ABSTRACT

Title of Dissertation: EXPLORING NEURAL REPRESENTATIONS IN MACAQUE PRIMARY VISUAL CORTEX THROUGH DATA-DRIVEN MODELS

Felix Bartsch
Doctor of Philosophy
2024

Dissertation directed by: Dr. Daniel Butts, Department of Biology

The study of the primary visual cortex (V1) holds profound significance for our understanding of the neural underpinnings of visual perception. Computational models have emerged as invaluable tools to decode the intricate computations occurring within V1 neurons. This dissertation embarks on a comprehensive exploration of V1 by fitting statistical models to electrophysiological data and scrutinizing model properties. My approach not only provides direct insights into how V1 neurons represent information but also furnishes mathematical descriptions of V1 computations, thereby contributing to the construction of a unified model of V1 function.

I begin in Chapter 2 by employing state-of-the-art statistical and machine learning techniques to unravel the high-resolution components of V1 receptive fields as they respond to random bar stimuli. I demonstrate how these models not only replicate classical findings but also offer superior explanations of the computations V1 undertakes. These results highlight how the simultaneous processing of multiple overlapping inputs enables cells to represent high-resolution information while also responding to full-field inputs, an intricate organization unattainable using conventional stimuli. In chapter 3, I expand this modeling approach by adding mechanisms for binocular integration and apply them to data obtained from random bar stimuli that also vary in binocular disparity. This approach reveals that V1 disparity selectivity is enhanced and well characterized using spatial convolutions. Finally, I further modify the approach in chapter 4 to map spatiotemporal receptive fields in luminance and color using data recorded from the fovea and present the first spatiochromatic measurements illustrating the scale of V1 processing at the fovea. I find that color signals operate at lower spatial scales compared to luminance signals, and that receptive field substructure can allow even cells with large receptive fields to represent fine-scale information throughout the fovea.

EXPLORING NEURAL REPRESENTATIONS IN MACAQUE PRIMARY VISUAL CORTEX THROUGH DATA-DRIVEN MODELS

by

Felix Bartsch

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

2024

Advisory Committee:
Professor Daniel A. Butts, Chair
Dr. Bevil R. Conway
Professor Elizabeth Quinlan
Professor Jonathan Z. Simon
Professor Timothy K. Horiuchi

Dedication

This thesis truly would not have been possible without the support of a plethora of truly wonderful people who helped me through the trials of graduate school and research, and I am more grateful to them all than a simple dedication page can convey.

First and foremost, this work would not have been possible without my advisor, Dr. Daniel Butts, who has been a steadfast source of mentorship not just for how to approach my specific research projects, but also how to think more deeply about scientific questions and frame them. His professional and personal support made the projects contained within this thesis as polished as they are.

Likewise, I thank Dr. Bevil Conway for facilitating the final and most involved chapter of this thesis, and for giving me endless advice for thriving in academia, and especially for sharing with me his hard-earned knowledge of presenting scientific information – a skill I will continue to hone for the rest of my career.

Next, I would like to thank Dr. Bruce Cumming for likewise providing excellent insights into how to best think about and study visual computations, as well as providing the data for the first two chapters.

Additionally, I cannot thank the various lab members of the Butts and Conway labs enough for their continued support and assistance, both scientifically and emotionally. Special considerations to Audrey Chang, for helping me stay sane while we were both in the trenches of recordings; to Dr. James Cavanaugh, for teaching me how to work with monkeys and always helping me find perspective; to Dr. Danny Garside, for always being there to listen and walk through color spaces with me; and to Matt Jacobsen, for struggling through the data alongside me.

Also, I am grateful to everyone in NACS and the UMD neuroscience community in general for all the conversations about science, grad school, and life that made my graduate school years enjoyable. Special considerations to Dr. Matthew Coon Manion, both for being an amazing friend and for introducing me to D&D and getting me into a hobby that saved my mental health during a pandemic; to Matt Evanusa and Dr. Greg Davis for the best, most insightful conversations about neural computations I've had; and to Drs. Andrew and Lauren Borrell, Dr. Greg Perrin, Dr. Jon Schenk, and Dr. Ruilong Hu, for bearing with me through it all.

Most importantly, I'd like to thank both my parents, Andreas and Stephanie, as well as my partner, Dr. Elizabeth Tarantino, for being my rocks. I could not have completed this work without your moral and intellectual support and your love, which I will never forget, and will always keep working to earn - no matter how much you tell me I don't need to.

Finally, I'd like to thank Lemeiux, Jr. Barnes, Jocamo, and Vincent for providing the neurons I studied in this thesis. You earned every last grape.

Table of Contents

List of Figures

List of Abbreviations

CSD	Current-source Density
CRT	cathode ray tube
D-O	Double-opponent
DSI	Direction selectivity index
DVF	Disparity variance fraction
E-I	Excitatory-Inhibitory
GQM	General quadratic model
K	Koniocellular
L-cone	Long- wavelength cone
L2/3	Layer 2/3 / supragranular layer
L4	Layer 4 / granular layer
L5/6	Layer 5/6 / infragranular layer
LCA	Longitudinal chromatic aberration
LGN	Lateral geniculate nucleus
Lin	Linear
LL	Log-likelihood
LN	Linear-nonlinear
Lum	Luminance
M	Magnocellular
M-cone	Medium-wavelength cone
NIM	Nonlinear input model
OSI	Orientation selectivity index
P	Parvocellular
PCNIM	Principal component NIM
Quad	Quadratic
ReLU	Rectified Linear Unit
S-cone	Short wavelength cone
SF	Spatial frequency
STA	Spike-triggered average
STC	Spike-triggered covariance
TCA	Transverse chromatic aberration
TF	Temporal frequency
V1	Primary visual cortex

1 Introduction

1.1 How do we see what we see?

The question of how we are able to form visual percepts goes back to antiquity, and although science has made significant progress in this matter, the road is long and so far, it looks processes represent visual information and facilitate perception?". In this thesis, I will focus on one major component of this question by investigating how visual signals are represented and processed in primary visual cortex, or V1, the first cortical processing stage of vision and the foundation of conscious visual perception (Leopold, 2012).

The link between neurons in V1 and vision was discovered by multiple researchers, starting with Bartolomeo Panizza in 1855, first through observational studies of stroke patients, and later through cortical lesioning studies in animals. By the 1920s, systematic observations of patients with cortical lesions had led to the discovery that V1 contained a precise map of the visual field (Gross, 1998). However, it was the work of Hubel and Wiesel in the 1950s that truly broke ground on the study of visual cortex when they published a seminal series of papers on V1's cellular responses, anatomy, development, and connectivity (Hubel and Wiesel, 1959, 1962, 1963, 1968, 1969, 1972; Hubel et al., 1978). They found that individual V1 neurons responded selectively to specific features at specific locations in the visual scene, such as oriented lines or edges. This finding helped contextualize the role of individual neurons in the complex web of connectivity that somehow generates conscious visual perception and thus constitutes not only a breakthrough for

vision research but for all neuroscience. Their work paved the way for substantial advances that continue to this day.

It is easy to see why V1 was an appealing initial target for study: it is well preserved across all mammalian species (Krubitzer, 2007), meaning that discoveries in an animal model likely translate to a better understanding of human neurophysiology; and (at least in most species) V1 is largely located on the cortical surface, making it easily accessible for researchers using different experimental approaches.

