On computing shape: a study of the neural processes concerning naturalistic boundary conformation within the ventral visual pathway

Timothy D. Oleskiw

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

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Eric Shea-Brown, Chair

Anitha Pasupathy

Wyeth Bair

J. Nathan Kutz

Program Authorized to Offer Degree:
Applied Mathematics

University of Washington

Abstract

On computing shape: a study of the neural processes concerning naturalistic boundary conformation within the ventral visual pathway

Timothy D. Oleskiw

Chair of the Supervisory Committee:
Associate Professor Eric Shea-Brown
Applied Mathematics

The perception of shape is a remarkable computation, solved rapidly by the brain to extract boundary features within natural scenes while being robust against many visual obstacles. Interestingly, while observers typically have an intuitive understanding of what shape is, it is in fact exceedingly difficult to mathematically describe shape in a natural context. In this dissertation I use computational modeling, primate electrophysiology, and mathematical analysis to study the neural processes underlying computations of shape within the ventral visual pathway.

I first present results from an investigation of the spectral receptive field (SRF) model, a Fourier-based method proposed to explain selectivity for boundary conformation of neurons in cortical area V4. Noting that spectral power coefficients are phase-invariant, I analyze the statistical properties of responses evoked by synthetic shape stimuli to demonstrate SRFs as incapable of capturing shape tuning seen in single-unit V4 responses.

Computational studies have shown that naturalistic images may be reconstructed from features common to physical scenes, namely contrast boundaries and blur, i.e. the spatial gradient of contrast across each boundary. I next present results from an electrophysiology study of primate V4 using stimuli in which object shape and boundary blur are systematically varied, reporting a population of V4 neurons that are tuned for intermediate blur magnitudes.

Importantly, I propose simple joint model in which blur multiplicatively scales shape-selective responses, capturing a range of observed behaviors and revealing distinct neural dynamics.

As we progress in understanding the computations of shape in neural networks, it is widely believed that the divisive inhibition of a neuron's firing rate is a fundamental operation, critical to computations of gain control and response invariance within the ventral pathway. However, the precise mechanisms underlying divisive inhibition are not completely understood, making it difficult to predict when this phenomenon will occur in spiking neural networks. Research has shown that balanced excitatory and inhibitory activity can divisively scale a neuron's spiking output, both *in vitro* and in conductance-based leaky-integrate-and-fire (LIF) model simulations. Interestingly, a review of these simulations suggests that gain modulation depends on another factor: the autocorrelation structure (timescale) of driving input. With numeric simulation and stochastic analysis I then show divisive inhibition to arise from an interaction between the neuron's input and membrane dynamics, developing an analytic approximation for the firing rate of a nonlinear stochastic LIF model wherein input timescale is explicitly parameterized. Furthermore, I demonstrate that by increasing synaptic timescales across a biophysically-plausible range of parameters, a neuron's response is divided by a novel circuit-level mechanism.

Background material for this interdisciplinary work is provided to clarify its relevance to neuroscience and mathematics communities.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

APC: Angular position curvature model (see Pasupathy and Connor, 2001).

APCB: Angular position, curvature, and blur model.

COM: Center-of-Mass, typically referring to the median value of integrating a neuron's tuning curve.

IT: Inferotemporal cortex.

LIF: Leaky integrate-and-fire, with regard to neuron models with membrane dynamics and spiking behavior.

NRMSE: Normalized root-mean-squared error.

PSTH: Peristimulus time histogram.

PC: Principal component.

PV: Principal value.

RF: Receptive field of a neuron, *i.e.* the portion of visual space to which a neuron will respond.

SRF: Spectral receptive fields, a Fourier-based method proposed by David et al. (2006) to explain tuning of V4 neurons in general and boundary conformation in particular.

V1: Striate cortex, the primary visual cortex.

V2: Prestriate visual cortex, the secondary visual area.

ACKNOWLEDGMENTS

I would like to acknowledge the support of my advisors, who gave me the unique opportunity to study at the intersection of multiple departments and disciplines. To Eric, Wyeth, and Anitha: I am grateful for your faith, encouragement, and mentorship.

To my friends and colleagues of the Washington National Primate Research Center and University of Washington's departments of Biological Structure and Applied Mathematics: thank you for the memorable years.

> *[to the player]* "Look at you, hacker: a pathetic creature of meat and bone, panting and sweating as you run through my corridors. How can you challenge a perfect, immortal machine?"
>
> — SHODAN, *System Shock*

DEDICATION

To Shannon

Chapter 1

MOTIVATION

When we view an image like Fig. 1.1, we immediately perceive the figure to be a person, and we arrange the position and orientation of the limbs to identify pose. What is striking about our experience of its perception is that the image entering our eyes is merely a collection of geometric features that comprise the object's shape. It has been known for some time that when confronting natural images, the human visual system uses shape information to quickly parse the image, segmenting it into various relevant objects (Attneave, 1954). Interestingly, these processes are extremely fast and robust against different visual obstacles, such as lighting, shading, and occlusion (Elder and Velisavljević, 2009).

What is truly fascinating about the problem of shape, that is to say, the computations underlying perception and representation boundary contours in naturalistic images, is that while observers typically have an intuitive understanding for what shape is, it is exceedingly difficult to mathematically define shape in a natural setting, or even to decompose a shape and enumerate its constituent parts (Oleskiw et al., 2010; Elder et al., 2013). As will be demonstrated, the problem of shape cannot be fully understood unless it is approached from all sides, namely through visual psychophysics, electrophysiology, computational modeling, and mathematical analysis. In this dissertation I study the neural processes concerning the perception of object boundaries in natural scenes, bringing together these different fields to gain insight into the computations of shape.

1.1 Background

For over a century now the Gestalt school of psychology has sought to describe the visual cues that facilitate the perceptual organization of a scene (Wagemans et al., 2012). Cues such

Figure 1.1: Matisse, H. (1952) *Blue Nude I*

as spatial proximity and similarity of color, size, or orientation, have produced theories of visual object perception that describe such phenomena as rising above the sum of elementary feature sensations. More recently have these principles been extended to model notions of object shape, such as the continuity of boundary contours, symmetry, and closure (Wagemans et al., 2012). A weakness of the descriptive models of phenomena surrounding the perception of object shape and boundary conformation is that they are merely qualitative and cannot provide a quantitative description of the underlying computations. While some quantitative models have been proposed to explain aspects of Gestalt phenomena, *i.e.* good continuation via association fields (Field et al., 1993), other global properties, such as closure, remain largely unexplained (Oleskiw, 2010). Given the above, I seek to study how shape is processed by biological vision systems in an attempt to reverse-engineer the computations responsible.

The ultimate goal of this research will be to further understanding of both how the primate brain has solved the problem of computing boundaries in natural scenes and how to replicate these computations toward applications of machine vision.

Toward this goal a focus of my investigation is cortical area V4, an intermediate visual area of the ventral pathway (Livingstone and Hubel, 1987). Here, much work has implicated V4 in the processing of visual shape (Kobatake and Tanaka, 1994), specifically through single-unit electrophysiology demonstrating neurons to respond selectively to features of boundary conformation, *i.e.* curvature (Pasupathy and Connor, 2001). Typically, such experiments rely on visual stimulation paradigms utilizing a limited and synthetic parameterized stimuli set to acquire data fit by simple descriptive models of selectivity. While useful in describing feature selectivity of single neurons, studies of this type do not result in image-computable models of activity, making inference of neuronal behavior with respect to boundary conformation in a naturalistic context difficult.

An alternative approach to physiology of V4 has employed rapid presentation of a multitude of naturalistic image frames to which image-computable neural networks are learned. A benefit of this paradigm, resulting from the sheer volume of stimuli presented, is that machine learning techniques can be employed to train networks with hundreds or thousands (even millions) of parameters. One such attempt to explain V4 selectivity is the spectral receptive field (SRF) model proposed by David et al. (2006). Here, spectral power coefficients, computed from the phase-invariant Fourier transform of stimuli, are fit to single-unit data in response to thousands of presented natural image patches. Such models are attractive, in part, due to their parsimonious explanation of V4 computation in relation to previous visual areas; spectral power coefficients are conceptually similar to phase-invariant complex cells that are tuned for spatial frequency and orientation within primary visual cortex V1. Therefore, an SRF-like model need only compute a linear combination of upstream V1-like spectral power coefficients to predict a V4 response.

Although SRF models are capable of capturing a significant amount of variance evoked by V4 responses to naturalistic stimuli, to what extend do SRF models resemble the computa-

tions underlying V4 tuning for boundary conformation? To study this question I hypothesize that by analyzing SRF models trained to a dataset of V4 neurons in response to a parametric set of shape stimuli, it may be determined if in fact such simple phase-invariant models like SRFs are capable of describing shape tuning within V4.

More recent models of visual computation have sought to loosely replicate the anatomy and physiology of the ventral pathway (Serre et al., 2005). For example, the image-computable V4 model of Cadieu et al. (2007) utilizes a bank of phase-dependant V1-like simple cells which are pooled and rectified. This application of feature selectivity, followed by rectifying non-linearities for invariance, has been argued as a canonical computation within the brain and the ventral visual pathway in particular (Carandini and Heeger, 2011). Specifically, as a stimulus is passing through and is processed by visual areas from V1 through inferotemporal cortex (IT), a cascade of selectivity and invariance operations are performed; within each layer, selectivity filters extracts useful feature information from the previous layer, while invariance generalizes that information, discarding feature dimensions that are less important to be passed to the layer above. These feed-forward neural networks, such as the HMAX model proposed by Riesenhuber and Poggio (1999), are strikingly successful in solving complex machine vision tasks, such as object recognition and scene categorization (Serre et al., 2005).

The extent to which computations within trained neural network layers reflect those observed in biological vision systems, however, remains to be fully understood. For example, while neurons in V4 are selective for boundary conformation, they are invariant to other stimulus dimensions such as color, luminance, position, and size (Bushnell et al., 2011a). Thus, answers to these questions, such as how position- or size-tolerant shape-selective responses could be computed in purely feed-forward networks, will constrain and inspire future experimentation and modeling efforts. One such experiment, to address computations of shape within V4, would fit image-computable feed-forward networks to data recorded from the presentation of naturalistic shape stimuli.

While presentation of natural images would allow for the volume of data necessary to fit

such models to be efficiently collected, we do not yet posses the tools necessary to rigorously define shape in naturalistic context (Oleskiw, 2010; Oleskiw et al., 2010; Elder et al., 2013). Further, while it is possible to parametrically define subsets of shape stimuli, these images are synthetic and do not possess the rich statistical structure of natural scenes (Pasupathy and Connor, 1999; Gallant et al., 1996). One possible avenue of reconciliation is to construct a stimulus set in which object shape is parameterized and well defined, but also possesses image features common to natural images (Elder, 1999; Rensink and Cavanagh, 2004). To bridge the gap between synthetic and naturalistic stimulus presentation paradigms, I hypothesize that by conducting a targeted experiment of V4 selectivity using a set of parametric stimuli in which object shape and naturalistic image features are systematically varied, one may further understanding of how shape is processed and represented by individual neurons.

Furthermore, when considering computations of feature selectivity in V4, and within the ventral pathway in general, it is important to also consider the biophysical limitations of these computations (Yu et al., 2002). One such constraint is that the output of a neuron, measured as the number of action potentials (spikes) per unit of time in response to some fixed input, has a limited dynamic range (Barlow, 1972). For example, consider the the firing rate r_i of the i^{th} neuron in some population, where the output firing rate of each neuron is limited to between 0-150 spikes per second. One way to account for a limited range is through normalization, where we imagine pooling the incoming activity with summation, and using division to scale down or modulate the gain of those responses, *i.e.*,

$$r_i^* = \frac{r_i}{\sigma + \sum_n r_n}, \tag{1.1}$$

where σ is a small constant to prevent division by zero in the absence of input (Heeger, 1992). Here, the normalized response r_i^* captures *equalized* across the population, so as to make full use of each neuron's limited range of available output activity.

Although not immediately clear, response normalization, in addition to selectivity, is also implicated in computations of invariance. To see this, note that invariance, or sustaining a response r' across an irrelevant stimulus dimension, can be implemented with the maximum

operator, *i.e.*,

$$r' = \max_i (r_i)\,. \tag{1.2}$$

While max is a nonlinear operation, it does have a convenient approximation (Yu et al., 2002). Using the formal definition of max as the infinity norm $\|\cdot\|_\infty$, one can write

$$\max_i (r_i) = \|(r_1, r_2, \ldots, r_N)\|_\infty = \lim_{p\to\infty} \left(\sum_i r_i^p \right)^{\frac{1}{p}}. \tag{1.3}$$

Then, taking an approximation for the rational exponent under finite p, (1.3) admits

$$r' \approx \sum_i \frac{r_i^{p+1}}{\sigma + \sum_n r_n^p}. \tag{1.4}$$

A key observation from (1.1) and (1.4) is that both require divisive inhibition, *i.e.* the scaling of a neuron's output response that is inversely proportional to some level of pooled population activity. While pooling of firing rates is simple to achieve through the accumulation (summation) of excitatory input, it is not immediately clear how to modulate the gain of a neuron's firing rate (multiplication) in response to some fixed level of input.

Despite much research into this topic, a clear description of the biophysically-plausible mechanisms for divisive inhibition of a neuron's output response in spiking circuits has yet to be fully developed (Holt and Koch, 1997; Doiron et al., 2001; Mitchell and Silver, 2003). Complicating study of these computations are the complexity of the underlying neural components; the nonlinear nature of spike-generating neural models are difficult to compute analytically (Brunel and Sergi, 1998), further confounded by the dimensionality of dynamical systems required to capture biophysical properties of membrane dynamics and neurotransmitter kinetics (Richardson, 2004).

Interestingly, physiological study and numeric simulations have implicated statistical properties of a neuron's input, *i.e.* variance, as a mechanism for divisive inhibition in biophysically realistic spiking neurons (Chance et al., 2002; Ayaz and Chance, 2009; O'Donnell and van Rossum, 2014). Furthermore, analytic studies have suggested that the autocorrelation structure, *i.e.* timescale, of a neuron's input may affect its output response (Brunel and

Latham, 2003; Brunel and Sergi, 1998; Brunel et al., 2001). Could timescale achieve divisive inhibition in spiking neural circuits? I hypothesize that by analyzing biophysically-realistic stochastic neuron models that capture the autocorrelation structure of input and membrane dynamics, I may identify and describe the regimes in which input timescale affect divisive inhibition in neural circuits.

1.2 Document structure

Chapter 1 motivates the study computations surrounding the perception of shape, and through a discussion of background material, presents three open questions of the neural processes concerning boundary conformation worthy of scientific inquiry. In the following chapters I answer these questions in sequence.

First, I describe in Chapter 2 results from a computational investigation into spectral receptive field models of V4 activity to determine if, and to what extent, shape selectivity can be explained in the context of phase-invariant spectral receptive fields. Through analysis and model fitting I conclude that SRFs in fact cannot account for the tuning boundary conformation exhibited by single-unit V4 responses. Secondly, having described computations which do *not* occur within the ventral pathway, I present in Chapter 3 results from an electrophysiology experiment of nonhuman primate, reporting selectivity in the the rate response of V4 neurons for the novel stimulus feature of boundary blur. Importantly, here I propose a joint code which captures selectivity for both shape and blur in a subset of neurons and argue for their utility as a neural basis for the sufficient representation of natural scenes and phenomena of perceptual organization. Finally, I address in Chapter 4 neural mechanisms governing gain modulation in spiking neurons. Through simulation I demonstrate divisive inhibition of a neuron's output rate response via the autocorrelation timescale of driving input, and provide a mathematical analysis to justify my findings in both single neurons and spiking circuits. We conclude this study with some remarks regarding the broad influence of my findings and potential topics of future research.

Chapter 2

SPECTRAL RECEPTIVE FIELDS DO NOT ACCOUNT FOR BOUNDARY CONFORMATION SELECTIVITY IN AREA V4

The mid-level visual cortical area V4 in the primate is thought to be critical for the neural representation of visual shape. Several studies agree that V4 neurons respond to contour features, *e.g.*, convexities and concavities along a shape boundary, that are more complex than the oriented segments encoded by neurons in the primary visual cortex. Here we compare two distinct approaches to modeling V4 shape selectivity: one based on a spectral receptive field (SRF) map in the orientation and spatial frequency domain and the other based on a map in an object-centered angular-position and contour curvature space. We test the ability of these two characterizations to account for the responses of V4 neurons to a set of parametrically designed two-dimensional shapes recorded previously in the awake macaque. We report two lines of evidence suggesting that the SRF model does not capture the contour sensitivity of V4 neurons. First, the SRF model discards spatial phase information, which is inconsistent with the neuronal data. Second, the amount of variance explained by the SRF model was significantly less than that explained by the contour curvature model. Notably, cells best fit by the curvature model were poorly fit by the SRF model, the latter being appropriate for a subset of V4 neurons that appear to be orientation tuned. These limitations of the SRF model suggest that a full understanding of mid-level shape representation requires more complicated models that preserve phase information and perhaps deal with object segmentation.

2.1 Introduction to spectral receptive field modeling

Visual object perception and recognition in primates is based on sensory information processing within the ventral visual pathway (Mishkin and Ungerleider, 1982; Felleman and Van Essen, 1991). Over the last half century, studies of the primary visual cortex (V1) have identified local orientation and spatial frequency as the basis dimensions of form representation at the early stages in the ventral pathway (Hubel and Wiesel, 1965, 1959, 1968; Campbell and Robson, 1968; Schiller et al., 1976; Movshon et al., 1978; De Valois and De Valois, 1990). At intermediate stages, in particular area V4, the representation has yet to be firmly established. Neurons in V4 have been shown to be selective for bars of different length, for radial or concentric gratings, for moderately complex shapes and specifically for the curvature of segments of the bounding contour of shapes (Desimone and Schein, 1987; Gallant et al., 1993; Kobatake and Tanaka, 1994; Pasupathy and Connor, 2001; Hegdé and Van Essen, 2007; Nandy et al., 2013). No single model is widely accepted to account for these observations, but a common approach to explaining extrastriate responses in both the dorsal and ventral pathways is to model them in terms of selectivity for simple combinations of the features that are represented at earlier levels. This amounts to using weighted combinations of V1-like channels to fit the observed data (David et al., 2006; Rust et al., 2006; Cadieu et al., 2007; Willmore et al., 2010; Vinch, 2013). Here, we examine whether an instance of this approach, known as the spectral receptive field (SRF) model (David et al., 2006), can account for complex curvature selectivity observed in V4 neurons.

The SRF model describes the tuning of V4 neurons in terms of a weighting function across orientation and spatial frequency bands in the power spectrum of the stimulus (David et al., 2006). This model has the elegant simplicity of combining V1-like signals in a manner that discards phase and thereby produces translation invariance, a key feature of V4 responses (Rust and Dicarlo, 2010; Pasupathy and Connor, 1999, 2001; Gallant et al., 1996). It has also been argued (David et al., 2006) that the SRF model can account for the ability of V4 neurons to respond to complex shapes in terms of contour features at a particular location within an

object-centered reference frame (Pasupathy and Connor, 2001). For example, some neurons may respond strongly to shapes with a sharp convexity to the upper right while others may respond to shapes with a concavity to the left. These patterns of selectivity are well modeled by 2D Gaussian tuning functions in a space defined by *i*) the curvature of the boundary and *ii*) angular positions relative to object center (Pasupathy and Connor, 2001). They are also well modeled by a hierarchical contour template model (Cadieu et al., 2007). Using the previously recorded data set on which both of these models were based, we examine whether the SRF model, the simplest of the three, can account for the contour selectivity observed in V4. We find that there are important features of the data that are not captured by the SRF model.

2.2 Materials and methods

2.2.1 Experimental procedures

All animal procedures for this study, including implants, surgeries and behavioral training, conformed to NIH and USDA guidelines and were performed under an institutionally approved protocol. The data analyzed here are derived from a previous study (Pasupathy and Connor, 2001) and consist of the responses of 109 single, well-isolated V4 neurons in two rhesus monkeys (Macaca mulatta) that were recorded while the animals fixated a 0.1° white spot on a computer monitor. After preliminary characterization of the receptive field (RF) location and preferred color of each cell, shape tuning was characterized with a set of 366 stimuli (Fig. 2.1). Each stimulus was presented in random order without replacement five times for most cells (91/109; nine cells had 4 repetitions and nine had 3 repetitions). Response rates were calculated by counting spike occurrences during the 500 ms stimulus presentation period. Spontaneous rates, calculated based on blank stimulus periods interspersed randomly during stimulus presentation, were subtracted from the average response rate for each stimulus.

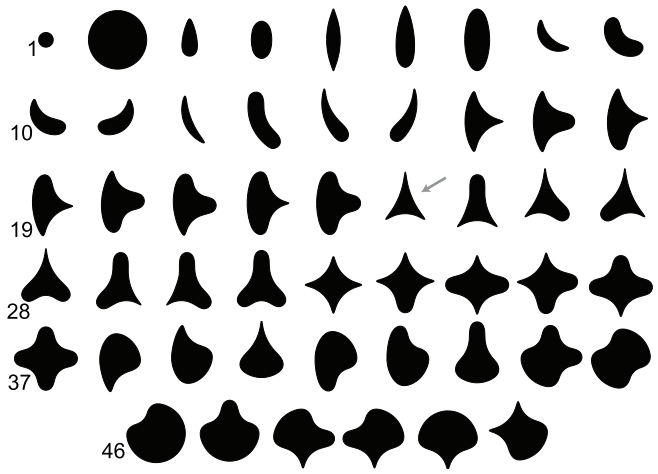

Figure 2.1: **Set of 51 shapes used to characterize V4 neurons.** Each shape was presented at each of eight orientations, or fewer for shapes with rotational symmetry. For example, the circles (top row, left side) were shown in just one conformation because all rotations are identical. Shapes are numbered left to right, top to bottom, starting with #1 in the upper left corner. Gray arrow marks shape #24, referred to in Results.

2.2.2 Stimulus design and representation

Stimulus design is described in detail in Pasupathy and Connor (2001). Briefly, stimuli were constructed by systematic combination of 4-8 contour segments each of which took one of five curvature values, resulting in 51 shapes (Fig. 2.1). To create radial variation, each shape is rotated by 8 increments of 45, discarding duplications due to rotational symmetry. Shape stimuli were presented in the center of the receptive field (RF) of the cell under study and were sized such that all parts of the stimuli were within the estimated RF of the cell. Specifically, the outermost stimulus edges were at a distance of 3/4 of the RF radius, which was estimated based on the reported relationship between eccentricity and RF size (Gattass et al., 1988).

For modeling and fitting, each shape was generated as a discretized binary mask of 128×128 pixels, then convolved with a Gaussian filter of standard deviation 1 pixel (*e.g.*, Fig. 2.2a). This image represents a 5×5 patch of the visual field to approximate the experimentally used resolution (Pasupathy and Connor, 2001). The cutoff frequency of

this representation is 12.8 cyc° (half of the 25.6 pixels/° resolution). Because the typical stimulus size was about 3° diameter in the electrophysiology study of Pasupathy and Connor (2001), we made the largest stimulus have a diameter of about 75 pixels within the 128 pixel field. Fourier transforms of stimulus images were computed using a 2D FFT algorithm. The magnitude of complex-valued Fourier components was subjected to the transformation $t(x) = log(x+1)$ to attenuate the low-frequency power that is largely similar across all shapes (Fig. 2.2b). Due to the limited number of stimuli and trial repetitions, power spectra were down-sampled to reduce the number of dimensions in the representation to facilitate model fitting (below). Specifically, a spectral power sample (Fig. 2.2c) was created by summing over 7×7 pixel blocks within the spectrum, with the middle block centered on the DC bin, to achieve a 17×17 grid (the extra few pixels at the margins were ignored). This limited our frequency representation to 0 to 12 cyc/°, which exceeds the range used in a comparable study (David et al., 2006). Because of the even symmetry of the power spectrum, this resulted in a 17×9 pixel representation as depicted in Fig. 2.2c, denoted $P = \{P_s\}$ for $s \in S$, where the set of all shapes is denoted S with $|S| = 366$. Overall, the aims of this representation were to approximate the methods used during the original data recording, to reduce the number of parameters to be fit (17×9), given the symmetry in the power spectra), and to represent the vast majority of the frequency range that would be available to the visual cortex at the relevant eccentricities.

2.2.3 SRF and APC models

Spectral receptive field As proposed by David et al. (2006) an SRF model performs a linear combination of the spectral power of the stimulus in discrete bands to predict neural activity. Using the spectral power sample, P_s, of each shape and observed neuronal responses, r_s, the SRF model seeks a set of weights, Φ^{SRF}, to minimize the residual error between model prediction, $P\Phi^{\mathrm{SRF}}$ and r. Finding such a template can thus be cast as a linear least-squares

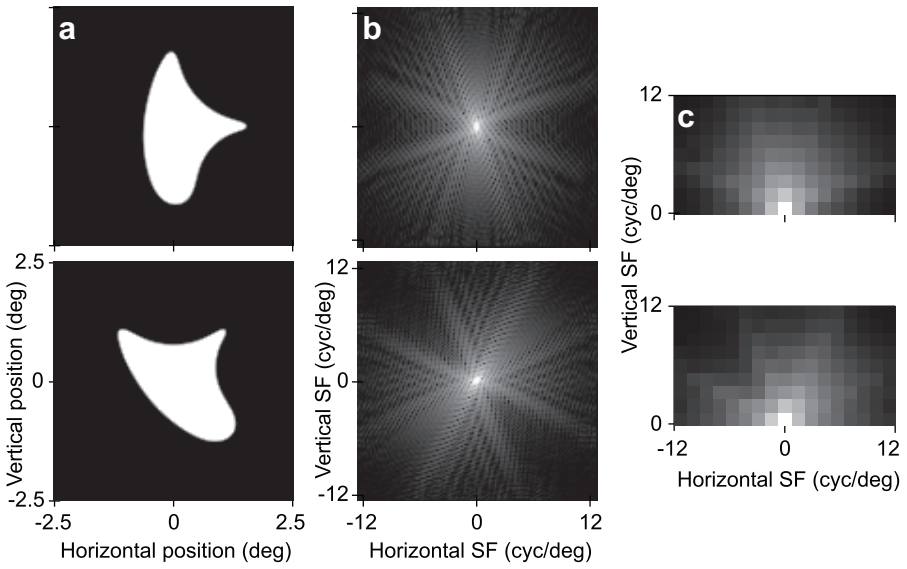

Figure 2.2: **Shape stimuli and their spectral power representation.** (a) Two rotations of shape #18 (Fig. 2.1) are shown at 128×128 pixel resolution, (b) the log of the corresponding Fourier power spectra, and resulting (c) down-sampled spectral power representation used for fitting.

optimization, *i.e.*,

$$\Phi^{\mathrm{SRF}} = \arg\min_{\Phi} \|P\Phi - r\|_2, \tag{2.1}$$

where $\|\cdot\|_2$ denotes the standard Euclidean norm. For procedural convenience, stimulus power spectra are encoded by a 153-element vector representing the coefficients of the 17×9 sampling of spectral power. As neural responses to 366 shape stimuli are considered, P is a 366×153 matrix. Vectors Φ and $r = r_s s \in S$ are of 153×1 and 366×1 elements, respectively.

Due to the high ratio of model parameters to stimuli and correlations among stimuli, the matrix P is ill-conditioned, making standard least-squares prone to over-fitting. To correct for this, we use Tikhonov regularization (Press et al., 2007), *i.e.*,

$$\Phi^{\mathrm{SRF}} = \arg\min_{\Phi} \left(\|P\Phi - r\|_2 + \|\lambda\Phi\|_2\right) \tag{2.2}$$

in place of Equation 2.1, where λ denotes the regularization factor. We tested values of λ from 0.01 to 100 using 100 points that were evenly spaced on a log scale. The data

was divided into 100 randomly chosen partitions of 75% training and 25% test data. Each partition was used to fit and test the model at each λ value. At each λ, we computed M^{λ}_{test} and M^{λ}_{train}, the average explained variance across all partitions in the testing and training data, respectively. For each cell, we defined λ' to be the value of that maximized M^{λ}_{test}, and then defined the training and testing performance to be M^{λ}_{test} and M^{λ}_{train}, respectively. To verify that these methods were sufficient to reveal SRF maps like those reported previously (David et al., 2006), we simulated SRFs having a variety of sizes and shapes, tested them with the same shape set used in the electrophysiology, and confirmed that we could recover the simulated fields.

