory inputs alone can approximately multiplicatively change the gain of a neuron's response to a stimulus-dependent input. These gain changes observed in our model are primarily dependent on two things: an expansive non-linearity (a power law with exponent substantially larger than one) relating the average membrane potential and the firing rate of the neuron, and an appropriately non-linear dependence (*e.g.*, sigmoidal or Gaussian) of the stimulus-dependent input rate on the corresponding stimulus parameter. Power law input-output functions are likely to be ubiquitous in cortex due to voltage noise (Anderson et al. 2000b, Arieli et al. 1996, Azouz and Gray 1999, Hansel and van Vreeswijk 2002, Hô and Destexhe 2000, Miller and Troyer 2002, Paré et al. 1998, Tsodyks et al. 1999), and cortical firing rates commonly have a sigmoidal or approximately Gaussian dependence on stimulus parameters. Hence, multiplicative gain changes should be common in cortex, and convergent inputs onto a neuron should multiplicatively modulate one another's gain, so long as one input (the "modulatory" input) is small relative to the peak input evoked by the other. This appears to be a natural result of the properties of cortical neurons and the input that they receive, and could help explain the ubiquity of such gain changes observed experimentally (Andersen et al. 1985, Boussaoud et al. 1993, Bremmer et al. 1997b, Galletti and Battaglini 1989, McAdams and Maunsell 1999a,b, Salinas and Thier 2000, Treue and Martinez-Trujillo 1999, Trotter and Celebrini 1999).

Predictions

The most obvious prediction of our model is that at least some of the multiplicative gain changes observed in cortex will be found to arise from purely excitatory or purely inhibitory modulation (or more generally, from unbalanced modulatory inputs yielding a net excitation or inhibition). In addition, our work suggests a number of clues that would be consistent with a mechanism involving net excitation or inhibition.

Net excitatory or inhibitory modulatory inputs should cause small systematic deviations from a purely multiplicative gain change: gain increases should lead to decreases in C_{50} and increases in tuning width, while gain decreases should yield opposite changes. However, the predicted deviations are small (approximately 10-20%) and may not be observable in practice. Even so, this is a fundamental prediction of our model and may become more

experimentally accessible in the future. Furthermore, the size of the deviations should be proportional to the size of the gain change, so larger changes in gain should give larger deviations from multiplication. This has two important implications. First, the deviations may become observable with sufficiently large gain changes. Second, modulation by unbalanced inputs should yield a correlation between the change in the amplitude of a tuning curve and the change in its width (for Gaussian tuning curves) or its C_{50} (for sigmoidal tuning curves).

We also predict that multiplicative gain changes *in vivo* induced by unbalanced inputs should be accompanied by an upward or downward shift in mean voltage, but little change in total conductance or response variability. Such gain changes should occur in neurons that have substantial voltage noise and thus have power-law input-output functions. If the gain change is due to a tonic excitation or inhibition (rather than a modulatory excitation or inhibition that only accompanies stimulus-induced input), then spontaneous activity levels should be modulated to the same degree as stimulus-induced responses.

Previous Theory and Experiment

We predict that tuning curves in cortex measured during intracellular current injection will exhibit multiplicative gain changes similar to those in our simulations. A gain decrease in the contrast response functions of neurons in primary visual cortex has been observed following injection of hyperpolarizing current (Fig. 2.3; Sanchez-Vives et al. 2000). Accompanying this gain decrease was a statistically significant increase in the parameter C_{50} describing the curve (Eq. 5), as predicted by our model.

Fox et al. (1990) examined the effects of externally applied glutamate-receptor-binding drugs on the responses of cortical neurons to visual stimulation. They found that NMDA increased the gain of the neuron's contrast response curve, while quisqualate shifted the curve upwards. Our model suggests that simple depolarization induced by NMDA application, rather than nonlinearities in NMDA-induced responses, caused the gain increase. The seemingly straightforward effects of quisqualate are more difficult to explain. They may have been confounded by the fact that quisqualate has a number of effects besides activating non-NMDA ionotropic glutamate receptors, including binding to metabotropic glutamate receptors (Chu and Hablitz 2000, Pin and Duvoisin 1995) and glutamate transporters (Chase et al. 2001). We predict that a selective agonist of AMPA receptors, such

as AMPA, would also cause a gain change.

A model proposed by Fox and Daw (1992) to account for their experimental results assumes that quisqualate acts exclusively at ionotropic non-NMDA receptors and that the firing rate of the neuron is linearly related to its membrane potential. In this case the quisqualate-induced shifts are easily explained. However, given a more realistic power law relationship between membrane potential and firing rate, it is difficult to account for the shifts induced by quisqualate based on its ionotropic action alone. To explain the effects of NMDA, they assumed cooperativity in binding between externally applied NMDA and synaptic glutamate released during visual stimulation, which requires that NMDA and glutamate bind to a common set of receptors. This was not always the case in their experiments; some cells exhibited gain increases during NMDA application, but no significant gain change during application of APV (an antagonist of NMDA receptors), implying that both bound to non-visual NMDA receptors. In light of this, depolarization seems a more plausible explanation for NMDA-induced gain changes.

Our model complements the recent model of Chance et al. (2002). They showed that a balanced change in inhibitory and excitatory inputs could cause a multiplicative gain change as assessed by a change in a neuron's curve of firing rate vs. current. We show that, when one instead considers the curve of firing rate vs. stimulus parameter, simple excitation alone or inhibition alone is sufficient to produce a gain change. This gain change is primarily dependent on the hyperpolarization or depolarization induced by the modulatory input, which causes only small changes in the cell's conductance. In contrast, the gain changes in Chance et al. (2002) require a relatively large change in both current noise and total conductance. A balanced change in excitation and inhibition together, as in Chance et al. (2002), also produces a gain change in our model (not shown). We conclude that a wide range of modulatory inputs, balanced or unbalanced, should induce a multiplicative gain change.

In cortical areas where eye position modulates neural responses to visual stimuli, some visually-driven neurons can also be driven directly by eye position alone (Boussaoud et al. 1993, Bremmer et al. 1997a,b, Squatrito and Maioli 1996, 1997). Our model explains both the modulatory and driving effects of eye position on these neurons as resulting from a single excitatory input. This requires that direct responses to eye position be small relative to visual responses in such neurons, which seems consistent with experiment (Bremmer et al. 1997a,b, Squatrito and Maioli 1997). This suggests some advantage of our proposal

over that of Chance et al. (2002), which seems to require two different types of eye-position-evoked inputs for such neurons: a set of balanced inhibitory and excitatory inputs that modulates the gain of visual responses, and another excitatory input that drives direct responses to eye position.

Changes in attention have been shown to multiplicatively scale the orientation and direction tuning curves of cortical neurons (McAdams and Maunsell 1999a,b, Treue and Martinez-Trujillo 1999). One interpretation of these experimental results is that the neurons are responding with increased gain to the same visually induced input, as in our model. However, more recent results (Martinez-Trujillo and Treue 2002, Reynolds et al. 2000) show that attention can cause a shift, rather than a gain change, in the contrast response function of these neurons. This suggests that the effects of attention may represent an increase in the effective contrast of the stimulus and not a change in the response gain of the neuron.

Treue and Martinez-Trujillo (1999) observed a slight (8%), but not statistically significant, widening of MT direction tuning curves during feature-based attention. This widening, if substantiated by more data, would be in accordance with our model, which predicts a widening of tuning curves with increasing gain (Fig. 2.4). Changes in attention do not appear to cause changes in the width of orientation tuning curves in area V4 (McAdams and Maunsell 1999a). Furthermore, although attention multiplicatively scaled tuning curves in V4 (including responses at non-preferred orientations), it did not systematically affect spontaneous activity in the absence of visual stimulation. A net excitatory modulatory input, if tonically active, would have scaled spontaneous activity in the same manner as stimulus evoked activity. Thus, if one models the effects of attention as a gain modulation rather than an increase in effective stimulus contrast, the model of Chance et al. (2002) may better explain the results from area V4. However, the widening of tuning curves in MT with attention is more in accord with unbalanced or purely excitatory modulation; it would be interesting to determine if spontaneous activity is affected by attention in MT, and to determine conclusively if attention affects the width of direction tuning curves in this area.

Our proposal is similar in spirit to that of Gabbiani et al. (2002), who suggested that an insect neuron does multiplication by subtracting one input from another at the level of voltages and having an exponential input/output relation (although they found that a power law fit their input-output relation better than an exponential). However, it differs in

being based on the properties of cortical neurons.

Since this work was completed, two papers have appeared that address the role of shunting inhibition in gain modulation. Both highlight the importance of voltage noise. Prescott and De Koninck (2003) showed in a modeling study that dendritic saturation of the excitatory input, along with voltage noise, could cause shunting inhibition to divisively alter firing rate. The effect of saturation seems similar to that of a non-linear, sigmoidal relationship between stimulus parameter and input rate in our model. Mitchell and Silver (2003) studied a cerebellar granular neuron that received relatively few excitatory synaptic inputs, all of which had large unitary conductances. As a result, an increase in input rate caused a significant increase in voltage noise. This in turn caused shunting inhibition to have a partially divisive effect on the curve of input rate versus output rate. If voltage noise does not increase significantly with input rate, as in the present and most previous studies of gain modulation, then shunting inhibition causes a subtractive shift in this curve (*e.g.*, Chance et al. 2002).

Implications for Neuronal Computation

The question of whether a single neuron can biophysically multiply its inputs has long been of interest to those concerned with the computational capabilities of single neurons (*e.g.*, Gabbiani et al. 2002, Koch 1998, Mel 1993, Torre and Poggio 1978). We are proposing that a cortical neuron that adds its inputs at the level of voltages, but raises this net input to a power significantly greater than one to produce an output, can effectively compute a multiplication of the inputs (or more strictly, of functions of the inputs: the output R is given by $R \approx f(i_1)g(i_2)$ where i_1 and i_2 are the inputs and f and g are some functions). Furthermore if the input voltages are nonlinear functions of a stimulus parameter, then this multiplication will not produce a mere left- or right-shift of the curve of output vs. parameter. Multiplication computed in this manner is only approximate. The approximation is accurate, though small systematic differences remain, when the "modulatory" input is substantially smaller than the "driving" input over the range of the tuning curve in which the driving input produces substantial responses.

Conclusion

Given a few basic assumptions about the common properties of cortical neurons, specifically the non-linear ways that their input firing rates depend on the properties of a stimulus and their output firing rates depend on their input, we have shown that it should be expected that a smaller input will multiplicatively modulate the gain of the response to a larger input. No special mechanisms are required to account for these multiplicative interactions. While other mechanisms may also play a role in experimentally observed gain changes, we are proposing that multiplicative gain changes are a normal property of the cortex, the natural outcome of these simple attributes of cortical neurons.

2.5 Acknowledgements

Chapter 3

Transient amplification

The material in this chapter has been submitted to *Nature* as:

Transient amplification is a new mechanism of pattern formation in neural activity

Brendan K. Murphy[1] and Kenneth D. Miller[2]

[1]Graduate Group in Biophysics

UCSF

[2]Center for Theoretical Neuroscience

Sloan Program in Theoretical Neuroscience

Columbia University

Email: murphy@phy.ucsf.edu, ken@phy.ucsf.edu

Abstract

In cerebral cortex, ongoing activity in the absence of a stimulus can resemble stimulus-driven activity both in size and structure (Anderson et al. 2000b, Arieli et al. 1996, Fiser et al. 2004, Kenet et al. 2003, Vincent et al. 2007). In particular, spontaneous activity in cat primary visual cortex (V1) has structure significantly correlated with evoked responses to oriented stimuli (Kenet et al. 2003). This suggests that, from unstructured input, cortical circuits selectively amplify activity patterns related to normal function. Current understanding of such amplification involves elongation of the lifetime of a neural pattern or "assembly" by mutual synaptic excitation among the neurons involved (Hebb

1949, Hopfield 1982). Here we introduce a new dynamical mechanism for selective amplification of neural activity patterns without elongation of lifetime: "transient amplification". We show in simple models how transient amplification arises, and in a detailed biophysical model that it can explain the observations from cat V1. Neurobiologically, strong transient amplification arises when strong feedback inhibition stabilizes strong recurrent excitation, a connectivity pattern likely to be typical of cortex (Chagnac-Amitai and Connors 1989, Haider et al. 2006, Ozeki et al. 2007, van Vreeswijk and Sompolinsky 1996). Thus, transient amplification should be ubiquitous in cortical circuits. Mathematically, transient amplification depends on the non-normal nature of synaptic connection matrices, in which individual neurons project only excitatory or only inhibitory synapses. This leads to a hidden feedforward connectivity between activity patterns, which underlies the amplification. Biological interaction matrices generally are non-normal, so these dynamics should have counterparts at all levels of biological organization.

3.1   Introduction

We focus on a well-studied example of selective amplification in V1. V1 neurons respond selectively to oriented visual stimuli. In cats, nearby neurons prefer similar orientations and there is a smooth map of preferred orientations across the cortical surface. Kenet et al. (2003) showed that, in the absence of a visual stimulus, spatial patterns of spontaneous activity across V1 upper layers showed greater similarity to the pattern evoked by an oriented stimulus than to similarly structured control patterns. This seems likely to result from the preferential cortical amplification of activity patterns in which neurons of similar preferred orientation are co-active. The substrate for such amplification is orientation-specific connectivity. Neurons in middle and upper layers of V1 receive both excitatory and inhibitory input predominantly from other neurons with similar preferred orientations (Anderson et al. 2000a, Martinez et al. 2002), and orientation-specific excitatory axonal projections can extend over long distances (Gilbert and Wiesel 1989). The amplified activity fluctuates with a time scale of about 80 ms (Kenet et al. 2003), comparable to the time scales over which inputs are correlated (DeAngelis et al. 1993, Wolfe and Palmer 1998), so the degree to which the cortical network can slow activity is limited (see Supplemental Materials).

In "Hebbian-assembly" models (*e.g.* Goldberg et al. 2004, Hebb 1949, Seung 2003),

selective amplification of an activity pattern is achieved by slowing its rate of decay. Given ongoing input that equally drives many patterns, patterns that decay most slowly will accumulate to the highest amplitude and so will dominate network activity. In the absence of intracortical connections, each pattern would decay with a time constant determined by cellular and synaptic time constants. A pattern that reproduces itself by passage through the recurrent circuitry will slow its decay rate or, if it reproduces itself faster than the intrinsic decay rate, will grow rather than decay. The latter condition, along with circuit nonlinearities, provides the basis for attractors, patterns that can persist indefinitely in the absence of specific driving input (Hopfield 1982).

In V1, recurrent excitation is strong but balanced by similarly strong feedback inhibition (Chagnac-Amitai and Connors 1989, Haider et al. 2006, Ozeki et al. 2007, van Vreeswijk and Sompolinsky 1996). Here we show that transient amplification is likely to be prominent in this type of network. During transient amplification, the recurrent circuitry transforms a small fluctuation in an "input pattern" into a large and transient fluctuation in a different "output pattern". The output pattern does not act back on the input pattern and neither pattern can significantly reproduce itself through the circuitry, so neither pattern shows slowed decay. Given ongoing input that equally drives many patterns, the output patterns of the largest transients will be most amplified in the network activity.

3.2   Linear Model

Essential features of this mechanism can be understood by considering a linear model of the firing rates of V1 neurons, in which a neuron's firing rate approaches its input with a time constant τ:

$$\tau \frac{d\mathbf{r}}{dt} = -\mathbf{r} + \mathbf{W}\mathbf{r} = -(\mathbf{1} - \mathbf{W})\mathbf{r} \tag{3.1}$$

Here, $\mathbf{r}$ is an N-dimensional vector representing the firing rates of a population of N neurons (the ith element r_i is the firing rate of the ith neuron). $\mathbf{W}$ is an $N \times N$ synaptic connectivity matrix (W_{ij} is the strength of connection from neuron j to neuron i). $\mathbf{W}\mathbf{r}$ represents input from other neurons within the network.

Dynamics of this model can be formally solved in terms of the *eigenvectors* $\mathbf{e}_i$ and *eigenvalues* λ_i of W, $i = 1, \ldots, N$. Each $\mathbf{e}_i$ represents a different pattern of activity across the neuronal population. They satisfy $\mathbf{W}\mathbf{e}_i = \lambda_i \mathbf{e}_i$; that is, each $\mathbf{e}_i$ reproduces itself, scaled

by the number λ_i, upon passage through the circuitry. Thus, eigenvalues with positive real part are the basis of Hebbian amplification.

However, for biological connection matrices, this formal solution hides key aspects of the dynamics. Because individual neurons project only excitatory or only inhibitory synapses, synaptic connection matrices have a characteristic structure, as follows. Let $\mathbf{r} = \begin{pmatrix} \mathbf{r}_E \\ \mathbf{r}_I \end{pmatrix}$, where $\mathbf{r}_E$ is the sub-vector of firing rates of excitatory neurons and $\mathbf{r}_I$ of inhibitory neurons. Let $\mathbf{W}_{xy}$ be a matrix with elements ≥ 0 describing the strength of connections from the cells of type y (E or I) to those of type x. Then the full connectivity matrix is $\mathbf{W} = \begin{pmatrix} \mathbf{W}_{EE} & -\mathbf{W}_{EI} \\ \mathbf{W}_{IE} & -\mathbf{W}_{II} \end{pmatrix}$. The left columns are non-negative and the right columns are non-positive. Such matrices are asymmetric and *non-normal*, meaning that their eigenvectors are not mutually orthogonal. Non-normal matrices can generate large transient amplification of small perturbations in ways not predicted by the eigenvalues (Trefethen and Embree 2005) (see Supplemental Materials). This has not previously been studied in a neural context, although other dynamical effects of the division of excitation and inhibition into distinct neuronal classes have been examined (Li and Dayan 1999, Wilson and Cowan 1972). We argue that such transient amplification results robustly from connectivity with strong recurrent excitation balanced by strong feedback inhibition.

The simplest example is a network with two populations of neurons, one excitatory and one inhibitory, each making projections that are independent of postsynaptic target. In terms of Eq. 3.1, $\mathbf{r} = \begin{pmatrix} r_e \\ r_i \end{pmatrix}$ and $\mathbf{W} = \begin{pmatrix} w_E & -w_I \\ w_E & -w_I \end{pmatrix}$. Here, r_e and r_i are the average firing rates of the excitatory and inhibitory populations and w_E and w_I are the strengths of projections of the excitatory and inhibitory populations respectively. We assume inhibition balances or dominates excitation, that is, $w_I \geq w_E$. Define $w_d = w_E - w_I$, $w_d \leq 0$, and $w_s = w_E + w_I$. Note that, if recurrent excitation and inhibition are both strong, then w_s is large. The eigenvalues of $\mathbf{W}$ are 0 and w_d, so $\mathbf{W}$ has no positive eigenvalues. When r_e and r_i are equal, the synaptic connections contribute net inhibition: that is, letting $\mathbf{r}_+ = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, then $\mathbf{W}\mathbf{r}_+ = w_d\mathbf{r}_+$. However, when the excitatory and inhibitory rates differ, then the rates are amplified by the synaptic connections. That is, letting $\mathbf{r}_- = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$,

then $\mathbf{W}\mathbf{r}_- = w_s\mathbf{r}_+$. We refer to $\mathbf{r}_+$ as a sum mode and $\mathbf{r}_-$ as a difference mode.