Today, following decades of study, additional factors have made V1 even more appealing for study: we understand the nature of its main inputs, we know what stimuli make its neurons fire, and we can easily create and manipulate those stimuli thanks to computer displays. As a result, V1 is one of the best-understood areas of the cerebral cortex and constitutes a prime workbench for the study of cortical circuits and computations.

1.2 Feedforward inputs into V1

V1 primarily receives visual information via afferents from the lateral geniculate nucleus (LGN) and transmits processed information to higher visual areas and back to subcortical structures via feedback connections (Felleman and Van Essen, 1991). The importance of the processing occurring in V1 should not be understated: while LGN processing is similarly fundamental to vision, V1 dramatically expands upon it thanks to its high neuronal density and large area. In humans, V1 contains approximately 140 million neurons per hemisphere, or about 40 V1 neurons for each LGN neuron (Wandell, 1995). In macaque monkeys, ~1.6 million retinal

ganglion cells (RGCs) project to a comparable number of neurons in the LGN, which in turn connect to over 120 **million** neurons in V1 (Van Essen et al., 1984). This divergence naturally provides ample opportunity for extensive processing of the visual information received from the LGN. Thus, understanding its geniculate afferents is an important first step toward understanding the diversity and depth of V1 processing.

There are three major types of LGN neurons projecting to macaque V1 through parallel pathways: magnocellular (M), parvocellular (P), and koniocellular (K). Each originates from specific types of RGCs: respectively, parasol cells, midget cells, and bistratified cells (Perry et al., 1984; Callaway and Callaway, 2005).

The M pathway is characterized by large neurons. Parasol RGCs project large-diameter axons to the two M layers of the LGN. The LGN M cells tend to have relatively large receptive fields, are wavelength insensitive, respond transiently to visual stimuli, prefer low spatial frequencies, and are relatively sensitive to luminance contrast. This makes them poorly suited for the analysis of fine shape or color but excellent for detecting subtle luminance changes or rapidly moving stimuli. The M pathway primarily projects into layer 4Cα of V1, although some projection to layer 6 has also been noted (Hubel and Wiesel, 1972; Sincich and Horton, 2005).

The P pathway originates from the much smaller midget RGCs, whose projections terminate in the four P layers of the LGN. Neurons in the P pathway are more numerous in the central retina and have smaller receptive fields than those in the M pathway, allowing them to convey more detailed fine spatial information. They also have color opponent receptive fields, allowing them to detect color contrast, an important feature for the later processing of color

information. P cells have more sustained visual responses than M cells, and their finer caliber axons have slower conduction velocities, making them less useful for the detection of rapid movement. Neurons in these layers send their axons primarily to layer 4Cβ of V1, along with weaker projections to layers 6 and 4A (Hubel and Wiesel, 1972).

The K pathway originates from a heterogeneous group of wide-field cells, including sparse cells and bistratified RGCs, which have the smallest diameter axons that innervate the thin "intercalated" layers, between the M and P layers of the macaque LGN. The K pathway remains more enigmatic than the relatively well-characterized M and P pathways, although some basic properties are well-documented. Bistratified cells carry short-wavelength signals and are used to generate center-only blue-ON/yellow-OFF receptive fields in the LGN (Hendry and Clay Reid, 2000). Sparse cells are presumed to transmit blue-OFF signals. The intercalated geniculate neurons then project to layer 1 and CO blobs in layer 2/3 of V1 (Livingstone and Hubel, 1982; Hendry and Clay Reid, 2000)

In summary, the only regions in macaque V1 that do not receive direct geniculate input are layers 4B and 5, and the CO interblobs in layers 2/3. Layers 4Cα and 4Cβ receive M and P input, respectively. Layer 6 receives both M and P input. Layer 4A receives just P input, and layer 2/3 blobs and layer 1 receive just K input.

1.3 Functional characterization of V1 neurons

Early studies of the primary visual cortex (V1) made major progress in the understanding of V1 utilizing simple parametric stimuli, primarily single bars and gratings.

Bar stimuli, effectively just oriented lines or edges on a uniform background, are highly controlled and easily manipulable. This allowed researchers to precisely vary the orientation, size, contrast, and other properties of the stimuli, as well as their combinations, all while remaining in a relatively low parameter space. Bar stimuli also make it easy to replicate the same stimulus conditions in multiple trials, making it easier to assess the reliability of neuronal responses and establish consistent patterns. So, by systematically varying one stimulus parameter at a time and carefully observing changes in the neural responses, Hubel and Wiesel and their contemporaries were able to uncover the basic building blocks of visual processing.

The other type of classical parametric stimuli is drifting gratings, which provide a more complex and dynamic visual input compared to single bars. They contain multiple visual attributes, making them useful for studying how V1 processes a variety of visual features simultaneously, without significantly complicating data analysis. Thus, grating stimuli, such as Hartley series, could be used as a general framework that allows the experimenter to incorporate a priori knowledge about the spatial properties of a cell and to restrict the study of the system to particular dimensions of interest. This effectively reduces the dimension of the input space and yields higher signal to noise ratios (Ringach et al., 1997). As a result, gratings were able to characterize further fundamental tuning properties of V1 neurons.

The beauty in this approach remains in its simplicity: Simple stimuli allowed researchers to design experiments with clear, well-defined questions, and resulted in straightforward and interpretable data. Additionally, this approach provided a common basis for comparative analysis.

Presenting the same set of stimuli to different neurons in V1 was the first step towards functional classification of neurons based on their response properties.

Orientation Selectivity. The most obvious such classification is orientation tuning. Most V1 neurons exhibit the highest response when presented with bars or edges oriented at specific angles, and their responses decrease as the orientation of the stimulus diverges from the preferred orientation of the neuron. This selectivity derives directly from the shape of their receptive field: these cells have elongated ON and OFF subregions, meaning that certain subregions within their receptive fields respond preferentially to light or dark stimuli.

Simple vs Complex. A second classification is the distinction between simple-like and complex-like cells (Fig. 1A). Simple cells are phase selective, meaning that they have a single set of ON and OFF subregions and thus respond to bars at a single location (Hubel and Wiesel, 1968; Movshon et al., 1978a). Complex cells are phase invariant, meaning they respond to bars of the preferred orientation at any position within their receptive field (Hubel and Wiesel, 1968; Movshon et al., 1978b). This follows straightforwardly from the descriptive model of complex cells: all the simple cells providing input to the complex cell are selective for the same orientation (Fig 1B), endowing the complex cell with the same orientation selectivity (Alonso and Martinez, 1998; Carandini, 2006). Notably, while early studies identified simple and complex cells as separate categories (Skottun et al., 1991), later studies revealed that while the distinction remains useful, V1 operates more as a spectrum of simple-like to complex-like cells, with many intermediates that show some degree of phase invariance but may still have a preferred phase (Mechler and Ringach, 2002).

Spatial frequency. In addition to orientation, V1 neurons are typically sharply selective for the spatial frequency of a stimulus. Spatial frequency is best defined for a grating pattern, where frequency is the inverse of the distance between bars. Just as for orientation, this selectivity arises naturally from the shape of the receptive fields (Fig. 1.1A). These receptive fields have multiple ON and OFF regions, and the more regions there are the more selective the neurons are for spatial frequency. Consequently, neurons in V1 tend to be much more selective for spatial frequency than LGN neurons (De Valois et al., 1982).