Angular position and curvature Pasupathy and Connor (2001) proposed an angular-position and curvature (APC) model that performs a non-linear computation over stimuli represented as a set of 4-8 points in the two dimensional space of angular position, θ, and contour curvature, κ. Neural responses are predicted by evaluating a two-dimensional Gaussian energy function (Von Mises in θ) at each of these points, and taking the maximum. In particular, $s_i = (\theta_i, \kappa_i)$ denotes the points defining a shape stimulus s for $i = \{1, \ldots, I_s\}$, where I_s is the number of points. An APC model seeks the energy function parameters $\Phi^{\mathrm{APC}} = (\alpha, \mu_\theta, \sigma_\theta, \mu_\kappa, \sigma_\kappa)$ that minimize the error with respect to the observed neural responses r_s. The APC model is fit through non-linear optimization, *i.e.*,

$$\Phi^{\mathrm{AC}} = \underset{(\alpha,\mu_\theta,\sigma_\theta,\mu_\kappa,\sigma_\kappa)}{\arg\min} \sum_{s\in S} \left[\max_{i\in I_s} \left(\alpha e^{-\frac{\sigma_\theta \cos(\theta_i-\mu_\theta)}{\sigma_\theta} - \frac{(\kappa_i-\mu_\kappa)^2}{\sigma_\kappa^2}} \right) - r_s \right]^2, \tag{2.3}$$

Unlike SRF modeling, a global optima cannot be found deterministically. We estimated the optimal model parameters by performing gradient descent on the objective function. To avoid locally optimal solutions, descent was repeatedly conducted from random initializations ($n = 100$) sampled from a uniform distribution over the angular position and curvature parameter space. Simulations reveal that global optima are consistently well approximated after only a few repeated descents.

Because responses of many V4 neurons depend on the curvature of three adjoining contour

segments centered at a specific angular position (Pasupathy and Connor, 2001), we also considered an APC model that includes three curvature dimensions and a single angular position dimension. We refer to this as the 4D APC model to distinguish it from the 2D APC model described above. The 4D APC model has nine parameters, which include the 4 additional parameters for the means and standard deviations of the Gaussian functions describing two adjoining curvature dimensions. We used the same 75%/25% data partition scheme for fitting and testing our APC models as described above for the SRF model.

2.3 Results

2.3.1 Response similarity for 180° *stimulus rotations*

The SRF model predicts responses of V4 neurons on the basis of the spectral power coefficients of the visual stimuli; therefore, any SRF-like neuron would naturally yield equivalent responses, up to noise, to stimuli having identical power spectra. It turns out that any stimulus rotated by 180° has the same spectrum as the original stimulus. This follows intuitively because any visual stimulus can be described by its Fourier (sine and cosine) components, and these components do not change their orientation, spatial frequency nor amplitude when rotated 180° in the spatial domain. Formally, denoting the Fourier transform $\mathcal{F}$ of a 2D shape image f

$$\mathcal{F}f(x,y) = \hat{f}(\xi,\zeta), \tag{2.4}$$

The spectral power of a 180° rotation of f, denoted f_R, is equal to the spectral power of f, *i.e.*

$$\begin{aligned} |\mathcal{F}f_R(x,y)|^2 &= |\mathcal{F}f(-x,-y)|^2 \\ &= |\hat{f}(-\xi,-\zeta)|^2 \\ &= |\overline{\hat{f}(\xi,\zeta)}|^2 \\ &= \overline{\hat{f}(\xi,\zeta)} \cdot \overline{\overline{\hat{f}(\xi,\zeta)}} \\ &= |\mathcal{F}f(x,y)|^2. \end{aligned} \tag{2.5}$$

The second step above follows from the time reversal property of the Fourier transform. The third step follows because the Fourier transform of a real-valued function is Hermitian (overline denotes the complex conjugate), and the fourth and fifth steps simply apply the definition of the squared norm as the product of a complex value and its conjugate, *e.g.*, $\|y\|2 = y\overline{y}$. This prediction of the SRF model, that neurons will respond the same to a shape and its 180° rotation, is counterintuitive in light of findings that many V4 neurons are tuned for the angular position of stimulus features around the boundary of a shape (Pasupathy and Connor, 2001), the latter being a property that is grossly changed by 180° rotation. For example, if a neuron is tuned for a sharp convexity to the right, it would respond strongly to a shape such as that in Fig. 2.2a (top), but not to the 180° rotation of that shape (not shown).

To test this prediction of the SRF model, we identified all pairs of shapes in our stimulus set that were 180° rotations of each other. For example, the shape in Fig. 2.2a was presented at 8 rotations and thus contributed 4 such 180° rotation pairs. We assessed the amount of correlation, $r_{180°}$ (Pearsons r-value) in these paired responses for each cell; data for three example cells are depicted in Fig. 2.3 (see legend for details). The first example cell (b1601, Fig. 2.3a) shows positive correlations for 180° rotations, cell a8602 (Fig. 2.3b) shows no correlation, and cell b2601 (Fig. 2.3c) shows anti-correlation. The first example would appear to be consistent with the idea that responses are similar for 180° rotations, whereas the third clearly contradicts this notion, suggesting that if a shape produces a larger than average response, its 180° rotation typically does not. However, the observed correlation must be interpreted relative to the amount of correlation between spectrally-dissimilar stimuli, *i.e.*, non-180° rotation pairs. To calculate this baseline correlation, r_{baseline}, we chose 4 of the 24 possible non-180° pairings at random for each shape (where 8 rotations were presented) and calculated the bootstrap distribution of r-values (Fisher z) from repeated simulations ($n = 100$, which proved to be convergent). Fig. 2.4a shows an example (cell a6802) where the response correlation for 180° rotations is significantly positive ($p < 0.05$) but not different from the correlation of non-180° pairings (Fig. 2.4b). It turns out that many cells show a

positive baseline correlation because they respond better to some shapes than others regardless of orientation. This can arise simply from shapes that have similar attributes repeated along their boundaries (*e.g.*, Fig. 2.1, shape 24 gray arrow) or from sensitivity to attributes that are not changed by rotation, such as surface area.

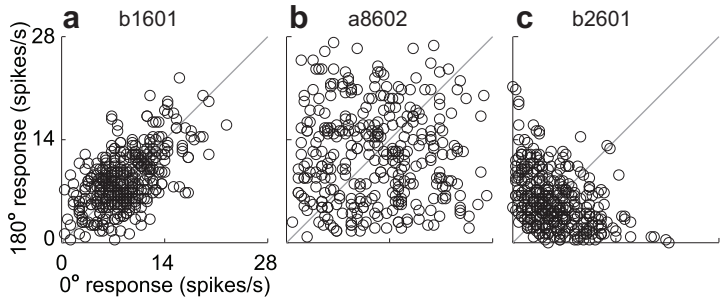

Figure 2.3: **Comparing responses to a shape and its 180° rotation.** Mean responses to a shape and its 180° rotation are plotted against each other for three example neurons. Data points are derived from the 42 shapes for which 8 rotations were presented. Each 180° rotation pair contributes two points, (x, y) and (y, x), for a total of 336 points per neuron. The SRF model predicts that $y = x$, up to noise, and thus a high positive correlation coefficient is expected. Nevertheless, our population contained neurons with (a) positive correlation ($r = 0.54$), (b) no correlation ($r = 0.00$) and (c) negative correlation ($r = -0.39$). Labels b1601, a8602, and b2601 indicate cell names, where the first letter indicates the animal identity.

The population results of this analysis for the dataset of 109 cells are shown in Fig. 2.5a, where $r_{180°}$ is plotted against r_{baseline}. The significance level is set at 2 of baseline correlation. Note that most neurons ($n = 68$) lie near the line of equality, *e.g.*, a6802 (from Fig. 2.4; point 4 in Fig. 2.5a). Interestingly, some cells, *e.g.* b1601 (from Fig. 2.3a, point 1 in Fig. 2.5a), fall significantly above equality, indicating possible selectivity for features that are preserved across 180° rotations and are potentially consistent with an SRF model.

We compared the scatter of data in Fig. 2.5a to that expected from an idealized SRF model that includes realistic (Poisson) noise. We did this by setting an underlying mean firing rate (target rate) for each shape and then deriving from it a measured rate by sampling a spike count from the target rate five times with Poisson statistics (variance equal to mean).

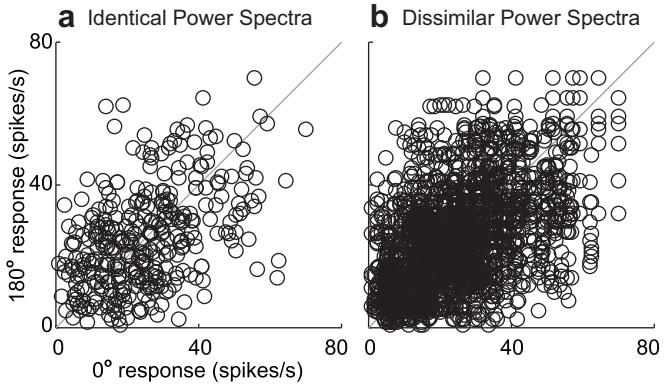

Figure 2.4: **Computing baseline correlation between non-180° shape rotations.** (a) Mean responses of neuron a6802 to pairings of stimuli that are spectrally identical, *i.e.*, 180° rotation, (same analysis and plot format as in Fig. 2.3), and again with (b) responses plotted for all pairings for a given shape that are not 180° rotations. There are 24 such pairings for each shape with 8 rotations, compared to only four pairings for 180° rotations. Correlation coefficients are similar in (a) and (b), $r = 0.46$ and $r = 0.47$ respectively, suggesting that for this neuron average responses are higher for some shapes than for others, regardless of rotation.

To embody the SRF model, we set the target rates equal for pairs of shapes that were 180° rotations, choosing randomly between the two experimentally observed rates. From these measured rates, we computed $r_{180°}$ as described above. We repeated this process 100 times and determined the average correlation (using Fisher z). In Fig. 2.5b, the results of this statistical simulation (diamonds) are plotted together with the actual data (filled gray circles) and against the same r_{baseline} values. The results indicate that hypothetical SRF units show much higher values of $r_{180°}$ than were observed in our data. This suggests that, while a few cells (*e.g.*, neuron b1601) show consistency with the SRF model, the vast majority of neurons from our population do not.

We performed a similar simulation using the response rates predicted by the APC model (see Section 2.2.3) for comparison to the neuronal data and to the responses of the idealized SRF model. Each cell was fit to the APC model (see Section 2.3.2) and the resulting predicted mean responses were used as the target rates. Observed rates were computed from

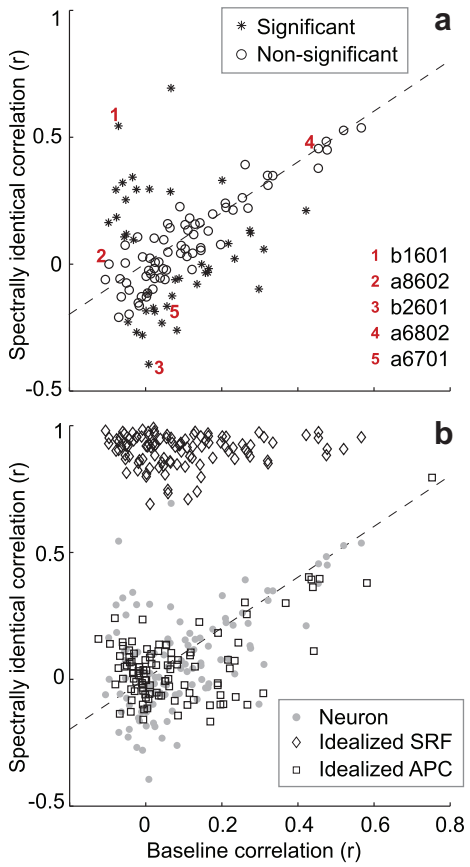

Figure 2.5: **Comparing response correlation for** 180° **rotations to baseline correlation** (a) For each of the 109 neurons, the correlation coefficient for responses to spectrally identical shapes (180° rotations) is plotted against the baseline correlation value (the correlation between responses to non-180° pairings, see Section 2.3.1). Points fall on both sides of the line of equality (dashed line). Points plotted as asterisks indicate neurons for which the ordinate deviates from the abscissa by more than two standard deviations, based on a bootstrap estimate of the baseline correlation distribution. The points numbered 1-3 correspond to the three example neurons in Fig. 2.3, point 4 corresponds to the example in Fig. 2.4 and point 5 corresponds to an example neuron shown later in Section 2.3.2. (b) The data from (a) are replotted here (gray filled circles) and are compared to predictions for an idealized SRF model (open diamonds) and an idealized APC model (open squares) for each of the 109 neurons. As expected, points for the idealized SRF prediction have much higher correlation for spectrally identical stimuli, because the SRF model predicts identical mean responses for such stimuli.

the average of five Poisson samples. The result (Fig. 2.5b, open squares) shows that the APC model predicts a much lower $r_{180°}$ value than does the SRF model, and that the predicted values are approximately consistent with the range of values found for the neurons.

In summary, the SRF model makes a distinct prediction about 180° rotations that the APC model does not, and with respect to this prediction, the SRF model is far less consistent with our data than is the APC model.

2.3.2 Model fitting and performance

Although the SRF model fails to predict the differences in neuronal responses to shapes and their 180° rotations, previous reports show that both the SRF and the APC models account for only part of the variance of V4 responses (Pearsons r-value of 0.32 for the SRF model of David et al., 2006; 0.57 for 4D APC model of Pasupathy and Connor, 2001). We thus wanted to establish, *i*) what fraction of the variance is captured by the SRF model across the entire set of shapes, and how does this compare to that previously reported for the SRF and APC models, and *ii*) are the cells that are well fit by the SRF model, in terms of amount of explained variance, also the ones that are well fit by the APC model?

We performed an empirical evaluation of both SRF and APC models by fitting to, and predicting, recorded neural responses to our stimuli. We partitioned our data into training and testing sets for cross-validation, and we measured model performance in terms of explained variance (r^2) for both sets. Bootstrap validation estimates (Fig. 2.6a) show that although the SRF model outperforms both APC models across training datasets, it underperforms both the 2D and 4D versions of the APC model on the test datasets. This is a hallmark of over-fitting: the SRF model has about thirty times the number of parameters (9×17 spectral weights) compared to the 2D APC model (5 parameters) and 17 times that of the 4D APC model (9 parameters). When comparing only the testing validation performance across all neurons (Fig. 2.6b), the responses of the majority of neurons (77 out of 109) are better predicted by the 4D APC model than the SRF model, with a significantly higher average explained variance (mean 0.09, standard deviation 0.13, paired t-test $p < 0.0001$).

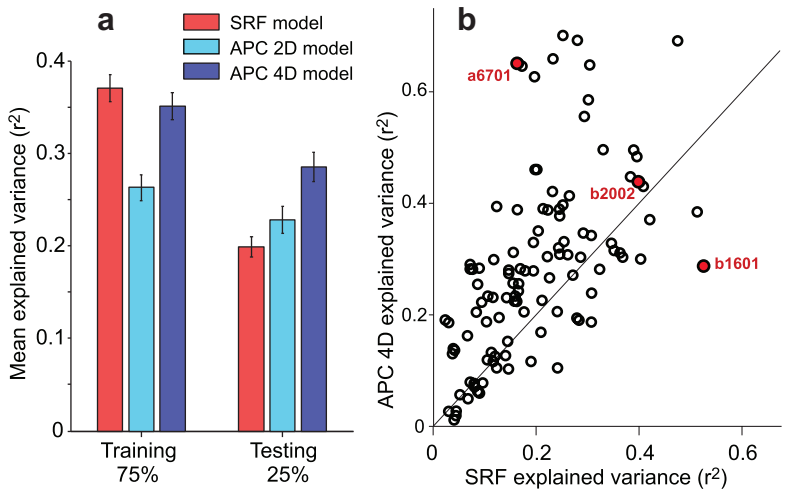

Figure 2.6: **Fit performance of SRF and APC models.** (a) The mean explained variance is plotted for the SRF model and the 2D and 4D versions of the APC model for the training data (left) and testing data (right). The training and testing partitions were 75% and 25% of the data, respectively. The SRF model explained more variance in the training data than did the APC models, but both APC models out-performed the SRF model on the testing data on average. Error bars show ±1 standard error on the mean. (b) Explained variance values for the 4D APC model are plotted against those for the SRF model for all 109 neurons. Three examples are indicated: b1601, which was better fit by the SRF model, a6701 which was better fit by the APC model, and b2002 which was about equally well fit by both models.

Although the performance of the SRF model was relatively weak, this does not appear to reflect particular limitations of our stimulus set, because the performance was favorable to, and in fact better than, that reported previously (David et al., 2006): our mean r-value was 0.43 ($n = 109$), compared to their mean of 0.32.

Another important feature of the scatter in Fig. 2.6b is the paucity of points near the upper right corner. This implies that the neurons best explained by the APC model are not also those that are best explained by the SRF model. For example, neuron b1601 (Fig. 2.6b, lower right) was among the most SRF- like cells: its responses were best fit by the SRF model ($r^2 = 0.54$) and were also among the most consistent with the prediction regarding 180° rotations examined above (Fig. 2.5a), but its responses were not well explained by the APC model ($r^2 = 0.2$). On the other hand, points do fall near the extreme lower left

in Fig. 2.6b, representing neurons that are poorly fit by both models. This is expected under the simple assumption that some neurons do not respond well to the stimulus set, or have very noisy responses. Discarding the neurons that were poorly fit by either model ($r^2 < 0.15$), there was no significant correlation between the explained variance of the APC and SRF models ($r = 0.17$, $p = 0.17$, $n = 65$). This suggests that these distinct models do not capture the same features of the response.

To understand the tuning properties of neurons that were well fit by the SRF model and compare them to those that were well fit by the APC model, it is useful to examine the raw responses and fit parameters for several example neurons. The responses of the SRF-like neuron, b1601, for each of the 366 shapes are plotted in Fig. 2.7a, where red indicates the strongest responses and blue the weakest. This neuron tended to respond most strongly to shapes that were oriented horizontally, and the strongest responses were often offset in the diagram by 4 rows (paired black arrows), which corresponds to 180° of stimulus rotation. We will see below (Fig. 2.8a) that the SRF map for this neuron reflects this apparent preference for horizontal orientation. A second example neuron (Fig. 2.7b) that was moderately well fit by both models (red point for b2002 in Fig. 2.6b) responded strongly to stimuli that were oriented vertically or tilting somewhat toward the right. Here, some but not all of the stimuli evoking the strongest responses were separated by 180° (arrow pairs), consistent with the moderate fit of the SRF model. A contrasting example (Fig. 2.7c) shows a neuron that did not display a clear preference for overall orientation. In particular, the strongest responses are not separated by 180° rotations (arrow pairs), consistent with the poor fit of the SRF model (red point for a6701 in Fig. 2.6b). All of the shapes that evoke strong responses from this cell include a concavity to the right side of the shape. This type of tuning is well captured by the APC model, as indicated by the relatively high explained variance value (a6701 in Fig. 2.6b).

The SRF maps for the example neurons just described are shown in Fig. 2.8. As described in Section 2.2.3, we fit SRF maps over a broad range of regularization values, λ, computing training and test performance at each value to assess and minimize the influences of over-

Figure 2.7: **Shape tuning maps for three example neurons.** (a) The mean firing rate of neuron b1601 to each shape (drawn in black) is indicated by the color surrounding the shape. Dark blue and dark red indicate the lowest and highest responses, respectively (see scale bar at bottom). This neuron responded best to shapes that had a horizontal alignment, and 180° rotations of the same shape often gave roughly similar responses (black arrow pairs). All rotations (up to 8) of each shape are arranged contiguously within a single column in one block. (b) Responses for example neuron b2002, which tended to prefer shapes with a vertical or right-leaning alignment. Sometimes responses to 180° rotations were similar (e.g., black arrow pair). (c) Responses for example neuron a6701, which was well fit by the APC model and poorly fit by the SRF model (Fig. 2.6b). Shapes associated with the strongest responses did not elicit strong responses when rotated by 180° (compare upper and lower arrows within each arrow pair).

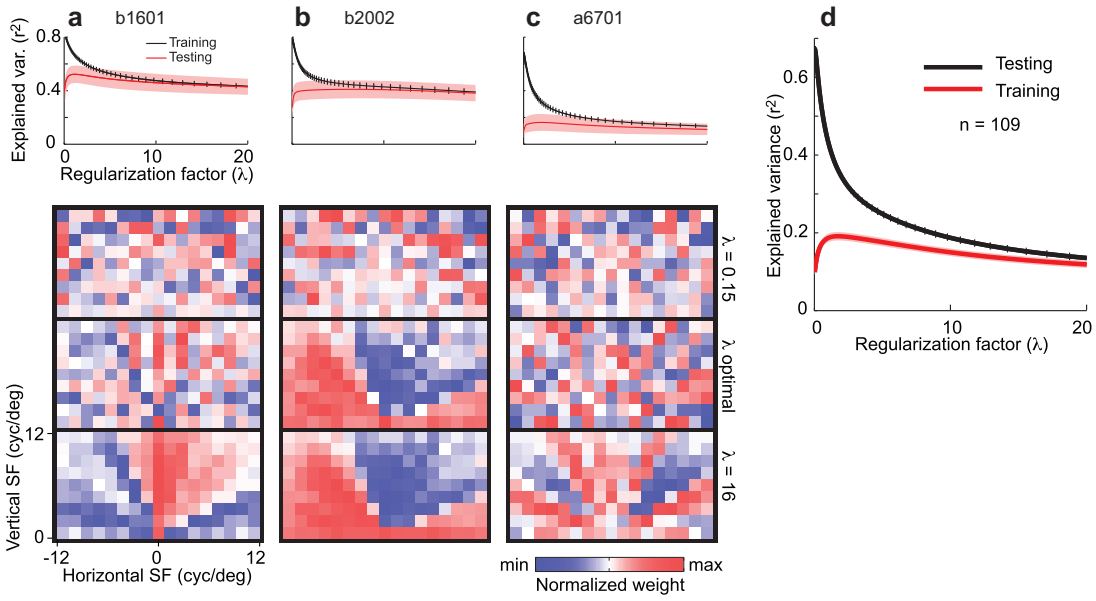

Figure 2.8: **SRF maps depend on regularization.** (a-c) SRF maps (image panels) are shown for the three example cells of Fig. 2.7 (columns), and for three levels of the regularization parameter (low to high, rows 1-3, respectively). Red indicates positive weights and blue indicates negative weights (see scale bar near bottom). The top row shows maps for low λ (0.15). These maps produce the best performance on the training data (black line, bottom panels, described below) but substantially worse performance on the test data (red line, bottom panels) because of overfitting. The second row shows maps at the optimal λ (best test performance) for each neuron ($\lambda = 1.52$, 5.09 and 0.38 for a-a, respectively). The third row shows maps for high λ (15.6). Each map shows an example of the SRF given a random selection of training/testing partition (75%/25%). The top panels plot the average performance (across 100 random training/test partitions) on the training (black) and test (red) data as a function of λ. Shaded area shows ± 1 standard deviation. (d) Average performance on the training (black) and test (red) data as a function of λ across all 109 neurons. Shaded area shows ± 1 standard deviation.

fitting. For neuron b1601 (Fig. 2.8a, top), the training performance (black line) declined with increasing λ, while the testing performance (red line) increased to a maximum and subsequently fell to an asymptote. This behavior is expected, and held for all neurons (Fig. 2.8d shows population average). For each neuron, SRF maps are shown (below the performance plots) for low, optimal (highest test performance) and high regularization values. Each map shows spectral weights as a function of horizontal and vertical spatial frequency.

In this representation, frequency increases with distance from the origin, and power at a particular orientation falls along a line radiating from the origin. At low λ (top row of maps), the maps have a salt-and-pepper appearance that fits the training data well, but they strongly underperform on the testing data and thus are not likely to reflect a true receptive field. At high λ (here $\lambda = 16$, but maps were similar over a broad range), the training and test performance become nearly equal, suggesting that the features remaining in the maps are those that best generalize beyond the training set. Indeed, the $\lambda = 16$ map for neuron b1601 (Fig. 2.8a, bottom) has a red streak along the vertical axis, indicating a preference for horizontal orientation, which is apparent in Fig. 2.7a. The high-λ map for neuron b2002 (Fig. 2.8b, bottom) has a red streak along the horizontal axis which expands upward in the left quadrant, indicating a preference for vertical to right-leaning orientation, as observed in Fig. 2.7b. In contrast, the SRF map for neuron a6701 (Fig. 2.8c) has red streaks at multiple orientations, and most notably, the performance (top panel), is substantially lower at all compared to the first two examples.

In summary, the correspondence between the coherent structure within the SRF maps (Fig. 2.8) and the raw shape responses (Fig. 2.7) suggests that our SRF fits provide a useful characterization for some neurons, but that these neurons also appear to be ones that display sensitivity to the overall orientation of a shape.

2.4 Discussion

We examined whether the selectivity of V4 neurons for boundary curvature can be simply explained in terms of tuning for the spatial frequency power spectrum as quantified by the SRF model. We found that the responses of curvature tuned V4 neurons are inconsistent with the SRF model on several counts. First, the SRF model predicts identical responses to 180° rotated stimuli but most V4 neurons, especially those that are curvature-tuned, do not exhibit this property. Second, in comparison to the curvature-based model, the SRF model captured significantly less of the variance in V4 responses for a set of parametrically designed two-dimensional complex shapes. Finally, the V4 neurons that were particularly

well fit by the SRF model were also those that could be roughly described as showing simple orientation tuning, and were not among the best fit by the curvature model.

A previous attempt to show that the SRF model could unify V4 neuronal selectivity from studies using disparate stimulus sets (David et al., 2006) was motivated by several attractive features of the model. The SRF model describes V4 tuning in terms of sensitivity to particular frequency bands within the power spectrum of the visual input. Because the frequency bands can be labeled in terms of orientation and spatial frequency, the SRF model can be viewed as a simple extension of the representation present in V1, where neurons are tuned to stimulus orientation (Hubel and Wiesel, 1968) and spatial frequency Campbell et al. (1969); Movshon et al. (1978); Albrecht et al. (1980); De Valois and De Valois (1990). This has the advantage that the circuit implementation of a V4 neuron in terms of the SRF model would be a relatively straightforward combination of V1 outputs. Another key feature of the SRF model is the second-order nonlinearity inherent to the power spectrum that discards phase information and can thereby produce phase and position invariant responses, approximating similar characteristics of V4 neurons (Gallant et al., 1996; Pasupathy and Connor, 2001, 1999; Rust and Dicarlo, 2010). However, the simplification of discarding phase information before integrating across frequency bands ignores a key feature of V4 curvature selectivity. Specifically, a V4 neuron may respond preferentially to a sharp convexity pointing upward relative to the object center but not to that same feature pointing downward; the SRF model cannot reproduce this important aspect of curvature tuning because phase-insensitive Fourier power models predict identical responses for pairs of stimuli that are 180° rotations of each other. We directly examined the responses of V4 neurons to such pairs of stimuli and found that this prediction did not hold, in contradiction to the SRF model. We conclude that a defining characteristic of the SRF model - that phase information is dropped before combining spatial frequency components across the image - is inconsistent with curvature selectivity in V4.

Because all current models of V4 have limitations, it is important to consider how the SRF model compares to alternatives in its ability to explain the variance of neuronal responses

to the same stimulus set. We fit SRF maps to V4 responses to a set of simple shapes that parametrically explored a space of contour curvature and angular position. Our SRF maps were roughly consistent with those reported previously (David et al., 2006; see Figures 1-3). Our maps often showed tuning for multiple orientations, similar to theirs, and our maps explained a larger fraction of the response variance than their maps did. One difference was that their spatial resolution was 12 cyc/RF, whereas ours was about three times higher (12 cyc/° with typical RF sizes approximately 3°). Nevertheless, the SRF model captured less response variance on average than did our APC model, which had far fewer parameters. Two observations are particularly worth noting. First, none of the cells best fit by the curvature model (20 cells for which $r^2 > 0.4$, 4D APC model), were better fit by the SRF model. This suggests that the SRF model does not capture the key features of curvature selectivity that are represented in the curvature model. Second, a closer examination of the cells best fit by the SRF model reveals that they would be well-described as orientation selective, consistent with examples of David et al., 2006 (see Figure 1b,3a). Thus, the SRF model does not provide a sufficient framework to understand curvature tuning in V4; nevertheless, it may serve an important role in describing cells in V4 whose tuning is largely in the orientation dimension. Future work will be required to understand how these different types of tuning operate together in V4.

Although the contour curvature model provides a good fit to the responses of many V4 neurons, it has the limitation of being a descriptive model and does not point to any obvious implementation in terms of biologically plausible circuitry. One model to derive curvature selectivity in V4 from inputs coming from V1 and V2 would involve first coarsely defining an object, *i.e.*, segmenting it, and then assessing the orientation progression along its boundary. The latter step is captured by the model of Cadieu et al., 2007 (discussed below). The former step, segmentation, is more challenging but could be achieved by a set of grouping cells like those proposed by Craft et al. (2007) as a mechanism for creating border ownership signals in V2. Grouping cells group together concentric contour segments, and a set of such cells captures the coarse shape of an object. This is equivalent to finding the set of largest discs

that would just fit within a bounding contour, a method proposed for computing the medial axis of a shape (Blum, 1967). Grouping cells are hypothesized in V4 and could send lateral connections to curvature-sensitive neurons. Inputs from the set of grouping cells would specify the centroid of the stimulus in a graded fashion. Further experiments are needed to explore this possibility, but preliminary results from our laboratory suggest that the earliest responses in V4 encode the overall size of the stimulus, which supports this hypothesis.