Writing $\mathbf{r}(t) = r_+(t)\mathbf{r}_+ + r_-(t)\mathbf{r}_-$, the dynamics are

$$\tau\frac{dr_+}{dt} = -(1 - w_d)r_+ + w_s r_- \tag{3.2}$$

$$\tau\frac{dr_-}{dt} = -r_- \tag{3.3}$$

The network, despite recurrent connectivity in which all neurons are connected to all others, is acting as a two-layer feedforward network. The difference mode activates the sum mode with connection strength w_s, but there is no feedback from the sum mode onto the difference mode. Small perturbations of r_- can be amplified into large responses in r_+, but, as expected for a feedforward network, the amplification scales linearly with the summed synaptic strength, w_s, and can be arbitrarily large without affecting the stability of the network.

To understand the response to ongoing random input, it suffices to know the response to the input at a single time, because responses to inputs at different times superpose. Solving Eqs. 3.2-3.3 gives the time course of response to a perturbation at time 0, $\mathbf{r}(0)$:

$$r_+(t) = r_+(0)e^{-(1-w_d)t/\tau} + r_-(0)(w_s/\tau)e^{-t/\tau}\frac{e^{w_d t/\tau} - 1}{w_d/\tau} \tag{3.4}$$

$$r_-(t) = r_-(0)e^{-t/\tau} \tag{3.5}$$

The summed synaptic strength w_s scales the size of the amplification but the time course, $e^{-t/\tau}\frac{e^{w_d t/\tau}-1}{w_d/\tau}$, is independent of w_s. Note that, for $w_d \to 0$, the time course becomes $te^{-t/\tau}$, while for w_d large and negative, the time dependence approaches $e^{-t/\tau}$. Thus, the time dependence, which in all cases is independent of w_s, can be thought of as interpolating between $te^{-t/\tau}$ and $e^{-t/\tau}$ for increasingly negative w_d.

Figure 3.1 plots the time course of the response of a network with $w_E = w_I = 5$ ($w_d = 0$) to an initial perturbation in which excitatory rates are initially larger than inhibitory ($r_+(0) = 0$, $r_-(0) = 1$). The difference in rates drives both rates up equally, that is, it drives the sum mode, while the difference itself decays. As the difference decays, so does the source of the amplification, so the overall activity of the network grows only transiently. This is the basic mechanism of transient amplification in circuits with strong, balancing excitation and inhibition: differences in excitatory and inhibitory activity drive modes with equal excitatory and inhibitory parts, while the difference itself decays.
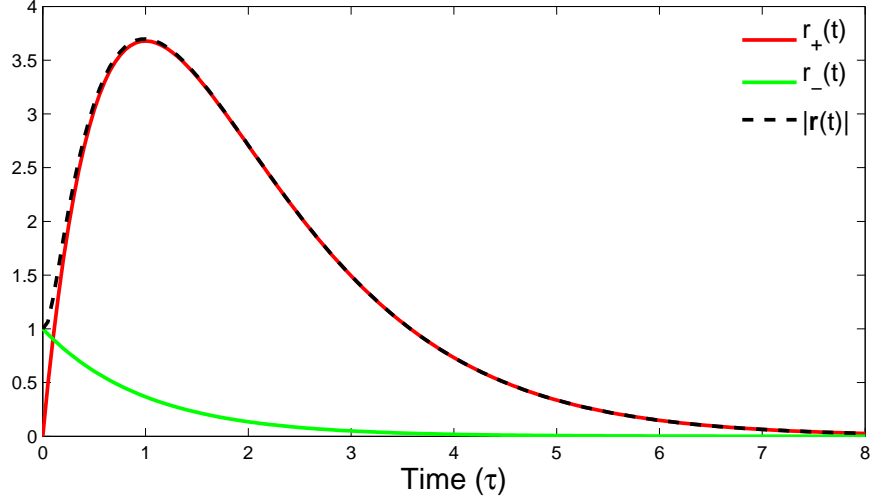
Figure 3.1: **Transient amplification in the two population case.** Plots of $r_+(t)$ and $r_-(t)$ in Eqs. 3.4 and 3.5 for the initial condition $r_+(0) = 0, r_-(0) = 1$, in which excitation is larger than inhibition. The dashed line is the total magnitude of the rate vector over time, $|\mathbf{r}(t)| = \sqrt{r_+(t)^2 + r_-(t)^2}$.

In spatially extended networks with many neurons, this process can selectively amplify specific spatial patterns of activity. For simplicity, take the number of excitatory and inhibitory neurons to be equal, and suppose that excitatory and inhibitory neurons, though making different patterns of projections, make projections that are independent of postsynaptic cell type. Then, if $\mathbf{A}$ describes the spatial pattern of excitatory projections and $\mathbf{B}$ of inhibitory projections, the full weight matrix is $\mathbf{W} = \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{A} & -\mathbf{B} \end{pmatrix}$. If $\mathbf{A}$ and $\mathbf{B}$ are $N \times N$, then $\mathbf{W}$ has N eigenvalues equal to the eigenvalues λ_i^D of $\mathbf{A} - \mathbf{B}$, and N eigenvalues equal to zero. We take inhibition to balance or dominate excitation, by which we mean the λ_i^D have real part ≤ 0. We let $\mathbf{e}_i^S$ be the eigenvectors of $\mathbf{A} + \mathbf{B}$ with eigenvalues λ_i^S, and note that these eigenvalues can be large. We define the difference modes $\mathbf{r}_i^- = \begin{pmatrix} \mathbf{e}_i^S \\ -\mathbf{e}_i^S \end{pmatrix}$ and the sum modes $\mathbf{r}_i^+ = \begin{pmatrix} \mathbf{e}_i^S \\ \mathbf{e}_i^S \end{pmatrix}$ and find that $\mathbf{W}\mathbf{r}_i^- = \lambda_i^S \mathbf{r}_i^+$. In turn, $\mathbf{W}\mathbf{r}_i^+$ gives a mixture of the $\mathbf{r}_i^+$ with weights linear in the eigenvalues λ_i^D, without feedback to the $\mathbf{r}_i^-$. Thus, each pair $\mathbf{r}_i^+$, $\mathbf{r}_i^-$ behaves much like the sum and difference modes $\mathbf{r}_+$, $\mathbf{r}_-$ in the simpler,

two-neuron model we studied previously, but with amplification weight λ_i^S instead of w_s. Spatial patterns with the largest eigenvalues of the average connectivity matrix $\mathbf{A} + \mathbf{B}$ will be most amplified, yet these eigenvalues do not affect the time course (further analyzed in Supplemental Materials). If connectivity is orientation-specific, then patterns resembling orientation maps are expected to be among those with largest λ_i^S.

We illustrate this by studying a simple model of synaptic connectivity based on known properties of V1. In this model, the strength of a synaptic connection between two neurons is determined by the product of Gaussian functions of distance and of difference in preferred orientation (see Methods in Supplemental Materials). The orientation map is a simple 4x4 grid of pinwheels shown in Figure 3.2a. The only difference between the patterns of excitatory and inhibitory synapses is that excitatory synapses extend over a much larger range of distances, as is true in layer II/III of V1 (Gilbert and Wiesel 1989). The orientation tunings of excitatory and inhibitory synapses are identical (Anderson et al. 2000a, Martinez et al. 2002). Inhibition is set strong enough that all the eigenvalues of the matrix have real part ≤ 0. We compute the eigenvectors of $\mathbf{A} + \mathbf{B}$, and illustrate the resulting difference modes $\mathbf{r}_i^-$ and sum modes $\mathbf{r}_i^+$ for the five largest amplification weights λ_i^S (Fig. 3.2b). We compare the output vectors – the sum modes – to stimulus-evoked orientation maps, each computed as the response of a rectified version of Eq. 3.1 to orientation-tuned feedforward input. In the mode corresponding to the largest transient, all the neurons increase or decrease their activity together. Kenet et al. (2003) filtered out such modes in their experiments because they can result from artifactual causes. The next two modes correspond to patterns that closely resemble evoked orientation maps. To characterize the time course of this amplification, we examine the time course of the overall size of the activity vector, $|\mathbf{r}(t)|$, in response to an initial perturbation consisting of one of the input vectors, *i.e.* one of the difference modes (Fig. 3.2c). The first mode follows the time course $te^{-t/\tau}$, while subsequent modes peak progressively sooner representing the movement from time course $te^{-t/\tau}$ toward time course $e^{-t/\tau}$.

In the more general case, when $\mathbf{W}_{EE}$, $\mathbf{W}_{EI}$, $\mathbf{W}_{IE}$, and $\mathbf{W}_{II}$ all have distinct structure, one cannot write a general solution, but one can infer that similar results should apply if excitation balances inhibition (see Supplementary Materials). More generally, any biological connection matrix has hidden feedforward connectivity. This is shown by the Schur decomposition, which for any matrix finds a (non-unique) orthonormal basis in which the matrix is upper triangular, with its eigenvalues on the diagonal. For a normal matrix,
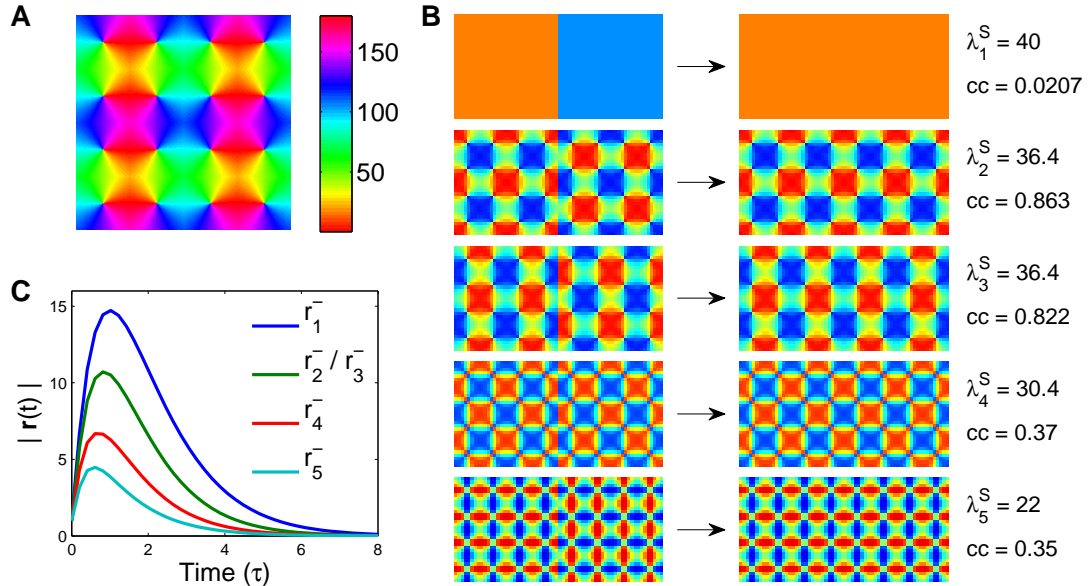
Figure 3.2: **Sum and difference modes are amplified in a spatially extended network. A)** Orientation map for both linear and spiking models. Color indicates preferred orientation in degrees. **B)** The sum (right) and difference (left) modes corresponding to the eigenvectors of $\mathbf{A} + \mathbf{B}$ with the largest eigenvalues. In each rectangle, the left half represents the 20×20 set of excitatory firing rates, while the right half represents the 20×20 set of inhibitory firing rates. In the difference modes (left), inhibitory rates are opposite to excitatory, while in the sum modes (right), inhibitory and excitatory rates are identical. Each difference mode is mapped into λ_i^S times the corresponding sum mode, where $i = 1$ to 5 labels the modes from top to bottom. Listed on the right are the amplification factors, λ_i^S, and the correlation coefficient (cc) of each sum mode with the evoked orientation map with which it is most correlated. The second and third patterns are strongly correlated with orientation maps. **C)** Plots of the time course of the magnitude of the activity vector, $|\mathbf{r}(t)|$, in response to an initial perturbation of unit length consisting of one of the difference modes shown in panel B. The perturbation labelled r_i^- corresponds to the ith mode in B.

the Schur decomposition is a diagonal matrix, but for a non-normal matrix, there are non-zero entries above the diagonal. These entries represent embedded feedforward connectivity between patterns: there can only be a connection from pattern i to pattern j if $i > j$. If the eigenvalues are small due to inhibition balancing excitation, but the original matrix entries are large, then there will be large entries off the diagonal in the Schur decomposition, because the sum of the absolute squares of the matrix entries is the same in any orthonormal basis. Thus, there will be large transient amplification.

3.3   Spiking Model

The linear rate model demonstrates the basic principles of transient amplification. To demonstrate that these principles apply to biological networks, in which neurons are nonlinear, spiking, and sparsely connected, we study a more detailed biophysical model capturing basic features of V1 connectivity. The model is highly simplified and is not meant to serve as a complete and accurate model of V1. It consists of 40,000 excitatory and 10,000 inhibitory integrate-and-fire neurons connected by fast conductance-based synapses. The excitatory and inhibitory neurons are each arranged on square grids spanning the orientation map used previously (Fig. 3.2a). The neurons are connected randomly and sparsely, with probabilities proportional to the weight matrix studied in the linear model, that is, dependent on distance and difference in preferred orientation. Each neuron receives feedforward Poisson input to generate sustained spontaneous activity. These input rates vary randomly with spatial and temporal correlations reflecting the likely structure of inputs to upper layers. During visually evoked activity each neuron receives a second Poisson input whose rate depends on the difference between the neuron's preferred orientation and the stimulus orientation. The network exhibits irregular activity as in other models of sparse balanced networks (Brunel 2000, van Vreeswijk and Sompolinsky 1996) (see Supplementary Figure 3.5).

By averaging the response of the network to a stimulus of a given orientation, we produce an evoked orientation map. Frames of spontaneous activity frequently resemble these evoked maps (Figs. 3.3a,b). We quantify the similarity between two patterns by the correlation coefficient between them. In agreement with the data of Kenet et al. (2003), frames of spontaneous activity show a distribution of correlation coefficients with a given evoked map that is about 2.5 times as wide as that for a control map (Fig. 3.3c)). That

is, the spontaneous activity shows greater similarity to the stimulus-evoked activity than expected by chance.

The timescale of the network activity is determined by two factors: the correlation time of the firing rates of the inputs to cortex, which is approximately 70ms, and the membrane time constant of the neurons, which is approximately 10ms during spontaneous activity. The former, being greater, largely determines the timescale of the fluctuations in the correlation coefficient (Figs. 3.3d,e). As long as inhibition balances or dominates excitation, the recurrent connectivity of the network has little effect on the timescale of network activity, just as in the linear model (see Supplemental Material).

3.4   Discussion

In cortical networks strong recurrent excitation coexists with strong feedback inhibition (Chagnac-Amitai and Connors 1989, Haider et al. 2006, Ozeki et al. 2007). This robustly produces an effective feedforward connectivity, in which small, patterned fluctuations in the differences between excitatory and inhibitory rates drive large, patterned fluctuations in their sum. Thus, transient amplification should be a ubiquitous feature of cortical networks, contributing both to spontaneous activity and to functional responses and their fluctuations. If inhibition balances or dominates excitation, then transient amplification occurs without elongation of lifetime. If some patterns show Hebbian slowing, then Hebbian amplification and transient amplification will coexist (see Supplemental Materials).

Given stochastic input, transient amplification produces orientation-map-like patterns in spontaneous activity (Kenet et al. 2003) in a network in which excitation and inhibition have similar orientation tuning. Previous work (Goldberg et al. 2004) found that these patterns could be explained by Hebbian slowing in a weakly recurrent, non-attractor network with "Mexican hat" connectivity in which inhibition is more broadly tuned for orientation than excitation. This creates positive eigenvalues for orientation-map-like patterns. However, intracellular recordings suggest that cells in V1 upper layers receive excitatory and inhibitory input with similar tuning (Anderson et al. 2000a, Martinez et al. 2002). Furthermore, the $\approx$ 80ms timescale of experimentally observed patterns (Kenet et al. 2003) places significant constraints on the degree of Hebbian slowing (see Supplemental Materials). Transient amplification represents a mechanism by which arbitrarily strong recurrent connectivity can shape activity while maintaining the fast dynamics normally associated
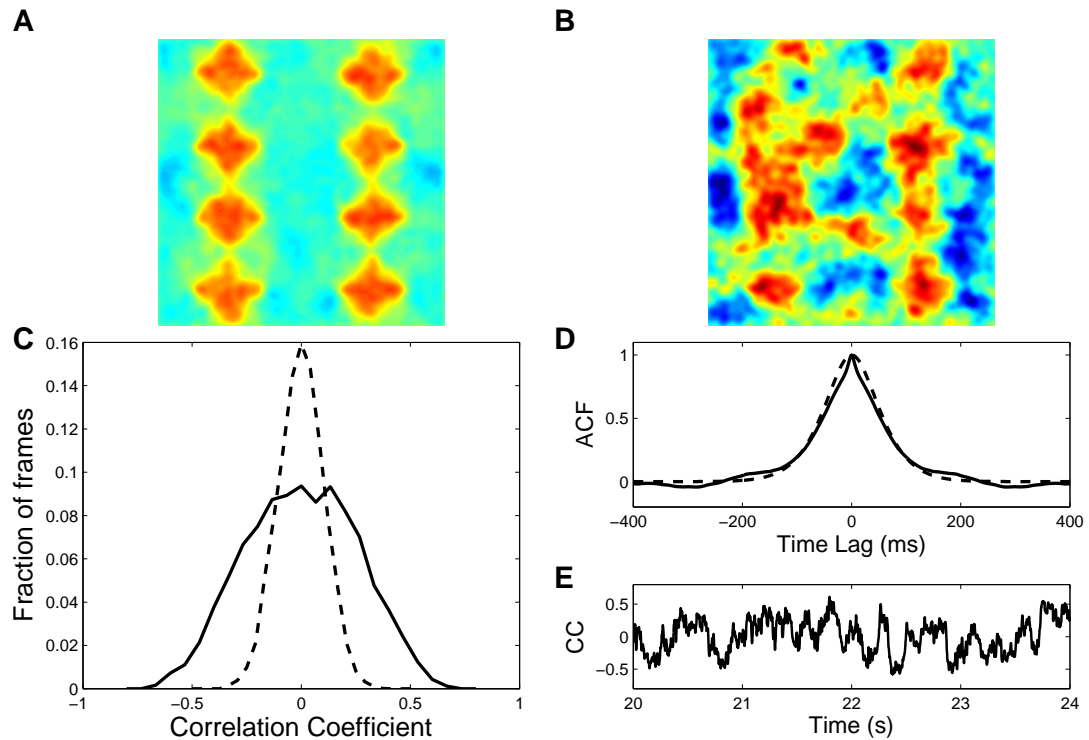
Figure 3.3: **Spontaneous patterns in a spiking model. A)** The 0° evoked map. **B)** Example of a spontaneous frame that is highly correlated with the 0° evoked map (correlation coefficient = 0.61). **C)** Distribution of correlation coefficients for the 0° evoked orientation map (solid line) and the shifted control (dashed line). The standard deviations of the two distribution are 0.25 and 0.1 respectively. The figure represents 40000 spontaneous frames corresponding to 40 seconds of activity. **D)** The solid line is the autocorrelation function of the time series of the correlation coefficient for the 0° evoked map and the spontaneous activity. It decays to 1/e of its maximum value in 70ms. The dotted line is the autocorrelation function of the input temporal kernel. It decays to 1/e of its maximum value in 73ms. **E)** A four-second-long example section of the full timeseries of correlation coefficients used to compute the autocorrelation function in panel D. All results are similar using an evoked map of any orientation.

with feed-forward networks.