Temporal Frequency. Additionally, V1 neurons are typically thought to be selective for temporal frequency, or inverse of the period between temporal oscillations between dark and light. V1 neurons typically prefer lower temporal frequencies than those that can drive LGN neurons but can exhibit much more intricate temporal response profiles compared to LGN.

*Direction***.** Cells in area V1 are commonly selective for direction of stimulus motion. Grating stimuli are particularly valuable for studying motion processing in V1, as the drifting motion of gratings allowed researchers to investigate how V1 neurons respond to the direction and speed of moving stimuli, revealing the mechanisms behind motion perception. Receptive fields of direction-selective V1 neurons can be thought of as slanted in space-time: a receptive field slanted towards the right in space-time confers selectivity for stimuli moving rightward. Support for this view comes from studies that have measured the space-time receptive fields of simple cells and found that they are indeed slanted in space-time (De Valois et al., 2000; Lochmann et al., 2013).

*Disparity***.** In animals with front-facing eyes (such as carnivores and primates), much of the visual field is covered jointly by both eyes. This poses a challenge as signals need to be integrated,

but also an opportunity for computing binocular depth (stereoscopy). The signals from corresponding regions in the two eyes are kept separate in the LGN and are combined in V1. Most neurons receive inputs from both eyes (Priebe et al., 2004), but the relative ability of each eye alone to drive spike responses varies: some neurons are driven mostly by one eye or the other, whereas other neurons are driven by both eyes (Hubel and Wiesel, 1962). In many neurons this arrangement is accompanied by selectivity for binocular depth, or disparity (Cumming and DeAngelis, 2001). The specifics of disparity tuning in V1 are discussed in detail in chapter 3.

*Color***.** In primates, the retina contains cones responsive to three bands of wavelengths (trichromacy). RGCs compare activation of these three cones to identify spectral differences in stimuli, also called cone-opponency. In the LGN, these cone-opponent signals are organized along three "cardinal directions", specifically Luminance (black vs white), L-M (red-green), and S+(L-M) vs S-(L-M) (blue-yellow) (Derrington et al., 1984). The chromatic selectivity of V1 neurons was initially thought to be much more varied (Gegenfurtner and Kiper, 2003), but later data indicated that it is also organized along "cardinal directions" though not necessarily ones corresponding to the LGN opponencies (Horwitz and Hass, 2012). The specifics of V1's selectivity to color are discussed in detail in chapter 4.

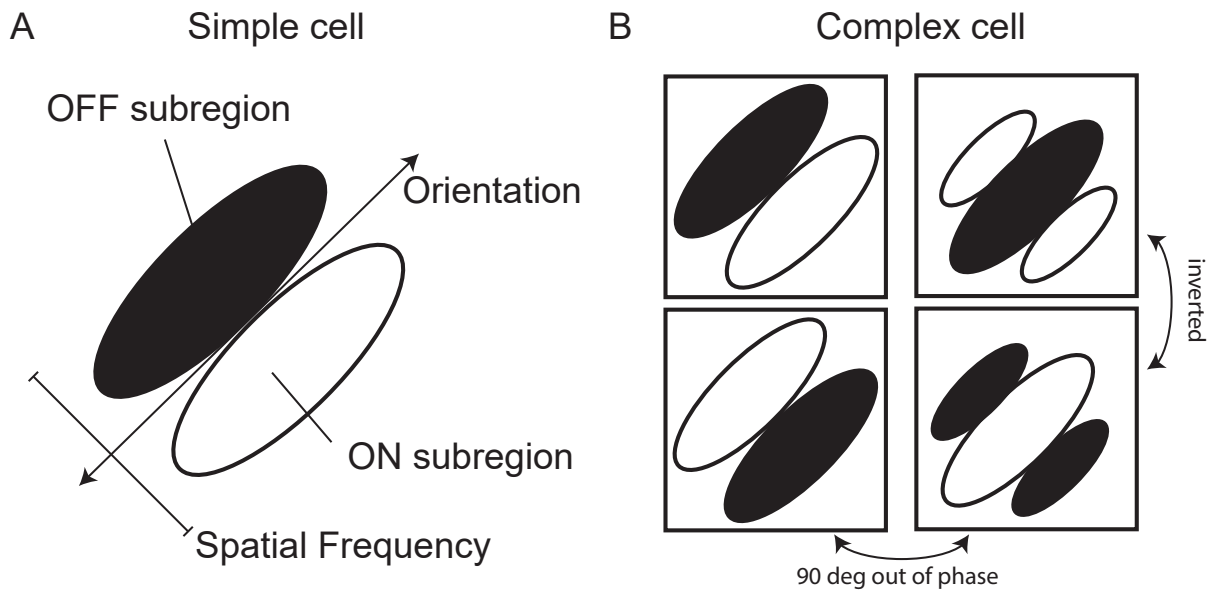

Figure 1.1: V1 receptive fields.

A: Example of a simple cell receptive field, preferring a 45-degree orientation with single ON and OFF domains. **B**: Example of simple cell RFs projecting to a complex cell.

1.4 Organization of functional domains in V1

1.4.1 Retinotopy

V1 is characterized by retinotopic organization, meaning it contains a complete map of the visual field from both eyes. This retinotopy is continuous, with nearby points in the cortex mapping to nearby visual positions (Bosking et al., 2002; Adams and Horton, 2003). However, while retinotopy is largely continuous, it is not even: One significant distortion is cortical magnification, which prioritizes the central visual field at the expense of the periphery. For example, approximately 50% of human V1 is dedicated to representing the central 2% of the visual field

(Wandell, 1995), while in macaques, about 42% is dedicated to the central 5 degrees (Van Essen et al., 1984; Tootell et al., 1988b). This emphasis is partly inherited from the retina (Adams and Horton, 2003), and further enhanced by the high neural density of V1.

Besides the retinotopic map of spatial location of receptive fields, there are also maps of specific feature selectivity, the most well-known of which is the pinwheel map of orientation. When mapped using electrode penetrations across the cortical surface, maps of orientation appear as slabs perpendicular to the cortical surface, with each slab containing an orderly sequence of slowly shifting orientation preferences - until a slab terminates when orientation preference suddenly shifts drastically (Hubel and Wiesel, 1974). Imaging using voltage-sensitive dyes further revealed that these orientation slabs are curved across cortical space (Blasdel and Salama, 1986; Obermayer and Blasdel, 1993) and revealed orientation "pinwheels" (see Fig. 1.2A) at the centers of circular orientation domains where many preferences converge (Das and Gilbert, 1997), but even here, orientation preference has been shown to be finely segregated (Ohki et al., 2006).

A second well-known organizational map consists of ocular dominance (OD) regions. Ocular dominance columns are striped regions of V1 that receive preferential input from one eye compared to the other, with cells at the edges of OD columns being more likely to be binocular (Hubel et al., 1978; LeVay et al., 1985; Tootell et al., 1988b). However, while these preferences are noteworthy, they are not prescriptive, as binocular cells can be found within OD columns and monocular cells can be found in intermediate regions (Adams and Horton, 2006). Comparing orientation maps with ocular dominance patterns from the same cortical regions, OD columns tend

to run perpendicular to orientation maps, forming a mosaic of selectivity (Obermayer and Blasdel, 1993).