Alternatives to the APC and SRF models considered here include a set of biologically inspired hierarchical models (Serre et al., 2005; Cadieu et al., 2007; Rodríguez-Sánchez and Tsotsos, 2012). The model of Cadieu et al., 2007 has been shown to account for the curvature tuning of V4 neurons using the same data set examined here - Cadieu et al., 2007 Figure 10A shows that their model performed similarly to the 4D APC model in terms of explained variance. This model, however, does not operate in an object-centered system and does not explicitly represent curvature. Curvature is built up as a combination of oriented segments, and translation invariance is achieved in small steps of positional invariance implemented by using the max-function. The model of Rodríguez-Sánchez and Tsotsos (2012) explicitly represents curvature-tuning at intermediate stages in the visual hierarchy and implicitly uses an object-centered coordinate system. These models may provide a useful foundation for testing the nature of an object-centered representation and for developing a more complete model that encompasses novel recent findings related to object segmentation in V4 that have yet to be modeled (Bushnell et al., 2011a).

In conclusion, it is essential to seek out the simplest models, and the SRF model is therefore an important point of comparison. However, responses of V4 neurons appear to reflect the solutions to some of the most difficult problems in visual object recognitionthat of translation invariance and object segmentation, so it may be unsurprising if simple combinations of V1 outputs do not account for V4 responses. To advance our understanding of V4, it will be important to: *i*) develop a mechanistic implementation that explains curvature responses, *ii*) extend such models to handle complex scenes, and *iii*) conduct experiments to further characterize those V4 neurons that are not well explained by either the APC or SRF models.

Chapter 3

JOINT CODING OF SHAPE AND BLUR IN AREA V4

Edge blur, a prevalent feature of natural images, is believed to facilitate multiple visual processes including segmentation and depth perception. Furthermore, image descriptions that explicitly combine blur and shape information provide complete representations of naturalistic scenes. Here we report the first demonstration of blur encoding in primate visual cortex: neurons in macaque V4 exhibit tuning for both object shape and boundary blur, with observed blur tuning not explained by potential confounds including stimulus size, intensity, or curvature. A descriptive model wherein blur selectivity is cast as a distinct neural process that modulates the gain of shape-selective V4 neurons explains observed data, supporting the hypothesis that shape and blur are fundamental features of a sufficient neural code for natural image representation in V4.

3.1 Introduction to shape and blur

In any natural scene, visual information is carried by boundaries of contrast that exist throughout an image (Attneave, 1954). For example, figure-ground contrast along the borders of solid objects may provide robust cues for object shape (Geisler et al., 2001), while internal boundaries and surface texture may reveal 3D structure and material composition of those objects (Ikeuchi and Horn, 1981; Fleming et al., 2011). Much work has quantified the extent to which these edge cues contribute to complex complex visual tasks, such as segmentation and recognition (Wagemans et al., 2012), and progress is being made toward understanding the neural mechanisms responsible (*e.g.*, see Hanazawa and Komatsu, 2001; Cadieu et al., 2007; Fleming et al., 2011; Tao et al., 2012; Drewes et al., 2016).

However, physical environments under naturalistic viewing conditions often produce edges

that are blurred, *i.e.* exhibit a spatial gradient of contrast orthogonal to the boundary contour (Fig. 3.1). These blurred boundaries can arise from a number of scenarios, such as defocus, cast shadows, or surface shading (Elder, 1999), and thus themselves convey relevant scene information such as object depth (Held et al., 2012; Burge and Geisler, 2014). Importantly, computational studies of luminance boundaries find that visual scenes may be sufficiently reconstructed from information contained in edge features, including the magnitude of blur at each edge (Elder, 1999). Further, psychophysical results demonstrate that in addition to shape (Elder and Velisavljević, 2009), the visual system is tuned to detect cast shadows during segmentation (Rensink and Cavanagh, 2004), of which shape and blur are diagnostic features.

While the visual system is adept at discriminating blur (Sebastian et al., 2015) and detecting blurred boundaries (Rensink and Cavanagh, 2004; Watson and Ahumada, 2011), neural mechanisms that underlie the computation and representation of blur remain unclear. Since sharp and blurred boundaries differ greatly in their high spatial frequency (SF) content, V1 populations tuned to various SFs implicitly encode blur. However, at intermediate stages of form processing, such as in area V4, simple gratings are ineffective at driving responses of shape-selective neurons (Gallant et al., 1996), and complex shape stimuli that *do* elicit responses have typically been defined by sharp boundaries (Kobatake and Tanaka, 1994; Carlson et al., 2011; Pasupathy and Connor, 2001). Thus, it is unknown if and how blur is encoded and combined with shape information along the ventral pathway to form a sufficient representation of natural scenes.

Here we present results from a study targeting single V4 neurons using customized sets of shape stimuli to test the hypothesis that V4 neurons jointly encode object shape and boundary blur. Our results demonstrate that shape-selective V4 neurons also exhibit tuning for blur and that single-unit responses are well-described by a joint model explicitly encoding both shape and blur information.

3.2 Materials and methods

3.2.1 Animals and surgery

Two rhesus monkeys (*Macaca mulatta*, one female and one male) were surgically implanted with custom-build head posts attached to the skull with orthopaedic screws. After fixation training, a recording chamber was implanted; a craniotomy ($\approx$ 10 mm diameter) was subsequently performed to expose dorsal area V4. See Bushnell et al. (2011b) for detailed surgical procedures. All animal procedures conformed to NIH guidelines and were approved by the Institutional Animal Care and Use Committee at the University of Washington.

Animals were seated in front of a CRT monitor at a distance of 57 cm and were trained to fixate on a 0.1° white dot within 0.5°-0.75° of visual angle for 3-5 seconds for water reward. Eye position was monitored using a 1 kHz infrared eye-tracking system (Eyelink 1000; SR Research). Stimulus presentation and animal behavior were controlled by customized software PYPE (originally developed in the Gallant Laboratory, University of California, Berkeley, Berkeley, CA). Each trial began with the presentation of a fixation spot at the center of the screen. Once fixation was acquired, four to six stimuli were presented in succession, each for 300 ms, separated by interstimulus intervals of 200 ms. Stimulus onset and offset times were based on photodiode detection of synchronized pulses in the lower left corner of the monitor.

3.2.2 Data collection

During each recording session, a single transdural tungsten microelectrode was lowered into cortex with an electromechanical microdrive system (Gray Matter Research). Electrode signals were amplified and single-unit activity was isolated using online spike sorting (Plexon Systems). Electrode penetrations targeted dorsal V4 from structural MRI scans localizing the prelunate gyrus. Single unit waveforms that responded briskly to the onset of shape stimuli were identified for further recording. After data collection, spikes were sorted offline with custom software (Plexon Systems) and exported for analysis.

3.2.3 Visual stimulation

For each recorded neuron, we first characterized the preferred RF location, size, luminance contrast and chromaticity with custom shape stimuli under mouse control. Shape stimuli were presented on an achromatic gray background of mean luminance 5.4 cd/m^2. Foreground luminance was chosen from four values (2.7, 5.4, 8.1, or 12.1 cd/m^2) that were darker, equiluminant or brighter than the background; chromaticity was selected from 25 gamma-corrected hues spanning the CIE color space (Bushnell et al., 2011a). Next, we assessed shape selectivity with a standard set of 366 shape stimuli generated by rotating 51 shapes (Fig. 3.1a) by increments of 45° (Fig. 3.1b), and discarding duplicates due to radial symmetry. The design of these stimuli is described in detail elsewhere (Pasupathy and Connor, 2001). All stimuli were presented in the center of the RF and were scaled such that all parts of the stimuli were within the estimated RF of the cell: the largest shape stimulus typically had outermost edges at a distance of 75% RF radius. Stimuli were presented in random order without replacement with three repeats per stimulus.

To assess how stimulus blur influences V4 responses, we identified 5-8 shape stimuli that evoked a range of responses, from weak (non-preferred) to strong (preferred), during the shape screen described above; we then studied responses to these shapes subjected to different levels of blur. As V4 neurons respond selectively for shape orientation (Pasupathy and Connor, 2001), it was often the case that preferred and non-preferred stimuli were chosen to be 180° rotation pairs of the same shape. This had the added benefit of controlling for spectral content, since such stimuli have identical spectral power (Oleskiw et al., 2014). Additionally, neutral curvature (circular) stimuli were also included for most neurons. Each of the chosen shapes were presented at up to 9 blur factors along an approximately exponential scale, *i.e.* $\beta \in \{0.005, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.48, 0.64\}$ (see Fig. 3.1e). Stimuli were blurred by convolving the discretized raster image with a circular 2D Gaussian blur kernel. The kernel standard deviation, denoted by a blur factor β, is written in units relative to the radius of the large circle (Fig. 3.1a, black arrow). Due to the limited color gamut of the display,

dithering was employed by noising each pixel with zero-mean Gaussian noise with standard deviation of two 8-bit greylevels. The resulting pixel intensities were linearly interpolated between background luminance and preferred color, rounded to the nearest calibrated RGB values. Finally, to prevent aliasing, stimuli were down-sampled by a factor of two. During the shape screen to assess shape selectivity, sharp stimuli were presented under a minimal blur factor of $\beta = 0.005$. Blur stimuli were randomly chosen without replacement with a median of 20 repetitions.

3.2.4 Control experiments

On a subset of cells, we conducted control experiments to evaluate whether preferred responses to intermediate blur factors could be explained on the basis of selectivity for stimulus size or stimulus contrast. Because stimulus blur increases the number of pixels distinct from the background (see Fig. 3.4a), preference for a specific level of blur could represent preference for stimulus size. If this were the case then blur preference will depend on the absolute size of the stimulus. To control for this, we presented blurred stimuli at multiple sizes and asked whether blur preference depended on the size of the stimulus. Size control stimuli were generated from up to three exemplar blur factors, consisting of the extremal blur levels $\beta = 0.005$ and 0.64, along with an intermediate factor, typically the blur factor that evoked the strongest responses from the neuron. Shape stimuli at each of the blur factors were resized with scaling factors of 0.9 and 1.1, *i.e.* scaled by $\pm 10\%$. These factors were chosen for their visual correspondence to stimuli subjected to intermediate blur factors of 0.08 and 0.16, *i.e.* Fig. 3.4a, and approximate contours generated from luminance thresholds at 1/3 and 2/3 of maximum stimulus intensity under blurring of $\beta = 0.16$. Each stimulus was presented randomly with 10 to 20 repeats. To evaluate whether preference for an intermediate level of blur could be explained on the basis of preference for average stimulus contrast, we also studied responses to contrast control stimuli generated by first computing mean stimulus intensity across the interior of a blurred shape stimulus, where shape interior is defined by the half-contrast level set. Then, for each blur factor, a control stimulus was generated with

a foreground intensity equal to the mean intensity within this level set (see Fig. 3.5). Control stimuli were subjected to only the minimal blur factor of $\beta = 0.005$, resulting in stimuli with sharp boundaries of reduced figure-ground contrast.

3.2.5 Analysis and model fitting

Neural responses to individual stimuli were calculated as the mean firing rate observed during stimulus presentation, 300 ms in duration with a 50 ms lag relative to onset, averaged across repeats. Peristimulus time histograms were computed for each stimulus by filtering spike rasters with a centered (noncausal) decaying exponential filter consistent with a membrane's integration time constant (37 ms).

Preferred-shape blur tuning curves (see Fig. 3.3 and Fig. 3.5) were constructed by first identifying preferred shapes, *i.e.* sharp stimuli ($\beta = 0.005$) that evoked a response greater than mean across shapes (2-4 preferred shapes per cell), and averaging across preferred shapes for each blur factor. A cubic spline is then fit to average preferred shape responses. Center-of-mass (CoM) was calculated by integrating preferred tuning curves across blur factors in log space (or similarly in intensity-matched factors during intensity control analysis), and returning the median cumulative factor. Significance of CoM measurements from intensity-matched controls was determined by bootstrapped estimates of tuning curves sampled from response distributions of recorded means and variances from each stimulus (100 repetitions) under a 2-way t-test of unequal variance.

The angular position and curvature (APC) model has been shown to accurately capture shape tuning properties of single-unit V4 spike-rate responses (Pasupathy and Connor, 2001; Oleskiw et al., 2014). Here, each shape stimulus is represented by 4-8 points in the space of angular position θ and curvature κ, corresponding to locations of curvature inflection along the contour. In particular, a stimulus $\Gamma = (\gamma^1, \gamma^2, \ldots, \gamma^n)$ is represented by n critical points $\gamma^i = (\gamma^i_\theta, \gamma^i_\kappa)$. The model predicts responses to shape stimuli by evaluating a Gaussian energy function (von Mises in periodic angular position) at each of these points, and returning the maximum. When applicable, Γ is augmented with a blur factor γ_β proportional to the kernel

size of a Gaussian-blurred shape stimuli. Thus,

$$APC(\Gamma;\omega,\alpha,\mu_\theta,\sigma_\theta,\mu_\kappa,\sigma_\kappa) = \omega + \alpha \exp\left(\frac{cos(\gamma_\theta^i - \mu_\theta)}{\sigma_\theta} - \frac{(\gamma_\kappa^i - \mu_\kappa)^2}{\sigma_\kappa^2}\right) \tag{3.1}$$

describes a model over dimensions of angular position and curvature. Note that the tuning peak μ and width σ are represented along each selectivity dimension θ and κ. Further, baseline parameter ω captures spontaneous activity in the absence of stimulation, and gain α is fit to produce maximal responses for preferred stimuli. We extend the APC model to also predict neural responses as a function of boundary blur. Here, blur selectivity is modeled as Gaussian in the logarithm of blur factors, *i.e.*

$$B(\Gamma;\mu_\beta,\sigma_\beta) = \exp\left(-\frac{(\log(\gamma_\beta) - \mu_\beta)^2}{\sigma_\beta^2}\right). \tag{3.2}$$

Again note preferred blur factor μ_β and blur tuning width σ_β. Every model is fit to minimize squared error between mean firing rate $\vec{r}$ evoked by stimuli $\vec{\Gamma}$, averaged across repetitions, and the responses predicted by the model. For example, a fit Θ^* of neural data to an angular position, curvature, and blur (APCB) model is written

$$\Theta^* = \arg\max_\Theta \left\| \max_i \left(APC(\vec{\Gamma};\Theta) * B(\vec{\Gamma};\Theta) \right) - \vec{r} \right\|_2, \tag{3.3}$$

such that the model predicts a response to any stimulus as the maximum of each critical point γ^i evaluated under that model. The optimal model therefore minimizes the L_2 norm between recorded $\vec{r}$ and predicted responses of stimulus set $\vec{\Gamma}$. Model fitting is complicated by the fact that optimization is highly non-convex. While standard gradient descent methods are quick to converge, solutions are typically only locally optimal: we employ a repeated randomized initialization procedure to approximate globally optimal fits, described elsewhere (Oleskiw et al., 2014).

For each blurred stimulus we computed the boundary contours by binary-thresholding images at a range of intensity levels. These contours were then represented as points in the 2D space of curvature and angular position. Each shape and its blurred counterparts were represented by the same set of angular position values but the curvature values were reduced under blur (see Fig. 3.6a,b).

To compare model fit performance we measure prediction accuracy to a subset of shape stimuli. For each stimulus, the set of critical points along the contour was evaluated under a trained APC model to produce a response prediction $\vec{p}$. We compute the normalized root-mean-squared error (NRMSE) between the predicted and recorded responses $\vec{r}$, *i.e.*

$$\frac{\|\vec{p}-\vec{r}\|_2}{\max(\vec{r})-\min(\vec{r})}. \tag{3.4}$$

NRMSE training estimates were computed via bootstrapping: a random subset of shape stimuli were selected, equal in number to the number of blurred stimuli recorded, and the NRMSE is again calculated between the best-fitting APC model prediction to responses elicited from sharp stimuli from the entire shape set.

3.3 Results

3.3.1 Selectivity for blur in area V4

To understand how blur, *i.e.* the gradient of boundary contrast, is encoded in the intermediate stages of the ventral pathway, we examined the responses of well-isolated V4 neurons to shape stimuli as a function of blur magnitude. For each neuron, we first assessed shape selectivity using a standard stimulus set (Fig. 3.1c,d; Pasupathy and Connor, 2001). Based on these responses, we identified a subset of preferred and non-preferred stimuli that evoked a range of responses for the neuron in question (see Materials and methods: Visual stimulation). We then examined the responses to this chosen subset of stimuli under various levels of blur (Fig. 3.1e).

Blurring a stimulus boundary, as implemented here (see Materials and methods: Visual stimulation), broadens the gradient of contrast across the shape's edges and thus decreases effective boundary contrast. Because the responses of roughly 80% of V4 neurons increase with increasing luminance contrast across the shape boundary (Bushnell et al., 2011a), one may expect blur to consistently reduce the responses of neurons. Indeed, many neurons in our population follow this trend. For example, cell a23 in Fig. 3.2a exhibited a range of responses

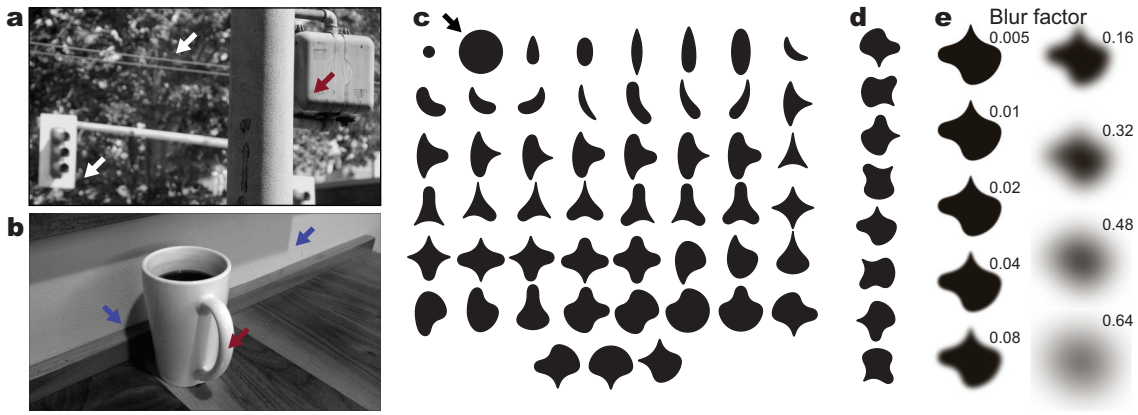

Figure 3.1: **Examples of blur in natural images and stimuli used to explore selectivity for shape and blur** (a, b) Examples of different types of blur in natural scenes. (a) Focal blur (white arrows) conveys information about depth while shading blur (red arrows) conveys information about 3D structure. (b) Penumbral blur is associated with cast shadows (blue arrows); during grouping, cast shadows do not interfere with perception of physical object boundaries and shading. (c-e) Stimulus set used to assess tuning for shape and blur in V4 neurons. (c) A standard set of 51 shapes were used to assess shape selectivity of V4 neurons. Stimulus size is defined relative to the diameter of the large circle (black arrow). (d) Each shape was presented at up to 8 unique orientations at 45° increments; all rotations for one example shape are shown. For shapes with radial symmetry, duplicates were excluded. (e) To assess tuning for blur, a subset of preferred and non-preferred shapes were presented at up to 9 levels of Gaussian blur (see Materials and methods: Visual stimulation). Example stimuli $\beta \geq 0.32$ were cropped here for display purposes.

from 15-45 spk/s for a variety of sharp stimuli subjected to minimal blur ($\beta = 0.005$). This response pattern was maintained for small amounts of blur ($\beta < 0.04$), but for intermediate and high blur factors ($\beta \geq 0.04$) responses to preferred stimuli gradually declined; at the highest levels of blur tested, *i.e.* $\beta = 0.64$, all stimuli were effectively amorphous with little discernible form (see Fig. 3.1e), and responses approach baseline (dashed line). Thus, for this neuron both response magnitude and shape selectivity declined with increasing levels of blur. Fig. 3.2b,c illustrate additional examples of this general behavior, but rather than a gradual decline as in Fig. 3.2a, cells a08 and b29 of Fig. 3.2b,c maintained shape selectivity up to blur factor $\beta = 0.16$ before transitioning sharply to a baseline level that is not selective

for shape. Note that limited shape selectivity may occur at high blur ($\beta \geq 0.32$) since all stimuli retain low SF oriented energy as blur magnitude is increased.

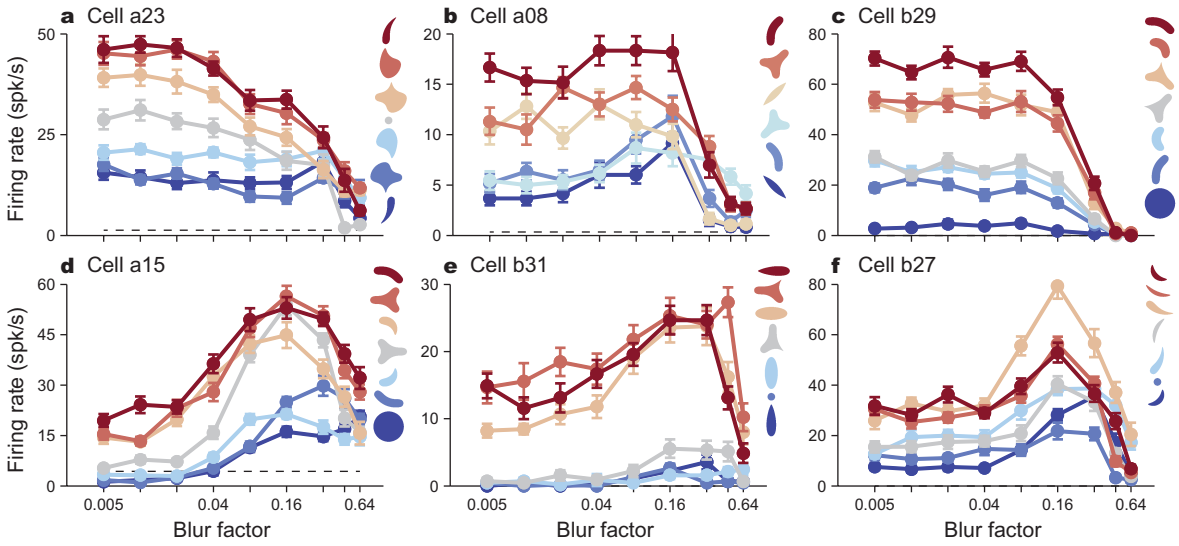

Figure 3.2: **Shape-selective V4 neurons are tuned for blur.** (a-f) For each neuron we plot the mean responses (y-axis) to several stimuli as a function of the magnitude of blur factor (x-axis, β). Line color indicates shape identity and is ordered from preferred (red) to non-preferred (blue) stimuli for each neuron based on responses to the sharp versions of each stimulus ($\beta = 0.005$). Error bars indicate standard error of the mean. (a) Responses of an example V4 neuron that was strongly selective to sharp stimuli, *i.e.* $\beta = 0.005$; responses declined gradually to baseline levels (dashed line) as blur magnitude was increased. (b,c) Two additional examples that also exhibited a monotonic decrease in responses with increasing blur. Unlike (a), these neurons maintained their response level across low blur levels, sharply declining to baseline beyond a critical blur factor ($\beta \approx 0.16$). (d-f) Example V4 neurons that respond best at intermediate levels of blur; responses for preferred stimuli dramatically increase for intermediate blur factors.

In striking contrast, many other V4 neurons exhibited a marked increase in response magnitude over intermediate blur levels (Fig. 3.2d-f). In other words, the activity of these three example neurons was non-monotonic as a function of blur. Further, this blur modulation appears dependent on stimulus shape, facilitating responses of preferred shapes more than non-preferred shapes. As a result, in these neurons shape selectivity was strongest at intermediate blur factors. For all three examples, responses to blurred stimuli are strongest

for blur values between $\beta = 0.16$ and $\beta = 0.32$, overlapping with the range of blur values associated with a decline in responses seen in Fig. 3.2a-c.

To quantify the effect of blur across the population of V4 neurons, we performed a model-free analysis of blur modulation illustrated in Fig. 3.3. For each neuron we first constructed an average tuning curve as a function of boundary blur, based on the interpolated responses to a subset of preferred stimuli (see Materials and methods: Analysis and model fitting). We then calculated two metrics from each tuning curve: the extremal blur factor (Fig. 3.3a-d, triangles) that is associated with the maximal response modulation relative to average non-blurred activity, and a modulation index taken as the tuning curve integral across blur factors (Fig. 3.3a-d, hatching). Fig. 3.3e depicts the modulation index as a function of extremal blur factor for all neurons in our population ($n = 65$); in this space our recorded data spans a continuum. Immediately visible is a sub-population ($n = 42, \approx 65\%$) with a high extremal blur factor ($\beta > 0.32$) and negative modulation index; these neurons exhibit responses that decrease with increasing blur, collapsing to a near-baseline response at highest blur levels (*e.g.* cell a19 of Fig. 3.3d). Conversely, other neurons ($n = 11, \approx 17\%$) exhibit a non-negative modulation index with intermediate extremal blur factor values ($0.1 < \beta < 0.32$), indicative of peak responses at intermediate blur values (*e.g.* cell a15 of Fig. 3.3a). Note that a few cells ($n = 5, \approx 8\%$) have weak tuning for intermediate blur coupled with a strong fall-off at high blur values to produce a positive modulation at high extremal blur factors (*e.g.* cell b13 of Fig. 3.3b). Interestingly, some cells ($n = 7, \approx 10\%$) demonstrate intermediate inhibition, *i.e.* negative modulation at intermediate blur values, demonstrating strong shape selectivity at both low and high blur magnitudes. In Fig. 3.3e the first principal component of our population in this space, calculated under a scaling to equalize variance along each dimension (shaded line), aids in segregating our neurons: cells with a positive principal value (PV, more red) respond best to intermediate blur, while those with a negative PV (more blue) show declining responses with increasing blur. While it is not perfect, we see in Fig. 3.3f that by superimposing blur tuning curves across the population, colored according to each neuron's PV, this measure is diagnostic of selectivity for low blur (more blue) and

intermediate blur (more red).

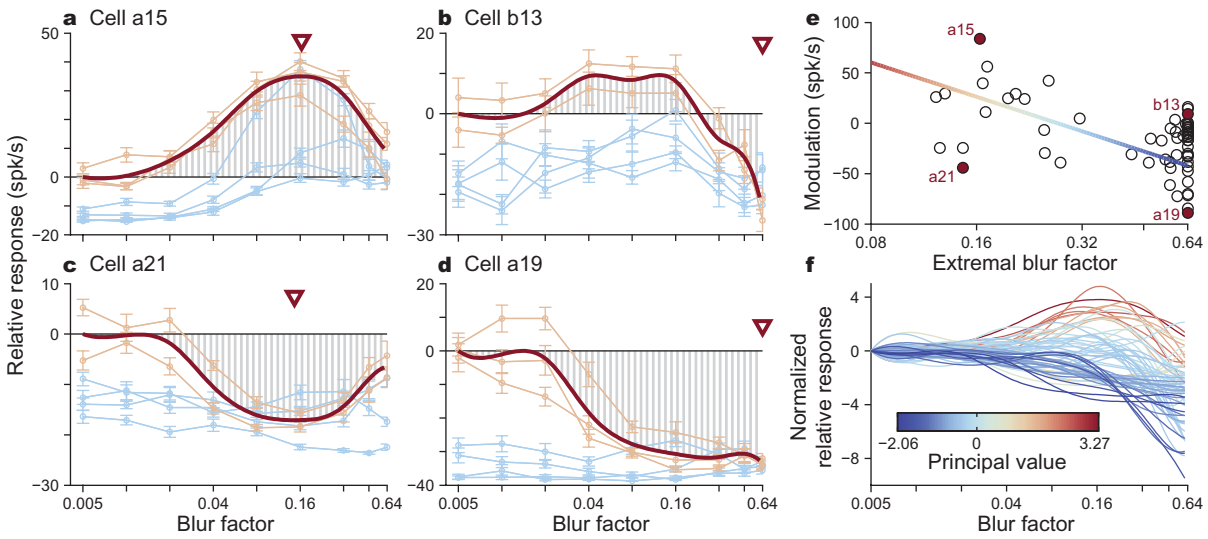

Figure 3.3: **Model-free analysis of blur selectivity across cells.** (a-d) Average blur tuning curves of four example neurons (dark red) constructed by averaging responses to preferred shape stimuli (light red). Relative response (y-axis) as a function of blur factor (x-axis, β) was computed with respect to mean response across sharp preferred stimuli, *i.e.* relative response is zero for lowest blur factors ($\beta = 0.005$). An extremal blur factor (triangle) was defined as the magnitude of blur that evoked the largest absolute deviation relative to responses to sharp stimuli. Response modulation was determined by calculating the integral of relative responses across blur factors (hatching). (e) Response modulation (y-axis) is plotted as a function of extremal blur factor (x-axis) for the population of neurons ($n = 65$) in our data set. The principal value, calculated from the first principal component of the population (shaded line), demarcates neurons with peak responses at intermediate blur values (more red) from those that show declining activity as blur increases (more blue). Example cells of (a-d) are filled and labeled. (f) Superposition of blur tuning curves computed in (a-d), scaled to have a unit-variance of relative response (y-axis) and colored according to (e), demonstrate complementary tuning with respect to blur magnitude (x-axis) within our population.

