Spatial frequency selectivity in macaque LGN and V1

by

Paul Levy

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Center for Neural Science

New York University

May 2023

J. Anthony Movshon

Eero P. Simoncelli

Acknowledgements

I would like to thank my PhD advisors, Tony Movshon and Eero Simoncelli, for many years of intellectual, moral, and institutional support. They guided my scientific development in countless ways: through close, direct feedback on this work; their permission of numerous excursions in my research; and in their own actions as members of the broader neuroscience community. In particular, their emphasis on clear, approachable scientific communication (and the strong understanding of the work that's required for that) continues to be an inspiration. Across many late nights, close deadlines, and points of confusion, they provided succinct and timely advice to enable my progress.

I also owe immense thanks to the remaining members of my advisory committee, Mike Hawken and Roozbeh Kiani. Much of the work in this thesis is built upon Mike's research, and I have learned a great deal from his emphasis on understanding the mechanisms and circuitry which give rise to the phenomena discussed here. He was also integral in the data collection process, where he shared with me and my colleagues countless tricks and techniques for conducting effective physiology experiments. During my rotation in his lab, Roozbeh's advice helped my scientific development immensely. Besides teaching me how to do psychophysics, he showed me how to write papers and organize the long timescale of research into manageable blocks.

I've been fortunate to witness different eras in both of my labs. In the Movshon-Kiorpes labs, I was lucky to have my earliest scientific explorations guided by Robbe Goris (who helped kick off this project), Romesh Kumbhani, Najib Majaj, Luke Hallum, and Christopher Shooner. None of the research in this thesis could exist without the incredible acute physiology and histology team: namely, Manu Raghavan, Jenna Kelly, and Mike Hasse, who led the bulk of the experiments described in this work. We spent many long, sleep-deprived weeks together on a mix of problems both difficult and strange. The earlier and later experiments here are also thanks to Luke, Christopher, Robbe, Corey Ziemba, and Sonica Saraf. I also owe many

thanks to my office-mate of many years, Tim Oleskiw, who has patiently advised me on a variety of scientific and modeling questions. The members of Eero's lab have inspired me with their important computational work and guided me through many puzzling programming issues. In particular, a lot of my education in computational neuroscience was helped by sharing space (and coffee) with Corey, Andrew Zaharia, Olivier Henaff, Alex Berardino, Kathryn Bonnen, Billy Broderick, and Zhuo Wang.

The NYU neuroscience community is a great place to grow as a scientist, critical thinker, and human being. The vision community, specifically, includes so many great scientists whose work provides the foundation for my own. My fellow cohort, in particular Gerick Lee, Ben Lu, and Saleh Esteki, made this PhD far more enjoyable and only encouraged my obsession with all things Queens. The department and program administration - especially Amala Ankolekar-Hinge, Jess Holman, Erik Cruz, and Heather McKellar - support the students in countless, important ways. The veterinary staff at CNS provided essential knowledge and support for this research. I especially want to thank Lee-Ronn Paluch and Ismael Quiñones for their close work with our acute experiments. The CNS tech support, especially Paul Fan, provided me with endless, timely support - even at odd hours! The dedicated staff of the NYU High Performance Computing team, especially Shenglong Wang, helped me troubleshoot many questions regarding the essential computing resources they provide for the research community.

Finally, I want to thank my family for always encouraging my curiosity. My parents, Richard and Geri, taught me the importance of reading, thinking, exploring, and looking up to educators. My sister, Alessandra, gave me my earliest math lessons and helped lead me to the University of Miami. I also want to thank my Aunt Diane for always being there with me in NYC, including my first few weeks at NYU when we were roommates. Our family time together - more recently with Andrea, too - for museum visits, playing music, and good Italian food has been invaluable. I had the unbelievable fortune of meeting my wife, Ash, in the first year of our PhDs. Together, we grew as people and scientists, and I couldn't have

finished this PhD without your love and support. You led by example, finishing your PhD (well) before me, and inspired me with how you managed your own challenging, international research. Ash's family has so thoroughly welcomed me into their own, and I thank Mumsa, Baba, and Aditya for making me feel at home wherever we are.

Abstract

Systems neuroscientists seek a mechanistic and computational understanding of neural activity. In the visual system, the jump from sub-cortical to cortical brings about many important changes in representation and circuitry. By focusing on the important transformation of spatial information between thalamus and cortex, this work provides a better understanding of the computations that underlie visual processing. We performed a series of experiments in anesthetized primates, recording from individual neurons in the lateral geniculate nucleus (LGN) of the thalamus and the primary visual cortex (V1) using single grating stimuli and mixtures of gratings. To make sense of these different recordings, we fit the measured responses to both mechanistic and more computational models.

In the first chapter, we bring together previous accounts - in both the LGN and in V1 - of shifts in spatial frequency tuning with image contrast. We use a common stimulus set comprised of sinusoidal gratings that vary in spatial frequency and contrast. Fitting canonical, mechanistic models which capture our understanding of each area's receptive field structure, we show that the tuning shifts in V1 are larger than those in the LGN. This result suggests that shifts in LGN selectivity are inherited in V1, but further intracortical processing contributes to the more pronounced tuning shifts.

In the second and third chapters, we turn our focus to stimuli of intermediate complexity. We used superimposed mixtures of gratings as well as a more direct masking experiment to measure the tuning of spatial frequency suppression. In the second chapter, we report stronger spatial frequency-dependent suppression in V1 than in LGN, and find that suppression is typically strongest for frequencies at or below the cell's preference. These stimulus sets were also designed to evoke a broad range of responses which help constrain our computational model of spatial frequency selectivity. In the third chapter, we fit this model to the observed neuronal responses. The model implements divisive normalization, a canonical computation in cortical processing that accounts for a wide variety of observed neural activity. In the

standard model of normalization, the response of a given neuron is normalized by the activity of nearby neurons that are selective across a wide range of stimulus values and features. We show that introducing a spatial-frequency tuned weighting of the normalization signal can preserve gain control while also better accounting for shifts in spatial frequency tuning and the observed suppression to complex stimuli. The tuning of the normalization was typically found to be stronger for frequencies below the cell's peak, highlighting the role of low frequency suppression in shaping selectivity.

Contents

List of Figures

List of Tables

Chapter 1

Introduction & General methods

1.1 General introduction

A fundamental goal of visual neuroscience is to understand how - beginning simply with the activation of photoreceptors in the retina – the visual system builds an exquisite set of representations that support an extraordinary range of perceptual capabilities and, in turn, visually-guided actions. Whether tracking a fast-moving baseball through the skies of the afternoon sun or appreciating both the constituent strokes and cohesive whole of a Pointillist painting, all visual processing begins with the activity of the same neural circuits. The early visual system - consisting of the retina, the dorsal lateral geniculate nucleus (LGN) of the thalamus, and primary visual cortex (V1) [23, 26] - provides the initial processing of incoming visual information and serves as the basis upon which more refined or selective representations are developed.

Influential early studies on the early visual system built an understanding that neuronal responses could be described by receptive field models with linear filtering [119, 43, 94, 165, 36] and, in some cases, further non-linear processing [135, 136, 93] to account for properties like response saturation and phase invariance. This powerful framework enabled researchers to understand neuronal responses using simple stimuli rather than complex natural images.

Crucially, experimenters can parametrically vary these stimuli along features of interest to understand what drives neural responses. In particular, all images have information contained at distinct spatial scales (i.e., spatial frequencies), as well as content at a variety of orientations, phases, and luminance values (e.g., [121, 89]). By measuring the response to variations in these stimulus properties, one can uncover the underlying filter properties, or tuning curve, of a cell.

While neurons are already selective for spatial frequency in the retina [43, 10], a marked shift occurs between the LGN, where neurons are weakly selective for lower frequencies [161, 70, 40], and V1, where selectivity sharpens and becomes bandpass [36, 48]. Similarly, stimulus contrast shapes responses from the earliest stages of the visual system, beginning in the retina where changes in contrast alter gain and other important response properties [134]. Thus, linear filtering for spatial frequency [85] and nonlinear gain control are present throughout the early visual system. Much focus has been dedicated to spatial frequency selectivity and gain control in the early visual system, and we seek to unify these findings by characterizing and modeling the interaction between spatial frequency tuning and contrast in both areas with a common stimulus set. We achieve this through a combination of physiology and modeling grounded in existing mechanistic and computational frameworks. In doing so, we add clarity to the picture of which computations give rise to spatial frequency selectivity and gain control in the early visual pathway. More broadly, because of our choice in stimulus properties and areas of study, this thesis work also offers a window into hierarchical sensory processing.

1.1.1 Spatial frequency selectivity

The mammalian visual system, including that of primates, includes circuitry connecting neurons within and across anatomically distinct areas (e.g., [158, 45, 157, 159]). The structure within areas and of the connections between them plays a profound role in shaping responses throughout the visual pathway [45], particularly the early visual system [23, 75]. These initial stages of processing relay incoming visual information downstream (*fig.* 1.1), where further,

specialized processing for different features of the visual world. However, in their role as the gateway to higher-order processing, understanding the computations and filtering performed here is essential to a complete picture of the visual system.

The cells in the output layer of the retina, retinal ganglion cells (RGC), have well-defined regions of visual sensitivity [76] whose profiles are well described by a pair of circularly symmetric, antagonistic Gaussians termed the difference of Gaussians (DoG) model [118, 43, 61, 83]. This simple model also captures the structure of neurons in the LGN, as first demonstrated in cats [145, 38] and then primates [40, 70]. Though LGN responses are not simple relays of their inputs [69], the foundations for their response properties are laid by their retinal afferents. Already in the primate retina, there is parallel processing by two classes of cells with different properties and thalamic targets which project only to parvocellular (P) or magnocellular (M) layers [72, 87]. Among other differences, cells in the parvo pathway have smaller receptive fields [32, 65] and are sensitive, at the same eccentricity, to slightly higher spatial frequencies [70, 40].

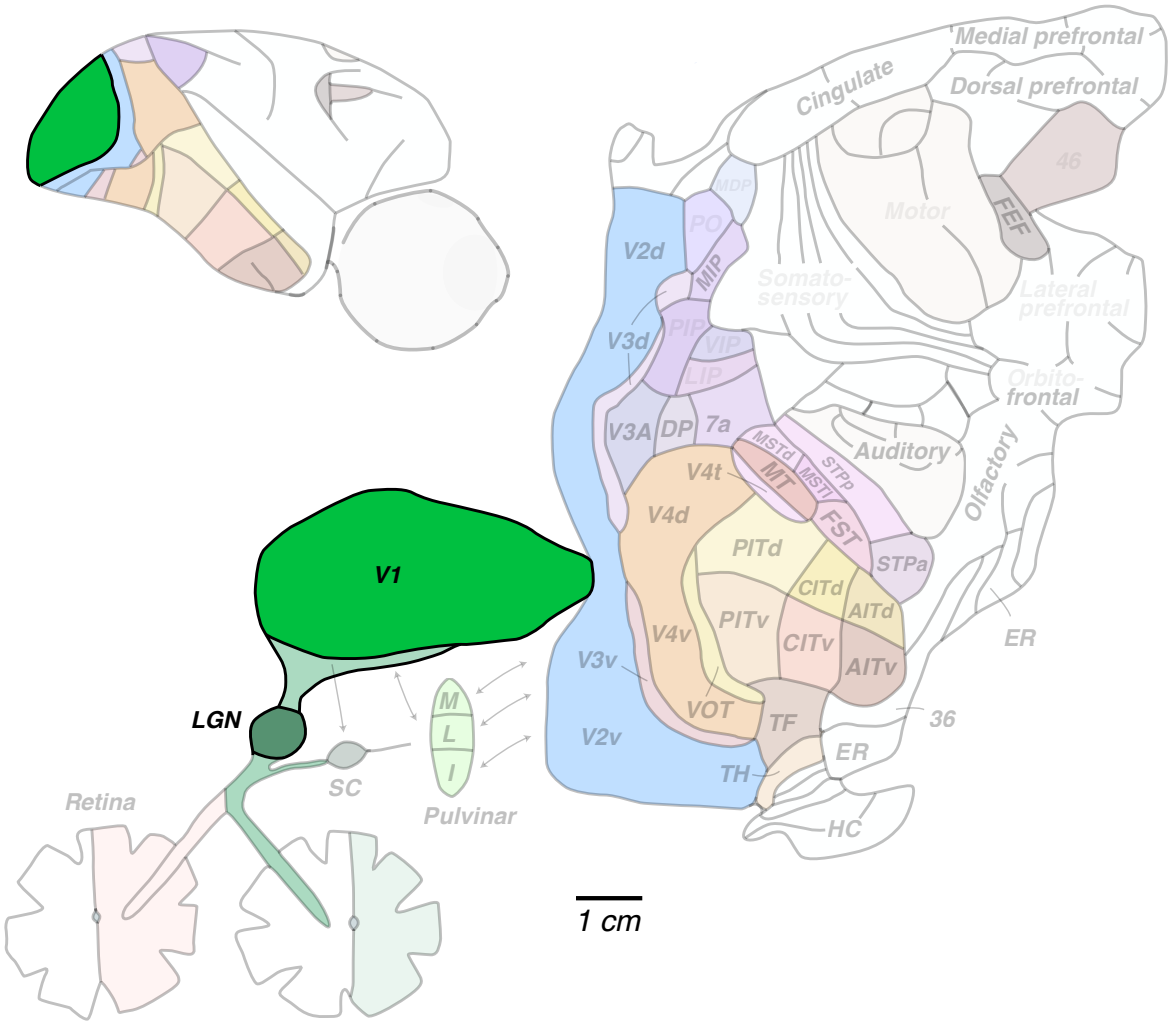

Figure 1.1: A depiction of a flattened macaque brain, adapted from Van Essen et al. [157]. The full-color areas (LGN and V1) are the focus of the this work, but the remaining transparent areas in color are part of the visual system. The inset (*top left*) shows a sagittal view of the intact brain with the eyes facing rightward.

However, the spatial filtering of both magno and parvo cells differ greatly from the tuning found in cortex. Unlike the low-pass, broad tuning curves of the LGN [161], V1 neurons can attain very sharp, bandpass tuning for spatial frequency [36]. Much in the way that the emergence of orientation tuning in cortex [64] demands an account of thalamocortical processing, understanding this marked change in tuning width also motivates a search for mechanism. A pair of early studies in cats [35, 13] focus on suppression as a possible mechanism which gives rise to sharper tuning. The hypothesis is straightforward - rather than relatively

stronger excitation for preferred stimuli, sharper tuning could come about through selective suppression of non-preferred frequencies. These studies differ in some key conclusions (likely due to methodological differences), but both demonstrate that suppression is most pronounced at frequencies away from the cell's preference, providing evidence for the hypothesis that inhibition plays a role in generating sharper cortical selectivity.

Later studies, including those in primates, have clarified suppression's role in shaping tuning. Studies of temporal dynamics show that cortical spatial frequency selectivity is initially broad and becomes narrower only later in the response [21, 91, 100]. Suppression also plays a role in shaping selectivity for orientation [18, 117], size [67, 27, 162, 144], and direction selectivity [109, 166] in V1 (see [66] for a more complete review). Functionally, there is a strong correlation between the measured strength of suppression and the width of tuning for both orientation and spatial frequency [115]. As for the origin of suppression, Webb et al. [162] argue for two distinct suppressive mechanisms: one which is orientation selective, binocular, susceptible to adaptation, acts at high contrast, and has sharp spatiotemporal tuning; and a second which is broadly tuned, monocular, insusceptible to adaptation, and acts at low contrast. Along with others [58, 59, 140], they propose that tuned suppression, which has a smaller spatial profile, is cortical in origin while the broader, untuned suppression arises from feedforward LGN input.

1.1.2 Contrast gain control

We now add a crucial dimension to the discussion of selectivity and suppression in the visual hierarchy: gain control. Linear filtering would suggest that response amplitude scales with stimulus intensity, yet this expectation often fails for both contrast and luminance [138] beginning at the very first area of visual processing. This short timescale adjustment of neuronal gain, which results in response saturation and preserves responsivity over a greater range of stimulus intensity, is termed gain control. Luminance and its local variation (contrast) are uncorrelated in natural images [49, 89], and their influence on gain has been shown to be

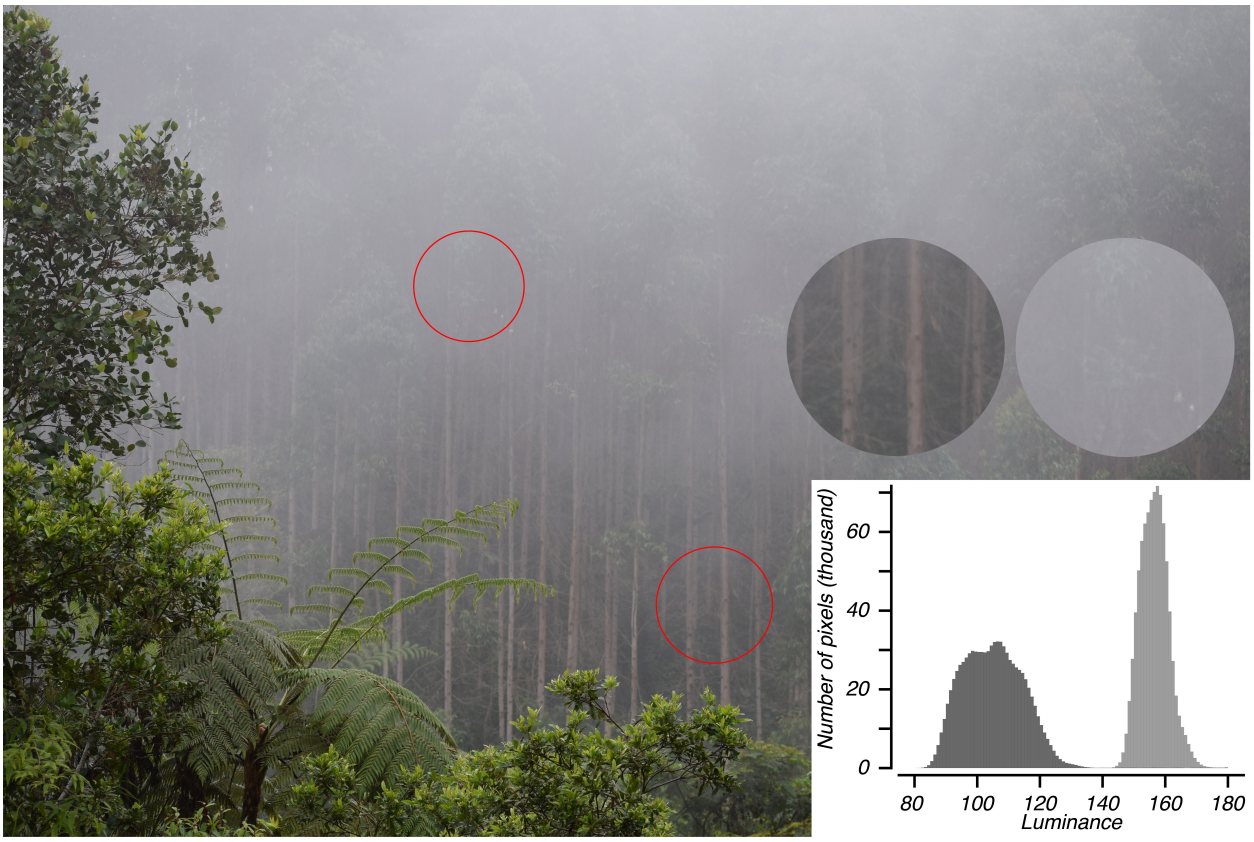

Figure 1.2: A photograph of a foggy, forested scene. The red circles highlight two patches in the background, shown zoomed in above the graph in the bottom right. Beneath the patches are the corresponding distributions of luminance values; the patch (*right*) taken from the upper part of the image has a higher luminance but smaller range, i.e., image contrast, than the other patch. The luminance values (horizontal axis) are the mean of the three RGB channel pixel values, scaled from [0, 255]. If contrast and spatial frequency tuning are non-separable (i.e., gain control influences spatial frequency tuning), then the spatial frequency tuning of an identical neuron "viewing" the two different patches would differ. Photo taken by the author.

separable in cat [89] and primate [112] LGN neurons. To that end, we will only consider the influence of contrast gain control in this work.

A key aim driving our research, then, is to characterize contrast-dependent changes in single neuron selectivity for spatial frequency and relate those findings to gain control mechanisms. To highlight a scenario in which contrast gain could influence selectivity, we show a natural image in *fig.* 1.1.2, with two selected patches differing in their luminance and contrast distributions. An essential aspect of understanding the relationship between contrast gain and frequency selectivity is to identify the origin of the relevant gain mechanisms.

Nonlinear response properties are observed beginning in the retina, and to differing extents in the two primary pathways of the early visual system. As first described in the cat retina, cells can be classified on the basis of their linearity in summation, with X, but not Y, cells responding linearly to combined stimulation of its on (sensitive to light increments) and off (decrements) regions [43]. These same classes differ in their sensitivity to low contrast stimuli and the strength of response saturation at higher contrast [134]. A consequence of this gain control is shorter response latency (phase advance) and changing temporal frequency selectivity with contrast, observed more strongly in Y cells than X [136]. Though not parallel to X and Y cells in cat, there are nonetheless distinctions in the primate LGN between different classes of cells: M cells show a stronger contrast sensitivity and gain [70, 40] and are selective for higher temporal frequencies [40] than P cells. As in cats, primate cells which show strong contrast gain control (namely, M cells) exhibit attenuation in high temporal frequency response with reduced contrast [15, 112]. Furthermore, the extent of contrast-dependent changes can be linked to the strength of gain control, as measured through phase advance [15].

Thus, already in the primate LGN, contrast gain shapes tuning. Though not as thoroughly studied as temporal frequency, LGN neurons also exhibit contrast-dependent spatial frequency [74, 102, 146, 101] and size [102, 19] tuning. Despite marked differences in gain control and suppression strength between M and P cells, contrast-dependent changes in receptive field size [148] and spatial frequency tuning [74, 146] appear to be comparable in the two cell types. This curious discrepancy offers the possibility that temporal frequency attention and contrast-dependent changes in spatial tuning differ in their origin.

Neurons in V1 show many of the same nonlinear response properties as their LGN inputs, including strong gain changes [3], reduction in preferred temporal frequency [110], as well changes in receptive field size [127, 68] with contrast. Precisely because spatial frequency tuning and contrast gain control are present at all levels of the visual hierarchy, identifying the respective roles of LGN and V1 circuitry in determining cortical tuning is difficult and

worth understanding in the context of sensory processing more broadly. For example, the precise spatial arrangement of thalamic inputs into cortex has long been theorized to produce orientation tuning [63]. This hypothesis has been largely borne out: LGN afferents are precisely aligned with the resulting V1 receptive fields [113], and cortical circuitry shapes [60] but doesn't create [46] orientation tuning (though see [141]).

An analogous approach for spatial frequency which focuses only on the arrangement of LGN inputs and the role of cortical processing would be incomplete; no account of cortical spatial frequency tuning can ignore the relevant processing - both linear and nonlinear - that occurs before V1. Briefly, there is plentiful evidence suggesting that cortical suppression and gain control arises from both feedforward and intracortical influences. Pharmacological silencing of intracortical inhibition in cat V1 changed neither the orientation tuning nor strength of surround suppression, suggesting that much of cortical surround suppression is inherited from the LGN [104]. Modeling and physiology in macaque also suggest that cortical suppression, at least for stimulus size, could be inherited from the LGN [154]. However, the tuning of suppression (e.g., its orientation and temporal selectivity) [28, 162, 58] provides strong evidence for a cortical-origin, more tuned suppression [59]. Histological reconstruction of the relevant circuitry [7, 6] and identification of the layer-specific properties of suppression [58, 59] support this argument. Thus, a more nuanced understanding of suppression in V1 has been developed, in which a set of partially overlapped mechanisms - both thalamic and cortical in origin - shape tuning (e.g., see review [8]).

The question of which influences shape V1 spatial frequency tuning is interconnected with the role of gain control. In the context of natural scenes, the wide range in values of spatial frequency [47, 121, 128] and contrast [49, 121, 128] begs the question of whether their influences are separable; that is, does contrast influence spatial frequency tuning. Orientation tuning has been reported as invariant to contrast [131, 20], meaning that its peak and bandwidth remain constant across a wide range of image contrasts. In addition to orientation tuning, spatial frequency tuning was also reported to be contrast invariant [20, 142]. However, follow-up

studies have demonstrated that contrast does influence spatial frequency bandwidth and, to a lesser extent, preferred frequency [126]. Taken together with the marked influence of contrast on size tuning (discussed above), we identify a need in more thoroughly understanding how contrast and contrast gain control shape spatial frequency tuning. Our work characterizes this relationship in the LGN and V1 with common stimuli, permitting a direct comparison in the strength of contrast-dependent spatial frequency tuning in the two areas and adding to our understanding of cascaded selectivity in the early visual system.

1.1.3 Models of V1, receptive field structure

To test whether they understand a system, scientists build models. Receptive field models encapsulate the linear systems approach that has long guided research in visual neuroscience (*fig.* 1.1.3). In the retina and LGN, receptive fields exhibit circular symmetry, with a weaker and more spatially extensive surround that acts in opposition to the stronger, more confined center mechanism - this can be modeled as two partially overlapping Gaussian profiles [43]. Simple cells, so termed because their receptive fields are comprised of separable regions sensitive to light increments ("on") or decrements ("off"), are sensitive to stimulus phase and have long been speculated to arise from the particular arrangement of LGN inputs into V1 [63]. Modeling [114] and physiological efforts [113, 46] have confirmed this hypothesis, and informed V1 receptive field models which are comprised of elongated regions of alternating on/off designation [94, 106]. This structure can also be modeled as either 2D Gabor or Gaussian derivative filters [163], or as arising from LGN-like subunits [53]. Akin to the distinction between X and Y cells, complex cortical cells are insensitive to stimulus phase and are typically modeled as nonlinear combinations of simple cells [63, 93]. In the above receptive field models, the initial response is governed by a linear filter. To better capture measured responses, however, the filter (e.g., Gaussian, Gabor, Gaussian derivative) response is then passed through an output nonlinearity, which maps excitatory drive to spike rate.

The sequential action of a linear filter and output nonlinearity is termed linear-nonlinear (LN) processing.

The LN model framework has proved useful in explaining V1 feature selectivity and receptive field structure [122, 160, 1]. However, left unmodified, LN models cannot account for contrast-invariant tuning, nor contrast gain effects, including experimentally observed changes in threshold with contrast [11] and the changes in filter (as discussed earlier) that underlie contrast-dependent tuning. The incompatibility of LN processing with contrast-invariant tuning, in particular, motivated the development of the normalization model of neuronal responses [55] in which a given neuron is inhibited by a pool of broadly tuned cells [18]. Normalization models have proven remarkably effective in accounting for changes in contrast gain [24], orientation tuning invariance [24, 52], and many other features. However, the role of normalization in accounting for contrast-*invariant* tuning is at odds with contrast-*dependent* tuning (e.g., [126]). To remedy this, researchers have added tuning to the broadly selective normalization signal, allowing for a more effective encapsulation of frequency-specific suppression in vision [99] and audition [130], as well as contrast-dependent changes in size tuning [130].

In our work, we turn our attention to the roles of suppression and gain control in spatial frequency selectivity directly, rather than the oft-studied size tuning. Despite our focus, the close relationship between size tuning and spatial frequency selectivity allows us to draw from the large literature on surround suppression. Specifically, as Fourier analysis suggests, one can derive the spatial frequency tuning curve from receptive field structure and size tuning [94, 151, 150]. A pair of earlier studies showed that V1 neuron receptive field sizes become larger at lower contrast [127, 68], implying that spatial frequency tuning should change with contrast, too; this effect has been observed in primate V1 [126]. However, one cannot exactly derive the spatial frequency tuning curve from size tuning. For one, V1 neurons are often best modeled as comprising of multiple subunits [160, 122] which complicates the translation between receptive field size and spatial frequency selectivity. Additionally, contrast-dependent

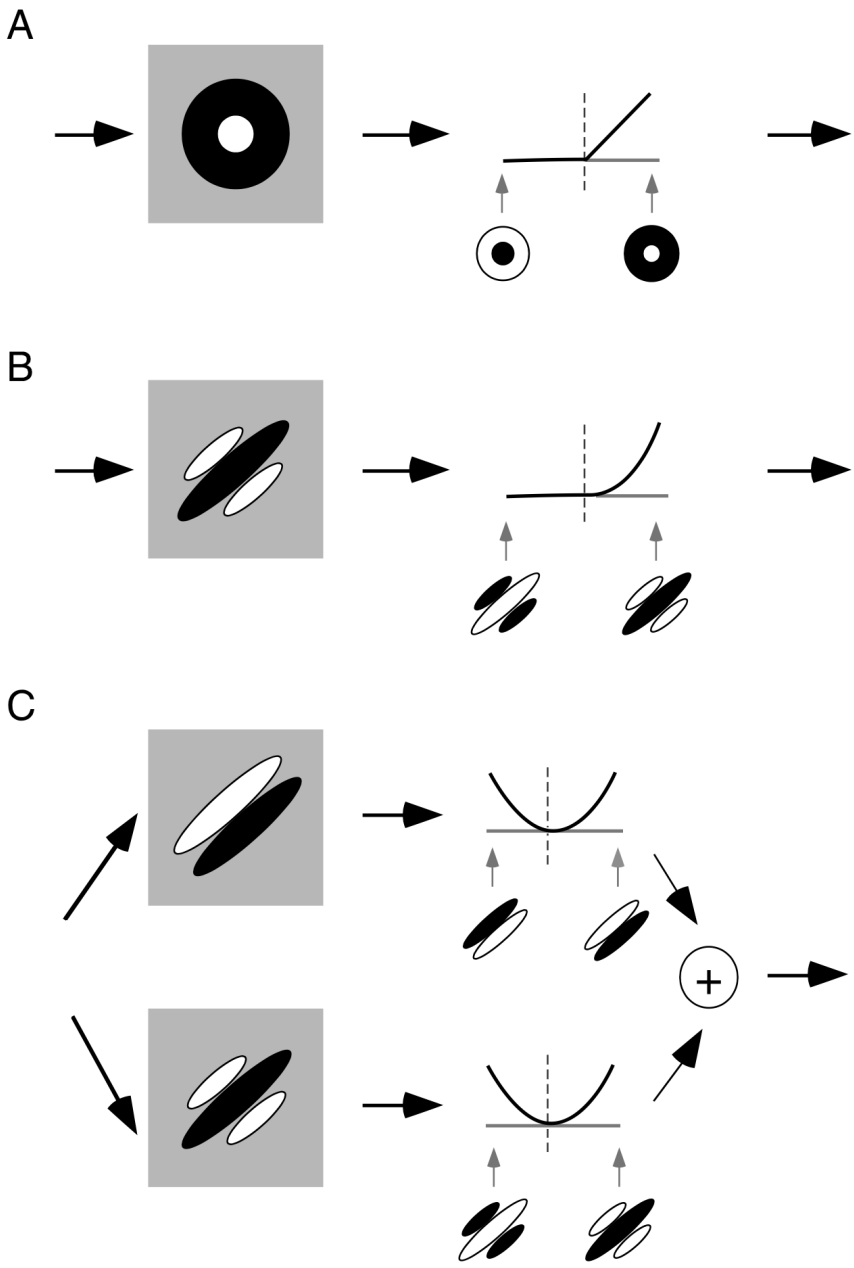

Figure 1.3: Taken from Carandini et al. [26], schematics for models of center-surround retinal ganglion cells (A), elongated V1 simple cells (B), and phase-insensitive V1 complex cells (C). In each case, the left column indicates the receptive field structure (linear filter), while the right column shows the output nonlinearity.

changes in spatial frequency tuning are more modest than changes in size tuning [127, 126], and those changes are asymmetric. Contrast induces greater changes on high spatial frequency selectivity than low, which Sceniak et al. postulate "might result from increases in the subunit size as well as an increase in the number of subunits within the receptive field" [126]. A close study of changes in spatial frequency tuning with contrast can provide greater context for these hypotheses.

In this work, we use two modeling approaches. The first is based on the difference-of-Gaussians model [43] and its cortical extension [53]. These parsimonious, mechanistic models allow the direct readout of receptive field subunit size as a function of contrast, facilitating comparisons with previous studies and between LGN and V1 properties. In the context of contrast gain control, though, these models are more limited in that they don't explicitly account for contrast. To better evaluate the role gain control spatial frequency tuning, we also implement a second, more computational model based on divisive normalization [55]. However, informed by prior work and our own findings of contrast-dependent spatial frequency tuning, we permit the gain control to be tuned as a function of spatial frequency. The model is simple - one excitatory filter and a few parameters governing the strength and tuning of normalization - yet allows us to account for responses across a range of image contrast and complexity with a set of fixed computations.

There are limitations, however, to both our modeling and electrophysiological approaches. The gain control model implemented doesn't directly identify or implicate particular thalamocortical or corticocortical circuitry. More mechanistic models, typically circuit-based, are able to capture gain control through the excitatory and inhibitory interactions of neurons [156, 56, 30]. Though none of these models have explicitly tackled the relationship between contrast gain and spatial frequency tuning, their consideration of the known connectivity and intra-network dynamics allows for a more direct link to specific mechanisms. Physiologically, the primary limitation of our approach is in their reliance on comparing LGN and V1 via response properties rather than through simultaneous recordings. Though the latter approach

would be experimentally difficult, it would help with a more direct identification of the time course and extent of LGN influence on V1 tuning.

1.1.4 Thesis overview

This rich tapestry of related prior research - the influence of suppression and gain control on tuning; the relationship between spatial frequency tuning and receptive field structure; the myriad models of cortical selectivity - provides the critical context to the work of this thesis. Critically, there is not yet a clear picture of how gain control mechanisms and thalamic influences shape V1 spatial frequency tuning. In short, there are several questions which we address and disparate findings which we attempt to unify using a set of informative stimuli, simple mechanistic models, and a tuned gain control model. First, we provide a direct, thorough comparison of the influence of contrast on spatial frequency tuning in LGN and V1. In doing so, we add to a body of literature which attempts to identify thalamic influences on cortical tuning. Furthermore, because our experiments widely sample contrast and go beyond single or paired sinusoidal gratings, we are able to identify the hallmarks of contrast gain control. Since our stimuli were designed with a model of tuned gain control in mind, we are able to fit the model directly to our data, which opens a direct link between the inferred tuning and strength of gain control and a rich set of measured responses.

Each chapter highlights a partially overlapping subset of the broader aim, which is to identify and model the influence of gain control on spatial frequency tuning in the early visual system. All experiments involved recording single LGN and V1 neurons in anesthetized macaque monkeys. In chapter 2, we characterize responses to single sinusoidal gratings which vary in spatial frequency and contrast using a pair of related, mechanistic models to elaborate on the relationship between LGN and V1 spatial frequency tuning and its contrast-dependence. Next, chapter 3 adds to our picture of spatial frequency selectivity by profiling summation and suppression using grating pairs and other superimposed grating mixtures. Again, the use of a common stimulus set permits a direct comparison between effect sizes in the two

areas of study. The final chapter considers a computational model of cortical gain control fit to single neuron responses. This model seeks to unify the results of the earlier chapters and previous studies by capturing the computations which underlie spatial frequency selectivity and gain control in the early visual system. Earlier versions of the work here were presented at Vision Sciences Society meetings in 2019 [79] and 2020 [80].

1.2 General methods

1.2.1 Electrophysiological recordings

Anesthetized preparation

We performed acute, anesthetized experiments in 13 (12 male, 1 female) macaque monkeys *(Macaca nemestrina, M. fascicularis, & M. mulatta)* according to the protocols previously outlined in greater detail [27, 12]. We administered intravenous sufentanil citrate (6 µg/kg/hr–30 µg/kg/hr) and vecuronium bromide (Norcuron, 0.1 mg/kg/h) to maintain anesthesia and paralysis, respectively. To ensure physiological health and a sufficient depth of anesthesia, we closely monitored and maintained a range of physiological signs, including heart rate and waveform, body temperature, lung pressure, urine output and specific gravity, and electroencephalogram (EEG) waveform. To keep the eyes protected and in-focus, we ensured the pupils were dilated by application of topical atropine, placed gas-permeable contact lenses on the eyes, and finally refracted with supplementary lenses chosen by direct ophthalmoscopy. Experiments typically lasted 4-7 days, after which the monkey was killed with an overdose of sodium pentobarbital. All animal care and experimental procedures were performed in accordance with protocols approved by the New York University Animal Welfare Committee and conformed to the NIH Guide for the Care and Use of Laboratory Animals.

Visual stimulus presentation

We presented stimuli on a gamma-corrected cathode ray tube (CRT) monitor (Eizo T966) placed 114 cm from the subject, with a mean luminance of 35 cd/m^2, a spatial resolution of 1280×960 pixels, and a temporal resolution of 120 Hz. Stimuli were generated with Expo (https://sites.google.com/a/nyu.edu/expo/home) running on an Apple Mac Pro. We qualitatively determined the ocular dominance for each neuron and occluded the non-dominant eye to present all stimuli monocularly.

Unit recording

We recorded neurons with one of two approaches - single channel microelectrodes or multi-channel electrode arrays. We performed a craniotomy and minimal durotomy placed to target the area of interest before lowering the electrode into the brain for recording. Over the course of each experiment, we made multiple electrode penetrations passing through the area of interest. In all cases, we targeted our recordings for one neuron at a time, starting by hand-mapping the neuron, and then running a suite of core programs to measure tuning for orientation, spatial and temporal frequency, size, and contrast (see *Basic tuning properties* section below, §1.2.2). The tuning preferences determined from these measurements were used to define the primary stimulus programs.

Single channel recordings were made either with platinum/tungsten (Thomas Recording; Giessen, Germany) (V1) or tungsten-in-glass [92] microelectrodes (LGN, V1). We tangentially advanced the electrode through the brain, stopping to isolate and record individual cells. The signals were amplified and bandpass-filtered between 300 Hz - 10 kHz before on-line spike detection, which was done via a dual-window time-amplitude discriminator. We saved the full spectrum waveform with a 0.1 mS temporal resolution for offline analysis.

Our multi-channel electrode recordings were made with Neuropixel linear arrays, where we recorded up to 384 channels simultaneously in V1. We typically placed the arrays perpendicular to the cortical lamination, settling at one depth for each penetration that spanned all of V1.

Then, we found individual channels with distinct, single neuron activity to which we tailored our primary stimulus program. We also saved the signals from all channels, and were thus able to also recover non-targeted, isolated neurons, either from the same channel or nearby ones (see *Spike sorting*, §1.2.1). We were able to also analyze the responses from non-targeted neurons when its tuning properties were similar to those of the targeted neuron.

Spike sorting

On-line spike discrimination was done via time-amplitude template matching in Expo. Offline sorting for V1 neurons, whether for recordings with poor isolation or multiple units, was completed with Mountainsort [31]. We started our offline sorting with the clusters identified by Mountainsort's algorithm. The default clusters are generally conservative, such that some cells are split across multiple of the initial clusters. We only merged clusters after ensuring the spike cross-correlations, waveform, and spike trains were consistent with only one underlying neuron. We focused our efforts on channels nearby the initially recorded neuron and, in particular, on channels we had noted as highly active during the initial recordings.

Eccentricity

The eccentricity of our LGN population spanned from 0.5 to 35°. Despite the wide range, the overwhelming majority of cells were within the central 8 degrees of the fovea. Our V1 neurons were largely parafoveal (within 2-8° of the fovea) though some more eccentric receptive fields were up to 20° from the fovea.

1.2.2 Stimulus sets

Basic tuning properties

Before conducting the core experiment, we performed a suite of basic tuning experiments for each cell, updating the optimal stimulus parameters after each. We first initialized the

Basic tuning experiments			
Stimulus Domain	Range & units	# samples	Additional comment
Orientation	$(0, 360)°$	16	Only for V1
Spatial frequency	$(0.3, 10)$ c/deg	11	
Temporal frequency	$(0.5, 30)$ Hz	11	
Size	$(0.1, 10)$ dva	10	Disk & annulus
Contrast	$(0.01, 1)$	11	

Table 1.1: Summary of basic tuning experiments in the typical order of presentation. With the exception of orientation, all stimulus domains were sampled logarithmically.

stimulus properties to match those we mapped by hand when first investigating a cell, and restarted the suite of basic tuning experiments if any values differed substantially from initial estimates. The stimulus sampling was logarithmic for all experiments except the orientation tuning, which was linearly sampled. The order and content of these experiments are listed in *Table 1.1*.

We note that for size tuning, annuli had a constant outer diameter of 10° and the sampling determined the variable-size inner diameter, while 'disk' simply refers to a standard, continuous sinusoidal grating.

For orientation, spatial, and temporal frequencies, the tuning was summarized by the cell's preference and tuning bandwidth as inferred from standard descriptive tuning functions fit to the responses. For contrast, we used the Naka-Rushton equation described below (1.2.3). For size tuning, we fit the ratio of Gaussians model described in [27] from which we could infer the grating summation field (stimulus extent which drives the maximum response), surround diameter (spatial extent of the surround), and size suppression index. The size suppression index is bounded from $(0, 1)$ where the extrema indicate, respectively, no surround suppression and the complete abolition of response with stimuli larger than the grating summation field.

Core stimulus sets

We used the properties measured with the basic tuning properties to determine the stimuli for the core experiment. All stimuli were achromatic and presented at the preferred orientation (if tuned) and location. With the exception of a portion of LGN recordings collected with a fixed 8° aperture, we chose the size to match the cell's preferred size (i.e., the grating summation field, beyond which increasing the stimulus size either suppresses or no longer increases the response) whenever possible, though we erred on the side of larger stimuli to ensure resolution of the lower spatial frequencies presented. However, for cells with smaller receptive fields and strong surround suppression, we had to strike a balance between low spatial frequency resolution and reliable firing activity. For our LGN dataset, all but 5 cells had stimulus sizes between 4°-8°. Across our V1 population, the mean stimulus size was 1.9°, and was on average 2 times larger than the receptive field. The implications of stimulus size are discussed in the chapters, as relevant. For each experiment, the order of conditions was pseudorandomized and blank trials (i.e., full field of mean luminance) comprised around 10% of conditions.

1.2.3 Data analysis and modeling

All analysis and models were written in Python 3.6.12 using a range of open-source packages including SciPy, NumPy, PyTorch, scikit-learn, Matplotlib, and Seaborn. Beyond what appears below, the remaining tuning functions and models are described in the relevant sections of the thesis.

Response-versus-contrast

To model the response versus contrast, we fit a Naka-Rushton curve [96], described as

$$R(c) = R_m \cdot (\frac{c^n}{c^n + c_{50}^n}) + b$$

For LGN neurons only, we also consider a function of the form

$$R(c) = R \cdot log(1 + \frac{c}{c_0}) + b$$

as described in Movshon et al. 2005 [95], since it permits a simple calculation of the contrast gain, or responsivity, by taking the slope of this function at 0 contrast.

Model fitting

Throughout this work, we fit a range of models to our data. The specific model formulations and the loss function used for each model are discussed in the relevant sections and chapters.

As an independent measure of model fit quality separate from the loss metric, we computed the explained variance between a given model fit and set of data. This metric is equivalent to R^2, and is described in [25], though we summarize it here:

$$v = 100 * (1 - \frac{d(m, r)}{d(r, \bar{r})}) \tag{1.1}$$

where

$$d(x, y) = \frac{1}{N} \sum_i (x_i - y_i)^2 \tag{1.2}$$

and m, r represent the mean model and data response for a given condition and $\bar{r}$ is the mean data response across all conditions.

Chapter 2

Contrast-dependent spatial frequency selectivity in macaque LGN and V1 neurons

2.1 Introduction

Unlike orientation tuning, which emerges anew in primary visual cortex [64], spatial selectivity first arises in the retina, is modified by the lateral geniculate nucleus (LGN), and is further sculpted by cortical mechanisms. In the LGN, both magno- and parvocellular (M&P) neurons are broadly tuned for spatial frequency, with largely low-pass tuning curves [40]. A difference-of-Gaussians model that consists of opponent linear center and surround mechanisms can describe this tuning [43]. On the other hand, the distribution of spatial frequency bandwidth and preference in V1 neurons is more varied, and tuning is generally much sharper than in the LGN [36]. The exact influence of thalamic spatial frequency selectivity - as opposed to intracortical processing – on the properties of V1 neurons is not fully understood. The goal of this work is to carefully document spatial frequency tuning under changes in contrast for

both LGN and V1 neurons. Furthermore, we relate these tuning properties to mechanisms proposed in commonly used receptive field models.

In the LGN, magno- and parvocellular neurons are selective for image contrast, color, and spatial and temporal frequencies [40, 39]. These responses are shaped by an excitatory center mechanism and two distinct suppressive mechanisms – one that is linear and part of the classical receptive field, and another that imbues thalamic responses with non-linear properties like contrast saturation [40] and size tuning [148]. These mechanisms have a similar spatial extent and may share feedforward retinal origin, but the extra-classical suppressive field is more broadly tuned for spatial and temporal frequency [147], and, unlike the linear surround, is sensitive to stimulus energy rather than intensity [19]. Furthermore, extra-classical suppression has been shown to give rise to contrast dependent tuning – for size – in the thalamus [148]. Previous studies in cat [102], macaque [146, 147], and marmoset [74] all document a contrast dependence in which receptive field size grows at lower contrast.

In V1, spatial tuning is also contrast dependent – reducing contrast increases the optimal stimulus size [127, 68] and decreases the preferred spatial frequency [126]. Mechanistically, one hypothesis is that the sizes of underlying excitatory center and suppressive surround mechanisms change with contrast [126]. An alternative explanation is that the center and surround are fixed in size, but their gains vary independently with contrast [27]. Our work will instead focus on the functional role of suppression by carefully measuring spatial frequency selectivity.

While studies of contrast's influence on spatial frequency have been conducted in primate LGN [146] and V1 [126], no studies have reported results from both areas. Here, we richly characterize the contrast-dependence of tuning in single neurons of anesthetized macaque monkeys using a common battery of sinusoidal grating stimuli which vary in spatial frequency and contrast. By fitting simple models with similar underlying mechanisms to data in both LGN and V1, we are able to relate the cortical results to the properties of their thalamic inputs. We replicate previous findings in LGN and V1 that highlight the reduction in bandwidth

and loss of high-frequency selectivity at lower contrasts. Our findings indicate that the contrast-dependency of LGN spatial frequency tuning is inherited in V1. The relatively larger effects in V1, for complex cells in particular, suggest that some additional cortical processing may further contribute to these contrast-dependent tuning properties.

2.2 Methods

2.2.1 Electrophysiological recordings

Anesthetized preparation

We performed acute, anesthetized experiments in 13 (12 male, 1 female) macaque monkeys *(Macaca nemestrina, M. fascicularis, & M. mulatta)* according to the protocols described in greater detail earlier in the thesis (§1.2.1) and as approved by the New York University Animal Welfare Committee.