In addition to other maps, a range of spatial frequencies are likewise represented independently within V1. Domains with SF preferences at the extremes of the SF continuum are often found clustered closely together at pinwheel centers in the cortical map of orientation preference (Issa et al., 2000). The organization of cortical maps permits nearly all combinations of orientation and SF preference to be represented in V1 (Tootell et al., 1988a).

A further organizational principle is the clustering into ON/OFF domains, where the responses of neurons are dominated either by the onset (ON) or offset (OFF) of luminance within their receptive fields (Kremkow et al., 2016). These domains tend to alternate at a high spatial resolution, with individual domains often being less than 100 microns wide - and thus, often corresponding to the scale of visual impetus into V1, meaning that the structure of ON/OFF domains is intrinsically linked to the structure of V1 receptive fields (Smith et al., 2015; Tring et al., 2022). This arrangement of V1 ON/OFF domains is also elegantly explained as a function of the LGN afferents into V1: axons carrying responses of one polarity bundle together and project to the same cortical locations (Jin et al., 2008, 2011).

More recently, these patterns of functional organization have been proposed to emerge from fundamental mechanisms of sensory learning: experience-dependent selective sampling of thalamic afferents. Sampling from a greater density of afferent connections from LGN (which themselves are organized by spatial position, ocularity and polarity) also enlarges the cortical domains representing the same visual point, facilitating the segregation of afferents and their

cortical targets along different stimulus dimensions such as spatial position, ocularity, luminance polarity, and orientation (Najafian et al., 2022).

1.4.2 Columnar organization

In addition to spatial organization across visual cortex, there is evidence for a second, perpendicular dimension of organization along the depth of cortex: cortical columns. The first functional evidence of columnar organization was made by Hubel and Wiesel demonstrating that cells in the same cortical location but different depths were likely to share preferences for spatial position and orientation, but would otherwise likely engage in very different computations (Hubel and Wiesel, 1963, 1968, 1969).

In accordance with the standard laminar classification of neocortex, area V1 is traditionally divided in 6 horizontal layers, with a characteristic distribution of inputs and outputs across layers (Douglas and Martin, 2004). Feed-forward inputs from LGN arrive in layer 4, with collaterals to layer 6, and feedback inputs from other cortical areas arrive mostly in superficial layers. Feed-forward outputs to other cortical areas depart from layer 2/3, feedback outputs to the thalamus depart from layer 6, and outputs to other subcortical targets depart from layer 5 (Hubel and Wiesel, 1972; Callaway, 1998). In primates, this distinction between layers can be further refined: layer 4 of V1 is divided into sublayers 4A, 4B, 4Ca, and 4Cß. The main LGN inputs arrive in 4C, and segregate depending on the source: magnocellular LGN cells to 4Ca and parvocellular cells to 4Cß (Lund et al., 2003). Additional LGN inputs of the koniocellular kind arrive in layer 4A, 3, and 1 (Nassi and Callaway, 2009). A system of horizontal connections then links neurons with co-

oriented, co-axially aligned receptive fields, thus linking columns of neurons with similar response characteristics, such as preferred orientation (Bosking et al., 1997). These columns can be specific at a fine scale: Neurons with opposite preferences for stimulus direction have been found to be segregated with extraordinary spatial precision in three dimensions, with columnar borders one to two cells wide (Ohki et al., 2005).

However, while the idea of cortical columns is appealing and laminar differences have been well demonstrated, the idea of columns as functional units of cortex has its challenges. Expression of columnar organization can be highly variable between individuals, or even in within V1 of the same individual, thus complicating to what degree the term is indeed useful (Horton and Adams, 2005).

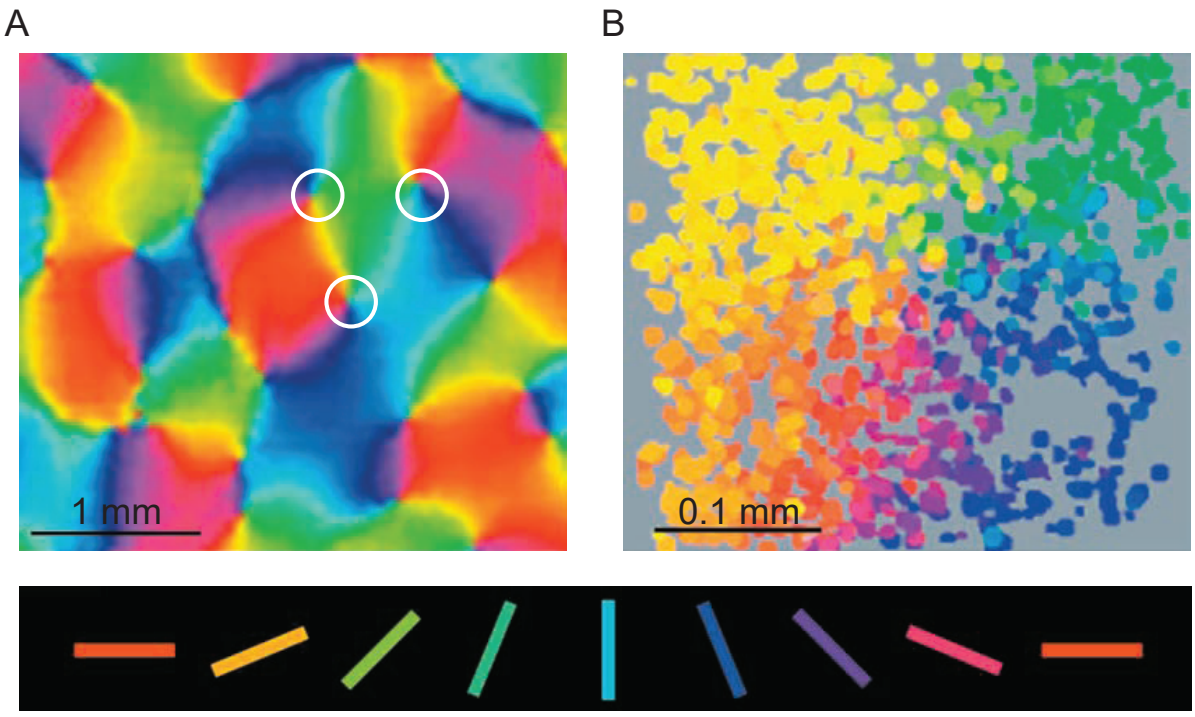

Figure 1.2: Organization of V1 orientation maps.

A: Mapping of orientation preference across the cortical surface. **B**: Labeling of individual cells by orientation. Image reproduced from Purves, Neuroscience 6th ed. (Purves et al., 2017).

1.5 Moving beyond classical functional characterization

The studies described above discovered some fundamental organizational and functional properties of V1 neurons, and although they may be oversimplifications of the true roles they play in vision, they are still the foundation of our understanding of visual cortex. The simplicity of these tuning properties makes them an extremely useful tool for us to comprehend the roles of single neurons in immensely complex systems. However, while the use of simple parametric stimuli was instrumental in laying the foundation for our understanding of the basic principles of visual processing in the brain and the role of V1 in this process, there were ultimately limits to the questions that could be asked with this approach.