3.3.2 Controlling for stimulus size

In addition to the gradient width of edge contrast, blurring alters many other stimulus characteristics. For example, as depicted in Fig. 3.4a, the foreground area of a blurred shape

stimulus, defined as the number of pixels distinct from the background, increases with blur magnitude. Thus, tuning for boundary blur might arise from a simple tuning for stimulus size. To test this hypothesis, in a subsequent size control experiment we presented shape stimuli that were first scaled by $\pm 10\%$ and then subjected to a diagnostic subset of blur levels (see Fig. 3.4b). If preference for intermediate blur was simply due to a preference for stimulus size we would expect a shift in the blur tuning peak as size was varied. In Fig. 3.4c,d we plot the responses of two example neurons that respond preferentially to intermediate levels of blur. For both examples the modulation of responses with respect to blur was consistent across changes in stimulus size. Across the neurons subjected to this size control ($n = 26$), in every case we found that blur modulation was similar across stimulus size. Importantly, while one might expect a systematic variation in responses across all stimuli with respect to size, we did not find a significant *interaction* effect between size and blur for any of the neurons (2-way ANOVA, $p > .14$ for all cells, median ≈ 0.96). Rather, blur accounted for a significant fraction of variance ($p < .05$) in the majority of these neurons ($n = 23, \approx 88\%$). As a result, our analysis suggests that selectivity for blur cannot be explained in the context of overall stimulus size.

3.3.3 Controlling for stimulus contrast

A further confound arising from blur is the amount of contrast across a shape's boundary. Under the hypothesis that boundaries are identified from the contiguous region of a certain stimulus intensity, from Fig. 3.1c we note that the average intensity within stimulus foreground, approximated as the set of pixels different from background with respect to an 8-bit stimulus gamut, decreases as blur magnitude increases. That is to say, blur *diffuses* stimulus intensity, reducing the average contrast between figure and ground. Therefore, the preference for an intermediate level of blur could arise from a preference for a specific average stimulus intensity which differs from that of sharp stimuli.

To test this hypothesis, for each blurred stimulus we constructed a non-blurred (sharp) version matched in mean intensity (Fig. 3.5a,b) and compared the responses to these two

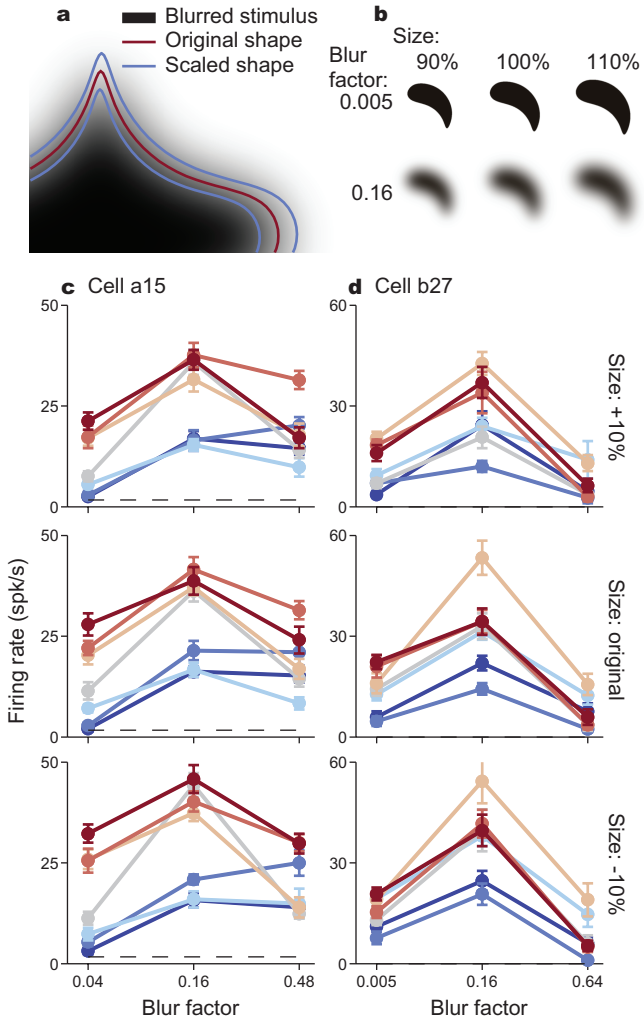

Figure 3.4: **Stimulus size does not explain blur selectivity.** (a) A blurred stimulus ($\beta = 0.16$) has an increased foreground area, defined as the number of pixels distinct from the background, compared to its original boundary prior to blurring (red). Stimuli scaled by $\pm 10\%$ are shown for comparison (blue), and correspond to luminance thresholds approximately 1/3 and 2/3 of maximum, respectively. (b) Example scaled and blurred stimuli used to asses a potential size confound. (c,d) Results of the size control experiment for cells a15 and b27 demonstrate increased responses for intermediate blur irrespective of stimulus size. Line color represents stimulus identity per Fig. 3.2d,f. For both neurons, responses were not significantly influenced by size ($p = 0.45$ and $p = 0.18$, respectively) but significant variance was found with respect to blur ($p < 0.0001$).

stimulus sets. Fig. 3.5c,d plots the results for two example neurons. For cell b26, the tuning curves as a function of blur and contrast are dramatically different: responses are strongest for intermediate blur but fall off as contrast is reduced for sharp stimuli, inconsistent with blur tuning explained by contrast. On the other hand, cell b32 demonstrates very similar selectivity across both stimulus sets, suggesting that blur selectivity in this case *could* be explained by a simple figure-ground contrast preference. To quantitatively compare the two tuning curves we calculated the center-of-mass (CoM) for each: tuning curves that peak at intermediate contrasts will garner a large CoM, and curves which monotonically decrease will retain smaller CoM values. Here, bootstrapping is employed to estimate the distribution of tuning curve CoM measurements under the variance of observed responses (see Materials and methods: Analysis and model fitting). Across our sub-population of neurons subjected to the intensity control (Fig. 3.5e, 34 cells), the majority ($n = 31, \approx 91\%$) had an intensity-matched CoM significantly different than that of blur (t-test, $p < .05$), indicating neural responses are not consistent with tuning for intermediate stimulus intensity. Thus, while selectivity for blur could be attributed in some neurons to a simple tuning for intermediate edge contrast, *e.g.* cell b32, the majority of cells cannot be explained in this context.

3.3.4 Controlling for attenuation of curvature

A more subtle confound of blurred stimuli arises when considering exactly how to define object shape in the presence of blur. Specifically, a blurred shape can be associated with any of a family of closed contours, each defined from the level set of stimulus intensity (see Fig. 3.6b). However, regardless of the intensity threshold, as blur increases, the magnitude of boundary curvature decreases. This is significant from the perspective of shape coding, as previous studies have leveraged the manner in which blurred boundary contours devolve into an ellipse to represent closed contours (Mokhtarian and Mackworth, 1986). Therefore, since the responses of many V4 neurons to shape stimuli can be explained in the context of tuning for boundary curvature (Pasupathy and Connor, 2001), selectivity for intermediate blur may be due to curvature-tuned neurons simply preferring attenuated curvature features

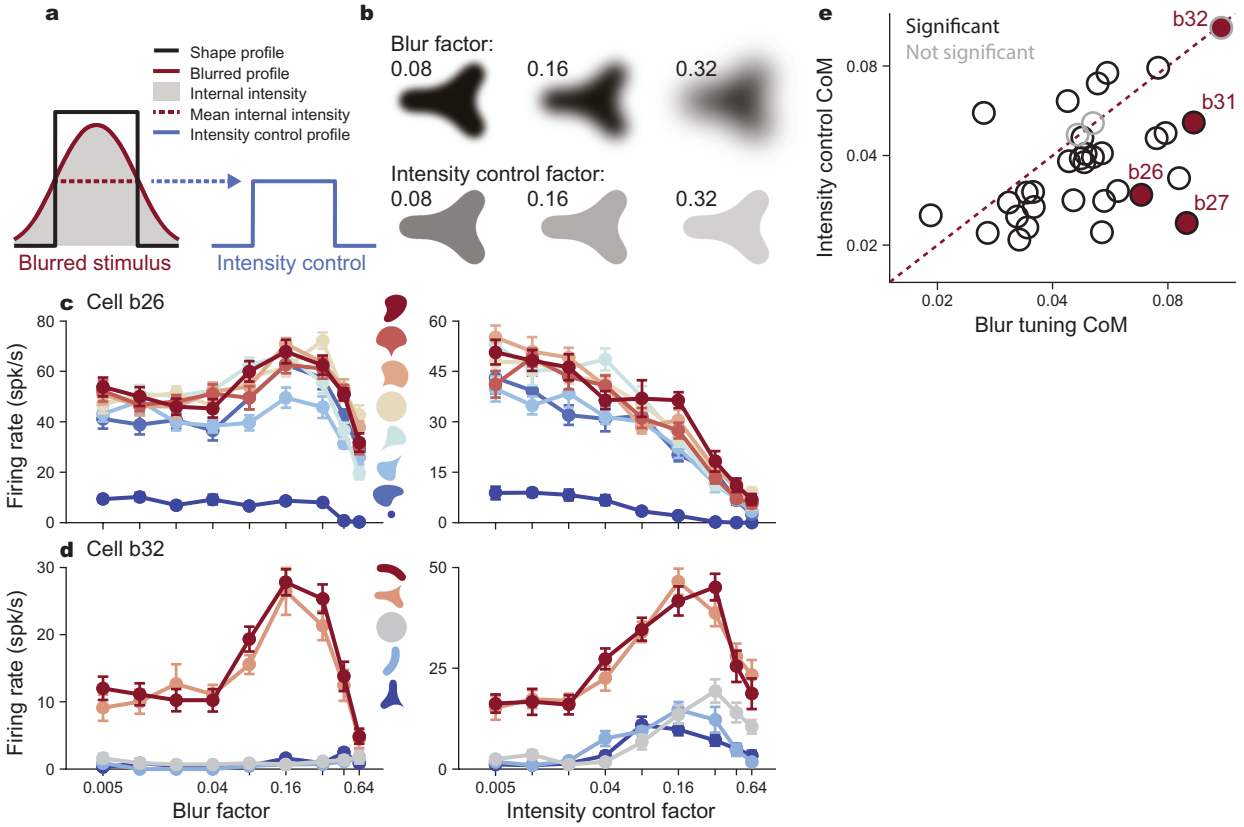

Figure 3.5: **Stimulus contrast does not explain blur selectivity** (a) Profile schematic of how a blurred stimulus (red) has a decreased mean foreground intensity relative to a sharp ($\beta = 0.005$) stimulus (black). An intensity control is constructed from a non-blurred shape with an identical mean foreground intensity (blue). (b) Example blurred stimulus and intensity-matched controls. (c, d) Responses of preferred (red) to non-preferred (blue) shapes that were presented either blurred (left) or as intensity-matched controls (right). While blur and contrast control tuning curves are remarkably different in (c), they are quite similar in (d). (e) Center-of-Mass analysis (see Materials and methods: Analysis and model fitting) reveals that for a majority of cells ($n = 31$ of 34) blur and contrast-control tuning curves have significantly different (black) tuning profiles (t-test, $p < 0.05$); other cells (gray) exhibited blur and contrast-control tuning curves with CoM values that were not significantly different. For these neurons, blur tuning may be explained in the context of intensity tuning. Neurons above the diagonal typically exhibited a tuning preference for intermediate intensities while remaining largely invariant to all but the highest levels of blur; a more conservative estimate that discounts these cells ($n = 6, \approx 18\%$) sill finds the majority of neurons unexplained by tuning for stimulus contrast ($n = 25, \approx 74\%$).

that arise under blur.

To test this hypothesis, we first described each neuron's shape preference in terms of tuning for boundary curvature. As done elsewhere (Oleskiw et al., 2014; Kosai et al., 2014; El-Shamayleh and Pasupathy, 2016) we identified the 2D Gaussian function in a shape space spanned by angular position and curvature (APC) that best predicts responses to the preliminary shape screen conducted using sharp stimuli (see Materials and methods: Visual stimulation). Bootstrapping was used to calculate the normalized mean-squared prediction error (Training NRMSE, Fig. 3.6d), as a measure of goodness of fit (see Materials and methods: Analysis and model fitting). We then evaluated how well this best-fitting APC model could predict responses to blurred stimuli by considering the curvature descriptions associated with each blurred stimulus at a range of intensity thresholds (see Fig. 3.6a for a schematic of this procedure). For each intensity threshold, we quantified goodness of fit as the normalized root mean-squared error (NRMSE) between the predicted and observed responses. Then, for each neuron, the intensity threshold that minimized NRMSE was selected as the exemplar threshold (Threshold Curvature NRMSE, Fig. 3.6c,d). In Fig. 3.6c, for each neuron we compare this threshold NRMSE against the NRMSE of a mean model that is agnostic to blur, *i.e.* a model that predicts identical responses to sharp and blurred versions of the same shapes. For most neurons in our population, model predictions derived from optimized intensity thresholds were associated with larger errors than simple predictions equal to mean responses across blur. Furthermore, Fig. 3.6d demonstrates that the APC model's failure to explain blur response variance is not due to an inability of the model to capture shape selectivity exhibited in our population; we find that, for the majority of neurons, prediction error of responses to blurred stimuli were higher than prediction error of models trained and validated on sharp stimuli alone.

3.3.5 Joint coding of shape and blur

Results thus far demonstrate that neuronal responses in V4 are modulated by boundary blur, and this modulation cannot be explained on the basis of tuning for size, contrast, or

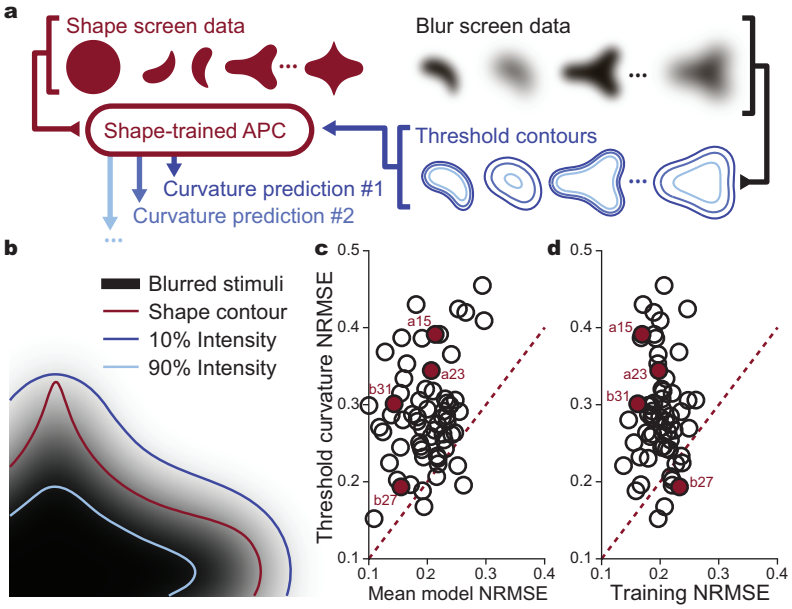

Figure 3.6: **Curvature attenuation does not explain blur selectivity** (a) Schematic of analysis performed across shape and blur datasets to assess the contribution of curvature attenuation toward blur selectivity. An APC model is fit to data collected from shape screening (red), which then predicts responses to curvature-attenuated threshold contours computed from blurred stimuli at different thresholds (blue). (b) A blurred stimulus ($\beta = 0.32$) generated from a shape contour (red) and a family of closed contours defined by the level set of an intensity threshold (blue), each with reduced curvature magnitudes. (c) For each cell, the minimum prediction error across all intensity thresholds (Threshold Curvature NRMSE; see Results) plotted as a function of a blur-invariant mean model's prediction error. The latter predicts responses to different shapes in accordance with the APC model ignoring blur; responses are identical over all blur levels for a given shape. Example cells are filled and labeled. (d) Threshold prediction error as a function of bootstrapped training error, *i.e.* a baseline estimate at how well the APC model predicts shape data.

curvature. Therefore, our findings support the hypothesis of an underlying neural code for object shape and boundary blur, *e.g.* a representation of both boundary conformation and spatial gradient of edge contrast in the rate responses of single V4 neurons.

To rigorously test this hypothesis we evaluate whether a *joint* model of shape and blur performs significantly better at predicting V4 responses than a *marginal* model that is tuned for shape alone. Importantly, this analysis builds on the well-studied APC model (Pasupathy

and Connor, 2001; Oleskiw et al., 2014; Kosai et al., 2014; El-Shamayleh and Pasupathy, 2016) that is known to capture tuning for boundary conformation in shape-selective V4 neurons (see Materials and methods: Analysis and model fitting). For each cell we first fit a standard APC model to blurred responses under the assumption that neurons are invariant to boundary blur of shape stimuli. Thus, the APC model fits to the mean response of each shape across blur factors. We then augment the APC model to include a blur-selective term, taken to be log-normal in blur factor β, that scales shape-selective responses. This angular position, curvature, and blur (APCB) model is fit to activity recorded from blurred shape stimuli. Note that APC and APCB models were fit to blurred data without including the preliminary screen data to ensure that any fit differences were not due to number and diversity of stimuli.

In Fig. 3.7a we plot accuracy as determined by prediction error (NRMSE) across all trained stimuli and responses for shape coding models with (APCB, x-axis) and without (APC, y-axis) the influence of blur. In every case, inclusion of blur information significantly reduced error in most neurons beyond what is expected due to additional parameters (F-test, $p < 0.05$ for $n = 61$ of 65, $\approx 94\%$). Examples in Fig. 3.7b-d illustrate the effectiveness of the APCB model: it accurately captures a range of response behaviors including selectivity for intermediate blur (Fig. 3.7d) and fall-off of responses to baseline at high blur levels (Fig. 3.7b,c).

This choice of model parameterization, wherein blur tuning multiplicatively scales selectivity for boundary curvature, is but one of many possible descriptive models. For example, another formulation could have a selectivity for blur that additively facilitates position and curvature tuning, *i.e.* $APC(\Gamma;\Theta) + B(\Gamma;\Theta)$. In Fig. 3.7e, however, we find that the multiplicative APCB model outperforms the additive variant in the majority of neurons. Thus, we conclude that neurons in V4 jointly encode information of shape and blur, and that this coding is best explained by a gain modulation, with tuning for blur multiplicatively scaling shape selectivity. It is important to note that, by construction, the APCB model is separable, *i.e.* does not rely on the interaction between boundary conformation and blur to accurately

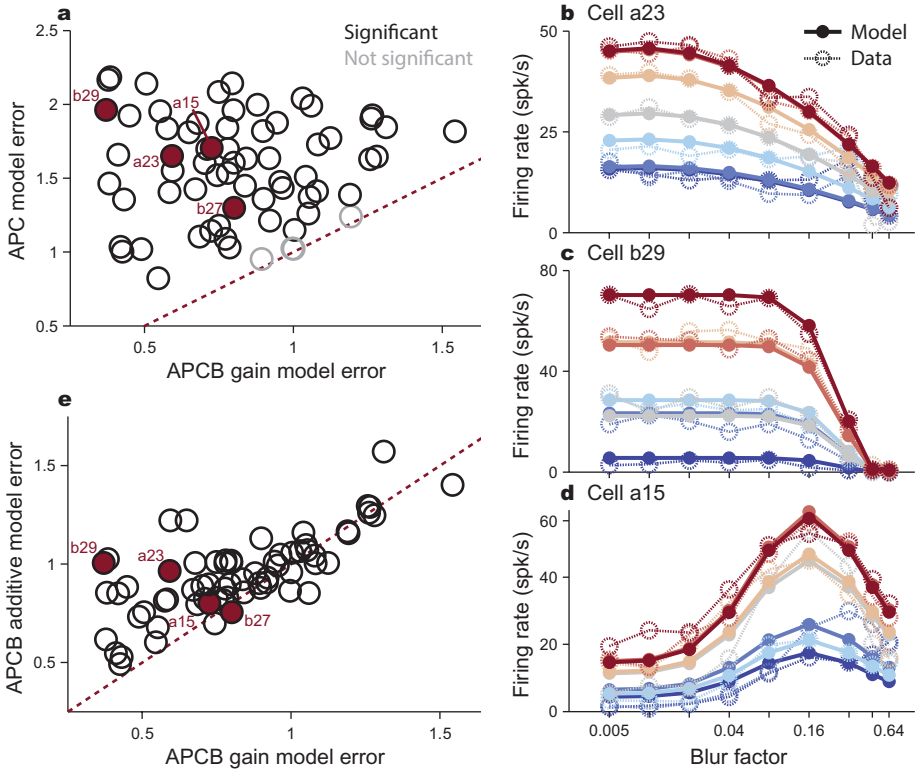

Figure 3.7: **Responses explained by a joint model of shape and blur** (a) NRMS fit error of an APC model plotted as a function of NRMS fit error of an APCB model for each neuron. Even though the APCB is a generalization of the APC model, *i.e.* APCB models form a superset of APC models with two additional parameters, the majority of neurons are fit significantly better by the APCB model, wherein shape-selective responses are scaled as a function of blur (F-test, $p < 0.05$ for $n = 61$ of 65). Example cells are filled and labeled. (b-d) Observed responses (open circles, dashed lines) and APCB model fits (filled circles, solid lines) for 3 example neurons (see Fig. 3.2a,c,d). Qualitative assessment of fits suggests that the APCB model captures blur-tuned response properties remarkably well. (e) Comparison of the APCB Gain model against an APCB Additive variant with equal degrees of freedom (see Materials and methods: Analysis and model fitting). Most neurons are better fit by the APCB Gain model, particularly when error is small, consistent with blur selectivity being explained by gain modulation.

predict neural response. As will now be addressed, this suggests a distinct neural mechanism regulating blur selectivity in V4.

3.3.6 Distinct dynamic properties of blur-selective responses

Finally, we ask how and when tuning for blur emerges in the the responses of individual neurons. In Fig. 3.8a we illustrate, for the preferred shape of an example neuron (cell a23), the peristimulus time histograms (PSTH) as blur is varied. For this neuron responses to blurred shape stimuli decreased with increasing blur (see Fig. 3.2a) and this decrease was uniform across the stimulus presentation interval. Given that shape and blur tuning is best explained by a modulation of gain, in Fig. 3.8c we quantify blur modulation by plotting the standard deviation of responses across blur factors for both preferred (black) and non-preferred (gray) shape PSTHs. The difference between these curves (hatching) then captures the timecourse over which blur modulation is applied to shape-selective activity. The utility of this analysis is seen when considering a blur-selective neuron (cell a15) in Fig. 3.8b. Here, the PSTH of preferred shape stimuli demonstrates a nonuniform increase in responses, most pronounced over the initial (transient) wave of activity (approximately $50 - 150$ ms after stimulus onset). This effect is captured by the transient response modulation seen in Fig. 3.8d. Thus, these example neurons suggest a potential difference in dynamics between two groups of cells: those that are selective for intermediate blur factors, and those that are not. If we demarcate neurons in our population with a blur-selective principal value (see Fig. 3.3e,f, shading) PV > 0.25 as tuned for intermediate blur from those tuned for sharp (minimal blur) stimuli, the average normalized PSTHs for intermediate- and sharp-selective sub-populations (Fig. 3.8e,f) exhibit the distinct qualitative differences observed in cells a23 (Fig. 3.8a) and a15 (Fig. 3.8b). For the blur selective sub-population (Fig. 3.8f) differences between responses with respect to blur are transient, restricted to the early time period, while such differences exhibited by the sharp-selective sub-popluation (Fig. 3.8e) are uniform across the stimulus presentation period.

To quantify these differences we first plot in Fig. 3.8g blur-selective PV versus the difference of blur modulation between preferred and non-preferred shape stimuli for each cell (Fig. 3.8c,d, hatching) over the sustained period of activity (shaded, $200 - 300$ ms). We note

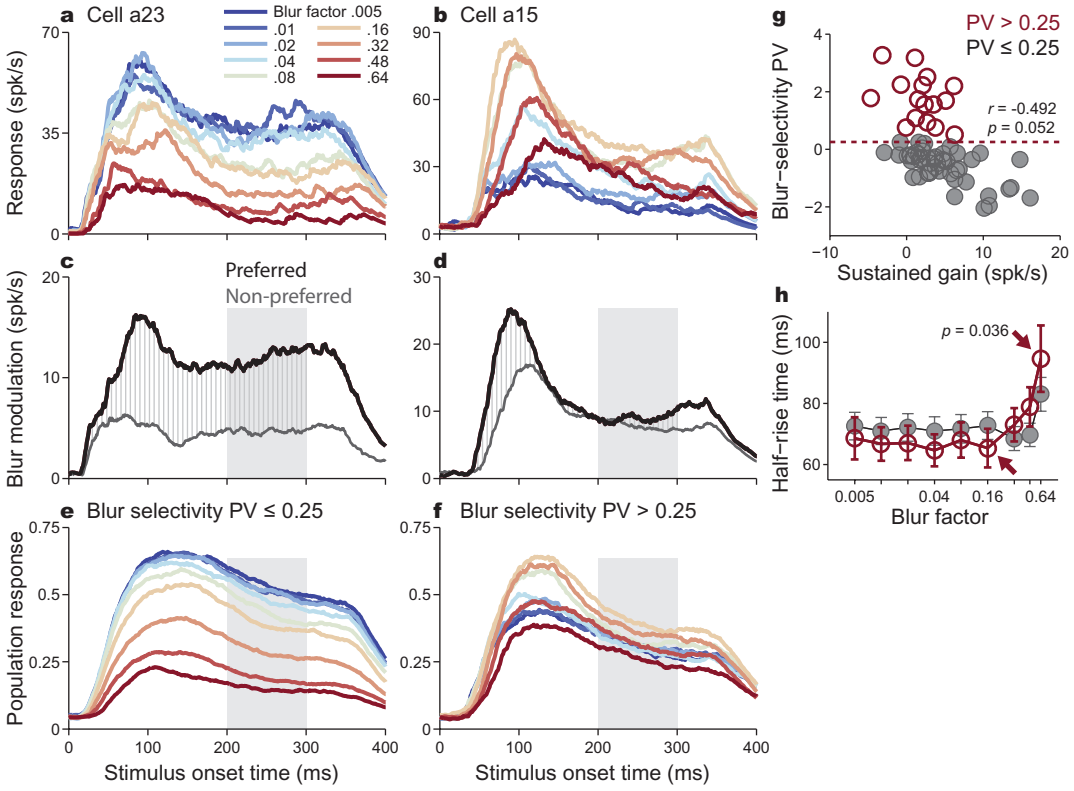

Figure 3.8: **Response dynamics differ with respect to blur selectivity.** (a,b) PSTH of preferred stimuli for low (blue) to high (red) blur factors for a sharp selective and a blur selective cell, respectively. (a) For the non-blur selective cell, increasing blur magnitude decreases responses throughout stimulus presentation. (b) For the blur selective cell, blur-dependent modulation is transient. (c,d) Blur modulation (y-axis), calculated as the standard deviation of responses with respect to blur, for preferred and non-preferred shape stimuli for example cells in (a) and (b) respectively. The difference in the timecourse (hatching) for preferred and non-preferred stimuli estimates the time at which blur scales shape-selective responses. (e,f) Average normalized PSTHs of non blur-selective and blur-selective cells, respectively, demonstrating qualitatively distinct response dynamics between sub-populations. (g) The principle value of blur selectivity (see Fig. 3.2e) plotted as a function of the average difference in blur modulation between preferred and non-preferred shape stimuli across the sustained period (200-300 ms; shaded region in e-f). (h) For each blur factor, mean latency to half-of-maximum response (half-rise time) across sub-populations. Error bars denote standard error of the mean. A significant increase in response latency (t-test, $p = 0.036$) occurs across cells selective for intermediate blur (PV > 0.25) between blur factors 0.16 and 0.64 (arrows).

a significant anti-correlation between these quantities (Pearson's $r = -0.492$, $p = .052$); for cells with a larger PV, sustained modulation of gain is smaller. There was also a significant difference in sustained blur modulation between the intermediate- and sharp-selective sub-populations (t-test, $p = 0.018$). Further, we found no significant difference in shape-dependent blur modulation during the initial period of activation between the two sub-populations (t-test, $p = 0.154$), suggesting that the more selective for intermediate blur a cell is (high blur-selective PV), the more transient blur modulation appears to be (low average sustained blur modulation).