Unit recording, spike sorting

The details regarding how we collected and spike sorted our extracellular recordings are given in greater detail earlier in the thesis (§1.2.1). Briefly, the LGN data presented here were collected using single-unit microelectrodes. The V1 data comes from both single-unit electrode and multi-unit array recordings. Data were initially spike-sorted online using time-amplitude matching templates within our joint stimulus presentation/recording software. For data with poor isolation or to maximize the yield from multi-unit electrode arrays, we also performed offline spike sorting in Mountainsort [31].

2.2.2 Stimulus set

Our core stimulus sets were comprised of single sinusoidal gratings which primarily varied in contrast and spatial frequency. All gratings, unless otherwise noted, were presented at the

preferred location, size, and, if applicable, orientation. Stimuli were shown with a randomized phase for each presentation, and the order of stimuli was pseudorandomized.

LGN

The majority of LGN neurons (40 parvo, 25 magno, and 9 unclassified, giving 74 in total) were collected using a stimulus set of 9 log-spaced contrasts, ranging from 0.05 to 1.0. For the great majority of these neurons, the 11 presented spatial frequencies spanned 4 octaves, centered either at 1, 1.5, or 2 c/deg, depending on the preliminary spatial frequency tuning curve. For neurons with very broad tuning and/or pass-bands shifted outside the range covered by our standard set, we shifted the center frequency accordingly and/or increased the span up to 5 octaves. These single grating stimuli were presented as part of a larger stimulus set, which required the temporal frequency of each spatial frequency to differ. However, the temporal frequency associated with each spatial frequency was constant across all repeats, and the values were chosen to be close to the peak of temporal frequency tuning to minimize the influence of this irrelevant stimulus dimension on the responses. Stimulus size was chosen to be at least as large as the receptive field, though we used larger sizes as needed to resolve lower spatial frequencies. Most stimuli were either 4° or 8° in size, with only 5 cells recorded using smaller stimuli.

A smaller fraction (14 parvo, 15 magno, and 5 unclassified) were collected by a previous PhD student in the Movshon lab, Sach Sokol. Due to the similarities between our experiments, we include his data here. This stimulus set sampled from a wider set of contrasts (11 log-spaced values from 0.01 to 1.0) and spatial frequencies (13 log-spaced frequencies centered according to each cell's tuning and spanning 6 octaves; and zero frequency). These stimuli were all presented at a size of 8° and with a fixed, integer temporal frequency closest to the cell's preference. Each stimulus was presented for 0.5 seconds at a time, with up to 20 repeats yielding 10s of total stimulus presentation per condition.

V1

Just over 50% of the data (47 out of 89 V1 neurons included) were collected using a similar 9 contrast by 11 spatial frequency matrix as in the LGN recordings. The contrasts again spanned 0.05 to 1.0, but we increased the default span of spatial frequencies to 5 octaves to account for the higher prevalence of responsivity at high spatial frequencies. Most neurons were sampled from 0.3 to 10 c/deg, but we shifted the 5 octave sample for cells with tuning distinctly outside of this range. Temporal frequency was chosen as described for the primary LGN stimulus set.

The remainder of the V1 data (42 neurons) were collected as part of a different stimulus set. The relevant stimulus subset for this analysis is a 7x7 contrast and spatial frequency grid. The contrasts ranged from 0.01 to 0.64, while the frequencies spanned 4.5 octaves, centered on the cell's preference. All gratings were presented at the integer temporal frequency closest to the cell's preference.

We chose the stimulus size to closely match the receptive field size while also seeking to minimize surround suppression when present. To ensure that we fully measured the tuning, we erred on the side of larger stimulus apertures which would allow for better resolution of low spatial frequencies. After balancing these factors, the average ratio between stimulus size and the receptive field size was 2.1 across our V1 population.

2.2.3 Spatial frequency tuning functions

Difference-of-Gaussians (DoG)

The most commonly descriptive tuning function used to fit spatial frequency tuning curves in pre-cortical areas is the difference-of-Gaussians model [43], parameterized in the frequency domain as

$$R(f) = K_c \cdot (e^{-(f\pi r_c)^2} - K_s e^{-(f\pi r_s)^2}) \tag{2.1}$$

where f is the spatial frequency, in c/deg. To simplify the optimization, we parameterized r_s as a scalar multiple of r_c, i.e., $r_s = r'_s * r_c$.

Difference of difference-of-Gaussians (d-DoG-S)

For cortical responses, we considered the DoG model and also a difference of difference-of-Gaussians (d-DoG-S) model [53], which can account for the greater variety of tuning properties observed in cortex.

In the spatial domain, the model is described as

$$R(x; g, S, \vec{\theta_a}, \vec{\theta_b}) = f(x; 0, \vec{\theta_a}) - g \cdot f(x; S, \vec{\theta_b}) - (1 - g) \cdot f(x; -S, \vec{\theta_b}) \tag{2.2}$$

where

$$f(x; x', \vec{\theta} = K_c, K_s, x_c, x_s) = K_c e^{-((x-x')/x_c)^2} - K_s e^{-((x-x')/x_s)^2} \tag{2.3}$$

Thus, there are two distinctly parameterized DoGs (i.e., $\vec{\theta_a}$, $\vec{\theta_b}$), as well as spatial offset (S) and symmetry (g) terms that determine the relative location of the flanking DoG and the receptive-field symmetry, resulting in 10 parameters. In the frequency domain, each DoG subunit in *eq.* 2.3 is instead expressed as in *eq.* 2.1.

2.2.4 Model optimization

Loss function

We applied the square-root transform to firing rates, as described in [111], and thus fit our model parameters $\vec{\theta}$ by optimizing:

$$\min_{\vec{\theta}} \sum_{\omega_i, c_j} \left(\sqrt{r_{\omega_i, c_j}} - \sqrt{R(\vec{\theta}; \omega_i)} \right)^2 \tag{2.4}$$

where w, c represent the frequency and contrast of the stimuli; r the measured response; and R the model response.

The square-root transformation works best for data, like neural firing rates, which are roughly Poisson in their noise. By taking the square-root, the transformation de-emphasizes high firing rates that are likely to be more variable. We fit our models to the mean firing rates per condition.

Model constraints

We chose our constraints to both reflect knowledge about receptive field structure from existing studies and to avoid extending model flexibility beyond what our stimuli constrained. For the DoG model, the surround radius and gain were by definition larger and weaker, respectively, than their corresponding center mechanisms. The surround radius was constrained to be no larger than 10 times the center; this encompasses the range of observed radius ratios [65, 32]. We did not target the phase-specific structure of the receptive field (RF) nor did we analyze the phase-specific response of the F1 response. Thus, in the d-DoG-S model, the symmetry of the RF (parameter g) in the model, was unconstrained. For simplicity, we kept $g = 0$, i.e., all model RFs were odd-symmetric. The center mechanism of the central DoG alone was constrained to give a response no greater than 1.5 times the maximum response [53]. We constrained the flanking DoG to have a center radius between 1-4 times that of the central DoG. This is in correspondence with the range of center sizes at a given eccentricity, accounting for the possibility that a given V1 neuron would receive input from both magno and parvo, which differ by roughly three-fold [32]. The spacing between the subunits was constrained to be no smaller than the smallest center radius and no greater than the sum of the two center radii.

Joint optimization, inclusion criterion

Neither the DoG nor d-DoG-S models explicitly consider contrast. Thus, our parameters $\vec{\theta}$ fit to different contrasts implicitly model the influence of contrast on the underlying mechanisms. The goal of the model comparison analysis described here is to understand how receptive

field mechanisms might plausibly change with contrast. To that end, we chose to jointly fit the responses across contrast while asserting certain fixed or smooth relationships between receptive field properties across contrast. These constraints were necessary due to the extent to which changes in certain DoG parameters can compensate for one another with minimal change in the overall loss function [83]; this is particularly the case for the surround radius and gain when fit without very low or zero spatial frequency stimuli. The models which we compared all serve to test which receptive field properties - e.g. surround gain or center radius - are most influential in capturing the changes in response properties that we observed with contrast.

As for the joint fitting procedure, we first fit one set of $\vec{\theta}$ for each contrast separately. Then, we also fit joint models in which we optimized $\vec{\theta}$ across all contrasts using the same loss function, *eq.* 2.4. For joint optimization, some of the parameters $\vec{\theta}$ which were previously fit separately for each contrast were instead fixed across all contrasts; the remaining parameters were allowed to vary with contrast, but the optimization was nonetheless run jointly across all contrasts. Our stimulus set consists of spatial frequency tuning curves measured at a range of contrasts; to avoid biasing the model fits, we excluded noisy or sub-threshold responses. These responses bias the model towards small center radii: the DoG model fits a very narrow receptive field to capture what is effectively a flat tuning curve. We thus restricted our fits to those conditions which were well accounted for by the DoG model when fit to those responses alone. Thus, we excluded all data below the lowest contrast for which the DoG model, fit only to the tuning curve at that contrast, accounted for at least 60% of the variance. This, and the requirement that at least 3 contrasts were included in the joint fits, comprised the inclusion criteria that we applied for all considered models and parameterizations.

For the DoG model, the separate fits ('*free*' model) arise from $4 \cdot n_{con}$ total parameters. We also considered joint models, each of which has the surround parameterized in terms of the center. That is, when we write that the surround gain or radius is fixed across contrast, that means fixed in constant proportion to its paired center mechanism. We considered the

following three joint parameterizations: surround radius fixed across all contrasts ('$\Delta r_c, vol$'; $3 \cdot n_{con} + 1$ total parameters); surround gain and surround radius fixed across all contrasts ('Δr_c'; $2 \cdot n_{con} + 2$ total parameters); and surround radius fixed across all contrasts with the center radius determined by a slope model ('Δr_c*(slope)*'; $2 \cdot n_{con} + 3$ total parameters).

The slope model assumes a power law relationship between contrast and center radius, which we can describe as a linear relationship in log-log space. That is, the center radius x at contrast c is given as

$$log_{10}(x(c)) = x_0 + \alpha \cdot log_{10}(c)$$

where the intercept (x_0) and slope (α) are jointly optimized across all contrasts for each cell. With this model, then, the center radii are determined from 2 parameters rather than n_{con} parameters. We verified the validity of this relationship by fitting a slope (via regression) to the change in center radius across contrast as evaluated for the '*free*' model. We found that the errors between the slope and measured center radii were symmetrically distributed (Wilcoxon signed-rank test, $p > 0.05$) for all tested contrasts. Further, the inferred slopes from this regression and from the slope model directly were in strong agreement.

For the d-DoG-S model, we kept the surround radii in constant proportion to the center radii. We fixed the spacing between subunits to be constant across all contrast, in keeping with the biophysical interpretation of each subunit as arising from some combination of LGN inputs whose receptive field centers do not move with contrast. Further, we kept the flanking DoG relative surround gain and radius equal to that of the central DoG. The flank center gain and radius were fixed relative to their central counterparts across all contrasts.

Model initialization

We ran multiple optimizations (3 for joint models, 10 for per-contrast fits) for each model, with randomly jittered starting parameters for each. For parameters which readily correspond to the data (e.g., response amplitude), we initialized based on the measured responses with

some jitter; for other parameters (e.g., surround radius), we chose randomly within ranges suggested by previous work [70, 4, 78, 149, 14, 95].

Besides a jittered start based on measured responses, we also initialized at least one optimization of each model on the basis of a successful optimization from a simpler model. For example, the central DoG of the d-DoG-S model was initialized (again with some jitter) on the basis of the DoG model fit. Likewise, we initialized jointly optimized fits on the basis of the individual fits and the simpler model fit. More specifically, joint parameters for the d-DoG-S model (e.g., characteristic frequency slope) would be initialized on the basis of the equivalent, optimized parameter for the DoG model; then, the per-contrast parameters (e.g., surround amplitude) would be initialized based on the separate, per-contrast fits.

2.2.5 Model comparison

To compare model performance between models with different numbers of total parameters, we performed leave one out cross-validation. Thus, we iteratively witheld a specific spatial frequency and contrast combination $\{f_i, c_j\}$, fitting the model to all other conditions and evaluating the loss, (*eq.* 2.4), on the held out condition. Finally, the total cross-validated loss is given as the sum of the evaluated test losses for all conditions.

2.2.6 Bootstrap resampling

We also performed bootstrap resampling of our data to better understand the variability in the underlying model parameters for each cell and fit. For this, we resampled (with replacement) from the existing trials, re-computed the condition averages, and then fit the model to these new mean responses.

2.3 Results

We recorded the spiking activity of 115 LGN and 115 V1 neurons in anesthetized macaque primates to sinusoidal grating stimuli which varied in spatial frequency (f) and contrast (c). All other stimulus dimensions (e.g., size, orientation) were fixed based on the neuron's preference. All LGN data were collected using single-channel electrodes, and we identified units as either magnocellular or parvocellular based on ocularity, post-experiment histological reconstruction, and a continuous tracking of layer and depth during the experiment. Neurons encountered in between magno and parvo layers (in the S layers) or that were otherwise difficult to classify were categorized separately. V1 data was primarily collected with multi-unit electrode arrays (the remainder with single-channel electrodes), but we identified individual neurons during each recording session and tailored our stimulus accordingly. We excluded 7 LGN neurons (3 parvo, 2 magno, and 2 unclassified) and 26 V1 neurons (12 simple, 14 complex) based on inclusion criteria (see §2.2.4), leaving 108 and 89 neurons, respectively. Most data were collected using a pre-defined set of logarithmically-spaced contrasts $\{c_i\}$ and spatial frequencies $\{f_j\}$ chosen to best sample the range of expected population responses. For a subset of V1 neurons (42 of those included), we chose a set of frequencies $\{f_j\}$ to best sample that neuron's passband. Each frequency and contrast combination was presented for a total of 10 seconds, acquired through randomized repetition of shorter, fixed-interval trials (1s or 2s). Throughout, we report the mean response averaged across stimulus repeats, using the F1 amplitude for LGN neurons and V1 simple neurons and the F0 (i.e., DC) for V1 complex cells.

We fit receptive field models for spatial frequency tuning to the each neuron's full set of responses. These mechanistically-driven models allow us to relate the observed responses to the underlying receptive field structure, and to compare between LGN and V1 neurons.

2.3.1 Quantifying spatial frequency tuning and its contrast-dependence in the LGN

We modeled each LGN neuron's spatial frequency tuning using a difference-of-Gaussians (DoG) model. Four example neurons with fitted DoG tuning curves are shown in *fig.* 2.1. This model, as has been demonstrated previously [43, 40], captures the measured spatial frequency tuning well. In conjunction with a model of the contrast response function, we replicated some of the hallmarks of LGN subtypes. In particular, we find that parvocellular neurons are more slightly more narrowly tuned, on average, and tuned to higher spatial frequencies (not significant; t-tests on distributions); furthermore, magnos have a much greater contrast sensitivity than parvos (population median contrast gain of 227 vs. 79 spikes/s/contrast; $p <$ 1e-4, t-test on distributions). Here, we report contrast sensitivity as the slope of the response-versus-contrast fit (responsivity) measured at threshold for the optimal spatial frequency, and our values are in agreement with prior studies (c.f. [78]). Beyond capturing basic tuning properties as measured at high contrast, the DoG model fit to the full stimulus set also permits a study of the changes in underlying center and surround mechanisms with contrast.

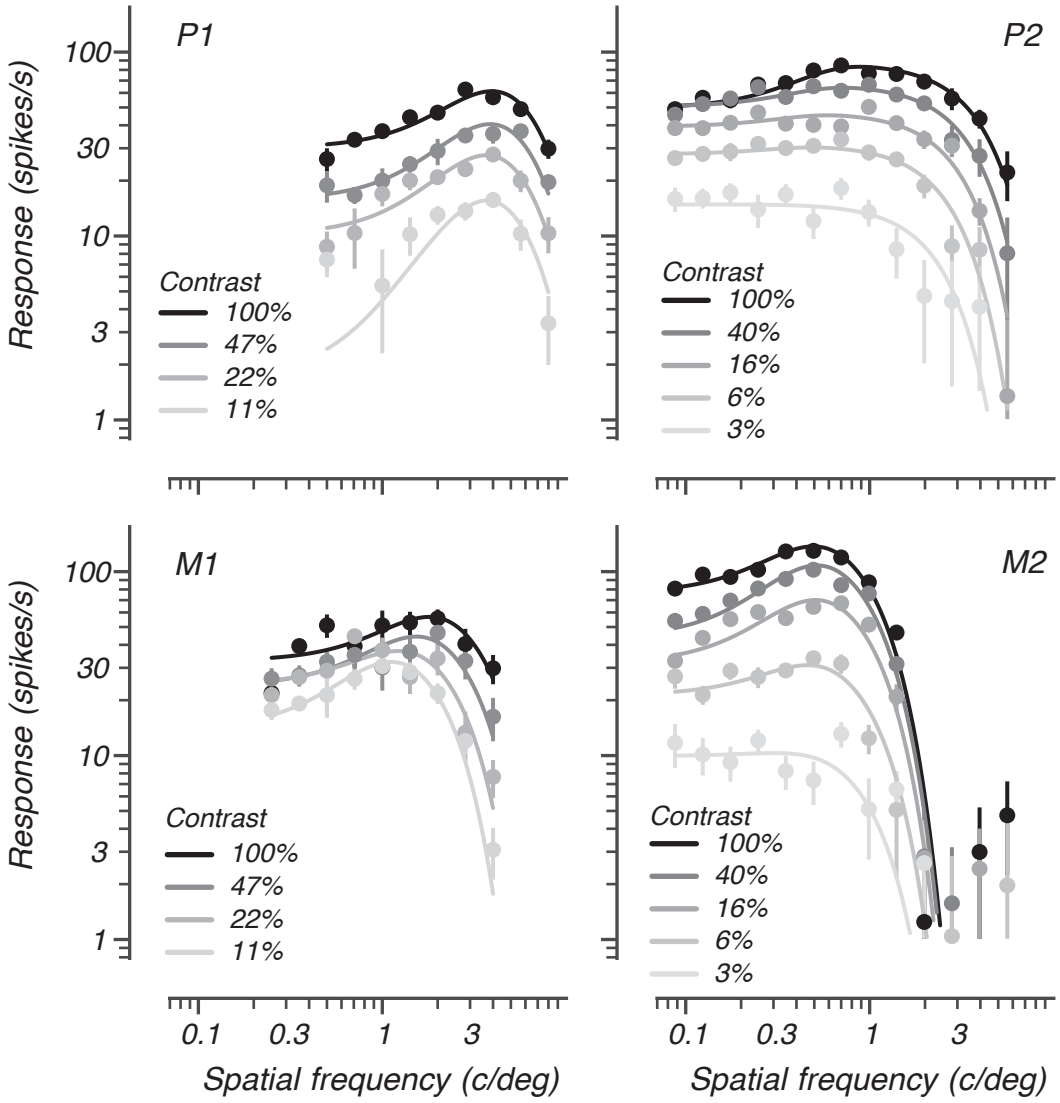

Figure 2.1: Four example LGN neurons with mean response and standard error of the mean (points and error bars, respectively) plotted on equal scale, double log axes. The tuning curves shown (lines) are DoG models fit with the chosen model (see model comparison analysis). The two neurons on the top are parvocellular, while the two on the bottom are magnocellular. The left two neurons are shown, shown from 0.11 to 1.0 contrast; the right, from 0.03 to 1.0. In each case, we depict only a subset of tested contrasts (skipping every other) for better legibility.

We next sought to better understand the influence that each DoG parameter has on the overall tuning, and, in particular, its contrast-dependence. As seen in *fig.* 2.2, each parameter shapes the tuning in different ways. Of great relevance here, only the center radius determines the extent to which the neuron responds at high spatial frequencies. On the contrary, the relative gain and radius of the surround both influence low spatial frequency responses, i.e.,

responses below the broad peak of selectivity. Though the surround gain and radius act on the tuning curve in somewhat orthogonal manners, they can nonetheless trade-off with one another given the spatial frequency sampling typically employed in these experiments - that is, adjustments in the surround gain can often be compensated for by surround radius [83]. With that non-orthogonality in mind, we employed a model comparison analysis to ensure that we chose a formulation of the DoG model which would yield explanatory power without overfitting. Next, we wanted to determine which succinct parameterization of the DoG model, fit jointly across contrast, would best capture our data.

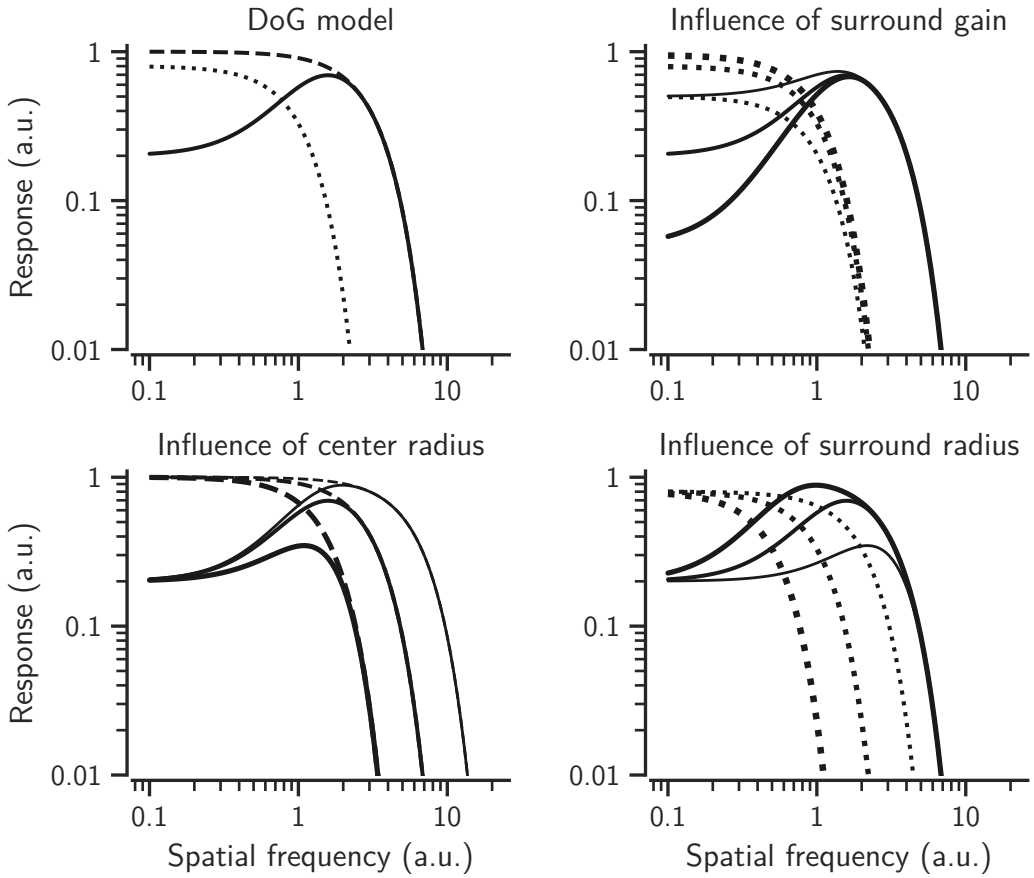

Figure 2.2: (*Top left*) An example DoG model tuning curve *(solid line)* with the underlying center (*dashed*) and surround Gaussian mechanisms (*dotted*) shown separately. Then, clockwise from the *top right*, we depict the resulting tuning curve and underlying mechanism when changing the relative surround gain, surround radius, and center radius. In all cases, lines of increased width correspond to an increase in the value of the particular mechanism.

Model comparison analysis

In this analysis, we simultaneously fit a series of linked DoG models to the spatial frequency tuning curves across contrast. We explored several distinct parameterizations to understand the influence of different DoG components on spatial frequency tuning and its change with contrast. The broader importance of this analysis, then, is to ensure that the chosen model is simple yet effective enough to allow a direct mapping between the measured tuning and underlying receptive field structure. In finding the most effective parameterization, we identify the aspects of tuning which change most dramatically with contrast and generate hypotheses for how changes in the receptive field give rise to that tuning. Given our focus on capturing changes in high-frequency selectivity with contrast, we only considered models which allowed the center radius of the DoG to vary with contrast. We then considered the following four model configurations - all 4 parameters free to vary with contrast ('*free*'); fixed surround radius, kept in constant proportion to the center radius across contrast ('$\Delta r_c, vol$'); surround gain and radius fixed in constant proportion to the center across contrast ('Δr_c'); and center radius inferred from a slope model, surround radius fixed ('Δr_c*(slope)*'). For further details on these models, see §2.2.4, *Joint optimization*.

We performed leave-one-out cross-validation to compare these models, which differ in their number of parameters. The center radius slope model, 'Δr_c*(slope)*', gave the lowest cross-validated test loss for more than 60% of the LGN neurons in our population (*fig.* 2.3), a percentage much greater than expected if all models were equally suited to the data. By contrast, fewer than 10% of cells each showed the lowest test loss with the '*free*' and '$\Delta r_c, vol$' models. Thus, within the context of our stimulus set, the additional model power granted by an unconstrained center radius, surround gain and/or surround radius are not necessary to account for the vast majority of neurons. The remainder, 20% of neurons, were better accounted for by an unconstrained center radius but fixed surround gain ('Δr_c') rather than a slope-constrained center radius and free surround gain ('Δr_c*(slope)*'). For the remainder of

the analyses, we consider the slope model, since it outperformed the other three models for the majority of neurons.

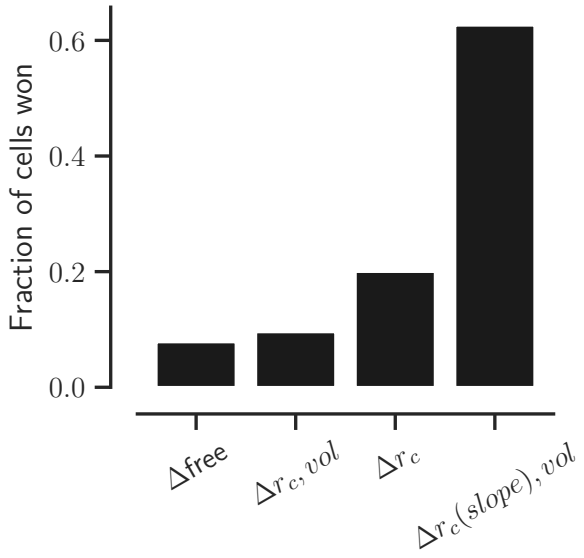

Figure 2.3: More than 60% of neurons had the lowest cross-validation test error with the slope model ('Δr_c*(slope)*'). Models with fewer free parameters outperformed those with more flexibility. For example, the fully unconstrained model (Δfree) was overparameterized for this stimulus set, as evidenced by the small fraction of cells it best accounted for.

LGN neurons show contrast-dependent spatial frequency tuning

As a metric for high-spatial frequency cut-off [65], we measured the inverse of the DoG center radius (*fig.* 2.4), which is measured in cycles per degree of visual angle (c/deg) and termed the characteristic frequency. The slope model provides a direct measure of how much the center radius changes with contrast; then, given the inverse relationship between center radius (x_c) and characteristic frequency (i.e., $f_c \propto \frac{1}{x_c}$), the change in characteristic frequency is simply the negation of the fitted center radius slope. For both magno and parvo neurons, the population distribution of these slopes were shifted away from zero (*fig.* 2.5). The population medians for magnos (0.12) and parvos (0.10) were comparable, and the distributions were not significantly different (t-test, $p = 0.76$). However, because magnocellular neurons generally have a higher signal-to-noise ratio than parvos, a greater fraction of individual magno neurons had significant shifts in characteristic frequency (65% vs. 15%). We measured significance in

individual neurons by computing a distribution of 100 bootstrap slopes and checking that no more than than 5% of slopes had negative values.

The reported slopes were fit in log-log space. To better understand these values, consider the following: a slope of 0.23 (example *M1* in *fig.* 2.1 and *fig.* 2.4) will yield a 70% increase in characteristic frequency from 0.1 to 1.0 contrast. The parvo population median (0.10) gives a 15% change from 0.25 to 1.0 contrast and 26% from 0.1 to 1.0 contrast.

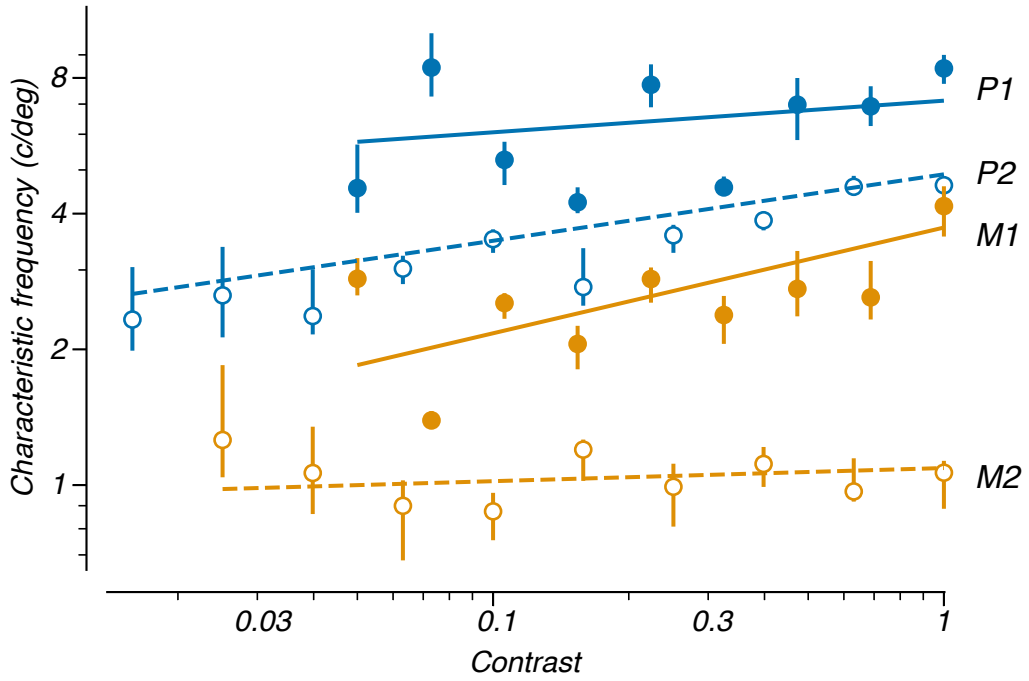

Figure 2.4: The characteristic frequency across contrast for the same 4 example neurons in *fig.* 2.1, notated correspondingly. The lines represent the slope as fit with the slope DoG model ('Δr_c*(slope)*'), while the points represent the median and interquartile range (error bars) across 100 bootstrap fits with the free model, in which the center radius was fit separately for each contrast. For legibility, we depict alternate cells with hollow/dashed or filled/solid points/lines. From top to bottom, the slopes are 0.07, 0.14, 0.23, 0.02. All slopes except for that of the bottom neuron were significantly greater than zero by bootstrap analysis.

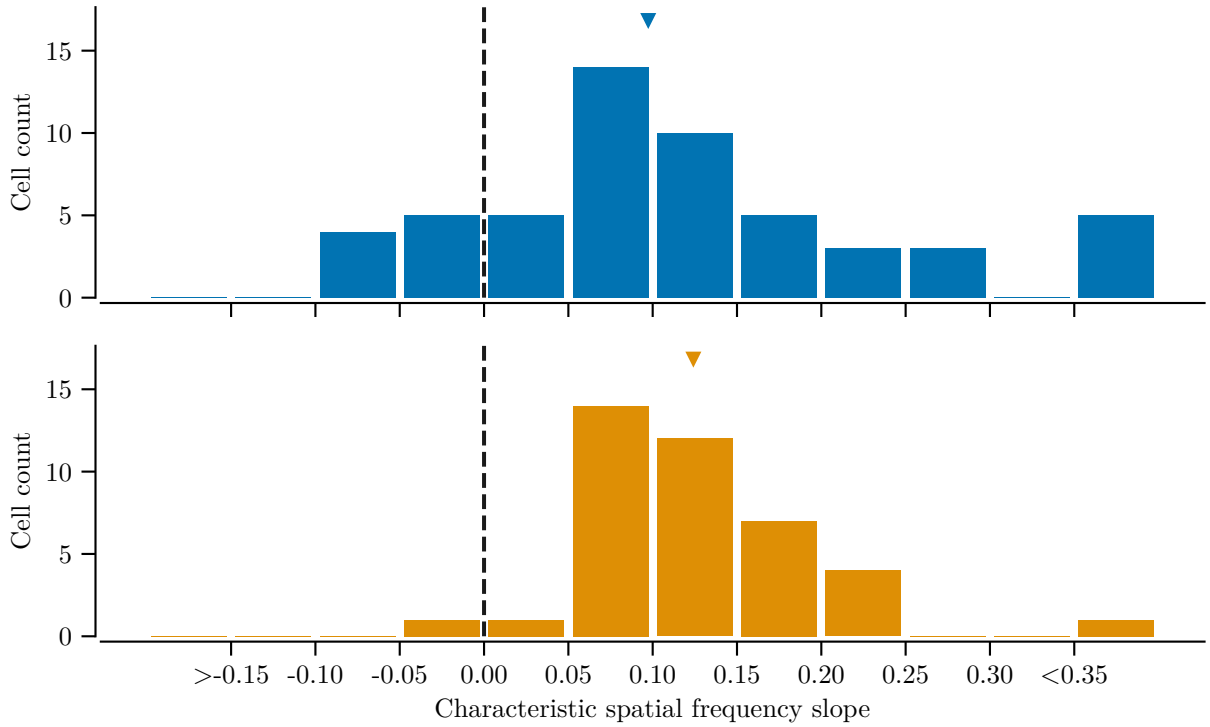

Figure 2.5: Distribution of the power law exponent determining the slope of characteristic frequency across contrast, split by cell class. A slope greater than 0 indicates an increasing characteristic frequency with contrast. Both parvo (*top*, n=54, median=0.10) and magno (*bottom*, n=40, median=0.12) distributions had means significantly greater than 0 (t-test; $p <$ 1e-7). The distributions were not significantly different from one another (t-test, $p = 0.76$). The median slope of 14 B-Y and unclassified LGN neurons (not shown) was 0.12.

In addition to a changing characteristic frequency, we also found tuning bandwidth to be contrast-dependent. Though often undefined for LGN neurons due to their broad, low-pass tuning, we measured the three-fourth bandwidth across contrast, and fit a power-law relationship between log-contrast and log-octave bandwidth for defined values. We found that parvocellular neurons had a median shift of 10% ($n = 51$) with a one-quarter reduction (e.g., 1.0 to 0.25 contrast) in contrast; the reduction was 3% for magnocellular neurons ($n = 34$). The effect size between these two populations was significant (t-test, p=.001) at three-fourth bandwidth, but not at half-height bandwidth (p=0.25), where the measured changes were more comparable. Finally, as with the cut-off frequency and center radius, we can relate the change in bandwidth to the DoG mechanisms. The intuition gained from the initial simulations of the DoG model are borne out: there was a strong correlation between measured

change in bandwidth and the relative strength of the surround. That is, the steeper the change in bandwidth, the more the relative strength of the surround increased at low contrast ($r = -0.59, p <$ 1e-10).

2.3.2 Quantifying spatial frequency tuning and its contrast-dependence in V1

Modeling V1 receptive fields

We performed an equivalent analysis of spatial frequency tuning across contrast for individual V1 neurons. We modeled V1 receptive fields using a model comprised of two LGN-like, DoG subunits [53] with a spatial offset. This difference of DoGs model, termed d-DoG-S, is grounded in both theoretical [114] and experimental [5] studies which show that V1 receptive fields - both simple and complex - are comprised of subunits [160, 122]. The d-DoG-S model simulates these subunits as spatially offset Difference-of-Gaussians. In addition to the influence of the underlying subunit parameters (*fig.* 2.2), the relative strength and spacing of the two subunits can shape the tuning (*fig.* 2.6). In particular, this model offers one mechanism by which the particular spatial arrangement of broadly tuned inputs can give rise to the sharper spatial frequency tuning observed in cortex. Most importantly for our analysis of frequency selectivity, the d-DoG-S model shares with its simpler DoG counterpart the feature that the center radius alone determines the high-frequency cut-off.

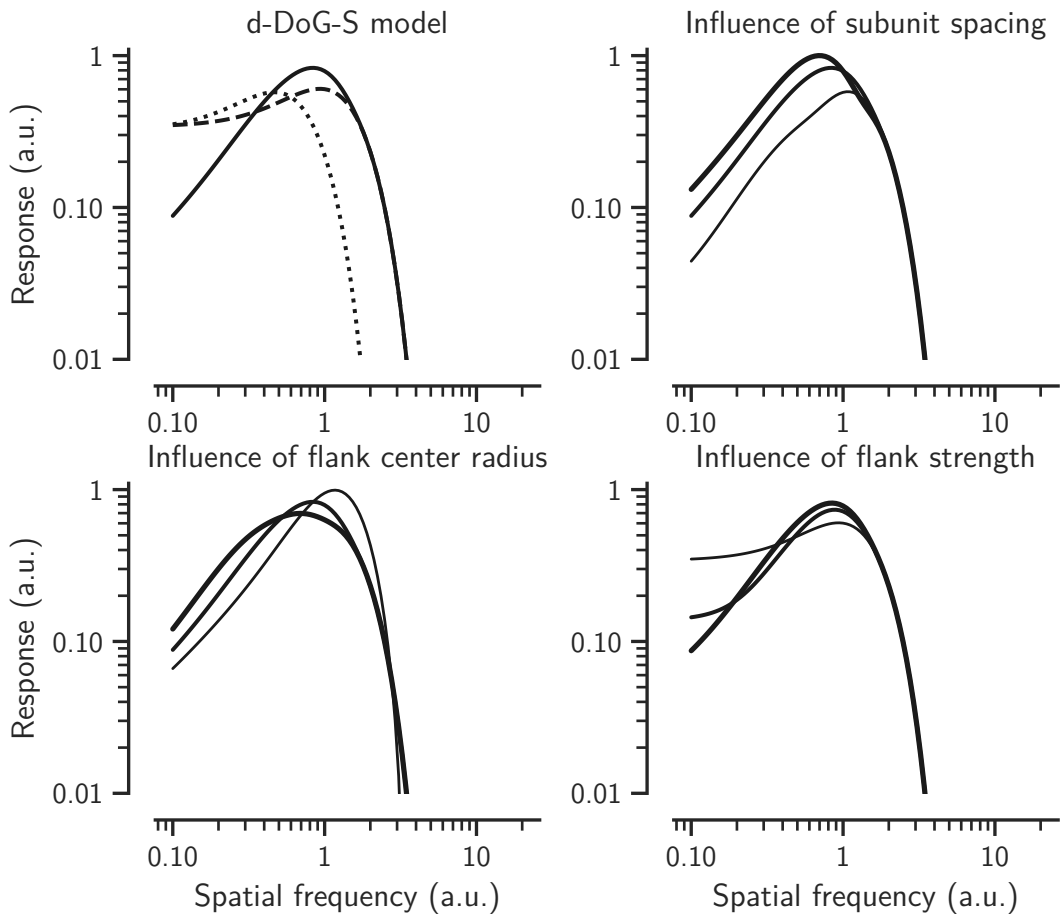

Figure 2.6: Simulations show the influence of various parameters in the d-DoG-S model. *Top left:* The central DoG (*dashed*), flanking DoG (*dotted*) and resultant model (*solid*) tuning curve in the frequency domain. The spatial offset of the DoG subunits creates bandpass tuning from two low-pass inputs. For the remaining three sub-plots, increased line thickness corresponds to an increase in the given parameter. *Top right:* Spacing the subunits further apart shifts the preferred frequency lower, but leaves the high-frequency limb of the tuning curve unaffected. *Bottom left:* Increasing the center radius of the flanking DoG again changes the preferred frequency and magnitude of tuning, but leaves the high-frequency cut-off unchanged. *Bottom right:* Increasing flank strength creates sharper tuning on the low-frequency end.

First, we ensured that our dataset was sufficiently rich enough to justify the extra parameters associated with the d-DoG-s model. We imposed constraints to simplify the optimization and interpretation of the model (see §2.2.4), and thus our d-DoG-S model had only 3 additional parameters compared to the DoG model. These three parameters granted a roughly 13% increase in explanatory power without resulting in overfitting (t-test on the distribution of relative cross-validation test errors yields $p = 0.16$, i.e., the models generalize

to unseen data comparably). Thus, we fit the d-DoG-S model to our V1 data rather than the simpler DoG model.

Nonetheless, the d-DoG-S model is a super-set of the DoG model, and some cells are best fit with a weak or absent flanking DoG. Though there were no statistical differences in the overall population distribution of flank strength between simple and complex cells (median strengths of 0.47 and 0.81, respectively; $p = 0.11$, t-test), we did observe that weak surrounds were more common in simple cells. Specifically, weak flanking DoG mechanisms were more common in simple cells (53% of neurons had a flanking DoG no more than half as strong as the central DoG) than in complex cells (30% with a weak flanking DoG), and model fits with effectively no second subunit (flank strength <= 5% of the central DoG) were again more prevalent in simple cells (13% of neurons, or 6 out of 45) than complex cells (only 2 neurons out of 44).

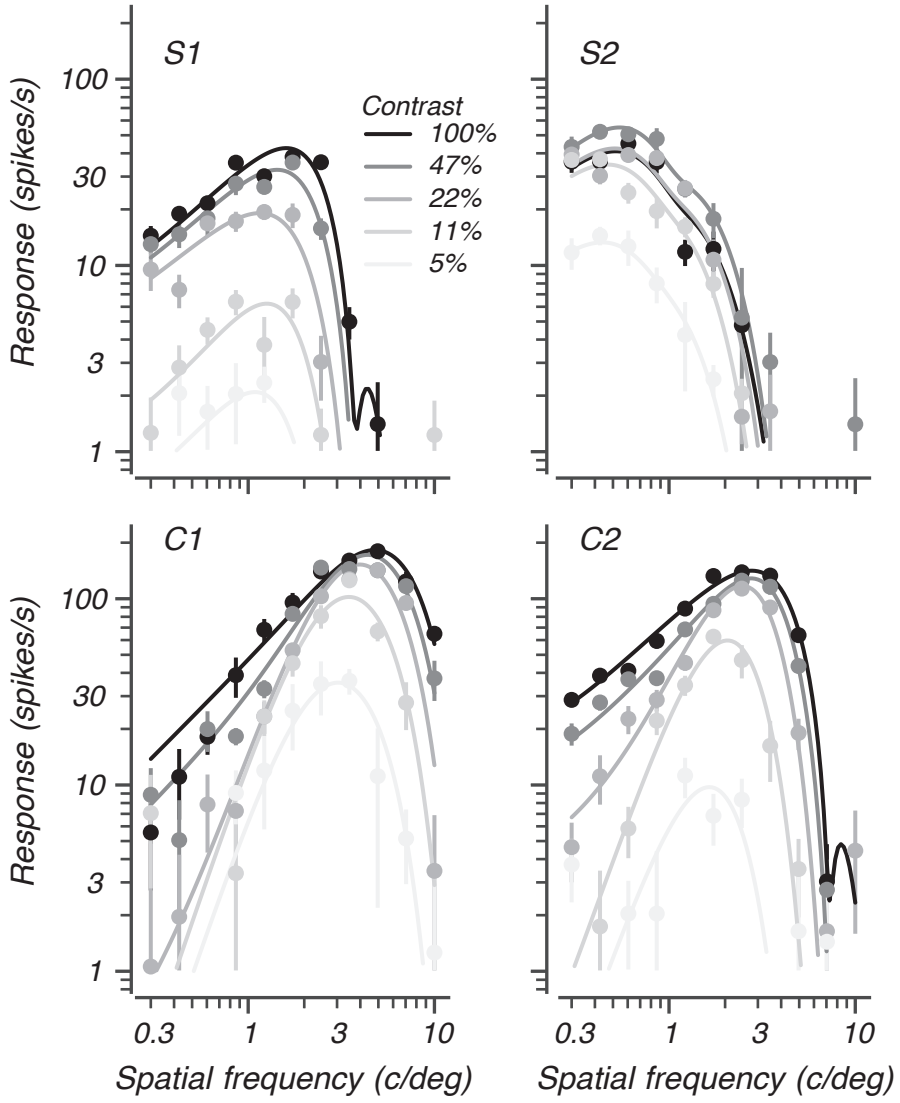

Figure 2.7: Four example V1 neurons with mean response and standard error of the mean (points and error bars, respectively) plotted on equal scale, double log axes. The tuning curves shown (lines) are d-DoG-S models fit jointly across contrast. The two neurons on the top are simple; the bottom two are complex cells. In each case, we depict only a subset of tested contrasts (skipping every other) for better legibility: 0.05 to 1.0 contrast for all neurons.

V1 neurons show contrast-dependent spatial frequency tuning

Example data and model fits (*fig.* 2.7) highlight the ability of this simple model to capture the variety of tuning behaviors in V1, both across cells and contrast. As with LGN neurons, V1 neurons show contrast-dependent spatial frequency tuning. For the same four example neurons, we show the trajectory of characteristic frequency with contrast (*fig.* 2.8). These

examples are reflective of the broader population. The simple and complex cell distributions are not significantly different (p=0.20, t-test), though the higher complex cell median (0.20 vs. 0.12 for simple cells) reflects the greater prevalence of steep slopes among complex cells compared to simple cells. More importantly, we see that the V1 population shows shifts comparable to, if not larger, than those observed in the LGN.

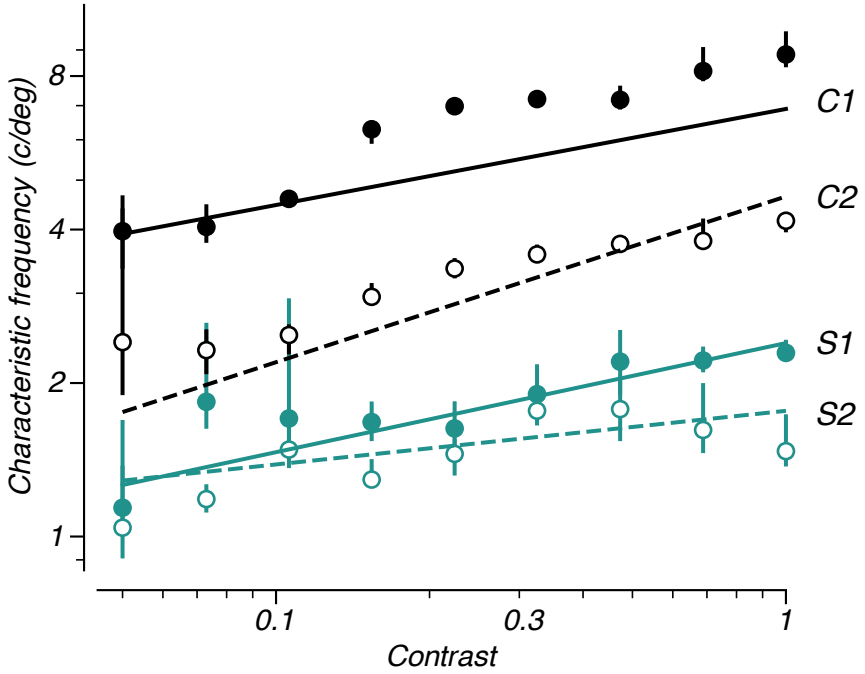

Figure 2.8: The characteristic frequency estimates (datum and error bars showing median and interquartile range; values from the 'free' model) and slopes (lines; values from the 'Δr_c*(slope)*' model) for the same example neurons in *fig.* 2.7. From top to bottom, the slopes are 0.11, 0.33, 0.19, 0.16. All slopes were significantly greater than zero by bootstrap analysis.