The role of excitatory and inhibitory neurons in generating transient amplification is specific to neural systems. However, the principles involved are general and have implications for any biological systems that can be described as networks of interacting elements. Non-normal transients can contribute both to deterministic response, as has been studied in ecological networks (Neubert and Caswell 1997, Townley et al. 2007), and to structured noise, as studied here. Studies of intracellular networks have largely focused on simple "motifs" with only a few elements (Alon 2007), but non-normal interaction structure can already be recognized in some (Barkai and Leibler 2000, Süel et al. 2006). These have been studied in strongly nonlinear regimes yielding oscillations, thresholds, and multistability (Barkai and Leibler 2000, Süel et al. 2006), but transient amplification in the linear regimes of these and larger-scale cellular networks may have unexpected roles to play in cellular dynamics.

3.5   Supplemental Material

3.5.1   Methods

Linear Model

The linear model consists of overlapping 32x32 grids of excitatory and inhibitory neurons. Each neuron is assigned an orientation according to an orientation map consisting of a 4x4 grid of pinwheels. If we assume that a single pinwheel is 1mm^2 the network as a whole covers an area of 4mm^2. Distances are expressed in these coordinates, with the size of one side of the grid of neurons being 4 mm.

The strength of a synaptic connection between two neurons in the network is determined by the product of Gaussian functions of distance (r) and difference in preferred orientation (θ):

$$W_{ij}(r_{ij}, \theta_{ij}) \propto e^{-r_{ij}^2/w_r^2} e^{-\theta ij^2/w_\theta^2} \tag{3.6}$$

For excitatory synapses $w_r = 4\text{mm}$ and $w_\theta = 20°$. For inhibitory synapses $w_r = 0.4\text{mm}$ and $w_\theta = 20°$. The synaptic strengths are normalized on a neuron by neuron basis such that the sum of the excitatory and inhibitory inputs are both equal to 20.

In order to generate evoked orientation maps for this model we run simulations of the response of a rectified version of the linear equation to an orientation-tuned feed-forward input. The rectified equation is:

$$\tau \frac{d\mathbf{r}}{dt} = -\mathbf{r} + \mathbf{W}\phi(\mathbf{r}) \tag{3.7}$$

$$\phi(\mathbf{r})_i = \begin{cases} 0 & r_i <= 0 \\ r_i & r_i > 0 \end{cases} \tag{3.8}$$

The size of the feed-forward input to each neuron is a Gaussian function of the difference in the preferred orientation of the neuron and the orientation of the stimulus:

$$R_{\text{evoked}} = 4e^{-\theta^2/w_\theta^2} \tag{3.9}$$

Excitatory and inhibitory neurons receive identical inputs. The evoked orientation map is the pattern of activity obtained once the simulation reaches a steady state.

Spiking Model

The network consists of forty thousand excitatory and ten thousand inhibitory integrate-and-fire neurons. The voltage of each neuron is described by the equation:

$$C\frac{dV}{dt} = g_{\text{leak}}(E_{\text{leak}} - V) + g_e(E_e - V) + g_i(E_i - V) \tag{3.10}$$

Here C is the capacitance, g_{leak} is the leak conductance, E_{leak} is the resting membrane potential, and g_e and g_i are the excitatory and inhibitory conductances with corresponding reversal potentials E_e and E_i. When the voltage reaches the spike threshold, V_{thresh}, it is reset to V_{reset} and held there for a refractory period t_{refract}. Parameters, except for C, are from previous work (Murphy and Miller 2003) and are the same for excitatory and inhibitory neurons: $g_{\text{leak}} = 10\text{nS}$, $C = 400\text{pF}$, $E_{\text{leak}} = -70\text{mV}$, $V_{\text{thresh}} = -54\text{mV}$, $V_{\text{reset}} = -60\text{mV}$ and $t_{\text{refract}} = 1.75\text{ms}$. Excitatory and inhibitory reversal potentials are: $E_e = 0\text{mV}$ and $E_i = -70\text{mV}$. The capacitance is set such that, taking into account mean synaptic conductances associated with ongoing spontaneous activity, the membrane time constant is about 10 ms. At rest, with no network activity, the membrane time constant is 40 ms.

Conductances

The time course of synaptic conductances is modeled as a difference of exponentials:

$$g(t) = \sum_{\Delta t_j} \bar{g}\left(e^{-\Delta t_j/\tau_{\text{fall}}} - e^{-\Delta t_j/\tau_{\text{rise}}}\right) \tag{3.11}$$

Here Δt_j is defined as $(t - t_j)$, where t_j is the time of the jth pre-synaptic action potential and $t_j < t$. We include only fast synaptic conductances, AMPA and GABA$_{\text{A}}$, with $\tau_{\text{rise}} = 1\text{ms}$ and $\tau_{\text{fall}} = 3\text{ms}$. We have chosen to model excitatory and inhibitory conductances as having fast, identical time constants for simplicity. This equality and speed are not necessary for our results. We have focused on the asynchronous regime in which neurons fire irregularly and without global oscillations in overall rate, for which time constants must be chosen appropriately (Brunel 2000, Shriki et al. 2003, Sompolinsky and White 2005, Wang 1999), but this is not a tight constraint. Although we have not explored the issue extensively, we imagine that, so long as firing remains in the asynchronous regime, differences in excitatory and inhibitory timescales can be compensated by changes in the synaptic connectivity, as in a linear rate model.[1]

The size of the synaptic conductances evoked by a pre-synaptic action potential, $\bar{g}$, are set in terms of the total conductance integrated over time evoked by one presynaptic action potential in units of $\text{nS} \cdot \text{ms}$. This total conductance is $\bar{g}\tau_{\text{int}}$ where $\tau_{\text{int}} = \int_0^\infty dt\left(e^{-t/\tau_{\text{fall}}} - e^{-t/\tau_{\text{rise}}}\right)$. The values used are $\bar{g}_i = 0.0575\text{nS} \cdot \text{ms}/\tau_{\text{int}}$ and $\bar{g}_e = 0.00325\text{nS} \cdot \text{ms}/\tau_{\text{int}}$. We have chosen the overall and relative sizes of $\bar{g}_e$ and $\bar{g}_i$ to produce a certain degree of orientation-map like patterns in the spontaneous activity, while maintaining an average conductance during ongoing spontaneous activity of roughly 3-4 times

---

[1]In the linear model, consider a 2×2 network with one excitatory and one inhibitory neuron, with time constants τ and $k\tau$ respectively:

$$\tau \begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix} \frac{d}{dt} \begin{pmatrix} \mathbf{r}_e \\ \mathbf{r}_i \end{pmatrix} = -\begin{pmatrix} 1 - w_{ee} & w_{ei} \\ -w_{ie} & 1 + w_{ii} \end{pmatrix} \begin{pmatrix} \mathbf{r}_e \\ \mathbf{r}_i \end{pmatrix} \tag{3.12}$$

This network is equivalent to a network with equal time constants and a modified connectivity matrix:

$$\tau \frac{d}{dt} \begin{pmatrix} \mathbf{r}_e \\ \mathbf{r}_i \end{pmatrix} = -\begin{pmatrix} 1 - w_{ee} & w_{ei} \\ \frac{-w_{ie}}{k} & \frac{1+w_{ii}}{k} \end{pmatrix} \begin{pmatrix} \mathbf{r}_e \\ \mathbf{r}_i \end{pmatrix} \tag{3.13}$$

In other words, suppose we begin with a network with equal excitatory and inhibitory time constants. If we then lengthen the inhibitory time constant, but also appropriately increase the $E \to I$ and $I \to I$ weights to compensate, the network behavior will be unchanged.

the resting leak conductance (Destexhe and Paré 1999). Increasing the overall size of the conductances or the ratio of excitation to inhibition increases the strength of the patterns.

Synaptic connectivity

The neurons are laid out in an evenly spaced grid, 200x200 for excitatory neurons and 100x100 for inhibitory neurons. As a result the space between inhibitory neurons is twice as large as between excitatory neurons. As in the linear model each neuron is assigned an orientation from an orientation map consisting of a 4x4 grid of pinwheels.

The synaptic connectivity is sparse but otherwise similar to that of the linear model, with the probability of a synaptic connection between two neurons (P_c) determined by the product of Gaussian functions of distance (r) and difference in preferred orientation (θ):

$$P_c(r, \theta) \propto e^{-r^2/w_r^2} e^{-\theta^2/w_\theta^2} \tag{3.14}$$

For excitatory synapses $w_r = 4$mm and $w_\theta = 20°$. For inhibitory synapses $w_r = 0.4$mm and $w_\theta = 20°$. P_c is normalized separately for excitatory and inhibitory connections for each neuron such that the expected number of connections received by each neuron (averaged over many random draws with the given neuron's P_c) would be $N_e = 100$ excitatory and $N_i = 25$ inhibitory connections.

Because the connections are random, some neurons will receive more or fewer connections, resulting in a range of average firing rates and mean voltages. In order to obtain similar firing rates for all neurons in the network we scale up or down the excitatory and inhibitory synaptic conductances received by each neuron. Specifically, let $R = \frac{N_e \bar{g}_e}{N_i \bar{g}_i}$ be the desired ratio of excitatory to inhibitory conductances onto a cell. Then we set the actual ratio to R for each cell, by scaling all the excitatory conductances onto a given neuron by f_e and all the inhibitory conductances onto the neuron by f_i according to the equations:

$$f_e = \frac{2.0}{1 + 1/x} \tag{3.15}$$

$$f_i = \frac{2.0}{1 + x} \tag{3.16}$$

with $x = N_e n_i/(N_i n_e)$. Here n_e and n_i are the actual number of excitatory and inhibitory synapses received by the given neuron. This sets $\frac{n_e f_e \bar{g}_e}{n_i f_i \bar{g}_i} = R$ for the cell, while also. setting

$(1.0 - f_e) = (f_i - 1.0)$. The latter condition is designed to imitate a homeostatic synaptic plasticity rule in which excitation and inhibition are increased or decreased proportionally in order to maintain a certain average firing rate.

Spontaneous Activity

During spontaneous activity each neuron receives a background feed-forward input consisting of an excitatory Poisson spike train. The rates of the background inputs are randomly determined by convolving white noise with a spatial and temporal filter. The spatial filter is a Gaussian function with a width of $200\mu m$. The temporal filter is given by the function:

$$K(t) = t^2 e^{-\gamma t} \tag{3.17}$$

Here γ determines the speed of the filter and is set to $\gamma = 40$Hz. This is slower than the average temporal kernel of LGN cells (Wolfe and Palmer 1998), and is closer in speed to the temporal kernels of simple cells in layer 4 (DeAngelis et al. 1993, 1999) that provide the main input to layers 2/3. For simplicity, we do not replicate the biphasic nature of real LGN or simple-cell temporal kernels, but simply try to capture the overall time scale.

We set the standard deviation of the the unfiltered input noise to 1250 Hz and normalize the integrals of the squares of the spatial and temporal filters to one to produce filtered noise with the same standard deviation. This rate noise is then added to a mean background rate of 10250 Hz. The size of the background synaptic conductance is set to 0.00025 nS · ms. Steady input at the mean background rate is sufficient to just barely make the neurons fire (less than 1Hz), while steady input at the mean plus three standard deviations would result in a firing rate of about 24Hz.

Evoked Maps

Visually evoked orientation maps are generated by averaging frames of network activity (see Comparison to Experiment below) for three seconds in response to a visually evoked input added to the background input. The evoked input is a Poisson spike train with a rate determined by a Gaussian function of the difference between the preferred orientation of each neuron and the orientation of the stimulus:

$$R_{\text{evoked}} = 10000e^{-\theta^2/w^2_{\text{evoked}}} \tag{3.18}$$

with $w^2_{\text{evoked}} = 20^o$. The size of the individual synaptic conductances is the same for the evoked and spontaneous inputs, $0.00025 \text{ nS} \cdot \text{ms}$.

Comparison to Experiment

To compare spontaneous and visually evoked activity we compute the correlation coefficient between frames of spontaneous activity and the visually evoked orientation map every millisecond. A frame consists of the shadow voltages of all the neurons filtered with a Gaussian filter with a width of 80 μm after subtracting the mean of each frame. The shadow voltage is simply the membrane potential of the neuron integrated continuously in time in the absence of a spike threshold, *e.g.* it is not reset when it reaches spike threshold. This is meant to approximate the voltage in the dendrites. The filter is used because the voltages of the individual neurons are very noisy and we are comparing to experimental data that does not resolve individual neurons. The filter width is chosen to conservatively underestimate the point spread function of the experimental images (Polimeni et al. 2005). Nonetheless, because both the filtering and the noise in the experimental system are unknown, it is difficult to directly compare the correlation coefficients from the model to those from the actual data. As a control we also compute the correlation coefficient between frames of spontaneous activity and a "shifted" visually evoked orientation map. The shifted map has been offset horizontally and vertically by 0.5mm, one half the width of a single pinwheel, with periodic boundary conditions. Although the width of the distribution of correlation coefficients depends on the width of the Gaussian filter used, the ratio of the widths of the real and control distributions does not (Fig. 3.4), so it is this ratio rather than the absolute width of the distribution that is best compared to experiment.

3.5.2 Asynchronous, irregular activity in the spiking model, and the correspondence between spiking and rate models

The spiking model studied here operates in the "asynchronous irregular" regime (Brunel 2000) characterized by irregular spiking response and absence of global rate oscillations (Fig. 3.5), as in models of sparse balanced networks with unstructured random
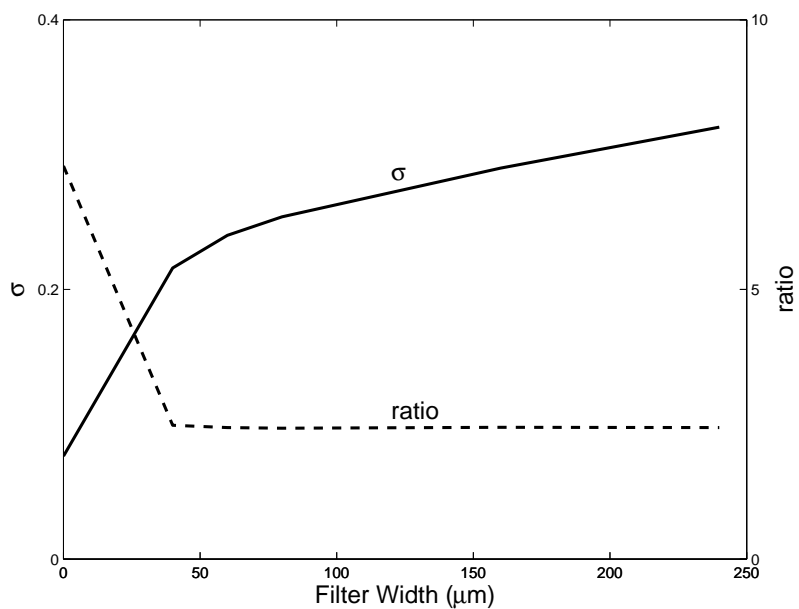
Figure 3.4: **Effects of smoothing on the correlation coefficient distribution.** Solid line is the width of the correlation coefficient distribution for the 0° evoked map plotted as a function of the filter width. Dashed line is the ratio of the standard deviation of the correlation coefficient distribution to the control distribution.

connectivity (Brunel 2000, van Vreeswijk and Sompolinsky 1996). The coefficient of variation for inter-spike intervals (ISIs) is around 1 (Fig. 3.5a), and the ISI distribution is essentially exponential (Fig. 3.5c), indicating Poisson-like firing. The average firing rate in spontaneous activity fluctuates around 15 Hz without oscillations (Fig. 3.5b).
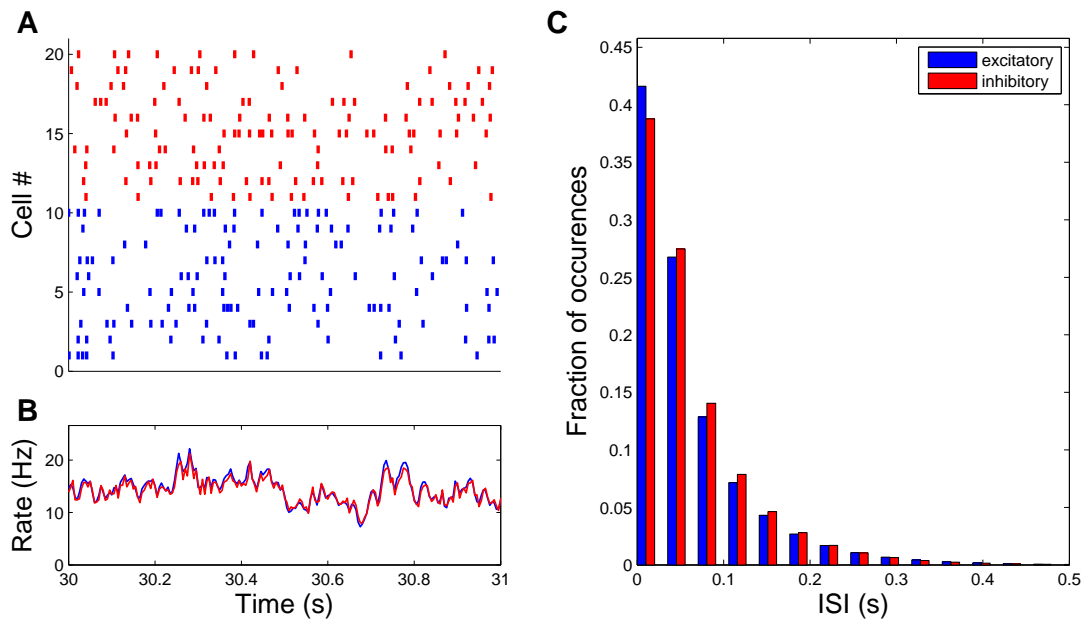
Figure 3.5: **Asynchronous, irregular activity in the spiking model.** During spontaneous activity excitatory neurons fired irregularly with a mean firing rate of 15 Hz and an average coefficient of variation (CV) for inter-spike intervals (ISI) of 1.0. Inhibitory neurons were similar with mean firing rates of 14.5 Hz and a CV of .95. ISIs for both types of neuron have a roughly exponential distribution. **A**) Spike raster plots over a one second long interval for 10 randomly selected excitatory (blue) and inhibitory (red) neurons. **B**) The average firing rate computed in 5ms bins of the entire population of excitatory (blue) and inhibitory (red) neurons for the same period. **C** Histogram showing the relative frequencies of different ISIs for excitatory (blue) and inhibitory (red) neurons.