While the response latency of cell a23 was unaffected by blur, cell a15 exhibited a consistent shift in tuning peaks: responses at high blur are significantly delayed. To examine this effect of response latency across the population for each neuron we quantified the half-rise time as the duration between stimulus onset and half-of-maximum response at each blur factor. At low blur factors ($\beta \leq 0.16$), blur-selective cells tend to have a slightly (though not significant) shorter mean half-rise time on average (Fig. 3.8h). However, half-rise time increased significantly (t-test, $p = 0.036$) for blur-selective cells as the magnitude of blur increased (arrows). To factor out increased response latency as a population artifact, we analyzed latency differences based on the ratio of half-rise times as a function of blur factor, *i.e.* response latency normalized by each neuron's baseline half-rise time to sharp stimuli ($\beta = 0.005$), and find the same trend to hold (not shown). This suggests the underlying circuitry of V4 neurons tuned for intermediate blur is distinct from those which consistently reduce activity as blur increases; blur-selective neurons receive a transient multiplicative facilitation to shape-selective response *as high SF content is removed.*

3.4 Discussion

We studied the responses of primate V4 neurons to determine if, and to what extent, blur influences neuronal activity. We found that many V4 neurons are jointly tuned to shape and blur, with responses explained by a blur-dependent gain-modulation of shape tuning. Importantly, our results are not a simple byproduct of changes in stimulus size, contrast, and

boundary curvature that co-vary with blur. This demonstration of blur tuning implicates a role for V4 in processes cued by blurred boundaries, *i.e.* segmentation, depth perception, and shading, and supports the hypothesis that V4 can provide an explicit and sufficient representation of natural scenes.

Our results identify blur as a novel tuning dimension in visual cortex; while some V4 neurons exhibit a monotonic decline in shape-selective responses with increasing levels of blur, others show maximal responses at intermediate blur levels. Specifically, this latter group of neurons respond best when high SF content is removed from a stimulus while preserving lower-band content. Thus, this effect is not explained by simple mid-band SF tuning, but rather indicates a preference for intermediate blur that requires high SF information to suppress shape-selective responses. In V4, intermediate blur tuning is associated with response peaks between blur factors $\beta = 0.08$ and 0.16. At 3° eccentricity, for example, this corresponds to a response enhancement to frequency content between 2.6 cyc/° and 1.2 cyc/°, respectively, which is consistent with SF tuning distributions observed in macaque V1 (Foster et al., 1985). Given that V4 responses are explained by a joint model of shape and blur, where shape-selective responses are multiplicatively scaled as a function of boundary blur, blur tuning in V4 may arise from the aggregation of SF information reported by V1, consistent with previous demonstrations of V4 selectivity for non-Cartesian gratings (Gallant et al., 1996) and illumination vectors (Hanazawa and Komatsu, 2001). Furthermore, tuned responses to shape occur at either low or intermediate levels of blur in each neuron, indicating that blur-tolerant shape identity may be decoded from a V4 population response. This is significant toward visual computation in natural environments, as defocus, due to a finite depth of field or improper accommodation, may introduce optical (Gaussian) blur within a scene; it has been argued that the visual system responds very differently to artificial versus naturalistic stimuli presented under Gaussian blur at various depths from the plane of focus Burge and Geisler (2014); Sebastian et al. (2015). Thus, V4 may play a critical role in not only detecting shapes subjected to blur for judgments of object depth, but for segmenting naturalistic scenes into blur-invariant object representations.

Blur is a critical cue for the perception of shadows. Shape theorists and psychophysicists have long argued that attached and cast shadows, formed on the occluding object or another surface, respectively, contribute to the perception of 3D shape and scene understanding. For example, the relative position of shadows may be used to infer the relative location of scene illuminants (Attwood et al., 2001; Ramachandran, 1988; Sun and Perona, 1996; Allen, 1999; Hubona et al., 2004). However, even though dark and blurry boundaries may be quickly identified as shadows (Elder et al., 2004), perceptual judgments on these shadows are difficult and slow (Rensink and Cavanagh, 2004; Porter et al., 2010). While some have argued that the poor access to shadow-specific information is due to low-level shadow detection and discounting (Rensink and Cavanagh, 2004), others have proposed a higher-level process related to the perceptual segregation of physical objects from nuisance factors related to illumination (Porter et al., 2010). Our results identify V4 as a plausible locus for processing shadows within the ventral stream, where shape and blur information coalesce in the activity of individual neurons. While the early emergence of blur-selective responses in V4 could underlie the rapid detection of shadows, more experiments are needed to determine whether V4 differentially encodes shadow versus non-shadow blurred boundaries, and how this difference could underlie the limited salience of shadows during perceptual judgments. An adaptive stimulus presentation protocol to first identify shape- and blur-selective V4 neurons followed by an investigation of how these neurons respond to naturalistic images will give insight into how neurons selective for blur and/or shape participate in the perceptual organization of natural scenes.

We note, however, that the descriptive model presented here does little to explain how computations encoding shape and blur arise *in vivo*. In particular, further modeling studies are required to determine how V4 selectivity for shape and blur could be constructed from upstream populations. Fortunately, the dynamics of blur selectivity reported here provide key insights into potential underlying mechanisms. Our results demonstrate that preference for intermediate levels of blur arise early, approximately 60 to 100 ms after stimulus onset, comparable to the time at which shape selectivity arises in V4 (Bushnell et al., 2011b). Fur-

thermore, this activity is transient, lasting until approximately 150 to 200 ms after stimulus onset. One candidate blur-selective circuit consists of a simple difference of SF power within V1, similar to Spectral Receptive Fields (SRF; David et al., 2006) and consistent with contrast energy models of blur discrimination (Watson and Ahumada, 2011). This blur signal could then bypass V2 to apply a fast reduction in gain of shape-selective V4 units. Recurrent inhibition, either within V4 or between V4 and previous areas, would then suppress the contribution of blur over sustained periods. Alternatively, preference for intermediate blur could arise from latent normalization of V1 activity, including faster magnocellular inputs, as high SF responses are removed. It is unknown, however, if such a normalization-based circuit architecture would reproduce these observed dynamics.

Computational studies have often argued that ideal representations within the earliest stages of visual processing are general-purpose codes, supporting a diversity of tasks, from which perception can emerge (Elder, 1999). Consistent with this argument, V1 receptive fields, tuned for local orientation (Hubel and Wiesel, 1959, 1965, 1968) and spatial frequency Movshon et al. (1978); Albrecht et al. (1980), form a wavelet-like representation of visual space Olshausen and Field (1997). Further, it has been shown that natural scenes can be efficiently decomposed into scale- and space- localized Gabor-like bases which are selective for orientation, remarkably similar to the receptive fields of V1 neurons (Bell and Sejnowski, 1997; Zylberberg et al., 2011). Local populations of V1 units therefore form a complete and efficient neural representation of naturalistic scenes Olshausen and Field (1996). Beyond V1, however, sensory representations of higher visual areas are thought to instead participate in solving specific visual tasks. For example, face-selective neurons in inferotemporal (IT) cortex may facilitate face recognition (Tsao et al., 2003), and border-ownership signals in V2 may underlie figure-ground organization (Zhou et al., 2000). Thus, rather than a general-purpose code, representations within each module beyond V1 appear to reflect the computations required to solve well-defined problems. Previous studies of V4 have shown that may neurons explicitly encode the curvature of object boundary fragments, thought to provide a structural code for complex object shape Kourtzi and Connor (2011). Our demonstration of tuning for

both shape and blur is especially significant since such an encoding framework may provide a sufficient representation of naturalistic scenes Elder (1999). Therefore, in addition to supporting a neural code for object recognition, V4 may also efficiently encode visual scenes for use in higher visual areas, *e.g.* IT. Further, while the representations of V1 and V4 may both be complete, single-unit V4 activity, unlike V1, includes an explicit code of object-centric boundary conformation. This interpretation is consistent with V1 encoding "stuff" and V4 building an intermediate annotation of "things", both of which are likely prominent in higher ventral computations (Adelson and Bergen, 1991). However, the notion of shape and blur underlying a sufficient representation of natural images does not imply these features to be the sole dimensions of selectivity within V4; while higher ventral visual areas like IT may in fact decode the entirety of an image from V4, an over-complete representation incorporating additional visual features may further benefit complex visual tasks such as object identification or scene categorization.

Chapter 4

DIVISIVE INHIBITION VIA INPUT TIMESCALE

Despite much evidence that divisive inhibition is important for regulating response amplitude within neural systems, we do not yet have a full understanding of the range of circuit mechanisms that can achieve this canonical computation. The key operation is modulating the gain of the output of a neuron, *i.e.* the multiplicative scaling of the average firing rate relative to a measure of the mean or tonic level of input. While it is well established that the variance of the input can impact the gain of neuronal output (Doiron et al., 2001; Mitchell and Silver, 2003), here we demonstrate that input *timescale*, as reflected by the input's autocovariance structure, can also have a significant effect on gain over a wide range of firing rates. Further, we provide an analytic solution for the firing rate of a conductance-based leaky integrate-and-fire (LIF) neuron in which both the membrane voltage and input timescales are explicitly parameterized. We find that divisive inhibition occurs when input timescale is increased over a biophysically plausible range of parameters and we illustrate how timescale can play a role in the normalization of responses in spiking networks. We conclude by demonstrating how correlated activity between afferent populations can serve as a means to shape input timescale and thereby affect circuit-level computation.

4.1 Introduction to divisive inhibition in spiking neurons

Information processing within neural networks is widely considered to be achieved by circuit computations in which the number of action potentials per second (firing rate), serves as the fundamental variable (Barlow, 1972). Therefore, by understanding how basic mathematical operations like addition and multiplication are applied to firing rates in networks, we may gain insight into fundamental mechanisms of neural computation (Carandini and Heeger,

2011; Yu et al., 2002; Angelucci and Bressloff, 2006; Sato et al., 2014). While addition of firing rates is easy to accomplish, *e.g.* via pooling of excitatory postsynaptic potentials, it is not clear how or to what extent multiplication or division of firing rates can be implemented in biophysically-realistic spiking networks (Yu et al., 2002; Sato et al., 2016; Somers et al., 1998; Rubin et al., 2015; Shushruth et al., 2012). Nevertheless, divisive inhibition of a neuron's firing rate enters into many circuit-level computations, including gain control and normalization (Carandini and Heeger, 2011).

In particular, given an input $\vec{x}$ representing the activity of multiple neurons, for divisive inhibition we seek to compute a normalized output $\vec{y}$,

$$\vec{y} = \frac{\vec{x}}{\phi + \|\vec{x}\|_1}, \tag{4.1}$$

with ϕ being some small constant to prevent division by zero (Heeger, 1992). From (4.1), the two requisite operations are addition, for computing the norm of $\vec{x}$, *i.e.*

$$\|\vec{x}\|_1 = \sum_i |x_i|, \tag{4.2}$$

and division. The neural implementation of the former is simple: pooling of excitatory activity, up to limits of rectification and saturation, effectively sums activity. The latter operation of division, however, is not so straightforward. Formally, suppose $f(I)$ represents an arbitrary f-I curve, *i.e.* firing rate versus input strength for some current I. To achieve divisive inhibition we seek to modify the output rate $\tilde{f}$ such that the response curve is *scaled* relative to net strength of input, *e.g.* $\tilde{f}(I) = \alpha f(I)$ where $0 < \alpha < 1$. Such a scaling naturally produces a change in the slope of the response curve — a modulation of gain — for fixed inputs. This is opposed to subtractive inhibition, in which output is *shifted* relative to input, *e.g.* $\tilde{f}(I) = f(I) - \alpha$ where $\alpha > 0$. While inhibition in a biophysical context is likely to involve of both scaling and shifting, we focus our investigation on dividing a neuron's spiking output with respect to net input strength. Therefore, we aim to understand mechanisms that produce a downward modulation in gain in conjunction with a net suppression of responses in the firing rate of single neurons.

Interestingly, much work has sought to understand gain modulation for purposes of multiplicative neural computation. Early investigation of single neurons suggested that shunting inhibition, the gating of excitation by inhibitory conductance with a reversal potential at resting equilibrium, could be an important mechanism of multiplication (Torre and Poggio, 1978; Koch et al., 1983). In-depth studies have revealed that while shunting divisively scales sub-threshold excitatory postsynaptic potentials, the cumulative effect on neuronal firing rate is largely subtractive (Gabbiani et al., 1994; Holt and Koch, 1997). However, when applied with dynamic noisy input, tonic shunting inhibition can divide a neuron's output response, at least over low firing rates (Mitchell and Silver, 2003; Doiron et al., 2001). Specifically, the variance of noisy input has been shown to affect the strength of division produced by tonic inhibition, both *in vitro* from direct measurements of cortical neurons via dynamic clamp, and in simulation of conductance-based LIF models in biophysically-realistic parameter regimes (Chance et al., 2002; Ayaz and Chance, 2009).

Beyond the mean and variance, another key property of stochastic inputs is the *timescale* of input fluctuations, which can be modulated by factors such as the spike timing and correlation of upstream activity (Moreno et al., 2002) or by the kinetics of synaptic filtering (O'Donnell and van Rossum, 2014). Input timescale has been shown in several studies to impact the firing rates of model neurons (Mitchell and Silver, 2003; Brunel and Latham, 2003; Moreno et al., 2002) as well as the gain and phase of their frequency response (Brunel et al., 2001; Ly and Doiron, 2009; Richardson, 2004). However, gain is typically quantified through a local linearization of responses, or at isolated locations in the space of model parameters. Thus, findings are not seen to generalize to a range of plausible output rates, nor are shown to be consistent across varying levels of excitatory input. Interestingly, some have reported timescale-based gain modulation to not always occurl response gain can also be quantitatively insensitive to the timescale of the driving input in certain parameter regimes (Ly and Doiron, 2009) or when voltage distributions are held constant (Alijani and Richardson, 2011; O'Donnell and van Rossum, 2014). Given this complex array of findings, we aim to develop a unifying and mechanistic understanding of the effect of input timescale on a neuron's firing

rate in a range of biophysically-plausible models and to demonstrate how timescale-based divisive inhibition can arise in spiking neural circuits.

4.2 Materials and methods

Simulations of all spiking LIF models were performed in MATLAB R2013b. Dynamics were evaluated numerically with the forward Euler method at a time step of 10 microseconds. After crossing threshold, a spike was recorded and membrane voltage was forced to reset potential instantaneously. Spike-rate response was determined from the mean spike frequency over a 100 second stimulation duration. Further mathematical analysis is given as supporting information. Code is available upon request.

4.3 Results

In this study we first explore the contribution of input statistics to the rate response of a very simple current-based leaky integrator neural model, for which we use simulation and analysis to observe that both input variance and timescale facilitate gain modulation across a wide range of output rates and parameter regimes. We then develop an analytic method to study the more realistic case of conductance-based input and show timescale to be a general mechanism for divisive inhibition across a range of neuron models. To apply our findings toward biophysically-realistic spiking networks, we study divisive inhibition though simulation of balanced input regimes, using our analytic methods to explain a range of observations. We conclude our study with the development of a novel circuit that shapes timescale by correlating input activity to achieve divisive inhibition in a biophysically-plausible conductance-based spiking network.

4.3.1 Input variance and timescale affect gain modulation in a leaky integrator

We begin with a simple demonstration of how input statistics of variance and timescale affect the gain of firing rates in response to tonic current injection. As shown previously, the variance of background noise, specifically inhibitory and excitatory input fluctuations

acting on a neuron, can modulate gain by divisively inhibiting the response to a tonic input (Chance et al., 2002). The relationship between variance and gain is demonstrated by a current-based LIF neuron, *i.e.*

$$C\frac{\partial}{\partial t}V = -V + I(t), \tag{4.3}$$

for some input current $I(t)$ and constant membrane capacitance C. When the membrane exceeds some voltage threshold, a spike is recorded and voltage is assigned to a reset potential. For our purposes we parameterize input current as $I(t;\mu,\sigma,\tau)$, where tonic drive μ is separated from stochastic background fluctuations of mean zero, variance σ^2, and timescale τ (Fig. 4.1a). Therefore, I is defined by the autocorrelation

$$\mathbb{E}[I(t)I(t+\delta)] - \mu^2 = \sigma^2 e^{-|\delta|/\tau}, \tag{4.4}$$

and neural response quantified by observing the spike rate as a function of tonic drive μ (Chance et al., 2002).

We observe from Fig. 4.1b,c that by increasing the standard deviation of background noise, the rate response curve is both shifted (the onset of spiking occurs at lower levels of mean current) and scaled (the rate of spiking increases more gradually as mean current is increased). This is consistent with the well-established effect of input variance toward gain modulation (Mitchell and Silver, 2003; Ayaz and Chance, 2009). Crucially, however, we find in Fig. 4.1d,e that a qualitatively similar shift in onset and scaling of response gain also occurs as the *timescale* of input fluctuations increases, even while holding input variance fixed.

Above, we simulated the separate effects of input variance and timescale on the gain of neural responses by varying each parameter independently. However, in a more natural setting, where input to a neuron arrives in the form of discrete synaptic events, these parameters change together. We consider one way in which this input timescale and variance are coupled under the constraint that the mean conductance delivered by synaptic events stays constant as synaptic timescales vary. Specifically, consider a unitary synaptic input that delivers an initial amount of conductance α that decays to zero over time constant τ, *e.g.* $\alpha\exp(-t/\tau)$

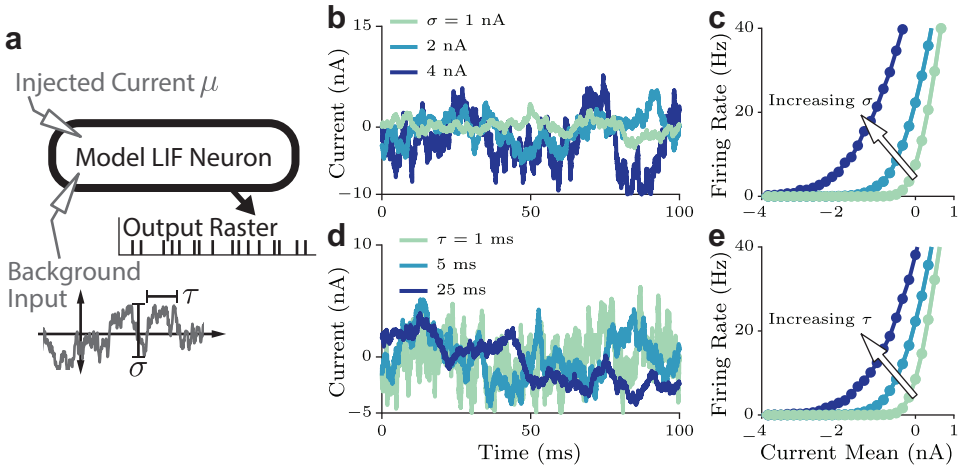

Figure 4.1: **The statistics of input currents affect spike-rate response gain.** (a) schematic of background fluctuations altering the gain of a current-based LIF neuron's spike-rate response to tonic current injection. Example traces of background current as (b) variance or (d) timescale is increased. Spike-rate response of a model LIF neuron as a function of tonic input drive as (c) variance or (e) timescale is increased, demonstrating a decrease in gain.

for $t \geq 0$. As the synaptic timescale is increased, so too is the total amount of conductance delivered over that timescale; if synaptic inputs then arrive at a constant rate R, such as in the case of Poisson shot noise, then the mean input is proportional to $R\alpha\tau$. To hold this mean input fixed, and without directly scaling input rate, we scale the magnitude of synaptic inputs so that $\tilde{\alpha} = \alpha/\tau$. This couples input timescale to variance, which is then proportional to $R\alpha^2/\tau$. Thus, as timescale is increased, variance is decreased so as to preserve mean input. A simulation of current traces in this scenario, and the spike-rate responses of a LIF neuron receiving these current traces, is given in Fig. 4.2. While in Fig. 4.1c,e we find that timescale and variance can each independently affect gain, in Fig. 4.2b we see that coupling these parameters, so that they change in opposite directions to hold mean input constant, has a net suppressive effect on rate response curves, both scaling and shifting spiking output as a function of mean input to achieve divisive inhibition. Note that here we are agnostic to how timescale or synaptic magnitude could realistically be manipulated in neural circuits;

below we will revisit the coupling of input variance and timescale when examining divisive inhibition in a conductance-based LIF neuron receiving excitatory and inhibitory spiking input.

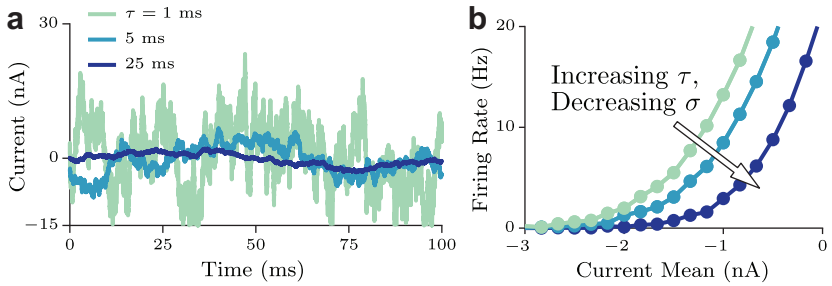

Figure 4.2: **Coupling the timescale and variance of background inputs achieves divisive inhibition.** (a) When the autocorrelation timescale is coupled to variance in a current-based LIF neuron, *i.e.* $\sigma^2 \propto 1/\tau$, (b) increasing timescale has a divisive effect on firing rates: both rate response slope and onset of spiking are decreased. This suggests timescale as a plausible mechanism for divisive inhibition in the setting of input modeled as discrete synaptic events, *e.g.* shot noise.

With Fig. 4.1 and Fig. 4.2 we have shown contributions of input variance σ and timescale τ to the modulation of rate response gain for specific parameter values. To more completely understand the relationship between these input statistics and gain, we simulate the rate response curve as a function of mean current μ across the space of σ and τ. As illustrated in Fig. 4.3a, we estimate gain with the slope of the piecewise-linear function $\lfloor m\mu + b \rfloor$, where $\lfloor \cdot \rfloor$ denotes half-wave rectification, fit via least squares to the rate response between 0 and 80 Hz. We observe in Fig. 4.3b that the rate response slope decreases with σ, in particular when the synaptic timescale is sufficiently large. This is consistent with findings of divisive inhibition occurring from an increase in conductance input variance during balanced synaptic activity (Ayaz and Chance, 2009). Fig. 4.3b also shows a modulation of gain with respect to timescale; for fixed current variance, the response slope is lowered for increasing input timescale, demonstrating that input autocorrelation independently contributes toward scaling of response curves, as suggested above. Additionally, the results observed in Fig. 4.3

are qualitatively identical to slope estimates constructed from a piecewise-power function (see Fig. A.1).

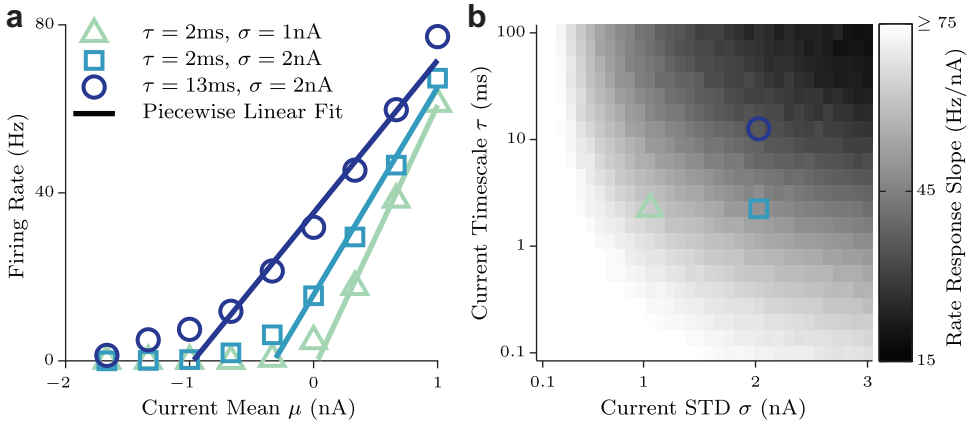

Figure 4.3: **Input variance and autocorrelation timescale shape response gain.** (a) Rate response of a current-based LIF neuron for multiple levels of stochastic background current variance and timescale, along with piecewise linear fit to assess gain of this response. (b) We find that the gain of a neuron's response to tonic input decreases as either current variance or timescale is increased.

Having shown that timescale affects gain in simulations of current-based LIF models, we desire a mechanistic explanation for how timescale can be leveraged to achieve divisive inhibition in spiking networks. Although we will give a more complete analysis further on, we first demonstrate the influence of input timescale on the rate response of a leaky membrane to provide an intuitive description of how input statistics affect the variance of membrane voltage and thereby modulate response gain. For this, consider a simplified current-integrator neuron model, *i.e.*

$$\tau_v \frac{d}{dt} v = -v + I(t; \sigma, \tau, \mu), \tag{4.5}$$

that corresponds to an O-U process with input current I, which is a Gaussian process of mean μ, variance σ^2, and autocorrelation timescale τ, acting on a leaky membrane of timescale τ_v. Since (4.5) is linear in v, the effect of current I on the membrane voltage is captured by a decaying exponential filter h_v of timescale τ_v.

As illustrated in Fig. 4.4, we now make a simplifying assumption that the unit's spiking activity R_v is proportional to the fraction of time that the filtered current is over some finite threshold I_{th}, *i.e.*

$$R_v = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathcal{U}\left(I(t;\sigma,\tau,\mu) * h_v(t) - I_{th}\right) dt, \tag{4.6}$$

where $\mathcal{U}$ is the unit step function. Since I is stationary, R_v can be computed from the Gaussian probability distribution $P_v(x) = Pr\{v = x\}$; calculating the statistics of v, we find

$$\mathrm{Var}[P_v] = \sigma^2 \frac{\tau}{\tau + \tau_\nu}. \tag{4.7}$$

Then, with R being proportional to the mass of probability over threshold, we have

$$R_v \propto \frac{1}{\sigma}\sqrt{\frac{\tau+\tau_v}{\tau}} \int_{I_{th}}^{\infty} \exp\left(-\frac{(x-\mu)^2(\tau+\tau_v)}{2\sigma^2\tau}\right) dx. \tag{4.8}$$

To estimate the gain of the rate response, we differentiate R_v with respect to μ and evaluate at $\mu = I_{th}$, finding the slope to be proportional to simply

$$\frac{1}{\sigma}\sqrt{\frac{\tau+\tau_v}{\tau}}. \tag{4.9}$$

Thus, in the limit $\tau << \tau_v$, this simplified analysis demonstrates response gain as propor-

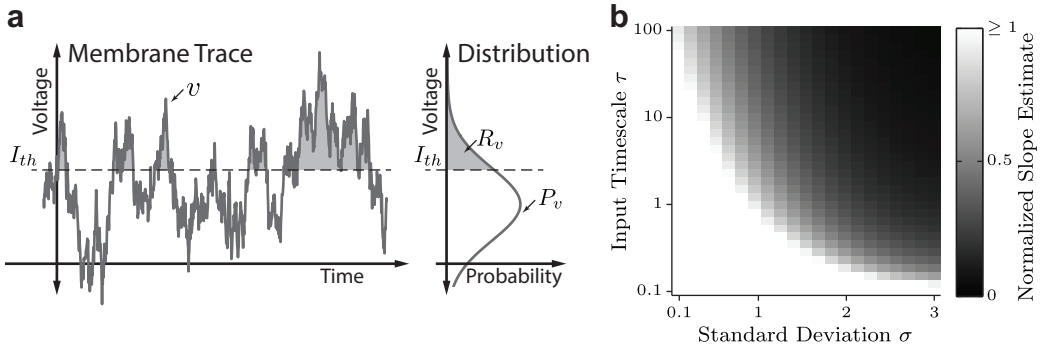

Figure 4.4: **Input timescale affects gain by shaping membrane voltage variance.** (a) Schematic of rate calculation from the probability distribution of membrane activity in the simplified current integrator model. (b) Estimate of the rate response gain as determined by (4.9), demonstrating this simplified current integrator to explain the qualitative effect of timescale on response gain.

tional to $\sqrt{\tau/\tau_v}$, *i.e.* the root of the ratio of synaptic and membrane timescales.

In Fig. 4.4 we see that as the input timescale τ rises for fixed membrane constant τ_v, the slope of the rate response curve will decrease, consistent with simulations shown in Fig. 4.3b. A biophysical interpretation of this simplified analysis is that a membrane's intrinsic properties apply a low-pass filter to stochastic input (O'Donnell and van Rossum, 2014) so that slower inputs produce a greater voltage variance. Most importantly, from (4.9) we explicitly connect input timescale and response gain; as the autocorrelation timescale of input is increased, the gain of spiking output is reduced. This result is consistent with simulations from Fig. 4.1 and Fig. 4.2, where gain is reduced as either input variance or timescale increases. In sum, by revealing how the autocorrelation structure of input can affect the slope of output activity, this analysis makes useful predictions regarding when and how gain modulation will occur in more realistic models.