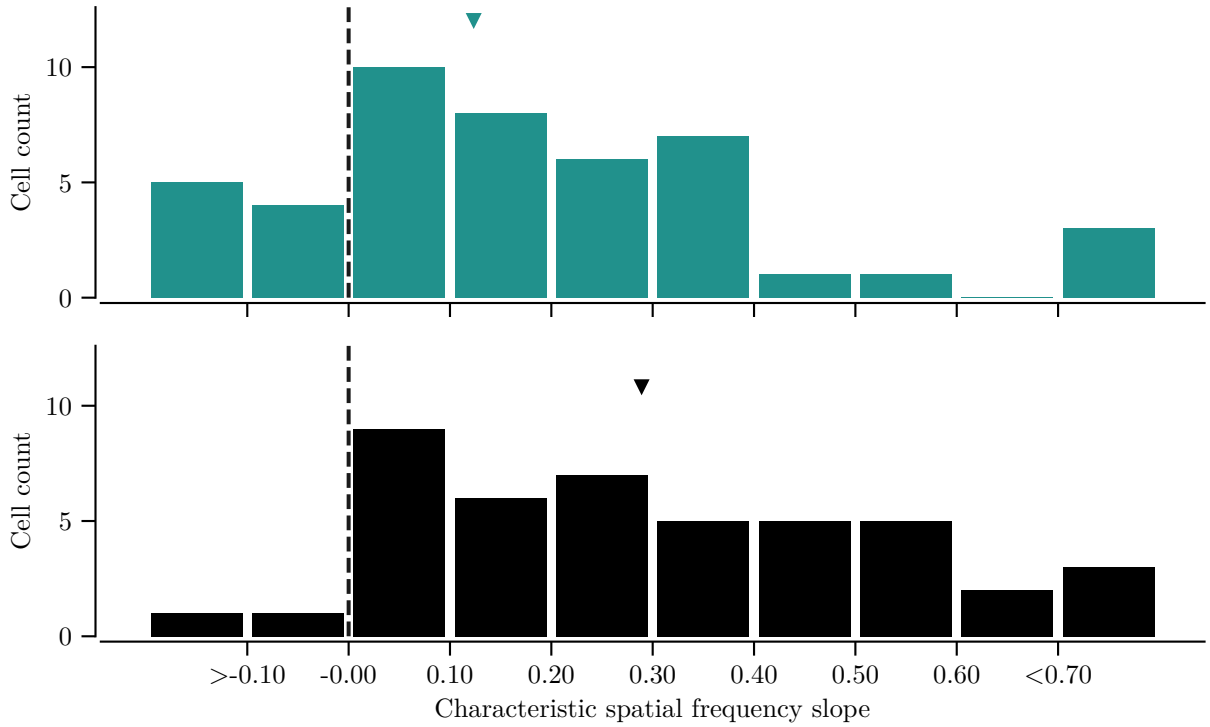

Figure 2.9: Distribution of the characteristic frequency slope for simple (*top*) and complex (*bottom*) V1 neurons. The distributions were not significantly different ($p = 0.2$, t-test), though the complex population median (0.20) was greater than the simple median (0.12). The two distributions were individually shifted away from zero ($p <$ 1e-3).

We find that V1 neurons, like those in the LGN, have lower bandwidths at low contrast than at high. We again evaluated the exponent of a power-law relationship between log-contrast and log-octave bandwidth for all conditions with a defined bandwidth. We find a 13% reduction for complex cells and a more modest (although not significantly so) 5% reduction for simple cells. To compare with a previous study [126], we also evaluated the discrete ratio of half-height octave bandwidths measured at the highest contrast and 25% of the maximum (the closest relative contrast to their 20% value for low contrast). With this metric, we report a geometric mean of 1.13, i.e., a 13% reduction from high to low contrast, in good agreement with their mean value of 1.24 (our non-geometric mean is 1.17). We again found the bandwidth to reduction to be larger for complex cells (1.15) than for simple cells (1.10), though as with the continuous metric, the populations were not significantly different ($p > 0.05$). As with the LGN model fits, we observed stronger surrounds at low contrast.

Furthermore, there was a marked correlation between the magnitude of reduced surround strength at low contrast and greater reduction in bandwidth ($r = -1.02, p <$ 1e-8).

2.4 Discussion

There has been extensive research on how the receptive fields of LGN and V1 neurons are influenced by image contrast. Our findings expand upon this work by using a common stimulus set and modeling framework to characterize the influence of contrast on spatial frequency tuning in LGN and V1. Most importantly, our approach allows for a quantitative comparison of tuning shifts in the two areas. We highlight two key trends - the reduction in high spatial frequency response and bandwidth with contrast - and relate them to changes in receptive field structure, namely larger center mechanisms and stronger surrounds at lower contrast. We find the tuning shifts in cortex to be as large, if not greater, than those in the LGN. These measures, inferred from fitted receptive field models which share the common difference-of-Gaussians subunit, suggest that tuning shifts in the LGN propagate forward to cortex, where they are further enhanced by local cortical computation.

2.4.1 Comparing LGN and V1 tuning shifts

We found no qualitative difference in LGN tuning shifts between magno and parvocellular neurons. We also found no significant difference between simple and complex cells in cortex. However, comparing all LGN neurons against all V1 neurons, we find the latter show significantly greater shifts in characteristic spatial frequency (*fig.* 2.10; $p < 0.05$, Mann-Whitney U test). Among pairwise comparisons between classes, only V1 complex cells show significantly different slopes, both to magno and parvo cells. The roughly two-fold magnitude of the (median complex cell slopes of 0.20 compared to 0.10-0.12 for LGN neurons) is in agreement with a previous study [148] which, though not recording from both areas, compared their LGN size shifts to those from a similar V1 experiment [127]. As evidenced by

the distributions in *fig.* 2.10, the variation in V1 shift size is greater than in either mango or parvo cells. For example, the least substantial shifts in both mango and complex V1 cells are comparable; however, we observe shifts at least twice as large in complex cells. In addition, we find a portion of simple cells with either no shift or even a reverse shift in spatial frequency tuning, such that the selectivity for high spatial frequencies is improved at lower contrasts.

We argue that the shifts observed in V1 are derived, in part, from the LGN inputs. That argument is bolstered by our use of a V1 model comprised of mechanistically-realistic, LGN-like inputs (i.e., Difference-of-Gaussians). Since our read-out measure is based on the center size of those sub-units, we were able to directly compare the relative effect size, and find strong support for V1 shifts which are at least as large as those in the LGN, and often larger. Besides the intuition derived from our model, previous modeling efforts also show that V1 spatial frequency cut-offs are determined by the LGN afferents [116]. Under such conditions, the loss of selectivity for high spatial frequencies at lower contrast will naturally propagate forward to V1 tuning.

Briefly, we comment on the circuit mechanisms that are necessary to support such shifts in tuning and receptive field size. As hypothesized in [127], weakened lateral connections due to depressed intracortical synapses at high contrast can explain the smaller cortical receptive fields at high contrast. However, because corticogeniculate feedback is typically restricted to the spatial extent of the recipient LGN field, [7] such a mechanism is unlikely to account for the intrathalamic changes. Nolt et. al. [102] hypothesize that thalamic changes likely originate in the retina: either due to changes in inter-photoreceptor gap junction strength or inner plexiform coupling strength. In the fovea, midget cells (i.e. parvocellular-projecting) often pool from only one photoreceptor; even then, however, evidence shows that connectivity mediated by gap junctions can result in more than one photoreceptor influencing a given midget RGC [33]. This allows for the possibility that even foveal parvo cells can still show contrast-dependent tuning shifts (perhaps more minimally than more eccentric cells) despite receiving the bulk of their input from a single photoreceptor. Nonetheless, a more careful

inspection of tuning shift effect size with eccentricity can help confirm or rule out some of these mechanistic hypotheses.

We also know that LGN and V1 spatial frequency tuning are starkly different, and it is therefore unsurprising that their contrast-dependence might also differ, at least in magnitude. Thus, we argue that further intracortical processing enhances the LGN-derived shifts. This hypothesis is bolstered by a concise, physiologically-inspired model developed by Tailby et al. [154], in which they show that cortical neurons can inherit the contrast dependent size tuning of realistic thalamic input. However, the authors take care to clarify that certain properties of V1 tuning (in particular, the spatial frequency and orientation specificity of suppression) rule out the possibility that such an LGN-derived model could account for all observed cortical shifts. Previous electrophysiology studies have also characterized the influence of LGN inputs on cortical tuning, and demonstrated how cortical computation further alters the low-pass inputs. This influence notably includes the sharpening of spatial frequency tuning by means of a delayed-onset suppression at low spatial frequencies [21]. Earlier onset, untuned suppression also contributes to the sharpening of tuning [133, 168], though from within the classical receptive field extent. These two suppressive mechanisms could contribute, in part, to the steeper tuning shifts observed in V1. Likewise, in the orientation domain, tuning starts as broad - reflecting untuned LGN inputs - and becomes sharper through suppression [117], suggesting that suppression of non-preferred stimuli is a common computational motif. In the context of our results, cortical suppressive fields with tuning for lower spatial frequencies [59] could contribute to the exacerbation of the inherited shifts through selective suppression of low spatial frequencies at higher contrast. A tuned mechanism with these properties could also serve as an explanation for the different contrast gains observed as a function of frequency (*fig.* 2.12, *right*), and in particular, the relatively higher contrast gains at high spatial frequency.

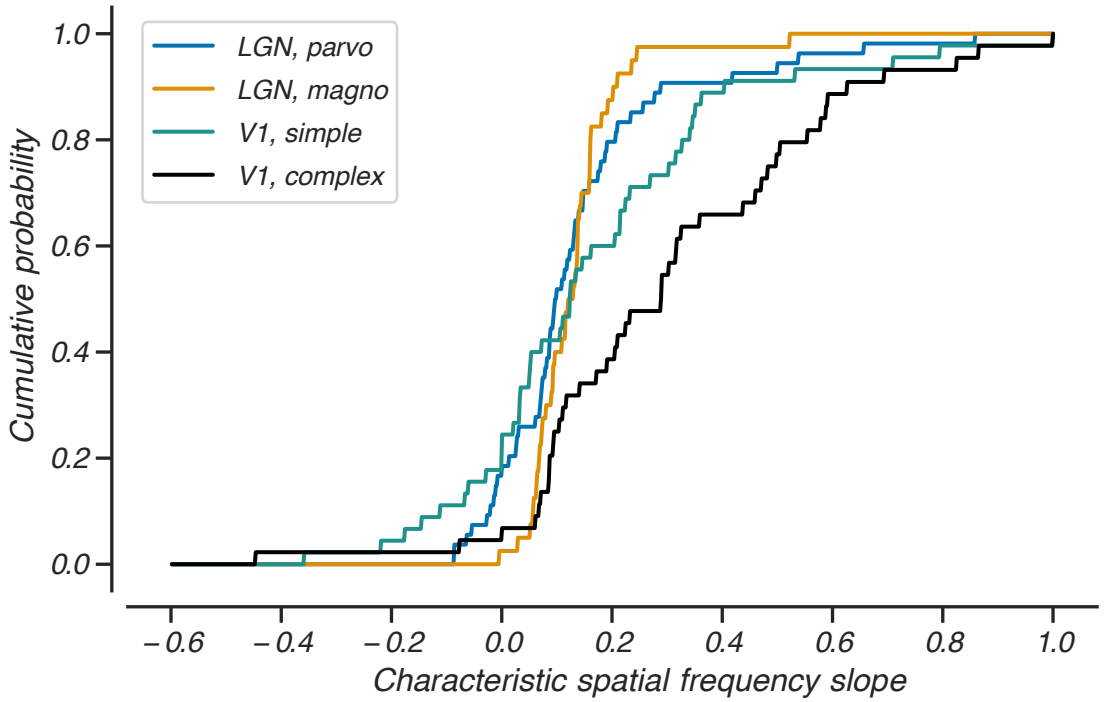

Figure 2.10: Cumulative distribution functions for the characteristic spatial frequency slopes, split by cell class. Unsurprisingly, given the distributions shown in *fig.* 2.5 and *fig.* 2.9, V1 cells (as a whole) have significantly steeper slopes than those in the LGN ($p < 0.05$, Mann-Whitney U test), though only complex cells are steeper than magno and parvo, when compared by class.

2.4.2 Laminar differences

Due to the substantial processing that happens within V1 (cf. [29]), it is natural to expect laminar differences in tuning that reflect, in part, cortical circuitry. This has been borne out for studies of orientation tuning [21], direction selectivity [54], orientation tuning and strength of surrounds [140], and, more recently, texture selectivity [169]. With respect to spatial frequency tuning, a previous study which characterized the tuning of the extraclassical receptive field found that tuned suppression was stronger in layers 2/3 and 4B [58]. This is in agreement with our preliminary finding of tuning shifts with contrast. Though we do not have histological reconstruction for all V1 neurons in our population, we did survey laminar differences by grouping cells as supragranular (layers 2/3, 4A, and 4B), infragranular (layers 5/6), and those belonging to the input layers (4C). We found that neurons in the

supragranular layers showed the strongest contrast-dependent tuning shifts (*fig.* 2.11). The observation that these effect sizes are at least partially accounted for by V1 layer, and therefore intra-cortical processing, lends further credence to the hypothesis that LGN input accounts for only part of the observed V1 effects.

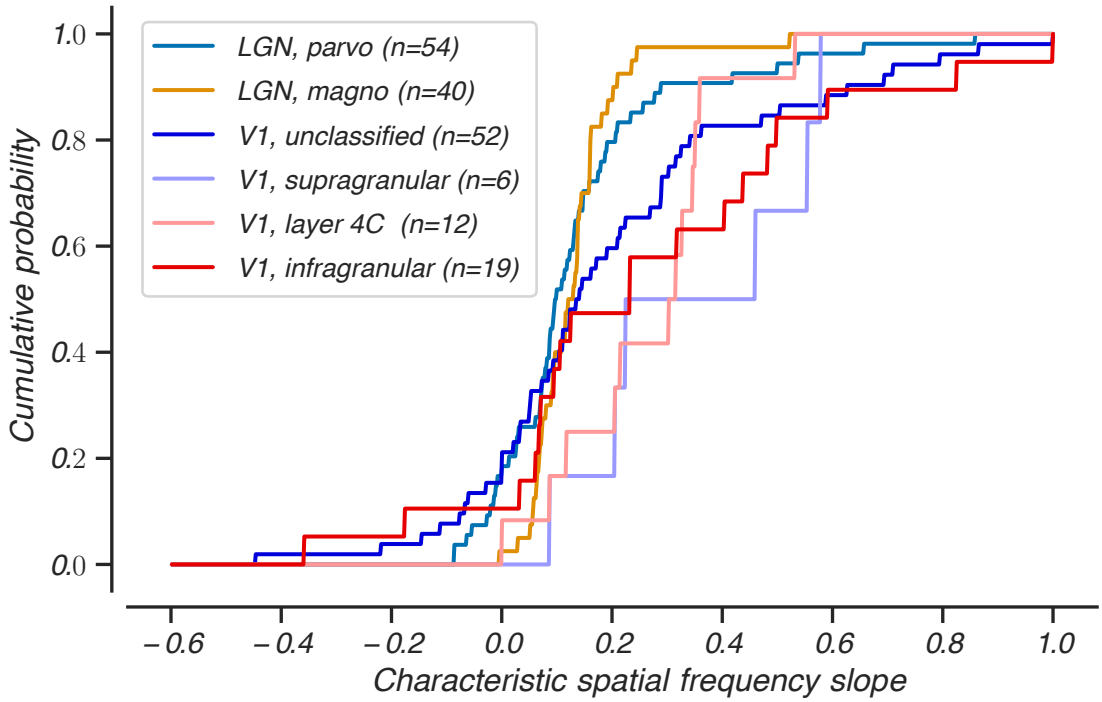

Figure 2.11: Here, we show the cumulative distribution of characteristic frequency slopes as grouped by area and layer. Supragranular layers are more likely to show steep shifts, and infragranular layers show comparable - if not slightly steeper - shifts than V1 input cells (layer 4C). All V1 cell distributions show steeper shifts than either magno or parvo LGN cells.

2.4.3 The relationship between contrast gain control and the observed tuning shifts

Our stimulus set was a two-dimensional matrix of spatial frequency and contrast. The primary analysis was based on fitting receptive field models of spatial frequency tuning to the responses at each contrast. However, the other slice of this 2D matrix allows us to see the response versus contrast (RVC) as a function of stimulus spatial frequency. We plot those data and inferred RVC curves based on the spatial frequency model fits to highlight how the contrast

response is spatial frequency dependent (*fig.* 2.12). The RVCs at different frequencies are not simply scaled versions of one another, suggesting that contrast gain is frequency dependent.

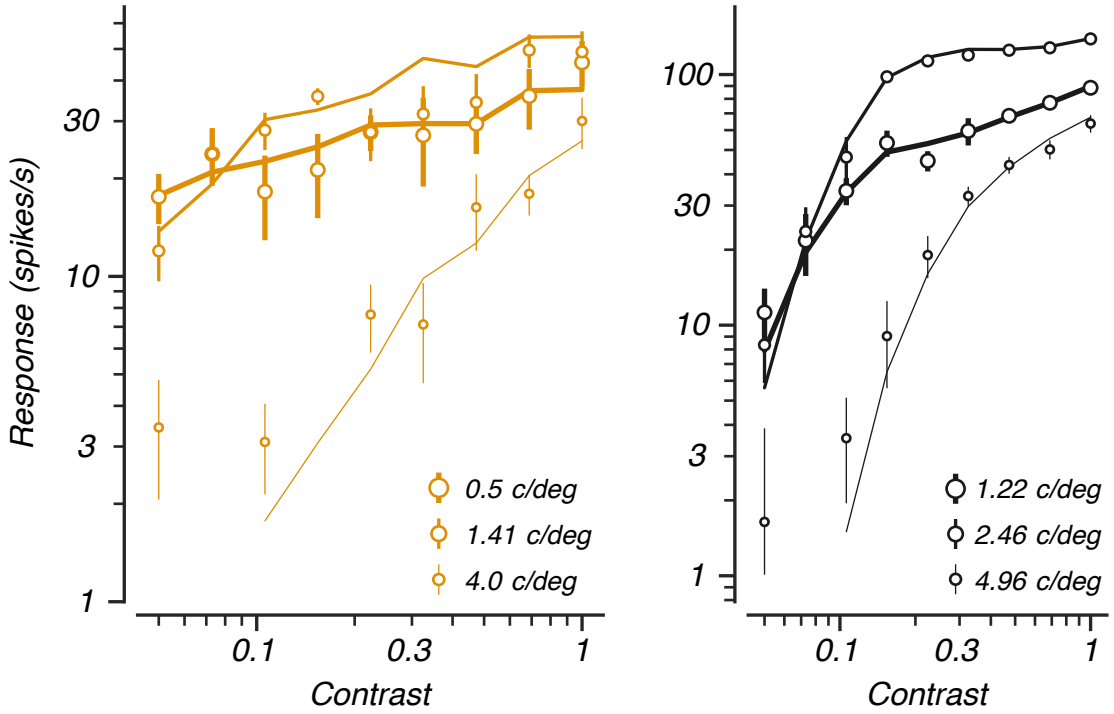

Figure 2.12: Response-versus-contrast curves for an LGN magno (*left*) and V1 complex cell (*right*) shown at three spatial frequencies, each. Model curves are drawn from the spatial frequency model fits, evaluated at each specific SF and contrast. The mean and standard error datum are from the spatial frequency tuning curves, but now shown as a function of contrast, split by spatial frequency. Spatial frequency increases are denoted by thinner curves and smaller datum. We show three spatial frequency values, highlighting the RVC below, near, and above the cell's preferred frequency. The RVC curves take different shapes at each spatial frequency, suggesting a contrast gain control mechanism which is frequency-dependent.

The disparate RVC shapes across spatial frequency are shown for one example magno cell (*fig.* 2.12, *left*). The saturation reflects the action of a contrast gain mechanism, and thus the distinct shapes suggest a frequency dependent contrast gain mechanism. One possible LGN mechanism which could account for this differential gain as a function of frequency is the stronger suppressive tuning observed at low spatial frequencies [19], at least in cat. Such an interpretation is in agreement with Mante et. al. [88], who attribute the contrast gain properties observed in the LGN to the suppressive field which is thalamic in origin.

Furthermore, frequency-dependent gain control has been observed for other stimulus domains, namely temporal frequency in the macaque retina [134, 15] and LGN [40, 78, 112].

Magno cells show stronger contrast gain effects [15], which would suggest stronger tuning shifts for these cells as opposed to parvocellular neurons. Interestingly, we do not find evidence for that disparity in our data, but we cannot discern whether this is due to the noisier parvo estimates (*fig.* 2.5) or the additional influence of other mechanisms on the shift in characteristic spatial frequency.

The non-uniformity of RVCs across spatial frequency is more pronounced in our example V1 complex cell (*fig.* 2.12, *right*). Here, the RVCs "cross", suggesting very different gains and thresholds as a function of spatial frequency. As for the possible mechanisms which give rise to gain changes with contrast, there is rich literature on contrast gain control and suppression, more broadly, in cortex. In particular, suppression is stronger at lower spatial frequencies and the nature of the its tuning for spatial frequency depends on the contrast in the surround [162], suggesting that this mechanism acts, in part, as a form gain control.

Chapter 3

Measuring the tuning of spatial frequency suppression

3.1 Introduction

Longstanding efforts have sought to understand the transformation of visual representation that happens between the lateral geniculate nucleus (LGN) and primary visual cortex (V1). These studies have characterized the spatial arrangement of LGN inputs to V1 [113, 5]; the relative contributions of feedforward thalamocortical [7, 124, 105] and intra-cortical circuitry [6, 97]; and the functional properties (e.g., tuning) of the excitatory and inhibitory mechanisms [100, 137, 18, 37] that shape V1 selectivity. For spatial frequency in particular, neurons in primary visual cortex are more sharply tuned for spatial frequency [62, 36] than their LGN inputs [70, 118]. One of the earliest studies on the role of suppression in refining cortical selectivity found suppression at high spatial frequencies to be most common [35]. Later reports suggested instead that suppression is matched to the cell's tuning - i.e., stronger at higher frequencies when a cell is tuned for low frequencies, and vice versa [13, 100]. However, across a range of stimulus protocols and in primates, particularly, experimenters have now converged on the understanding that suppression at higher frequencies is less

common. Instead it is stronger suppression at low spatial frequencies which shapes tuning: this has been demonstrated with reverse correlation stimuli [21, 59] and with masking gratings [144].

A deeper understanding of how spatial frequency selectivity is shaped across the visual hierarchy requires discerning *de novo* properties from those which are inherited from feed-forward inputs. Crucially, the finite spatial extent of receptive fields ensures that spatial frequency selectivity is present at all stages, complicating the identification of specific influences. However, great progress has been made on this front with respect to the emergence of more narrow spatial frequency tuning in cortex. Careful examination of the temporal dynamics of spatial frequency selectivity highlights an initially broad tuning which is refined over the response time course due to the late-onset suppression of low spatial frequencies [21, 91]. Furthermore, isolating suppression from the receptive field surround reveals a tuning which is contrast-dependent: orientation-selective and more narrowly frequency tuned at high contrast; broader and unselective for orientation at low contrast [162]. Taken together with the properties of LGN tuning for orientation (untuned) and spatial frequency (low-pass), these results motivate the hypothesis that some - but not all - suppression is inherited from subcortical inputs to cortex. More recent work has built upon these findings, using histological analysis and a profiling of the spatial extent and temporal dynamics of suppression in V1 to offer a more precise hypothesis: namely, that untuned suppression originates largely from magnocellular input while tuned suppression is cortical [58, 59]. A more prominent role for magnos in driving cortical suppression is in alignment with their stronger non-linearities compared to parvos.

The aim of the experiments described below is twofold: first, to provide a comparison between suppression as observed with two distinct stimulus sets; and the second is to help constrain a computational model of tuned gain control which we cover later in the thesis. Our first experiment directly tested the linearity of summation, and differed from many previous studies by presenting broad spectral content, with some stimuli spanning up to

2 octaves in spatial frequency. Given the relatively narrow tuning of excitation compared to suppression in V1, we hypothesized that such a stimulus would elicit strong suppressive effects. The second experiment was a masking stimulus modeled after a previous study on the refinement of spatial frequency tuning [13]. Our interest in gain control motivated a more systematic sampling of mask contrast than in many previous studies; additionally, the previous experiment was performed in cat, and thus our work allows us to relate that study to more recent research conducted in primates. Additionally, these stimuli were presented as part of the same experiment described earlier (ch. 2), permitting a comparison between suppressive effects and contrast-dependent changes in spatial frequency tuning.

We find strong agreement between the spatial frequency tuning of suppression we report here and previous reports in primate. The mixture summation experiment reveals a more prominent role of suppression for V1 spatial frequency tuning compared to in LGN. We also compare the strength of suppression's frequency dependence and other tuning properties, finding, finding correlations with surround suppression and contrast-dependent tuning shifts. The masking experiment offers a finer resolution characterization of suppression's spatial frequency and contrast selectivity, through which we find low-frequency suppression which becomes more broadband at higher contrast. We end by discussing the differences in results between the experiments, which we argue likely derive from both design-dependent properties and the engaging of different mechanisms.

3.2 Methods

For an overview of experiment preparation, recording techniques, spike sorting procedure, and other general methods, see the General Methods section of the thesis. Briefly, all data reported below are from anesthetized macaque experiments, and were collected from multi-unit electrode arrays (V1) or single-unit electrodes (all LGN data and a few V1 cells). We first characterized the basic tuning characteristics of each cell before running the primary

experiments described below. We analyzed the DC or F1 response based on the cell's classification [143]. For stimuli with multiple sinusoidal gratings, the overall F1 response was taken as the sum of the F1 measured for each component.

For both stimulus sets described below, we use drifting sinusoidal gratings presented at the cell's optimal size and orientation. Each stimulus presentation lasted either 1s or 2s, with a corresponding number of repeats to ensure 10s total for each unique condition. Temporal frequency was varied, as described in more detail below, but were chosen such that all gratings in a given stimulus presentation had unique, non-harmonic temporal frequencies. Furthermore, each particular spatial frequency was only presented at one temporal frequency throughout the experiment, regardless of the overall stimulus configuration.

For LGN neurons, we only presented the mixture stimulus set described below (§3.2.1). In V1 neurons, we recorded either the mixture (§3.2.1) or masking (§3.2.2) experiment, but never both for the same cell.

3.2.1 Stimulus set I: Mixtures

The mixture stimulus set was based on a 2D matrix of sinusoidal gratings that varied in spatial frequency and contrast. The spatial frequencies and contrasts were log-spaced, and spanned the range of selectivity for most neurons: typically 11 spatial frequencies spanning 4 or 5 octaves, and 9 contrasts ranging from 0.05 to 1. Stimuli were presented in a pseudo-randomized order, and consisted of single gratings (described previously in §2.2.2), and superpositions of multiple gratings. The superposition mixtures were comprised of 3, 5, or 7 individual gratings, and each mixture was uniquely specified by the central spatial frequency, total contrast, and number of gratings used (*fig.* 3.1). The spatial frequencies were chosen to be sequential from the set of frequencies $\{f_i\}$, e.g., $\{f_{j-1}, f_j, f_{j+1}\}$ for three gratings.

The total contrast for mixtures was given as the linear sum of the component contrasts. The mixtures were presented at 4, 3, and 2 total contrasts for mixtures of 3, 5, and 7 components, respectively. The total mixture contrasts were drawn from the same set of

values as used for single gratings, descending from 1, 0.69, 0.47, and 0.33. In all cases, the contrasts of the flanking (i.e., non-central) gratings fell-off as a function of distance from the central spatial frequency (*fig.* 3.1*C*), and the relative contrast of each grating compared to the central grating was constant regardless of the total contrast. The relative contrasts were approximated by a Gaussian centered at the central spatial frequency with an increasing width for mixtures with more components. To ensure that all components of mixtures were also presented in isolation, each grating had a contrast nearest to that which was used for the isolated stimuli. The temporal frequencies were chosen from a set of non-overlapping, non-harmonic values near the cell's preference. Further, we chose this set to ensure an integer number of cycles for each presentation. For example, with a temporal frequency preference of 8Hz and 2s stimulus presentations, the 7 temporal frequency values would span from 6.5-9.5Hz in 0.5Hz steps. This arrangement allowed us to average over the relative phase arrangements between gratings, and also permitted a unique identification of the response to each individual grating for simple cells, even when presented as a compound stimulus.

Quantifying mixture summation

Each component of a given mixture stimuli was also presented in isolation, permitting a test of superposition, or linearity. For DC values, all responses (i.e., both mixtures and isolated gratings) were baseline subtracted. For LGN data and V1 simple cells, no such adjustment was necessary, since the F1 already reflects only stimulus-driven activity. The response-summation analysis which follows is based on work by Levine and Abramov [77]. In their work, they considered the response to illumination on one or two non-overlapping spots of the goldfish retina. Rather than separating our stimuli in space, we performed a superposition test in spatial frequency. Response summation for a given mixture was evaluated by comparing the sum of its component responses when presented in isolation (predicted response) to the

response to the mixture stimulus as a compound. That is,

$$r_{pred} = \sum_{i=1..n} r(g_i) \tag{3.1}$$

$$r_{mix} = r(g_1 + g_2 + ... + g_n) \tag{3.2}$$

where r(g_i) is the response to a single grating, g_i, drawn from the set of contrasts and spatial frequencies. Then, the sublinearity of the response, s, could simply be measured as

$$s = \frac{r_{mix}}{r_{pred}} \tag{3.3}$$

Unsurprisingly, many neurons in the LGN and in V1 show sub-linear summation for mixtures, i.e., $s < 1$ for many stimuli. This is akin to the saturation seen in a contrast-response function, where now the predicted response is a proxy for overall input drive, i.e., contrast in the typical view of response saturation.

Our primary goal was to analyze how response summation might be dependent on stimulus properties like the mixture's total contrast and its spatial frequency content. Furthermore, we needed to isolate the influence of the stimulus properties from the input drive, per se, i.e., r_{pred} (also referenced as r_p). Thus, our benchmark against which to compare summation for a given condition was computed relative to the average saturation rather than linear summation. To determine a cell's overall saturation profile, we fit a Naka-Rushton function to the mixture responses, r_{mix}, based on each mixture's predicted response, r_p:

$$\hat{R}_{mix}(r_p) = R_0 * (\frac{r_p^n}{r_p^n + r_{50}^n}) \tag{3.4}$$

However, we fixed the exponent, n, at 1, reducing the fit to two parameters - overall gain, R_0, and the r_{50}. Here, the predicted response serves the role that contrast typically plays in the Naka-Rushton function, and thus r_{50} is equivalent to c_{50}, the semi-saturation contrast. This model allows us to then predict the mixture response given the response to its components.

We evaluated the quality of this model fit by computing the explained variance (*eq.* 1.1) of the model's predicted mixture response against the measured mixture responses.

Inclusion criteria We performed 250 bootstrap repetitions (resampling trial-by-trial responses) and calculated the distribution of explained variance across each fit. Only cells where at least 95% of bootstrap fits explained at least some of the variance (i.e., explained variance $> 0\%$) between the predicted and measured mixture responses were included in the population summaries.

Overall summation index We devised a single metric to capture the non-linearity of each cell's mixture summation, i.e., the mapping between the response generated from a stimuli's inputs and the response to their superposition. To measure the overall summation profile, we measured the area under the Naka-Rushton fit and compared it against the area under a linear prediction for the same responses. The area beneath a given mixture prediction curve, $\hat{R}$, is computed by

$$A_{\hat{R}} = \int_0^R \hat{R}(r_p)dr \tag{3.5}$$

where R is the largest sum of mixture components, i.e., $\max(r_p)$. For the linear prediction, $\hat{R}_{lin}(r_p) = r_p$, whereas our saturation model prediction ($\hat{R}_{mix}$) is given in *eq.* 3.4. Finally, the overall summation index, S_i, is given as

$$S_i = \frac{A_{mix}}{A_{lin}} \tag{3.6}$$

A value of S_i near 1 reflects a cell which sums its inputs linearly across the range of tested stimuli. Lower S_i reflect a stronger contrast gain mechanism, i.e., a consistent tendency for inputs to add sub-linearly ($r_m << r_p$).

Suppression index The model in *eq.* 3.4 and the associated summation index capture the summation averaged across all conditions. To analyze the deviation from the overall summation for a given condition, i, we compute the error between the mixture response and the prediction based on its inputs, i.e., $\delta_i = r_{mix} - \hat{r}_{mix}(r_p)$. To better compare across stimuli that drove different response amplitudes, we computed a normalized suppression index for each condition:

$$\delta_N = \frac{r_{mix} - \hat{r}_{mix}}{r_{mix} + \hat{r}_{mix} + 2\epsilon} \tag{3.7}$$

where

$$\epsilon = std(\sum_i \delta_i)$$

i.e., the standard deviation of the residual error across all conditions for a given cell. This small factor, ϵ, helps put δ_N in terms of the overall variability in δ_i for a given cell, akin to the correction in [111]. That is, more variable cells and noisier conditions (e.g., those with relatively low firing rates) will have more muted δ_N values. In principle, when $\epsilon \approx 0$, the suppression index is bounded between -1 and 1, with positive values indicating relatively less suppression than would be predicted based on that cell's summation. In practice, however, the bounds will be less than ± 1, since $\epsilon > 0$.

As a summary statistic, we compute the weighted variance of the suppression index grouped by spatial frequency, f:

$$f_{\text{var}} = \frac{\sum_f w_f(\delta_f - \Delta)}{(N-1) * \bar{w}} \tag{3.8}$$

where $w_f = \frac{1}{\text{sem}(w_f)}$, $\Delta = \text{mean}(\delta_f)$, and $\bar{w} = \text{mean}(w_f)$. Larger values indicate a stronger frequency-dependence of summation.

3.2.2 Stimulus set II: Masking

As in the mixture stimulus set, the primary component of this stimulus set was a 2D matrix of sinusoidal gratings that varied in spatial frequency and contrast. Here, the matrix was 7x7, with contrasts ranging from 0.01 to 0.64 in log steps, and frequencies spaced in steps of 0.75 octaves and centered at either 1 or 2 c/deg, depending on the cell's preference. The second primary component of this stimulus set was a base stimulus which was chosen to have a spatial frequency near the cell's preference and a contrast which elicited a strong, but not saturated, response (e.g., near c_{25} or c_{50}). Because the two stimuli were presented simultaneously for some conditions, we refer to the grating that varied in frequency and contrast as the mask stimulus. In a pseudorandomized order, we presented the base alone, the mask alone, or the mask and base stimuli together. The masks and base were presented at different temporal frequencies, both near the cell's preference, permitting the unique identification of the response elicited from each component for simple cells. Each mask and mask-base pair was presented for a total of 10s across 5 or 10 stimulus repetitions; the base alone was overrepresented, and had 70s of total stimulus presentation.

Measuring frequency-dependent masking suppression

First, we measured spatial frequency selectivity and its contrast dependence for each neuron by analyzing the mask responses. However, this stimulus set was designed specifically to measure the tuning of spatial frequency suppression, and thus we analyzed the influence of the mask on the base stimulus response. Because an accurate measure of the base stimulus response was so crucial, we overrepresented this condition in the experiment. Thus, we had a reliable estimate for the mean and variability of the base stimulus response against which we could measure the influence of the mask.

For simple cells, characterizing suppression was straightforward - we simply measured the response at the base temporal frequency in the presence of different masks. Suppression

manifests as a decrease in the base response below the base response in isolation. Masks which left the base response unaltered from its response in isolation were considered non-suppressive.

For complex cells, however, one cannot disentangle the responses driven by the mask and base stimuli. Thus, the only directly measurable suppression occurs when the response to the mask and base together is less than the response to the base alone, i.e., $R(m+b) < R(b)$.

For all conditions, we quantified suppression for a given mask, m, as

$$S(m) = \frac{R(m+b) - R(b)}{\text{std}(R(b))} \tag{3.9}$$

with larger negative values indicating stronger suppression relative to the base only response.

Criteria for determining which cells showed suppression: To enable a summary of only cells with more marked suppression, we wanted to classify cells by whether or not they showed suppression from the mask stimuli. We set a threshold of $S(m) \leq -0.5$ for considering a given mask m to drive suppression. Then, we used this threshold to evaluate whether suppression was observed for each stimulus in the 7x7 matrix of mask conditions. To avoid including cells without lawful suppression, we counted as suppressed only cells with suppression for at least two adjacent mask stimuli, i.e., the same spatial frequency at adjacent contrasts, or two spatial adjacent spatial frequencies at the same contrast. This reduced the number of suppressed cell designations by nearly half. However, we argue this criterion is justified, since the excluded cells had fewer than half of the number of conditions with measurable suppression (medians of 8 conditions vs. 3, respectively), as well as weaker and more variable base responses (medians firing rates of 15.3 vs. 4.5 spikes/s).

3.3 Results

We designed two stimulus sets comprised of both single sinusoidal gratings and mixtures of those gratings to characterize the spatial frequency tuning of suppression. In both stimulus

sets, we sampled the cell's tuning using single sinusoidal gratings that varied in both spatial frequency and contrast. These experiments richly characterized tuning and elicited a broad range of responses through which we were able to measure the influence of additional stimuli on spatial frequency selectivity. All LGN stimuli were at least 3° in size, with all but one cell at least 4° or larger. In our V1 population, due to stronger surround suppression on average, we chose smaller stimuli than for the LGN. The average stimulus size was 2.2°, with all but 7 cells presented with stimuli at least 1° in size. The average ratio between stimulus size and receptive field size was 2.54.

3.3.1 Measuring summation and its frequency-dependence with mixture stimuli

The mixture experiment permitted a direct test of superposition, in which we could compare the response to multiple superimposed gratings (i.e., presented simultaneously with the same spatial extent) against the responses of those same gratings presented in isolation. The experiment is described in more detail in the methods above (§3.2.1), but we briefly summarize its structure here. The core of the experiment is a 2D matrix of sinusoidal gratings with 11 spatial frequencies (in steps slightly less than 0.5 octaves) and 9 contrasts (log-spaced between 0.05 and 1). We presented either single gratings or mixtures of gratings in pseudorandomized order, always drawing from the components in the core matrix. Mixtures were comprised of 3, 5, or 7 drifting gratings, all with the same stimulus size and orientation, and drawn from adjacent spatial frequencies (*fig.* 3.1*A*). We chose the contrasts of the individual components such that the total linear contrast sum was either 0.33, 0.47, 0.69, or 1 (logarithmically spaced; see *fig.* 3.1*B*). Furthermore, the central grating always had the highest contrast of the mixture, with contrast reducing symmetrically for components with relatively lower or higher spatial frequency (*fig.* 3.1*C*). All mixtures were drifting at different, non-harmonic temporal frequencies. We considered the F1 for LGN neurons and V1 simple cells and the baseline-subtracted DC responses for V1 complex cells.

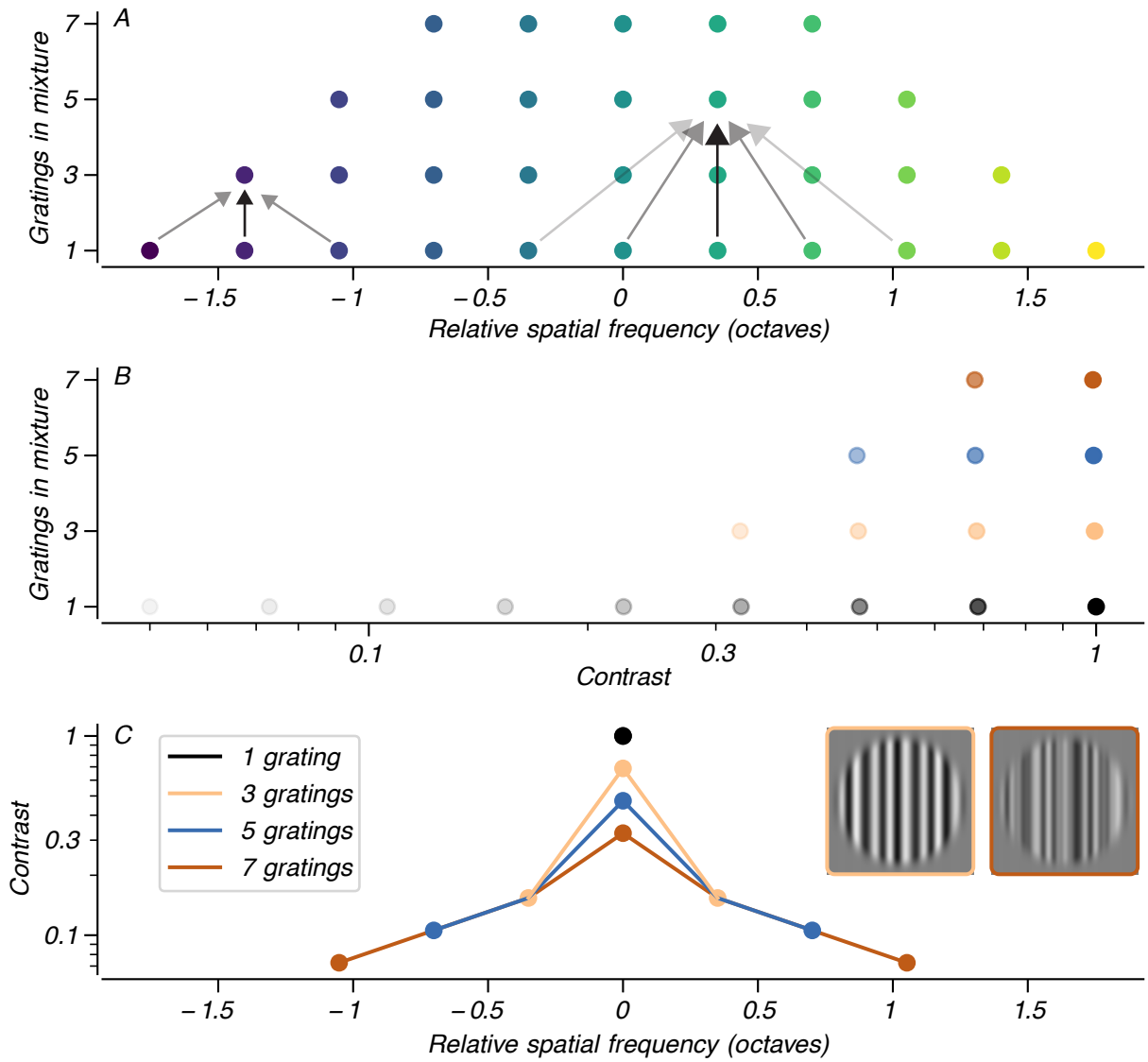

Figure 3.1: Schematic of the mixture experiment, detailing the center spatial frequencies (A) and total contrasts (B) used for each of four dispersion levels. The arrows in (A) show the construction for two example mixtures of 3 and 5 gratings, respectively. The relative contrast for each mixture component (C) is shown for a total contrast of 1. On the right of C, we also show two frames taken from a movie of 3 and 7 drifting grating mixtures. Each condition is uniquely specified by the center spatial frequency, total contrast, and dispersion level (i.e., the number of gratings in the mixture).

As a preliminary step towards determining if there is a spatial frequency dependence to the way that mixtures add, we first assessed the overall summation. A perfect linear prediction would yield a mixture response r_{mix} equal to the sum of its components r_{pred}, and we could capture any deviations from linearity with the simple ratio $\frac{r_{mix}}{r_{pred}}$. In practice, many neurons

in the LGN and V1 show sublinear response saturation, so our analysis instead considers any frequency-dependence relative to a given cell's overall profile of mixture summation.

To better understand how we characterize overall mixture summation, first consider the example cell in *fig.* 3.2, an LGN neuron for which only a subset of the full stimulus set was presented. We highlight two mixtures, each with three gratings and with one component with the same spatial frequency present in both stimuli. Though the sum of their respective component responses (grating prediction, r_{pred}) are similar, the mixture with lower spatial frequencies (*M1*) shows more sublinear summation and $r_{M1} < r_{M2}$. We capture each cell's summation properties by measuring the relationship between all r_{mix}, r_{pred} pairs.

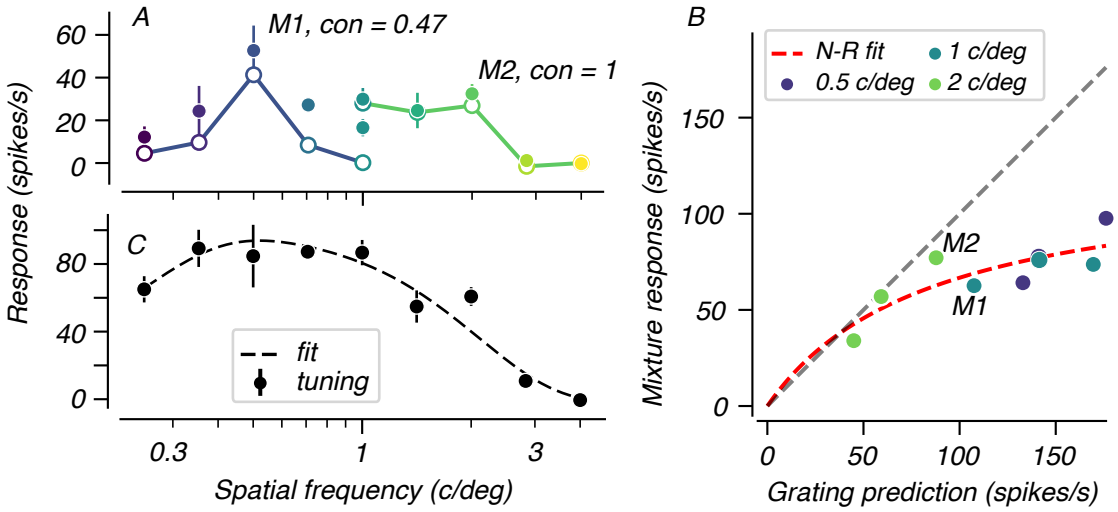

Figure 3.2: Analysis of the mixture summation experiment for an example LGN neuron (magnocellular) tested on a subset of the full stimulus set. The isolated points (*filled*) and connected, solid points (*hollow*) in *A* represent the responses to the same grating presented in isolation and as part of a superimposed mixture, respectively. We only show error bars (standard error of the mean across condition repeats) for the isolated stimuli. The two mixtures, *M1* and *M2*, have different total contrasts and center spatial frequencies. The grating prediction (abscissa in *B*) is given by the sum of the isolated responses to the components that comprise a given mixture; the response to the mixture as a composite is the ordinate. Each mixture is assigned a color based on its center spatial frequency. Here, the components of *M1* & *M2* drive similar responses, but the summation for *M2* is more nearly linear. The overall summation was modeled as a Naka-Rushton function ('N-R'; *red*), while the prediction of linear summation is shown the dashed gray line. For reference, we also include the spatial frequency tuning curve at full contrast (*C*).