In the asynchronous irregular regime, mean field theory can be applied to derive expressions for firing rates from a spiking model (*e.g.* Brunel 2000, Lerchner et al. 2006, Shriki et al. 2003, Sompolinsky and White 2005). Furthermore, for a statistically stationary input for which the system is fluctuating relatively weakly around the mean rates it would have in response to the mean input, as is the case for the spontaneous activity studied here, one can derive linear dynamical equations for the rate (although the best linear description

has a band-pass temporal filter, rather than the low-pass filter used here for the rate model) (Shriki et al. 2003, Sompolinsky and White 2005).[2] One imagines that the mean connectivity matrix from which the sparse random connectivity is drawn should provide a reasonable description of the connectivity in this linear model, again by mean field arguments (given enough inputs, the input a neuron receives from the sparse random sampling should show small deviations from the input it would receive under the mean connectivity matrix). Together these provide an intuitive but speculative reasoning as to why the linear rate model we studied should capture key aspects of the behavior of the spiking model we studied. Obviously, these ideas need more careful study.

3.5.3 Time constant of activity in the spiking model

As shown in Fig. 3.3d, the activity in the spiking network model studied in the main text is not slowed relative to the feedforward inputs to the network. This does not depend on the input time scale, as it is also true for a simulation in which the input rates fluctuate on a substantially faster timescale (Fig. 3.6). This matches the prediction of the linear model for a balanced network in which all eigenvalues are less than or equal to zero, as in Figure 3.2.

Strictly, the linear model predicts that the autocorrelation of the correlation coefficients in the spontaneous maps should be given by the autocorrelation of the convolution of the input temporal kernel with some function ranging in time course from $te^{-t/\tau}$ (when the orientation-map-like patterns have eigenvalues equal to 0) to $e^{-t/\tau}$ (when the orientation-map-like patterns have eigenvalues significantly less than 0). In Figs. 3.3d and 3.6, we simply compared to the autocorrelation of the input temporal kernal. In Fig. 3.7, we show the effects of including the convolutions. For the longer-time-course input temporal kernel used in the text, the differences are indiscernible (Fig. 3.7a). For the shorter-time-course input temporal kernel of Fig. 3.6, small differences can be seen, and indeed the time course seems best predicted by convolution with $e^{-t/\tau}$ (Fig. 3.7b).

---

[2]This correspondence was derived by (Shriki et al. 2003, Sompolinsky and White 2005) on the assumption that a neuron receives a large enough number of uncorrelated pre-synaptic spikes in one integration time that fluctuations in this number for a fixed network firing rate can be neglected. We speculate that, even for the sparsely connected network studied here, this approximation is sufficient to explain why a simple linear model captures key aspects of spiking model behavior, although this requires further study. Our neurons have about a 10 ms time constant, so at a 15 Hz average firing rate, with 100 excitatory and 25 inhibitory connections, they will receive a mean of about 15 excitatory and 3.8 inhibitory inputs in one integration time. Fluctuations in number, relative to the mean N, are expected to be of size $1/\sqrt{N}$, that is, about 25%
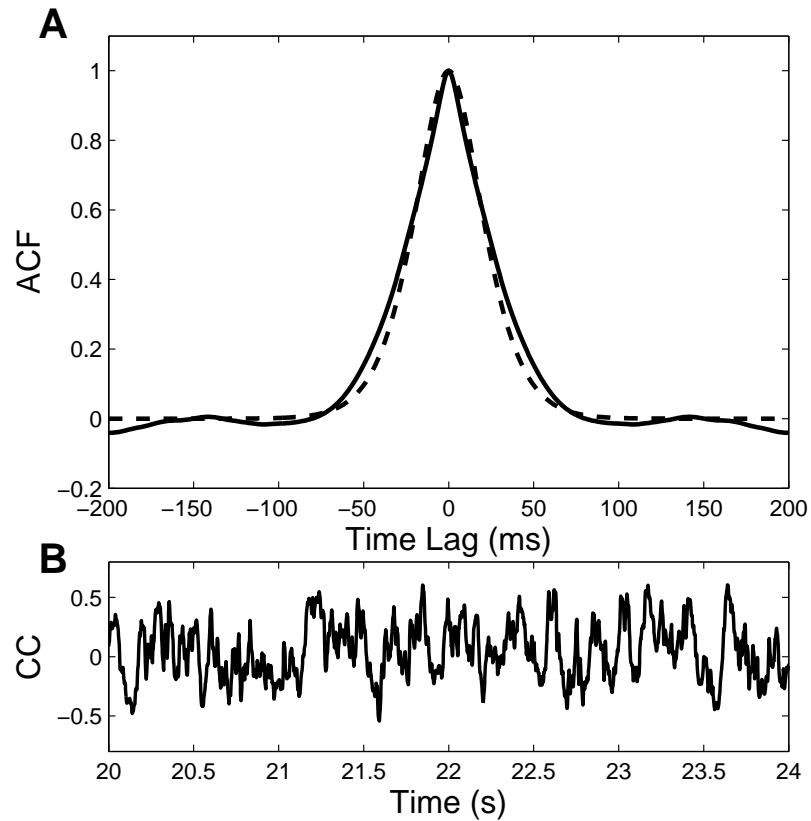
Figure 3.6: **Time constant of activity in the spiking model with faster input.**
A) Autocorrelation function of the correlation coefficient time series for a simulation with
faster fluctuations in the rate of background inputs ($\gamma = 100Hz$ vs $40Hz$, see Methods)
but otherwise identical to that presented in Figure 3. The solid line is the autocorrelation
function of the time series of the correlation coefficient for the 0° evoked map and the
spontaneous activity. It decays to 1/e of its maximum value in 33ms. The dotted line is
the autocorrelation function of the input temporal kernel. It decays to 1/e of it's maximum
value in 29ms. **B**) A four-second long example section of the full timeseries of correlation
coefficients used to compute the autocorrelation function in panel A.
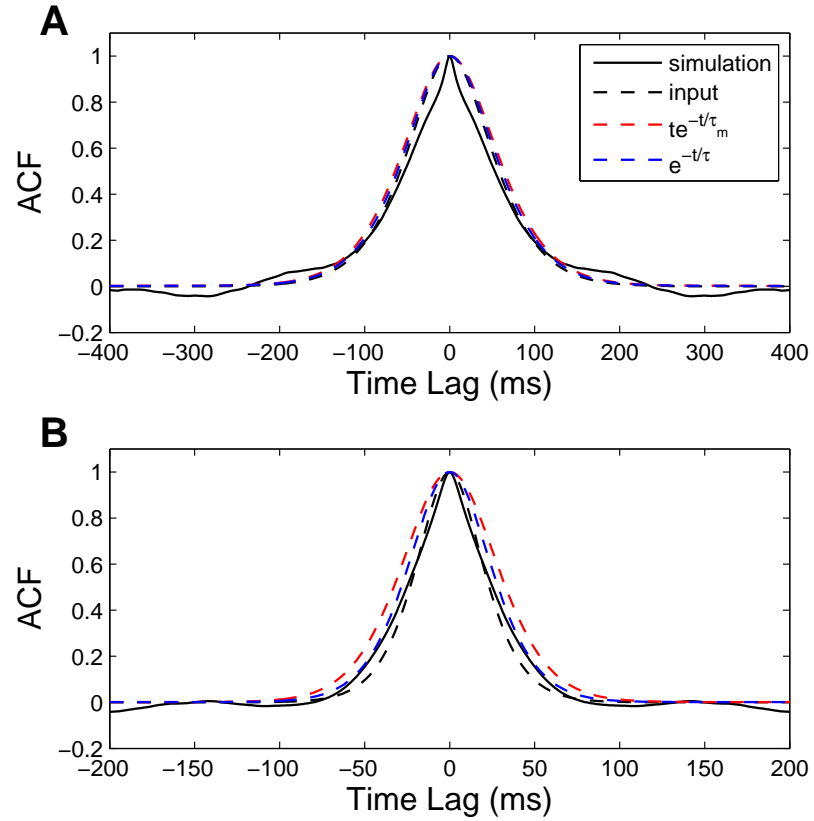
Figure 3.7: **Comparision of the ACF of the spiking model with the predictions of the linear model.** Comparison between the ACF for the correlation coefficient between the spontaneous activity and the 0° evoked map (solid black line), the input temporal kernel (dashed black line), and the input temporal kernel convolved with te^{-t/τ_m} (red dashed line) or e^{-t/τ_m} (blue dashed line). In all cases $\tau_m = 10$ms. Panel A corresponds to the simulation in Figure 3 with $\gamma = 40$Hz. The half width at $1/e$ heights for the curves are 73 ms, 70 ms, 76 ms, and 79 ms for the dashed black, solid black, dashed blue and dashed red lines respectively. Panel B corresponds to the simulation in Figure 3.6 with $\gamma = 100$Hz. The half width at $1/e$ heights of the curves are 29 ms, 33 ms, 35 ms, and 40 ms for the dashed black, solid black, dashed blue and dashed red lines respectively.

We can put the idea that the mechanisms of the linear model underly the results of the spiking model to a stronger test, by showing that the time course of the amplified patterns is not slowed even as the strength of the recurrent connectivity and thus the strength of the amplification is increased. The linear model predicts that, for a pattern of recurrent circuitry that has no positive eigenvalues, all amplification is due to the transient amplification mechanism. The strength of the amplification will increase as the strength of the recurrent circuitry is scaled up, but the time scale of the activity will be unaffected (more precisely, the activity might be slightly sped up, due to the possible effects of an increasingly negative eigenvalue, which as just noted can cause small changes in the function that convolves the input temporal kernel and also due to decreases in the cellular time constant τ induced by increased activity; but at any rate, the time scale will not be slowed down). On the other hand, if the amplified pattern has a positive eigenvalue, then the Hebbian amplification mechanism and the transient amplification mechanism will both contribute. If the strength of the recurrent circuitry is scaled up, the positive eigenvalue will be increased, and so the time scale of the amplified pattern will be slowed, which is the signature of the involvement of the Hebbian amplification mechanism.

To test this, we consider two models. One is the model of Figs. 3.2-3.3, in which excitation and inhibition have identical orientation tuning. In this case, all eigenvalues in the linear model are ≤ 0. In the second model, a "Mexican hat" connectivity is used, in which inhibitory connections have wider orientation tuning than excitatory inputs. In the linear version of this model, orientation-map-like patterns have positive eigenvalues. Each neuron in the second model receives exactly the same summed excitatory and summed inhibitory input as in the first model (both in the linear versions of the models and in the spiking versions of the models). However, in the second model, the excitatory input a cell receives comes from cells with a narrower range of preferred orientations than the inhibition it receives, so that more excitation than inhibition is received from cells with nearby preferred orientations and more inhibition than excitation is received from cells with more distant preferred orientations. Thus, orientation-map-like patterns, in which neurons with similar preferred orientation have positive activity and neurons with more distant preferred orientations have negative activity, acquire positive eigenvalues. In this case, as we increase the strength of recurrent connections, the dynamics of the orientation-map-like

_____

for excitatory inputs and about 50% for inhibitory inputs.

patterns should become noticeably slowed.

In Figure 3.8 we compare the effect of increasing recurrent strength on these two types of network. The blue lines in Figure 3.8 correspond to the first model, in which excitatory (w_θ^e) and inhibitory (w_θ^i) orientation tuning widths are equal, with $w_\theta^e = w_\theta^i = 20°$. For this network, 100% recurrent strength on the x-axis corresponds to the network presented in Figure 3.3. The green lines correspond to a network in which inhibitory tuning is wider, with $w_\theta^i = 50°$. Other than the width of the inhibitory tuning the networks have identical parameters.
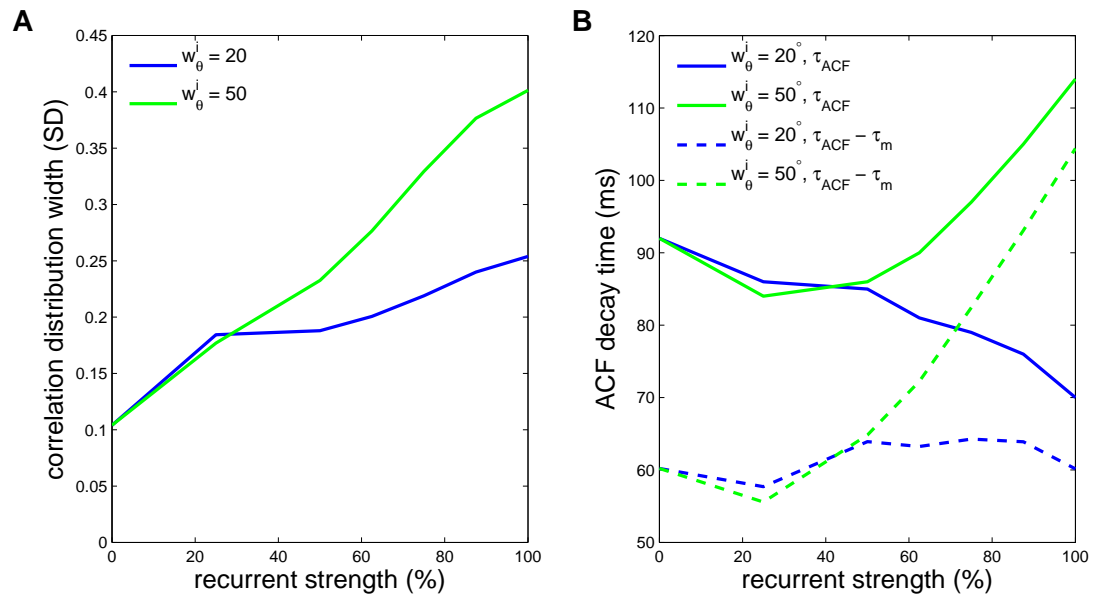
Figure 3.8: **Effects of increasing recurrent strength in networks with and without positive eigenvalues.** A comparison of the effects of increasing the strength of recurrent synapses in a network with equal excitatory and inhibitory tuning widths (blue line, $w_\theta^e = w_\theta^i = 20°$) or wider inhibitory tuning (green line, $w_\theta^e = 20°$, $w_\theta^i = 50°$). **A)** Plot of the effect of increasing recurrent strength on the width of the distribution of correlation coefficients with the $0°$ evoked map. **B)** Plot of the effect of increasing recurrent strength on the time constant of network activity as measured by the time required for the auto-correlation function of the correlation coefficient timeseries to decay to $1/e$ of its maximum value (τ_{ACF}). The membrane time constant of the neurons (τ_m), taking into account the average synaptic conductance associated with ongoing spontaneous activity, decreases with increasing recurrent strength. The blue and green dashed lines plot $\tau_{ACF} - \tau_m$ for $w_\theta^i = 20°$ and $w_\theta^i = 50°$ respectively. In both panels a strength of 100% corresponds to the synaptic strengths in the network presented in the paper (Figure 3).

For both networks, increasing the strength of recurrent connections increases the width of the distribution of correlation coefficients with the 0° evoked map (Fig. 3.8a). The width increases more strongly for the $w_\theta^i = 50°$ network, as expected for two reasons. First the patterns in the $w_\theta^i = 50°$ network are amplified both by slowing associated with a positive eigenvalue and by transient amplification, while the $w_\theta^i = 20°$ network has only the transient amplification. Second, one expects the correlation coefficient to grow to a plateau with increasing recurrent strength for the network that only has transient amplification, but not for the network that also has Hebbian amplification, for the following reason. As can be seen in Eq. 4 and more generally below in section 3.5.6, two terms contribute to the size of a given pattern in the output: one just represents the input and its decay, one represents the transient amplification. With no recurrent circuitry, only the first term contributes. If there are no positive eigenvalues, then as the recurrent circuitry is scaled up, the second term becomes more and more dominant, until for sufficiently strong recurrent circuitry the first term is negligible. Then the amplitude of each pattern should be proportional to the time integral of the second term, which itself should asymptote for sufficiently strong recurrent circuitry.[3] The correlation coefficient of the spontaneous activity with one particular pattern just represents the square of that pattern's amplitude divided by the sum of the squares of the amplitudes of all of the patterns, so it should asymptote when the amplitudes of the patterns asymptote. This is not the case in the $w_\theta^i = 50°$ network where there is a positive eigenvalue. The size of the eigenvalue (λ) grows linearly with the recurrent strength, but the degree of selective amplification is proportional to $1/(1-\lambda)$. For λ close to one the pattern associated with the largest eigenvalue will be amplified much more than other patterns and will dominate the activity of the network.

To compare the degree of slowing in the two networks we plot the time required for the autocorrelation function of the timeseries of correlation coefficients between the spontaneous activity and the 0° evoked map to decay to 1/e of it's maximum value (τ_{ACF})

---

[3]Using the language of Eq. 4: For $w_d = 0$, the term grows with w_s and thus grows linearly with recurrent strength. For $w_d < 0$, both $|w_d|$ and w_s grow linearly with recurrent strength, and for sufficiently strong recurrent strength the term becomes proportional to w_s/w_d, which is constant with recurrent strength. Thus, the amplitude of any patterns with $w_d = 0$ will grow with increasing recurrent strength relative to patterns with $w_d < 0$, but if all patterns have $w_d < 0$ then their relative strengths will asymptote. For our full model circuit, w_d is replaced with λ_j^D and w_s with λ_i^S, where j and i specify the output and input patterns of the amplification (section 3.5.6), but the analysis is much the same. In our model circuit, all the λ_j^D have real part < 0 except one, that corresponding to the first pattern shown in Fig. 3.2b, which is a spatially uniform or "DC" pattern and has $\lambda^D = 0$. However, following the methods used in experiments, we subtract the DC component from the frames before computing their correlation coefficient with the evoked map, so this pattern does not contribute to the correlation coefficient.

versus recurrent strength in Figure 3.8b. Increasing the strength of recurrent connections causes a decrease in τ_{ACF} for the $w_\theta^i = 20°$ network. Intially τ_{ACF} decreases for the $w_\theta^i = 50°$ network as well, but for larger values of recurrent strength τ_{ACF} shows a significant increase. The decrease in τ_{ACF} is the result of decreasing membrane time constants (τ_m) for the neurons in the two networks as mean conductances increase. With no positive eigenvalue this is the dominant effect of increasing recurrent strength for the $w_\theta^i = 20°$ network. For the $w_\theta^i = 50°$ network the activity patterns that are correlated with the evoked map have positive eigenvalues and their time constants therefore grow as $1/(1 - \lambda)$, while λ increases linearly with recurrent strength. Initially the decrease in the membrane time constant is dominant, but this effect is soon overwhelmed by the slowing effect of the positive eigenvalues. To isolate the effects of the recurrent connectivity beyond simply decreasing the membrane time constant we can subtract the numerically measured average membrane time constant associated with ongoing background activity at each recurrent strength. The resulting curves are shown as blue and green dashed lines for $w_\theta^i = 20°$ and $w_\theta^i = 50°$ respectively. The curve for $w_\theta^i = 20°$ is essentially flat, while the curve for $w_\theta^i = 50°$ increases quickly once recurrent strength gets above about 30%.

3.5.4 Constraints on models from the time scales observed in Kenet et al. (2003)

In the text, we stated "the ≈ 80ms timescale of experimentally observed patterns (Kenet et al. 2003) places significant constraints on the degree of Hebbian slowing". Here we amplify this thought.