4.3.2 Analyzing divisive inhibition in a conductance-based model

So far we have examined the response of simple integrate-and-fire models in which a parameterized input current is injected into the model membrane. However, a more complete understanding of divisive inhibition requires a study of conductance-based input as well. To explain effects of input variance and timescale on both current- and conductance-based LIF neurons, we seek analytic solutions to the firing rate of a general LIF model that explicitly includes input timescale and conductance-based inputs (i.e., multiplicative noise). This is a challenging task, as including separate timescales for membrane and input dynamics makes exact solutions difficult to calculate.

Previous work with current-based models, however, shows a way forward by demonstrating that a change in the timescale of input currents affects a shift in the effective voltage threshold and reset potentials of a model membrane (Brunel and Sergi, 1998; Fourcaud and Brunel, 2002). Analysis of the authors' solutions reveals these shifts to be consistent with a divisive modulation of firing rates. However, when moving beyond current injection it is unclear exactly how conductance timescales will interact with spiking activity since spike-

generated voltage fluctuations significantly alter the input current statistics that a neuron receives. In this section we examine a biophysically-realistic stochastically-driven LIF model neuron and develop an analytic solution for the firing rate, studying both current- and conductance-based inputs.

To calculate the firing rate of a general LIF model, *i.e.* a spiking membrane receiving current or conductance input of arbitrary timescale, we define the dynamics of membrane voltage v and synaptic input s as

$$\begin{aligned} \tau_m \frac{d}{dt} v &= A(v) + B(v)s(t) \\ \tau_s \frac{d}{dt} s &= -s + \sqrt{\tau_m}\eta(t) \end{aligned} \tag{4.10}$$

where η is standard white noise. Note that the synaptic input s, defined by the SDE

$$ds = \frac{-s}{\tau_s}dt + \frac{\sqrt{\tau_m}}{\tau_s}dW_t, \tag{4.11}$$

is an OU-process of zero mean and variance $\tau_m/2\tau_s$. An important feature of this formulation is that since the membrane terms A and B are functions of voltage, they can be chosen in such a way as to capture the behavior of a current- or conductance-based neuron. To model action potentials, when the membrane potential v exceeds threshold v_{th}, a spike is emitted and reset to v_{re}. Building from methods presented elsewhere (Brunel and Sergi, 1998; Brunel and Latham, 2003; Fourcaud and Brunel, 2002), in (A.1)-(A.22) we derive the mean firing rate R, calculated up to first order in the timescale ratio $k = \sqrt{\tau_s/\tau_m}$, given by

$$\frac{1}{R} = 2\tau_m \int_{v_{re}^{eff}}^{v_{th}^{eff}} \frac{dz}{B(z)W(z)} \int_{-\infty}^{z} \frac{W(x)}{B(x)}dx, \tag{4.12}$$

where

$$W(v) = \exp\left(2\int^{v} \frac{A(u)}{B^2(u)}du\right) \tag{4.13}$$

and where

$$\begin{aligned} v_{th}^{eff} &= v_{th} + B(v_{th})\frac{\alpha}{2}k \\ v_{re}^{eff} &= v_{re} + \frac{B^2(v_{re})}{B(v_{th})}\frac{\alpha}{2}k \end{aligned} \tag{4.14}$$

are the effective membrane threshold and reset potentials. Note that $\alpha = -\sqrt{2}\zeta(\frac{1}{2})$ for the Riemann zeta function ζ (Hagan et al., 1989). Importantly, the dependence of R on timescale ratio k is consistent with (4.9) as determined from our simplified analysis above.

Following Ayaz and Chance (2009), a biophysically-realistic model neuron observes the membrane equation

$$C\frac{\partial}{\partial t}V = g_l(v_l - V) + g_e(v_e - V) + g_i(v_i - V) + I, \tag{4.15}$$

where the membrane constant τ_m is defined by $\tau_m = C/gl = 37$ ms, from $g_l = 20$ nS, with reversal potentials $v_l = -70$ mV, $v_e = 0$ mV, and $v_i = -80$ mV. Membrane threshold and reset potentials are $v_{th} = -52mV$ and $v_{re} = v_l$, respectively. Therefore, to implement a simple current-driven neuron, we define an input current I, *e.g.*

$$I = (\sigma s + \mu)g_l(v_e - v_0), \tag{4.16}$$

where stochastic input s is scaled by σ and shifted by μ to achieve arbitrary variance and mean. Here, v_0 is a 'typical' membrane voltage expected in the absence of input, which we take here as $v_0 = -62$ mV. Note that a tonic inhibitory (shunting) conductance of $3g_l$ is included to prevent the emission of spikes in the presence of noisy excitatory input. In this regime, for mean excitatory current $\mu = 0$, average membrane voltage is ≈ -80 mV, *i.e.* near v_i. To approximate the response of this current-driven system we thus use

$$\text{Current} \triangleq \begin{cases} A(v) = (v_l - v) + 3(v_i - v) + \mu(v_e - v_0) \\ B(v) = \sigma(v_e - v_0). \end{cases} \tag{4.17}$$

for the parameterizations of (4.10). In Fig. 4.5a we compare the analytic approximation of (4.17) to simulation, finding divisive inhibition to occur for increasing synaptic timescale τ_s over a wide range of firing rates.

Given that conductance is the biophysically-realistic manner in which neuron's receive excitation and initiation, we now implement conductance-based input by eliminating current injection of (4.15) and instead introduce stochastic excitatory conductance, *i.e.* $I = 0$ and

$g_e = (\sigma s + \mu) g_l (v_e - V)$. Thus, for a conductance-based formulation, (4.10) takes the form

$$\text{Conductance} \triangleq \begin{cases} A(v) = (v_l - v) + 3(v_i - v) + \mu(v_e - v) \\ B(v) = \sigma(v_e - v). \end{cases} \tag{4.18}$$

In Fig. 4.5b we compare the response curves generated by this conductance-based formulation in both simulation and analytic approximation, noting divisive inhibition to again occur as timescale τ_s is increased. Importantly, both simulations of Fig. 4.5 are qualitatively identical to Fig. 4.2 for coupled timescale and variance due to the stochastic input process s defined by (4.11) having a variance that scales with $1/\tau_s$. Since the expression for the firing rate (4.12) is only guaranteed to be accurate up to first order in the timescale ratio k, we next assess its predictive power. Specifically, we compare predicted response curves to simulation of current (Fig. 4.5a) and conductance (Fig. 4.5b) inputs according to (4.17) and (4.18), respectively. We find the analytic solution to accurately predict LIF firing rates in both cases. Therefore, divisive inhibition via input timescale is captured by the approximation (4.12), demonstrating timescale-based divisive inhibition as a general effect for LIF neurons receiving stochastic input.

4.3.3 Achieving timescale-based divisive inhibition in a spike-driven neuron

As recent studies emphasize the use of variance toward gain modulation in spike-driven networks (Mitchell and Silver, 2003; Chance et al., 2002; Ayaz and Chance, 2009), it is important to now consider how timescale affects neuronal response in the biophysical context of fluctuating background excitatory and inhibitory input. In particular, we wish to examine gain modulation in spiking networks that are balanced, *i.e.* where neurons receive complementary magnitudes of excitation and inhibition, while systematically varying input timescale. In this section we explore the interaction of input timescale and balanced activity toward divisive inhibition, demonstrating how synaptic timescale can affect division in spiking networks.

To build upon recent findings we will follow a biophysically-realistic model that has been previously studied (Ayaz and Chance, 2009); we consider a conductance-based neuron with

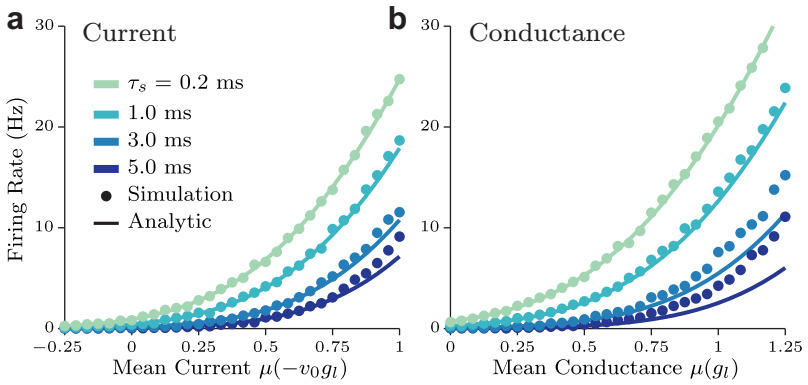

Figure 4.5: **Analytic approximation captures timescale modulation of firing rate response in biophysically-realistic LIF cells.** The rate response of (a) noisy current injection of (4.17) for varying input timescale τ_s, with $\sigma = 0.4$ and $v_0 = -62$ mV. Thus, mean current spans $-.31$ to 1.24 nA, generating the expected output range of 0 to 30 Hz. (b) A conductance-based neuron of (4.18) has similar divisive behaviour for increasing timescale τ_s. Approximate analytic solution (4.12) for the firing rate accurately predicts divisive inhibition across models as input timescale is increased.

membrane and synaptic dynamics defined by (4.15) with excitatory and inhibitory synaptic dynamics

$$\tau_n \frac{d}{dt} g_n = -g_n \quad | \quad n \in \{e, i\}, \tag{4.19}$$

where excitatory and inhibitory spikes instantaneously increase conductance upon arrival, *i.e.*

$$g_n(t) \leftarrow g_n(t) + \frac{\alpha_n}{\tau_n} \mathcal{P}(t; R_n). \tag{4.20}$$

Note that $\mathcal{P}$ is a train of unit impulse functions generated from a homogeneous Poisson process of rate R_n. Depicted in Fig. 4.6a, we consider this model as having two sources of input: background and drive. Background consists of excitatory and inhibitory population activity, while drive represents the signal of synaptic excitation that is to be divided.

Formally, the gain of a neuron is determined from the slope firing rate as a function of additional synaptic drive R_{ex}. As depicted in Fig. 4.6a, this is explored for different values of total balanced background activity R_{total}. Thus, for a fixed level of background R_{total} and drive R_{ex}, the rate of excitatory and inhibitory synaptic events required to simulate

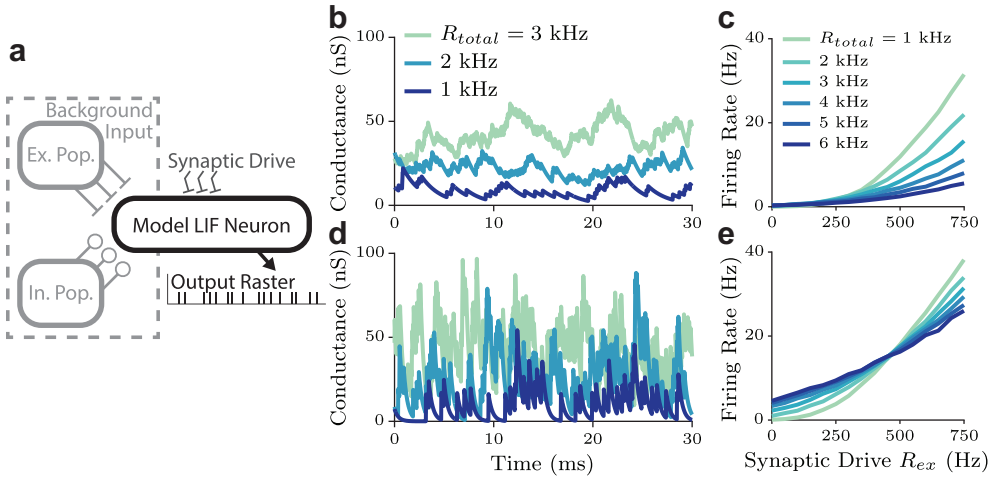

Figure 4.6: **Divisive inhibition via balanced synaptic activity.** (a) Schematic of balanced background activity as proposed by Ayaz and Chance (2009) to achieve divisive inhibition of a conductance-based LIF neuron's rate response. Example excitatory conductance traces at multiple firing rates are given for synaptic timescales of (b) 5 ms and (d) 1 ms. (c) By increasing the amount of excitatory and inhibitory conductance at a synaptic timescale of 5 ms, the rate response to additional synaptic drive is divisively inhibited. (e) Modulation of gain becomes non-monotonic over input for different synaptic timescales, *i.e.* 1 ms. Parameters of (4.15) and (4.20) are identical to Ayaz and Chance (2009), unless otherwise specified, and agree with known biophysical quantities. Note that in (4.20) we scale each incoming conductance event by $1/\tau_n$ to fix net conductance while timescale is varied (see Fig. 4.2).

Fig. 4.6a under (4.15), (4.19), and (4.20) is $R_e = \frac{1}{2}R_{total} + R_{ex}$ and $R_i = \frac{1}{2}R_{total}$, respectively. Importantly, the relative strength of excitation and inhibition, *i.e.* magnitude of conductance and associated reversal potentials, are set so as to maintain resting-state membrane voltage as balanced activity is increased; for $R_{ex} = 0$, the time-averaged voltage does not change as $R_{total} \geq 0$ is varied, with the membrane in neither a hyperpolarized nor spiking state (Chance et al., 2002).

To first understand the effect of balanced synaptic activity activity, we demonstrate divisive inhibition in a conductance-based spike-driven context as background input R_{total} is increased. In Fig. 4.6b,c we find this effect to be qualitatively identical to results of current-based simulations previously reported (Ayaz and Chance, 2009). Note here that the

synaptic time constants τ_e and τ_i are fixed at 5 ms, unlike in Fig. 4.2 and Fig. 4.5, with division occurring as the total amount of background synaptic inputs $R_{total} = R_i + R_e - R_{ex}$ is increased.

We next assess the contribution of input timescale by altering the constants of synaptic dynamics. In Fig. 4.6b,d we illustrate conductance traces generated for different spike-rates with 5 ms and 1 ms synaptic timescales, respectively (see Fig. 4.6 for details). Interestingly, in Fig. 4.6e we see that for the case of fast synapses, *e.g.* $\tau_e = \tau_i = 1$ ms, rate response is no longer scaled equivalently for all levels of synaptic drive. While the slope of the resulting curve does decrease with balanced synaptic activity, consistent with a modulation of gain, spiking onset occurs at lower levels of excitation. Therefore, for fixed synaptic timescales, balanced synaptic activity requires a precise tuning of excitatory and inhibitory magnitudes; evident in Fig. 4.6e, such mistuning results in balanced activity *increasing* rate response at low levels of net synaptic drive. Nevertheless, Fig. 4.6 suggests timescale to modulate the gain of a neuron receiving background synaptic activity.

This suggestion is confirmed by direct demonstration of timescale as a mechanism for gain modulation in a conductance-based LIF neuron. Unlike in Fig. 4.6, where division inhibition is achieved via balanced activity for fixed synaptic timescales, in Fig. 4.7 excitatory and inhibitory synaptic timescales are varied while holding the number of synaptic events and net conductance fixed. As predicted by our analysis of simplified current models in Fig. 4.3, the extent to which synaptic timescales modulate response gain is determined by the magnitude of total balanced activity, with a stronger divisive timescale effect occurring when the number of balanced excitatory and inhibitory synapses, and hence background synaptic variance, is high.

From this observation we reconsider the analytic approximation (4.10) depicted in Fig. 4.5. While the solution of (4.12) accurately reproduces rate response curves for increasing μ, we observe only a modest divisive effect with increasing synaptic timescale τ_s. From a comparison to Fig. 4.7 we attribute this to a low overall input variance due to a lack of balanced activity, suggesting timescale-based rate modulation, a potential mechanism for division in

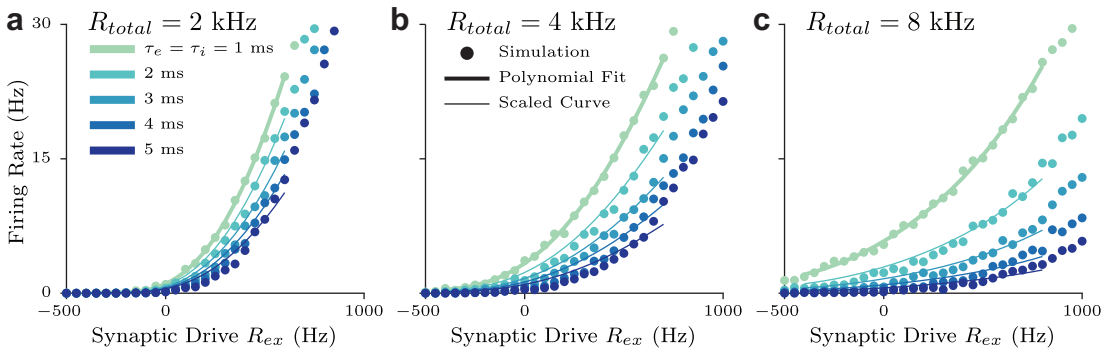

Figure 4.7: **Divisive inhibition achieved by increasing input timescale**. Time constants of excitatory τ_e and inhibitory τ_i synapses are varied while holding net conductance fixed. In each plot, a 4^{th}-order polynomial is fit to $\tau_n = 1$ ms simulation data (thick line). Then, best-fitting scaling factors are computed to minimize least-squared error between simulation and scaled curve predictions (thin lines). Note that the *amount* of timescale-induced division is proportional to total activity, *i.e.*, (a) $R_{total} = 2$ kHz, (b) 4 kHz, and (c) 8 kHz.

spiking networks, is most prominent in regimes of high balanced background input. We next test this hypothesis by exploring an analytic formulation of balanced activity developed above.

4.3.4 Analyzing divisive inhibition via input timescale in regimes of balanced activity

To explain the interaction between input timescale and balance we examine analytic solutions to (4.12) that replicate an increased background activity regime. Specifically, as in Fig. 4.7, we aim to study the behavior of response curves as synaptic timescale and total balanced background activity rates are varied. Unfortunately, a conductance-based membrane equation like (4.15), containing both excitatory and inhibitory synaptic conductance, cannot be written in the form of (4.10) due to the presence of multiple types of synaptic inputs with distinct reversal potentials. In fact, this rate response approximation would require an analytic solution to a three-dimensional stochastic system, which is beyond the scope of the systems investigated here. Rather, the effects of balanced activity may be seen by introducing to (4.18) an excitatory conductance γg_l and inhibitory conductance $(3\gamma + 3)g_l$, defined

to be proportional to parameter γ that scales in tandem the amount of balanced activity on membrane voltage. Note that the relative magnitude of excitation and inhibition, *i.e.* $1:3$, is chosen to prevent spontaneous spikes for $\mu = 0$, analogous to the ratio of synaptic magnitudes used in Fig. 4.6 (Ayaz and Chance, 2009). Thus, we consider a modified version of the conductance-based model from (4.18):

$$\begin{cases} A(v) = (v_l - v) + (3\gamma + 3)(v_i - v) + (\gamma + \mu)(v_e - v) \\ B(v) = \sigma(v_e - v). \end{cases} \tag{4.21}$$

In Fig. 4.8a, we compare the analytic approximation of rate responses from (4.21) against numeric simulation. We find the output rates predicted by (4.12) to be accurate over the range of input timescales in this regime. Further, divisive inhibition is most prominent for large γ (Fig. 4.8a, $\gamma = 1$) analogous to Fig. 4.7c where the amount of balanced synaptic activity is high. While there is a small systematic error in the analytic approximation at higher timescales, the qualitative trend indicates significant modulation of responses as τ_s is increased over a small range relative to τ_m. Thus, while not identical to the full simulation of (4.15), our analytic approximation of balanced activity replicates qualitative behaviors.

We next examine numerical solutions to the integrand of (4.12) over voltage, which is inversely proportional to the firing rate R. Fig. 4.8b depicts the integrand as a function of mean synaptic drive over $\gamma = 0.25$, 0.5, and 1. Since from (4.14) we find input timescale to affect rate response R through a shift in the bounds of integration, we plot the effective reset and threshold potentials for $\tau_s = 0.2$ ms and 5 ms. While an increase in τ_s significantly shifts the effective spiking threshold, at small γ the integrand is near zero when mean input is nonzero. Conversely, high balanced activity, as implemented by large γ of (4.21), shifts mean activity curves such that they are nonzero primarily in the region of voltage included by the effective bound of integration as input timescale is increased. Thus, the strong inhibitory effect of input timescale can be explained by a shift in the effective spiking threshold of a neuron receiving sufficient balanced activity.

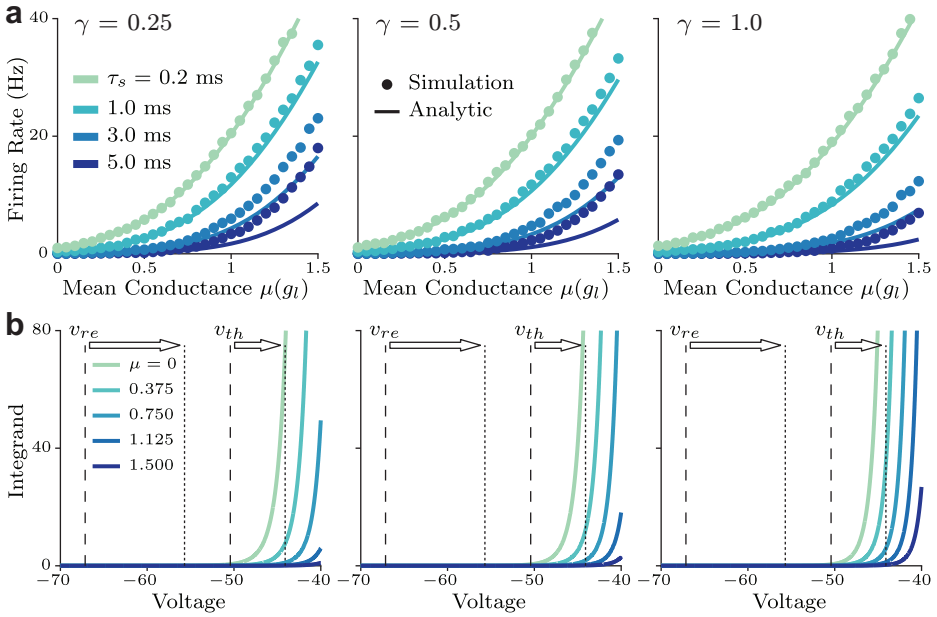

Figure 4.8: **Balanced activity promotes divisive inhibition via input timescale.** (a) Analytic approximation to the rate response of a conductance-based LIF neuron, plotted as functions of mean stochastic conductance for multiple synaptic timescales τ_s and balanced activity parameter γ. Note the strong agreement to simulation, particularly when τ_s is small. (b) Depiction of the integrand in (4.12) as a function of voltage for various levels of mean conductance, illustrating the effect of increasing synaptic timescale on the effective voltage reset and threshold between $\tau_s = 0.2$ ms (dashed) and 5 ms (dotted); the response curves to mean input are compressed as γ increases, allowing for the shift in effective reset v_{re} and threshold v_{th} as τ_s increases (arrows) to modulate responses.

4.3.5 A circuit implementation of timescale-based gain modulation

When considering how gain modulation via input timescale may arise in neural networks, it is important to ask how networks of neurons could shape the statistical structure of their input, even when synaptic timescales and levels of activity are fixed. One possibility, as found by several studies into the role of correlated and balanced input (Moreno et al., 2002; Brunel and Wang, 2001; Litwin-Kumar et al., 2011), is that correlated spiking among upstream neurons can change the timescale of net activity. We now explore how this phenomenon may lead to divisive inhibition, using incoming pools of balanced synaptic activity that are

correlated by a common drive.

The phenomenon of spike time correlations affecting input timescale is illustrated in Fig. 4.9a, where two pools of units are synchronized by a common source. Here, the output rate of each pool (R_1 and R_2) is driven by an independent and a shared source of activity (R_c) so that total output activity between pools, *i.e.* $R_1 + R_2$, is fixed. As the relative strength of common versus independent activity is varied, correlations are induced between output spikes. Importantly, the common input source contains a delay Δ, resulting in two characteristic spiking events in the summed activity: independent spikes occurring in either pool, and correlated events, offset by Δ, occurring between both pools.

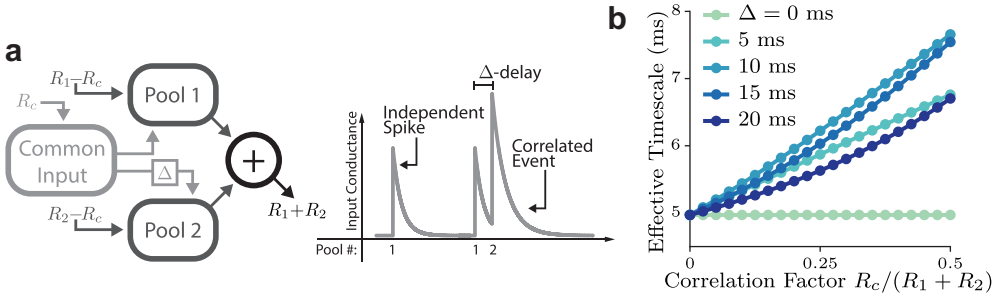

Figure 4.9: **Input autocorrelation timescale shaped by correlated activity.** (a) For two pools of driven neurons, as the amount of correlated activity R_c increases, the occurrence of correlated synaptic events, separated by a delay of Δ, increases. (b) By increasing the correlation factor $Rc/R_1 + R_2$, *i.e.* the amount of common input relative to fixed activity between pools, the effective autocorrelation of the summed pool output is increased.

Since each class of spiking event is independent, the autocorrelation of the resulting summed conductance process is a linear combination of the autocorrelation for the two distinct event types. These autocorrelation functions are in turn determined exactly by the synaptic timescale and delay. By fitting a decaying exponential to these autocorrelations (see Fig. A.2), we estimate the effective timescale of the summed process as a function of the proportion of correlated activity $R_c/(R_1 + R_2)$ for various values of delay Δ. We see in Fig. 4.9b that for amounts of delay similar to the synaptic timescale, *i.e.* $0 < \Delta \approx \tau = 5$ ms, correlated inputs display a modest increase in timescale relative to independent inputs.

Simulation utilizing a non-constant distribution of delays, *i.e.* synaptic jitter, produces a similar increase in effective timescale (not shown); here we remain agnostic to the specific biophysical mechanism responsible for spike-time latency and assume constant delay time for simplicity.

For a neuron driven by excitatory and inhibitory populations, correlations can occur either within or between each population. In the case of excitation and inhibition that are separately correlated, the resulting effect on firing rate is easy to interpret: correlations increase both the variance and effective timescale of each conductance process, resulting in a decrease in the slope of the resulting rate response curve as illustrated in Fig. 4.1 and Fig. 4.3. However, this decrease in slope is accompanied by an overall increase in spiking activity, as in Fig. 4.1c,e, resulting in a net excitatory effect that is inconsistent with division.

We instead focus on the case of correlation between populations. As we will show, this scenario yields a spiking circuit, depicted in Fig. 4.10a, that successfully implements divisive inhibition. While it is unlikely that such a circuit is literally implemented in biological neural networks, it simply serves as a schematic to demonstrate how network properties can influence gain modulation. Here, the response of a conductance-based LIF neuron to an excitatory drive signal is modulated by correlating background activity. Prior to correlation, both excitatory and inhibitory populations are independently driven. To introduce synchronized synaptic events, a rate of excitatory and inhibitory spikes encoding a modulator signal then titrates the ratio of independent and common activity, *i.e.* R_c/R_{total}, with common inhibitory synapses occurring 10 ms after excitatory counterparts.

Crucially, as the amount of correlation is increased, the level of activity in either excitatory and inhibitory population is constant, with only the occurrence of correlated synaptic events increasing. Results of this network simulation are given in Fig. 4.10b, illustrating the excitatory drive signal to be divisively inhibited as the correlation between excitatory and inhibitory synapses is increased. Again, both the synaptic timescale and total amount of excitatory and inhibitory synapses are fixed, indicating correlated activity as a means of leveraging input timescale to achieve a useful computation with spiking neural networks.

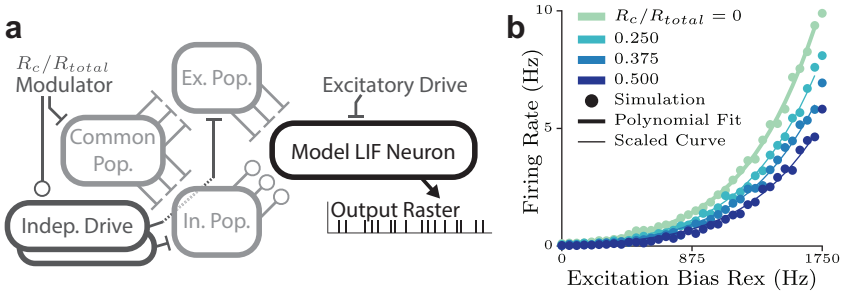

Figure 4.10: **Divisive inhibition achieved by correlating input activity.** (a) A simple network is constructed to achieve divisive inhibition by leveraging correlated activity to alter input timescale. The output rate in response to an excitatory drive signal is divisively inhibited by a modulator signal proportional to the amount common activity R_c relative to the total excitatory and inhibitory drive $R_{total} = R_e + R_i = 12$ kHz. (b) Simulation of the rate response to excitatory drive as R_c/R_{total} is increased. As in Fig. 4.7, a 4^{th}-order polynomial is fit to uncorrelated simulation data (thick line). Best-fitting scaling factors are found to demonstrate division of responses for increasing R_c/R_{total} (thin lines).