Overall summation

To characterize the response linearity for each neuron, we fit a Naka-Rushton function to the full set of mixture responses, r_{mix} (*eq.* 3.4), where the grating prediction, r_{pred} (*eq.* 3.1), acts akin to contrast in the standard contrast-response formulation. Cells with linear summation are those where $r_{mix} \approx r_{pred}$, while cells with stronger saturation have $r_{mix} << r_{pred}$; this is analogous to contrast, where often $r(2c) << 2r(c)$ for a given contrast value c. Then, we summarized the overall non-linearity of summation by calculating the area under the Naka-Rushton fit normalized by a perfectly linear prediction (*eq.* 3.5). We call this ratio the summation index (S_i) with values closer to 1 indicating linear summation and lower values signifying more sublinear summation, i.e., response saturation. As an example, the magno cell in *fig.* 3.2 shows moderate saturation, with $S_i = 0.64$. However, the small set of conditions in this first example highlights the necessity for a sampling of mixtures across a wider range of both spatial frequency and contrast.

For all subsequent analyses, we only consider neurons tested on the full stimulus set. This was only a fraction of LGN cells (13 out of 81 with the full stimulus set; 4 magno, 7 parvo, and 2 unclassified), but included all V1 neurons (63 cells; 43 complex and 20 simple). The two example LGN neurons in *fig.* 3.3 are representative of their respective classes. Both neurons show relatively linear summation for mixtures comprised of lower contrast gratings, or those which do not otherwise drive a strong response (e.g., towards the edge of the cell's spatial frequency passband). However, across the full range of mixtures, the parvocellular neuron (*fig.* 3.3, *left*) shows more linear mixture summation compared to the magno cell (*right*). This visual impression is reflected in their summation indices, and as one would predict on the basis of the typical response-versus-contrast of each class (e.g., [70]). In both cases, the Naka-Rushton fit captures the relationship between the component responses and overall mixture response, accounting for 80% and 86% of the variance, respectively.

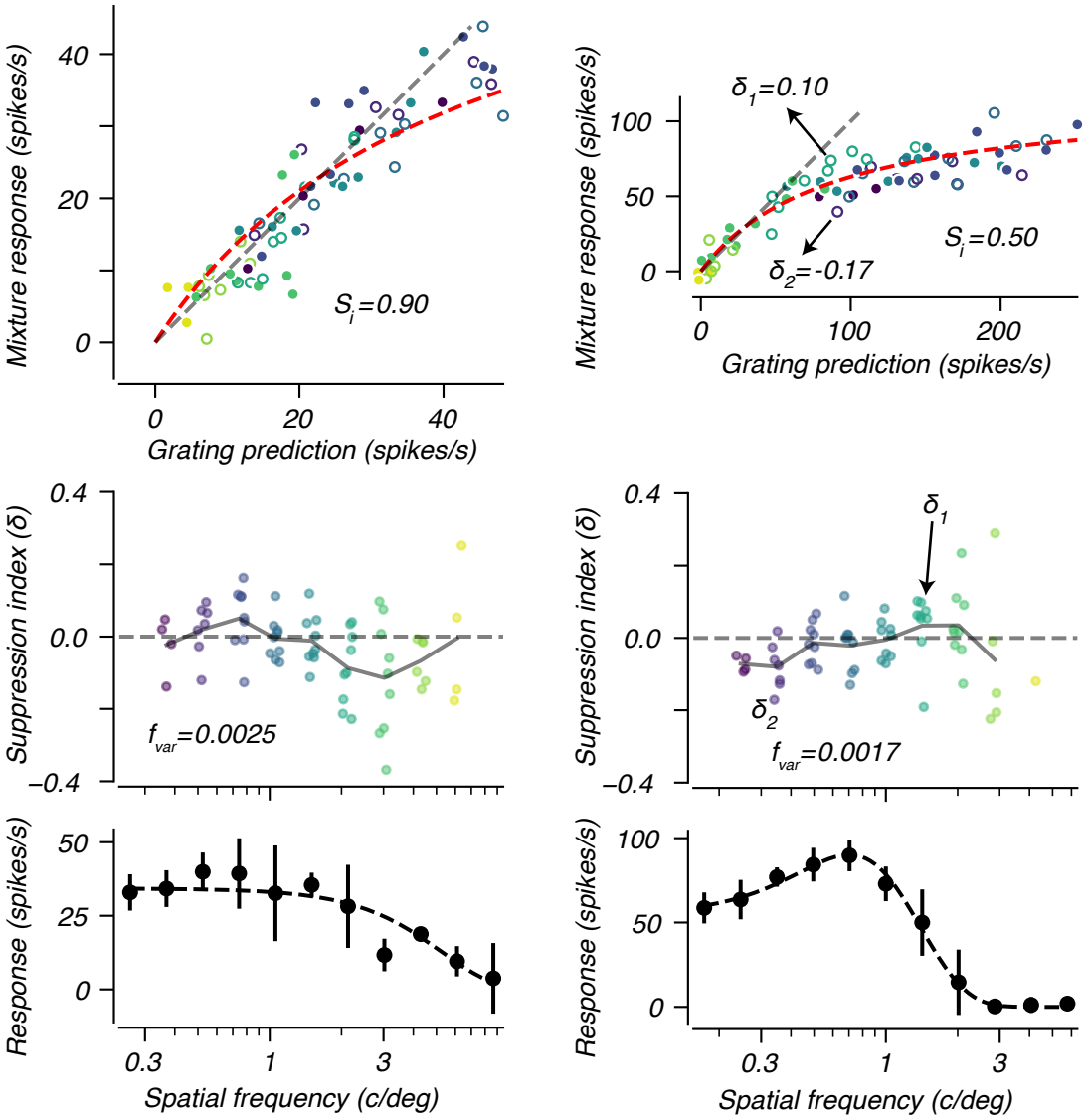

Figure 3.3: Mixture summation for one parvo (*left*) and one magno (*right*) neuron. We annotate two conditions with their suppression index, S_i, in the top right panel. The Naka-Rushton fit to the overall saturation profile captures 80% and 86% of the variance in the two cells, respectively. In the *middle* panel, we show the suppression index per mixture (i.e., unique dispersion-sf-contrast condition), grouped by center spatial frequency with the gray connecting line indicating the average per spatial frequency. The weighted variance of δ_i across spatial frequency is f_{var}. The mapping between mixture center spatial frequency and color is the same in both panels, though we alternate filled and hollow markers in the top panel for enhanced visibility. *Bottom*: For reference, we include the single grating spatial frequency response at full contrast with a Difference-of-Gaussians fit.

Both our characterization of superposition responses and any subsequent frequency-dependent deviations are contingent upon the Naka-Rushton fit to the overall response

summation profile. Thus, we excluded cells whose superposition responses were poorly explained by this model by using an inclusion criterion based on the variance explained by the Naka-Rushton fit (see Methods, §3.2.1). Cells where more than 5% of bootstrap Naka-Rushton fits explained no summation variance were discarded from population summaries. This additional criterion excluded only one additional LGN neuron but eliminated more than half of complex cells (only 20 out of 43 passed this criterion) and 7 out of 20 simple cells. The excluded cells had statistically lower maximum firing rates and less reliable tuning, as assessed by the explained variance of d-DoG-S fits to their single grating responses. The majority of remaining cells in both the LGN and in V1 showed sublinear summation (*fig.* 3.4*B*), with S_i population medians of 0.78 and 0.71, respectively. Parvo cells summed mixtures more linearly (median $S_i = 0.90$) than magno ($S_i = 0.50$); likewise with simple ($S_i = 0.80$) and complex ($S_i = 0.63$) cells in V1, though not statistically so ($p = 0.08$, t-test). Furthermore, there was a strong, positive correlation ($r = 0.64$, $p <$ 1e-5; Pearson correlation coefficient) between S_i and the semi-saturation contrast (c_{50}) as measured for single gratings at the cell's preferred frequency (*fig.* 3.4*C*). Such a result suggests that the nonlinearity of mixture summation can be at least partially predicted based on the cell's contrast saturation.

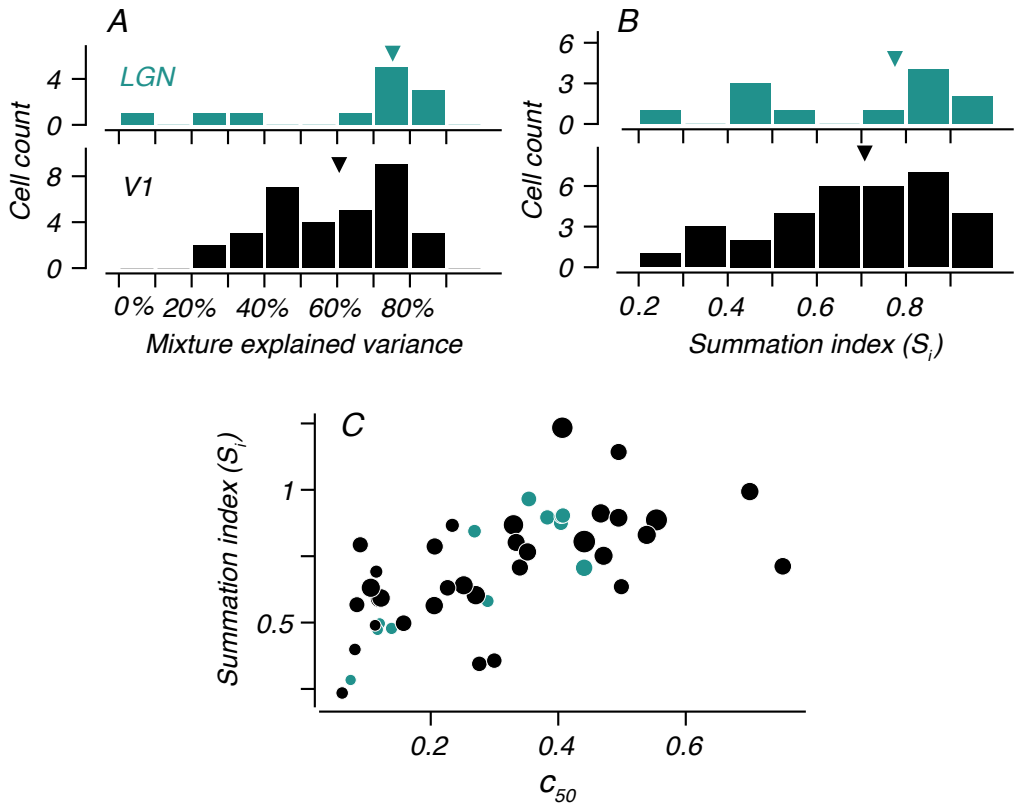

Figure 3.4: Summary of the Naka-Rushton function which we use to characterize mixture summation for 12 LGN (*green*) and 33 V1 (*black*) neurons. (A) Distribution of variance explained by the Naka-Rushton (N-R) fit across cells. (B) Distribution of summation indices, S_i for each cell, based on the N-R fit. (C) Relationship between c_{50} (as measured for single gratings at the cell's preferred spatial frequency) and S_i ($r = 0.64$, $p <$ 1e-5). The size of each point is inversely proportional to the standard deviation of bootstrap S_i values.

Our initial characterization of a cell's summation treats all stimuli equivalently, i.e., the model's prediction for the mixture response, $\hat{R}_{mix}$, is dependent only on the individual component responses, r_p (*eq.* 3.4). However, we can next ask whether summation is dependent on what is being summed, i.e., the stimulus properties. To assess any spatial frequency dependence specifically, we first calculated the difference between the measured mixture response, r_{mix}, and the model's predicted response, $\hat{R}_{mix}$, for each mixture condition. Then, to account for the wide range in response magnitudes, we normalize this deviation δ_N by the sum of the predicted and measured responses (*eq.* 3.7). This creates an index bounded from (-1,1), though in practice the range is less due to an additional noise-dependent constant in the normalization (see Methods, §3.2.1). Two conditions with example suppression indices are

highlighted in *fig.* 3.3. For both example LGN neurons, summation is relatively more linear ($\delta \geq 0$) for spatial frequencies near the cell's peak compared to lower spatial frequencies.

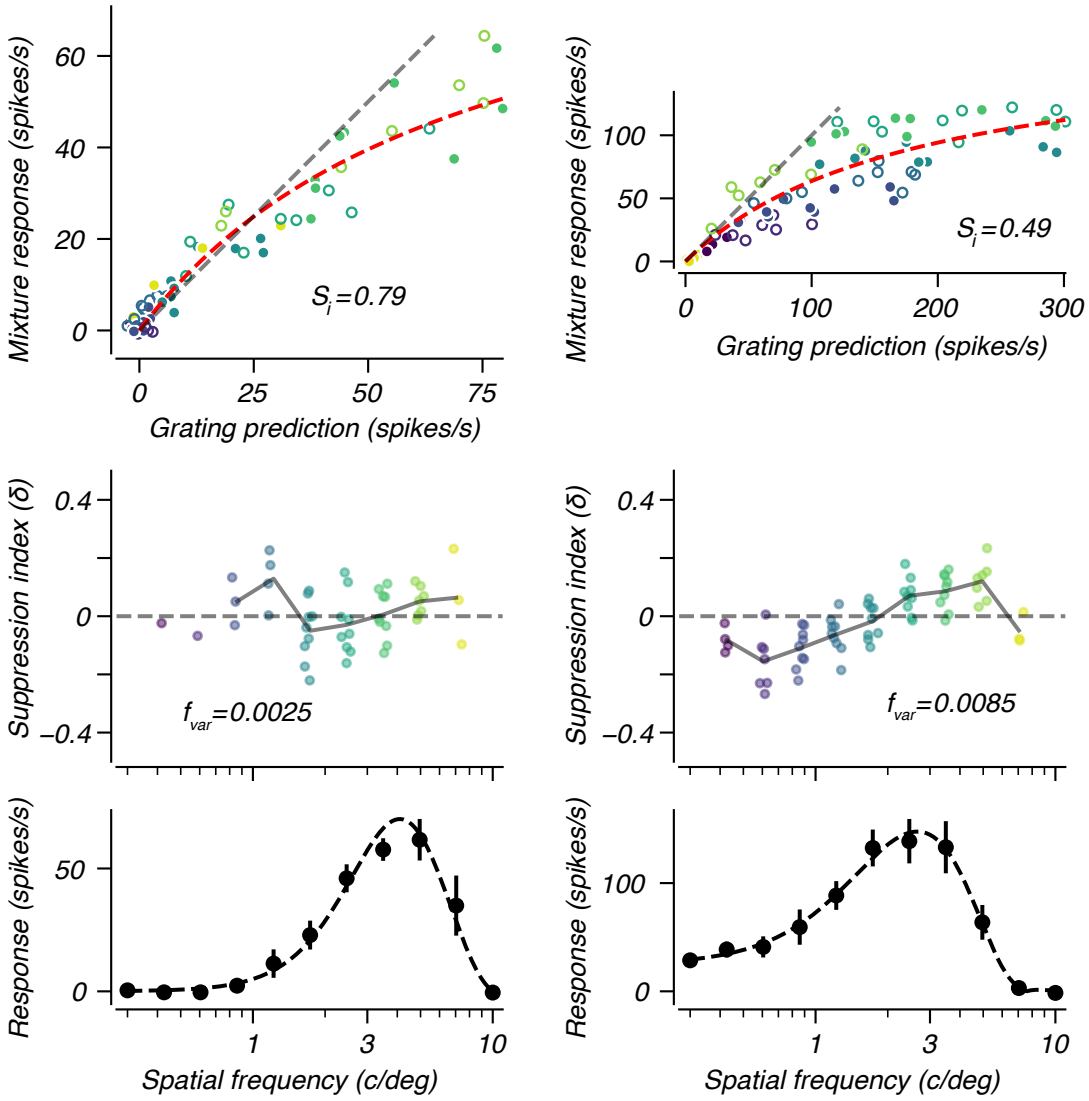

Figure 3.5: As in *fig.* 3.3, but for two V1 complex cells. Both neurons are bandpass, but the neuron on the left is more sharply tuned, and thus has fewer mixture conditions driving strong responses. The variance in suppression index, δ, across spatial frequency is given by f_{var}. The neuron on the right exhibits a stronger, more lawful influence of center spatial frequency on relative summation, with less suppression (δ) at frequencies at and above the peak compared to those on the low-frequency limb of the cell's tuning curve. For reference, we also include the cell's overall summation index, S_i.

However, compare the summation in the two LGN examples to an example V1 complex cell which exhibits strong, systematic deviations in suppression with spatial frequency. The

frequency dependence is visually striking in the summation plot (*fig.* 3.5, *top right*), where the vertical separation between mixtures with similar grating predictions is well accounted for by center spatial frequency. That is, for a set of mixtures with similar grating predictions, this cell reliably shows more linear summation for higher spatial frequencies. This is also borne out in the distribution of suppression indices (*fig.* 3.5, *middle right*), which shows less within-frequency and more across-frequency variation compared to the LGN examples in *fig.* 3.3 and to an additional example complex cell (*fig.* 3.5, *left*). We also present two V1 simple cell responses (*fig.* 3.6) - one shows low-pass tuning for spatial frequency, linear mixture summation, and no strong frequency dependent suppression; the bandpass cell again shows strong suppression at low frequencies compared to those at and above the cell's peak frequency. The presence of relatively stronger suppression for low spatial frequencies was common among cells which showed spatial frequency dependent suppression. For cells with weaker suppression, we often observed supralinear summation at low response amplitudes. This likely reflects a high response threshold, which is barely exceeded for weak stimuli and beyond which the response grows more linearly. However, as shown by the distribution of summation indices, S_i, systematic facilitation (supralinearity, with $S_i > 1$) across the full stimulus set was very rare (*fig.* 3.4).

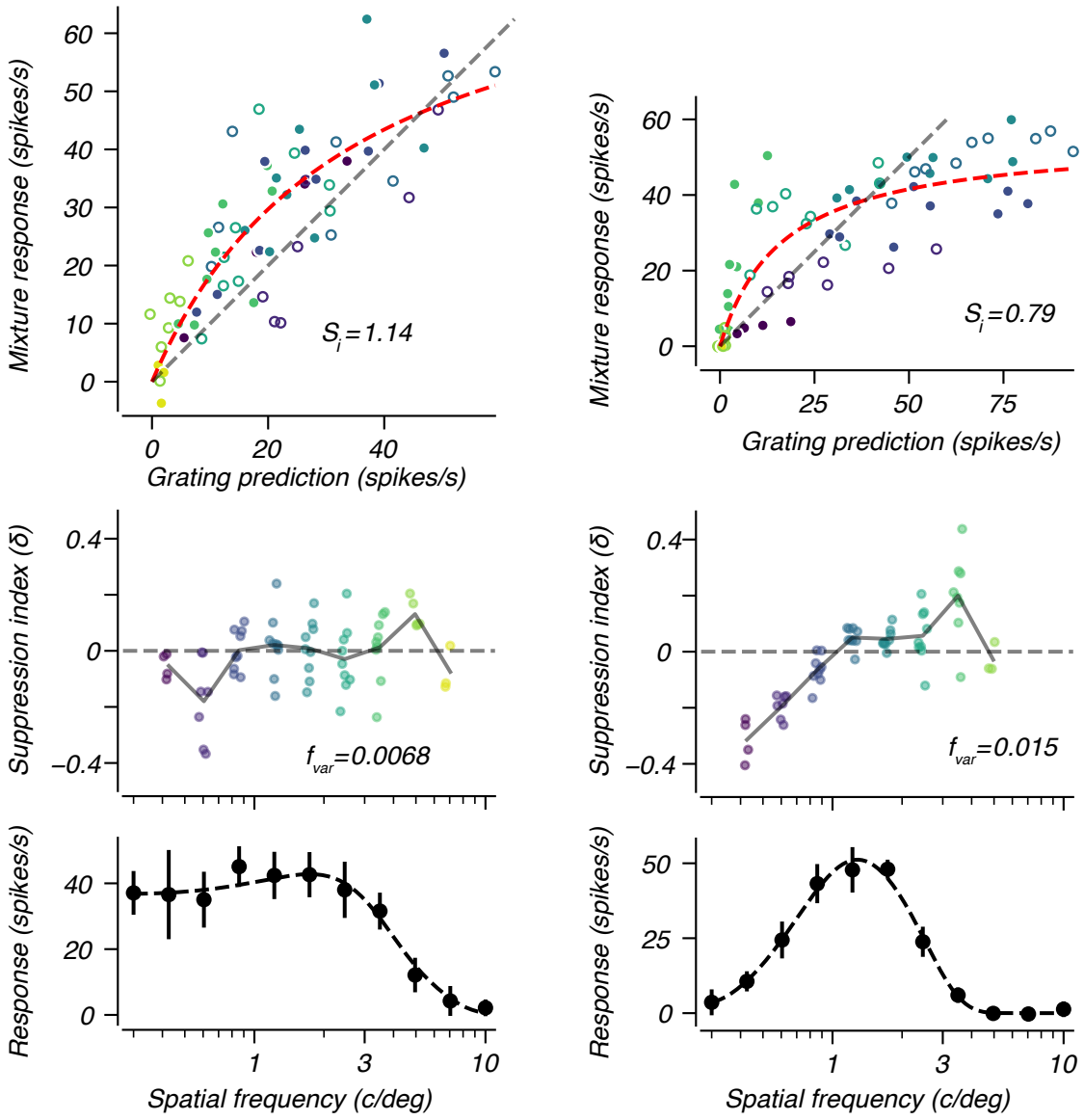

Figure 3.6: As in *fig.* 3.3, but for two V1 simple cells, and with descriptive fits from a d-DoG-S model (dashed line, *bottom*). Though both neurons have relatively linear summation (S_i), the bandpass neuron (*right*) shows stronger saturation and frequency-dependent suppression (f_{var}). The supralinear summation shown here (i.e. responses above the diagonal) was not uncommon for low amplitude responses, and potentially reflects a strong response threshold.

Frequency-dependence of summation

We can summarize the spatial frequency dependence of suppression for each cell by averaging δ_i for each spatial frequency and taking the weighted variance across those values (*eq.* 3.8). We include this weighted variance, f_{var}, in the plots for each of our example cells, and found

the summary results which follow to be equivalent whether we used the original value or those taken from bootstrap resampled estimates. We find that across the population, V1 cells show greater frequency-dependent variation in mixture suppression compared to those in the LGN (*fig.* 3.7*A*), with an population median ($f_{var} = 0.0071$) more that twice that of the LGN ($f_{var} = 0.0035$; t-test comparison yields $p = 0.013$). However, the discrepancy between LGN and V1 is largely driven by complex cells, which show stronger frequency-dependence (median $f_{var} = 0.0096$) than simple cells ($f_{var} = 0.0050$).

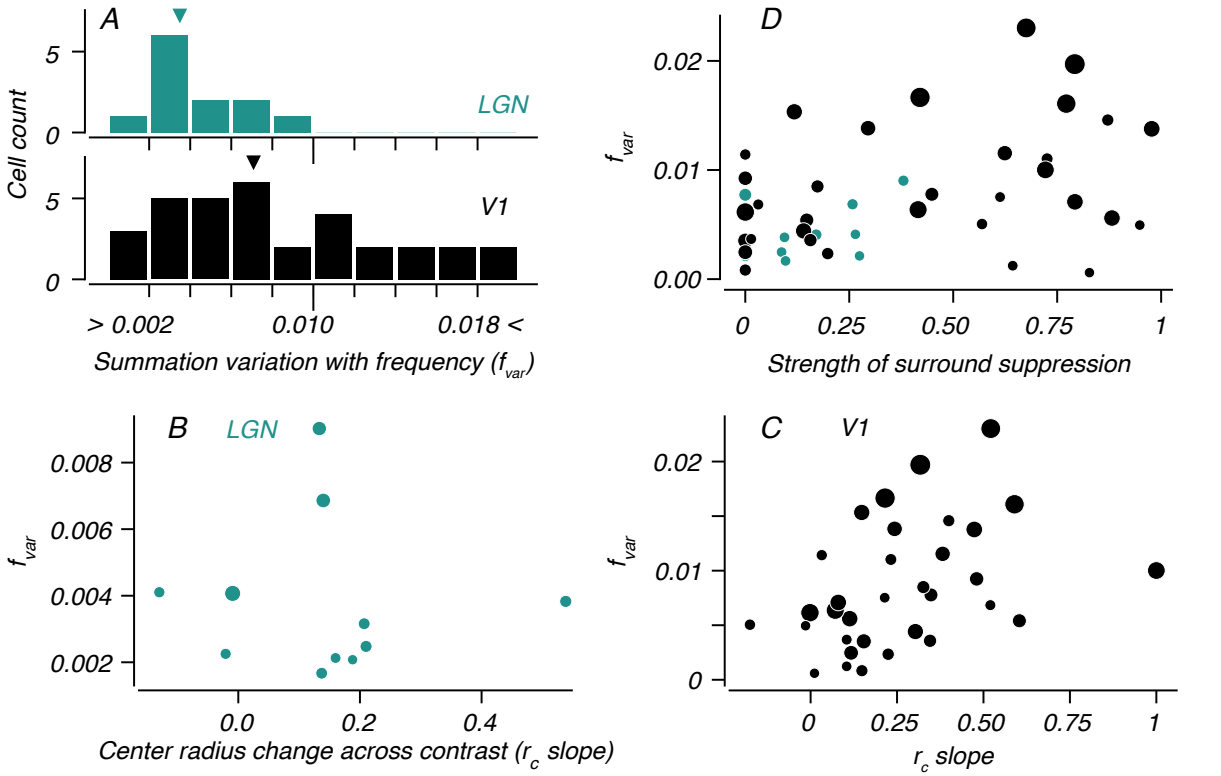

Figure 3.7: Summary of the observed frequency-dependence of mixture summation for LGN (*green*) and V1 (*black*) neurons. (A) Distribution of how strongly summation varies with spatial frequency (f_{var}). The variation is stronger in V1 (population median of 0.0071) than in the LGN (0.0035). For all scatter plots, the size of each datum is inversely proportional to the standard deviation of bootstrap f_{var} estimates. (B,C) Correlation between f_{var} and a previous measure of contrast-dependent spatial frequency tuning, r_c slope. No relationship is observed within LGN neurons (n=11; $r = 0.01, p = 0.9$), but V1 neurons show a relationship between these two measures of spatial frequency-contrast interactions (n=33; $r = 0.41, p = 0.02$). (D) Across all cells, we found a correlation between the strength of surround (size) suppression and the magnitude of frequency-dependent mixture summation (n=44; $r = 0.43, p = 0.003$).

Beyond cataloging the difference in frequency dependent suppression between LGN and V1 neurons, we also sought to understand what properties might account for these differences. We first asked whether this measure of frequency-dependence was related to contrast-dependent shifts in single grating spatial frequency tuning (as explored in the previous chapter, ch.2). This measure (previously called the characteristic frequency slope, referred to as r_c slope here) reflects the change in center-size mechanism as a function of contrast. In the difference-of-Gaussians (LGN) and paired DoG (V1) models we fit to our data, the size of the receptive field's center mechanism determines the high-spatial frequency cut-off of the tuning curve. We found that V1 neurons with larger r_c slope values - and therefore greater contrast-dependent tuning shifts - also showed stronger frequency-dependent summation, f_{var}, (*fig.* 3.7*B*) in our mixture experiment ($r = 0.41, p = 0.02$; Pearson's r). There was no such relationship for LGN neurons (*fig.* 3.7*C*), though the sample size was smaller (11 LGN neurons vs. 33 in V1).

Due to the different summation properties of LGN and V1 neurons, we compared underlying response properties between the two classes. From a set of properties including maximum response amplitude, contrast saturation, as well as tuning for orientation, spatial frequency, and stimulus size, we found that only preferred spatial frequency and the strength of surround suppression differed between the included LGN and V1 populations. Neither of these differences are surprising, given the known tuning properties in the two areas. As expected, the population of V1 neurons was tuned to higher spatial frequencies than those in our LGN dataset. (Note that while spatial frequency tuning is of course broader in the LGN than in V1, half-height octave bandwidth is often undefined for LGN neurons, and thus we omit that metric.) We also found that V1 neurons showed stronger surround suppression (mean size suppression index of 0.42 vs. 0.14 for LGN neurons; $p = 0.007$, t-test), in agreement with previous studies [162].

Based on our observation of stronger f_{var} in V1 and the differences in spatial frequency tuning and surround suppression between our LGN and V1 samples, we asked whether there was any parametric relationship between these tuning measures and frequency-dependent

summation (f_{var}). We found that more surround suppressed neurons exhibited a stronger frequency dependence to summation (*fig.* 3.7*D*; $r = 0.43, p = 0.003$). However, the relationship was largely driven by the differences in LGN and V1, with a weaker intra-V1 correlation ($r = 0.33, p = 0.06$). We also found a strong, positive correlation (plot not shown) between preferred spatial frequency and f_{var}, whether evaluated across all cells, or within either the LGN ($r = 0.65, p = 0.029$) or V1 ($r = 0.52, p = 0.002$). Due to the range in tuning bandwidth in cortex, we also queried the relationship between tuning bandwidth and f_{var} and found no correlation ($r = 0.08, p = 0.72$).

3.3.2 Measuring the frequency and contrast tuning of suppression using masks

The superposition experiment offered a survey of summation's dependence on the spatial frequency and contrast content of an image. While that experiment provided evidence for spatial-frequency dependent suppression, the construction of the mixture stimuli required broad spectral content, i.e., a range of spatial frequencies. In the prior analyses, we grouped mixture stimuli by their central spatial frequency. Precisely because of the broad spectral content of the mixtures, however, it is difficult to pinpoint the exact tuning of spatial frequency suppression. To more precisely target the tuning of suppression, we performed a masking experiment in which only a maximum of two sinusoidal gratings appeared simultaneously. Due to our finding that frequency dependence was stronger in V1 than in the LGN, we recorded only in V1, collecting data from 47 cells. Of the 47 cells, 34 had the experiment tailored to their tuning; the remaining 13 neurons were units isolated during the sorting process.

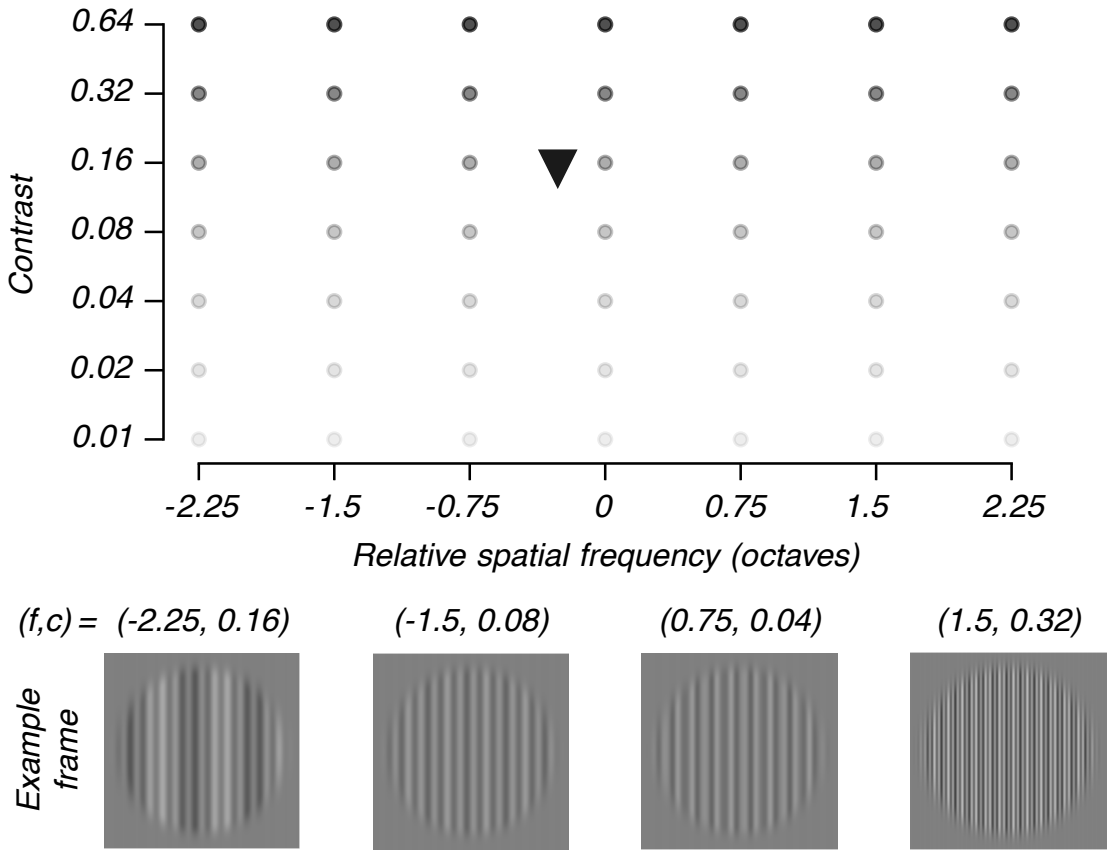

Figure 3.8: (*Top:*) The 7x7 grid of mask contrast and spatial frequency for the masking experiment. We note an example base stimulus with the inverted triangle at a contrast of 0.16 and with a spatial frequency just below the center spatial frequency of the mask sampling. Stimuli consisted of either a mask alone, the base stimulus alone, or a paired presentation. For each cell, the base stimulus was chosen to be the cell's tuning peak, while we chose the center of the mask frequency sampling (relative spatial frequency = 0, in this plot) to be 1 or 2 c/deg, depending on the cell's tuning. (*Bottom:*) We show a frame from four joint mask + base conditions, as indicated (f, c) by their relative mask spatial frequency and mask contrast.

The stimulus set described here is modeled after a previous masking experiment performed in cat [13]. As in the mixture summation experiment, the core element of the stimulus set is a 2D matrix of sinusoidal gratings that vary in spatial frequency and contrast (*fig.* 3.8). For this experiment, we sampled 7 spatial frequencies centered near the cell's preference with 0.75 octave steps and 7 contrasts ranging from 0.01 to 0.64. Presenting these gratings allowed allowed us to determine the cell's spatial frequency tuning, and any influence of contrast on that tuning. These same gratings also served as a masking stimulus, since we presented them in combination with a base grating of fixed spatial frequency and contrast.

The base grating was set at the cell's preferred spatial frequency (except in the case of the cells acquired through sorting) and with a contrast in the accelerating portion of the response-versus-contrast function, i.e., above threshold but below the saturation point. Because we kept our set of mask contrasts fixed with a maximum of 0.64, the maximum permitted base contrast was 0.36.

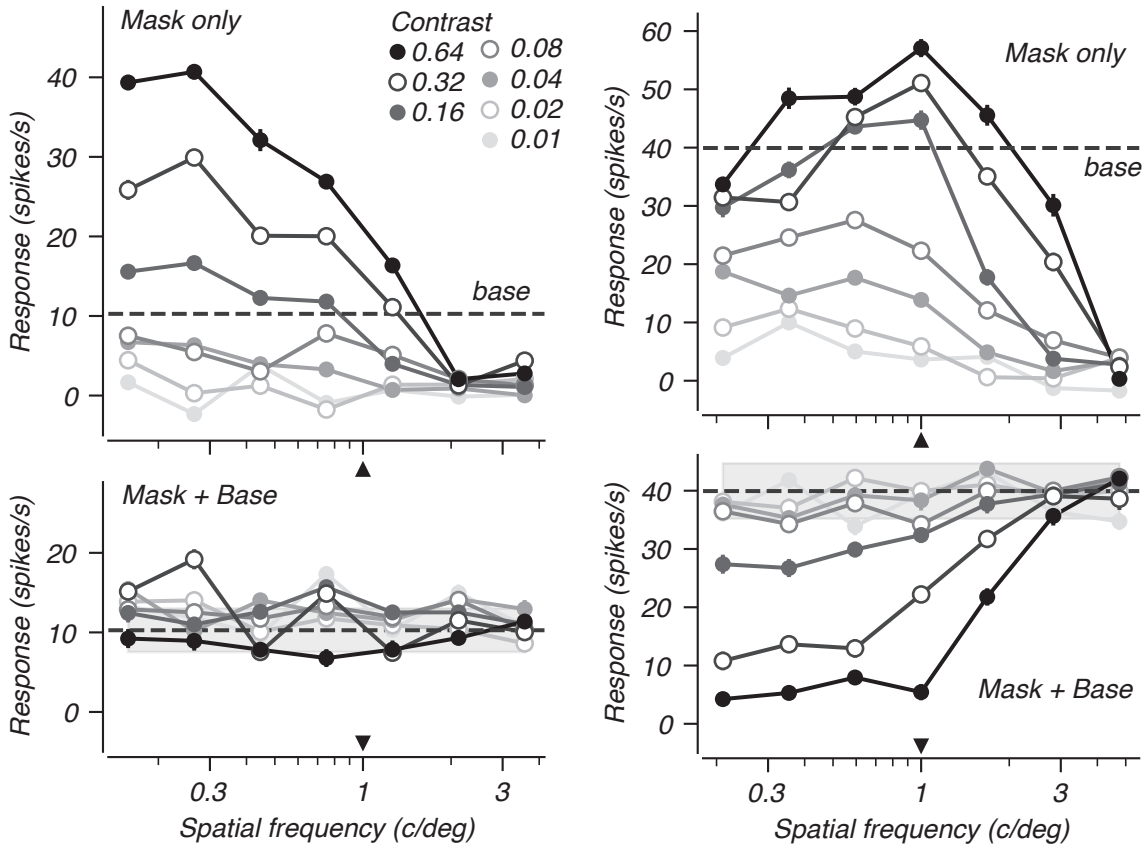

Figure 3.9: Mask only (*top*) and base response in the presence of the mask (*bottom*) for two V1 simple cells (*left, right*). The color of each line indicates the mask contrast, ranging from 0.01 to 0.64; for enhanced legibility, we alternate hollow and filled points for each contrast. In the bottom panel, we indicate the mean base response *dashed line* and show 1 std. (± 0.5) (*gray band*). The neuron on the left shows minimal, if any, suppression. The more bandpass cell (*right*) shows suppression that grows with contrast and is strongest for spatial frequencies at or below the cell's preference. The base stimuli were 1 c/deg for each, though at contrasts of 0.23 and 0.16, respectively, with frequency indicated by the triangles along the horizontal axis. The stimulus sizes were 1.5° and 2°, respectively, due to small receptive fields (1° and 0.6°) with at least modest surround suppression.

We analyzed the response of base and mask stimuli alone, as well as the response to their joint presentation. The responses of two example V1 simple cells are shown in *fig.* 3.9. The

mask-only responses in the top panel show the spatial frequency tuning across a range of contrasts, providing a view into the variety of observed tuning behaviors - one is low-pass tuned with low contrast sensitivity (*left*) while the other is bandpass with a higher sensitivity for contrast (*right*) and sharper tuning at high contrast. These example cells are simple, and thus we analyze the F1 response; furthermore, the base and mask stimuli were always presented at differing temporal frequencies, allowing us to disentangle the response attributed to each stimulus component. The response to the base in the presence of the mask is shown in the bottom panels of *fig.* 3.9. A lack of suppression to the masks manifests as base responses which are unchanged from the isolated base stimulus response. However, when suppression is present, we can analyze the properties of the suppressive masks to characterize its tuning. The bandpass example cell on the right of *fig.* 3.9 shows exactly this: for low mask contrast, the base response is relatively unchanged; however, as the mask contrast increases, there is marked suppression below the base firing rate. In this case, suppression first appears at a contrast of 0.16, and is strong only at the lowest mask spatial frequencies. However, as contrast grows, the strength and tuning of suppression grows, encompassing all masks up nearly 2 c/deg at the highest mask contrast of 0.64. We highlight two additional example cells in *fig.* 3.10. For all four example simple cells, the receptive fields were small (less than 1°) and surround suppressed, which required us to keep our stimulus sizes small to maintain sufficient response amplitudes. However, because of the small stimulus sizes (between 1.5-2° for these examples), the lower spatial frequencies were poorly resolved; that is, the spectral content for low spatial frequencies was blurred, limiting our ability to determine the tuning of suppression at the lowest spatial frequencies.

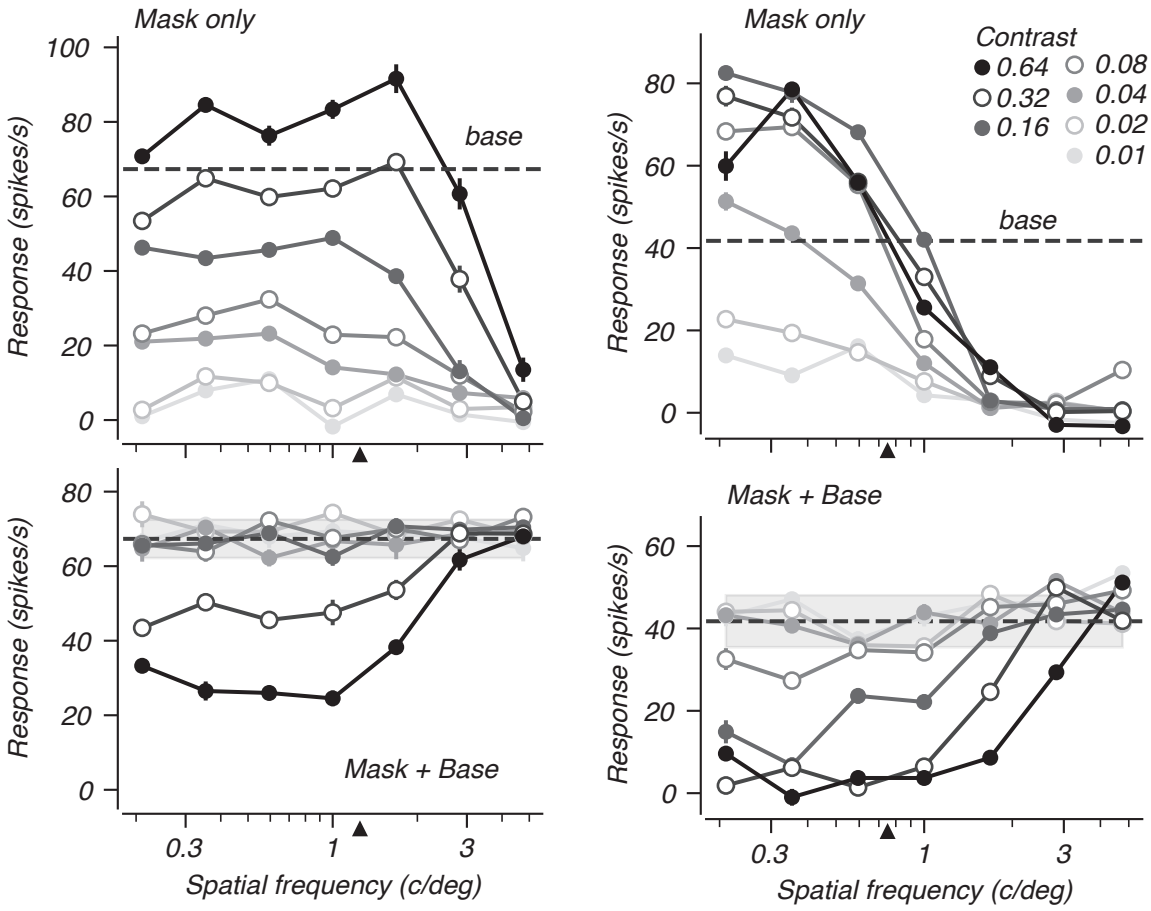

Figure 3.10: As in *fig.* 3.9 for two additional simple cells. Here, both cells show suppression which first emerges at intermediate contrast and for low spatial frequencies. As contrast increases, the suppression becomes broader through the emerging influence of masks at higher spatial frequency. We used stimulus sizes of 1.5° and 1.6°, respectively, due to small receptive fields (0.9° and 0.45°) with at least modest surround suppression.

Analyzing complex cell responses to the masking stimuli is less straightforward, since we cannot separate the component of the response attributed to each grating. For that reason, the observations here provide only a lower bound on the underlying extent and strength of suppression. To better understand this, consider the two complex cells in *fig.* 3.11. Both neurons here are bandpass tuned, as evidenced by the responses to the mask stimulus alone. Considering the response to the mask and base stimuli presented together, however, we see that for the majority of mask conditions, the responses are similar to those of the mask alone. The influence of masks that are near the preferred spatial frequency and presented at

high enough contrast is always facilitation, with joint response amplitudes greater than the base-only responses. For masks of the highest contrast, the joint response is often equivalent to the mask alone, due to response saturation (e.g., note the responses at the preferred spatial frequency and highest mask contrast for both cells). However, for intermediate mask contrasts (e.g., 0.08 and 0.16), the joint responses are greater than to either stimulus alone, as would be expected for summation which has not yet saturated entirely. Finally, for masks of very low contrast, particularly for cells with relatively low contrast sensitivity like these two examples, the joint response is often unchanged from the base response alone; that is, the mask has no marked influence on the response.

However, for masks presented outside the cell's passband we often observed suppression which grows stronger with contrast. This can be seen in one example cell (*fig.* 3.11, *right*), where there is strong suppression beneath the base response at the highest mask contrasts, but only for spatial frequencies at or below the cell's low spatial frequency cutoff. This was a typical, albeit stronger, example among complex cells which showed measurable suppression. One additional example cell - with both DC and F1 responses - is shown in *fig.* 3.16.