There are several considerations that need to be taken into account in the paradigm shift beyond classical functional characterization. First is the design of stimuli that explore more complex visual features than bars or gratings. While it is trivial to design more complex stimuli, it is not trivial to design complex stimuli with features that can be interpreted to drive meaningful responses: while complex stimuli drive more complex neural responses, they also sacrifice the simplicity and ease of comprehension that make simple parametric stimuli appealing. The specific considerations about stimulus choice are discussed in section 1.6 below. Second, any set of

complex stimuli will necessarily have high spatial dimensionality which makes them prone to noise from eye movements, thus requiring good eye tracking. Previous studies avoided this issue by recording in anaesthetized animals, but this process only induces other complications and questions about whether observed activity will generalize to vision in normal contexts. As such, it is more informative to record in awake animals and account for eye movements – approaches to do this are outlined in section 1.7. Finally, the complexity of noise stimuli means they require more complex analyses to interpret the relationships between stimuli and neural responses, which can be solved through the application of models. The historical and current applications of computational models to better understand visual processing are discussed in section 1.8.

1.6 V1 characterization using complex stimuli

Obviously, the stimuli most closely eliciting responses informative about typical visual processing are natural scenes: images or movies of the subject's normal visual environment. However, while natural movies elicit the most realistic visual responses, they are so dynamic and complex that analyzing neural responses they elicit becomes intractable. On the other hand, neural responses elicited by simple parametric stimuli like bars or gratings are not predictive of responses to natural images or movies (David et al., 2004; Russ and Leopold, 2015) as natural stimuli drive visual neurons more strongly and have much lower signal-to-noise ratio (Touryan et al., 2005). Thus, gaining a better understanding of visual processing requires a class of stimuli that is complex and dynamic, containing a wide range of visual features that drive the same mechanisms as natural

scenes, but remains low-dimensional enough to maintain enough interpretability and stimulus control to be useful.

The stimuli that emerged to fit this description are varieties of noise stimuli, which provide a versatile and ecologically valid approach to studying the response properties, selectivity, and computational mechanisms of V1 neurons (Chichilnisky, 2001; Nykamp and Ringach, 2002). Noise stimuli offer a high level of complexity and diversity, making them suitable for exploring the wide range of selectivity and response properties exhibited by V1 neurons (Touryan et al., 2002). Because they potentially can drive combinations of visual features at the same time, they can also shed light on how V1 neurons combine and process information, contributing to the understanding of feature integration in the visual system. Furthermore, they can easily be adapted to match stimulus statistics reflecting certain aspects of natural scenes, making them more ecologically relevant (Coen-Cagli et al., 2012). Noise stimuli thus help bridge the gap between controlled laboratory experiments and the complexity of the real visual world, contributing to a more comprehensive understanding of visual processing in the primary visual cortex.

1.7 Effects of eye movements on V1

Saccades are rapid eye movements that involve shifting the gaze from one point of interest to another. Their purpose in vision is to redirect the fovea quickly and accurately toward specific objects or areas of interest in the visual field. Saccades play a critical role in visual exploration, allowing us to scan our surroundings, track moving objects, and bring important details into focus (Martinez-Conde and Macknik, 2008; Martinez-Conde et al., 2009).

Even when fixating on a particular visual target, these eye movements do not cease to happen; instead, saccades simply decrease in magnitude. These small saccades are often referred to as microsaccades, but they exist on the same continuum as regular saccades and should not be considered categorically different. They move the eyes just enough to prevent perceptual fading (Martinez-Conde et al., 2006) and possibly reposition receptive fields to enhance information about the visual target (Rucci et al., 2007; Poletti et al., 2013). Because the magnitude of these microsaccades is still greater than the resolution of many receptive fields but smaller than the resolution of most eye trackers, this means that they can be a major source of noise when displaying high-resolution stimuli.

The visual system transforms this constantly shifting visual input into a continuous visual experience by integrating the processing of retinal signals with information about eye movements. Such extra-retinal signals associated with saccades have been described in the LGN (Reppas et al., 2002), V1 (Martinez-Conde et al., 2000; Kagan et al., 2008; Rajkai et al., 2008) and higher-order visual areas (Sommer and Wurtz, 2006; Ibbotson and Krekelberg, 2011). In V1, saccadic modulation manifests as biphasic modulation of average firing rates composed of two distinct effects on stimulus processing: an initial stimulus-independent increase in firing rates followed by divisive gain suppression (McFarland et al., 2015), effects which are likely inherited from the LGN.

This means that eye movements significantly complicate studies of visual processing by 1) constantly changing the inputs into V1 receptive fields independently of presented stimuli, and 2) directly modulating visual activity independently of presented stimuli. One way the field has dealt

with these complications is by studying anesthetized animals which cannot move their eyes, but this solution also changes neural activity in complex ways and thus adds additional uncertainty. Another possible solution is presenting stimuli in the parafovea, where receptive field resolution is lower than the effects of fixational eye movements. This solution comes at the cost of studying the center of gaze, which has the highest spatial resolution and underlies some of the most important visual processing.

However, yet another solution is to not see these complications as problems to be avoided. In fact, activity observed in V1 can also be used to infer the actual position of the eyes. By comparing observed neural activity with predictions of neural activity for different possible eye positions, the most likely actual eye position can be inferred with high confidence and spatial resolution (McFarland et al., 2014). This facilitates studying visual processes in awake, behaving animals at fine scales, even during free viewing (Yates et al., 2021). The work presented in this thesis makes extensive use of this technique to refine data and treat eye movements as an integral component of vision, not just a source of noise.

1.8 Computational models of neural activity

Mathematical descriptions of neurons have always been desirable. As explained by David Marr, understanding complex systems such as the visual system requires not only an understanding of the neurophysiological implementation of processes, but also the underlying computations and the algorithms implementing them (Marr, 1982). For this reason, there is a long history of

neuroscientists devising mathematical descriptions, or models, of observed neural processes. These models can be broadly separated into two categories: Analytical and Statistical.

Analytical models relate variables together based on a physical understanding of the world. These models are typically generated by researchers who observe activity in neural populations, and then theorize possible computations these populations might perform that would explain the observed activity. Critically, while these models are made to describe observed data, they are not directly "fit" to the observed data. This top-down approach corresponds to forward engineering, which is the general goal of cognitive science, computer vision, machine learning, and artificial intelligence.

Statistical models relate predictor variables to a target outcome based on a non-physical model. These models are typically enabled by electrophysiology recordings carried to try to reverse-engineer the computations performed by the visual system by systematically estimating the neurophysiological properties of neural circuits. It is worth noting that although biophysical models such as Hodgkin-Huxley have an important place in the computational literature, I will be primarily discussing models concerned with spikes, operating under the assumption that they are the base computational unit of the cortex.