The 80ms number reported by Kenet et al. (2003) is rough. It was not a direct measure of the autocorrelation time of the correlation coefficient, but instead was the "average transition time" in preferred orientation as measured by the template from a Kohonen-map algorithm that best matched the instantaneous snapshot of activity.

One expects the autocorrelation time of the correlation coefficient for a pattern to be given roughly by the sum of the correlation time of the inputs and the time constant of the network activity for that pattern. We expect the correlation time of inputs to upper layers to be many tens of ms, based on the temporal kernels of inputs from lateral geniculate nucleus to layer 4 of V1 (Wolfe and Palmer 1998) or of simple cells in V1 (DeAngelis et al. 1993, 1999), which should provide the dominant input to V1 upper layers (*e.g.* Martinez

et al. 2005). The 70 ms time we used in the main paper seems reasonable based on the studies of simple cells, but shorter times are also reasonable. Overall, we can say that the network time constant for the pattern should not account for much more than half of the roughly 80 ms time constant, and probably less.

The network time constant in a Hebbian-assembly model is given by $\tau/(1 - \lambda)$ where τ is the intrinsic decay time of cortical activity without recurrent connections, as in Eq. 1, and λ is the eigenvalue of $\mathbf{W}$ of the pattern. Goldberg et al. (2004) showed that $\lambda = 0.6$ in a Hebbian-assembly model would give a widening of the distribution of correlation coefficients comparable to that of Kenet et al. (2003). This gives a network time constant of 2.5τ. Thus, the intrinsic decay time, τ, needs to be on the order of 10 ms in order for a Hebbian-assembly explanation of the amplification to be workable. This intrinsic decay time is certainly plausible, but it is also plausible that the intrinsic time scale might be considerably larger. In reducing spiking models to rate models, the intrinsic time scale typically receives a key contribution from the time scale of the synaptic conductances (Ermentrout 1994, Shriki et al. 2003, Sompolinsky and White 2005). Intracortical excitatory synapses clearly receive significant contribution from NMDA receptor-mediated conductances (Feldmeyer et al. 1999, 2002, Fleidervish et al. 1998), although the exact contribution depends on the voltage of recording and Mg^{++} and glycine concentrations in the bath. These conductances have slow dynamics, with time constants > 100ms at physiological temperature (Monyer et al. 1994). If the effective τ were several 10's of ms or larger, the Hebbian-assembly scenario would produce too long a time scale.

3.5.5 Non-normal matrices

In the main text we make some brief statements about the behavior of non-normal matrices that may be puzzling and surprising to those whose previous experience with linear systems has been largely restricted to normal matrices (a group that would have included ourselves until recently). For this audience, we here present in more detail some basic facts about the behavior of non-normal matrices. A much fuller treatment can be found in Trefethen and Embree (2005). Our treatment differs from theirs in that we emphasize the role of the Schur decomposition, which provides a natural interpretation in terms of networks.

Normal matrices are matrices $\mathbf{M}$ that satisfy $\mathbf{M}^\dagger\mathbf{M} = \mathbf{M}\mathbf{M}^\dagger$, or equivalently

matrices that have a complete orthonormal basis of eigenvectors, where $\mathbf{M}^\dagger$ is the complex conjugate of the transpose of $\mathbf{M}$.[4] For real matrices, $\mathbf{M}^\dagger = \mathbf{M}^T$, the transpose of $\mathbf{M}$.

Why do transients arise that are not predicted by the eigenvalues?

Assuming $\mathbf{W}$ has a complete basis of eigenvectors, Eq. 1 can be formally solved by transforming to the eigenvector basis. Let $\mathbf{W}$ have eigenvectors $\mathbf{e}_i$ with eigenvalues λ_i. Then the solution is:

$$\mathbf{r}(t) = \sum_i \mathbf{e}_i r_i(0) e^{-(1-\lambda_i)t/\tau} \tag{3.19}$$

Here, $r_i(0)$ is the magnitude of $\mathbf{r}(0)$ in the direction of the ith eigenvector, such that $\mathbf{r}(0) = \sum_i r_i(0)\mathbf{e}_i$; it is given by $r_i(0) = \left[\mathbf{C}^{-1}\mathbf{r}(0)\right]_i$ where $\mathbf{C}$ is the matrix whose columns are the eigenvectors. If $\mathbf{W}$ is non-normal, this solution can include large transient responses to small perturbations as we have shown in Eqs. 4-5. How can this be consistent with Eq. 3.19 when all $\Re(\lambda_i) < 1$ (where $\Re(x)$ is the real part of x)?

Consider a two-dimensional case. We let $\mathbf{M} = -(\mathbf{1} - \mathbf{W})$, with eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_2$ and corresponding eigenvalues $\beta_1 = \lambda_1 - 1$ and $\beta_2 = \lambda_2 - 1$. We assume $\mathbf{M}$ is non-normal, which is true if and only if $\mathbf{W}$ is non-normal. Equation 3.19 becomes

$$\mathbf{r}(t) = \mathbf{e}_1 r_1(0) e^{\beta_1 t/\tau} + \mathbf{e}_2 r_2(0) e^{\beta_2 t/\tau} \tag{3.20}$$

Naively, for $\Re(\beta_1) < 0$ and $\Re(\beta_2) < 0$, this appears to describe the monotonic decay of two independent components. The fallacy is that, because $\mathbf{M}$ is non-normal, $\mathbf{e}_1$ and $\mathbf{e}_2$ are not orthogonal, so the decaying components are not independent. We take $|\mathbf{e}_1| = |\mathbf{e}_2| = 1$. Then,

$$
\begin{aligned}
|r(t)|^2 &= r_1(0)^2 e^{2\beta_1 t/\tau} + r_2(0)^2 e^{-2\beta_2 t/\tau} + 2\Re\left(\mathbf{e}_1 \cdot \mathbf{e}_2\right) r_1(0) r_2(0) e^{(\beta_1+\beta_2)t/\tau} & (3.21) \\
&= r_1(0) e^{\beta_1 t/\tau} \left(r_1(0) e^{\beta_1 t/\tau} + r_2(0)\Re\left(\mathbf{e}_1 \cdot \mathbf{e}_2\right) e^{\beta_2 t/\tau}\right) & (3.22) \\
&+ r_2(0) e^{\beta_2 t/\tau} \left(r_2(0) e^{\beta_2 t/\tau} + r_1(0)\Re\left(\mathbf{e}_1 \cdot \mathbf{e}_2\right) e^{\beta_1 t/\tau}\right) & (3.23)
\end{aligned}
$$

If $r_1(0)$ and $r_2(0)$ have opposite signs, this gives two weighted differences of exponentials, which can grow enormously in time if $r_1(0)$ and $r_2(0)$ are large and $\Re(\mathbf{e}_1 \cdot \mathbf{e}_2)$ is of order 1,

---

[4] The overall idea underlying this equivalence is: the right eigenvectors of $\mathbf{M}^\dagger$ are the conjugate transpose of the left eigenvectors of $\mathbf{M}$. Two matrices share a common basis of eigenvectors if and only if they commute. Thus, iff $\mathbf{M}^\dagger$ and $\mathbf{M}$ commute, the right and left eigenvectors of $\mathbf{M}$ are identical (meaning that one set is the conjugate transpose of the other). These are mutually orthonormal, so iff they are identical, they constitute an orthonormal basis.

before eventually decaying to zero. Of course, for a normal matrix, $\mathbf{e}_1 \cdot \mathbf{e}_2 = 0$ and so there is monotonic decay.

Furthermore, as the eigenvectors approach being in the same direction (*i.e.*, become very nonorthogonal; $\Re(\mathbf{e}_1 \cdot \mathbf{e}_2) \approx 1$), it is generic for $r_1(0)$ and $r_2(0)$ to be large and of opposite signs. The initial condition will generically contain some component orthogonal to the average of $\mathbf{e}_1$ and $\mathbf{e}_2$, but if $\mathbf{e}_1$ and $\mathbf{e}_2$ are almost in the same direction, this component can only be expressed as a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$ by combining a large multiple of $\mathbf{e}_1$ with a large and opposite multiple of $\mathbf{e}_2$, so that the two largely cancel leaving only the component of the correct size orthogonal to their average. These large and canceling components of the initial condition then decay at different rates, giving a difference of exponentials that grows into a large transient before ultimately decaying back to zero.

More insight can be obtained as follows. If we made a transformation to the eigenvector basis, a basis in which the eigenvectors are orthogonal, then in this basis Eq. 3.20 would describe monotonic decay. However, because the eigenvectors are not orthogonal in the original basis, the transformation to the eigenvector basis is not unitary, that is, it does not preserve the lengths of or angles between vectors. Thus, what appears to be monotonic decay of vector length in the eigenvector basis can correspond to transient increase in vector length in the original basis. If we wish to understand the changes in vector length that occur in the original basis, we must restrict ourselves to transformations that preserve vector length and angles, that is, to unitary transformations, which transform to a basis that is orthonormal as judged in the original basis. If $\mathbf{M}$ is non-normal, it cannot be diagonalized by such a transformation – it is diagonalized by the basis of eigenvectors, which are not orthogonal in the original basis. How close to diagonal can we make the matrix by transformation to an orthogonal basis? The answer, given by the *Schur decomposition*, is that we can make the matrix upper triangular, with the eigenvalues on the diagonal and all other nonzero entries above the diagonal; this matrix will be diagonal (no nonzero entries above the diagonal) if and only if the matrix is normal (Horn and Johnson 1985).[5]

We interpret the Schur decomposition as follows. The strictly upper triangular part of the matrix (excluding the diagonal) corresponds to a strictly feedforward hierarchy

---

[5]The Schur Decomposition should not be confused with the Jordan normal form of a matrix. The Jordan normal form involves non-unitary transformations, and is diagonal for any matrix, non-normal or normal, with a complete basis of eigenvectors. It has nonzero entries above the diagonal only for matrices that are missing one or more eigenvectors. The Schur Decomposition involves only unitary transformations, and is diagonal only for normal matrices; it has nonzero entries above the diagonal for all non-normal matrices.

of connections: connectivity flows from node j to node i only for $j > i$. The diagonal entries correspond to recurrent connectivity: node i connects to itself with a strength corresponding to an eigenvalue. In the transformed orthonormal basis in which $\mathbf{M}$ is upper triangular, each node corresponds to an activity pattern. Thus, non-normal matrices, in addition to the recurrent connectivity represented by the eigenvalues, have a hidden feedforward connectivity pattern between activity patterns, which results in transient amplification not predicted by the eigenvalues. We will say more below about how to characterize this hidden feedforward connectivity.

For the generic case in which a matrix has a complete basis of eigenvectors, the Schur decompositions are found by transforming to an orthogonal basis obtained by Gram-Schmidt orthonormalization of the eigenvector basis. We illustrate the Schur Decomposition for the 2-dimensional case, starting with the solution in the eigenvector basis as in Eqs. 3.19 or 3.20:

$$\mathbf{r}(t) = e^{\beta_1 t}\mathbf{e}_1\mathbf{f}_1^\dagger\mathbf{r}(0) + e^{\beta_2 t}\mathbf{e}_2\mathbf{f}_2^\dagger\mathbf{r}(0) \tag{3.24}$$

Here the $\mathbf{f}_i$ are the left eigenvectors, and the right and left eigenvectors satisfy $\mathbf{f}_i \cdot \mathbf{e}_j = \delta_{ij}$. The orthogonal basis obtained by applying the Gram-Schmidt process to $\{\mathbf{e}_1, \mathbf{e}_2\}$ is $\{\mathbf{e}_1, \mathbf{q}\}$, where $\mathbf{q} = \frac{\mathbf{e}_2 - \mathbf{e}_1(\mathbf{e}_1 \cdot \mathbf{e}_2)}{\sqrt{1 - |\mathbf{e}_1 \cdot \mathbf{e}_2|^2}}$; one can easily verify, given $|\mathbf{e}_1| = |\mathbf{e}_2| = 1$, that $|\mathbf{q}| = 1$ and $\mathbf{e}_1 \cdot \mathbf{q} = 0$ We write Eq. 3.24 in terms of the orthonormal basis $\{\mathbf{e}_1, \mathbf{q}\}$. To do this, we note $\mathbf{e}_2 = \mathbf{q}\sqrt{1 - |\mathbf{e}_1 \cdot \mathbf{e}_2|^2} + \mathbf{e}_1(\mathbf{e}_1 \cdot \mathbf{e}_2)$; $\mathbf{f}_2 = \frac{\mathbf{q}}{\sqrt{1 - |\mathbf{e}_1 \cdot \mathbf{e}_2|^2}}$; and $\mathbf{f}_1 = \frac{\mathbf{e}_1 - \mathbf{q}(\mathbf{e}_1 \cdot \mathbf{e}_2)}{\sqrt{1 - |\mathbf{e}_1 \cdot \mathbf{e}_2|^2}}$. Substituting all of this into Eq. 3.24 and rearranging we get:

$$\mathbf{r}(t) = e^{\beta_1 t}\mathbf{e}_1(\mathbf{e}_1 \cdot \mathbf{r}(0)) + e^{\beta_2 t}\mathbf{q}(\mathbf{q} \cdot \mathbf{r}(0)) + \frac{(\mathbf{e}_1 \cdot \mathbf{e}_2)}{\sqrt{1 - |\mathbf{e}_1 \cdot \mathbf{e}_2|^2}}(e^{\beta_2 t} - e^{\beta_1 t})\mathbf{e}_1(\mathbf{q} \cdot \mathbf{r}(0)) \tag{3.25}$$

The first two terms describe the monotonic decay of two orthogonal modes, $\mathbf{e}_1$ and $\mathbf{q}$, at rates given by the two eigenvalues. The third term describes a transient, involving a mapping of fluctuations in the $\mathbf{q}$ direction in the input into the $\mathbf{e}_1$ direction in the output. The transient term is large when $|\mathbf{e}_1 \cdot \mathbf{e}_2|$ is close to one, i.e. when the angle between the eigenvectors is small; on the other hand, it becomes zero when the matrix is normal, so that $|\mathbf{e}_1 \cdot \mathbf{e}_2| = 0$.

These results can alternatively be stated as follows: in the $\{\mathbf{e}_1, \mathbf{q}\}$ basis, $\mathbf{M} = \begin{pmatrix} \beta_1 & \gamma \\ 0 & \beta_2 \end{pmatrix}$ where $\gamma = (\beta_2 - \beta_1)\frac{(\mathbf{e}_1 \cdot \mathbf{e}_2)}{\sqrt{1 - |\mathbf{e}_1 \cdot \mathbf{e}_2|^2}}$, and $\mathbf{r}(t) = e^{\mathbf{M}t}\mathbf{r}(0)$ where the matrix $e^{\mathbf{M}t} =$

$$\begin{pmatrix} e^{\beta_1 t} & \gamma \frac{e^{\beta_2 t} - e^{\beta_1 t}}{\beta_2 - \beta_1} \\ 0 & e^{\beta_2 t} \end{pmatrix}.$$ $\gamma = 0$ if and only if $\mathbf{M}$ is normal.[6] In particular, for $\mathbf{M} = \mathbf{W} - \mathbf{1}$ with

$\mathbf{W} = \begin{pmatrix} w_{EE} & -w_{IE} \\ w_{EI} & -w_{II} \end{pmatrix}$, the effective feedforward connection strength is $\gamma = w_{IE} + w_{EI}$,

and $\beta_1 = \beta_-$, $\beta_2 = \beta_+$ where $\beta_\pm = \frac{w_{EE} - w_{II} - 2}{2} \pm \frac{\sigma}{2}$ with $\sigma = \sqrt{(w_{EE} + w_{II})^2 - 4w_{EI}w_{IE}}$.
Thus, the effective feedforward connection strength in the 2×2 case is just given by the summed strength of the terms representing feedback inhibition.

In sum, the eigenvalue spectrum tells us about the asymptotic behavior of Eq. 1: when the real parts of all the eigenvalues of $\mathbf{M} = -(\mathbf{1} - \mathbf{W})$ are less than zero, the initial condition, $\mathbf{r}(0)$, decays to 0 as $t \to \infty$. It does not tell us what happens at finite times, when there may be large transients that are not predicted by the eigenvalues. Unfortunately there is no simple, straightforward property of the connectivity matrix that tells us about the size of the transient amplification. Below we will discuss several methods that provide some clues about interesting transient behavior and allow us to place upper and lower bounds on the size of the transient amplification.

Bounds on transient size

The solution to Eq. 1 can be written $\mathbf{r}(t) = e^{\mathbf{M}t}\mathbf{r}(0)$, where $\mathbf{M} = -(\mathbf{1} - \mathbf{W})$. The largest possible amplification at time t is given by:

$$\rho(t) \equiv \max_{\mathbf{r}(0)} \frac{|\mathbf{r}(t)|}{|\mathbf{r}(0)|} = \max_{\mathbf{r}(0)} \frac{|e^{\mathbf{M}t}\mathbf{r}(0)|}{|\mathbf{r}(0)|} \tag{3.26}$$

The norm $\|\mathbf{A}\|$ of a matrix $\mathbf{A}$ is defined as $\max_{\mathbf{v}} |\mathbf{A}\mathbf{v}|/|\mathbf{v}|$, and is equal to the maximum singular value of $\mathbf{A}$ (the square root of the largest eigenvalue of $\mathbf{A}\mathbf{A}^\dagger$ or of $\mathbf{A}^\dagger \mathbf{A}$), which we can write[7] as $\sigma_{\max}^{\mathbf{A}}$. Thus the largest possible amplification at time t is $\rho(t) = \|e^{\mathbf{M}t}\| = \sigma_{\max}^{e^{\mathbf{M}t}}$, and the largest overall amplification is $\rho_{\max} \equiv \max_t \rho(t)$. The eigenvalues of $e^{\mathbf{M}t}$ are $e^{\beta_i^{\mathbf{M}} t}$ where the $\beta_i^{\mathbf{M}}$ are the eigenvalues of $\mathbf{M}$, and $e^{\mathbf{M}t}$ is normal if and only if $\mathbf{M}$ is normal. For a normal matrix, the singular values are the absolute values of the eigenvalues, and so $\rho(t) = \sigma_{\max}^{e^{\mathbf{M}t}} = |e^{\beta_{\max}^{\mathbf{M}} t}|$ where $\beta_{\max}^{\mathbf{M}}$ is the eigenvalue of $\mathbf{M}$ with maximum real part.

---

[6]If the eigenvalues are equal, then γ can equal 0 even if $\mathbf{e}_1 \cdot \mathbf{e}_2 \neq 0$, but in this case the matrix is normal: when the eigenvalues are equal, any vectors in the space spanned by the two eigenvectors is also an eigenvector, so in particular $\mathbf{q}$ is also an eigenvector, so the matrix has a set of orthogonal eigenvectors and is normal.

[7]Mathematically one can assume many different norms for a vector, but we will always assume the Euclidean or L^2 norm $|v| = \sqrt{\sum_i v_i^2}$; the equivalence $\|\mathbf{A}\| = \sigma_{\max}^{\mathbf{A}}$ depends on this.

For a nonnormal matrix, the largest singular value is greater than or equal to the largest eigenvalue absolute value (Horn and Johnson 1985), so the largest amplification can be larger than expected from the eigenvalues: $\rho(t) \geq |e^{\beta_{\max}^{\mathbf{M}} t}|$.