The effect demonstrated in the circuit of Fig. 4.10a cannot be attributed entirely to input timescale, since correlations in the spiking circuit are between excitatory and inhibitory populations that have an opposed effect on voltage. When excitatory and inhibitory synapses are perfectly synchronized, the net effect is a *veto* on synaptic transmission, inhibiting the rate response in a manner distinct from modulations of input variance or timescale. However, our simulations show that for $\Delta \neq 0$, firing rate division persists. Rather, even though variance of conductance-driven voltage fluctuations are reduced by synchronizing between populations, the increase of input timescale is sufficient for divisive inhibition of rate response to occur. The circuit proposed in Fig. 4.10a, therefore, serves as a proof of concept for correlated synaptic activity achieving divisive inhibition, providing a possible connection between observations of attention affecting correlations and synchrony within neural populations (Litwin-Kumar et al., 2011; Ruff and Cohen, 2014), and the theory of normalization as a model for attention (Cohen and Kohn, 2011; Reynolds and Heeger, 2009; Rabinowitz et al., 2015).

4.4 Discussion

While divisive inhibition is widely accepted to be a fundamental operation in neural information processing, much work remains to be done in understanding how it occurs in biologically realistic circuits. Our investigation of the role of autocorrelation timescale gives insight into how stochastic properties of input, including variance, modulate the gain of a neuron's firing rate to produce a division of spike response. We show how this occurs over plausible parameter regimes in both current- and conductance-based LIF neurons. Further, by constructing simple circuits leveraging correlated activity, we demonstrate how variation in input timescales can produce divisive inhibition in spiking networks.

Our analyses and methods build upon previous findings of divisive inhibition in simulations and give tools to explain precisely why gain modulation occurs under stochastic input, particularly in the presence of balanced background synaptic activity (Chance et al., 2002; Ayaz and Chance, 2009). Our investigation of input timescale is motivated by prior work showing timescale to significantly impact firing rates of model neurons (Mitchell and Silver, 2003). In particular, when studying the quadratic integrate-and-fire (QIF) neuron, Brunel and Latham (2003) found analytic solutions to a neuron's firing rate driven by stochastic input with both the membrane and synaptic time constants explicitly parameterized. Moreover, when examining solutions in the large synaptic timescale limit, increasing input variance was found to suppress QIF output rate. Indeed, spike-rate response approximations of current-based LIF model have previously been derived, but these have only explored limited input parameters and local effects frequency response modulation (Brunel and Sergi, 1998; Fourcaud and Brunel, 2002). In the present study we generalize these results to explain a range of current- and conductance-based input regimes, and most importantly, study their application across a wide range of output rates toward timescale-based gain modulation for the purposes of divisive inhibition.

Additionally, our investigation reconciles an apparent conflict within the literature regarding how input autocorrelation structure affects the rate response of LIF neurons by

considering timescale as a mechanism for divisive inhibition in a spike-driven setting. Prior work has typically used specific current injection parameterizations at isolated levels of input strength, either overlooking timescale-based rate modulation (Mitchell and Silver, 2003; Richardson, 2004), or noting effects that were relatively minor in chosen parameter regimes (Brunel and Latham, 2003). By examining the qualitative behavior of the output response curve at steady state across a range of activity, our investigation demonstrates a clear and complete relationship between input timescale and response gain, both in simulation and analysis. Further, studies that suggest no significant effect of input timescale may have drawn these conclusions due to analysis utilizing sinusoidal driving signals which possess a periodic autocorrelation structure (Ly and Doiron, 2009) or with input magnitude carefully scaled so as to hold membrane voltage at constant variance (Alijani and Richardson, 2011; O'Donnell and van Rossum, 2014). Our investigation reveals that input timescale achieves gain modulation precisely because membrane voltage statistics are affected; the effective autocorrelation structure of the input, together with the filtering properties of the membrane, act to set voltage variance and thereby shift effective spiking thresholds to modulate gain. Thus, our study resolves various disparate findings regarding input timescale.

Overall, our study highlights the need to consider input autocorrelation structure when investigating neural computation either *in vivo* or simulation. For example, circuits constructed from simplified neuron models, such as the generalized linear model (GLM), may not exhibit timescale-based gain modulation; an interesting and open question is to what extent properties of a GLM's spatiotemporal stimulus filter and spike history can replicate timescale-based computations present in LIF networks. Conversely, even if the mechanism of input timescale is not utilized *in vivo*, our findings indicate that effects of gain control will occur in network simulations of LIF model neurons where properties such as correlated upstream activity or multiple synaptic time constants can shape the autocorrelation structure of input activity.

Importantly, our work allows for testable predictions to be made regarding circuitry underlying common computations in biological networks. In particular, the mechanism of input

timescale could be leveraged to modulate gain in neural networks by utilizing neurotransmitters of varying synaptic timescales. For example, recruiting fast AMPA versus slower NMDA excitatory synapses could mediate timescales of excitatory inputs; the same is true of faster GABA A versus slower GABA B synapses. Additionally, while normalization is believed to play a significant part in selectivity and invariance computations in sensory systems, it is unclear exactly how pooled activity across a population modulates the gain of individual neurons or recurrent connectivity (Serre et al., 2005; Yu et al., 2002). By combining the present work on mechanisms of input variance or timescale with metrics such as accuracy or metabolic cost, future work could identify circuit motifs that are the most effective or efficient at achieving divisive inhibition in sensory systems.

CONCLUDING REMARKS

In this dissertation I seek to further understanding of the remarkable computational processes that underly perception of naturalistic shape in biological vision systems. These problems are remarkably broad and complex, spanning multiple fields of research. To that end, this work is an attempt to bring together a range of perspectives to shed light on questions that may otherwise remain unresolved.

In my study of spectral receptive fields in Chapter 2 I conclude that, at least for the majority of V4 units, such simple phase-invariant models cannot account for the tuning of boundary conformation within V4. Although a few cells may in fact posses this behavior, they are by no means representative of the population. However, this is not to say that linear combinations of phase-invariant inputs tuned for spatial frequency and orientation have no place in the ventral stream. To see this, one must consider that V4 cells selected for recording are typically mapped and identified using stimuli similar to those used during data collection. Thus, the population of V4 cells used in Chapter 2 may have been biased toward highly shape-tuned cells, missing other single-unit responses that may, in fact, be well-modeled by SRF-like computations.

The utility of SRF-like neurons or signals may be seen when considering blur-selective responses described in Chapter 3. My analysis reveals that tuning for intermediate blur is consistent with shape-selective activity being multiplicatively scaled by responses selective for intermediate spatial frequencies. Naturally, one simple mechanism to achieve this intermediate blur signal is an excitatory low spatial frequency signal being inhibited by a high spatial frequency signal, very much in the flavor of SRF models limited to two spatial frequencies with positive and negative coefficients, respectively. This signal would then multiplicatively scale the response of a shape-selective neuron to achieve the separable neu-

ral code implicated by the proposed APCB model. Thus, while the notion of SRFs as a mechanism for shape selectivity has largely been ruled out, it will be fascinating to consider future studies in which SRF-like computations do occur with respect to naturalistic shape and sufficient scene representations within V4.

With respect to the sufficient coding of natural scenes, my study lacks the sample size to make a clear statement regarding the proportion and distribution of intermediate-blur-selective cells within V4. One would assume that if V4 neurons tile the space of boundary conformation across the visual field, there would also exist a similar range of tuning for various magnitudes of blur. One intriguing study would investigate the distribution of shape and blur features that exist in natural images, to see if V4 might possess (in some sense) an optimal distribution of blur-tuning.

It is clear, however, that the limited dataset as presented in Chapter 3, containing only a handful of cells tuned for intermediate blur magnitudes, is not sufficient to answer such questions. Part of this may be due to our experimental procedures, which identified and mapped individual neurons using only sharp shape stimuli with low blur magnitudes. Thus, cells which may have been highly shape-selective at intermediate or high magnitudes of blur may have been underrepresented. Similarly, exemplar preferred and non-preferred shape stimuli were chosen from responses to the entire shape set at the minimum blur magnitude. Having found blur to be an important stimulus feature, an experimental paradigm to classify V4 selectivity ought to span dimensions of both object shape and blur, presenting stimuli that adequately span each feature dimension. A careful exploration of such procedures, perhaps utilizing adaptive stimulus presentation schemes, is left for future work.

Perhaps the most fertile ground for further investigations lies in the relationship between blur information as multiplicative scaling shape selectivity and mechanisms for gain modulation in spiking neural circuits. In Chapter 4, while I develop a framework for analyzing biophysically-realistic scenarios in which divisive inhibition can occur, I have yet to investigate their specific application toward specific computations. Considering the case of blur-selectivity in V4, gain modulation may occur at multiple levels. For example, blur

modulation may act on a single neuron receiving shape selective input, or it may modulate the gain of upstream units that are themselves contributing to shape selectivity. To compound this issue, a V4 unit selective for shape and blur may itself modulate the gain of its responses relative to the activity within its local population, *e.g.* via response normalization across multiple cells. However, exactly how these different forms of gain modulation interact, and to what extent these interactions will affect efficient neural coding, remains to be seen.

The framework presented in Chapter 4 will aid in the construction of models that can achieve these computations. For example, the transient nature of blur selectivity may imply some form of lateral or feed-forward gain modulation, while sustained reduction of responses may rely on recurrent inhibition. Thus, my results suggest potential mechanisms for blur modulation, including the decorrelation of feed-forward boundary feature responses which are then filtered out through recurrent processes. Although merely speculation, one may imagine potential modeling experiments that utilize different network configurations; my analytic formulation of divisive inhibition via input variance or timescale may suggest or rule out potential circuit-level mechanisms.

In any event, questions rising from this study seem to greatly outnumber those which have been answered. While we have clarified some of the underlying processes surrounding the perception of naturalistic boundary conformation, much work remains to be done before the construction of image-computable biophysical models of blur- and shape-selectivity, for example, could begin. The overall view of this research, however, is encouraging, as it suggests that progress has been made toward understanding the remarkable computations of shape.

BIBLIOGRAPHY

Adelson EH, Bergen JR (1991) The plenoptic function and the elements of early vision. In Landy MS, Movshon JA, editors, *Computational models of visual processing*, pp. 3–20. The MIT Press, Cambridge, MA, US.

Albrecht DG, De Valois RL, Thorell LG (1980) Visual cortical neurons: are bars or gratings the optimal stimuli? *Science* 207:88–90.

Alijani AK, Richardson MJE (2011) Rate response of neurons subject to fast or frozen noise: From stochastic and homogeneous to deterministic and heterogeneous populations. *Physical Review E* 84:011919.

Allen BP (1999) Shadows as sources of cues for distance of shadow-casting objects. *Perceptual and motor skills* 89:571–84.

Angelucci A, Bressloff PC (2006) Chapter 5 Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in Brain Research* 154:93–120.

Attneave F (1954) Some informational aspects of visual perception. *Psychological Review* 61:183–193.

Attwood CI, Harris JP, Sullivan GD (2001) Learning to search for visual targets defined by edges or by shading: Evidence for non-equivalence of line drawings and surface representations. *Visual Cognition* 8:751–767.

Ayaz A, Chance FS (2009) Gain modulation of neuronal responses by subtractive and divisive mechanisms of inhibition. *J Neurophysiol* 101:958–68.

Barlow HB (1972) Single units and sensation: A neuron doctrine for perceptual psychology? *Perception* 1:371–394.

Bell AJ, Sejnowski TJ (1997) The 'independent components' of natural scenes are edge filters. *Vision Research* 37:3327–3338.

Blum H (1967) *A transformation for extracting new descriptors of shape*, Vol. 19.

Brunel N, Chance FS, Fourcaud N, Abbott LF (2001) Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Physical Review Letters* 86:2186–2189.

Brunel N, Latham P (2003) Firing rate of the noisy quadratic integrate-and-fire neuron. *Neural computation* 15:2281–306.

Brunel N, Sergi S (1998) Firing frequency of leaky intergrate-and-fire neurons with synaptic current dynamics. *Journal of theoretical biology* 195:87–95.

Brunel N, Wang XJ (2001) Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *Journal of computational neuroscience* 11:63–85.

Burge J, Geisler WS (2014) Optimal disparity estimation in natural stereo images. *J Vision* 14:1–.

Bushnell BN, Harding PJ, Kosai Y, Bair W, Pasupathy A (2011a) Equiluminance Cells in Visual Cortical Area V4. *J Neurosci* 31:12398–12412.

Bushnell BN, Harding PJ, Kosai Y, Pasupathy A (2011b) Partial occlusion modulates contour-based shape encoding in primate area V4. *J Neurosci* 31:4012–24.

Cadieu C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T (2007) A model of V4 shape selectivity and invariance. *J Neurophysiol* 98:1733–50.

Campbell FW, Cooper GF, Enroth-Cugell C (1969) The spatial selectivity of the visual cells of the cat. *The Journal of physiology* 203:223–35.

Campbell FW, Robson JG (1968) Application of Fourier analysis to the visibility of gratings. *The Journal of physiology* 197:551–66.

Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. *Nature reviews. Neuroscience* 13:51–62.

Carlson ET, Rasquinha RJ, Zhang K, Connor CE (2011) A sparse object coding scheme in area V4. *Current Biology* 21:288–93.

Chance FS, Abbott LF, Reyes AD (2002) Gain modulation from background synaptic input. *Neuron* 35:773–82.

Cohen MR, Kohn A (2011) Measuring and interpreting neuronal correlations. *Nature neuroscience* 14:811–819.

Craft E, Schu H, Niebur E (2007) A Neural Model of Figure Ground Organization. *J Neurophysiol* 2685:4310–4326.

David S, Hayden B, Gallant J (2006) Spectral receptive field properties explain shape selectivity in area V4. *J Neurophysiol* pp. 3492–3505.

De Valois RL, De Valois KK (1990) *Spatial Vision*, Vol. 12.

Desimone R, Schein SJ (1987) Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *J Neurophysiol* 57:835–868.

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

Drewes J, Goren G, Zhu W, Elder JH (2016) Recurrent Processing in the Formation of Shape Percepts. *J Neurosci* 36:185–192.

El-Shamayleh Y, Pasupathy A (2016) Contour Curvature As an Invariant Code for Objects in Visual Area V4. *J Neurosci* 36:5532–43.

Elder JH (1999) Are edges incomplete? *International Journal of Computer Vision* 34:97–122.

Elder JH, Oleskiw TD, Yakubovich A, Peyré G (2013) On growth and formlets: Sparse multi-scale coding of planar shape. *Image and Vision Computing* 31:1–13.

Elder JH, Trithart S, Pintilie G, MacLean D (2004) Rapid processing of cast and attached shadows. *Perception* 33:1319–1338.

Elder JH, Velisavljević L (2009) Cue dynamics underlying rapid detection of animals in natural scenes. *J Vision* 9:7.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1:1–47.

Field DJ, Hayes A, Hess RF (1993) Contour integration by the human visual system: evidence for a local "association field". *Vision research* 33:173–93.

Fleming RW, Holtmann-Rice D, Bulthoff HH (2011) Estimation of 3D shape from image orientations. *Proceedings of the National Academy of Sciences* 108:20438–20443.

Foster KH, Gaska JP, Nagler M, Pollen Da (1985) Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *J Physiol* 365:331–63.

Fourcaud N, Brunel N (2002) Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural computation* 14:2057–2110.

Gabbiani F, Midtgaard J, Knöpfel T (1994) Synaptic integration in a model of cerebellar granule cells. *J Neurophysiol* 72:999–1009.

Gallant J, Braun J, Van Essen DC (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259:100–3.

Gallant J, Connor CE, Rakshit S, Lewis JW, Van Essen DC (1996) Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol* 76:2718–39.

Gardiner C (2009) *Stochastic Methods: A Handbook for the Natural and Social Sciences*.

Gattass R, Sousa aP, Gross CG (1988) Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci* 8:1831–45.

Geisler WS, Perry JS, Super BJ, Gallogly DP (2001) Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research* 41:711–24.

Hagan PS, Doering DA, Levermore CD (1989) Mean Exit Times for Particles Driven by Weakly Colored Noise. *SIAM J. Appl. Math.* 49:1480–1513.

Hanazawa A, Komatsu H (2001) Influence of the direction of elemental luminance gradients on the responses of V4 cells to textured surfaces. *J Neurosci* 21:4490–7.

Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Visual neuroscience* 9:181–97.

Hegdé J, Van Essen DC (2007) A comparative study of shape representation in macaque visual areas V2 and V4. *Cerebral Cortex* 17:1100–1116.

Held RT, Cooper EA, Banks MS (2012) Blur and Disparity Are Complementary Cues to Depth. *Current Biology* 22:426–431.

Holt GR, Koch C (1997) Shunting Inhibition Does Not Have a Divisive Effect on Firing Rates. *Neural computation* 9:1001–1013.

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 148:574–91.

Hubel DH, Wiesel TN (1965) Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat. *J Neurophysiol* 28:229–89.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195:215–43.

Hubona GS, Shirah GW, Jennings DK (2004) The effects of cast shadows and stereopsis on performing computer-generated spatial tasks. *IEEE Transactions on Systems, Man, and Cybernetics* 34:483–493.

Ikeuchi K, Horn BK (1981) Numerical shape from shading and occluding boundaries. *Artificial Intelligence* 17:141–184.

Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71:856–67.

Koch C, Poggio TA, Torre V (1983) Nonlinear interactions in a dendritic tree: localization, timing, and role in information processing. *Proceedings of the National Academy of Sciences* 80:2799–2802.

Kosai Y, El-Shamayleh Y, Fyall AM, Pasupathy A (2014) The role of visual area V4 in the discrimination of partially occluded shapes. *J Neurosci* 34:8570–84.

Kourtzi Z, Connor CE (2011) Neural representations for object perception: structure, category, and adaptive coding. *Annual review of neuroscience* 34:45–67.

Litwin-Kumar A, Oswald AMM, Urban NN, Doiron B (2011) Balanced synaptic input shapes the correlation between neural spike trains. *PLoS Computational Biology* 7.

Livingstone MS, Hubel DH (1987) Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *J Neurosci* 7:3416–68.

Ly C, Doiron B (2009) Divisive gain modulation with dynamic stimuli in integrate-and-fire neurons. *PLoS computational biology* 5:e1000365.

Mishkin M, Ungerleider LG (1982) Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav Brain Res* 6:57–77.

Mitchell SJ, Silver RA (2003) Shunting inhibition modulates neuronal gain during synaptic excitation. *Neuron* 38:433–45.

Mokhtarian F, Mackworth A (1986) Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8:34–43.

Moreno R, de la Rocha J, Renart A, Parga N (2002) Response of Spiking Neurons to Correlated Inputs. *Physical Review Letters* 89:288101.

Movshon JA, Thompson ID, Tolhurst DJ (1978) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J Physiol* 283:53–77.

Nandy AS, Sharpee TO, Reynolds JH, Mitchell JF (2013) The Fine Structure of Shape Tuning in Area V4. *Neuron* 78:1102–1115.

O'Donnell C, van Rossum MCW (2014) Systematic analysis of the contributions of stochastic voltage gated channels to neuronal noise. *Frontiers in computational neuroscience* 8:105.

Oleskiw TD (2010) On the perceptual organization and computational representation of natural planner shape Masters, York University.

Oleskiw TD, Elder JH, Peyre G (2010) On growth and formlets: Sparse multi-scale coding of planar shape In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 459–466. IEEE.

Oleskiw TD, Pasupathy A, Bair W (2014) Spectral receptive fields do not explain tuning for boundary curvature in V4. *J Neurophysiol* 112:2114–22.

Olshausen BA, Field DJ (1996) Natural image statistics and efficient coding. *Network: Computation in Neural Systems* 7:333–339.

Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research* 37:3311–25.

Pasupathy A, Connor CE (1999) Responses to contour features in macaque area V4. *J Neurophysiol* 82:2490–502.

Pasupathy A, Connor CE (2001) Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86:2505–19.

Porter G, Tales A, Leonards U (2010) What makes cast shadows hard to see? *J Vision* 10:13.1–18.

Press WH, Teukolsky Sa, Vetterling WT, Flannery BP (2007) *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Vol. 1.

Rabinowitz NC, Goris RL, Cohen M, Simoncelli E (2015) Attention stabilizes the shared gain of V4 populations. *eLife* 4:e08998.

Ramachandran VS (1988) Perception of shape from shading. *Nature* 331:163–166.

Rensink Ra, Cavanagh P (2004) The influence of cast shadows on visual search. *Perception* 33:1339–1358.

Reynolds JH, Heeger DJ (2009) The Normalization Model of Attention. *Neuron* 61:168–185.

Richardson MJE (2004) Effects of synaptic conductance on the voltage distribution and firing rate of spiking neurons. *Physical Review E* 69:051918.

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nature neuroscience* 2:1019–25.

Rodríguez-Sánchez AJ, Tsotsos JK (2012) The Roles of Endstopped and Curvature Tuned Computations in a Hierarchical Representation of 2D Shape. *PloS One* 7:e42058.

Rubin DB, VanHooser SD, Miller KD (2015) The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* 85:402–417.

Ruff Da, Cohen MR (2014) Attention can either increase or decrease spike count correlations in visual cortex. *Nature neuroscience* 17:1591–7.

Rust NC, Dicarlo JJ (2010) Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. *J Neurosci* 30:12978–95.

Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nature neuroscience* 9:1421–1431.

Sato TK, Haider B, Häusser M, Carandini M (2016) An excitatory basis for divisive normalization in visual cortex. *Nature neuroscience* 19:2–7.

Sato TK, Häusser M, Carandini M (2014) Distal connectivity causes summation and division across mouse visual cortex. *Nature neuroscience* 17:30–2.

Schiller PH, Finlay BL, Volman SF (1976) Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *J Neurophysiol* 39:1288–319.

Sebastian S, Burge J, Geisler WS (2015) Defocus blur discrimination in natural images with natural optics. *J Vision* 15:16.

Serre T, Kouh M, Cadieu C, Knoblich U, Kreiman G, Poggio TA (2005) A Theory of Object Recognition : Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex. *CBCL Paper* #259/AI Me.

Shushruth S, Mangapathy P, Ichida JM, Bressloff PC, Schwabe L, Angelucci A (2012) Strong Recurrent Networks Compute the Orientation Tuning of Surround Modulation in the Primate Primary Visual Cortex. *J Neurosci* 32:308–321.

Somers DC, Todorov EV, Siapas AG, Toth LJ, Kim DS, Sur M (1998) A local circuit approach to understanding integration of long-range inputs in primary visual cortex. *Cerebral Cortex* 8:204–217.

Sun JY, Perona P (1996) Preattentive perception of elementary three-dimensional shapes. *Vision Research* 36:2515–2529.

Tao X, Zhang B, Smith EL, Nishimoto S, Ohzawa I, Chino YM (2012) Local sensitivity to stimulus orientation and spatial frequency within the receptive fields of neurons in visual area 2 of macaque monkeys. *J Neurophysiol* 107:1094–1110.

Torre V, Poggio TA (1978) A Synaptic Mechanism Possibly Underlying Directional Selectivity to Motion. *Proceedings of the Royal Society B: Biological Sciences* 202:409–416.

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RBH (2003) Faces and objects in macaque cerebral cortex. *Nature neuroscience* 6:989–995.

Vinch B (2013) Structured Hierarchical Models for Neurons in the Early Visual System Dissertation, New York University.

Wagemans J, Elder JH, Kubovy M, Palmer SE, Peterson Ma, Singh M, von der Heydt R (2012) A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin* 138:1172–217.

Watson AB, Ahumada AJ (2011) Blur clarified: a review and synthesis of blur discrimination. *J Vision* 11:10–10.

Willmore BDB, Prenger RJ, Gallant JL (2010) Neural representation of natural images in visual area V2. *J Neurosci* 30:2102–14.

Yu AJ, Giese MA, Poggio TA (2002) Biophysiologically plausible implementations of the maximum operation. *Neural computation* 14:2857–81.

Zhou H, Friedman HS, von der Heydt R (2000) Coding of border ownership in monkey visual cortex. *J Neurosci* 20:6594–611.

Zylberberg J, Murphy JT, DeWeese MR (2011) A Sparse Coding Model with Synaptically Local Plasticity and Spiking Neurons Can Account for the Diverse Shapes of V1 Simple Cell Receptive Fields. *PLoS Computational Biology* 7:e1002250.

Appendix A

SUPPORTING INFORMATION

A.1 Supplementary figures

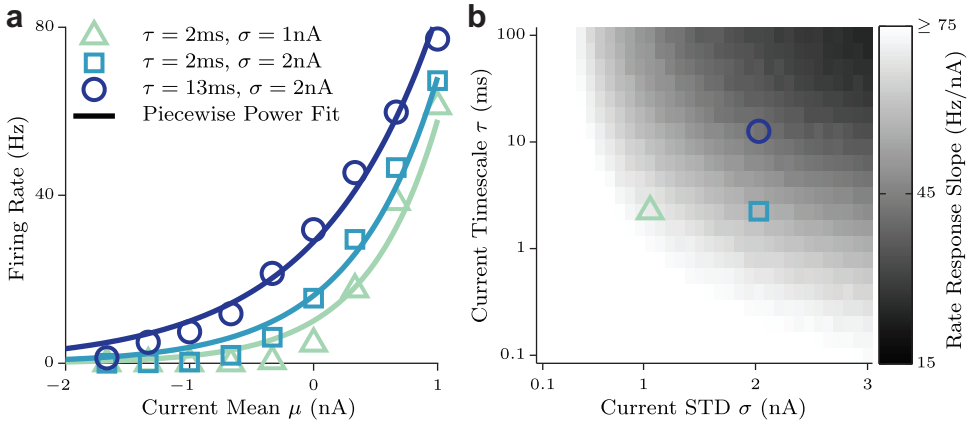

Figure A.1: **Response gain estimated from a piecewise-power fit.** (a) Rate response of a current-based LIF neuron for multiple levels of stochastic background current variance and timescale, along with a piecewise-power fit, *i.e.* $\lfloor \mu^a + b \rfloor$, to assess gain of this response. (b) We find that the gain of a neuron's response to tonic input, computed from the slope of the fit at median response (40 Hz), decreases as either current variance or timescale is increased, qualitatively identical to the piecewise-linear fit of Fig. 4.3.

A.2 Derivation of the generalized LIF model rate response

The following is an analytic method developed by Brunel *et al.* (see Brunel and Sergi, 1998 for a similar derivation). Here we demonstrate the key steps to approximating the firing rate of the general LIF system described by (4.10). The associated equilibrium Fokker-Planck equation for the distribution P of voltage V and input $z = ks$, written with timescale ratio

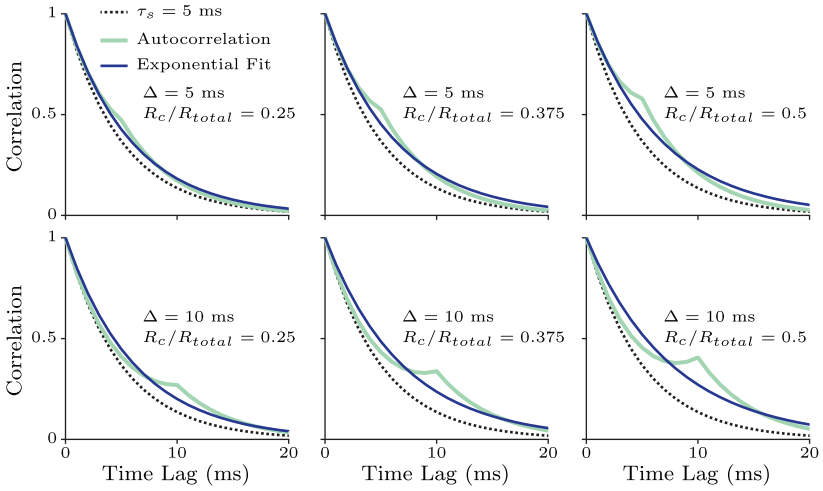

Figure A.2: **Effective timescale of correlated synaptic activity.** For fixed synaptic timeconstant $\tau_s = 5$ ms and delay Δ, increasing the proportion of correlated synaptic events R_c/R_{total} shapes autocorrelation structure. Simulated autocorrelation of correlated synaptic activity (teal) is approximated via least-squares fit of a decaying exponential (blue), compared against the intrinsic synaptic autocorrelation structure (black). Rows correspond to $\Delta = 5$ and 10 ms, and columns correspond to $R_c/R_{total} = 0.25, 0.375$, and 0.5, respectively.

parameter $k = \sqrt{\frac{\tau_s}{\tau_m}}$, is given by Gardiner (2009)

$$\mathcal{L}P - kz\frac{\partial}{\partial V}(B(V)P) - k^2\frac{\partial}{\partial V}(A(V)P) = 0, \tag{A.1}$$

with the differential operator $\mathcal{L}$ defined as

$$\mathcal{L}P = \frac{1}{2}\frac{\partial^2 P}{\partial z^2} + \frac{\partial}{\partial z}(zP). \tag{A.2}$$

The probability flux in voltage V is therefore

$$J_V = \frac{1}{\tau_m}\left(A(V) + B(V)\frac{z}{k}\right)P, \tag{A.3}$$

which cannot be negative at spiking threshold $V = V_{th}$, giving rise to the boundary conditions

$$\begin{aligned} P(V_{th}, z) &= 0 \quad z < -k\frac{A(V_{th})}{B(V_{th})} \\ P(V_{th}, z) &\geq 0 \quad z > -k\frac{A(V_{th})}{B(V_{th})}. \end{aligned} \tag{A.4}$$

The strategy is to find solutions in boundary layers, as in Hagan et al. (1989) and Fourcaud and Brunel (2002). We compute the solution in three regions: (i) in the *outer* region far from both threshold and reset, (ii) in the *threshold layer* when V is close to emitting a spike, and (iii) in the *reset layer* when V is close to its refractory reset potential.