1.8.1 Analytical Models of V1

The first analytical models of circuits in the visual cortex date back to Hubel & Wiesel (Hubel and Wiesel, 1962), who described their results using the framework of a qualitative models of simple and complex cells. Their model for simple cells explains some of the key operations they

observed, such as how orientation selectivity could be obtained via selective pooling mechanisms over afferent LGN neurons with center-surround-like receptive field organizations aligned along a preferred axis of orientation. Similarly, their model for complex cells showed how the tolerance to position of the complex cells could be obtained by considering afferent cells with the same preferred orientation but with receptive fields at slightly different positions. Fifty years after it was originally proposed, these models continue to be a key element in both machine and biological models of object recognition.

Hubel & Wiesel's original complex cell model was later refined by Adelson & Bergen into the energy model (Adelson and Bergen, 1985), an influential model that could account for the phase-tolerance of cortical complex cells. At the algorithmic level, the model is based on the summation of quadrature pairs of spatiotemporal filters (i.e., simple cells) with the same orientation to build complex cells that are invariant to changes in the phase of the stimulus (and contrast reversal). Many subsequent papers have expanded upon this model, including by translating the basic energy model computations in frequency space to predict tuning properties in V1 population responses (Mante and Carandini, 2005), and by separating the spatiotemporal filters for each eye to describe the energy computations underlying disparity selectivity in V1 (Ohzawa et al., 1997; Read and Cumming, 2003a).

However, there is still a lack of formal analytical models of visual circuits, in that there are few grand theoretical models that can be used to make predictions for experimentalists to test.

1.8.2 Statistical models of V1

Much historical work has primarily relied on spike-triggered averaging (STA) (Reid et al., 1997; Chichilnisky, 2001), or reverse correlation analysis (Ringach and Shapley, 2004). Effectively, these studies consisted of presenting varieties of noise stimuli at time bins t (which can be expressed as the vector s_t), observing the elicited neural responses r_t (typically a spike count measured via extracellular recording), and averaging the stimuli preceding neural responses across N spikes (eq 1).

$$STA = \frac{1}{N}\sum_{t=1}^{N} \overrightarrow{s_t} \cdot r_t \qquad \text{eq. 1}$$

However, a single stimulus feature is usually not sufficient to describe a neuron's response characteristics. Many V1 neurons are sensitive to several features and integrate their selectivity to these features nonlinearly, meaning that STA analysis will cause overlapping features to cancel each other out and result in uninterpretable receptive field maps.

For these reasons, spike-triggered covariance (STC) analysis emerged as a popular extension of the STA (De Ruyter Van Steveninck and Bialek, 1988; Touryan et al., 2002; Rust et al., 2004). STC aims to detect differences in variance between the distribution of all stimuli in the presented set, and the distribution of the subset of stimuli that elicited spikes. The STC matrix is computed as the sum of the outer product of the spike-triggered stimulus vectors with the STA subtracted out, and is equivalent to a second-order Wiener or Volterra kernel.

Extracting the eigenvectors of this matrix (effectively applying PCA) reveals the primary vectors of variance within the stimulus space that drive the observed neural responses.

Eigenvectors with eigenvalues significantly larger or smaller than the eigenvalues of the raw stimulus covariance correspond to stimulus dimensions along which the neural response is enhanced or suppressed. In other words, it recovers a base set of filters which together describe the cell's receptive field. Although alternative techniques exist that are applicable under more general stimulation (Paninski et al., 2004; Sharpee et al., 2004; Pillow and Simoncelli, 2006; Park and Pillow, 2011), STC analysis has retained considerable popularity, just like the STA, because of its relative computational simplicity (Samengo and Gollisch, 2013; Sandler and Marmarelis, 2015).

The accuracy of STA/STC filter estimates depends on three elements: (1) the dimensionality of the stimulus space, (2) the total number of spikes collected, and (3) the signal to noise ratio of the observed spikes. In other words, it works best when the stimulus dimensionality is low, there are many spikes, and those spikes are well isolated. However, in many experimental conditions, observed neural behavior includes significant amounts of noise due to uncontrolled variables and natural stochasticity in spike generation.

1.8.3 Using maximum likelihood estimation to directly fit statistical models

However, modern computer power enables us to directly use neural data to identify the underlying parameters relating a neuron's inputs to its responses. Most approaches to model parameter optimization in neuroscience fall under the category of *maximum a posteriori* (MAP) estimation (Butts, 2019). For a given mathematical form of a model, the best description of the

data can be defined as the most probable choice of parameters that generated the data. The point-process log-likelihood per spike is given by:

$$LL = \frac{1}{N_{spk}} \left[\sum_{t_s} log_2 \Pr\{spk|t_s\} - \sum_t \Pr\{spk|t\} \right] \quad \text{eq. 2}$$

where $\{t_s\}$ are the observed spike times, and N_{spk} is the number of observed spikes. Parameter values corresponding to the linear terms can be simultaneously optimized by maximizing the likelihood of the model given the data. All model estimation uses gradient ascent of the log-likelihood (*LL*), which, given the spiking nonlinearity shown above, has no local maxima other than the single global maximum (Paninski et al., 2004). Also, for a given choice of model parameters, both the *LL* and its gradient can be directly calculated, resulting in efficiently computed ascents to the globally maximum likelihood.

MAP estimation is equivalent to maximum likelihood estimation if there are no assumptions made about the model parameters. If the noise stimuli used can be assumed to have a Gaussian distribution, MAP is also equivalent to spike-triggered approaches minimizing the mean-squared error between predicted and observed responses. Additionally, maximizing the LL is equivalent to maximizing the single-spike information between the model prediction and data (Sharpee, 2013). As such, MAP estimation is a highly flexible tool that has been used to fit parameters in a number of different model architectures to describe neural activity.

LN models. Until relatively recently, Linear-Nonlinear-Poisson (LNP) models have dominated sensory neuroscience (Nykamp and Ringach, 2002; Simoncelli et al., 2004). LNP models consist of three stages: The first stage consists of a linear filter *k*, which is a vector

describing a feature the neuron responds to, i.e. its receptive field, in terms of the dimensions of the presented stimulus *s*. The output of this filter then passes through a nonlinear function f[.]. This response is then passed through a Poisson process to generate spikes (eq 3).

$$P(spike) \propto f(\vec{k} \cdot \vec{s}) \qquad \text{eq. 3}$$

These models can be easily fit by passing a filter estimated through spike-triggered averaging through a spiking nonlinearity. As a result, they were a powerful initial tool to fit stimulus filters simultaneously with linear weights on other experimental observables (Weber and Pillow, 2017), such as activity observed in related neurons (Butts et al., 2016) or the neuron's own past history of spiking (Paninski et al., 2004; Truccolo et al., 2005). Indeed, LN models have been used throughout the visual system with great success, from the retina (Chichilnisky, 2001), to the LGN and V1simple cells (Carandini et al., 2005). However, their reliance upon a single linear filter means LN models inherit many of the limitations of STA – an inability to capture selectivity to overlapping features, a core component of complex cells' phase invariance. As a result, LN models have been largely replaced by more powerful models, which also handle responses to more complex stimuli (David and Gallant, 2005; Sharpee, 2013).

A particular realization of this process, known as the 'Nonlinear Input Model' (NIM), uses a variable number of rectified linear units (ReLU), then constrains the output to be either positive or negative, thus approximating excitation or inhibition (McFarland et al., 2013). This model architecture acts as a universal function approximator due to the properties of ReLU subunits

(Kriegeskorte, 2015) and can capture arbitrary computations linking stimuli and observed responses. Due to its versatility, this architecture is the basis of much of the modeling presented in this thesis.