Further bounds on $\rho(t)$ can be placed as follows (note, it is also easy to compute $\rho(t)$ numerically for a given matrix $\mathbf{M}$). If $\mathbf{M}$ has a complete basis of eigenvectors, then, letting $\mathbf{C}$ be the matrix whose columns are the eigenvectors, we have $\mathbf{M} = \mathbf{C}\mathbf{D}\mathbf{C}^{-1}$ where $\mathbf{D}$ is a diagonal matrix with the eigenvalues on the diagonal. Now $\|e^{\mathbf{M}t}\|$ can be written as $\|e^{\mathbf{C}\mathbf{D}\mathbf{C}^{-1}t}\| = \|\mathbf{C}e^{\mathbf{D}t}\mathbf{C}^{-1}\| \leq \|\mathbf{C}\|\|\mathbf{C}^{-1}\|\|e^{\mathbf{D}t}\|$. $\kappa_{\mathbf{C}} \equiv \|\mathbf{C}\|\|\mathbf{C}^{-1}\|$ is known as the condition number of $\mathbf{C}$; it is 1 if $\mathbf{M}$ is normal, but can be very large if $\mathbf{M}$ is non-normal. Futhermore, $\|e^{\mathbf{D}t}\| = |e^{\beta_{\max}^{\mathbf{M}} t}|$. Thus an upper bound on $\rho(t)$ is given by $\kappa_{\mathbf{C}}|e^{\beta_{\max}^{\mathbf{M}} t}|$. Combining this with the lower bound we have $|e^{\beta_{\max}^{\mathbf{M}} t}| \leq \rho(t) \leq \kappa_{\mathbf{C}}|e^{\beta_{\max}^{\mathbf{M}} t}|$. Unfortunately, these are typically not tight bounds (Trefethen and Embree 2005).

We can calculate another upper bound by noting that:

$$\frac{1}{\mathbf{r}}\frac{d|\mathbf{r}|}{dt} = \frac{\mathbf{r}^T \mathbf{M}^S \mathbf{r}}{|\mathbf{r}^2|} \tag{3.27}$$

where $\mathbf{M}^S$ is the symmetrized matrix $(\mathbf{M} + \mathbf{M}^\dagger)/2$. The maximum value of $\frac{1}{\mathbf{r}}\frac{d|\mathbf{r}|}{dt}$ is $\beta_{\max}^{\mathbf{M}^S}$, the largest eigenvalue of $\mathbf{M}^S$ (it takes this value when $\mathbf{r}$ is the corresponding eigenvector of $\mathbf{M}^S$). Thus, for a vector of unit length, $\beta_{\max}^{\mathbf{M}^S}$ is the largest possible instantaneous rate of change. If this is larger than $\beta_{\max}^{\mathbf{M}}$, this rate of change cannot be sustained - the vector direction as well as its magnitude will change in time, and any other direction has a smaller rate of change. This gives another upper bound, $\rho(t) \leq e^{\beta_{\max}^{\mathbf{M}^S} t}$, which again is typically not a tight bound (Trefethen and Embree 2005). For a normal matrix, the eigenvalues of $\mathbf{M}^S$ are just the real parts of the eigenvalues of $\mathbf{M}$, so $e^{\beta_{\max}^{\mathbf{M}^S} t} = |e^{\beta_{\max}^{\mathbf{M}} t}|$; but for a nonnormal matrix, the largest eigenvalue of $\mathbf{M}^S$ is greater than or equal to the largest real part of an eigenvalue of $\mathbf{M}$.

Tighter bounds, but ones that are less intuitive, can be obtained by the study of pseudospectra, a generalization of the theory of eigenvalue spectra (Trefethen and Embree 2005). The spectrum of $\mathbf{M}$, $\sigma(\mathbf{M})$, is the set of complex numbers z for which there is some unit vector $\mathbf{v}$ such that $\mathbf{M}\mathbf{v} - z\mathbf{v} = 0$. The ϵ-pseudospectrum, $\sigma_\epsilon(\mathbf{M})$, is the set of z for which there is some unit vector $\mathbf{v}$ such that $|\mathbf{M}\mathbf{v} - z\mathbf{v}| < \epsilon$. This can be more formally stated as $\|(z\mathbf{1} - \mathbf{M})^{-1}\| > 1/\epsilon$. For normal matrices, $\sigma_\epsilon(\mathbf{M})$ is just the set of points within ϵ of $\sigma(\mathbf{M})$, but for non-normal matrices, $\sigma_\epsilon(\mathbf{M})$ can wander far from $\sigma(\mathbf{M})$, and this is revealing of the behavior of $\mathbf{M}$. We mention just one result from the study of pseudospectra. Define

the *Kreiss constant* $\mathcal{K}(\mathbf{M}) \equiv \sup_{\epsilon>0} \mathrm{Re}(z \in \sigma_\epsilon(\mathbf{M}))/\epsilon$ where $\mathrm{Re}(x)$ is the real part of x. Then the maximum amplification $\rho_{\max}$ satisfies $\mathcal{K}(\mathbf{M}) \leq \rho_{\max} \leq eN\mathcal{K}(\mathbf{M})$, where N is the dimension of $\mathbf{M}$. This means that, if all eigenvalues have negative real part, but the pseudospectra for a given ϵ wanders into the right half complex plane by amounts much greater than ϵ, then there will be large transients.

Schur decomposition and hidden feedforward connectivity

We described above that the Schur decomposition reveals that non-normal matrices have both a recurrent component, represented by the eigenvalues, and a feedforward connectivity between activity patterns, represented by the nonzero entries above the diagonal in the Schur decomposition. A problem with the Schur decomposition is that it is not unique. For a non-normal matrix, each ordering of the non-orthogonal eigenvectors typically leads, under the Gram-Schmidt orthonormalization process, to a distinct orthogonal basis. Since there are $N!$ possible orderings of the eigenvectors, a non-normal matrix typically has $N!$ distinct Schur decompositions (not counting decompositions that differ only by a re-ordering of the orthonormal basis vectors). Thus, we cannot describe a unique feedforward structure between activity patterns that characterizes a given matrix.

However, we can uniquely characterize the overall strength of the feedforward connectivity of a matrix. All the different Schur decompositions of a matrix are related to one another by unitary transformations. The sum of the absolute squares of all of the elements of $\mathbf{M}$ is a unitary invariant (unchanged by unitary transformations of $\mathbf{M}$, and thus identical for all Schur decompositions of $\mathbf{M}$), and is equal to $\mathrm{Tr}\,\mathbf{MM}^\dagger$, where Tr is the trace, which in turn is equal to the sum of the squares of the singular values $\sigma_a^{\mathbf{M}}$ of $\mathbf{M}$. The eigenvalues of $\mathbf{M}$ are also unitary invariants, and so in particular the sum of the absolute squares of the eigenvalues of $\mathbf{M}$, $\sum_a |\beta_a^M|^2$, is a unitary invariant. But since all Schur decompositions have the eigenvalues on the diagonal, this is the sum of the absolute squares of the diagonal elements of any Schur decomposition of $\mathbf{M}$. Thus, the sum of the absolute squares of the off-diagonal or feedforward elements of any Schur decomposition of $\mathbf{M}$, as a proportion of the sum of the absolute squares of all of the elements, is $f^{\mathbf{M}} = \left(\mathrm{Tr}\left(\mathbf{MM}^\dagger\right) - \sum_a |\beta_a^M|^2\right)/\mathrm{Tr}\left(\mathbf{MM}^\dagger\right) = 1 - \frac{\sum_a |\beta_a^M|^2}{\sum_a (\sigma_a^{\mathbf{M}})^2}$. The size of $f^{\mathbf{M}}$ is a measure of the strength of hidden feedforward connectivity and thus of the strength of transient response and of the non-normality of the matrix.

3.5.6 Exact solution for the network in Figure 3.2

If the matrices $\mathbf{A} + \mathbf{B}$ and $\mathbf{A} - \mathbf{B}$ are normal, we can find an exact solution in an orthonormal basis for the network with connectivity matrix $\mathbf{W} = \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{A} & -\mathbf{B} \end{pmatrix}$.[8] As in the main text with slight additions, we let $\mathbf{e}_i^D$ be the eigenvectors of $\mathbf{A} - \mathbf{B}$ with eigenvalues λ_i^D, and $\mathbf{e}_i^S$ be the eigenvectors of $\mathbf{A} + \mathbf{B}$ with eigenvalues λ_i^S. Then $\mathbf{e}_i^{\mathbf{W}} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{e}_i^D \\ \mathbf{e}_i^D \end{pmatrix}$ is an eigenvector of $\mathbf{W}$ with eigenvalue λ_i^D. If $\mathbf{A}$ and $\mathbf{B}$ are $N \times N$, there are N such eigenvectors, and the other N eigenvectors of $\mathbf{W}$ all have eigenvalue 0 (because the top N and bottom N rows of $\mathbf{W}$ are identical). We define the difference modes $\mathbf{r}_i^- = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{e}_i^S \\ -\mathbf{e}_i^S \end{pmatrix}$ and the sum modes $\mathbf{r}_i^+ = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{e}_i^S \\ \mathbf{e}_i^S \end{pmatrix}$ and find that $\mathbf{W}\mathbf{r}_i^- = \lambda_i^S \mathbf{r}_i^+$, that is, patterns of the difference between excitation and inhibition are converted into patterns of the sum and amplified by λ_i^S, which can be large.

These sum patterns $\mathbf{r}_i^+$ can in turn be written as linear combination of the eigenvectors $\mathbf{e}_i^{\mathbf{W}}$. Both the $\mathbf{e}_i^D$ and the $\mathbf{e}_i^S$ are complete bases for the $N \times N$ vector space, so we can write $\mathbf{e}_i^S = \sum_j d_{ij} \mathbf{e}_j^D$ where d_{ij} is the component of $\mathbf{e}_i^S$ in the $\mathbf{e}_j^D$ direction. (Given our assumptions of normality, the $\mathbf{e}_i^D$ and the $\mathbf{e}_i^S$ are both orthonormal bases and so d_{ij} has a simple form, $d_{ij} = (\mathbf{e}_j^D)^\dagger \mathbf{e}_i^S$.) From this, we find $r_i^+ = \sum_j d_{ij} \mathbf{e}_j^{\mathbf{W}}$. We let $\mathbf{D}$ be the matrix with elements $D_{ij} = d_{ji} \lambda_j^S$, and let $\mathbf{L}^D$ be the diagonal matrix of the the λ_i^D. Then in the orthonormal basis $\{\mathbf{e}_1^{\mathbf{W}}, ..., \mathbf{e}_N^{\mathbf{W}}, \mathbf{r}_1^-, ..., \mathbf{r}_N^-\}$, the matrix $\mathbf{W}$ becomes $\begin{pmatrix} \mathbf{L}^D & \mathbf{D} \\ 0 & 0 \end{pmatrix}$. It is then not hard to compute $e^{-(\mathbf{1}-\mathbf{W})t/\tau}$. Let $\mathcal{L}^D$ be the diagonal matrix of $e^{\lambda_i^D t/\tau}$, and define $\mathbf{K}$ as the matrix with entries $K_{ij} = d_{ji} \lambda_j^S (e^{\lambda_i^D t/\tau} - 1)/\lambda_i^D$. Then $e^{-(\mathbf{1}-\mathbf{W})t/\tau} = e^{-t/\tau} \begin{pmatrix} \mathcal{L}^D & \mathbf{K} \\ 0 & \mathbf{1} \end{pmatrix}$.

Thus, differences between excitation and inhibition, $\mathbf{r}_i^-$, that involve spatial patterns $\mathbf{e}_i^S$ with large λ_i^S get amplified into large transients that are identical in excitation and

---

[8]If $\mathbf{A} + \mathbf{B}$ or $\mathbf{A} - \mathbf{B}$ are not normal this solution is still valid, but it is no longer in an orthogonal basis and will be misleading in the same way that the solution in the eigenvector basis is misleading when the eigenvectors are not orthogonal. The $\mathbf{A}$ and $\mathbf{B}$ matrices we used in Fig. 3.2 are slightly nonnormal, because the normalization of total excitatory and inhibitory weights onto each neuron (see Methods) results in small asymmetries. However, this non-normality is very small, as assessed by measures such as $f^{\mathbf{M}}$, so the vast majority of the non-normality of the overall matrix $\mathbf{W}$ is the result of the arrangment of the submatrices, not the non-normality of the submatrices themselves.

inhibition, with a strength along the spatial pattern $\mathbf{e}_j^D$ proportional to λ_i^S times the coefficient d_{ij} of $\mathbf{e}_i^S$ in the $\mathbf{e}_j^D$ direction, and with a transient time course $\left(e^{-(1-\lambda_j^D)t/\tau} - e^{-t/\tau}\right)/\lambda_j^D$. For $\lambda_j^D \to 0$, this time course is just $(t/\tau)e^{-t/\tau}$, while for $\Re(\lambda_j^D)$ large and negative and $t \gg \tau/(1 + |\Re(\lambda_j^D)|)$, it goes to $-e^{-t/\tau}/\lambda_j^D$. More generally, given ongoing noisy input, the contribution of this transient term to the amplitude of $\mathbf{e}_j^D$ in the output will be given by the time integral of the term, which is $\lambda_i^S d_{ij}\tau/(1 - \lambda_j^D)$. If $|\lambda_j^D| \ll 1$, then the size does not depend significantly depend on λ_j^D. In summary, only the λ_i^D influence the time course of the transient, while the λ_i^S and to some extent λ_i^D determine the size.

3.5.7 The general case of distinct $\mathbf{W}_{EE}$, $\mathbf{W}_{EI}$, $\mathbf{W}_{IE}$, and $\mathbf{W}_{II}$

In the general case in which $\mathbf{W} = \begin{pmatrix} \mathbf{W}_{EE} & -\mathbf{W}_{EI} \\ \mathbf{W}_{IE} & -\mathbf{W}_{II} \end{pmatrix}$, where each of the submatrices $\mathbf{W}_{XY}$ has non-negative entries, we cannot form a general solution or make a general argument as to the size of the transient amplification that will arise. However, we can make a number of more limited arguments to suggest that, when recurrent excitation is large but is balanced by large feedback inhibition, we should expect large transient amplification. As always, let $\mathbf{M} = -(\mathbf{1} - \mathbf{W})$. We think of $\mathbf{W}$ as the mean connectivity matrix in the linear model, which defines the probabilities from which the sparse random connectivity of the spiking model was drawn. However, some of our arguments would also apply to a sparse random connectivity matrix.

The first argument is that presented in the main text, which we slightly amplify here. As discussed above, the sum of the absolute squares of the matrix entries of $\mathbf{M}$ is a unitary invariant. Since both excitation and inhibition are strong, this sum is large. In the basis of a Schur decomposition, this is equal to the sum of the absolute squares of the eigenvalues plus the sum of the absolute squares of the effective feedforward connections. If we define balanced inhibition to mean that all of the eigenvalues are small, then it follows that there will be large effective feedforward connections and therefore large transient amplification. However, some connectivities that might otherwise be interpreted as "balanced inhibition" might produce eigenvalues with large negative real parts and/or large imaginary parts, and conceivably these eigenvalues could be large and/or numerous enough to account for most of the sum, leaving only relatively small feedforward connections; we cannot rule this out or state conditions under which it will or will not happen.

Second, consider the case in which $\mathbf{W}_{EE}$, $\mathbf{W}_{EI}$, $\mathbf{W}_{IE}$, and $\mathbf{W}_{II}$ can all be si-

multaneously diagonalized in an orthonormal basis. This is not unreasonable for the mean connectivity; for example, if each mean connectivity submatrix has entries that depend only on differences of spatial positions and/or preferred features of the neurons, then, except possibly for boundary effects, all can be simultaneously diagonalized by the Fourier transform, which represents a transformation to an orthonormal basis. In this case, we can show quite generally that, when recurrent excitation is large but is balanced by large feedback inhibition, there will be large amplification.

Let $\mathbf{e}_i$ be the orthonormal basis of the $N \times N$ subspace in which all of the submatrices are diagonal. Let $\mathbf{D}_{EE}$ be the diagonalized version of $\mathbf{W}_{EE}$ and $D_{EE}(k)$ its kth diagonal entry, and similarly for the other submatrices. Define orthonormal basis vectors of the full space $\mathbf{e}_i^U = \begin{pmatrix} \mathbf{e}_i \\ \mathbf{0} \end{pmatrix}$ and $\mathbf{e}_i^D = \begin{pmatrix} \mathbf{0} \\ \mathbf{e}_i \end{pmatrix}$, where $\mathbf{0}$ is the N-dimensional vector of all 0's, and work in the basis $\{\mathbf{e}_1^U, \ldots, \mathbf{e}_N^U, \mathbf{e}_1^D, \ldots, \mathbf{e}_N^D\}$. In this basis, the matrix $\mathbf{W}$ becomes $\begin{pmatrix} \mathbf{D}_{EE} & -\mathbf{D}_{EI} \\ \mathbf{D}_{IE} & -\mathbf{D}_{II} \end{pmatrix}$.

This means that the dynamics of Eq. 1 break up into independent two-dimensional subspaces of the form $\{\mathbf{e}_k^U, \mathbf{e}_k^D\}$ for a given k; subspaces with different k do not interact. The dynamics of the kth subspace is governed by the 2×2 matrix $\begin{pmatrix} D_{EE}(k) - 1 & -D_{EI}(k) \\ D_{IE}(k) & -(D_{II}(k) + 1) \end{pmatrix}$. In section 3.5.5, we computed the Schur decomposition for this 2×2 matrix, and showed that the feedforward connection strength is $\gamma = D_{EI}(k) + D_{IE}(k)$. That is, the size of the transient amplification in each subspace is just given by the sum of the two terms in that subspace that constitute the feedback inhibition. Since the feedback inhibition overall is large, it must be large in particular in some of the subspaces (*e.g.*, the sum of the absolute squares of the $D_{EI}(k)$'s is equal to the sum of the absolute squares of the elements of $\mathbf{W}_{EI}$, etc.), so there will be large transient amplification.