A.2.1 Outer solution

The outer solution, far from reset and threshold, is obtained by expanding the probability distribution P in powers of k, *i.e.* $P = P_0 + kP_1 + k^2P_2 + \ldots$. Substituting this expansion into (A.1), we find a recurrence relations for the distribution terms P_i given by

$$\begin{aligned} \mathcal{L}P_0 &= 0 \\ \mathcal{L}P_1 &= z\frac{\partial}{\partial V}(BP_0) \\ \mathcal{L}P_2 &= z\frac{\partial}{\partial V}(BP_1) + \frac{\partial}{\partial V}(AP_0), \\ &\ldots \end{aligned} \tag{A.5}$$

which leads to

$$\begin{aligned} P_0 &= \frac{e^{-z^2}}{\sqrt{\pi}}Q_0(V) \\ P_1 &= \frac{e^{-z^2}}{\sqrt{\pi}}Q_1(V) - \frac{ze^{-z^2}}{\sqrt{\pi}}\frac{\partial}{\partial V}(BQ_0) \\ P_2 &= \frac{e^{-z^2}}{\sqrt{\pi}}Q_2(V) - \frac{ze^{-z^2}}{\sqrt{\pi}}\frac{\partial}{\partial V}(BQ_1). \end{aligned} \tag{A.6}$$

To find a solution for P_2 that satisfies the boundary conditions on z we need to impose the condition

$$\frac{1}{2}\frac{\partial}{\partial V}\left(B(V)\frac{\partial(B(V)Q_0)}{\partial V}\right) - \frac{\partial(A(V)Q_0)}{\partial V} = 0, \tag{A.7}$$

which can be solved to reveal a solution for Q_0. As expected, (A.7) coincides with the Fokker-Plank equation in the white noise limit using Stratonovich calculus. Going to third

order, we find that Q_1 obeys (A.7) as well. Thus,

$$\begin{aligned} Q_0(V) &= \begin{cases} \alpha_0 R(V) & V < V_{re} \\ \beta_0 R(V) + \gamma_0 S(V) & V > V_{re} \end{cases} \\ Q_1(V) &= \begin{cases} \alpha_1 R(V) & V < V_{re} \\ \beta_1 R(V) + \gamma_1 S(V) & V > V_{re} \end{cases} \end{aligned} \tag{A.8}$$

for the voltage reset potential V_{re}, where

$$\begin{aligned} R(V) &= \frac{W(V)}{B(V)} \\ S(V) &= \frac{W(V)}{B(V)} \int_V^{V^t} \frac{du}{B(u)W(u)} \end{aligned} \tag{A.9}$$

for W given in (4.13). Furthermore, the solutions Q_0 and Q_1 have to obey the normalization conditions

$$\begin{aligned} \int Q_0(V) dV &= 1 \\ \int Q_1(V) dV &= 0. \end{aligned} \tag{A.10}$$

A.2.2 Inner solutions

Solutions to the inner threshold and reset layers are found using similar techniques. To construct solutions within the threshold layer, we need to rescale the voltage as $V = V_{th} - kxV_{th}$. It will also be convenient to define a substitution

$$z' = z + k\frac{A(V_{th})}{B(V_{th})}, \tag{A.11}$$

simplifying the boundary conditions and operator $\mathcal{L}'$. Therefore, rewriting (A.1) for the threshold distribution P^T, we have

$$\mathcal{L}' P^T + z' \frac{\partial P^T}{\partial x} - k\left(\frac{A(V_{th})}{B(V_{th})}\frac{\partial P^T}{\partial z'} + \frac{B(V_{th})}{B(V_{th})} z' \frac{\partial}{\partial x}(xP^T)\right) + O(k^2) = 0, \tag{A.12}$$

which again can be solved by expanding $P^T = P_o^T + kP_1^T + \ldots$, satisfying the boundary condition $P^T(0, z') = 0$ for $z' < 0$. The probability flux at $(0, z')$ is given by

$z'P^T(0,z')B(V_{th})/(\tau_m k)$, implying $P_0^T = 0$. The firing rate terms at zero and first orders are therefore

$$\begin{aligned}\nu_0 &= \frac{B(V_{th})}{\tau_m}\int_0^\infty z' P_1^T(0,z')dz' \\ \nu_1 &= \frac{B(V_{th})}{\tau_m}\int_0^\infty z' P_2^T(0,z')dz'.\end{aligned} \tag{A.13}$$

A solution to P_1^T has previously been found by Hagan et al. (1989) to be

$$P_1^T = \frac{e^{-z'^2}}{\sqrt{\pi}}\rho_1^T\left(\tilde{\alpha} + x + z + U(x,z)\right). \tag{A.14}$$

While specifics of $\tilde{\alpha}$ and $U(x,z)$ are described in detail in derivations of Hagan et al. (1989), since U decays exponentially to zero for large x, and since $\int z e^{-z'^2}U(x,z) = 0$, we conclude that $\rho_1^T = 2\nu_0\tau_m/B(V_{th})$ (Hagan et al., 1989).

The reset layer can be dealt with exactly in the same way as the threshold layer. One finds the solutions to the left and right of the reset, *i.e.* V_{re}^- and V_{re}^+, coincide at zero order, but that the difference between these solutions obeys (A.14).

A.2.3 Matching outer and inner layers

To match the outer and threshold layers we use the change of variables $V = V_{th} - kxB(V_{th})$, $z' = z + kA(V_{th})/B(V_{th})$. The outer solution becomes

$$P(x,z') = \frac{e^{-z'^2}}{\sqrt{\pi}}\left(Q_0(V_{th}) + k\left[Q_1(V_{th}) - xB(V_{th})Q_0'(V_{th}) + \frac{z'\gamma_0}{B(V_{th})}\right]\right) \tag{A.15}$$

This solution must match P_1^T in the large x limit. Hence, we have $Q_0(V_{th}) = 0$ which implies $\beta_0 = 0$. We also have

$$\begin{aligned}\gamma_0 &= 2\nu_0\tau_m \\ Q_1(V_{th}) &= \tilde{\alpha}\rho_1^T = \frac{2\tilde{\alpha}\nu_0\tau_m}{B(V_{th})},\end{aligned} \tag{A.16}$$

which leads to

$$\beta_1 = \frac{2\tilde{\alpha}\nu_0\tau_m}{W(V_{th})}. \tag{A.17}$$

Matching of the reset and outer layers is done in a similar way. One find that Q_0 has to be continuous in V_{re}, implying

$$\alpha_0 = \int_{V_{re}}^{V_{th}} \frac{du}{B(u)W(u)} \tag{A.18}$$

which, together with the normalization condition for Q_0, leads to the equation for the zeroth order firing rate ν_0 as expected. One then finds that Q_1 is instead discontinuous in V_{re}, with

$$Q_1(V_{re}^+) - Q_1(V_{re}^-) = \tilde{\rho}_1^T = \beta_1 R(V_{re}) + \gamma_1 S(V_{re}) - \alpha_1 R(V_{re}). \tag{A.19}$$

Including now the normalization condition for Q_1, this gives us two equations for the last two remaining unknowns, γ_1 and α_1. In particular, we find that

$$\gamma_1 = 4\tilde{\alpha}(\nu_0\tau_m)^2 B(V_{th}) \left(\frac{B(V_{re})^2}{B(V_{th})^2}\psi(V_{re}) - \psi(V_{th}) \right) \tag{A.20}$$

where

$$\psi(V) = \frac{1}{B(V)W(V)} \int_{-\infty}^{V} R(u)du \tag{A.21}$$

A.2.4 First order correction to the firing rate

The last step is to compute ν_1. From (A.13) it would seem that we need to compute P_2^T. Fortunately we only need the term proportional to z in this equation, as it is the only term that contributes to the firing rate. Because of the matching condition with the outer solution that term must match with the corresponding term in the outer solution, which is proportional to γ_1. Therefore, the correction is

$$\nu_1 = \frac{\gamma_1}{2\tau_m} = 2\tilde{\alpha}\nu_0^2\tau_m B(V_{th}) \left(\frac{B(V_{re})^2}{B(V_{th})^2}\psi(V_{re}) - \psi(V_{th}) \right). \tag{A.22}$$

Finally, in the $k \ll 1$ limit, the firing rate up to first order (4.12) is found to be consistent with the expression (4.14).

A.3 General LIF approximation is limited to one stochastic conductance source

A conductance-based membrane equation with two sources of input conductance (*i.e.* excitatory and inhibitory synaptic events) takes the form

$$\tau_m \frac{d}{dt} v = -v + (v - e_1)g_1 + (v - e_2)g_2 \tag{A.23}$$

for conductance processes g and reversal potentials e. In order to cast a system of this type into the generalized input timescale formulation,*i.e.*

$$\tau_m \frac{d}{dt} v_t = A(v) + B(v)s_t \tag{A.24}$$

$$\tau_s \frac{d}{dt} s_t = -s + \sqrt{\tau_m}\xi_t, \tag{A.25}$$

we would like to arbitrarily scale and shift each input from standard OU processes s_i of arbitrary timescale, *i.e.*

$$g_i = \mu_i + \sigma_i s_i. \tag{A.26}$$

Expanding the membrane equation we see

$$\tau_m \frac{d}{dt} v = -v + (\mu_1 + \mu_2)v + (-e_1\mu_1 - e_2\mu_2) + (\sigma_1 s_1 + \sigma_2 s_2)v + (-e_1\sigma_1 s_1 - e_2\sigma_2 s_2), \tag{A.27}$$

which can be factored, using the components

$$A(v) = -v + (\mu_1 + \mu_2)v - (e_1\mu_1 + e_2\mu_2) \tag{A.28}$$

$$s = \sigma_1 s_1 + \sigma_2 s_2 \tag{A.29}$$

$$p = -e_1\sigma_1 s_1 - e_2\sigma_2 s_2. \tag{A.30}$$

The problem is that, in general, $s \neq p$. The membrane equation becomes

$$\tau_m \frac{d}{dt} v = A(v) + sv + p, \tag{A.31}$$

which cannot be further reduced into a term of the form $B(v)s$, necessary for the generalized analytic solution. However, an approximation can be made: looking at the statistics of s and p we see both have zero mean and variance of $\sigma_1^2 + \sigma_2^2$ and $e_1^2\sigma_1^2 + e_2^2\sigma_2^2$ respectively. Therefore, we can match the variance with a gain term

$$\alpha^2 = \frac{e_1^2\sigma_1^2 + e_2^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{A.32}$$

and using a matched-variance approximation of p as $-\alpha s$, define $B(v) = v + \alpha$ and write

$$\tau_m \frac{d}{dt} v = A(v) + B(v)s, \tag{A.33}$$

Which is then easy to analytically approximate. The error of this approximation is found by considering

$$\xi = \|\alpha s + p\| = \|\alpha\sigma_1 s + 1 + \alpha\sigma_2 s_2 - e_1\sigma_1 s_1 - e_2\sigma_2 s_2\| \tag{A.34}$$

$$= \|(\alpha - e_1)\sigma_1 s_1 + (\alpha - e_2)\sigma_2 s_2\|. \tag{A.35}$$

Requiring ξ to be zero shows that the only perfect approximation for nonzero σ_1, σ_2 is achieved when $e_1 = e_2 = 0$. However, ξ appears to be proportional to $(e_1 - \alpha)\sigma_1 + (e_2 - \alpha)\sigma_2$, seeming to make an intelligent choice of parameters possible.

A.4 Derivation of a bound on the response of simple neuron models

let f be a stochastic voltage signal defined by an arbitrary autocovariance. We make explicit the parameterization of signal variance σ^2 and autocorrelation h with timescale τ, *i.e.*

$$f(t; \sigma, \tau) = \sigma\xi(t) * h(t; \tau), \tag{A.36}$$

where ξ is zero-mean white noise of unit variance. We assume that the firing rate R of a neuron is proportional to the amount of 'voltage power' above some spiking threshold v_{th}, estimated to be

$$R = \lim_{T\to\infty} \frac{1}{T} \int_0^T |\exp\{f(t; \sigma, \tau) - v_{th}\}|^2 dt \tag{A.37}$$

$$\leq e^{-2v_{th}} \lim_{T\to\infty} \frac{1}{T} \int_0^T |\exp\{f(t; \sigma, \tau)\}|^2 dt \tag{A.38}$$

$$\leq e^{-2v_{th}} \lim_{T\to\infty} \frac{1}{T} \int_0^T \exp\{|f(t; \sigma, \tau)|^2\} dt \tag{A.39}$$

Note the use of exponentiation to approximate half-wave rectification. This is an approach often used, *e.g.* the Generalized Linear Model (GLM), but any convex non-negative non-decreasing function is valid. Next, we define the Fourier transform of f to be

$$\hat{f}(z; \sigma, \tau) = \int_{-\infty}^{\infty} f(t; \sigma, \tau) e^{-itz} dt, \tag{A.40}$$

and using the fact that spectral power is equal to the fourier transform of the autocorrelation, we find

$$|f(t;\sigma,\tau)|^2 = \left|\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(z;\sigma,\tau)e^{itz}dz\right|^2 \tag{A.41}$$

$$\leq \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\left|\hat{f}(z;\sigma,\tau)\right|^2 dz \tag{A.42}$$

$$= \frac{\sigma}{4\pi^2}\int_{-\infty}^{\infty}\hat{h}(z;\tau)dz. \tag{A.43}$$

Assume then that the autocorrelation is a two-sided decaying exponential, *i.e.*

$$h(t,\tau) = \exp\left\{-\frac{t}{\tau}\right\}, \tag{A.44}$$

Which has the Fourier transform

$$\hat{h}(z;\tau) = \frac{2\tau^{-1}}{\tau^{-2}+z^2} \tag{A.45}$$

This is a convenient form that can be easily integrated in z, giving

$$\int_{-\infty}^{\infty}\hat{h}(z;\tau)dz = \left[\arctan\frac{z}{\tau}\right]_{z=-\infty}^{z=\infty} = 2. \tag{A.46}$$

Thus, we find

$$R \leq e^{-2v_{th}}\lim_{T\to\infty}\frac{1}{T}\int_0^T \exp\left\{\frac{\sigma}{2\pi^2}\right\}dt \tag{A.47}$$

$$= e^{\frac{\sigma}{2\pi^2}-2v_{th}}. \tag{A.48}$$

Since the voltage threshold can be interpreted as the negation of signal mean, we preform the substitution

$$\mu = -2v_{th} \tag{A.49}$$

giving the final bound for the firing rate R as

$$R \leq e^{\frac{\sigma}{2\pi^2}+\mu}. \tag{A.50}$$

A surprising finding is that the expression no longer depends on the autocorrelation timescale τ. By calculating the derivative with respect to mean input voltage μ we easily see that

the slope of a firing rate curve, estimated by varying μ invariant to τ, matching numeric simulations.

Suppose instead that the autocorrelation is parameterized as a piecewise-linear tent function, *i.e.* a triangle centered about the origin of unit height and base of length 2τ. Such a function is constructed by convolving two rectangular boxcar functions of width τ. If we first scale time by τ, the rectangular functions have unit width, with the well known transform being a a sinc function, *i.e.*

$$\hat{h}(z;\tau) = \tau \int_{\infty}^{\infty} \text{rect}\left(\frac{t}{\tau}\right) * \text{rect}\left(\frac{t}{\tau}\right) e^{-itz} dt \tag{A.51}$$

$$= \frac{\tau \sin^2(\pi z \tau)}{(\pi z \tau)^2}. \tag{A.52}$$

Integration of this transformed autocorrelation for real z, as in (A.46), evaluates to $\frac{1}{\tau}$, canceling out the previous scaling factor. Thus, the rate bound for a triangular autocovairance is also τ-invariant, an observation confirmed by simulation.

We can leverage utility of the triangular hat function to make a claim on *arbitrary* autocovariance functions. Suppose h' is a smooth continuous symmetric function of finite support. We can then construct an approximation of h' with a piecewise-linear function h, constructed from a series of N appropriately translated and scaled hat functions, *i.e.*

$$h(t, \vec{\sigma}, \vec{\mu}, \vec{\tau}) = \sum_{n=1}^{N} \sigma_n \text{tri}\left(\frac{t - \mu_n}{\tau_n}\right). \tag{A.53}$$

The Fourier transform is thus

$$\hat{h}(z, \vec{\sigma}, \vec{\mu}, \vec{\tau}) = \sum_{n=1}^{N} \sigma_n \frac{\tau_n \sin^2(\pi z \tau_n)}{(\pi z \tau_n)^2} e^{-\mu z}, \tag{A.54}$$

and one can easily see how, by taking another absolute value of the rate, and integrating for real z, the upper bound will be proportional to $\|\vec{\sigma}\|_1$, which is invariant to time-dilation of the autocovariance. Hence, the bound appears to depend on autocovariance shape, and not its underlying mass.

This suggests that simple neural models, such as GLM, cannot utilize the previously described autocorrelation mechanisms of neural computation, and narrows the potential mechanisms responsible for that computation to voltage reset and leak.

However, we note that increasing timescale τ tends to have a suppressive effect on rate response. Thus, this limit may very well be an upper bound on firing rate for white noise, *i.e.* infinitesimal autocorrelation timescale present in white noise. Thus it will be helpful to study the impact of input variance and timescale on the firing rate of GLM neurons, particularly for certain history filters to determine if these models can replicate observed phenomena of gain modulation.

A.5 Derivation of response gain in a simplified current integrator

An explicit solution can be achieved if instead we assume the unit's firing rate to be proportional to the *time* a signal is above threshold. This is equivalent to setting the spiking nonlinearity to be a unit step function after threshold translation, *i.e.*

$$R = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathcal{U}\left(f(t;\sigma,\tau) - v_{th}\right). \tag{A.55}$$

Tis expression can be interpreted as collapsing the signal into a stationary probability distribution and integrating it to threshold. We assume that f has zero mean, and seek to compute its variance. Naturally, as the variance increases, the slope of R with respect to signal mean, *i.e.* $-v_{th}$, will decrease.

We begin by assuming that nonlinear aspects of the neuron can be captured solely by a membrane filter $\nu(t)$. The resulting rate expression is

$$R = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathcal{U}\left(I(t;\sigma,\tau) * \nu(t) - I_{th}\right) dt. \tag{A.56}$$

for response R to threshold I_{th}. We assume I to again be a stationary process having zero mean, *i.e.*

$$I(t;\sigma,\tau) = \sigma\xi(t) * h(t;\tau), \quad \text{where } \int_{-\infty}^{\infty} |h(t;\tau)|^2 dt = 1, \tag{A.57}$$

where ξ is unit white noise. Therefore, R can be computed simply from the probability distribution of the input current convolved with the membrane filter. We consider the probability distribution P of the stationary random variable $I(t;\sigma,\tau) * (\nu)$, and compute the variance

$$\begin{aligned}
\mathrm{Var}[P] &= \int_{-\infty}^{\infty} |I(t;\sigma,\tau) * \nu(t)|^2 dt && && \text{(A.58)}\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} |\hat{I}(\omega;\sigma,\tau)\cdot\hat{\nu}(\omega)|^2 d\omega && \text{Parseval's theorem} && \text{(A.59)}\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} |\sigma\hat{\xi}(\omega)\cdot\hat{h}(\omega;\tau)\cdot\hat{\nu}(\omega)|^2 d\omega && \text{convolution property} && \text{(A.60)}\\
&= \frac{\sigma^2}{2\pi}\int_{-\infty}^{\infty} |\hat{h}(\omega;\tau)\cdot\hat{\nu}(\omega)|^2 d\omega. && \xi(t) \text{ is unit white noise} && \text{(A.61)}
\end{aligned}$$

Here we see that the variance of the driving current, that is the input current after membrane filtering, depends on the resonance between the input statistics and membrane filter in the frequency domain. To further understand this effect, we establish $\nu(t)$ as a biophysically plausible membrane filter. Specifically, if a current integrator V is modeled as leaky, the driving signal will observe

$$\tau_\nu \frac{\partial}{\partial t} V = -V + I, \tag{A.62}$$

which can be easily solved with Green's functions and principle of superposition, giving

$$V(t) = \mathcal{U}(t)\exp\left\{-\frac{t}{\tau_\nu}\right\} * I(t;\sigma,\tau). \tag{A.63}$$

Thus, the leaky current integrator model has the effect of low-pass filtering the input current. We define $\nu(t;\tau_\nu)$ to be a decaying exponential with timescale τ_ν and calculate exactly the effect on the current integrator's gain as a function of input timescale. Assuming the input current to have the same autocorrelation shape of timescale τ while observing (A.57), the input and membrane filters have the Fourier transforms

$$\hat{h}(\omega;\tau) = \frac{\sqrt{2\tau}}{i\tau\omega + 1} \quad \text{and} \quad \hat{\nu}(\omega;\tau_\nu) = \frac{\tau_\nu}{i\tau_\nu\omega + 1}. \tag{A.64}$$

The variance of P is then calculated to be

$$\mathrm{Var}[P] = \frac{\sigma^2}{2\pi}\int_{-\infty}^{\infty}\left|\frac{\sqrt{2\tau}}{i\tau\omega+1}\cdot\frac{\tau_\nu}{i\tau_\nu\omega+1}\right|^2 d\omega \tag{A.65}$$

$$= \sigma^2\frac{\tau}{\tau+\tau_\nu}, \tag{A.66}$$

which can be seen to increase with input variance σ^2 or timescale τ.

A.6 Derivation of shaping autocorrelation timescale via correlated activity

We conder the system of stochastic differential equations

$$\begin{cases} \tau dI_c(t) = (-I_c + R_c)dt + \sqrt{2\tau}R_c dW_t^c \\ \tau dI_n(t) = (-I_n + R_n - R_c + I_c)dt + \sqrt{2\tau}(R_n - R_c + I_c)dW_t^n \\ I_{total} = I_e + I_i \end{cases} \tag{A.67}$$

for $n \in \{e, i\}$ to replicated correlated drive to excitatory and inhibitory neural populations. Here, the correlated population is driven by a tonic rate R_c. The excitatory and inhibitory populations are then driven by the activity of I_c, and independent drive $Rn - Rc$ such that the net mean activity to I_e an I_i is R_e and R_i, respectively.

It is simple to solve for the activity of I_c for arbitrary time t, *i.e.*

$$I_c(t) = I_c(t_0)e^{-t/\tau} + R_c(1 - e^{-t/\tau}) + e^{-t/\tau}\int_{t_0}^{t}\sqrt{2/\tau}R_c e^{u/\tau}dW_u^c, \tag{A.68}$$

using the substitution $k = I_c(t)e^{t/\tau}$. For solutions to I_n we can use a similar substitution for k to yield

$$\frac{d}{dt}k = \frac{d}{dt}I_n e^{t/\tau} + \frac{1}{\tau}e^{t/\tau}I_n, \tag{A.69}$$

which leads to

$$\begin{aligned} I_n = {} & I_n(t_0)e^{(t_0-t)/\tau} + (R_n - R_c)(1 - e^{(t_0-t)/\tau} + \sqrt{2/\tau}e^{-t/\tau}(R_n - R_c)\int_{t_0}^{t} e^{u/\tau}dW_u^n \ldots \\ & + \frac{1}{\tau}e^{-t/\tau}\int_{t_0}^{t} I_c e^{u/\tau}du + \sqrt{2/\tau}e^{-t/\tau}\int_{t_0}^{t} I_c e^{u/\tau}dW_u^n. \end{aligned} \tag{A.70}$$

We first assume $t_0 = 0$, $I_c(0) = R_c$, and $I_n(0) = R_n$ to examine the system in the steady state. This leads to the simplification, after substitution for I_c, of

$$\begin{aligned} I_n = R_n + \frac{\sqrt{2}}{\sqrt{\tau}} e^{-t/\tau} R_n \int_0^t e^{u/\tau} dW_u^n - \frac{\sqrt{2}}{\sqrt{\tau}} \frac{R_c}{\tau} e^{-t/\tau} \int_0^t v e^{v/\tau} dW_v^c \\ + \frac{\sqrt{2}}{\sqrt{\tau}} \frac{R_c}{\tau} e^{-t} t \int_0^t e^{v/\tau} dW_v^c + \frac{2}{\tau} R_c e^{-t/\tau} \int_0^t \int_0^u e^{v/\tau} dW_v^c dW_u^n. \end{aligned} \tag{A.71}$$

We then wish to calculate the autocovariance of $I_{total} = I_e + I_i$, which has the disadvantage of possessing many non-obvious terms. However, by using symbolic solvers we can conclude I_{total} has variance

$$\sigma_{total}^2 = R_e^2 + R_i^2 + 4R_c^2, \tag{A.72}$$

and autocorrelation at lag δ of

$$e^{-\delta/\tau} \left(1 + 2 \frac{\delta}{\tau} \frac{R_c^2}{\sigma_{total}^2} \right). \tag{A.73}$$

Thus we see that as correlation R_c is increased, so too is effective autocorrelation timescale of the resulting combined process I_{total}.

Appendix B

FUNDING AND COLLABORATION

T.D.O. was funded in part by the National Institutes of Health (Computational Neuroscience Training Grant, 5R90DA033461-03) and by the Natural Sciences and Engineering Research Council of Canada (NSERC, PGS-D).

Chapter 2: Spectral receptive fields

This work was funded by the National Eye Institute (R01 EY018839, A.P.), the National Science Foundation (CRCNS grant IIS-1309725, W.B. and A.P.), and the NIH Office of Research Infrastructure Programs (grant RR00166, A.P.).

Chapter 3: Shape and blur in V4

This work was funded by the the National Eye Institute (R01 EY018839, A.P.), NEI Center Core Grant for Vision Research (P30 EY01730) to the University of Washington, and the National Institutes of Health Office of Research Infrastructure Programs (grant P51 OD010425) to the Washington National Primate Research Center. We thank and acknowledge Amy Nowack for partial data collection and animal handling.

Chapter 4: Divisive inhibition

This work was funded by the National Science Foundation (CRCNS grant IIS-1309725, W.B.; grand DMS-1514743, E.S.-B.). T.D.O. was supported partially by the National Eye Institute (R01 EY018839). E.S.-B. acknowledges the Simons Fellowship in Mathematics. We thank and acknowledge Nicolas Brunel for derivation of some analytic methods.

Appendix C

VITA

Education

M.Sc. Applied Mathematics University of Washington, 2013

M.Sc. Computer Science York University, 2010

Thesis: On the Organization and Representation of Natural Planar Shape

Advisor: Dr. James H. Elder

B.Sc. Hons. Mathematics & Computer Science University of Regina, 2007

Graduated with honours *summa cum laude* minoring in philosophy

Awards and scholarship

2013-2014 **Natural Sciences and Engineering Research Council of Canada PGS-D**

NSERC Postgraduate Scholarship held at foreign institution (CAD $42,000)

2011-2012 **Computational Neuroscience Training Grant**

University of Washington (Full Ph.D. Funding)

2008-2010 **Natural Sciences and Engineering Research Council of Canada CGS-M**

Alexander Graham Bell Canadian Graduate Scholarship (CAD $35,000)

2009 **Best Student Poster Award**

GEOIDE Annual Scientific Conference 2009 Poster Session

2008 **York University Entrance Scholarship**

2006-2007 **NSERC Undergraduate Student Research Award**

2003-2007 **Undergraduate Awards**

University of Regina Academic Silver Scholarship, 2005-2007; University of Regina Dean's Honors List 2003-2007; Campion College Dean's Honors List, 2003-2007; Campion College Alpha Sigma Nu Scholarship, 2006; Centennial Merit Scholarship, 2003