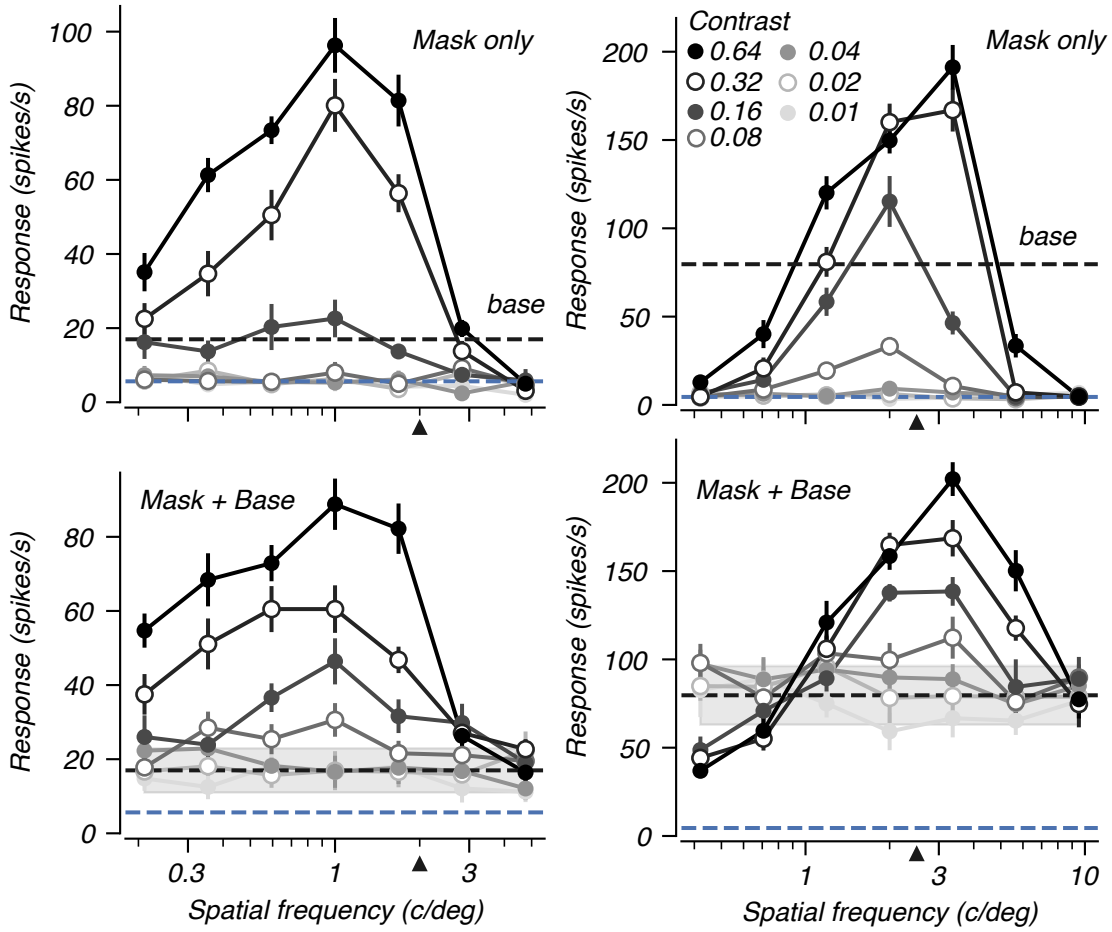

Figure 3.11: As in *fig.* 3.9 for two V1 complex cells. For DC responses, we cannot disentangle the response due to each stimulus component, and thus the mask + base response shows facilitation for most masks within the cell's passband (*bottom*). However, for one cell here (*right*), there is marked suppression at low spatial frequencies. The base stimuli have a contrast and spatial frequency of (0.2, 2 c/deg) and (0.14, 2.5 c/deg), respectively, with frequency indicated by the inverted triangles along the horizontal axis. The dashed blue line at the bottom of all four panels represents the baseline (i.e., mean luminance field) firing rate. The receptive field and stimulus sizes were (0.85°, 0.9°) and (1.3°, 2.5°), respectively, for the two cells.

To summarize the spatial frequency tuning of suppression, we first computed the difference between the base alone and all joint mask-base responses. We used the standard deviation of the base responses to normalize these difference values (*eq.* 3.9). The normalization serves two purposes: one is to take into account the overall amplitude of the differences; the second is to penalize cells with more variable responses relative to their firing rate. This suppression measure, $S(m)$, effectively z-scores the mask-induced changes in response. The distribution

across cells for the strongest observed suppression across all mask conditions was heavy-tailed, with most cells showing suppression weaker than 1 or 1.5 times the base standard deviation (*fig.* 3.12*A*). For the purposes of classifying cells as either showing suppression or not, we applied the following criteria: at least two consecutive masks (be it for adjacent mask frequencies at the same contrast, or the same frequency at two adjacent mask contrasts) should show suppression above a threshold. We set this threshold to be 0.5, such that a mask was counted as suppressing a cell only if the joint mask-base response deviated from the base response alone by more than 1 standard deviation (± 0.5). This procedure excluded more than half of all cells (26 out of 47) as showing no suppression, though disproportionately complex cells (13 out of 20 excluded, compared with only 13 out of 27 simple cells). Furthermore, cells which were not the target of the experiment and were instead identified as distinct, responsive units during spike sorting, were also excluded at a higher rate (8 out of 13) than the remaining cells (18 out of 34). For further details on the properties of excluded cells, see the Methods paragraph on inclusion criteria and the classification procedure (§3.2.2).

Considering only cells which showed suppression, we then asked which mask spatial frequencies most commonly suppressed the cell's firing. Because each cell's tuning was different, we grouped mask frequencies by their spatial frequency relative to each cell's preference. Additionally, due to the differing ways in which suppression is measured for DC and F1 responses, we split this analysis by simple-complex designation. We found that for simple cells, suppression was most common at mask frequencies below or near the peak (*fig.* 3.12*B*), with a gradual reduction in prevalence for higher mask frequencies. Nearly all simple cells showed suppression for masks just below and above the cell's peak. Suppression was less commonly observed overall for complex cells, with mask frequencies ± 2.5 octaves from the cell's peak most likely to suppress responses. The complex cell findings are due, in part, to the inability of directly measuring suppression unless it is so strong as to reduce the response below the level of the base stimulus alone.

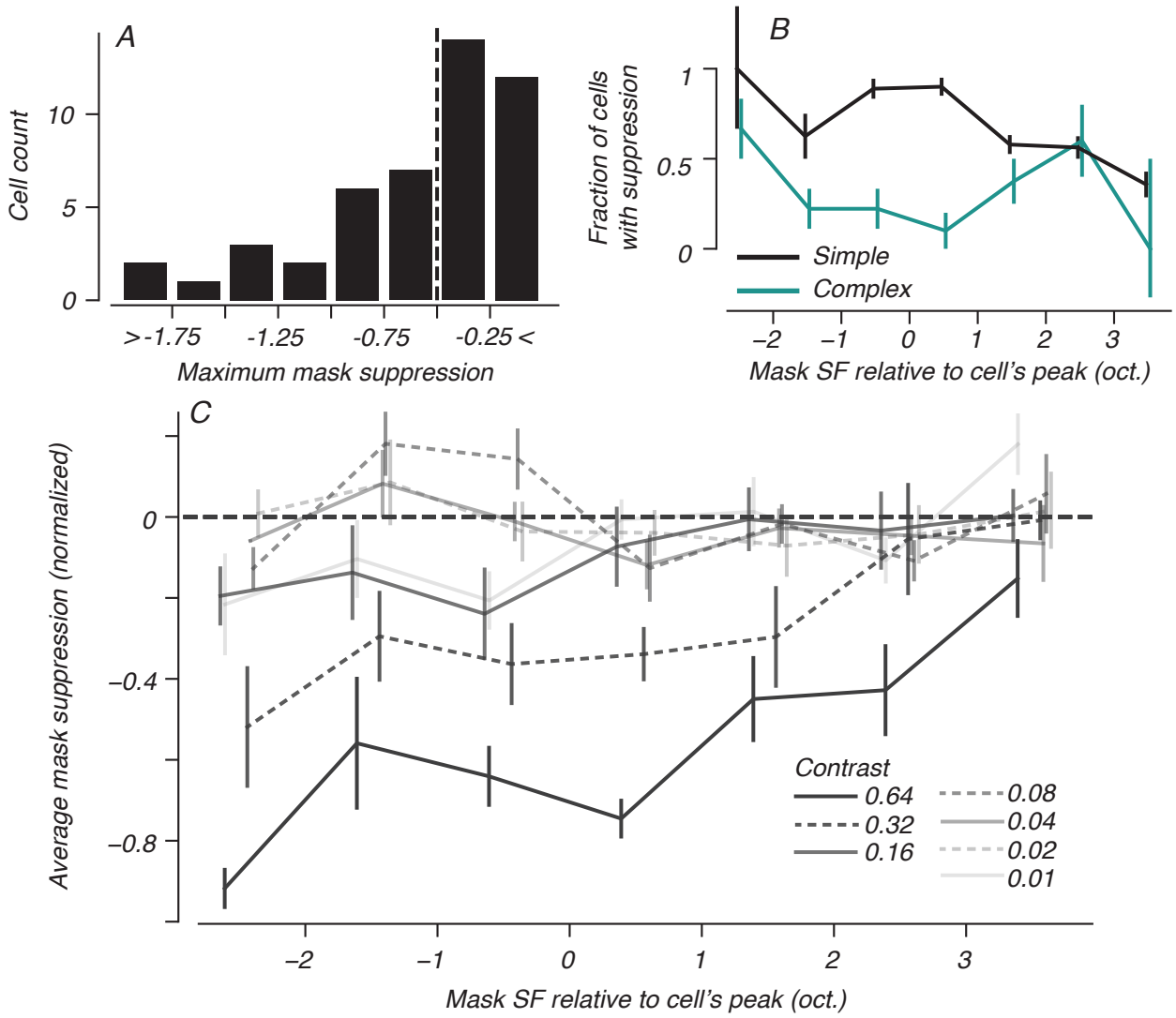

Figure 3.12: Summary of suppression in the masking experiment. (A) Distribution of the maximum observed suppression for any mask across 47 V1 cells. Among cells which show suppression, the fraction which show suppression at a given spatial frequency relative to the cell's peak (B), split for simple (n=14) and complex (n=7) cells. Error bars are inversely proportional to the number of cells with mask spatial frequencies in that range. Suppression for lower spatial frequencies was most common, while the second peak at high spatial frequency for complex cells reflects suppression which is made visible outside the cell's passband. (C) Average suppression by relative mask spatial frequency and mask contrast for simple cells (n=14); error bars are standard error of the mean. At a population level, only masks of 0.32 and 0.64 contrast show prominent suppression. As hinted at in B, suppression is strongest at or below the cells peak. We added random jitter to the plotted spatial frequency locations in the two errorbar plots to minimize data overlap.

Finally, because suppression was more widely observed for simple cells, we then asked how the strength of suppression varied as a function of the mask stimulus. To measure the average

profile of suppression, we binned mask conditions by contrast and relative spatial frequency, computing the average normalized suppression, $S(m)$, within each bin. In agreement with suppression being most commonly observed at lower spatial frequencies, we found that suppression was also strongest at or below the cell's peak (*fig.* 3.12*C*). Furthermore, we found no systematic facilitation by the masking stimuli. The profile of averaged suppression also highlights the influence of mask contrast on suppression strength: low contrast masks exert little to no effect on base responses, but masks of contrast 0.32 or higher typically drive suppression. As observed in the example cells above, the exact minimum mask contrast at which suppression is first apparent, of course, varies across cell and spatial frequency. The population averaged suppression also complements the observation we draw from the example simple cells: suppression first appears at lower spatial frequencies, but suppression at intermediate frequencies becomes apparent or relatively stronger with higher mask contrast.

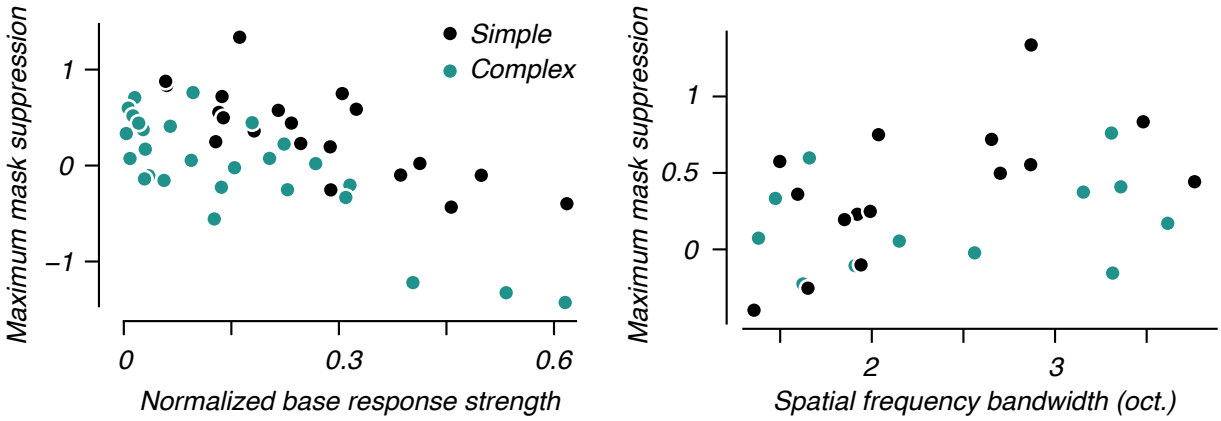

Figure 3.13: Relationship between the logarithm of the strongest observed suppression per cell (negative values indicate stronger suppression) and the strength of the base response (*left*) or spatial frequency bandwidth at half-height (*right*). The base response strength is given by the base-alone response normalized by the maximum observed mask-alone response. Note that more negative values indicate stronger suppression, and we use the log of the maximum suppression measure to better handle the long-tailed distribution of values. Cells whose base responses were relatively stronger showed more suppression ($r = -0.60, p <$ 1e-5). For the 27 cells with defined half-height spatial frequency bandwidth at high contrast, more broadly tuned cells showed weaker suppression ($r = 0.39, p = 0.045$).

We next sought to consider properties of the cell and/or the experimental design which might have influenced the strength of suppression. The choice of base stimulus was crucial

for this experiment, since the base set the cell's gain state and all suppression was measured relative to the base response. We evaluated the strength of the base stimulus by normalizing its mean firing rate by the maximum mask-only response (i.e., the highest single grating response). If the base stimulus drove too weak a response, the cell might be in a more high gain state, and therefore suppression might be less visible or only apparent for the strongest mask conditions, if at all. We evaluated this relationship across all cells - even those excluded from the previous suppression summary - and found a strong correlation (*fig.* 3.13, *left*) between the strength of the base response and observed suppression ($r = -0.60, p <$ 1e-5). Whether due to more variable responses or poor base stimulus choice, this relationship suggests that at least some cells might have either shown more extensive and/or stronger suppression with a different base stimulus. This relationship also explains the lower prevalence of suppression in cells which were identified only post-experiment through spike sorting, since the base stimulus was not chosen specifically for their tuning and therefore drove weaker responses.

In the previous stimulus set, we uncovered a relationship between frequency-dependent mixture summation and both surround suppression and contrast-dependent spatial frequency tuning. We sought to understand whether these same relationships held for the suppression observed with the masking experiment. Interestingly, neither surround suppression ($r = -0.10, p = 0.50$) nor the slope of contrast-induced tuning changes ($r = 0.05, p > 0.75$) were predictive of the measured suppression strength. However, we did note a modest correlation with half-height octave bandwidth (*fig.* 3.13, *right*), with more narrowly tuned cells showing relatively stronger suppression ($r = 0.39, p = 0.045$). We also found that cell's with higher preferred temporal frequency showed stronger masking suppression (not shown; $r = -0.42, p < 0.005$).

3.4 Discussion

Using two distinct stimulus sets, we characterized the spatial frequency dependence of suppression. In the first experiment, we measured suppression in LGN and V1 neurons by calculating the discrepancy between the observed and expected responses of compound stimuli which varied in the center frequency, spectral spread, and total contrast. We first showed that mixture summation behaved comparably to contrast summation, i.e., response saturation as measured with response-versus-contrast curves. Then, we quantified the strength of frequency-dependent deviations in summation and found that cortical cells - complex ones, particularly - showed a stronger dependence on frequency. Typically, these cells summed more linearly for frequencies at or above the cell's preference and showed stronger sub-linearities for low frequencies. Having observed strong frequency dependent summation in V1, we then used a masking experiment to further characterize cortical suppression. The experiment's structure allowed us to directly measure suppression, which we found to be more readily observed in simple cells than in complex cells. For simple cells, we measured a broadly tuned suppression which was strongest for frequencies at and below the cell's preference. However, for complex cells, suppression was most apparent at the edges of the cell's passband. The results from the two experiments provide complementary, though not always concordant, insights into spatial frequency tuning.

3.4.1 Relating the two experiments and potential mechanisms

With the mixture stimuli, we found that summation was more sublinear for low frequency mixtures. Given the relationship between summation and suppression, this finding is in alignment with the result of strong suppression for low spatial frequency masks in the second experiment. However, though the two experiments offer the same core finding, there are other important differences between the results. For one, mixture summation showed stronger frequency-dependence for complex cells than simple cells, yet complex cells were less likely to

Relationships between mixture/masking metrics and other tuning properties						
Stimulus	Metric	prefSf	sfBW	tfPref	Size suppr.	r_c slope
Mixtures	SF-dep. summation, f_{var}	+	None	None	+	+
Masks	Suppression strength	None	–	+	None	None

Table 3.1: Positive (+), negative (-), and not significantly correlated (None) relationships between tuning properties and two independent measures of suppression/summation: the spatial frequency-dependent variation in summation (f_{var}) and the strength of masking suppression. None of the correlated metrics were in common between the two experiments.

show suppression in the masking experiment. We believe this is due to the nature of how we measure suppression, and not a reflection of cell's properties, but see the discussion below for elaboration (§3.4.3). There were also marked differences in the relationships between other tuning properties and the results of these two experiments. The strength of frequency-dependent mixture summation was found to be positively correlated with preferred spatial frequency, the strength of size suppression, and the magnitude of contrast-dependent tuning shifts we observed in the previous chapter (*fig.* 3.7). No such correlations were observed with respect to the strength of masking suppression. An additional discrepancy is noted for bandwidth, where we noted stronger suppression for narrowly tuned cells (*fig.* 3.13, *right*); no relationship was found between bandwidth and frequency-dependent summation. These relationships are summarized in *table* 3.1.

Several of our findings are in agreement with previously reported studies. For one, the tight correlation between tuning bandwidth and the strength of masking suppression is in agreement with prior work which argues that suppression in V1 is responsible for sculpting broad LGN inputs into bandpass selectivity [13, 21, 115, 100, 139]. Given the positive correlation we observe between preferred temporal frequency and the strength of masking suppression, we also draw a connection between this finding and the proposed mechanism for suppression which is untuned for orientation. Namely, the work of Henry et al. (2020) [59] argues that

this form of untuned suppression is derived from magnocellular inputs, which are tuned to higher temporal frequency [70, 40].

As for the mixture experiment, we draw a connection between the influence of size tuning on summation and orientation-selective suppression [162, 27]. First, we argue that the qualitative difference in spatial frequency-dependent summation between LGN and V1 aligns with a previous adaptation study in cat [42], in which only V1 neurons showed frequency-dependent adaptation. This is the first piece of evidence which suggests the frequency-dependent summation is cortical in origin. Additionally, previous findings show that suppression is found to be stronger for more narrowly tuned cells (whether for orientation or spatial frequency [115]), and cells more narrowly tuned in one domain are typically sharper in the other, too [163, 167] (though see [91, 36]). Finally, we note that strong size suppression is commonly found in layers 2/3 and 4b [125, 140] where Henry et al. (2013) [58] also found the suppression to be most strongly tuned for orientation. Thus, taken together, we suggest that relative to the masking experiment, our frequency-dependent summation either recruits cortical-origin (and therefore more sharply tuned) suppression. Alternatively, our population for the masking experiment might have contained relatively more supragranular neurons. These alternative hypotheses could be addressed with an orientation varying set of mixtures and histological reconstruction, respectively.

The positive correlation between frequency-dependent summation (f_{var}) and contrast-dependent tuning shifts (r_c slope) in V1 bolsters the case for a cortical-origin effect in the mixture experiment. Both f_{var} and r_c slope were stronger in V1 than in LGN, and were uncorrelated with one another in LGN neurons. These observations, taken with the above commentary on cortical-origin suppression, suggest that even though neurons in both areas show contrast-dependent spatial frequency tuning, additional cortical processing contributes to the enhancement of both effects.

To summarize the above, we have argued on the basis of relationships with other tuning properties that the mixture experiment primarily recruits cortical-origin suppression and

the masking stimulus results instead reflect a thalamic-origin. Finally, we show how this hypothesis can be applied to the discrepancy between how spatial frequency preference and bandwidth relate to mixture summation and masking suppression. Namely, this hypothesis would posit that bandwidth, but not peak frequency, is primarily influenced by untuned (for orientation) suppression which is at least partially driven by thalamic inputs; on the other hand, preferred frequency in V1 is further shaped by cortical mechanisms. This claim is in agreement with physiological [113] and modeling [154, 106] results which suggest that the spatial arrangement of inputs to V1 already sharpens spatial frequency tuning and only through delayed-onset mechanisms does tuning shift towards higher spatial frequencies [21].

3.4.2 Dynamics of suppression

Our analysis ignores response dynamics. Prior work has shown that time is an important dimension for understanding tuning and suppression in cortical responses. Early onset responses are less selective, with delayed responses showing higher preferred spatial frequencies [21, 91] through suppression of low spatial frequencies [21]. Our stimuli were typically 1s in duration, and in some cases 2s, while suppression typically becomes apparent between 50-100ms after response onset [144, 168, 21]. Given that our analysis averages the onset as well as sustained responses during which suppression is more apparent, we wanted to consider different response epochs to better understand suppression.

We show two examples - one simple *fig.* 3.14 and one complex cell *fig.* 3.15 from the masking experiments - which show dynamics commensurate with prior reports. The early responses show little to no suppression which is concentrated only at the lowest spatial frequencies. Later on in the response, suppression is stronger and encompasses a wider band of spatial frequencies, in alignment with prior reports on tuned, cortical-origin suppression [58, 162]. In addition to more measurable suppression later in the response - and at least in part, as a consequence of this suppression - spatial frequency tuning becomes sharper. This is most readily observed in *fig.* 3.15, where responses to the preferred spatial frequencies

become relatively stronger at later times. A more systematic analysis of response dynamics could yield a more nuanced picture of how different forms of suppression - both tuned and untuned - contribute to spatial frequency tuning.

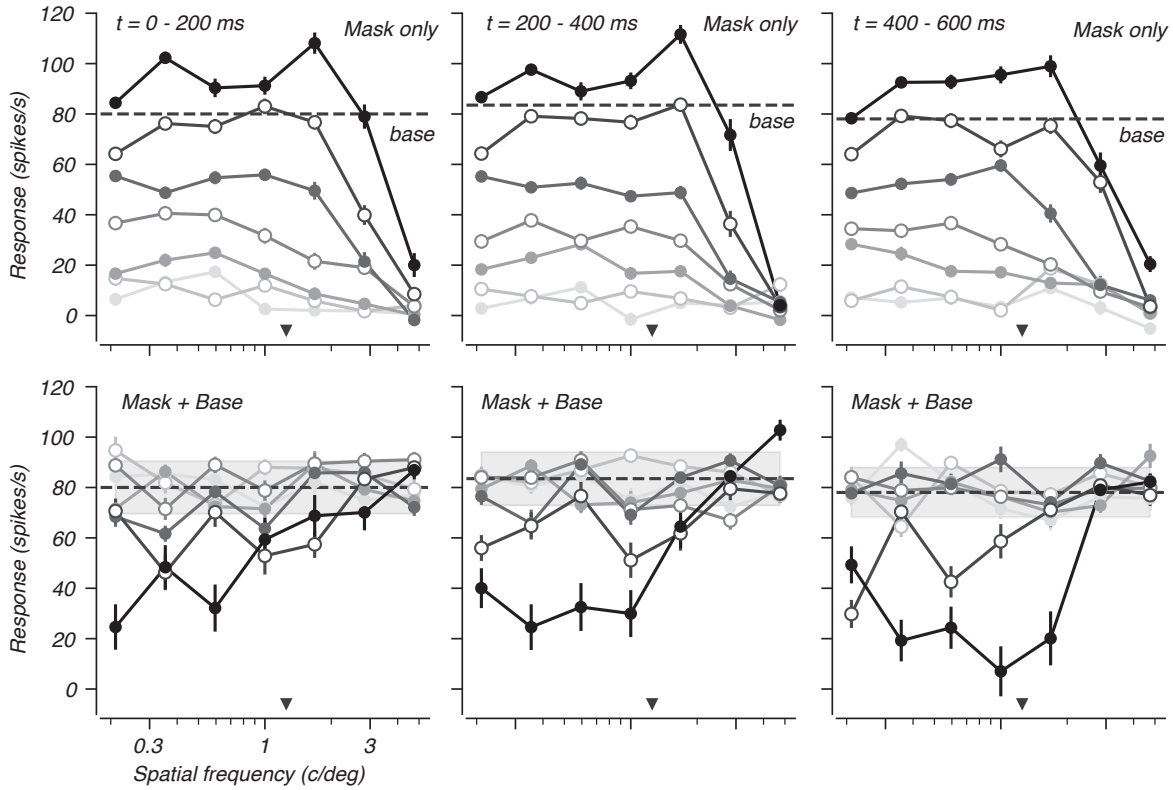

Figure 3.14: The same cell as in *fig.* 3.10 (*left*), but shown for three distinct time periods of the stimulus presentation: 0-200ms (*left*), 200-400ms (*center*), and 400-600ms (*right*). The suppression is weak and present only for very low spatial frequencies earlier on; later in the stimulus, the suppression is stronger and present even at intermediate spatial frequencies.

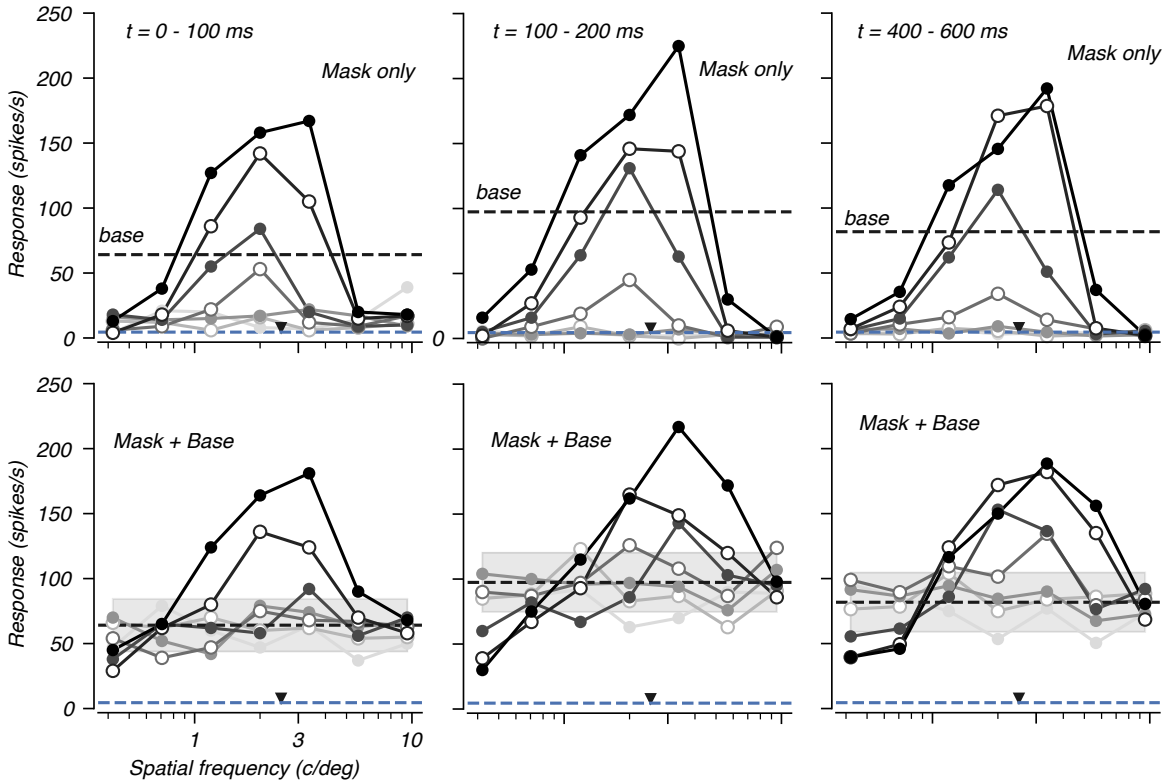

Figure 3.15: As in *fig.* 3.11 (*right*) but at three distinct time periods. Minimal suppression is visible in the first 100ms (*left*); in the 100-200ms (*center*) and 400-600ms (*right*) periods, suppression is observed at low spatial frequencies. Tuning also becomes sharper later on in the response - note the increased response near the cell's preference around 3 c/deg, particularly at contrasts of 0.32 and 0.64 (highest two contrasts).

3.4.3 Limitations of the experimental designs

A key limitation for both stimulus sets was our choice of receptive field size. The LGN data were collected with large stimuli (at least 4° patches for all but one cell), but stronger surround suppression in some V1 cells required us to use smaller stimuli. Presenting gratings with small apertures results in the blurring of low spatial frequencies - for a stimulus of 2°, any spatial frequencies below 0.5 c/deg, for example, have fewer than 1 complete cycle in the stimulus aperture and therefore have blurred spectra. This aberration hampers the resolution with which we can measure the tuning of suppression and summation for low spatial frequencies, since the spectra of gratings at 0.25 c/deg and 0.5 c/deg are less distinct than they would otherwise be with appropriately sized stimuli (e.g. at least 4°). In the

context of the masking experiment, 3 of the 4 example simple cells (*figs.* 3.9, 3.10) show largely flat suppression for low spatial frequency masks; if the stimulus sizes were larger and therefore the spectral resolution sharper, we might observe suppression getting weaker again at the lowest spatial frequencies. In the mixture experiment, the spectral blurring is less consequential because the stimuli themselves are already constructed to contain a range of spatial frequencies. Nonetheless, the ratio between our stimuli and receptive field sizes (mean and medians of 2.5 and 1.8, respectively) are on par with those used in prior experiments investigating suppression and spatial frequency tuning (e.g. [115], where the typical ratio was 1.5-3).

The work of Xing et. al. [168] suggests that if we used larger stimuli, we would find stronger suppression in certain cases. They report that untuned suppression is unaffected by stimulus size, but note stronger tuned suppression with larger stimuli [168]. In the context of our experiments, recruiting stronger tuned suppression from beyond the extra-classical receptive field should result in stronger frequency-dependent variations in summation that we measured with the mixture experiment. Per our arguments above, the masking experiment results are consistent with untuned suppression as the primary driving factor; thus, our masking results should be less sensitive to changes in stimulus size, though larger stimuli might recruit stronger tuned suppression.

Masks: Measuring suppression for simple cells was far easier than for complex cells in the masking experiment. For complex cells, the near perfect alignment with suppression at the ends of the passband suggests that suppression might be more broadly tuned, but is simply not apparent using this coarse measure of suppression beneath the firing rate. One possible approach for measuring suppression - even in the case of masks which do not drive responses below the base level - would be evaluating the deviation from linear summation, i.e., $R(m+b) - R(m) - R(B)$. For suppressed complex cells, this measure showed strong sublinear summation for a wider range of low frequencies than the primary analysis

above would suggest. However, as discussed in the summation analysis from the mixture experiment, deviation from linearity alone is not informative as to the tuning of suppression, since many cells show sublinear summation regardless of frequency. The masking stimulus set, however, does not permit a measurement of overall summation and subsequent analysis of any frequency-dependent deviations that might be observed.

Fortunately, one cell in our dataset had an F1:F0 ratio near unity and a base stimulus which drove strong DC and F1 responses (*fig.* 3.16). This cell showed the characteristic absence of DC suppression for conditions within the passband, with very strong suppression only observed for masks near the high-frequency cutoff of the cell's tuning. However, the F1 responses show a comparably strong yet far broader suppression which was visible at most low spatial frequencies. Taken alongside the observation in simple cells of consistent suppression for masks near and below the cell's peak, it is likely that complex cell suppression is more broadly tuned that we could measure here. We offer that this result unifies a previous finding of high-spatial frequency suppression found in cat [13] with the preponderance of evidence that suppression is strongest at low spatial frequencies, as observed in primate [21, 59, 162]. Specifically, we suspect that their observation of high spatial frequency suppression - which they found to be more common in complex cells and low-pass cells more generally - is due in part to difficulty in measuring suppression within a cell's passband.

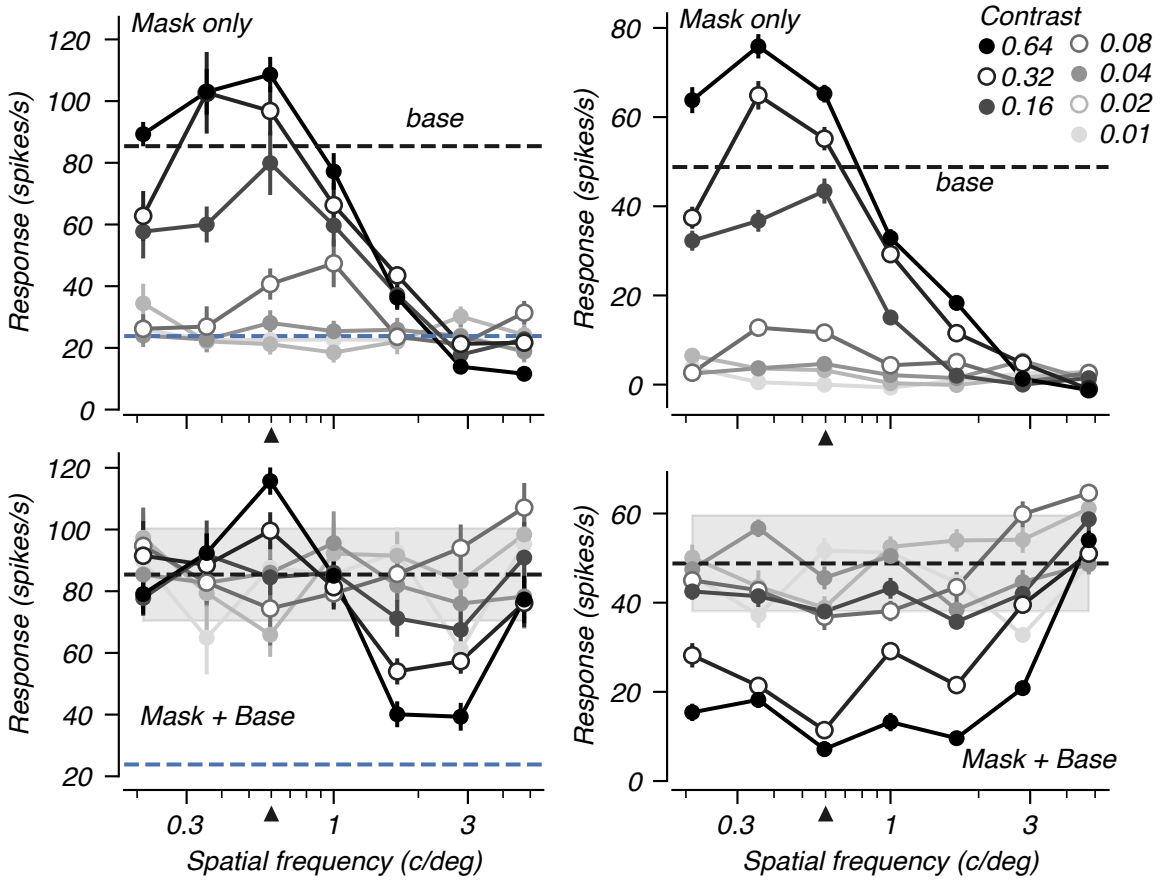

Figure 3.16: DC (*left*) and F1 (*right*) responses shown for one cell whose $\frac{F1}{F0}$ was close to 1. Though suppression is only apparent for spatial frequencies at and above the cell's cutoff based on the DC response, the F1 responses show that suppression is more broadly tuned, including for low spatial frequencies. The base stimulus had a spatial frequency of 0.6 c/deg and a contrast of 0.18. The receptive field size was 1.6° and we used a stimulus size of 2°.

Mixtures: In the mixture analysis, we are unable to disentangle frequency-dependence from sub-threshold effects, i.e., R(a+b) > R(a) + R(b) for high threshold neurons. Furthermore, a stronger analysis than our suppression index, δ_N, and its paired weighted variance across frequency, f_{var}, would be a more direct measure of frequency dependent summation. For the mapping between predicted and measured mixture responses, one could compute how much additional variance is accounted for by knowing the spatial frequency identify of each mixture. This could be achieved by allowing the Naka-Rushton fit of overall summation to vary with spatial frequency. However, despite our fairly exhaustive stimulus set, there are too few conditions to constrain such a model without, at the

minimum, assuming a particular functional form (e.g., some low-order polynomial) for how frequency alters summation linearity. Finally, our analysis groups mixtures by their central spatial frequency to reduce the 3, 5, or, 7 constituent frequencies down to 1 value. However, this simplification could be ignoring features (e.g., lowest or highest spatial frequency) besides the mixture's central frequency which might also be indicative of summation and its frequency dependence.

Chapter 4

A computational model for tuned gain control

4.1 Introduction

The remarkable range of visual neuron response properties has long motivated an understanding through the construction of models. Earlier models of primary visual cortex (V1), including those seeking to understand receptive field construction [63]; both linear [94] and nonlinear [93] response properties; as well as gain control [17, 55] provided great insight. These models went beyond simply explaining the observed responses and advanced the field by permitting simulations and, in turn, generating testable predictions. More recent single neuron, receptive field models have built upon the framework of paired linear-nonlinear (LN) processing to explain selectivity for stimulus features including orientation [107], spatial frequency [126], size [154], and color [34], among others. Models of a wide variety have accounted for response properties more generally, including models which explicitly identify the subunits of neurons [122, 160], network models [52, 29, 57, 120], and models which account for suppression through divisive gain control [25, 129]. An additional category of models are cascaded ones, in which the processing of more than one layer or area is considered, enabling

an identification of the origins and relative influences of specific computations. Such models have sought to understand how lateral geniculate nucleus (LGN) inputs shape V1 [114, 41], captured the influence of thalamic input on cortical network behavior and responses [29], and summarized inter-laminar differences in spatial integration [81].

The previous chapters of this thesis have highlighted the need for a computationally rich model of spatial frequency tuning in V1. Our own results and the work of others have shown that spatial frequency tuning varies with contrast, an observation which contradicts the invariant tuning yielded by divisive normalization. Even the earliest work within the normalization framework [55] notes that certain tuning properties (e.g., contrast adaptation [2]) are unaccounted for by a uniform gain control signal. Future studies highlighted the explanatory power of tuned gain control [99, 129] and showed its relationship to the statistics of natural scenes in both the visual and auditory domains [129]. In addition to the relationship between gain control and spatial frequency tuning, experiments designed to measure the tuning of suppression have shown (again, both in this thesis and in the work of other authors) that inhibition helps to sharpen broad LGN inputs into bandpass cortical tuning. Our work seeks to leverage a series of rich stimulus sets to understand the interplay between inhibition, contrast-dependent tuning, and normalization through a computational model of tuned gain control.

We constructed a linear-nonlinear receptive field model with gain control to account for a range of V1 response properties to stimuli that vary in spatial frequency and contrast. Using divisive normalization [55, 24], our model is comprised of a few simply parameterized components. To encapsulate the excitatory and suppressive signals that shape selectivity, our model includes a single linear excitatory filter, a gain control (or normalization) signal, and a thresholding non-linearity. We fit the model to single neuron spiking responses from three different experiments, each of which use stimuli varying in complexity to uncover linear and nonlinear computations. Crucially, we compare models with and without tuned gain control, and find that the majority of cell responses are better accounted for with a normalization

signal which is tuned for spatial frequency. In particular, we show that only the tuned gain control model can capture the observed contrast-dependent spatial frequency tuning, grating mixture summation properties, and frequency-tuned suppression. In agreement with previous studies [21, 162, 58], we find that the frequency tuning of normalization is near, but typically strongest below, the excitatory filter preference. Finally, we present an initial attempt at a cascaded model for spatial frequency tuning which takes into account import pre-cortical computations. Though the LGN-V1 model is insufficient in its current form, our results suggest that modest changes could help characterize the way in which thalamic selectivity for spatial frequency influences cortical tuning.

4.2 Methods

4.2.1 Stimulus sets

Three distinct stimulus sets make up the data considered in this chapter. The first two are the mixture and masking experiments discussed in the previous chapter (ch. 3), as well as their single grating counterparts (ch. 2). The only stimulus set not covered in the previous chapters is itself closely related to the mixture experiment. As in the mixture experiment, the stimuli in this experiment (which we term the 'dispersion' experiment) are comprised of either single sinusoidal gratings or mixtures of gratings. Rather than focusing on a wide range of contrast, this experiment instead spans a wider range of spectral dispersion. As a consequence, the dispersion stimulus does not test superposition; i.e., not all mixtures have components which were presented in isolation. In the dispersion stimulus set, all stimuli - whether single gratings or mixtures - appear at either full or one-third total linear contrast (1 or 0.33) and with spatial frequency centers ranging from 0.3 to 10 c/deg. We sample this range with 11 spatial frequencies, spaced half an octave apart. Like the previous mixture experiment, the central spatial frequency is presented with the highest contrast, with a Gaussian fall off away from the center frequency. Furthermore, this experiment uses mixtures of 3, 5, 7, or

9 gratings spaced 0.375 octaves apart; therefore, the mixtures span 0.75, 1.5, 2.25, and 3 octaves respectively. Lastly, unlike the mixture experiment in which temporal frequency was fixed throughout the experiment for a given condition, each grating presentation had a drift rate sampled from a Gaussian centered around the cell's preference and with a standard deviation of 20% of the distribution mean (e.g., 2Hz if the cell's preference was 10Hz). Due to variation in temporal frequency and the non-integer cycle length, we only analyzed DC responses for this stimulus, regardless of simple-complex designation.

4.2.2 Computational model of V1

Our computational model of V1 spatial frequency tuning is a linear-nonlinear model implementing contrast gain control, akin to [55]. Optionally, we include a preliminary LGN stage of filtering meant to capture the spatial frequency and contrast response properties of magno- and parvocellular neurons. The model response was time-varying, i.e., computed on a frame-by-frame basis, where each frame consisted of one or more superimposed sinusoidal gratings that varied in spatial frequency, contrast, temporal frequency, and phase.

This model was selective for spatial frequency but not orientation, stimulus size, spatial location, or temporal frequency. Any stimulus feature to which the model was insensitive was either constant (orientation, stimulus size, spatial location) or spanned a small range of values near the cell's preference (temporal frequency), across which we assumed the cell's response was constant. The contrast of each stimulus component was treated linearly by the (linear) filters but, through rectification and the normalization term, the model could generate a non-linear response to stimulus contrast.

The model was written in Python, and implemented to take advantage of Pytorch, which provides powerful optimization tools. The total number of parameters differed depending on the particular form of the model, but was between 5-9 parameters for all variations. We applied the square-root transform to the measured firing rates, r, as described in [111], and thus fit our model, whose response was given as $R(\vec{s_i};\vec{\theta})$, by optimizing the parameters, $\vec{\theta}$:

$$\min_{\vec{\theta}} \sum_{\vec{s_i}} \left(\sqrt{r_{\vec{s_i}}} - \sqrt{R(\vec{s_i}; \vec{\theta})} \right)^2 \tag{4.1}$$

For complex cells, r was simply the spike rate for a given stimulus presentation. For simple cells, r was itself a vector of length N_i, the number of sinusoidal gratings for stimulus, s_i. Thus, we computed the error between model and data on a per-component basis.

Stimulus encoding

We represented each component of the stimulus by its spatial frequency, contrast, temporal frequency, and phase. These values were encoded as numbers (e.g., 3 cycles/degree or 5.5Hz) which were passed through the corresponding filter(s) or used to encode the stimulus through time. First, we ensured the stimulus was accurately represented as a time-varying signal by progressing the phase of each grating i independently using its starting phase, ϕ_0, and temporal frequency, Φ_t

$$\phi_i(t) = \quad \text{mod } (\phi_0 + 2\pi * \Phi_t * t, 360) \tag{4.2}$$

We will discuss the inclusion of stimulus contrast and spatial frequency in more detail in the following two sections. Briefly, each contrast value served as a multiplicative term on the filter response; contrast was handled linearly in the first stage of the model (either the LGN filters or the V1 excitatory filter) and the normalization pool; spatial frequency was passed through the spatial frequency selectivity of each filter. Each component of the stimulus was "seen" by all filters in the model, which allowed for stimulus interactions, i.e., interference., when multiple components were within a filter's passband.

LGN stage

We simplified the LGN to two filters, one each to encompass magnocellular and parvocellular properties. All properties of this LGN front-end were fixed except for (1) the relative weighting

of magno and parvo subunits and (2) the absolute placement of the filters in spatial frequency. The relative weighting was parameterized as the proportion of LGN input given by the magno channel and was bounded between 0 and 1.

The contrast response properties of each channel were based on the relative sensitivity of magno and parvo neurons at a given eccentricity [65, 95, 149]. The contrast response functions were given by

$$R(c) = R \cdot log(1 + \frac{c}{c_0})$$

with c_0 as 0.05 (magnos) and 0.50 (parvos) to reflect the substantially stronger contrast gain of magnos. The spatial frequency selectivity of each filter was governed by a Difference-of-Gaussians (DoG) equation. The spatial frequency tuning of the magno and parvo channels were in fixed relationship to one another, with the center radius of the magnocellular filter always three times larger than that of the parvo [65, 32]. Furthermore, the gain and shape (i.e., relative surround radius and gain) of each DoG was fixed.Thus, the only optimized feature of the filters was their absolute spatial frequency tuning. We parameterized this placement as the central spatial frequency about which the parvo and magno filters were symmetrically above and below, respectively. Thus, the LGN front-end itself required only two additional parameters for optimization.

V1 linear filter

We used the derivative of a 1D Gaussian function to describe the V1 linear, excitatory filter. The Gaussian derivative is defined by two parameters - the preferred spatial frequency, ω_e, and the derivative order, α_e:

$$h(\omega; \omega_e, \alpha_e) = [\frac{\omega}{\omega_e} \exp(-\frac{1}{2}(\frac{\omega}{\omega_e})^2]^{\alpha_e} \tag{4.3}$$

Larger α_e yielded a narrow filter. This response, h, and all other stimulus-dependent responses are time-varying, though not explicitly labeled as such here. The time-varying response is

due to the independently drifting gratings (*eq.* 4.2), which result in the effective contrast varying with time. We note that the Gaussian derivative form above was used for multiple aspects of the model response - the excitatory filter described here, the normalization pool filters, as well as the frequency-dependent weighting applied to the normalization filters.

We threshold the time-varying filter response at 0 and apply a squaring operation to the activation response (half-wave rectified). For complex cells, we generated a phase-independent response by taking a quadrature set of the linear filter, i.e., taking the same filter and computing the response of 4 copies, each with a 90 degree phase offset. For simple cells, we compute just one filter with a fixed (i.e., non-optimized) phase. We do not optimize for the simple cell's phase selectivity, since we compute our error only on the response magnitude and ignore response phase.

Normalization pool

The normalization signal was modeled as a set of filters, g, - linear in contrast - that differed only in their spatial frequency preference, as given by a Gaussian derivative model (*eq.* 4.3). We created two pools of neurons, corresponding to narrow and broad channels, with 15 and 12 filters, respectively, whose center spatial frequencies spanned from 0.1 to 30 cpd. The narrow and broad filters had half-octave bandwidths of 2 and 3 octaves, respectively.

The contribution of these filters to the overall response was given by a weighted sum over the two pools of filters. The weight applied to a given filter was based on its preferred spatial frequency alone. In the case of flat (i.e., untuned) gain control, all filters were weighted equally. For models with an LGN front-end, the overall spatial frequency selectivity given by the particular placement and weighting of the magno and parvo channels was also applied to the normalization pool filters before further cortical weighting, if any. Finally, for the cortically-tuned gain control model, we applied a weighting to these filters, K, itself given by a Gaussian derivative for which we optimized the center and derivative order. Regardless of

the weighting chosen, the overall normalization signal was necessarily at least as broad as the narrowest filter in the normalization pool, i.e., 2 octaves.