As a third argument, we compute the invariant $f^{\mathbf{M}}$ defined above, which measures the relative strength of the effective feedforward connectivity and thus of the transient amplification, in a special case: we assume that all of the eigenvalues of $\mathbf{M}$ are real. In this case, the sum of the absolute squares of the eigenvalues is given by $\text{Tr}\left(\mathbf{M}^2\right)$, so $f^{\mathbf{M}} = \text{Tr}\left(\mathbf{M}\mathbf{M}^\dagger - \mathbf{M}^2\right) / \text{Tr}\left(\mathbf{M}\mathbf{M}^\dagger\right)$. Let $\mathbf{W}_{EE}^A = (\mathbf{W}_{EE}^\dagger - \mathbf{W}_{EE})/2$ and $\mathbf{W}_{II}^A = (\mathbf{W}_{II}^\dagger - \mathbf{W}_{II})/2$

be the antisymmetric parts of $\mathbf{W}_{EE}$ and $\mathbf{W}_{II}$ respectively. Then we can compute

$$f^{\mathbf{M}} = \frac{\mathrm{Tr}\left(\mathbf{W}_{EI}\mathbf{W}_{EI}^{\dagger} + \mathbf{W}_{IE}\mathbf{W}_{IE}^{\dagger} + \mathbf{W}_{EI}\mathbf{W}_{IE} + \mathbf{W}_{IE}\mathbf{W}_{EI} + 2(\mathbf{W}_{EE} - \mathbf{1})\mathbf{W}_{EE}^{A} + 2(\mathbf{W}_{II} + \mathbf{1})\mathbf{W}_{II}^{A}\right)}{\mathrm{Tr}\left(\mathbf{W}_{EI}\mathbf{W}_{EI}^{\dagger} + \mathbf{W}_{IE}\mathbf{W}_{IE}^{\dagger} + \mathbf{W}_{EE}\mathbf{W}_{EE}^{\dagger} + \mathbf{W}_{II}\mathbf{W}_{II}^{\dagger} + 2(\mathbf{W}_{II} + \mathbf{W}_{II}^{A} - \mathbf{W}_{EE} - \mathbf{W}_{EE}^{A} + \mathbf{1}))\right)}$$

(3.28)

In particular, if all of the submatrices $\mathbf{W}_{XY}$ are symmetric, this becomes

$$f^{\mathbf{M}} = \frac{\mathrm{Tr}\left((\mathbf{W}_{EI} + \mathbf{W}_{IE})^{2}\right)}{\mathrm{Tr}\left(\mathbf{W}_{EI}^{2} + \mathbf{W}_{IE}^{2} + \mathbf{W}_{EE}^{2} + \mathbf{W}_{II}^{2} + 2(\mathbf{W}_{II} - \mathbf{W}_{EE} + \mathbf{1})\right)}$$

(3.29)

Thus, if (1) feedback inhibitory terms $\mathbf{W}_{IE}$ and $\mathbf{W}_{EI}$ are at least comparable in size to the recurrent terms $\mathbf{W}_{EE}$ and $\mathbf{W}_{II}$, as they must be for inhibition to balance excitation, and (2) the submatrices are large, as they must be if recurrent excitation and feedback inhibition are both strong, so that the linear and constant terms can be neglected relative to the quadratic terms, then the numerator of $f^{\mathbf{M}}$ should be comparable to the denominator and $f^{\mathbf{M}}$ should be significantly nonzero. This becomes particularly clear in the symmetric case, in which $f^{\mathbf{M}}$ becomes essentially a measure of the size of the feedback inhibition relative to the overall connectivity, similar to our finding in the case in which the different submatrices can be simultaneously diagonalized.

Finally, we make an argument based on the overall structure of $\mathbf{W}$, namely its two nonnegative submatrices and two nonpositive submatrices. We focus on $\mathbf{W}$, assuming that the $\mathbf{1}$ in $\mathbf{M} = \mathbf{W} - \mathbf{1}$ does not substantially change things. Let $\{\mathbf{e}_i\}$ be any set of N orthonormal N-dimensional basis vectors. Form the $2N$-dimensional orthonormal basis consisting of the vectors $\mathbf{e}_i^{+} = \frac{1}{\sqrt{2}}\begin{pmatrix}\mathbf{e}_i \\ \mathbf{e}_i\end{pmatrix}$ and $\mathbf{e}_i^{-} = \frac{1}{\sqrt{2}}\begin{pmatrix}\mathbf{e}_i \\ -\mathbf{e}_i\end{pmatrix}$. In any $2N$-dimensional orthonormal basis $\mathbf{f}_i$, $\mathbf{W}$ can be written $\mathbf{W} = \sum_{ij} W_{ij}\mathbf{f}_i\mathbf{f}_j^{\dagger}$ where $W_{ij} = \mathbf{f}_i\dagger\mathbf{W}\mathbf{f}_j$. So in particular, we can write

$$\mathbf{W} = \sum_{ij} W_{ij}^{++}\mathbf{e}_i^{+}\mathbf{e}_j^{+\dagger} + \sum_{ij} W_{ij}^{--}\mathbf{e}_i^{-}\mathbf{e}_j^{-\dagger} + \sum_{ij} W_{ij}^{+-}\mathbf{e}_i^{+}\mathbf{e}_j^{-\dagger} + \sum_{ij} W_{ij}^{-+}\mathbf{e}_i^{-}\mathbf{e}_j^{+\dagger} \qquad (3.30)$$

We define $\mathbf{W}^{++} = \sum_{ij} W_{ij}^{++}\mathbf{e}_i^{+}\mathbf{e}_j^{+\dagger}$ and similarly for the other three terms. Then it is easy to show that these matrices have the form $\mathbf{W}^{++} = \begin{pmatrix}\mathbf{A} & \mathbf{A} \\ \mathbf{A} & \mathbf{A}\end{pmatrix}$, $\mathbf{W}^{--} = \begin{pmatrix}\mathbf{B} & -\mathbf{B} \\ -\mathbf{B} & \mathbf{B}\end{pmatrix}$, $\mathbf{W}^{+-} = \begin{pmatrix}\mathbf{C} & -\mathbf{C} \\ \mathbf{C} & -\mathbf{C}\end{pmatrix}$, $\mathbf{W}^{-+} = \begin{pmatrix}\mathbf{D} & \mathbf{D} \\ -\mathbf{D} & -\mathbf{D}\end{pmatrix}$ for some submatrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$. From

the fact that $\mathbf{W} = \mathbf{W}^{++} + \mathbf{W}^{--} + \mathbf{W}^{+-} + \mathbf{W}^{-+}$, we can find

$$
\mathbf{A} = \frac{1}{4} \left(\mathbf{W}_{EE} - \mathbf{W}_{EI} + \mathbf{W}_{IE} - \mathbf{W}_{II} \right) \tag{3.31}
$$

$$
\mathbf{B} = \frac{1}{4} \left(\mathbf{W}_{EE} + \mathbf{W}_{EI} - \mathbf{W}_{IE} - \mathbf{W}_{II} \right) \tag{3.32}
$$

$$
\mathbf{C} = \frac{1}{4} \left(\mathbf{W}_{EE} + \mathbf{W}_{EI} + \mathbf{W}_{IE} + \mathbf{W}_{II} \right) \tag{3.33}
$$

$$
\mathbf{D} = \frac{1}{4} \left(\mathbf{W}_{EE} - \mathbf{W}_{EI} - \mathbf{W}_{IE} + \mathbf{W}_{II} \right) \tag{3.34}
$$

$\mathbf{C}$ is the average of the four nonnegative submatrices of $\mathbf{W}$, so it is nonnegative and it will have large entries if $\mathbf{W}$ does. $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$ all are averages of two of these submatrices and the negatives of two others, meaning that $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$ should be relatively small by some measure (for example, in the case of sparse random submatrices, then $\mathbf{C}$ would have leading eigenvalue of order N while $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$ would have leading eigenvalue of order $\sqrt{N}$ (Rajan and Abbott 2006)). Thus, the dominant contribution to $\mathbf{W}$ should be from $\mathbf{W}^{+-}$, which has the same structure of signs as $\mathbf{W}$. The contributions from $\mathbf{W}^{++}$, $\mathbf{W}^{--}$, and $\mathbf{W}^{-+}$ should be relatively small: these account for the differences between $\mathbf{W}_{EE}, \mathbf{W}_{EI}, \mathbf{W}_{IE}$ and $\mathbf{W}_{II}$, that is, for their deviations from their average, while $\mathbf{W}^{+-}$ accounts for their average. But $\mathbf{W}^{+-}$ is a sum of terms of the form $W_{ij}^{+-} \mathbf{e}_i^+ \mathbf{e}_j^{-\dagger}$, that is, it represents patterns of the difference between excitation and inhibition $\mathbf{e}_j^-$ being converted into patterns of the sum of excitation and inhibition, $\mathbf{e}_i^+$. Since these terms make the dominant contribution to $\mathbf{W}$ and $\mathbf{W}$ overall is large (involving large recurrent excitation balanced by large feedback inhibition), we expect $\mathbf{W}$ to involve large transient amplification in which small patterns of differences between excitation and inhibition are amplified into large patterns of the sum of excitation and inhibition.

None of these arguments are general or definitive, but all are consistent with the hypothesis that large transient amplification should be expected when large recurrent excitation is balanced by large feedback inhibition. It obviously remains an important open question to define more precisely when this will or will not be true.

3.5.8   Coexistence of Hebbian and transient amplification

We have focused on the case in which there are no positive eigenvalues, so that there is no Hebbian slowing and the only mechanism of amplification is transient amplification. However, it is important to point out that transient amplification and amplification by

slowing down will coexist if there are eigenvalues of $\mathbf{W}$, λ_i, with positive real part but in the stable regime, $0 < \Re(\lambda_i) < 1$.

This is illustrated by the solution given in section 3.5.5 for the Schur decomposition in the 2×2 case, with $\mathbf{M} = \mathbf{W} - \mathbf{1}$. We saw that, in the orthonormal $\{\mathbf{e}_1, \mathbf{q}\}$ basis, $\mathbf{r}(t) = e^{\mathbf{M}t}\mathbf{r}(0)$ where the matrix $e^{\mathbf{M}t} = \begin{pmatrix} e^{\beta_1 t} & \gamma \frac{e^{\beta_2 t} - e^{\beta_1 t}}{\beta_2 - \beta_1} \\ 0 & e^{\beta_2 t} \end{pmatrix}$ and $\beta_i = \lambda_i - 1$. If either eigenvalue satisfies $0 < \Re(\lambda_i) < 1$, then the corresponding diagonal term(s) and the transient term will both have a correspondingly slowed time course. That is, Hebbian slowing will amplify both the recurrent terms traditionally considered in Hebbian assembly models and the transient, feedforward term that arises from non-normality. Given ongoing noisy input, the contribution of each term to the output amplitude is given by its time integral. The diagonal terms thus contribute $\frac{1}{1-\lambda_i}$, while the off-diagonal term contributes $\frac{\gamma}{(1-\lambda_1)(1-\lambda_2)}$.

In the general $N \times N$ case, $e^{\mathbf{M}t}$ in the Schur basis will always have $e^{\beta_i t}$ on the diagonals, and the time-dependence of the off-diagonal terms will involve various linear combinations of the $e^{\beta_i t}$, such as the differences of exponentials in the 2×2 case. In general, the ij term can involve linear combinations of the $e^{\beta_x t}$ for $i \leq x \leq j$ (in particular, it will involve precisely those $x's$ for which there is a feedforward path from j to i that includes x). So, the same basic analysis will apply – if the kth eigenvalue satisfies $0 < \Re(\lambda_k) < 1$, Hebbian slowing will amplify both the kth recurrent term and any transient terms that contain $e^{\beta_k t}$. This can also be seen in the general solution for the matrix $\mathbf{W} = \begin{pmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{A} & -\mathbf{B} \end{pmatrix}$ given in section 3.5.6. There, for the ij transient term, the time course involves only $\lambda_i = \lambda_i^D$ and $\lambda_j = 0$, so the feedforward terms have the same time dependence as just discussed for the 2×2 case with $\lambda_2 = 0$.

3.6 Acknowledgements

Chapter 4

Conclusion

Understanding the functional purpose of the highly recurrent circuitry found through-out cerebral cortex is a formidable scientific problem, one that will occupy many people for a long time to come. In some sense, this circuitry is responsible for everything the cortex does, from processing sensory input to motor control to cognition and memory. This dissertation has primarily been about the mechanisms that cortical circuits might use to accomplish these diverse computational goals.

In chapter 2, we showed that multiplicative gain changes are a natural consequence of the way single cortical neurons with noisy membrane potentials combine their inputs. Gain changes are thought to be an important part of the mechanism by which the cortex computes coordinate transformations and also play in role in the neural mechanisms of attention.

More generally, in Chapter 3, we discuss transient amplification, which is an en-tirely new mechanism by which recurrent networks can shape neural activity. Transient amplification is fundamentally distinct from the Hebbian assembly mechanism currently understood to underlie amplification in these networks. It is difficult to distinguish between these two mechanisms by looking at the existing data on spontaneous activity in cat V1 and, in fact, they may both be present in cortex. The most conceptually simple experiment to test for transient amplification would be to transiently stimulate distinct patterns of excitatory and inhibitory neurons over an area of cortex and record the resulting activity. Our model predicts that large transient responses will occur when excitatory and inhibitory neurons receive opposite input patterns, and that the response of the network will cause excitatory and inhibitory activity to rise together. This type of experiment may be possible

in the near future using calcium sensitive dyes to record activity and genetically targeted optical stimulation to induce it. Hebbian assemblies were proposed with a specific functional purpose in mind - maintaining a pattern of neural activity after a stimulus has gone, as in working memory. Although we have shown that transient amplification is likely to be ubiquitous in cortex, perhaps the most interesting future work will be in determining its possible functional roles.

Bibliography

B. Ahmed, J. C. Anderson, R. J. Douglas, K. A. Martin, and J. C. Nelson. Polyneuronal innervation of spiny stellate neurons in cat visual cortex. *J Comp Neurol*, 341:39–49, Mar 1994. ISSN 0021-9967. doi: 10.1002/cne.903410105.

D. G. Albrecht and D. B. Hamilton. Striate cortex of monkey and cat: contrast response function. *J. Neurophysiol.*, 48:217–237, 1982.

Uri Alon. Network motifs: theory and experimental approaches. *Nat Rev Genet*, 8:450–461, Jun 2007. ISSN 1471-0056. doi: 10.1038/nrg2102.

R.A. Andersen and V.B. Mountcastle. The influence of the angle of gaze upon the excitability of the light- sensitive neurons of the posterior parietal cortex. *J. Neurosci.*, 3:532–48, 1983.

R.A. Andersen, G.K. Essick, and R.M. Siegel. Encoding of spatial location by posterior parietal neurons. *Science*, 230:456–8, 1985.

J. S. Anderson, M. Carandini, and D. Ferster. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J Neurophysiol*, 84:909–926, Aug 2000a. ISSN 0022-3077.

J. S. Anderson, I. Lampl, D. C. Gillespie, and D. Ferster. The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science*, 290:1968–1972, Dec 2000b. ISSN 0036-8075.

A. Arieli, A. Sterkin, A. Grinvald, and A. Aertsen. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science*, 273:1868–1871, Sep 1996. ISSN 0036-8075.

R. Azouz and C. M. Gray. Cellular mechanisms contributing to response variability of cortical neurons. *J. Neurosci.*, 19:2209–2223, 1999.

N. Barkai and S. Leibler. Circadian clocks limited by noise. *Nature*, 403:267–268, Jan 2000. ISSN 0028-0836. doi: 10.1038/35002258.

R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky. Theory of orientation tuning in visual cortex. *Proc Natl Acad Sci U S A*, 92:3844–3848, Apr 1995. ISSN 0027-8424.

L. S. Benardo. Separate activation of fast and slow inhibitory postsynaptic potentials in rat neocortex *in vitro*. *J. Physiol.*, 476:203–215, 1994.

O. Bernander, C. Koch, and R.J. Douglas. Amplification and linearization of distal synaptic input to cortical pyramidal cells. *J. Neurophysiol.*, 72:2743–53, 1994.

W. H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J Neurosci*, 17: 2112–2127, Mar 1997. ISSN 0270-6474.

D. Boussaoud, T.M. Barth, and S.P. Wise. Effects of gaze on apparent visual responses of frontal cortex neurons. *Exp. Brain Res.*, 93:423–34, 1993.

F. Bremmer, C. Distler, and K.P. Hoffmann. Eye position effects in monkey cortex. ii. pursuit- and fixation- related activity in posterior parietal areas lip and 7a. *J. Neurophysiol.*, 77:962–77, 1997a.

F. Bremmer, U.J. Ilg, A. Thiele, C. Distler, and K.P. Hoffmann. Eye position effects in monkey cortex. i. visual and pursuit-related activity in extrastriate areas mt and mst. *J. Neurophysiol.*, 77:944–61, 1997b.

N. Brunel. Dynamics of networks of randomly connected excitatory and inhibitory spiking neurons. *J Physiol Paris*, 94:445–463, Sep/Dec 2000. ISSN 0928-4257.

E. M. Callaway. Local circuits in primary visual cortex of the macaque monkey. *Annu Rev Neurosci*, 21:47–74, 1998. ISSN 0147-006X. doi: 10.1146/annurev.neuro.21.1.47.

G. Carmignoto and S. Vicini. Activity-dependent decrease in NMDA receptor responses during development of the visual cortex. *Science*, 258,:1007–1011, 1992.

S. Cash and R. Yuste. Input summation by cultured pyramidal neurons is linear and position- independent. *J. Neurosci.*, 18:10–5, 1998.

S. Cash and R. Yuste. Linear summation of excitatory inputs by ca1 pyramidal neurons. *Neuron*, 22:383–94, 1999.

J.R. Cavanaugh, W. Bair, and J.A. Movshon. Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. *J. Neurophysiol.*, 88:2530–46, 2002.

Y. Chagnac-Amitai and B. W. Connors. Horizontal spread of synchronized activity in neocortex and its control by gaba-mediated inhibition. *J Neurophysiol*, 61:747–758, Apr 1989. ISSN 0022-3077.

F.S Chance, L.A. Abbott, and A.D. Reyes. Gain modulation from background synaptic input. *Neuron*, 35:773–782, 2002.

L.A. Chase, R.J. Roon, L. Wellman, A.J. Beitz, and J.F. Koerner. L-quisqualic acid transport into hippocampal neurons by a cystine- sensitive carrier is required for the induction of quisqualate sensitization. *Neurosci.*, 106:287–301, 2001.

Z. Chu and J.J. Hablitz. Quisqualate induces an inward current via mglur activation in neocortical pyramidal neurons. *Brain Res.*, 879:88–92, 2000.

S. Chung and D. Ferster. Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. *Neuron*, 20:1177–1189, Jun 1998. ISSN 0896-6273.

B. W. Connors, R. C. Malenka, and L. R. Silva. Two inhibitory postsynaptic potentials, and $GABA_A$ and $GABA_B$ receptor-mediated responses in neocortex of rat and cat. *J. Physiol.*, 406:443–468, 1988.

M. C. Crair and R. C. Malenka. A critical period for long-term potentiation at thalamo-cortical synapses. *Nature*, 375:325–328, 1995.

G. C. DeAngelis, I. Ohzawa, and R. D. Freeman. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. i. general characteristics and postnatal development. *J Neurophysiol*, 69:1091–1117, Apr 1993. ISSN 0022-3077.

G. C. DeAngelis, G. M. Ghose, I. Ohzawa, and R. D. Freeman. Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. *J Neurosci*, 19:4046–4064, May 1999. ISSN 1529-2401.

A. Destexhe and D. Paré. Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *J Neurophysiol*, 81:1531–1547, Apr 1999. ISSN 0022-3077.

A. Destexhe, M. Rudolph, J.M. Fellous, and T.J. Sejnowski. Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons. *Neurosci.*, 107:13–24, 2001.

A. Destexhe, M. Rudolph, and D. Paré. The high-conductance state of neocortical neurons in vivo. *Nat Rev Neurosci*, 4:739–751, Sep 2003. ISSN 1471-003X. doi: 10.1038/nrn1198.