1.8.4 Neural network models: the state of the art

Given the recent, rapid acceleration of developments in machine learning, computational neuroscience has likewise applied new tools and achieved new discoveries. As the hierarchy of the visual system served as early inspiration for deep models, this is certainly gratifying: The brain can be described as a deep and complex recurrent neural network, and thus it can seem appealing to try and apply neural network tools to understand it.

However, while these new tools are extremely appealing in their ability to capture neural responses, they largely do so at the expense of interpretability. While the simple parametric studies described above were certainly limited in their ability to probe many features of neural processing, they did yield clear, interpretable predictions. Many deep networks today can describe neural activity with astounding accuracy but do so by replacing the black box of the brain with a black box in silica. Of course, there is still much to be learned even by simply understanding which neural populations are able to represent certain types of information even without understanding the exact algorithms or implementation used by the brain. However, there is still a significant gap in the current state of the field for an approach that forms a happy medium between model complexity/power and interpretability. More closely approaching such a goldilocks model of V1 is one of the core motivators of this thesis.

1.9 Thesis overview

The overarching goal of this dissertation is to identify computations performed in V1 by fitting statistical models to electrophysiological data and interpreting the model properties. My approach both yields direct insight into how V1 neurons represent information and provides mathematical descriptions of how V1 computations can be effectively described, building up the components of a unified model of V1 function. To this end, I present different ways to identify computations performed by V1 in increasingly complex stimulus contexts.

In chapter 2, I apply state-of-the-art statistical and machine learning approaches to identify high-resolution components of V1 receptive fields as mapped using 1-dimensional luminance stimuli and analyze the properties of these models to quantify the computations these cells perform. I present model-based measures of feature selectivity and relate them to classical measures of V1 tuning, showing how models can not only provide the same information gained by classical measures, but explain the actual computations performed by V1 even better. Specifically, I will show how multiple overlapping inputs can allow cells to represent high-resolution information while also responding to full-field inputs, an arrangement that could not be discovered using classical stimuli.

In chapter 3, I expand upon this process by applying similar modeling techniques to data from 1-D luminance stimuli presented over multiple binocular disparities to elucidate how V1 neurons combine binocular inputs. I identify computations underlying disparity tuning by fitting single-unit and shared nonlinear models for V1 cells, and show how spatial convolutions can amplify selectivity to disparity. I then explain how interactions between excitatory and inhibitory

signals further enhance disparity selectivity without interfering with selectivity to particular spatial patterns. I finally show how these results are an efficient implementation of classical hypotheses of disparity selectivity in V1.

In chapter 4, I present novel data recorded during presentation of 2-D achromatic and color stimuli at the fovea and further expand my modeling approaches to identify the computations through which V1 neurons represent chromatic inputs at the center-of-gaze. I provide the first spatiochromatic measurements of the scale of V1 processing at the center of gaze, showing that foveal V1 neurons have a variety of receptive fields subtending the same spatial regions, thus facilitating hyperacuity. I then characterize the computations used by V1 color-sensitive neurons to combine cone inputs to represent spatial and chromatic stimuli, showing that luminance signals carry much higher spatial resolution than color inputs. Finally, I show how RF subregion structure can allow foveal cells with large RFs to maintain high acuity, as well as jointly represent color and spatial opponency.

2 Model-based characterization of the selectivity of neurons in primary visual cortex

This work has already been published and was written together with Dan Butts and Bruce Cumming – see (Bartsch et al., 2022).

2.1 Introduction

The activity of neurons in primary visual cortex (V1) set the foundation for the cortical processing of vision. Thus, detailed characterization of their stimulus selectivity is of great intrinsic utility. Since the initial discovery of their basic selectivity to oriented bars (Hubel and Wiesel, 1962), most V1 neurophysiology has used precise manipulations of simple parametric stimuli such as flashed or drifting gratings (Movshon et al., 1978c; Maunsell and Newsome, 1987; Alonso and Martinez, 1998) to make measurements of dominant aspects of V1 visual selectivity, such as receptive field (RF) size and location, orientation tuning, and spatial frequency and temporal frequency tuning (Hubel and Wiesel, 1962; DeAngelis et al., 1994; Sceniak et al., 1999; Chen et al., 2009). Because how V1 processes these simple parametric stimuli has limited power in predicting their responses to more complex time-varying stimuli (David and Gallant, 2005; Sharpee, 2013; Butts, 2019), a large number of increasingly sophisticated models has been used to understand their responses to more complex stimuli (Butts, 2019), using techniques ranging from spike-triggered covariance (Touryan et al., 2002; Rust et al., 2005) up through modern machine-learning methods (e.g. (Cadena et al., 2019)). While the necessity of more sophisticated models to capture neural computation in complex stimulus contexts suggests that these previous measures of

V1 selectivity offer an incomplete picture of its processing of natural vision, it is not clear whether more sophisticated approaches are simply measuring different aspects of processing, and exactly what is gained by them relative to more canonical measurements.

Indeed, the success of more sophisticated models in explaining V1 neural responses is often at the expense of the interpretability that defined more canonical approaches to characterizing V1 neurons (Butts, 2019). Likewise, few studies analyzing models to gain insight into neural properties have attempted to directly relate modern nonlinear modeling approaches to classical measures of cell (although see (Almasi et al., 2020)). Here, we focus on the most interpretable of these modern models: "subunit" models such as the LNLN cascade (Park and Pillow, 2011; McFarland et al., 2013; Vintch et al., 2015), which consist of a linear combination of nonlinear spatiotemporal stimulus filters (or "subunits") that are each selective to a particular stimulus feature that contributes to the neuron's response (Vintch et al., 2012, 2015; McFarland et al., 2013; Park et al., 2013). The array of spatiotemporal filters generated by these models are determined by optimization of model parameters using data recorded during any sufficiently general spatiotemporally varying visual stimulus paradigm, and captures the set of stimulus features that affect the neuron's response. We present several measures that can then be derived from the resulting models, which can be directly related to canonical aspects of neural tuning, while capturing other interpretable aspects of its selectivity. Notably, while we apply these measures to the models measured in this particular context, such measures can be generalized to any "image-computable" model of V1 neurons (Butts, 2019). Because such more sophisticated characterizations of V1 processing implicitly capture V1 neural selectivity to multiple features as

well as interactions between selectivity to related features, the resulting descriptions of V1 computation go beyond what can be achieved with simple parametric stimuli, and set a new bar for measurements of V1 neural selectivity.

2.2 Materials and methods

2.2.1 Neurophysiology experiments.

We used previously collected recordings, as described in detail in (McFarland et al., 2014). Briefly, multi-electrode recordings were made from primary visual cortex (V1) of two awake head-restrained male rhesus macaques (Macaca mulatta; 13- to 14-years old) while the animals performed a simple fixation task, where they were required to maintain gaze within a small window around a fixation target to obtain a liquid reward after each completed 4-s trial. Eye position was monitored continuously using scleral search coils in each eye, sampled at 600 Hz, and a model-based eye-position correction algorithm (see below) was applied to accurately recover eye position at sufficiently high spatial resolution (McFarland et al., 2014, 2016). Trials during which the monkey broke fixation were discarded from analysis. Additionally, the first 200 ms and last 50 ms of each trial were excluded from analysis to minimize the impact of onset transients.