The overall gain control signal was the sum of the stimulus-dependent component described above, and a normalization constant term, c_{50}. First, each normalization pool filter g_i is given a weight based on evaluating the weighting function, K, at the filter preferred frequency ω_i, i.e,

$$k_i = K(\omega_{gc}, \alpha_{gc})|_{\omega_i} \tag{4.4}$$

Then, the stimulus-dependent normalization response G, is

$$G(\omega, c) = \sum_{i=1..n} k_i \cdot g_i(\omega, c) \tag{4.5}$$

and the overall normalization response to a stimulus ω, c is then

$$N(\omega, c; \omega_{gc}, \alpha_{gc}, c_{50}) = G(\omega, c; \omega_{gc}, \alpha_{gc}) + c_{50} \tag{4.6}$$

As with the excitatory filter response, each g_i varies with time due to the drifting gratings and the interactions between different stimulus components. Likewise, we also apply half-wave rectification and take the response of a quadrature set for each filter.

Overall response, non-linearities

Thus far, we've described the linear-nonlinear LGN stage and linear processing of the V1 filter and normalization pool. The model contains two additional stimulus-independent terms which shape the response: one additive constant and one multiplicative scalar. The additive constant is added to the response after normalization, and allows the model to accurately capture the baseline response in the data. The multiplicative scalar is applied after computing the overall response, as given by the response of the linear filter normalized by the gain control signal. This parameter allowed the model to match the overall firing rates. Thus, the

overall response, R, given all model parameters, $\vec{\theta}$, and stimuli, $\vec{s_i}$, is computed by dividing the filter response h by the normalization term N,

$$r_t(\vec{s_i};\vec{\theta}) = \frac{h(\vec{s_i};\vec{\theta})}{N(\vec{s_i};\vec{\theta})} \tag{4.7}$$

$$R(\vec{s_i}) = b_0 + k \cdot \mathcal{F}(r_t(\vec{s_i};\vec{\theta})) \tag{4.8}$$

where the initial, un-transformed amplitude of the model response is given by the Fourier transform $\mathcal{F}$ of the time varying response r_t. The LGN influence is not explicitly shown, but would simply act as a scalar multiple on the underlying filters in h & N. Note that the basis of the model response, r_t is time-varying due to the stimuli being comprised of drifting gratings.

Plotting model responses

Our model allows us to present stimuli of arbitrary spatial frequency and contrast, so we present smooth model curves by simulating the model at frequencies and contrasts not presented in our stimulus sets. For plotting, we sampled spatial frequency tuning and contrast response curves at 50 log-spaced values across the stimulus range, though the response could be simulated for stimuli outside the tested range, too. We report only the response amplitude (i.e., not the response phase), and thus the model response to single gratings is independent of the stimulus temporal frequency or phase. However, for compound stimuli, the particular arrangement of component phases, as well as spatial and temporal frequencies, introduced unique irregularities in the stimulus which influence the model response. We attempted to minimize the consequence of this interference by randomizing the phase of each grating on each stimulus presentation, and accumulating a total of 10s of presentation for each stimulus condition. However, our model allows us to simulate the responses for arbitrary stimuli, so we can mitigate the influence of stimulus interference by sampling more stimuli. For all plotted

mixture model responses, we generated 100 distinct trials of each condition by generating random phases for each trial and showing the within-condition average of these responses. Despite this resampling, some model response curves for mixtures of many gratings (e.g., 5 or more components) nonetheless show some irregularities; these non-smooth deviations were caused precisely by the nature of stimulus interference.

4.2.3 Model comparisons, analysis

Cross-validation

Our models differed in their number of parameters, and thus we could not directly compare the loss or explained variance. We performed k-fold cross-validation with k=5. That is, we randomly trained on 80% of the data and evaluated the loss on the remaining 20%; we performed this process 5 times, where each holdout sample was comprised of unique trials. To avoid the influence of any longer time-scale changes in response, the folds were comprised of shuffled (i.e., not continuous in time) data. We report the cross-validated test loss as the average of the 5 holdout loss values.

Analysis of normalization tuning

The weighting of the normalization pool acts upon the overall response indirectly, i.e., as passed through the normalization pool filters. Thus we couldn't directly measure the relationship between the excitatory filter and the normalization tuning. We instead quantified the relationship between the normalization tuning and the excitatory filter via simulation. We densely sampled the model response at full contrast across the range (f_{min}, f_{max}) of spatial frequencies presented in each experiment and mean-centered the normalization pool response, $N(\omega)$, i.e.

$$N'(\omega) = N(\omega) - \overline{N}$$

Then, we calculated the area under the curve separately for frequencies above and below the excitatory filter's peak, f_0, normalized by the range of frequencies and took their difference, n_t,

$$n_t = \frac{1}{b}\int_{f_{min}}^{f_0} N' d\omega - \frac{1}{a}\int_{f_0}^{f_{max}} N' d\omega \tag{4.9}$$

where a, b are the octave ranges sampled above and below the filter's peak, respectively. To minimize the influence of frequencies outside the filter's passband, and where the relative tuning of normalization would therefore be irrelevant to the overall response, we cut back the integration bounds f_{min}, f_{max} where the excitatory filter response was less than 5% of the maximum. Larger values represent more asymmetric normalization tuning relative to the filter preference, with positive values indicating stronger normalization below the excitatory peak. If due to the filter placement, f_0 was equal to either f_{min} or f_{max}, we left n_t as undefined.

4.3 Results

4.3.1 A computational gain control model of V1 spatial frequency tuning

We recorded the responses of 166 V1 neurons to three related sets of experiments which consisted of sinusoidal gratings and compound grating mixtures that varied in contrast and spatial frequency. These stimuli elicited a wide range of responses and tested both linear (e.g., summation of multiple stimuli) and nonlinear response mechanisms (e.g., contrast gain). To model the excitatory and inhibitory computations that account for these responses, we fit the data with a linear-nonlinear computational model of V1 spatial frequency tuning. Our model implements gain control via divisive normalization [55] and is computed on a frame-by-frame basis, where each frame consisted of one or more superimposed sinusoidal drifting gratings that varied in spatial frequency, contrast, temporal frequency, and phase. This model was selective for spatial frequency but not orientation, stimulus size, spatial location, or temporal

frequency. The stimulus features to which the model was insensitive were either constant for all stimuli or chosen to span a small range of values near the cell's preference, across which we assumed the cell's response was constant.

The two primary components of the model are a linear filter and set of self-similar filters whose summed activity make up a normalization pool, as schematized in (*fig.* 4.1*A*-*C*). Both the excitatory filter, h (*fig.* 4.1B), and the normalization filters $\{g_i\}$, were parameterized as Gaussian derivatives (*eq.* 4.3), where only two parameters - a preferred frequency, ω, and derivative order, α, for determining tuning width - describe the tuning. We fixed two sets of normalization filters - one relatively narrower than the other (*fig.* 4.1*C*) but otherwise identical - which span a wide range of spatial frequency selectivity, and required no optimization. The model operated on the stimulus values for a given component, i, i.e., contrast, c_i; spatial frequency, ω_i; temporal frequency Φ_i, and stimulus phase, ϕ_i. We can represent the full set of stimulus specifications as $\vec{s_i}$. Thus, the stimulus-dependent response can be given as

$$R(\vec{s_i};\omega,\alpha) = \frac{h(\vec{s_i};\omega,\alpha)}{G(\vec{s_i}) + c_{50}} \tag{4.10}$$

where c_{50} acts as a stimulus-independent, normalization constant term. Furthermore, the stimulus-dependent normalization response, G, is given by the sum of the response of each filter, g_i in the pool

$$G(\vec{s_i}) = \sum_{i=1..n} g_i(\vec{s_i}) \tag{4.11}$$

All stimuli were drifting gratings, and thus the model response varied on a frame-by-frame basis (*eq.* 4.2). For single gratings, each filter's linear response was therefore a sinusoid matched to the grating drift rate. However, for mixtures of gratings, interaction between the constituent gratings interfered destructively such that $c_{1+2} < c_1 + c_2$, with the exact magnitude of interference dependent on the relationship between temporal frequencies and phase. All filters in the model (h, g_i) have their linear, frame-by-frame responses half-wave rectified (i.e., threshold at 0, then squared).

The stimulus-dependent response component requires the optimization of only three parameters: two which shape the excitatory filter's preference and width; and an additional stimulus-independent constant term. The formulation above (*eq.* 4.10) is akin to a Naka-Rushton equation (*eq.* 3.4) for contrast responses, as are other models which implement divisive normalization [55, 129, 160, 52]. To better match the measured response magnitudes, we include two additional scalar values, one which scales the response, R, and the other which serves as a constant offset to better match the cell's baseline firing rate. Thus, the core of the gain control model consists only of five parameters.

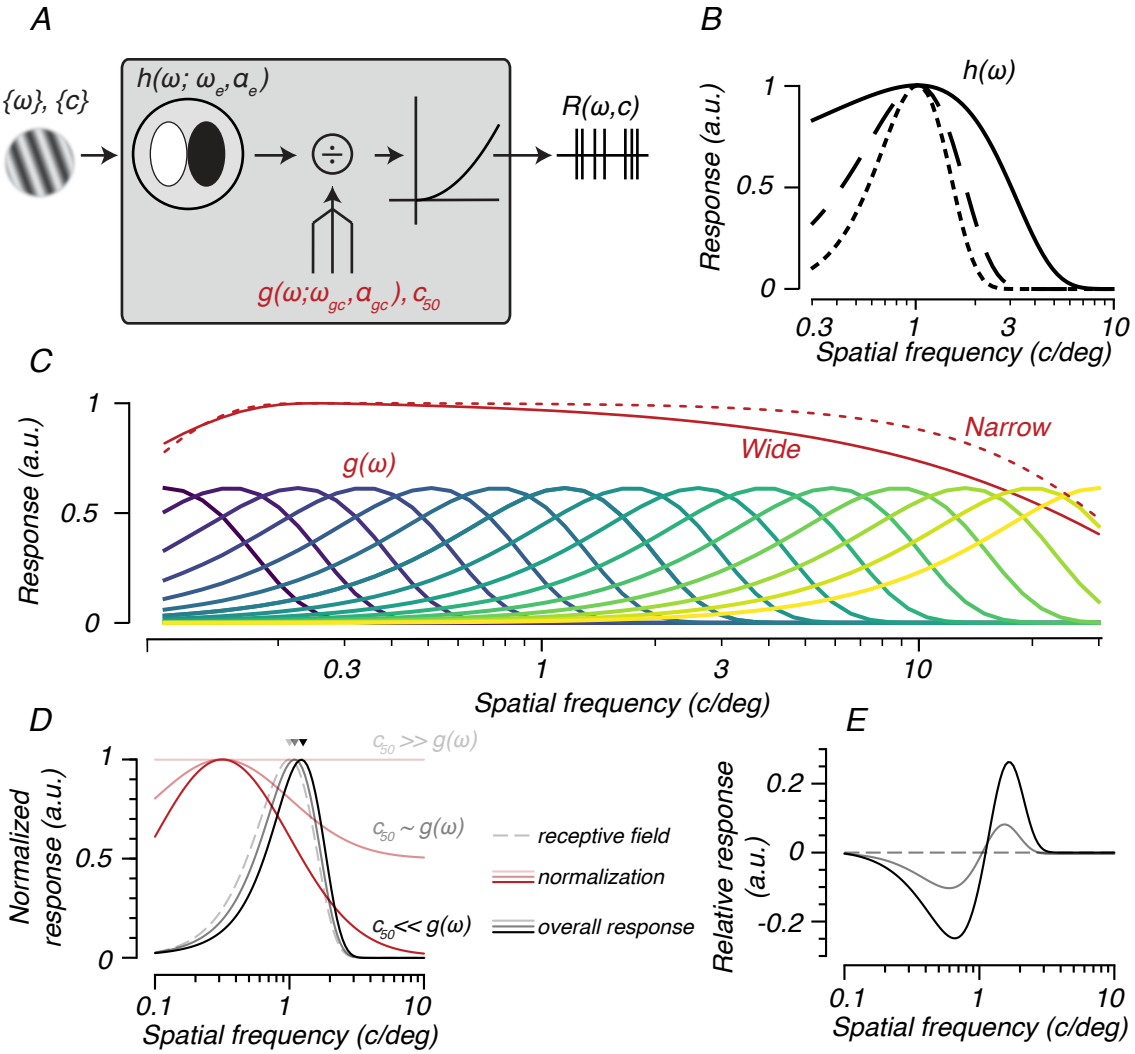

Figure 4.1: (A) Schematic of the LN computational model, which takes as inputs a set of sinusoidal gratings described by their spatial frequency (ω) and contrast (c). The linear filter response, h, is divided by the response of a normalization pool of self-similar filters, g, plus a constant term, c_{50}. The resulting response is passed through a static nonlinearity to yield a response rate, R. (B) The linear filter, h, is parameterized as a Gaussian derivative, where increasing derivative order yields narrower tuning (solid, dashed, dotted in increasing derivative order). (C) A set of self-similar normalization filters, g; the overall tuning is shown for both the narrow and wide set of filters (here, we show the individual filters from the narrow set). (D) Intuition for how the normalization response shapes the overall response given the stimulus contrast (increasing opacity) and a fixed excitatory filter (dotted line, termed receptive field). At low contrast (faint line), the constant term dominates the normalization signal. (E) Relative to the excitatory filter (dashed horizontal line at 0), we show the normalized, overall response of the example tuned gain control model. The tuning shifts towards higher spatial frequencies with increasing contrast.

We refer to the simplest model - in which all normalization filters are weighted equally - as one with flat or untuned normalization; the normalization signal is flat over a broad range

of spatial frequencies, with a gentle reduction at extremely low or high frequencies outside the passband of many cells (*fig.* 4.1*C*). Next, we also investigated a formulation of the model in which the response of each normalization filter g_i was weighted based on its preferred spatial frequency (*eq.* 4.5). The optimized weighting was determined by a Gaussian derivative (2 additional parameters, yielding 7 in total), and allowed for the normalization response to vary in strength as a function of spatial frequency (*fig.* 4.1*D*). A broader filter weighting yields a normalization term which is similar to a frequency-independent gain control, while sharper tuning can alter the overall model responses from the excitatory filter expectation. In particular, because the normalization response is comprised of a constant term and a stimulus-dependent term, the tuning of normalization will vary as a function of contrast - flatter for contrasts which do not elicit strong responses and more tuned for stimuli driving stronger responses. If the normalization response is relatively stronger above or below the excitatory filter, the tuning can shift as a function of contrast (*fig.* 4.1*D,E*).

4.3.2 Tuned outperforms untuned gain control

To better understand the relative capabilities of the tuned and untuned normalization models, we can first consider fits to example data. The dispersion experiment (*fig.* 4.2) differs from the previous mixture experiment (3.3.1) in that the compound stimuli are not superpositions. Unlike the mixture experiment which was designed to permit a direct data-only test of linearity, this stimulus was structured to strongly test normalization and thus serves as a useful comparison for our two models. Briefly, this dispersion-focused experiment (see Methods for greater detail 4.2.1) consisted of either 1, 3, 5, 7, or 9 superimposed sinusoidal gratings presented with a total, summed linear contrast of 0.33 or 1 (*fig.* 4.2*A*) and with center spatial frequencies taking any of 11 log-spaced values between 0.3 and 10 c/deg. As with the previous experiment, component spatial frequencies were logarithmically spaced around a given center frequency, and grating contrast was lower for components farther away from the central frequency. The most dispersed stimulus spanned 3 octaves of spatial frequency, thus

providing an important test for understanding the relative tuning of excitatory filter and normalization signal. The temporal frequency was randomized for each stimulus presentation, but chosen to be near the cell's preference. As with all previous experiments, orientation and size were kept constant at the cell's preference; phase was randomized; and the order of conditions was pseudorandomized.

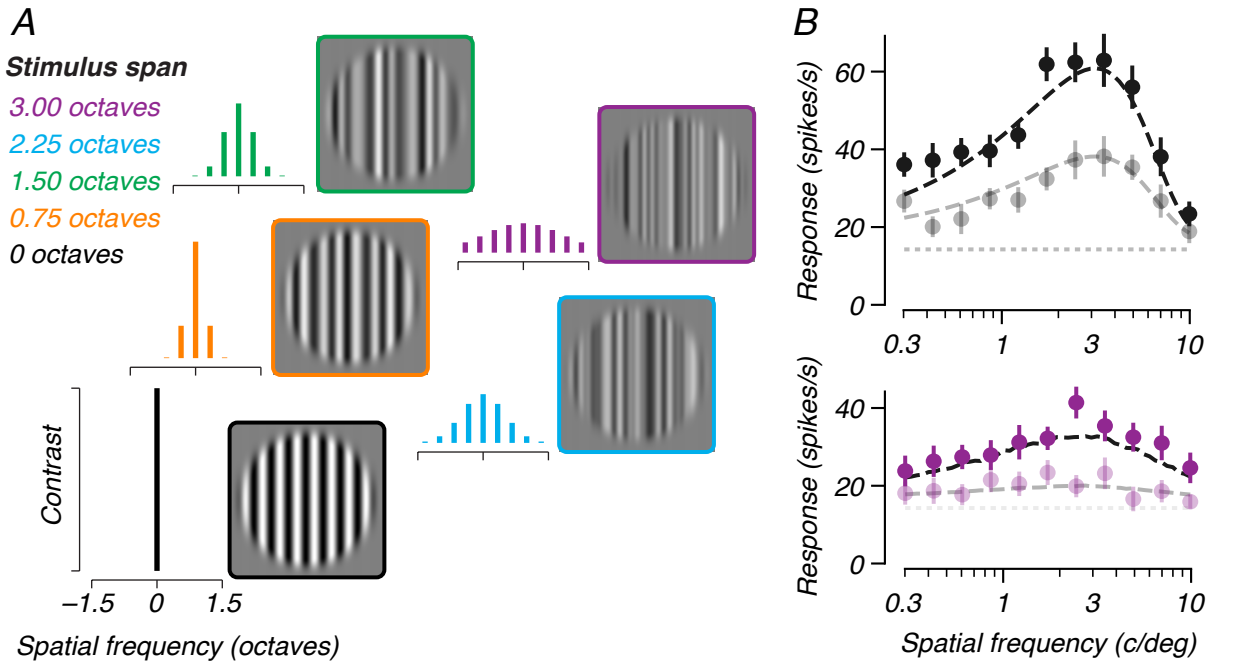

Figure 4.2: (A) Schematic of the dispersion-focused experiment, which varies from 1 to 9 superimposed sinusoidal gratings. The grating frequencies are symmetric in log space about a center grating. The total contrast always sums to either 0.33 or 1, and the grating contrasts are also symmetric, with the central grating at the highest contrast. Next to each frequency-contrast profile is an example frame from a given mixture condition presentation. (B) Selected responses from one example V1 neuron at two dispersion levels and two contrasts, 0.33 and 1, with untuned normalization model fits (*dashed lines*). Here, and throughout this work, we label and plot mixtures based on their central spatial frequency. This model accurately captures both the single grating tuning (*black*) and the responses to more spectrally complex mixtures (*purple*). The baseline firing rate is shown with the dotted line.

Four example conditions (single gratings and the most dispersed mixture, each at two contrasts) are shown for one cell in *fig.* 4.2*B*. The range of frequencies allows for a characterization of the spatial frequency tuning for single gratings and mixtures. We generate tuning curves for compound stimuli by simply grouping the mixtures by their central spatial frequency. In this example cell, the tuning to the most dispersed set of mixtures is broader

and of a lower response magnitude than for single gratings. These response properties were well-captured by our simply parameterized, untuned gain control model, which accounted for nearly 91% of the variance across all stimulus conditions.

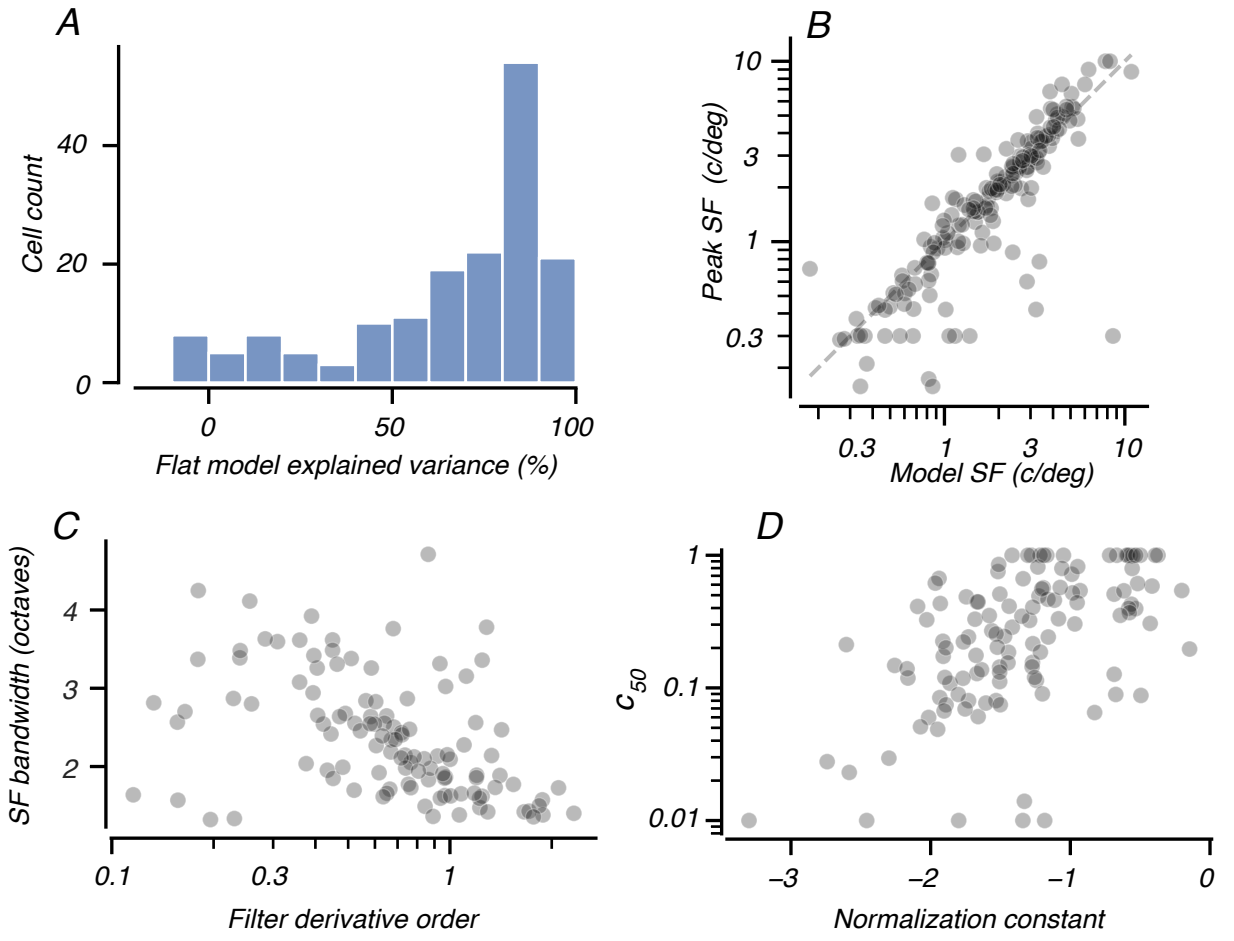

Figure 4.3: (A) Distribution of cross-validated explained variance for the flat/untuned gain control model. Population median of 75.4% (n=166). All correlations in panels B-D are significant via Pearson's correlation coefficient. (B) Strong correlation ($r = 0.86$; n=166) between the peak spatial frequency of the model's excitatory filter and the cell's preferred spatial frequency at full contrast as estimated from a descriptive function fit. (C) Strong negative correlation ($r = -0.45$; n=113) between the filter's derivative order and the cell's half-height spatial frequency bandwidth measured at full contrast. (D) Strong correlation ($r = 0.52$; n=122) between the model's normalization constant and the c_{50} of a Naka-Rushton fit to contrast responses near the cell's preferred spatial frequency.

Overall, the flat (i.e., untuned) gain control model was able to account for responses across all experiments. The median model explained variance on cross-validated data was 75% (*fig.* 4.3*A*). The cells which were poorly fit (38 cells, or 23%, had an explained variance

$< 50\%$) had statistically lower maximum and mean firing rates, as well as lower explained variance for a descriptive tuning function fit to the spatial frequency tuning curve at full contrast. We take this result as an indicator that lower model explained variances largely reflect noisier responses rather than model failures.

We next asked whether we could relate the core model parameters to independent measures of the cell's tuning properties (*fig.* 4.3*B-D*). First, we noted a strong correlation ($r = 0.86$; Pearson's correlation coefficient) between the excitatory filter's preferred spatial frequency and an estimate derived from a descriptive function fit to the cell's single grating spatial frequency tuning. Likewise, the filter derivative order behaved as expected: higher derivative orders were found for cells with narrower tuning as estimated by half-height spatial frequency bandwidth ($r = -0.45$). Finally, the normalization constant behaved comparably to the semi-saturation constant from a Naka-Rushton fit, as evidenced by their strong correlation ($r = 0.52$). Because the constant term is fit for each cell, the strength of the normalization signal is flexible. In particular, this means that for contrast-insensitive cells, the model normalization is dominated by the constant term (rather than the stimulus-dependent normalization pool responses) and the strength of gain control is therefore weak. Together with the distribution of model explained variances, we show that the model not only captures the wide range of responses across cells and experiments, but does so using the model parameters as anticipated.

Having established the lawful behavior of the gain control model, we can now directly compare the flat and tuned model responses for a series of example cells. Continuing with the dispersion-focused experiment, consider *fig.* 4.4. This narrowly tuned complex cell shows a differing influence of contrast on single grating and mixture responses. For single gratings, the responses at low contrast are nearly identical in magnitude to those at full contrast; yet for the dispersed stimuli, the responses at low contrast are both lower in amplitude and also narrower. While neither model perfectly captures the responses (e.g., note the failure to exactly match the single grating tuning at full contrast), only the tuned gain control model is able to capture the broader tuning for dispersed stimuli. In the context of our

normalization models, we might expect this discrepancy between the two if the tuned gain control was narrower, and thus less influential for high spatial frequencies compared to an untuned gain control. This is precisely what we observe in the underlying model filters (*fig.* 4.9*f4*). Nonetheless, because our tuned gain control model has both a stimulus-dependent and independent component, the normalization signal is effectively flatter at lower contrasts, where the former term is weaker. This expected property aligns with the responses, where dispersion tuning is narrower at low contrast and the flat and tuned models yield more similar responses.

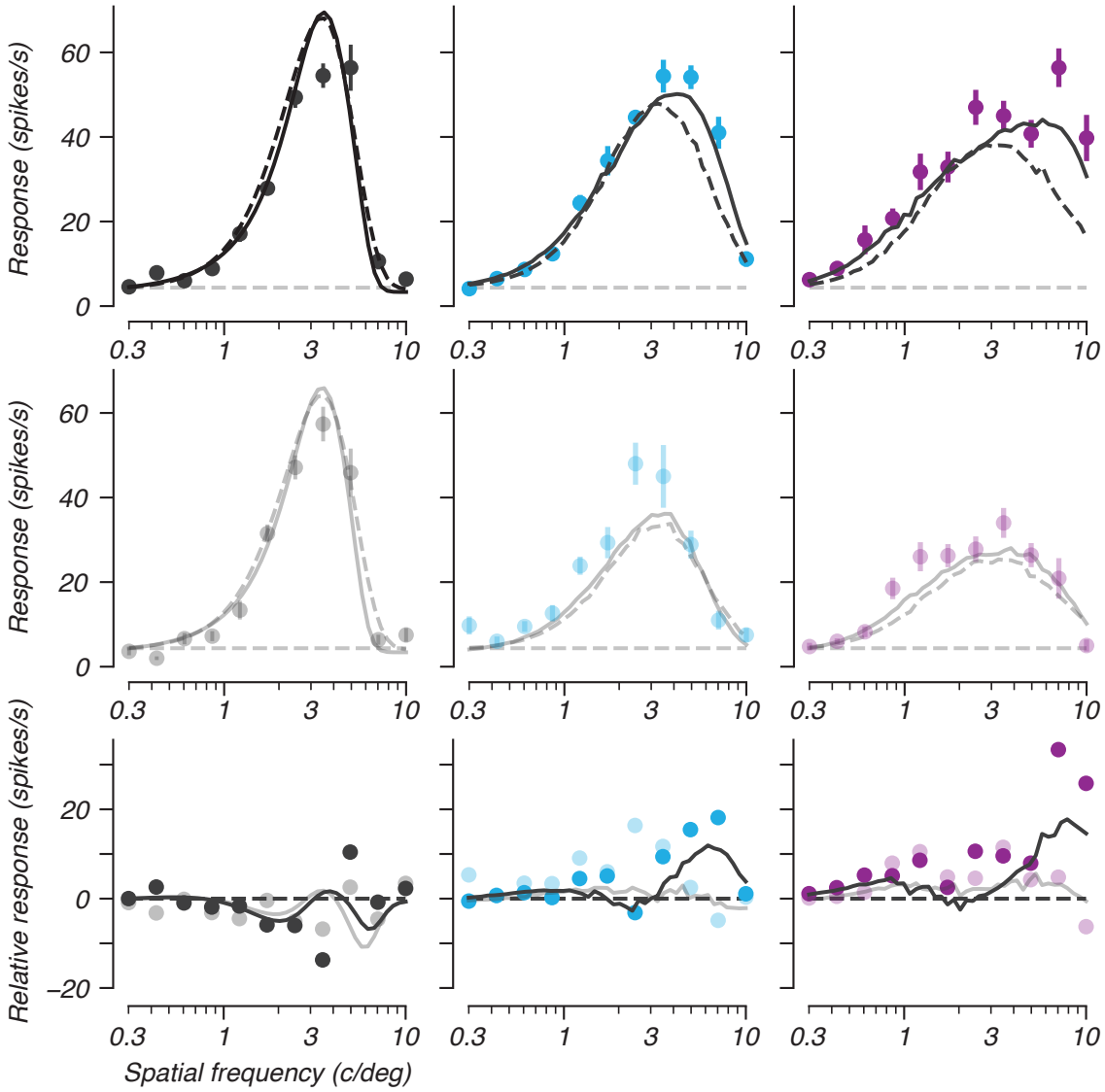

Figure 4.4: Model fits to a full dispersion experiment for one V1 complex cell. (*Top, middle*) Mean and standard error of responses alongside the flat (dashed) and tuned (solid) gain control model fits to single gratings and mixtures of 7 and 9 gratings, from left to right. The top and middle rows are for stimuli of 1 and 0.33 total contrast, respectively. Color and opacity refer to dispersion (number of gratings in stimulus) and contrast, respectively, but see *fig.* 4.2 for details. While even the untuned gain control model captures the tuning well, it fails to capture the broader tuning for mixtures at high spatial frequency. The tuned model had a holdout loss 17% lower than the flat gain control model. (*Bottom*) Compared to the flat model fit, the relative response and tuned model firing rates for the same conditions as plotted above.

The mixture experiment uses compounds of fewer gratings and samples spatial frequency less widely, but does offer a new view into the model behavior by more widely sampling contrast. Over the range of sampled contrasts, the example cell in *fig.* 4.5 shows a shift in

spatial frequency tuning, with reduced responsivity for high spatial frequencies. The flat gain control model has no mechanisms that can capture contrast-dependent tuning, and thus exhibits an identical shape for all contrasts. On the contrary, the tuned gain control model is able to capture tuning shifts for single gratings, as well as the relatively broader tuning for mixture gratings, as also shown in the previous dispersion example. This can be seen by the systematic nature of that model's failures (see the bottom panels of *fig.* 4.5), all of which are at least partially compensated for by adding a tuned gain control. Though not perfectly fitting the responses, the tuned gain control model is better able to capture tuning shifts for single gratings, as well as the relatively broader tuning for mixture gratings. The wide contrast sampling also permits an analysis of the contrast-response functions for different spatial frequencies. Once again, the flat model curves are re-scaled replicas at different frequencies, yet the data show differing contrast gain depending on frequency. The tuned gain control model is able to capture the higher contrast threshold and weaker saturation for higher spatial frequencies (*fig.* 4.5*B*). In our model framework, we should expect the strength of contrast gain and saturation to align with the strength of normalization; this precisely matches the underlying model filters for this cell (*fig.* 4.9*f5*), which show stronger normalization at lower spatial frequencies.

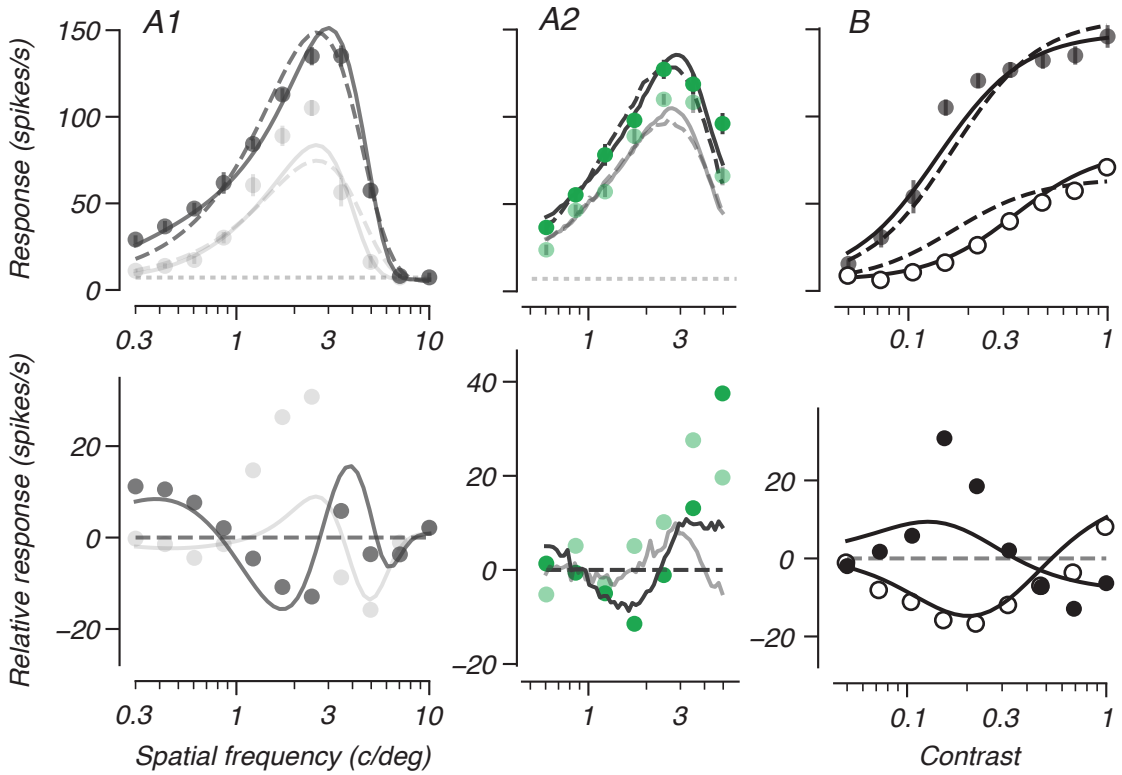

Figure 4.5: As in *fig.* 4.4, but for an example complex cell from the mixture experiment. Again, the tuned model (*solid curves*) captures contrast- and mixture-dependent response properties that the flat gain control model (*dashed*) cannot (holdout loss 13% lower for the tuned model). Though the tuned model cannot capture the tuning shifts entirely (note the under prediction of responses at low contrast near the cell's peak), it matches the data more closely. *A1* and *A2* are responses to single gratings and mixtures of 5 gratings, respectively at contrasts of 0.15 and 0.69 (refer to *fig.* 4.2 for details on mixture construction). *B* shows the response-versus-contrast for single gratings at 2.46 c/deg (filled) and 4.96 c/deg (hollow). The failures of the flat model highlight the shift in contrast gain across these two frequencies.

Finally, we turn our attention to the third stimulus set, a masking experiment. More than the other two stimuli, this experiment offers the most direct window into suppression, particularly for simple cells. Two example simple cells (*fig.* 4.6) show suppression which grows with mask contrast and is stronger for lower spatial frequencies. While even the untuned normalization model can capture masking suppression, the resulting suppression is largely independent of the mask frequency. However, as in these examples, we found that mask suppression is typically restricted in its frequency tuning (see the previous chapter, ch. 3, for details). The tuned gain control model, on the other hand, can capture both the

contrast-dependence and frequency-selectivity of suppression. As with the prior examples, the model responses are reflected in the tuning of the underlying filters (4.9*f6*).

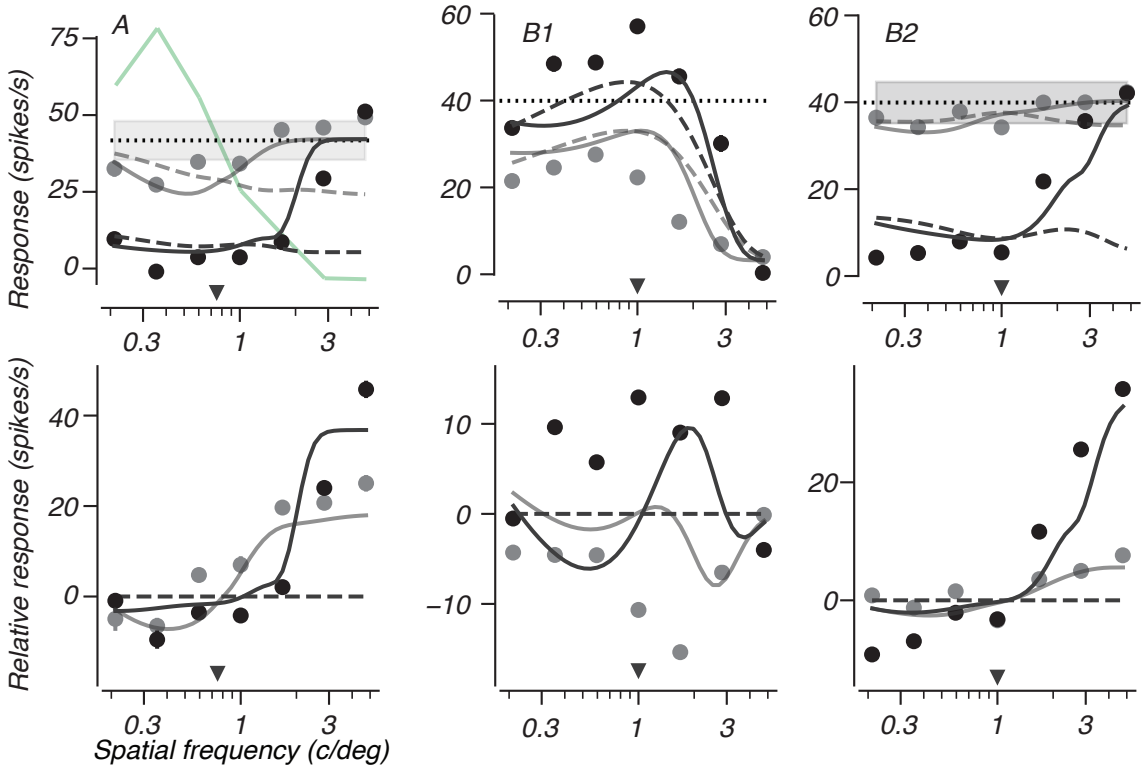

Figure 4.6: As in *fig.* 4.4, 4.5 for two simple cells from the masking experiment. Here, both cells are shown at contrasts of 0.64 (black) and 0.08 (gray). The black dotted line and gray bounding box represent the mean and ±0.5 std. of the base stimulus response; the triangles along the horizontal axis indicate the base spatial frequency. As a reference, the faint green curve on the top of figure A is an interpolation of the mask-only spatial frequency tuning at 0.64 contrast. Akin to the prior figures, the top row shows the data and model fits; the bottom row shows data and the tuned model response relative to the flat model prediction. For both cells, the joint mask+base responses (*A, B2*) show tuned suppression which is stronger for frequencies at or below the cell's preference. Suppression falls off steeply beyond the cell's passband. The mask-only response is shown only for cell B (*B1*). Frequency-dependent suppression is only captured by the tuned gain control model, which had improvements of 19% and 26% for the two cells, respectively, over the untuned model.

Because we can access the underlying tuning properties of our simple gain control models, we can infer the excitatory and inhibitory selectivity of cells even when not directly measurable. This is particularly useful for DC responses in the masking experiment, since most mask stimuli simply add to the the base response rather than suppress firing. As suggested by other studies and our own work, we expected that the tuning of masking suppression is likely

broader than can be observed directly for complex cells. We highlight two example complex cells (*fig.* 4.7) which both show suppression only for masks outside the cell's passband - one each for frequencies above and below the cell's preference. While the untuned gain control model shows suppression for masks of high contrast, the resulting suppression is observed for all frequencies outside the cell's passband. Again, the tuned gain control model is able to capture the frequency selectivity of the masking suppression. Furthermore, measuring the tuning of the underlying normalization filters confirms the hypothesis of broader suppression - particularly at low and intermediate spatial frequencies - than is directly observed (*fig.* 4.9*f7*).

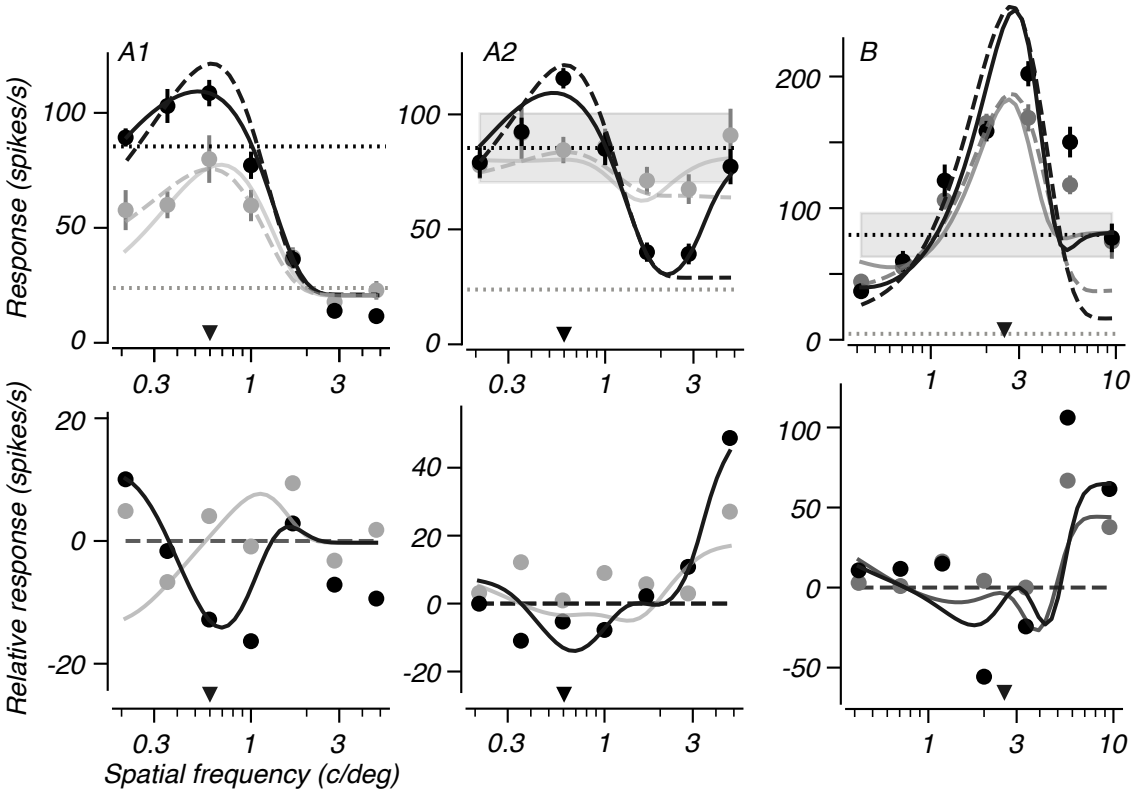

Figure 4.7: As in *fig.* 4.6 but for two complex cells. The dotted lines represent the base only (*black*) and blank stimulus (*gray*) responses, respectively. *A1* shows the mask only responses and model fits for one example cell. In *A2*, we plot the joint mask+base response, which shows strong suppression for frequencies above the cell's peak that only the tuned model (*solid curve*, as before) captures. (*B*) Joint mask+base responses for an additional cell, highlighting the inability of the flat gain control model to show suppression only at low spatial frequencies. The tuned model shows a lower cross-validated test loss by 8% and 18% for the two cells, respectively.

To compare overall model performance, we performed k-fold cross-validation (with $k = 5$) and calculated the relative improvement compared to the model without tuned gain control as

$$I_{tuned} = \frac{\epsilon_{tuned} - \epsilon_{flat}}{\epsilon_{flat}} \tag{4.12}$$

where ϵ is the holdout test loss (see (4.1) in Methods for further detail). The tuned model consistently outperformed the flat model, with 81% of cells having a lower cross-validated loss with the tuned gain control model (*fig.* 4.8) and the distribution of loss improvement, I_{tuned}, significantly shifted away from 0 (median improvement of 2.58%, $p <$ 1e-10; Wilcoxon rank-sum). As a control, we also fit a version of the model in which the filter determining the weighting over the normalization filters was not separately optimized, but instead matched to the V1 excitatory filter; we term this model the 'matched' model, in contrast to the 'tuned' model discussed throughout. This model also outperformed the untuned model, and significantly so, with better cross-validated loss in 79% of cells and a median improvement, $I_{matched}$, of 1.84% ($p <$ 1e-10). Comparing the tuned and matched models directly, we found the median absolute improvement, $I_{tuned} - I_{matched}$, yielded by the tuned model was 0.54%. Though the additional improvement yielded by the optimized-for tuned gain control was often modest compared to the matched model, the former was consistently better. In our population, 72% of cells showed lower cross-validated test loss with the tuned rather than the simpler, matched model (*fig.* 4.8, *right*; $p <$ 1e-8).