B. Doiron, A. Longtin, N. Berman, and L. Maler. Subtractive and divisive inhibition: effect of voltage-dependent inhibitory conductances and noise. *Neural Comput*, 13:227–48, 2001.

B. Ermentrout. Reduction of conductance based models with slow synapses to neural nets. *Neural Comput.*, 6:679–695, 1994.

D. Feldmeyer, V. Egger, J. Lubke, and B. Sakmann. Reliable synaptic connections between pairs of excitatory layer 4 neurones within a single 'barrel' of developing rat somatosensory cortex. *J Physiol*, 521 Pt 1:169–190, Nov 1999. ISSN 0022-3751.

Dirk Feldmeyer, Joachim Lübke, R. Angus Silver, and Bert Sakmann. Synaptic connections between layer 4 spiny neurone-layer 2/3 pyramidal cell pairs in juvenile rat barrel cortex: physiology and anatomy of interlaminar signalling within a cortical column. *J Physiol*, 538:803–822, Feb 2002. ISSN 0022-3751.

D. Ferster and K. D. Miller. Neural mechanisms of orientation selectivity in the visual cortex. *Annu Rev Neurosci*, 23:441–471, 2000. ISSN 0147-006X. doi: 10.1146/annurev.neuro.23.1.441.

D. Ferster, S. Chung, and H. Wheat. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature*, 380:249–252, Mar 1996. ISSN 0028-0836. doi: 10.1038/380249a0.

József Fiser, Chiayu Chiu, and Michael Weliky. Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature*, 431:573–578, Sep 2004. ISSN 1476-4687. doi: 10.1038/nature02907.

I. A. Fleidervish, A. M. Binshtok, and M. J. Gutnick. Functionally distinct nmda receptors mediate horizontal connectivity within layer 4 of mouse barrel cortex. *Neuron*, 21:1055–1065, Nov 1998. ISSN 0896-6273.

K. Fox and N. Daw. A model for the action of NMDA conductances in the visual cortex. *Neural Comput.*, 4:59–83, 1992.

K. Fox, H. Sato, and N. Daw. The effect of varying stimulus intensity on NMDA-receptor activity in cat visual cortex. *J. Neurophysiol.*, 64:1413–1428, 1990.

F. Gabbiani, H.G. Krapp, C. Koch, and G. Laurent. Multiplicative computation in a visual neuron sensitive to looming. *Nature*, 420:320–4, 2002.

C. Galletti and P.P. Battaglini. Gaze-dependent visual neurons in area v3a of monkey prestriate cortex. *J. Neurosci.*, 9:1112–25, 1989.

C. D. Gilbert and T. N. Wiesel. Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. *Nature*, 280:120–125, Jul 1979. ISSN 0028-0836.

C. D. Gilbert and T. N. Wiesel. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J Neurosci*, 9:2432–2442, Jul 1989. ISSN 0270-6474.

Joshua A. Goldberg, Uri Rokni, and Haim Sompolinsky. Patterns of ongoing activity and the functional architecture of the primary visual cortex. *Neuron*, 42:489–500, May 2004. ISSN 0896-6273.

Bilal Haider, Alvaro Duque, Andrea R. Hasenstaub, and David A. McCormick. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J Neurosci*, 26:4535–4545, Apr 2006. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.5297-05.2006.

D. Hansel and C. van Vreeswijk. How noise contributes to contrast invariance of orientation tuning in cat visual cortex. *J. Neurosci.*, 22:5118–5128, 2002.

DO Hebb. *The organization of behavior: A neurophysiological approach.* Wiley, 1949.

M.L. Hines and N.T. Carnevale. The neuron simulation environment. *Neural Comput*, 9: 1179–209, 1997.

J. A. Hirsch, J.-M. Alonso, R. C. Reid, and L.M. Martinez. Synaptic integration in striate cortical simple cells. *J. Neurosci.*, 18:9517–9528, 1998.

N. Hô and A. Destexhe. Synaptic background activity enhances the responsiveness of neocortical pyramidal neurons. *J. Neurophysiol.*, 84:1488–1496, 2000.

G. R. Holt and C. Koch. Shunting inhibition does not have a divisive effect on firing rates. *Neural Comput.*, 9:1001–1013, 1997.

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A*, 79:2554–2558, Apr 1982. ISSN 0027-8424.

R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1985.

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160:106–154, Jan 1962. ISSN 0022-3751.

B. Jagadeesh, H.S. Wheat, and D. Ferster. Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex. *Science*, 262: 1901–4, 1993.

B. Jagadeesh, H.S. Wheat, L.L. Kontsevich, C.W. Tyler, and D. Ferster. Direction selectivity of synaptic potentials in simple cells of the cat visual cortex. *J. Neurophysiol.*, 78: 2772–89, 1997.

C. E. Jahr and C. F. Stevens. Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. *J. Neurosci.*, 10:3178–3182, 1990.

Tal Kenet, Dmitri Bibitchkov, Misha Tsodyks, Amiram Grinvald, and Amos Arieli. Spontaneously emerging cortical representations of visual attributes. *Nature*, 425:954–956, Oct 2003. ISSN 1476-4687. doi: 10.1038/nature02078.

C. Koch. *Biophysics of Computation: Information Processing in Single Neurons.* Oxford Univ. Press, Oxford, 1998.

A. E. Krukowski and K. D. Miller. Thalamocortical NMDA conductances and intracortical inhibition can explain cortical temporal tuning. *Nature Neuroscience*, 4:424–430, 2001.

A. Lerchner, G. Sterner, J. Hertz, and M. Ahmadi. Mean field theory for a balanced hypercolumn model of orientation selectivity in primary visual cortex. *Network*, 17:131–150, 2006.

R. A. Lester, J. D. Clements, G. L. Westbrook, and C. E. Jahr. Channel kinetics determine the time course of NMDA receptor-mediated synaptic currents. *Nature*, 346:565–567, 1990.

S. LeVay and C. D. Gilbert. Laminar patterns of geniculocortical projection in the cat. *Brain Res*, 113:1–19, Aug 1976. ISSN 0006-8993.

Z. Li and P. Dayan. Computational differences between asymmetrical and symmetrical networks. *Network*, 10:59–77, Feb 1999. ISSN 0954-898X.

Luis M. Martinez, José-Manuel Alonso, R. Clay Reid, and Judith A. Hirsch. Laminar processing of stimulus orientation in cat visual cortex. *J Physiol*, 540:321–333, Apr 2002. ISSN 0022-3751.

Luis M. Martinez, Qingbo Wang, R. Clay Reid, Cinthi Pillai, José-Mañuel Alonso, Friedrich T. Sommer, and Judith A. Hirsch. Receptive field structure varies with layer in the primary visual cortex. *Nat Neurosci*, 8:372–379, Mar 2005. ISSN 1097-6256. doi: 10.1038/nn1404.

J. Martinez-Trujillo and S. Treue. Attentional modulation strength in cortical area mt depends on stimulus contrast. *Neuron*, 35:365–70, 2002.

C.J. McAdams and J.H. Maunsell. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *J. Neurosci.*, 19:431–41, 1999a.

C.J. McAdams and J.H. Maunsell. Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron*, 23:765–73, 1999b.

B. W. Mel. Synaptic integration in an excitable dendritic tree. *J. Neurophysiol.*, 70:1086–1101, 1993.

M. Migliore. On the integration of subthreshold inputs from perforant path and schaffer collaterals in hippocampal ca1 pyramidal neurons. *J. Comput. Neurosci.*, 14:185–92, 2003.

K. D. Miller and T. W. Troyer. Neural noise can explain expansive, power-law nonlinearities in neural response functions. *J. Neurophysiol.*, 87:653–659, 2002.

S.J. Mitchell and R.A. Silver. Shunting inhibition modulates neuronal gain during synaptic excitation. *Neuron*, 38:433–45, 2003.

H. Monyer, N. Burnashev, D. J. Laurie, B. Sakmann, and P. H. Seeburg. Developmental and regional expression in the rat brain and functional properties of four nmda receptors. *Neuron*, 12:529–540, Mar 1994. ISSN 0896-6273.

J.R. Muller, A.B. Metha, J. Krauskopf, and P. Lennie. Local signals from beyond the receptive fields of striate cortical neurons. *J. Neurophysiol.*, 90:822–31, 2003.

B.K. Murphy and K.D. Miller. Multiplicative Gain Changes Are Induced by Excitation or Inhibition Alone. *Journal of Neuroscience*, 23(31):10040–10051, 2003.

J.S. Nettleton and W.J. Spain. Linear to supralinear summation of ampa-mediated epsps in neocortical pyramidal neurons. *J. Neurophysiol.*, 83:3310–22, 2000.

M.G. Neubert and H. Caswell. Alternatives to Resilience for Measuring the Responses of Ecological Systems to Perturbations. *Ecology*, 78(3):653–665, 1997.

H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Submitted to* Nature, 2007.

L.A. Palmer and J.S. Nafziger. Effects of surround motion on receptive-field gain and structure in area 17 of the cat. *Vis. Neurosci.*, 19:335–53, 2002.

S.E. Palmer and K.D. Miller. The effects of inhibitory gain and conductance fluctuations in a simple model for contrast invariant orientation tuning in cat v1. In *2002 Abstract Viewer and Itinerary Planner*. Washington, DC: Society for Neuroscience, 2002., 2002.

D. Paré, E. Shink, H. Gaudreau, A. Destexhe, and E. J. Lang. Impact of spontaneous synaptic activity on the resting properties of cat neocortical pyramidal neurons in vivo. *J. Neurophysiol.*, 79:1450–1460, 1998.

A. Peters and B. R. Payne. Numerical relationships between geniculocortical afferents and pyramidal cell modules in cat primary visual cortex. *Cereb Cortex*, 3:69–78, Jan/Feb 1993. ISSN 1047-3211.

J.P. Pin and R. Duvoisin. The metabotropic glutamate receptors: structure and functions. *Neuropharmacol.*, 34:1–26, 1995.

P. Poirazi, T. Brannon, and B.W. Mel. Arithmetic of subthreshold synaptic summation in a model ca1 pyramidal cell. *Neuron*, 37:977–87, 2003a.

P. Poirazi, T. Brannon, and B.W. Mel. Pyramidal neuron as two-layer neural network. *Neuron*, 37:989–99, 2003b.

J.R. Polimeni, D. Granquist-Fraser, R.J. Wood, and E.L. Schwartz. Physical limits to spatial resolution of optical recording: Clarifying the spatial structure of cortical hypercolumns. *Proceedings of the National Academy of Sciences*, 102(11):4158–4163, 2005.

A. Pouget and T. J. Sejnowski. Spatial transformations in the parietal cortex using basis functions. *J. Cogn. Neurosci.*, 9:222–237, 1997.

A. Pouget and L.H. Snyder. Computational approaches to sensorimotor transformations. *Nat. Neurosci.*, 3 Suppl:1192–8, 2000.

S.A. Prescott and Y. De Koninck. Gain control of firing rate by shunting inhibition: roles of synaptic noise and dendritic saturation. *Proc. Nat. Acad. Sci. USA*, 100:2076–81, 2003.

D. Purves. *Neuroscience*. Sinauer Associates Inc, 2001. ISBN 0878937269.

K. Rajan and LF Abbott. Eigenvalue Spectra of Random Matrices for Neural Networks. *Physical Review Letters*, 97(18):188104, 2006.

R. C. Reid and J. M. Alonso. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378:281–284, Nov 1995. ISSN 0028-0836. doi: 10.1038/378281a0.

A. Reyes. Influence of dendritic conductances on the input-output properties of neurons. *Annu. Rev. Neurosci.*, 24:653–75, 2001.

J.H. Reynolds, T. Pasternak, and R. Desimone. Attention increases sensitivity of v4 neurons. *Neuron*, 26:703–14, 2000.

E. Salinas and L.F. Abbott. Coordinate transformations in the visual system: how to generate gain fields and what to compute with them. *Prog. Brain Res.*, 130:175–90, 2001.

E. Salinas and L.F. Abbott. Transfer of coded information from sensory to motor networks. *J. Neurosci.*, 15:6461–74, 1995.

E. Salinas and L.F. Abbott. A model of multiplicative neural responses in parietal cortex. *Proc. Nat. Acad. Sci. USA*, 93:11956–61, 1996.

E. Salinas and P. Thier. Gain modulation: a major computational principle of the central nervous system. *Neuron*, 27:15–21, 2000.

M. V. Sanchez-Vives, L. G. Nowak, and D. A. McCormick. Membrane mechanisms underlying contrast adaptation in cat area 17 *in vivo. J. Neurosci.*, 20:4267–4285, 2000.

P.C. Schwindt and W.E. Crill. Synaptically evoked dendritic action potentials in rat neocortical pyramidal neurons. *J. Neurophysiol.*, 79:2432–46, 1998.

G. Sclar, J. H. Maunsell, and P. Lennie. Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res.*, 30:1–10, 1990.

H. S. Seung. *The Handbook of Brain Theory and Neural Networks: Second Edition*, pages 94–97. MIT Press, 2003.

M. N. Shadlen and W. T. Newsome. Noise, neural codes and cortical organization. *Curr Opin Neurobiol*, 4:569–579, Aug 1994. ISSN 0959-4388.

M. N. Shadlen and W. T. Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci*, 18:3870–3896, May 1998. ISSN 0270-6474.

M. J. Shelley and L. Tao. Efficient and accurate time-stepping schemes for integrate-and-fire neuronal networks. *J Comput Neurosci*, 11:111–119, Sep/Oct 2001. ISSN 0929-5313.

O. Shriki, D. Hansel, and H. Sompolinsky. Rate models for conductance-based cortical neuronal networks. *Neural Comput.*, 15:1809–1841, 2003.

M.R. Smith, A.B. Nelson, and S. Du Lac. Regulation of firing response gain by calcium-dependent mechanisms in vestibular nucleus neurons. *J. Neurophysiol.*, 87:2031–42, 2002.

W. R. Softky and C. Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *J Neurosci*, 13:334–350, Jan 1993. ISSN 0270-6474.

D. C. Somers, S. B. Nelson, and M. Sur. An emergent model of orientation selectivity in cat visual cortical simple cells. *J Neurosci*, 15:5448–5465, Aug 1995. ISSN 0270-6474.

H. Sompolinsky and R. Shapley. New perspectives on the mechanisms for orientation selectivity. *Curr Opin Neurobiol*, 7:514–522, Aug 1997. ISSN 0959-4388.

H. Sompolinsky and O. L. White. Theory of large recurrent networks: From spikes to behavior. In C. Chow., B. Gutkin, D. Hansel, C. Meunier, and J. Dalibard, editors, *Methods and Models in Neurophysics, Volume Session LXXX Lecture Notes of the Les Houches Summer School 2003*, pages 267–340. Elsevier, 2005.

S. Squatrito and M.G. Maioli. Gaze field properties of eye position neurones in areas mst and 7a of the macaque monkey. *Vis. Neurosci.*, 13:385–98, 1996.

S. Squatrito and M.G. Maioli. Encoding of smooth pursuit direction and eye position by neurons of area mstd of macaque monkey. *J. Neurosci.*, 17:3847–60, 1997.

M.V. Srinivasan and G.D. Bernard. A proposed mechanism for multiplication of neural signals. *Biol. Cybern*, 21:227–36, 1976.

M. Steriade, I. Timofeev, and F. Grenier. Natural waking and sleep states: a view from inside neocortical neurons. *J Neurophysiol*, 85:1969–1985, May 2001. ISSN 0022-3077.

G. Stuart and N. Spruston. Determinants of voltage attenuation in neocortical pyramidal neuron dendrites. *J. Neurosci.*, 18:3501–10, 1998.

Gürol M. Süel, Jordi Garcia-Ojalvo, Louisa M. Liberman, and Michael B. Elowitz. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440: 545–550, Mar 2006. ISSN 1476-4687. doi: 10.1038/nature04588.

K. Tanaka. Cross-correlation analysis of geniculostriate neuronal relationships in cats. *J Neurophysiol*, 49:1303–1318, Jun 1983. ISSN 0022-3077.

V. Torre and T. Poggio. A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond. B*, 202:409–416, 1978.
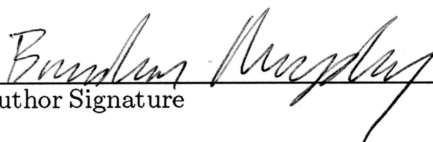
S. Townley, D. Carslake, O. Kellie-Smith, D. McCarthy, and D. Hodgson. Predicting transient amplification in perturbed ecological systems. *J. Appl. Ecol.*, May 2007. doi: 10.1111/j.1365-2664.2007.01333.x.

Lloyd N. Trefethen and Mark Embree. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators.* Princeton University Press, July 2005. ISBN 0691119465.

S. Treue and J.C. Martinez-Trujillo. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399:575–9, 1999.

Y. Trotter and S. Celebrini. Gaze direction controls response gain in primary visual-cortex neurons. *Nature*, 398:239–42, 1999.

T. W. Troyer and K. D. Miller. Physiological gain leads to high ISI variability in a simple model of a cortical regular spiking cell. *Neural Comput.*, 9:971–983, 1997.

T. W. Troyer, A. E. Krukowski, N. J. Priebe, and K. D. Miller. Contrast-invariant orientation tuning in cat visual cortex: Feedforward tuning and correlation-based intracortical connectivity. *J. Neurosci.*, 18:5908–5927, 1998.

M. V. Tsodyks, T. Kenet, A. Grinvald, and A. Arieli. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*, 286:1943–1946, 1999.

M. C. W. van Rossum, G. G. Turrigiano, and S. B. Nelson. Fast propagation of firing rates through layered networks of noisy neurons. *J Neurosci*, 22:1956–1966, Mar 2002. ISSN 1529-2401.

C. van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274:1724–1726, Dec 1996. ISSN 0036-8075.

J. L. Vincent, G. H. Patel, M. D. Fox, A. Z. Snyder, J. T. Baker, D. C. Van Essen, J. M. Zempel, L. H. Snyder, M. Corbetta, and M. E. Raichle. Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, 447:83–86, May 2007. ISSN 1476-4687. doi: 10.1038/nature05758.

T. P. Vogels and L. F. Abbott. Signal propagation and logic gating in networks of integrate-and-fire neurons. *J Neurosci*, 25:10786–10795, Nov 2005. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.3508-05.2005.

X. J. Wang. Synaptic basis of cortical persistent activity: the importance of nmda receptors to working memory. *J Neurosci*, 19:9587–9603, Nov 1999. ISSN 1529-2401.

J. Waters and F. Helmchen. Background synaptic activity is sparse in neocortex. *J Neurosci*, 26:8267–8277, Aug 2006. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.2152-06.2006.

D.S. Wei, Y.A. Mei, A. Bagal, J.P. Kao, S.M. Thompson, and C.M. Tang. Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons. *Science*, 293:2272–5, 2001.

H.R. Wilson and J.D. Cowan. Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. *Biophysical Journal*, 12(1):1, 1972.

J. Wolfe and L. A. Palmer. Temporal diversity in the lateral geniculate nucleus of cat. *Vis Neurosci*, 15:653–675, Jul/Aug 1998. ISSN 0952-5238.

D. Zipser and R. A. Andersen. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331:679–684, 1988.

UCSF Library Release

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Brandon Murphy _____   12/14/2007 _____
Author Signature   Date