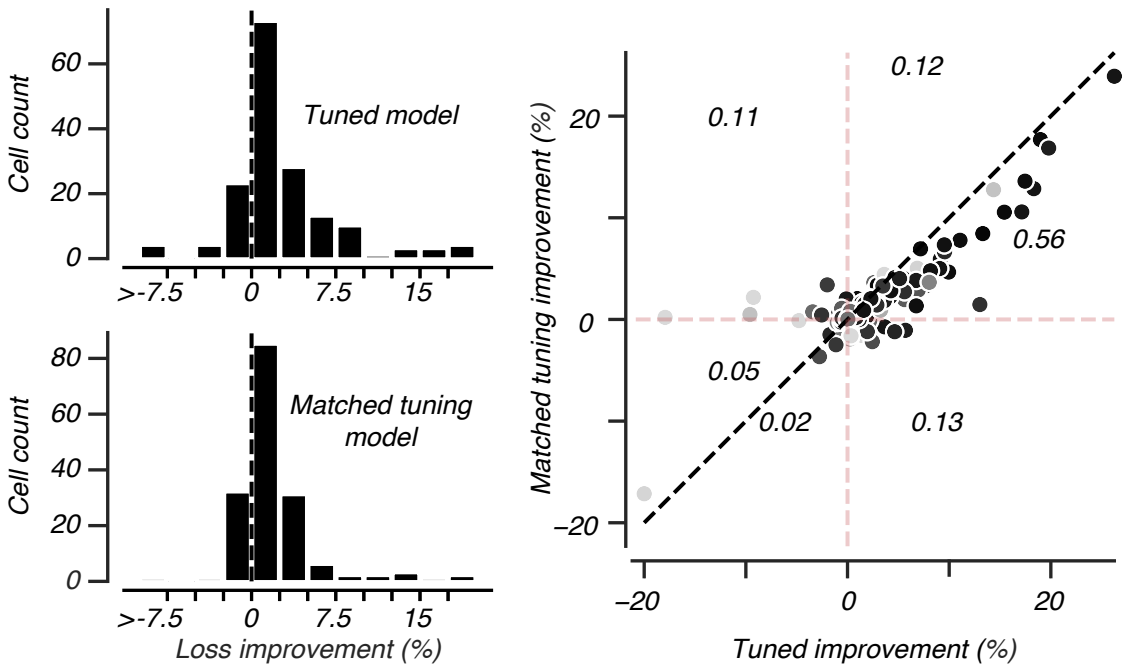

Figure 4.8: (*Left*) Relative cross-validated loss improvement compared to the flat gain control model for the tuned (*top*; normalization tuning unconstrained) and matched (*bottom*; normalization tuning matched to the excitatory filter) models. Both models consistently perform better than the flat model. (*Right*) Direct comparison of the relative improvements of the two models, with opacity proportional to the model explained variance; numeric values reflect the fraction of the total cell count present within each (sub)quadrant. Roughly 68% of cells are better explained by both of these models (upper-right quadrant) compared to the flat gain control model. The prevalence of points below the diagonal demonstrates that the majority of cells were better explained by the tuned model than the matched tuning model. Conversely, only 7% of cells were better explained by the flat model, and the fits located deepest into the lower-left quadrant were generally poorly fit (less opaque).

Having established the success of the tuned gain control model, we then sought to identify the reason for the improved fit quality by quantifying the tuning of the normalization signal relative to each fit's excitatory filter. Briefly, we computed the relative strength of the normalization pool below and above the filter's preference. We termed this metric, n_t, i.e., the normalization tuning summary (see Methods for further detail, 4.2.3). Positive values indicate stronger normalization below the excitatory peak and the magnitude of n_t reflects the normalization's strength and selectivity. The distribution of n_t (*fig.* 4.9*A*) was centered just above 0, with 57% of cells showing stronger normalization below the peak (median of 2.6e-3; $p = 0.03$, Wilcoxon rank-sum). Crucially, there was a modest, yet reliable

correlation between n_t and the relative improvement yielded by the tuned gain control model ($r = 0.17; p = 0.03$). Restricting the comparison to only well-fit cells (here, those with explained variance greater than 50%), the relationship was stronger, still ($r = 0.24, p = 0.007$).

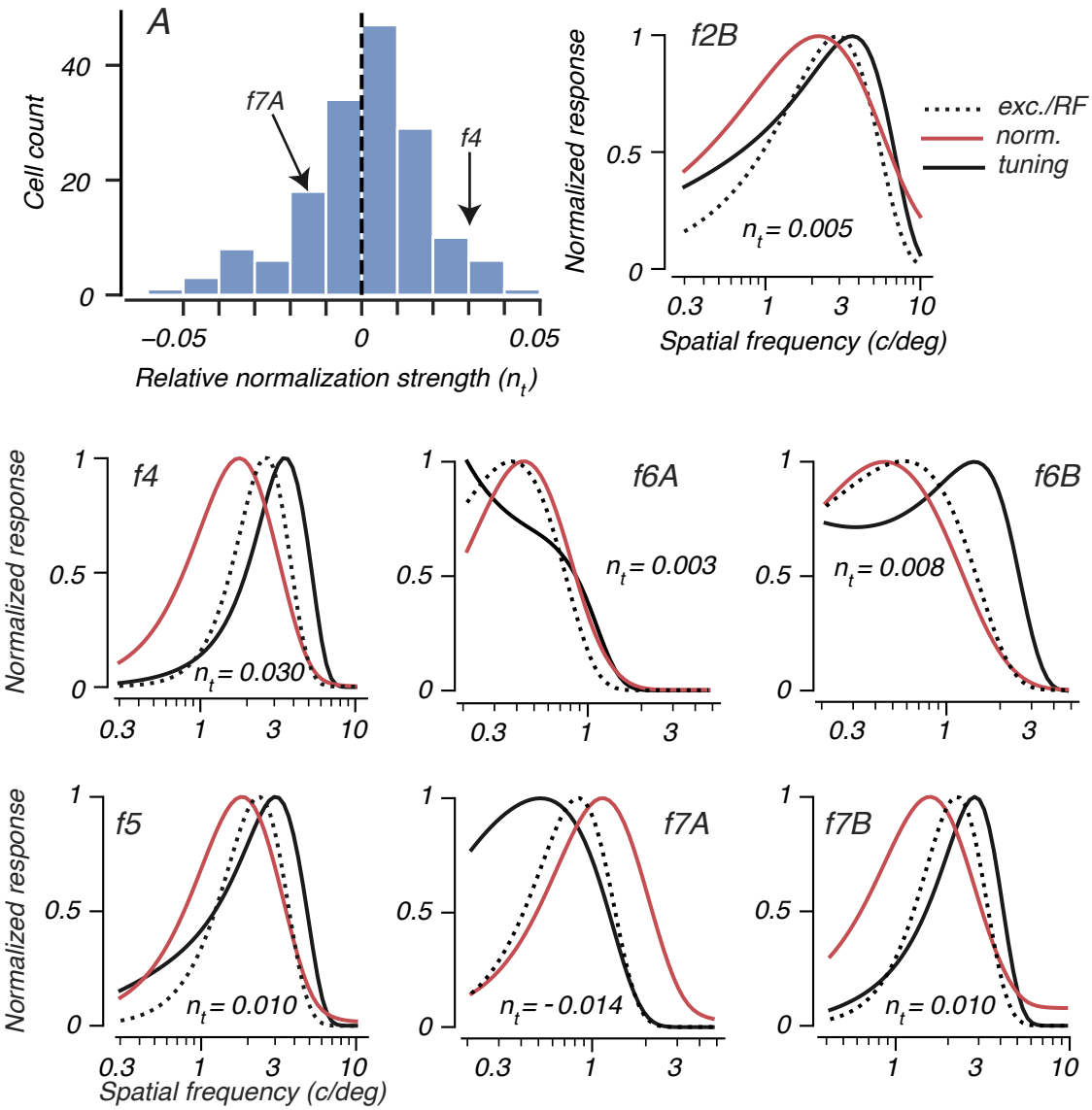

Figure 4.9: In A, we show the population distribution of normalization strength relative to the excitatory filter, termed n_t, where positive values indicate stronger normalization below the filter peak. The distribution is centered near, but typically greater than, zero (median of 0.0026; $p = 0.03$, n=163; Wilcoxon rank sum test). We also highlight the location in the n_t distribution for the two most extreme values shown in this plot. The remaining panels ($f*$) show the relative, rescaled tuning of the excitatory filter (i.e. receptive field), normalization response, and overall cell response for each of the example cells in the previous figures, along with the n_t value.

We then determined whether any underlying cell properties were related to either the strength and bias of normalization tuning or the overall model improvement garnered by a tuned gain control. This search was focused, in part, on the observations we made in previous chapters (3.1). Namely, we uncovered relationships between key tuning properties (e.g. preferred spatial and temporal frequency, surround suppression) and the strength of frequency-dependent summation and suppression. For the analyses below, we only considered models which were well fit (i.e., the 127 cells, 77% of the total, with explained variance $> 50\%$), though the results were qualitatively similar, albeit weaker, when considering all fits. We found the relative strength and placement of normalization tuning, n_t, to be predictive of several cell properties, including those which were related to the strength of frequency-dependent mixture summation, f_{var}. Complex cells showed stronger, more biased normalization tuning compared to simple cells, with n_t population medians of 4.5e-3 and 0.3e-3, respectively ($p = 0.04$). Neurons with higher preferred spatial frequencies also showed stronger relative normalization tuning, particularly among well-fit cells ($r = 0.28, p = 0.002$). As predicted in our model schematic (*fig.* 4.1*D,E*), contrast-dependent spatial frequency shifts were also correlated with n_t ($r = 0.33, p <$ 1e-3): cells with stronger shifts in spatial frequency tuning with contrast showed normalization which was more biased for frequencies below the peak. No other tuning property (e.g., orientation selectivity, size tuning, temporal frequency preference) showed a strong relationship with n_t.

Curiously, neither the strength of frequency-dependent summation nor the presence/absence of masking suppression (both covered in chapter 3) were directly correlated with model normalization tuning, n_t. However, if we consider only simple cells (due to the difficulty in measuring suppression for complex cells), we find that cells classified as suppressed showed stronger, more biased normalization tuning ($p = 0.03$).

The extent to which the tuned model improved fits over the untuned gain control, I_{tuned}, was also linked with several underlying tuning properties. Namely, cells which were more contrast sensitive per the semi-saturation contrast c_{50} ($r = -0.27$), narrower in spatial

frequency tuning ($r = 0.30$), and selective for higher temporal frequencies ($r = 0.17$) all showed greater fit improvement with the tuned model. The latter two metrics were also positively correlated with the strength of masking suppression. Together with the correlations between the underlying parameters of the tuned normalization model and frequency-dependent summation (f_{var}), these findings suggest the following: not only does the tuned gain control model capture important contrast-dependent response properties, but it does so in a way that reflects other important, related tuning measures. In particular, the findings for cells with high contrast sensitivity and temporal frequency selectivity suggest that our model is capturing the hallmarks of untuned, magno-driven suppression [59]. This form of suppression was well-targeted by our masking experiment in the prior chapter (see discussion, §3.4.1), and fits to those data here show marked, readily visible improvement with a tuned gain control.

4.3.3 Introducing an LGN front-end

We now present a preliminary attempt at a cascaded model of V1 spatial frequency tuning. Rather than only cortical filters which directly 'see' the stimulus, this form of the model begins with an LGN front-end stage whose responses then provide input to the V1 stage described above. We present a very simplified version of an LGN front-end here which serves as an initial attempt to more accurately capture the hierarchical nature of sensory processing. In particular, by modeling the input responses to V1, a cascaded model can better characterize the relative contributions of pre-cortical and *de novo* cortical influences on the excitatory and inhibitory computations we have characterized with our tuned gain control model. We present this effort as a step towards a more refined version which should match, if not outperform, the V1-only model and therefore permit a quantitative comparison of thalamic and cortical influence in V1 spatial frequency.

We model the LGN with two simple filters, one each for capturing the basic tuning properties of magno (M) and parvocellular (P) neurons. The front-end stage required only two parameters to be optimized - one which determined the absolute placement of their spatial

frequency selectivity, and one for weighting their responses. The LGN filters are selective only for stimulus contrast and spatial frequency, and their parameters were - with one exception, noted below - fixed (i.e., non-optimized). Unlike the V1 stage, where contrast gain arises from the division of excitatory responses by a normalization pool, we explicitly model contrast saturation using a simple form given as $R(c) = k \cdot log(1 + \frac{c}{c_0})$ [43]. To reflect the observation that magno contrast sensitivity is 5-10 greater than that of parvos [71, 95], we fixed (k, c_0) at $(12.5, 0.05)$ and $(17.5, 0.50)$ for the M & P filters, respectively. This parameterization results in a relatively stronger influence of magno input at lower contrasts. In addition to the contrast gain of the LGN stage, we also imposed tuning for spatial frequency. We modeled magno and parvo as Difference-of-Gaussians (see 2.2.3) which differed only in their center size and surround strength. We fixed the relative center size in a 3-fold relationship [65, 32], where parvo centers were smaller, and therefore more sensitive to high spatial frequencies. The only optimized parameter for the LGN filters was the center frequency about which the magno and parvo filters were centered. Parvo surround subunits are typically stronger than those of magnos [32], which we also took into account with our DoG parameter choice.

The responses of the two LGN filters were linearly combined and served as input to the V1 filters by acting as the effective stimulus contrast. This linear combination was governed by the second and final LGN-stage parameter that determined the relative weighting of the M & P filters. The weight parameter was bounded between $(0, 1)$ where the extrema indicate only parvo or magno input, respectively. Beyond this initial filtering, the remaining filters and computations of the model were unchanged. We tested three forms of the LGN model which mirrored the variations of the V1-only model - one without tuned gain control ('flat'), one with tuning matched to the V1 excitatory filter ('matched'), and one with an optimized gain control tuning ('tuned'). The LGN front-end applies to all filters, including those in the normalization pool; thus all gain control tuning is layered on top of the LGN spatial frequency filtering.

An example LGN front-end is shown in *fig.* 4.10*A*, with separate panels for the magno and parvo filters, as well as their combined response. The combined response provides the initial filtering of the stimulus before any of the cortical filters. Crucially, for a cell with relatively balanced M & P input (the example shown has a 30:70 split), the overall shape of the LGN spatial frequency filter will change as function of contrast, as observed in this example. First, we ensured that the LGN front-end didn't interfere with the overall model behavior by assessing the relationship between three key parameters and their analogous tuning properties (filter preference and spatial frequency preference; model derivative order and bandwidth; normalization constant and c_{50}). As with the V1-only model (*fig.* 4.3), these relationships were strong, albeit slightly weaker for the filter preference ($r = 0.86$ vs. $r = 0.74$ for the LGN model) and normalization constant ($r = 0.52$ vs. $r = 0.40$). All properties reported below are for the 'matched tuning' LGN model, though the results were comparable across all three versions (flat, matched, tuned). We next asked whether the two additional LGN-specific parameters were also related to tuning properties (*fig.* 4.3*B,C*). First, we noted a strong correlation between the spatial frequency placement of the LGN filters and an independent measure of each cell's preferred spatial frequency ($r = 0.49; p <$ 1e-10), suggesting that the model correctly used the LGN front-end to start shaping the overall frequency selectivity. Given the relative contrast sensitivity of magno and parvo units, we would expect cells with low contrast thresholds to receive more magno input. We found exactly this relationship, with a negative correlation between c_{50} and the relative strength of magno input in the model ($r = -0.27; p = 0.0025$). We did find, however, that our distribution of magno weightings was skewed, with a strong bias towards mostly parvocellular input (population median of 0.15).

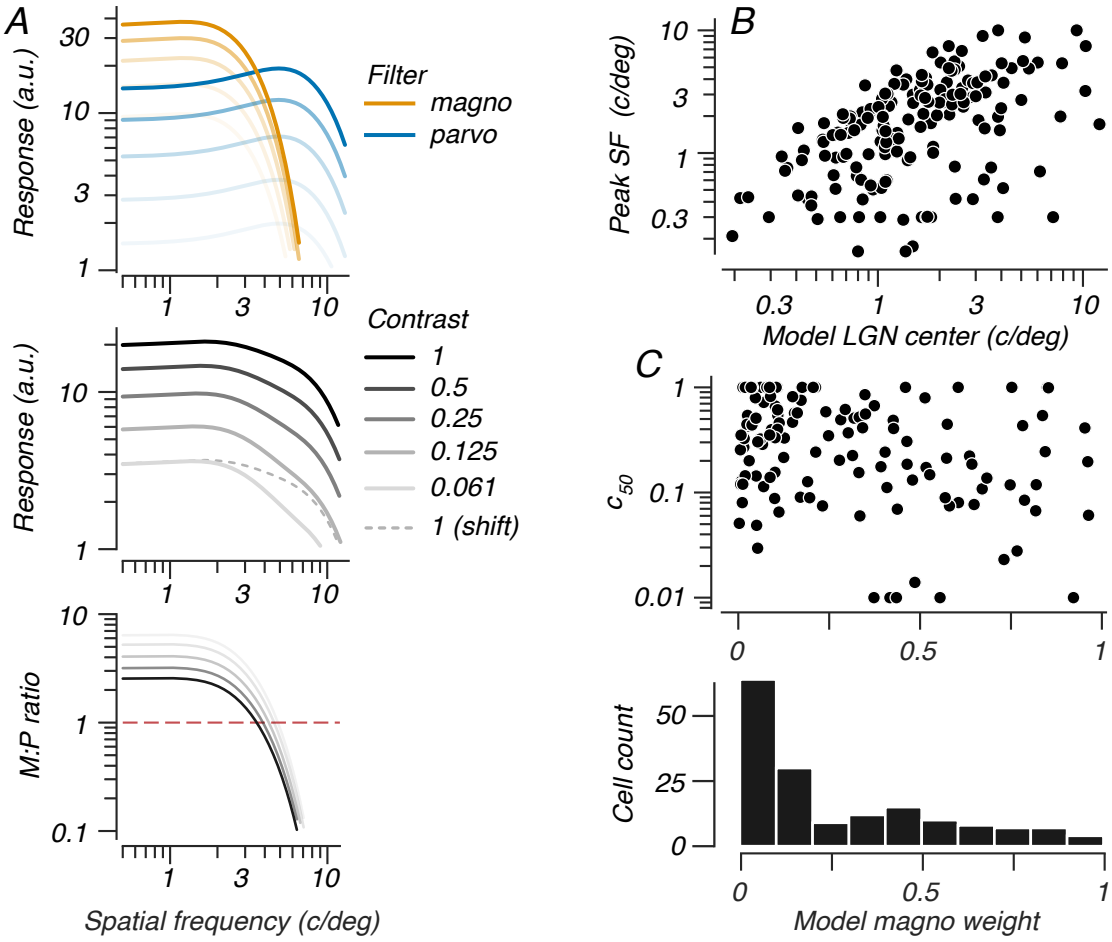

Figure 4.10: (*A*) The separate (*top*) and combined (*middle*) response of two LGN filters for one example neuron on equal-scale log axes, with contrast indicated by opacity. The filters are combined with magno/parvo weights of 0.3/0.7. To highlight the shift towards lower spatial frequency tuning at lower contrast, the dotted line (*middle*) shows a vertical-only shift of the combined filter tuning at full contrast. The relative strength of magno input (*bottom*) is stronger for low spatial frequencies and at lower contrast, accounting for the shift in tuning noted in the panel above (*middle*). (*B, C*) Relationships between two LGN-specific parameters and their closest equivalent tuning property: LGN center and preferred spatial frequency (*B*; $r = 0.49; p <$ 1e-10) and magno input strength compared to c_{50} (*C*; $r = -0.27; p = 0.0025$). The bottom panel of *C* shows the distribution of model magno weights, which was heavily skewed towards 0 (population median of 0.15), i.e., typically more parvocellular input to V1.

Next, we sought to compare LGN model performance to the V1-only fits. For all three forms of gain control, the LGN front-end made the fits worse as measured by relative cross-validation loss with their equivalent V1 model (*fig.* 4.11*A*). For the untuned gain control model, the LGN front-end hampered fit quality but not statistically so (median cross-validated loss reduction by 0.3%; 46% of cells worse, $p = 0.53$ Wilcoxon rank-sum). Thus, across the

full population of cells, the inclusion of an additional linear-nonlinear (fixed) stage, at best, failed to improve the model fits. For the matched and tuned models, the influence of the LGN front-end was more detrimental per the population distribution of relative loss ($p <$ 1e-3 for both distributions; Wilcoxon rank-sum).

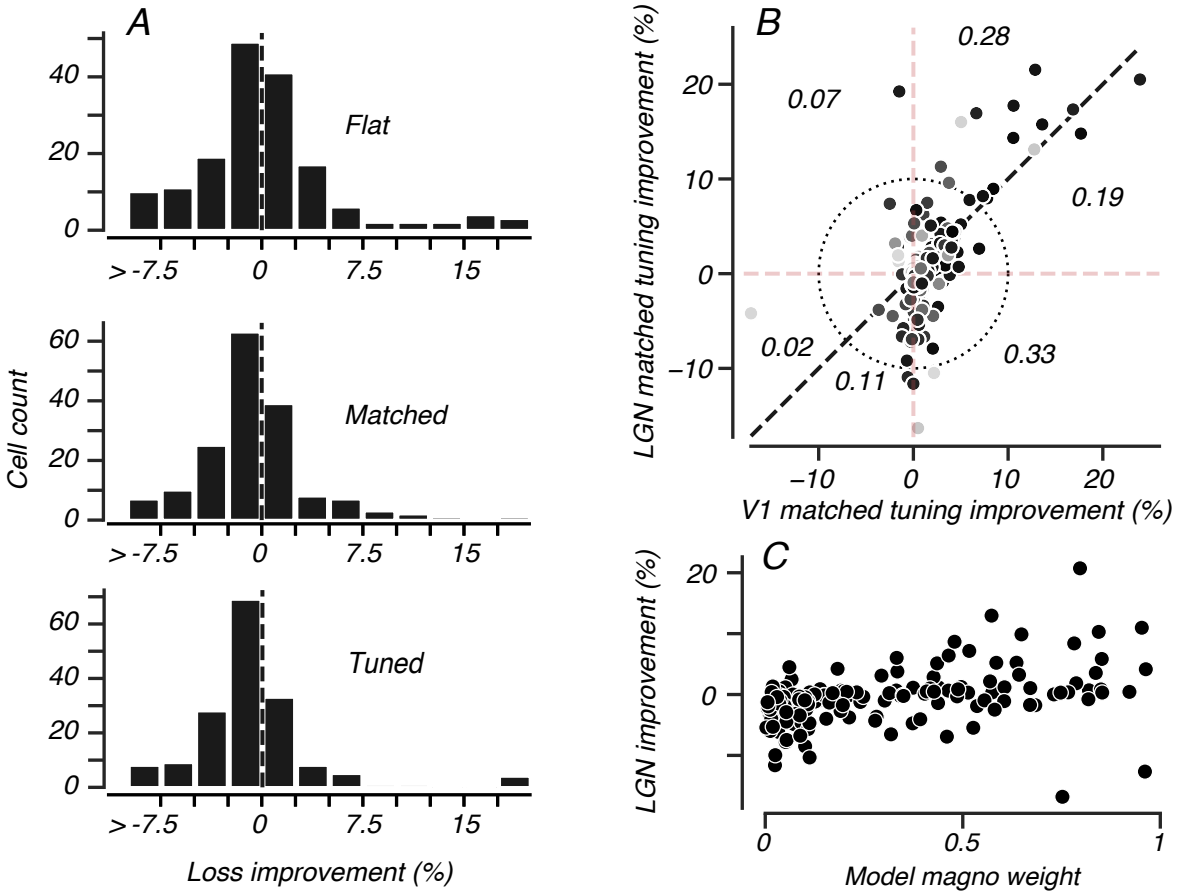

Figure 4.11: (A) Three histograms with the LGN model loss improvements relative to their V1-only equivalent. Positive values indicate cells for which the addition of the LGN front-end improved fit quality. From top to bottom, 46.4%, 36.7%, and 31.3% of cells had a better cross-validated loss with the LGN front-end. For the models with matched and tuned gain control, the LGN fits were significantly worse ($p <$ 1e-3 for both distributions; Wilcoxon rank-sum). (B) Relative improvements in cross-validated loss compared to the V1 untuned model for V1 and LGN matched models. For visual reference, the dotted circle marks a 10% change in loss; numeric values reflect the fraction of cells present within each (sub)quadrant; opacity is proportional to fit quality for the untuned V1-only model. (C) Scatter plot of the model's magno weighting and the extent to which the LGN front-end improves model fit quality; the correlation was strong ($r = 0.41; p =$ 4e-8).

Given the varied, albeit typically worse, fit quality of the two stage models, we sought to understand why certain cells were better suited to the LGN model than others. To isolate

the LGN front-end influence on fit quality, we now focus on comparing models with matched gain control tuning (i.e., the weighting over the normalization pool is matched to the V1 excitatory filter). This comparison avoids the additional cortical shaping of the LGN tuning that is permitted if we compare the influence of the LGN front-end for the (freely) tuned model; likewise, we do not handicap the V1 model by forcing a flat normalization signal. We plot the relative improvement compared to the untuned V1 model for these two models in *fig.* 4.11*B*, where as discussed above, the majority of cells show better fits without the LGN front-end. However, we did find that of the cells which showed more substantial improvement with the inclusion of some gain control tuning (see reference circle in *fig.* 4.11*B*), many were equally well, if not better, explained by the model with an LGN front-end; conversely, many of the cells for which the LGN front-end detracts from fit quality were cells with more modest improvements from a V1-only matched tuning gain control.

The most rigid aspect of the LGN front-end was the contrast saturation introduced by the fixed M & P non-linearities. We also found that the distribution of relative magnocellular weighting was neither uniform nor bimodal (*fig.* 4.10*C*), suggesting that this parameter was poorly utilized. Finally, we noted two modest relationships related to each cell's contrast gain control (c_{50}): first, for the V1-only model, cells with stronger saturation yielded greater improvement with tuned gain control; second, the model optimization chose stronger magno input for cells with stronger saturation. Motivated by this set of relationships, we asked whether the model's weighting of magno and parvo inputs influenced the change in fit quality brought about by the addition of LGN filtering. We found a significant correlation ($r = 0.41; p = 4\text{e-}8$) in which cells with stronger magnocellular input to V1 showed greater improvements over the untuned V1-only model (*fig.* 4.11*C*). We elaborate further in the discussion, but propose that for cells with non-linearities well-matched to our fixed LGN properties, the cascaded filtering helps; for other cells, the LGN front-end only reduces fit quality.

4.4 Discussion

The results above cover three different experiments, each with a distinct set of stimulus properties designed to probe excitatory and divisive computations for V1 spatial frequency tuning. Both a standard, untuned gain control and one which is tuned for spatial frequency capture the overall profile of contrast sensitivity and tuning. However, in each of the example responses we highlighted, the tuned gain control model captures features of responses that an untuned normalization signal could not. Across two experiments which widely sampled contrast (mixtures and masks), we observed shifts in single grating tuning with contrast (see chapter 2) which the untuned model cannot capture due to a fixed tuning shape. For the dispersion- and mixture-focused experiments, we found broader tuning to compound stimuli - particularly at high contrast - which were also unaccounted for by the untuned normalization. The masking experiments provide an even more direct test of tuned gain control: here, we found that the measured suppression and the underlying tuning of the model's normalization were in strong agreement throughout. Across our population of V1 cells, there was a modest but consistent trend for the normalization signal to be stronger for frequencies below the cell's preference. The tuned and contrast-dependent action of the model normalization also offers an answer to an earlier posed question on why contrast-dependent changes in spatial frequency are asymmetric, with greater change at the high spatial frequency limb [126]. Finally, we presented a version of the model which attempts to encapsulate hallmarks of pre-cortical filtering for spatial frequency and contrast. Though in its current form, this model typically performs worse than its V1-only equivalent, we note the potential for improved versions of this model to afford a detailed, quantitative exploration of how pre- and intra-cortical computation shape tuning together.

4.4.1 Comparison with other models

In an earlier chapter of the thesis (ch. 2), we fit a mechanistic difference of DoG (d-DoG-S) model to single grating spatial frequency tuning. We simplified the model by imposing constraints on how center size (and therefore high-frequency selectivity) changed as a function of contrast and jointly optimized to keep some parameters fixed throughout. This modeling providing a set of receptive field structure snapshots at different contrasts. To that end, the d-DoG-S model proved very informative for understanding how receptive fields change as a function of contrast, but is less practical for simulating or modeling responses to arbitrary stimuli. For one, since not all parameters were constrained to change in a prescribed manner with contrast, only contrasts that were presented and fit in our stimulus set could be simulated. Similarly, one would have to choose which receptive field snapshot to use for a mixture comprised of different grating contrasts. Finally, even the simplest model considered in the d-DoG-S analysis had $2 * n_{con} + 3$ parameters for a stimulus set with n_{con} contrast values. The models described here require far fewer parameters (5-7 without the LGN front-end) and can process a greater variety of stimuli because contrast is explicitly handled, rather than implicitly through changing RF structure.

We can relate our model and findings to several other computational, filter-level models of sensory tuning and gain control. The most directly related work is a spatial-frequency focused investigation of the standard Heeger normalization model [55]. Via variations in the spatial frequency tuning of normalization and the response nonlinearity, Nestares and Heeger [13] found that the two model components can trade-off for *some*, but not all, response features [99] of a masking experiment. Our model differs in the implementation of the normalization pool (they use only 3, wider filters) and selection of parameters (they simulated from chosen parameters). However, constant tuning bandwidth across contrast argues, in their model, for a suppression which is broad (i.e., mostly untuned) rather than frequency-specific near the flanks. Our results suggest something in-between: we know from our own and other measurements

[126] that there are contrast-dependent tuning changes in V1, and our modeling shows these cannot be accounted for with an untuned gain control. Furthermore, we find that the tuning of gain control - as well as direct measurements of suppression - is broad and primarily at low spatial frequencies rather than at the cell's tuning flanks. Though we report results here only with a fixed nonlinearity (exponent of 2), we did make versions of the model with a free, optimized nonlinearity and found no qualitative difference in the optimal parameters.

Next, we consider two models whose formulations differ from our own, yet share important computations with the work presented here. By measuring the statistical structure of inputs to different sensory domains, Schwartz and Simoncelli [129] uncover non-uniform normalization and highlight the ability of such a model to capture a variety of physiological response properties, including contrast gain control from masking. In particular, they show an analogous action of tuned, divisive gain control for auditory frequency as we show here for spatial frequency (*fig.* 4.1) - at low stimulus intensity, the (flat) constant term dominates the normalization, while higher stimulus intensities recruit the tuned, stimulus-dependent normalization signals, resulting in changes in selectivity with intensity. Like our two-stage model here, Chariker et al. [29] attempt to understand cortical processing by modeling LGN inputs. Despite their richer, more network- and population-based focus, their model also demonstrates how inhibitory cortical mechanisms can shape broad LGN inputs into more selective V1 tuning.

4.4.2 Model limitations

The model fits - particularly with the inclusion of tuned gain control - capture the variety of responses we found across cells and stimuli. As evidenced by the success of even an untuned gain control, however, the efficacy of normalization models in capturing stimulus selectivity is not new. Instead, we used selected model variations to understand what computations are missing from the standard framework. While the tuned gain control did improve fits by capturing properties (e.g., shifts in tuning with contrast, tuned suppression, broader tuning

to mixtures) unaccounted for by untuned normalization, there were still some model failures. As evidenced by our example cells, these were typically quantitative, not qualitative failures. More specifically, the tuned model matched the contours of suppression's tuning, but not always sharply enough (e.g., *fig.* 4.6); or captured the change in contrast gain with frequency, but not as steeply as required (e.g., *fig.* 4.5C). Beneath, we discuss some potential avenues of improvement for our model and reference some relevant, previous modeling efforts.

In most cases, the improvements involve increasing the model complexity (i.e., number of parameters or filters). Given the small number of model parameters relative to stimulus conditions and the minimal change between training and test loss in our cross-validation exercise, we know that the model is not in a regime of overfitting. The experiments conducted in this work should easily be able to support more nuanced, parameter-rich models.

Excitatory filter

We chose a very simple parameterization (Gaussian derivative) for our filters throughout, including the primary excitatory filter. Though largely effective, previous research suggests that a more detailed representation of the filter would provide better model fits. For one, mechanistic models like the d-DoG-S [53] capture sub-units which reflect their LGN inputs. Other computational models with more flexible forms also argue for the inclusion of multiple subunits to better capture V1 tuning. Rust et al. [122] find an average of 1 to 3 excitatory subunits for simple cells, and up to 8 for complex cells. While the work in Rust relies on finding subunits with different tuning, another study shows that even self-similar, spatially offset subunits in a divisive normalization framework can greatly improve fit quality [160]. Both studies point towards either more filters or a more richly parameterized excitatory filter.

Suppression, normalization response

As with the excitatory filter, we settled on a simple representation of the normalization pool subunits. Any of these choices (number and width of the filters; weighting function) could be

interrogated to potentially improve the model fits. Additionally, we modelled inhibition as arising only through divisive action despite the physiological [109, 24] and model-based [160, 29] evidence that inhibition also involves additive/subtractive mechanisms. Rather than only division, we could first subtract a scaled version of the normalization pool response from the excitatory filter. This might help with cases like like the mask-only tuning example shown in *fig.* 4.6*B1*: there, the suppression tuning is very well matched, but so strong that at high contrast the tuning curve 'dips' rather than capturing the more gentle tuning on the low frequency limb of the tuning curve.

Our choice of normalization pool tuning breadth handicaps the ability of our standard normalization model to fit the data. As shown in *fig.* 4.1C, we chose a pool of filters which is responsive beyond 20 c/deg. We constructed the pool this way to ensure that all stimuli were seen by the normalization pool (e.g. our dispersion-focused experiment includes stimuli with spatial frequencies at and above 10 c/deg). However given the spatial clustering of spatial frequency preference [98] and the closer-range connections which are postulated to contribute to the normalization signal, it is unlikely that filters with such tuning contribute to gain control for cells with low spatial frequency preference. Thus, our normalization signal is artificially wider than is physiologically plausible, particularly at high spatial frequencies; however, the consequence of such a mismatch is at partially minimized because the normalization signal is irrelevant for stimuli outside a cell's passband. Nonetheless, the wide normalization pool tuning can yield relatively stronger normalization at the high frequency flank of a cell's selectivity. This handicapping of the untuned normalization model is supported by the improvements yielded by the matched model (*fig.* 4.8) that shapes the normalization signal to more closely match a cell's spatial frequency tuning.

Spatial and temporal considerations

We briefly note that our model of V1 does not explicitly consider space. There is a large literature discussing the spatial extents of different suppressive mechanisms in cortex (e.g.

see [8] for a review). In particular, several studies have converged on the notion that cortical-origin, tuned suppression is more spatially extensive than feedforward untuned suppression [168, 59]. Our stimuli were on average 2.5 times larger than the receptive fields, and they therefore recruited this tuned, extraclassical surround. If we explicitly accounted for the relationship between receptive field size and stimulus size, our model would reflect this recruitment and should therefore produce stronger suppressive effects than have been shown here. Likewise, we model the temporal dynamics of the stimuli but not any of the underlying model components. As discussed in an earlier chapter (§3.4.2), suppression is weak early in the response (e.g. first 100ms) and only later on becomes stronger and apparent at a wider range of spatial frequencies. Our model mechanisms (namely the normalization signal) have no adjustment with time; thus, our model fits are slightly hampered by needing to account for the time-averaged influence of suppression. Nonetheless, we imagine the impact of such dynamics are likely small given that the bulk of the response is dominated by the steady-state period with stronger suppression.

Cascaded models (LGN front-end)

The overarching goal of our modeling effort is to understand the computations which give rise to the variety of spatial frequency selectivity observed in cortex. The V1-only version of the model ascribes all shaping - both linear and nonlinear, excitatory and suppressive - to cortical mechanisms. However, we know from a rich literature that there are strong non-linearities and stimulus filtering in the retina and the LGN. To better understand how pre-cortical processing shapes V1 selectivity, we introduced a pre-cortical filtering stage which captures the key properties of magno and parvo neurons to spatial frequency and contrast. This cascaded model better captures the actual flow of information through the visual hierarchy, and in a sense, reduces the burden of complexity placed upon V1 computation. More plainly, it is possible that the ultimate shaping of V1 spatial frequency tuning by cortical mechanisms is more nuanced than our results here suggest, since some of those computations might occur

before cortex. This is in line with a series of studies on the suppressive mechanisms which shape V1 tuning (cf. [21, 59]).

In its current form, however, our LGN model fails to improve upon the V1-only stage, and is thus limited in its diagnostic power. One potential culprit for at least part of this failure can be surmised from the following: for one, the distribution of relative magno/parvo weighting was non-uniform and skewed towards mostly parvo input; second, models with stronger magno input to V1 yielded larger improvements compared to models without the LGN filtering. Given the correlation between lower contrast thresholds and stronger model magno input, we suspect that for many cells, the pre-cortical non-linearities were a hindrance to be overcome, not an aid in capturing nonlinear response properties unaccounted for by V1 normalization. For these cells, models landed on strong/exclusive parvocellular input in order to choose the filter with a more linear contrast response function. We hypothesize that were the LGN filters more flexible, many cells would be better modeled with more linear contrast response functions. We cannot discern, however, whether this would reflects the sampling of relatively more linear inputs by certain V1 cells or an additional transformation of thalamic afferents not represented in our model.

Chapter 5

Conclusions

The primary discussion of our results appears at the end of each chapter. However, we also summarize the findings here and discuss issues which are outside of the scope of the individual sections. In the three chapters of this thesis, we addressed related aspects of spatial frequency tuning and its underlying computations in the early visual system of macaque monkeys. We employed a combination of electrophysiology and modeling - both mechanistic and computational - to address important gaps in the field's knowledge of how contrast influences spatial frequency tuning; the tuning and form of the computations which underlie that influence; and how tuning changes and suppression differ in the LGN and V1. In our careful construction of stimulus sets which widely sampled contrast, spatial frequency, and spectral spread, we were able to fit two forms of models which each addressed a different aspect of how spatial tuning varies with contrast. In successive chapters, our primary aims were to compare the magnitude of contrast-dependent spatial frequency tuning in LGN and V1 neurons; measure the strength and tuning of spatial frequency suppression; and account for our findings with a tuned gain control model.

Our first set of results, presented in chapter 2, concerns the contrast-dependence of spatial frequency tuning. Though previous studies have characterized the increase in receptive field size [127, 102, 74] and decrease in high spatial frequency selectivity [146, 126] with reduced

contrast, no direct comparisons between these tuning shifts in LGN and V1 have been made. Furthermore, motivated by our interest in expressing these shifts in terms of gain control, we sample contrast more widely than many previous studies. We find that neurons in both areas show a contrast-dependent reduction in tuning bandwidth and high frequency selectivity, but that the effects are more pronounced in cortex. Within each area, we find no difference between magno and parvo cells, but do note that complex cells show stronger shifts than simple cells. A preliminary analysis by layer also finds that supragranular layers have the strongest contrast-dependencies.

The second set of results builds upon our clarification of contrast-dependent spatial frequency tuning by investigating suppression and summation. To bolster our argument for stronger contrast-dependent tuning in V1, we record LGN and V1 activity using a common set of grating mixtures. In both areas, summation ranges in its nonlinearity: more nonlinear for magnos and complex cells than parvos and simple cells, respectively. We further analyzed summation nonlinearity by measuring its frequency dependence. We find that V1 cells - in particular complex cells - show a greater variation in summation linearity with spatial frequency compared to LGN cells, and that summation was more linear for frequencies near or above the cell's peak. To attain a more fine-grained characterization of contrast-spatial frequency interactions in V1, we used a masking stimulus paradigm. Through this experiment, we replicate earlier results which suggest a low-frequency suppression - strongest below the cell's peak tuning - which grows with contrast. By considering these findings with our reported single grating tuning changes, we assemble a more clear picture: spatial frequency tuning in both LGN and V1 exhibits contrast dependence, but only in V1 do those dependencies reflect the action of a strongly tuned suppression.

Finally, we unified the findings of the previous chapters by fitting a model of tuned gain control to the recorded responses of each V1 cell. We used cross-validation to compare models of tuned and untuned normalization, finding that the former performs better in the vast majority of cells. This model is able to capture the array of response properties outlined in

our experiments: shifts in spatial frequency tuning with contrast, more linear summation to mixtures with high spatial frequency content, and low-frequency tuned, contrast-dependent suppression with masking stimuli. Crucially, this model achieves these results with a fixed linear filter and pool of normalization filters; all that's needed is for the gain control signal to be tuned. In line with the tuning properties that we and others have demonstrated, our model typically optimized for a stronger weighting of low spatial frequencies in the normalization signal. To better quantify the influence of LGN vs. V1 mechanisms on spatial frequency tuning, we also present a cascaded gain control model which encapsulates the spatial frequency and contrast selectivity of magno and parvo LGN cells. Though the current form fails to improve upon the V1-only model, we note that its current parameterization is very simple; in future work, we will carefully refine the LGN stage to attain better performance. In doing so, this LGN-V1 model would permit the distinct identification and simulation of the thalamic and cortical mechanisms which shape contrast-dependent spatial frequency tuning.

5.1 Relationship to psychophysics

The goal of vision is to support perception and perception-enabled action. Left undiscussed to this point is how the background and results of this thesis relate to previous psychophysical studies. However, a large body of literature provides interesting context to our findings, which we briefly outline here.

Fourier analysis of an image results in an amplitude and phase associated with each spatial frequency. Per this analysis, natural images contain a wide range of spatial frequencies, typically following a $1/f$ distribution [47, 121, 128]. The several octave spread of spectral content in images, then, poses an interesting question as to what drives a viewer's percept. Many studies have shown that an image is perceived largely on the basis of its phase [103, 108, 155]. However, despite the outsize role of phase, "the contribution of the amplitude spectrum

towards the appearance of an image cannot be regarded as irrelevant." [153]. For example, spatial frequency content is used by observers to discriminate textures [82] and letters [86].

Given the role that spatial frequency perception plays in natural vision, we then ask how contrast-dependent spatial frequency tuning might influence perception. Earlier psychophysical studies shed light on this question. The work of Georgeson [50, 51] shows that contrast influences spatial frequency judgements. In a matching experiment, subjects reliably perceived lower contrast gratings as higher spatial frequency than veridical; additionally, this misjudgement varied parametrically with contrast, rather than appearing only at the very lowest contrasts [50]. Most relevant to our work is his hypothesis that if perception works in a labeled-line manner (e.g., [90]), one should expect to observe preferred higher spatial frequencies at higher contrast [50]. In this thesis (chapter 2, as in [126]) we find a compatible result: the high spatial frequency cutoff of neurons in the early visual system increases with contrast.

While the agreement between psychophysics and physiology is exciting, the question remains whether these effects are relevant to vision outside the laboratory. For one, both effects are fairly modest: psychophysically, a 10% shift over a 3-4 octave reduction in contrast [51] (though see the larger shift reported in [50]); in neurons (e.g., complex cells), an average shift of 13% over a 4+ octave range. However, we argue that rather than looking only for direct psychophysical evidence of contrast-dependent shifts in spatial frequency perception, we can also survey the literature for signs of tuned gain control. Using natural images rather than laboratory stimuli reveals a strong adaptation effect selective for low spatial frequencies [164] (and adaptation is believed to reflect gain control [138]). A study which further explored these influences argued that this biased adaptation reflects a low-frequency tuned gain control which seeks to preserve sensitivity to all frequencies across contrast [16]. This is in agreement with our findings (and those previously reported) which show stronger suppression and gain control for low spatial frequencies. The preservation of frequency sensitivity across contrast -

despite the observed tuning shifts in V1 neurons - might also play a role in natural vision (e.g., to preserve texture discrimination [82] across a wider range of contrast.)

Georgeson's work showed that a wide range of experimental changes caused an increase in apparent spatial frequency, leading to his conclusion that: "...any other manipulation that takes the stimulus closer to threshold also increases its apparent spatial frequency." [50] This result suggests that spatial frequency tuning can be influenced by changes in grating orientation and presentation time, too, among other features. Orientation and spatial frequency tuning have been shown to be separable [91], though these stimuli were only presented at high contrast. Perceptually, subjects perform worse at detection and discrimination for gratings at oblique angles [9, 22]. Though convincing proof of the expected underlying physiology had long eluded scientists, a recent imaging study demonstrates stronger population responses at cardinal orientation [44]. The Georgeson hypothesis, then, would suggest that perhaps at lower stimulus contrast, the separability of orientation and spatial frequency tuning becomes less clear, at least at a population level; that is, the contrast-dependent shifts we and others have observed might be more pronounced at oblique orientations. We know of no studies which study spatial frequency tuning as a function of stimulus presentation. However, investigations into the dynamics of tuning have shown that spatial frequency tuning is tuned to lower frequencies earlier on in the response, and shifts towards higher spatial frequencies with time [21, 91]. We hypothesize, then, that a temporal masking experiment with variations in presentation time might reveal stronger contrast-dependent tuning shifts for shorter presentation times.

In the context of these perceptual results, there are alternatives to the labeled line hypothesis discussed above. Rather than focusing only on single neuron interpretations, we also note the potential for population-based decoding models (e.g. [84, 123]) to explain the alignment between our physiological results and those in prior perceptual studies. Of particular relevance to our work is how changes in tuning, in part brought about by gain modulation, might alter population decoding. One key study [132] found that maximum likelihood estimation based on fixed encoders - rather than assuming the downstream decoder

is aware of gain changes - yields predictions that agree with prior experimental results. Using a Bayesian population decoder, Stocker and Simoncelli [152] - among others, including in different sensory domains [73] - show that for stimuli with greater uncertainty, perception more closely matches a prior based on the distribution of stimuli in the natural world. For spatial frequency perception as a function of contrast, that prior would be for low spatial frequencies [47, 121, 128], which also aligns with our findings.

Appendix A

Adjusting simple cell responses

A.1 Why model the phase-amplitude relationship?

Neurons in the lateral geniculate nucleus (LGN) of macaque monkeys are sensitive to the stimulus phase of sinusoidal gratings presented within their receptive field [70]. Simple cells in cortex also exhibit this property. Despite this linear response property, these neurons show a hallmark of contrast gain control, which is a non-linear computation. In particular, the response phase for LGN neurons systematically advances with contrast [134]. More specifically, this relationship can be modelled as the change in response phase with response amplitude; this is discussed and modelled in [25]. Furthermore, this property varies in effect size between magno and parvo cells [40]. Regardless, we can use this property to correct F1 responses for noise by correcting for the difference in the expected and observed response phase.

In one set of our series of spatial frequency mixture (*sfMix*) experiments, we present sinusoidal gratings - both in isolation and in superimposed mixtures - at a variety of contrasts, spatial frequencies, and temporal frequencies. For a given spatial and temporal frequency combination, we have stimulus repetitions at a range of contrasts. Over each repetition, the response amplitude and phase will be subject to variation, i.e., noise. By explicitly modeling the relationship between the response phase and amplitude (the latter of which is a function

of stimulus contrast), we can determine how much of the response is not stimulus driven (i.e., noise), and correct the responses accordingly.

A.1.1 Stimuli, notation

We used circular, sinusoidal drifting gratings in this experiment. All stimuli - both those presented in isolation and in superimposed mixture - have stimulus properties drawn from a set of contrasts $\{c_i\}$, spatial frequencies $\{f_i\}$, and temporal frequencies $\{\omega_i\}$. Before running any *sfMix* experiment, we first characterize the basic response properties (e.g., contrast response; spatial/temoral frequency, orientation, and size tuning) of the cell under consideration. For the *sfMix* experiment, the orientation is set to be the optimal for the cell (or vertical if untuned, e.g., LGN neurons), as is the stimulus size. The phase of each stimulus component is randomized for each presentation.

Bolded quantities, like $\mathbf{r}$, are vectors. I will use i as an index into vectors or sets, such as ω_i referring to the temporal frequency of a given stimulus component. Quantities about the response are given a "hat", e.g., to distinguish $\hat{\phi}_i$ - the response phase for stimulus component i - from the stimulus phase of that same component ϕ_i. I also use the subscript "m" (e.g., R_m) to refer to modelled quantities.

A.2 Method

A.2.1 Phase-amplitude relationship

Given a vector $\mathbf{r}$ of spike times in a given trial, we first create a peri-stimulus time histogram *fig.* A.1 by binning the spikes into non-overlapping bins of width t_{bin} (e.g., 1 ms). Then, we

compute the Fourier transform $\hat{f}(\mathbf{r})$ to get

$$\mathbf{A}(\mathbf{r}) = |\hat{f}(\mathbf{r})| \tag{A.1}$$

$$\phi(\mathbf{r}) = \arg(\hat{f}(\mathbf{r})) \tag{A.2}$$

Here, (A.1) is a vector of response amplitudes at each integer frequency, and (A.2) is the same, but for the phase of the response. Then, we want determine the response amplitude

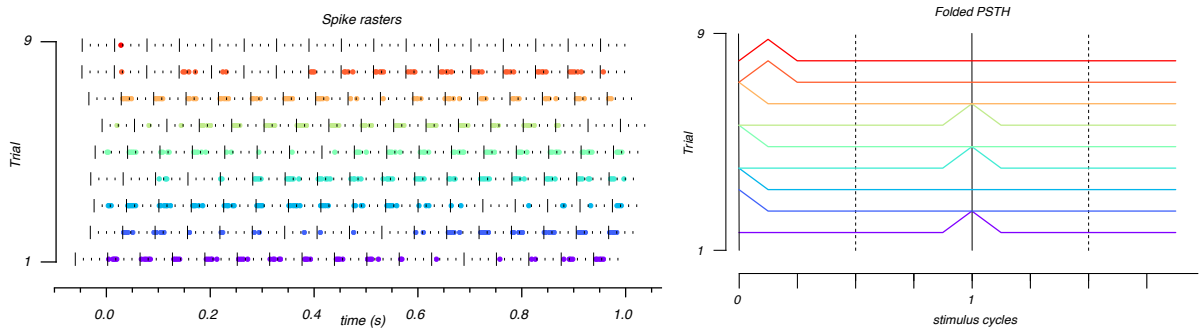

Figure A.1: (*Left*) Spike rasters for an example LGN parvo cell. Each row is one repetition of a stimulus presentation identical in every way except for the starting phase. The long bars in each trial indicate when the stimulus phase is 0; each subsequent tick indicates an additional 90° advance. (*Right*) Peri-stimulus time histogram for the same trials. Here, we fold the stimulus into 2 cycles with bins of 1 ms. The spike rates are normalized within each trial.

and phase (in degrees) *relative* to the stimulus. The analysis that follows applies only to gratings presented in isolation, and thus the stimulus has its own (single) temporal frequency ω_s and start phase ϕ_s.

$$\hat{R} = \mathbf{A}(\mathbf{r})\Big|_{\omega_s} \tag{A.3}$$

$$\hat{\phi} = \phi(\mathbf{r})\Big|_{\omega_s} - \phi_s \pmod{360} \tag{A.4}$$

The amplitude $\hat{R}$ is considered the F1 - as opposed to the DC (or F0) - response since we are evaluating the power at the stimulus temporal frequency. To visualize the response (*fig.* A.2), we can use a 2D polar plot with the coordinates $(\hat{R}, \hat{\phi})$. Then, we model the effect of

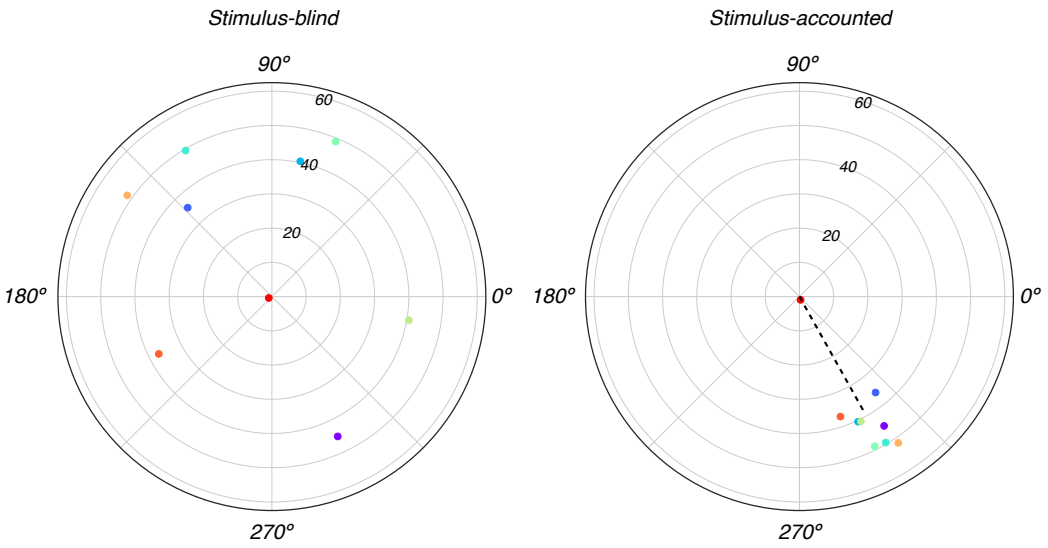

Figure A.2: For the same data as above, polar plots showing the amplitude R (distance from the origin) and phase θ (counter-clockwise angle) of the response in multiple trials before and after accounting for the stimulus phase. These values are obtained by evaluating the Fourier transform of the neuron's response at the stimulus temporal frequency. The length and angle of the dashed line in the right plot indicate the mean amplitude and phase of these 10 responses, respectively.

response amplitude on response phase [95] as

$$\phi_m(\hat{R}; \phi_0, p) = \phi_0 + p \cdot \hat{R}(c) \pmod{360} \tag{A.5}$$

Implicitly, $\hat{R}$ is a function of stimulus contrast c, and we seek to minimize the mean squared error between measured and modeled phase for a series of responses $\{\hat{\mathbf{R}}(c_i)\}$ for a set of contrasts $\{c_i\}$ by optimizing for ϕ_0, the phase at a response amplitude of 0, and p, the slope which determines how quickly phase changes with response amplitude. The measured response $\hat{R}(c)$ for a given contrast c_i is averaged over all trials of that stimulus condition, since our model makes a prediction for each stimulus, not each trial.

A.2.2 Adjusting the responses

With a model for the expected response phase as a function of response amplitude, we can then correct the measured response to reflect only stimulus-driven activity. This is computed

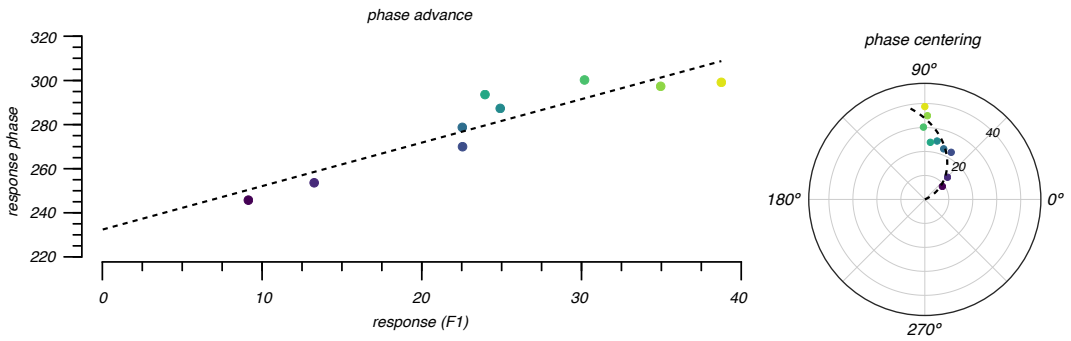

Figure A.3: The data in *fig.* A.1 and *fig.* A.2 - from one stimulus - are the point of maximum response here (yellow). (*Left*) Response phase plotted as a function of F1 amplitude; the dashed line is the fit for (A.5). (*Right*) The same data and model plotted in a 2D polar phase, with the phase values re-centered such that the phase at the highest amplitude response is at 90°.

simply as

$$R^* = \hat{R} \cdot \cos(\hat{\phi} - \phi_m) \tag{A.6}$$

the cosine of the angle difference between the phase of the response relative to the stimulus and the phase as predicted by the model (A.5). This ensures that $R^* \leq \hat{R}$, i.e., the adjusted response can only reduce in magnitude due to deviations from the expected phase ϕ_m. These adjusted responses can be computed on a trial-by-trial basis or after averaging the responses from all trials for a given condition. All analysis and model fitting with these data are based on the responses as adjusted by this process.

A.2.3 Example: Response-versus-contrast

As an example, we can fit a model of response versus contrast [43] to these adjusted responses $\{R^*\}$,

$$R_m(c; b, k, c_0) = b + k \cdot \log(1 + \frac{c}{c_0}) \tag{A.7}$$

$$L(\{\mathbf{R}^*\}; b, k, c_0) = \sum_i \frac{1}{\mathrm{Var}[\mathbf{R}_\mathbf{i}^*]} \cdot (\mathrm{Mean}[\mathbf{R}_\mathbf{i}^*] - R_m(c_i; b, k, c_0))^2 \tag{A.8}$$

by optimizing for the baseline response b, a scaling constant k and the logarithmic saturation contrast c_o. As in [95], we restrict $b \leq 0$, and almost always $b = 0$ since a negative b is needed only for cells with little-to-no spontaneous activity, which is rare in LGN neurons [149, 12] though more common in cortex [12].

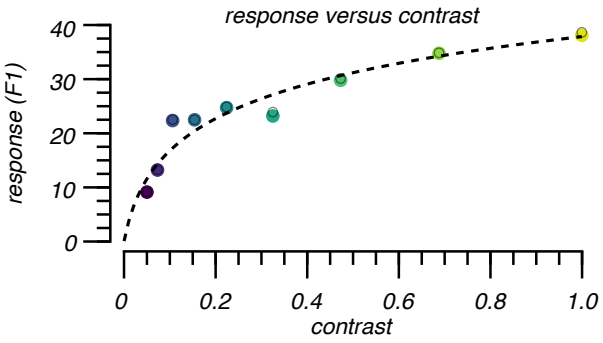

Figure A.4: Original (open circle) and adjusted responses (filled circle) with an optimized response-versus-contrast model curve (dashed line). The responses of this cell, the same as in the above figures, were not adjusted substantially, as evidenced by the similarity between the originally measured and adjusted data points.

The data in *fig.* A.4 are from a parvocellular neuron. An additional example, a magno cell, is presented in *fig.* A.5, while an example of a noisier parvo cell recording is presented in *fig.* A.6.

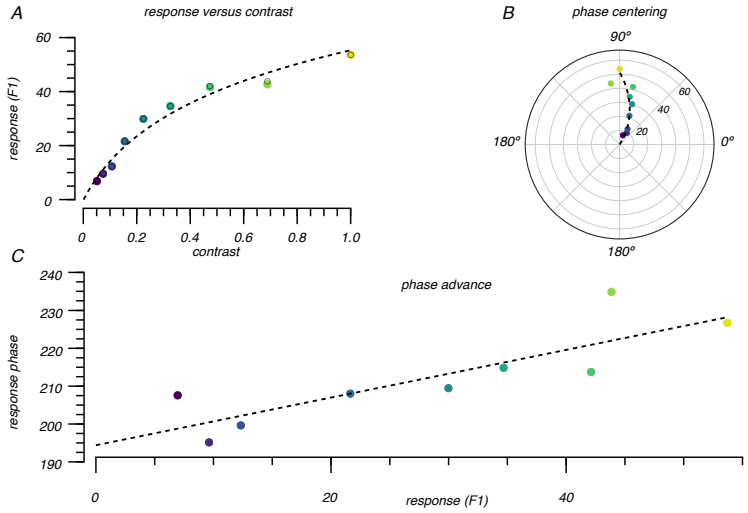

Figure A.5: (*A*) As in *fig.* A.4; again, note the minimal change in response amplitude. (*B, C*) As in *fig.* A.3; the minor deviations from a linear progression in phase with response amplitude explain the small adjustments in the response-versus-contrast plot.

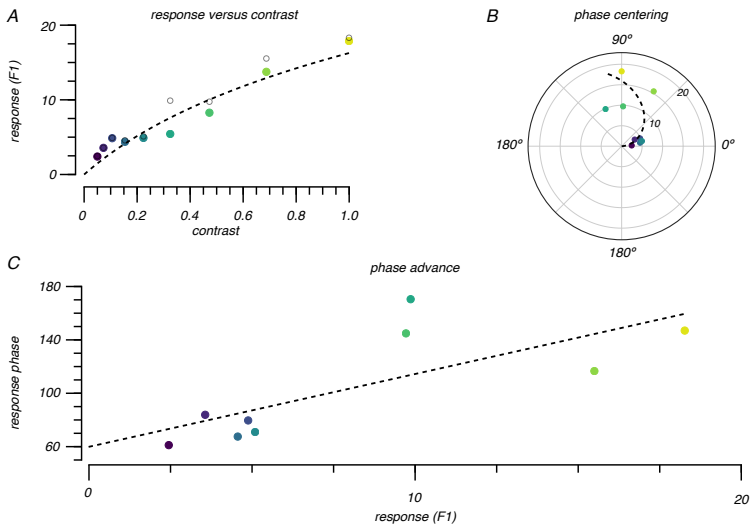

Figure A.6: Note the steeper slope and large deviations from the model in (C). This results in larger response adjustments in (A).

A.2.4 Adjusting mixture stimuli responses

To adjust the responses to stimuli comprised of multiple gratings, we can use the results of the single grating analyses described above. Throughout the experiment, each spatial frequency f_i appears only at one temporal frequency, which is unique to that f_i. Thus, each component of a mixture stimulus has a distinct temporal frequency ω_i, enabling us to identify the response attributable to that stimulus component. For a given mixture stimulus with spike times $\mathbf{r}$, we evaluate (A.1), (A.2) at the temporal frequencies $\{\omega_i\}$. Since each component presented in a mixture is also presented in isolation during other trials, we can then use the phase-amplitude relationship optimized in (A.5) to determine the expected phase ϕ_m given the measured (unadjusted) response amplitudes $\{R_i\}$. We then adjust the response of each component as in (A.6). To get the overall stimulus response, we simply take the sum of all adjusted responses.

Succinctly,

$$\begin{pmatrix} \hat{R}_i \\ \\ \hat{\phi}_i \end{pmatrix} = \begin{pmatrix} \mathbf{A}(\mathbf{r})\Big|_{\omega_i} \\ \\ \phi(\mathbf{r})\Big|_{\omega_i} - \phi_i \pmod{360} \end{pmatrix} , \forall i \tag{A.9}$$

$$R_i^* = \hat{R}_i \cdot \cos(\hat{\phi}_i - \phi_{m_i}) , \forall i \tag{A.10}$$

$$R^* = \sum_i R_i^* \tag{A.11}$$

These adjustments can be done on a trial-by-trial basis, or on the average amplitude and phase for all trials in a condition. All analyses, then, can be done with R^* or R_i^* if considering the component responses.

References

[1] Adelson, E. H. and Bergen, J. R. “Spatiotemporal energy models for the perception of motion”. In: *J Opt Soc Am A* 2.2 (1985), pp. 284–99.

[2] Albrecht, D. G., Farrar, S. B., and Hamilton, D. B. “Spatial contrast adaptation characteristics of neurones recorded in the cat’s visual cortex”. In: *J Physiol* 347 (1984), pp. 713–39.

[3] Albrecht, D. G. and Hamilton, D. B. “Striate cortex of monkey and cat: contrast response function”. In: *J Neurophysiol* 48.1 (1982), pp. 217–37.

[4] Alitto, H. J. et al. “A comparison of visual responses in the lateral geniculate nucleus of alert and anaesthetized macaque monkeys”. In: *Journal of Physiology* 589.1 (2011), pp. 87–99.

[5] Alonso, J. M., Usrey, W. M., and Reid, R. C. “Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex”. In: *J Neurosci* 21.11 (2001), pp. 4002–15.

[6] Angelucci, A. and Bressloff, P. C. “Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons”. In: *Prog Brain Res* 154 (2006), pp. 93–120.

[7] Angelucci, A. and Sainsbury, K. “Contribution of feedforward thalamic afferents and corticogeniculate feedback to the spatial summation area of macaque V1 and LGN”. In: *J Comp Neurol* 498.3 (2006), pp. 330–51.

[8] Angelucci, Alessandra et al. “Circuits and mechanisms for surround modulation in visual cortex”. In: *Annual review of neuroscience* 40 (2017), pp. 425–451.

[9] Appelle, S. “Perception and discrimination as a function of stimulus orientation: the "oblique effect" in man and animals”. In: *Psychol Bull* 78.4 (1972), pp. 266–78.

[10] Attwell, D. and Wilson, M. “The spatial frequency sensitivity of bipolar cells”. In: *Biol Cybern* 47.2 (1983), pp. 131–40.

[11] Baccus, S. A. and Meister, M. “Fast and slow contrast adaptation in retinal circuitry”. In: *Neuron* 36.5 (2002), pp. 909–19.

[12] Bair, W. et al. “The timing of response onset and offset in macaque visual neurons”. In: *J Neurosci* 22.8 (2002), pp. 3189–205.

[13] Bauman, L. A. and Bonds, A. B. “Inhibitory refinement of spatial frequency selectivity in single cells of the cat striate cortex”. In: *Vision Research* 31.6 (1991), pp. 933–944.

[14] Benardete, E. A. and Kaplan, E. “The receptive field of the primate P retinal ganglion cell, I: Linear dynamics”. In: *Visual Neuroscience* 14 (1997), pp. 169–185.

[15] Benardete, E. A. and Kaplan, E. “The dynamics of primate M retinal ganglion cells”. In: *Visual Neuroscience* 16 (1999), pp. 355–368.

[16] Bex, P. J., Solomon, S. G., and Dakin, S. C. “Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure”. In: *J Vis* 9.10 (2009), pp. 1 1–19.

[17] Bonds, A. B. “Temporal dynamics of contrast gain in single cells of the cat striate cortex”. In: *Vis Neurosci* 6.3 (1991), pp. 239–55.

[18] Bonds, AB. “Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex”. In: *Visual neuroscience* 2.1 (1989), pp. 41–55.

[19] Bonin, V., Mante, V., and Carandini, M. “The suppressive field of neurons in lateral geniculate nucleus”. In: *J Neurosci* 25.47 (2005), pp. 10844–56.

[20] Bradley, A. et al. “Visual Orientation and Spatial Frequency Discrimination: A Comparison of Single Neurons and Behavior”. In: *Journal of Neurophysiology* 57.3 (1987), pp. 755–772.

[21] Bredfeldt, Christine E and Ringach, DL. “Dynamics of spatial frequency tuning in macaque V1”. In: *Journal of Neuroscience* 22.5 (2002), pp. 1976–1984.

[22] Caelli, T. et al. “Discrimination thresholds in the two-dimensional spatial frequency domain”. In: *Vision Research* 23.2 (1983), pp. 129–133.

[23] Callaway, E. M. “Structure and function of parallel pathways in the primate early visual system”. In: *J Physiol* 566.Pt 1 (2005), pp. 13–9.

[24] Carandini, M. and Heeger, D. J. “Summation and division by neurons in primate visual cortex”. In: *Science* 264.5163 (1994), pp. 1333–6.

[25] Carandini, M., Heeger, D. J., and Movshon, J. A. “Linearity and normalization in simple cells of the macaque primary visual cortex”. In: *J Neurosci* 17.21 (1997), pp. 8621–44.

[26] Carandini, M. et al. “Do we know what the early visual system does?” In: *J Neurosci* 25.46 (2005), pp. 10577–97.

[27] Cavanaugh, J. R., Bair, W., and Movshon, J. A. “Nature and Interaction of Signals From the Receptive Field Center and Surround in Macaque V1 Neurons”. In: *Journal of Neurophysiology* 88 (2002), pp. 2530–2546.

[28] Cavanaugh, J. R., Bair, W., and Movshon, J. A. “Selectivity and Spatial Distribution of Signals From the Receptive Field Surround in Macaque V1 Neurons”. In: *Journal of Neurophysiology* 88 (2002), pp. 2547–2556.

[29] Chariker, L., Shapley, R., and Young, L. S. “Orientation Selectivity from Very Sparse LGN Inputs in a Comprehensive Model of Macaque V1 Cortex”. In: *J Neurosci* 36.49 (2016), pp. 12368–12384.

[30] Chariker, L., Shapley, R., and Young, L. S. “Contrast response in a comprehensive network model of macaque V1”. In: *J Vis* 20.4 (2020), p. 16.

[31] Chung, J. E. et al. “A Fully Automated Approach to Spike Sorting”. In: *Neuron* 95.6 (2017), 1381–1394 e6.

[32] Croner, L.J. and Kaplan, E. “Receptive Fields of P and M Ganglion Cells Across the Primate Retina”. In: *Vision Research* 35.1 (1994), pp. 7–24.

[33] Dacey, D. M. “Parallel pathways for spectral coding in primate retina”. In: *Annu Rev Neurosci* 23 (2000), pp. 743–75.

[34] De, A. and Horwitz, G. D. “Coding of chromatic spatial contrast by macaque V1 neurons”. In: *Elife* 11 (2022).

[35] De Valois, KAREN K and Tootell, RB. “Spatial-frequency-specific inhibition in cat striate cortex cells”. In: *The Journal of Physiology* 336.1 (1983), pp. 359–376.

[36] De Valois, R. L., Albrecht, D. G., and Thorell, L. G. “Spatial frequency selectivity of cells in macaque visual cortex”. In: *Vision Research* 22 (1982), pp. 545–559.

[37] DeAngelis, G. C. et al. “Organization of suppression in receptive fields of neurons in cat visual cortex”. In: *J Neurophysiol* 68.1 (1992), pp. 144–63.

[38] Derrington, A. M. and Fuchs, A. F. “Spatial and temporal properties of X and Y cells in the cat lateral geniculate nucleus”. In: *J Physiol* 293 (1979), pp. 347–64.

[39] Derrington, A. M., Krauskopf, J., and Lennie, P. “Chromatic mechanisms in lateral geniculate nucleus of macaque”. In: *J Physiol* 357 (1984), pp. 241–65.

[40] Derrington, A. M. and Lennie, P. “Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque”. In: *Journal of Physiology* 357 (1984), pp. 219–240.

[41] Dhruv, N. T. and Carandini, M. “Cascaded effects of spatial adaptation in the early visual system”. In: *Neuron* 81.3 (2014), pp. 529–35.

[42] Duong, T. and Freeman, R. D. “Spatial frequency-specific contrast adaptation originates in the primary visual cortex”. In: *J Neurophysiol* 98.1 (2007), pp. 187–95.

[43] Enroth-Cugell, C. and Robson, J. G. “The contrast sensitivity of retinal ganglion cells of the cat”. In: *J Physiol* 187.3 (1966), pp. 517–52.

[44] Fang, C., Cai, X., and Lu, H. D. “Orientation anisotropies in macaque visual areas”. In: *Proc Natl Acad Sci U S A* 119.15 (2022), e2113407119.

[45] Felleman, D. J. and Van Essen, D. C. “Distributed hierarchical processing in the primate cerebral cortex”. In: *Cereb Cortex* 1.1 (1991), pp. 1–47.

[46] Ferster, D., Chung, S., and Wheat, H. “Orientation selectivity of thalamic input to simple cells of cat visual cortex”. In: *Nature* 380 (1996), pp. 249–252.

[47] Field, D. J. “Relations between the statistics of natural images and the response properties of cortical cells”. In: *J Opt Soc Am A* 4.12 (1987), pp. 2379–94.

[48] Foster, K. H. et al. “Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey”. In: *J Physiol* 365 (1985), pp. 331–63.

[49] Frazor, R. A. and Geisler, W. S. “Local luminance and contrast in natural images”. In: *Vision Res* 46.10 (2006), pp. 1585–98.

[50] Georgeson, M. A. “The perceived spatial frequency, contrast, and orientation of illusory gratings”. In: *Perception* 9 (1980), pp. 695–712.

[51] Georgeson, M. A. “Apparent spatial frequency and contrast of gratings: separate effects of contrast and duration”. In: *Vision Res* 25.11 (1985), pp. 1721–7.

[52] Goris, R. L., Simoncelli, E. P., and Movshon, J. A. “Origin and Function of Tuning Diversity in Macaque Visual Cortex”. In: *Neuron* 88.4 (2015), pp. 819–31.

[53] Hawken, M. J. and Parker, A. J. “Spatial properties of neurons in the monkey striate cortex”. In: *Proc R Soc Lond B Biol Sci* 231.1263 (1987), pp. 251–88.

[54] Hawken, M. J., Parker, A. J., and Lund, J. S. “Laminar organization and contrast sensitivity of direction-selective cells in the striate cortex of the Old World monkey”. In: *J Neurosci* 8.10 (1988), pp. 3541–8.

[55] Heeger, D. J. “Normalization of cell responses in cat striate cortex”. In: *Vis Neurosci* 9.2 (1992), pp. 181–97.

[56] Heeger, D. J. and Mackey, W. E. “Oscillatory recurrent gated neural integrator circuits (ORGaNICs), a unifying theoretical framework for neural dynamics”. In: *Proc Natl Acad Sci U S A* 116.45 (2019), pp. 22783–22794.

[57] Heeger, D. J. and Zemlianova, K. O. “A recurrent circuit implements normalization, simulating the dynamics of V1 activity”. In: *Proc Natl Acad Sci U S A* 117.36 (2020), pp. 22494–22505.

[58] Henry, C. A. et al. “Functional characterization of the extraclassical receptive field in macaque V1: contrast, orientation, and temporal dynamics”. In: *J Neurosci* 33.14 (2013), pp. 6230–42.

[59] Henry, C. A. et al. “Distinct spatiotemporal mechanisms underlie extra-classical receptive field modulation in macaque V1 microcircuits”. In: *Elife* 9 (2020).

[60] Hirsch, J. A. et al. “Functionally distinct inhibitory neurons at the first stage of visual cortical processing”. In: *Nat Neurosci* 6.12 (2003), pp. 1300–8.

[61] Hochstein, S. and Shapley, R. M. “Linear and nonlinear spatial subunits in Y cat retinal ganglion cells”. In: *J Physiol* 262.2 (1976), pp. 265–84.

[62] Hubel, D. H. and Wiesel, T. N. “Receptive fields of single neurones in the cat’s striate cortex”. In: *The Journal of Physiology* 148.3 (1959), pp. 574–591.

[63] Hubel, D. H. and Wiesel, T. N. “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”. In: *J Physiol* 160.1 (1962), pp. 106–54.

[64] Hubel, D. H. and Wiesel, T. N. “Ferrier Lecture: Functional architecture of macaque monkey visual cortex”. In: *Proc. R. Soc. Lond. B* 198 (1977), pp. 1–59.

[65] Irvin, G. E., Casagrande, V. A., and Norton, T. T. “Center/surround relationships of magnocellular, parvocellular, and koniocellular relay cells in primate lateral geniculate nucleus”. In: *Vis Neurosci* 10.2 (1993), pp. 363–73.

[66] Isaacson, J. S. and Scanziani, M. “How inhibition shapes cortical activity”. In: *Neuron* 72.2 (2011), pp. 231–43.

[67] Jones, H. E. et al. “Surround suppression in primate V1”. In: *J Neurophysiol* 86.4 (2001), pp. 2011–28.

[68] Kapadia, M. K., Westheimer, G., and Gilbert, C. D. “Dynamics of spatial summation in primary visual cortex of alert monkeys”. In: *Proc Natl Acad Sci U S A* 96.21 (1999), pp. 12073–8.

[69] Kaplan, E., Purpura, K., and Shapley, R. M. "Contrast affects the transmission of visual information through the mammalian lateral geniculate nucleus". In: *J Physiol* 391 (1987), pp. 267–88.

[70] Kaplan, E. and Shapley, R. M. "X and Y cells in the lateral geniculate nucleus of macaque monkeys". In: *J Physiol* 330 (1982), pp. 125–43.

[71] Kaplan, E. and Shapley, R.M. "The primate retina contains two types of ganglion cells, with high and low contrast sensitivity". In: *Proc Natl Acad Sci U S A* 83 (1986), pp. 2755–2757.

[72] Kaplan, Ehud, Lee, Barry B., and Shapley, Robert M. "New views of primate retinal function". In: *Progress in retinal research* 9 (1990), pp. 273–336.

[73] Kording, K. P. and Wolpert, D. M. "Bayesian integration in sensorimotor learning". In: *Nature* 427.6971 (2004), pp. 244–7.

[74] Kremers, J., Silveira, L.C.L., and Kilavik, B.E. "Influence of Contrast on the Responses of Marmoset Lateral Geniculate Cells to Drifting Gratings". In: *Journal of Neurophysiology* (2001), pp. 235–246.

[75] Kremkow, J. and Alonso, J. M. "Thalamocortical Circuits and Functional Architecture". In: *Annu Rev Vis Sci* 4 (2018), pp. 263–285.

[76] Kuffler, S. W. "Discharge Patterns and Functional Organization of Mammalian Retina". In: *Journal of Neurophysiology* 16.1 (1953), pp. 37–68.

[77] Levine, M. W. and Abramov, I. "An analysis of spatial summation in the receptive fields of goldfish retinal ganglion cells". In: *Vision Res* 15.7 (1975), pp. 777–89.

[78] Levitt, J. B. et al. "Visual response properties of neurons in the LGN of normally reared and visually deprived macaque monkeys". In: *J Neurophysiol* 85.5 (2001), pp. 2111–29.

[79] Levy, Paul G., Simoncelli, Eero P., and Movshon, J. Anthony. "Contrast-dependent spatial frequency selectivity in macaque V1 neurons explained with tuned contrast gain control". In: *Journal of Vision* 19.10 (2019), 43a–43a.

[80] Levy, Paul G. et al. "Differing mechanisms for contrast-dependent spatial frequency selectivity in macaque LGN and V1 neurons". In: *Journal of Vision* 20.11 (2020), pp. 1579–1579.

[81] Li, Y. et al. "Cascaded normalizations for spatial integration in the primary visual cortex of primates". In: *Cell Rep* 40.7 (2022), p. 111221.

[82] Lieber, J. D. et al. “Sensitivity to naturalistic texture relies primarily on high spatial frequencies”. In: *J Vis* 23.2 (2023), p. 4.

[83] Linsenmeier, R. A. et al. “Receptive field properties of x and y cells in the cat retina derived from contrast sensitivity measurements”. In: *Vision Res* 22.9 (1982), pp. 1173–83.

[84] Ma, W. J. et al. “Bayesian inference with probabilistic population codes”. In: *Nat Neurosci* 9.11 (2006), pp. 1432–8.

[85] Maffei, L. and Fiorentini, A. “The visual cortex as a spatial frequency analyser”. In: *Vision Res* 13.7 (1973), pp. 1255–67.

[86] Majaj, N. J. et al. “The role of spatial frequency channels in letter identification”. In: *Vision Res* 42.9 (2002), pp. 1165–84.

[87] Malpeli, J. G. and Baker, F. H. “The representation of the visual field in the lateral geniculate nucleus of Macaca mulatta”. In: *J Comp Neurol* 161.4 (1975), pp. 569–94.

[88] Mante, V., Bonin, V., and Carandini, M. “Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli”. In: *Neuron* 58.4 (2008), pp. 625–38.

[89] Mante, V. et al. “Independence of luminance and contrast in natural scenes and in the early visual system”. In: *Nat Neurosci* 8.12 (2005), pp. 1690–7.

[90] Marr, D. “Vision”. In: *Minds, Brains, and Computers: The Foundations of Cognitive Science*. Ed. by R. Cummins and D. D. Cummins. BLackwell, 1982.

[91] Mazer, James A et al. “Spatial frequency and orientation tuning dynamics in area V1”. In: *Proceedings of the National Academy of Sciences* 99.3 (2002), pp. 1645–1650.

[92] Merrill, E. G. and Ainsworth, A. “Glass-coated platinum-plated tungsten microelectrodes”. In: *Med Biol Eng* 10.5 (1972), pp. 662–72.

[93] Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. “Receptive field organization of complex cells in the cat’s striate cortex”. In: *J Physiol* 283 (1978), pp. 79–99.

[94] Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. “Spatial summation in the receptive fields of simple cells in the cat’s striate cortex”. In: *J Physiol* 283 (1978), pp. 53–77.

[95] Movshon, J. A. et al. “Functional maturation of the macaque’s lateral geniculate nucleus”. In: *J Neurosci* 25.10 (2005), pp. 2712–22.

[96] Naka, K. I. and Rushton, W. A. “S-potentials from luminosity units in the retina of fish (Cyprinidae)”. In: *J Physiol* 185.3 (1966), pp. 587–99.

[97] Nauhaus, I. et al. “Neuronal selectivity and local map structure in visual cortex”. In: *Neuron* 57.5 (2008), pp. 673–9.

[98] Nauhaus, I. et al. “Orthogonal micro-organization of orientation and spatial frequency in primate primary visual cortex”. In: *Nat Neurosci* 15.12 (2012), pp. 1683–90.

[99] Nestares, O. and Heeger, D. J. “Modeling the apparent frequency-specific suppression in simple cell responses”. In: *Vision Res* 37.11 (1997), pp. 1535–43.

[100] Ninomiya, T., Sanada, T. M., and Ohzawa, I. “Contributions of excitation and suppression in shaping spatial frequency selectivity of V1 neurons as revealed by binocular measurements”. In: *J Neurophysiol* 107.8 (2012), pp. 2220–31.

[101] Nolt, M. J., Kumbhani, R. D., and Palmer, L. A. “Suppression at high spatial frequencies in the lateral geniculate nucleus of the cat”. In: *J Neurophysiol* 98.3 (2007), pp. 1167–80.

[102] Nolt, M.J., Kumbhani, R.D., and Palmer, L.A. “Contrast-dependent spatial summation in the lateral geniculate nucleus and retina of the cat”. In: *Journal of Neurophysiology* 92 (2004), pp. 1708–1717.

[103] Oppenheim, Alan V. and Lim, Jae S. “The importance of phase in signals”. In: *Proceedings of the IEEE* 69.5 (1981), pp. 529–541.

[104] Ozeki, H. et al. “Relationship between excitation and inhibition underlying size tuning and contextual response modulation in the cat primary visual cortex”. In: *J Neurosci* 24.6 (2004), pp. 1428–38.

[105] Paik, S. B. and Ringach, D. L. “Retinal origin of orientation maps in visual cortex”. In: *Nat Neurosci* 14.7 (2011), pp. 919–25.

[106] Parker, A. J. and Hawken, M. J. “Two-dimensional spatial structure of receptive fields in monkey striate cortex”. In: *J Opt Soc Am A* 5.4 (1988), pp. 598–605.

[107] Persi, E. et al. “Power-law input-output transfer functions explain the contrast-response and tuning properties of neurons in visual cortex”. In: *PLoS Comput Biol* 7.2 (2011), e1001078.

[108] Piotrowski, L. N. and Campbell, F. W. “A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase”. In: *Perception* 11.3 (1982), pp. 337–46.

[109] Priebe, N. J. and Ferster, D. “Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex”. In: *Neuron* 45.1 (2005), pp. 133–45.

[110] Priebe, N. J., Lisberger, S. G., and Movshon, J. A. "Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex". In: *J Neurosci* 26.11 (2006), pp. 2941–50.

[111] Prince, S. J. D. et al. "Quantitative Analysis of the Responses of V1 Neurons to Horizontal Disparity in Dynamic Random-Dot Stereograms". In: *Journal of Neurophysiology* 87 (2002), pp. 191–208.

[112] Raghavan, R. T. et al. "Contrast and luminance gain control in the macaque's lateral geniculate nucleus". In: *eneuro* (2023), ENEURO.0515–22.2023.

[113] Reid, R. C. and Alonso, J.-M. "Specificity of monosynaptic connections from thalamus to visual cortex". In: *Nature* 378 (1995), pp. 281–284.

[114] Ringach, Dario L. "Haphazard wiring of simple receptive fields and orientation columns in visual cortex". In: *Journal of neurophysiology* 92.1 (2004), pp. 468–476.

[115] Ringach, Dario L et al. "Suppression of neural responses to nonoptimal stimuli correlates with tuning selectivity in macaque V1". In: *Journal of Neurophysiology* 87.2 (2002), pp. 1018–1027.

[116] Ringach, Dario L. "On the origin of the functional architecture of the cortex". In: *PLoS One* 2.2 (2007), e251.

[117] Ringach, Dario L., Hawken, M. J., and Shapley, R. "Dynamics of Orientation Tuning in Macaque V1: The Role of Global and Tuned Suppression". In: *Journal of Neurophysiology* 90 (2003), pp. 342–352.

[118] Rodieck, R. W. and Stone, J. "Analysis of receptive fields of cat retinal ganglion cells". In: *J Neurophysiol* 28.5 (1965), pp. 833–49.

[119] Rodieck, R. W. and Stone, J. "Response of cat retinal ganglion cells to moving visual patterns". In: *J Neurophysiol* 28.5 (1965), pp. 819–32.

[120] Rubin, D. B., Van Hooser, S. D., and Miller, K. D. "The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex". In: *Neuron* 85.2 (2015), pp. 402–17.

[121] Ruderman, D. L. and Bialek, W. "Statistics of natural images: Scaling in the woods". In: *Phys Rev Lett* 73.6 (1994), pp. 814–817.

[122] Rust, N. C. et al. "Spatiotemporal elements of macaque v1 receptive fields". In: *Neuron* 46.6 (2005), pp. 945–56.

[123] Salinas, E. and Abbott, L. F. "Vector reconstruction from firing rates". In: *J Comput Neurosci* 1.1-2 (1994), pp. 89–107.

[124] Sceniak, M. P., Chatterjee, S., and Callaway, E. M. “Visual spatial summation in macaque geniculocortical afferents”. In: *J Neurophysiol* 96.6 (2006), pp. 3474–84.

[125] Sceniak, M. P., Hawken, M. J., and Shapley, R. “Visual spatial characterization of macaque V1 neurons”. In: *J Neurophysiol* 85.5 (2001), pp. 1873–87.

[126] Sceniak, M. P., Hawken, M. J., and Shapley, R. “Contrast-Dependent Changes in Spatial Frequency Tuning of Macaque V1 Neurons: Effects of a Changing Receptive Field Size”. In: *Journal of Neurophysiology* 88 (2002), pp. 1363–1373.

[127] Sceniak, M. P. et al. “Contrast’s effect on spatial summation by macaque V1 neurons”. In: *Nature* 2.8 (1999), pp. 733–739.

[128] Schaaf, A. van der and Hateren, J. H. van. “Modelling the power spectra of natural images: statistics and information”. In: *Vision Res* 36.17 (1996), pp. 2759–70.

[129] Schwartz, O. and Simoncelli, E. P. “Natural signal statistics and sensory gain control”. In: *Nature Neuroscience* 4.8 (2001), pp. 819–825.

[130] Schwartz, Odelia. “Modeling and Characterization of Neural Gain Control”. Thesis. 2002.

[131] Sclar, G. and Freeman, R. D. “Orientation selectivity in the cat’s striate cortex is invariant with stimulus contrast”. In: *Exp Brain Res* 46.3 (1982), pp. 457–61.

[132] Series, P., Stocker, A. A., and Simoncelli, E. P. “Is the homunculus "aware" of sensory adaptation?” In: *Neural Comput* 21.12 (2009), pp. 3271–304.

[133] Shapley, R., Hawken, M. J., and Ringach, D. L. “Dynamics of Orientation Selectivity in the Primary Visual Cortex and the Importance of Cortical Inhibition”. In: *Neuron* 38 (2003), pp. 689–699.

[134] Shapley, R. M. and Victor, J. D. “The effect of contrast on the transfer properties of cat retinal ganglion cells”. In: *J Physiol* 285 (1978), pp. 275–98.

[135] Shapley, R. M. and Victor, J. D. “Nonlinear spatial summation and the contrast gain control of cat retinal ganglion cells”. In: *J Physiol* 290.2 (1979), pp. 141–61.

[136] Shapley, R. M. and Victor, J. D. “How the contrast gain control modifies the frequency responses of cat retinal ganglion cells”. In: *Journal of Physiology* 318 (1981), pp. 161–179.

[137] Shapley, R. M. and Xing, D. “Local circuit inhibition in the cerebral cortex as the source of gain control and untuned suppression”. In: *Neural Netw* 37 (2013), pp. 172–81.

[138] Shapley, Robert and Enroth-Cugell, Christina. “Visual adaptation and retinal gain controls”. In: *Progress in Retinal Research* 3 (1984), pp. 263–346.

[139] Shooner, Christopher, Kiorpes, Lynne, and Movshon, J. Anthony. “Binocular combination and contextual masking in normal and abnormal visual cortex”. Thesis. 2015.

[140] Shushruth, S. et al. “Different orientation tuning of near- and far-surround suppression in macaque primary visual cortex mirrors their tuning in human perception”. In: *J Neurosci* 33.1 (2013), pp. 106–19.

[141] Sillito, A. M. et al. “A re-evaluation of the mechanisms underlying simple cell orientation selectivity”. In: *Brain Res* 194.2 (1980), pp. 517–20.

[142] Skottun, B. C. et al. “The effects of contrast on visual orientation and spatial frequency discrimination: A comparison of single cells and behavior”. In: *Journal of Neurophysiology* 57.3 (1987), pp. 773–786.

[143] Skottun, B. C. et al. “Classifying simple and complex cells on the basis of response modulation”. In: *Vision Res* 31.7-8 (1991), pp. 1079–86.

[144] Smith, Matthew A., Bair, Wyeth, and Movshon, J. Anthony. “Dynamics of Suppression in Macaque Primary Visual Cortex”. In: *The Journal of Neuroscience* 26.18 (2006), pp. 4826–4834.

[145] So, Y. T. and Shapley, R. “Spatial properties of X and Y cells in the lateral geniculate nucleus of the cat and conduction veolcities of their inputs”. In: *Exp Brain Res* 36.3 (1979), pp. 533–50.

[146] Sokol, Sach H. et al. “Selectivity and dynamics of responses in the central visual pathway of the macaque”. Thesis. 2009.

[147] Solomon, S. G., Lee, B. B., and Sun, H. “Suppressive surrounds and contrast gain in magnocellular-pathway retinal ganglion cells of macaque”. In: *J Neurosci* 26.34 (2006), pp. 8715–26.

[148] Solomon, S. G., White, A. J., and Martin, P. R. “Extraclassical receptive field properties of parvocellular, magnocellular, and koniocellular cells in the primate lateral geniculate nucleus”. In: *J Neurosci* 22.1 (2002), pp. 338–49.

[149] Spear, P.D. et al. “Effects of Aging on the Primate Visual System:Spatial and Temporal Processingby Lateral Geniculate Neurons in Young Adult and Old Rhesus Monkeys”. In: *Journal of Neurophysiology* 72.1 (1994), pp. 402–420.

[150] Spitzer, H. and Hochstein, S. “A complex-cell receptive-field model”. In: *J Neurophysiol* 53.5 (1985), pp. 1266–86.

[151] Spitzer, H. and Hochstein, S. “Simple- and complex-cell response dependences on stimulation parameters”. In: *J Neurophysiol* 53.5 (1985), pp. 1244–65.

[152] Stocker, A. A. and Simoncelli, E. P. “Noise characteristics and prior expectations in human visual speed perception”. In: *Nat Neurosci* 9.4 (2006), pp. 578–85.

[153] Tadmor, Y. and Tolhurst, D. J. “Both the phase and the amplitude spectrum may determine the appearance of natural images”. In: *Vision Res* 33.1 (1993), pp. 141–5.

[154] Tailby, C. et al. “Two expressions of “surround suppression” in V1 that arise independent of cortical mechanisms of suppression”. In: *Visual Neuroscience* 24 (2007), pp. 99–109.

[155] Thomson, Mitchell G. A., Foster, David H., and Summers, Robert J. “Human sensitivity to phase perturbations in natural images: a statistical framework”. In: *Perception* 29.9 (2000), pp. 1057–1069.

[156] Troyer, Todd W., Krukowski, Anton E., and Miller, Kenneth D. “LGN Input to Simple Cells and Contrast-Invariant Orientation Tuning: An Analysis”. In: *Journal of Neurophysiology* 87.6 (2002), pp. 2741–2752.

[157] Van Essen, D. C., Anderson, C. H., and Felleman, D. J. “Information processing in the primate visual system: an integrated systems perspective”. In: *Science* 255.5043 (1992), pp. 419–23.

[158] Van Essen, D. C. and Zeki, S. M. “The topographic organization of rhesus monkey prestriate cortex”. In: *J Physiol* 277 (1978), pp. 193–226.

[159] Van Essen, D. C. et al. “Cerebral cortical folding, parcellation, and connectivity in humans, nonhuman primates, and mice”. In: *Proc Natl Acad Sci U S A* 116.52 (2019), pp. 26173–26180.

[160] Vintch, B., Movshon, J. A., and Simoncelli, E. P. “A Convolutional Subunit Model for Neuronal Responses in Macaque V1”. In: *J Neurosci* 35.44 (2015), pp. 14829–41.

[161] Von Blanckensee, H. T. “Spatio-temporal properties of cells in monkey lateral geniculate nucleus.” Thesis. 1980.

[162] Webb, B. S. et al. “Early and late mechanisms of surround suppression in striate cortex of macaque”. In: *J Neurosci* 25.50 (2005), pp. 11666–75.

[163] Webster, M. A. and De Valois, R. L. “Relationship between spatial-frequency and orientation tuning of striate-cortex cells”. In: *J Opt Soc Am A* 2.7 (1985), pp. 1124–32.

[164] Webster, M. A. and Miyahara, E. “Contrast adaptation and the spatial structure of natural images”. In: *J Opt Soc Am A Opt Image Sci Vis* 14.9 (1997), pp. 2355–66.

[165] Wiesel, T. N. and Hubel, D. H. “Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey”. In: *J Neurophysiol* 29.6 (1966), pp. 1115–56.

[166] Wilson, D. E., Scholl, B., and Fitzpatrick, D. “Differential tuning of excitation and inhibition shapes direction selectivity in ferret visual cortex”. In: *Nature* 560.7716 (2018), pp. 97–101.

[167] Xing, D. et al. “Correlation of local and global orientation and spatial frequency tuning in macaque V1”. In: *J Physiol* 557.Pt 3 (2004), pp. 923–33.

[168] Xing, D. et al. “Effect of stimulus size on the dynamics of orientation selectivity in Macaque V1”. In: *J Neurophysiol* 94.1 (2005), pp. 799–812.

[169] Ziemba, C. M. et al. “Laminar Differences in Responses to Naturalistic Texture in Macaque V1 and V2”. In: *J Neurosci* 39.49 (2019), pp. 9748–9756.