2.2.2 Recordings

In one animal, we implanted a 96-electrode planar Utah array (Blackrock Microsystems; 400 μm spacing), in the other, a linear electrode array (V-probe, Plexon; 24 contacts, 50 μm spacing) was passed through the dura at the start of each day. Spike waveforms were detected

online using Spike2 software (Cambridge Electronic Design) and recorded at 32 kHz (Utah array) or 40KHz (V-Probe) sampling rate. Single-units (SUs) were identified through off-line spike-sorting with custom software. Spike clusters were modelled using Gaussian mixture distributions that were based on several different spike features, including principal components, voltage samples and template scores. The features providing the best cluster separation were used to cluster single units. Cluster quality was quantified using a variety of measures including 'L-ratio', 'isolation distance', as well as a variant of 'd-prime'. Only spike clusters that were well isolated using these measures—confirmed by visual inspection— and had average firing rates during stimulus presentation of at least 5 spikes per second were used for analysis. Voltage signals were also separately low-passed at 100 Hz and used to compute the inverse current-source density (iCSD) profile of each recording, in order to categorize each isolated single unit by layer (Pettersen et al., 2006; Chen et al., 2011). All protocols were approved by the Institutional Animal Care and Use Committee and complied with Public Health Service policy on the humane care and use of laboratory animals.

2.2.3 Stimuli

We presented a ternary bar noise stimulus, consisting of random patterns of black, white, and grey bars (matching the mean luminance of the screen). These stimuli were displayed on a cathode ray tube (CRT) monitor at a 100 Hz refresh rate; the monitors subtended 24.3°×19.3° of visual angle. RFs were manually mapped on-line, and stimuli were then centered over the mapped RF location. Stimuli consisted of 36 individual bars that had a width of 0.057° for the Utah array

recordings and ranged from 0.038° to 0.1° for the linear array recordings, depending on the RF sizes of the recorded neurons. The bar length was adjusted to yield square stimulus patches such that stimuli subtended between 1.368°-3.6°. Bar patterns were uncorrelated in space and time. The probability of a given bar being non-grey (i.e., black or white) was set to 12% (i.e. 88% grey, relatively sparse), although in several experiments we used a denser distribution (33% grey), which yielded similar results. For the linear array recordings, the orientation of the bars was chosen to be within 15° of the preferred orientation of the majority of units recorded in a given session. In Utah array recordings, where the neurons typically had a full range of preferred orientations due to being located across multiple orientation columns, we performed the experiments with two different bar orientations (vertical and horizontal) that alternated on a block-by-block basis, where blocks were about 10 minutes of recording time in duration. In this case we modeled the orientation that best drove neural activity, and the properties of cells recorded using this approach were comparable to those recorded during presentation of more optimally oriented bar stimuli.

2.2.4 Model-based eyetracking

Due to fixational eye movements, the stimulus projected on the retina is shifted over time, which can result in mis-estimation of model parameters (McFarland 2014; 2016). As a result, we estimated the precise position of the eye perpendicular to our 1-dimensional stimuli as a function of time using model-based eye tracking (McFarland et al., 2014). Briefly, the algorithm uses the observed firing rates during each fixation period to estimate the probability of each potential eye position, and — with enough neurons simultaneously recorded — these probabilities can be

combined to accurately infer the most likely eye position, which can be accurate to arc-min resolution. Using this estimated position, we then adjusted the presented stimulus by the eye shifts, and used the resulting adjusted stimulus to fit the models, as described in detail in (McFarland et al., 2014).

2.2.5 Model structure

We used the Nonlinear Input Model (NIM) to model V1 neurons (McFarland et al., 2013). The NIM (Fig. 2.1A) operates on the stimulus $\mathbf{s}(t)$ with a number of linear-nonlinear (LN) subunits, each consisting of a stimulus filter $\mathbf{k}$ (Fig. 2.1B) and a rectifying nonlinearity $f(.)$. Subunit outputs are then weighted by either +1 or -1 (depending on whether each is excitatory or inhibitory, respectively), and summed together and passed through a final spiking nonlinearity $F[.]$ to result in a predicted firing rate $r(t)$: given stimulus $\mathbf{s}(t)$ is given as:

$$r(t) = F[\sum_n w_n f_n(\mathbf{k_n} \cdot \mathbf{s}(t))] \quad (2.1)$$

where w_n is +1 or -1 as described above. We use a "ReLU" (rectified linear) function for the subunit nonlinearity $f()$, and a softplus nonlinear function for the spiking nonlinearity: $F(x) = \log[1+\exp(x)]$ as in our previous work (McFarland et al., 2013).

We initially fit NIMs to each neuron using standard maximum *a posteriori* optimization in the context of sparseness (L1) and smoothness (Laplacian) regularization. Both the number of subunits and the regularization parameters were optimized through nested cross-validation, as described previously (McFarland et al., 2013; Butts, 2019). We then leveraged the resulting collection of models to yield additional model improvements, using dimensionality reduction on

the spatiotemporal filters pooled across all models. Specifically, this database (over all neurons in the dataset fit using standard NIMs) consisted of 2061 filters, which were then centered to remove any spatial shifts between them. We computed the principal components (PCs) of this filterbank, finding that 80% of its variance could be explained by the first 50 PCs, with diminishing returns from including more (see Appendix Figure 2A1A). We then applied these top 50 filters to the spatiotemporal stimulus **s**, resulting in a 50-dimensional "dimensionality-reduced" stimulus in place of the 360-dimensional "full" stimulus that was composed of 36 spatial dimensions and 10 lags.

The resulting dimensionality-reduced stimulus was used in eq. 1 in place of the full stimulus, and NIMs were refit using the filtered, dimensionality-reduced stimulus (which we refer to as Principal Component NIMs, or PCNIMs). The fit model subunits thus had a set of coefficients corresponding to the weights on the PCs, and the full spatiotemporal filters could be recovered by matrix multiplication. This allowed for the potential inclusion of many more subunits without overfitting (see below). Furthermore, because the PCs did not reflect noise in the filter estimation that was not shared among many filters, it was not necessary to apply any regularization on subsequent stages of fitting. Representing model filters using PCA yielded better model performance (Appendix Fig. 2A1C) than the standard NIM approach, including finding more subunits that improve model performance that would have been suppressed by regularization penalties for the standard NIM.

In addition to using the standard structure of the NIM with rectified subunits, we also compared LNLN cascades with linear and quadratic subunit nonlinearities $f_n(.)$ (eq. 1) (McFarland

et al., 2013; Park et al., 2013; Park & Pillow, 2011) (Fig. 2.1C). This "quadratic" model is also commonly used to model V1 neurons because it embodies the "energy model" (Adelson and Bergen, 1985; Mante and Carandini, 2005), although it is less flexible in approximating more general nonlinearities (McFarland et al., 2013; Butts, 2019). We applied a subspace approach to fitting quadratic models as well, but found that both full and subspace quadratic models converged on equivalent subunit structures with similar cross-validated performance. As a result, we report results from the more straightforward quadratic model fits that do not rely on the filter subspace.

2.2.6 Model parameter optimization

The parameters of the filters are fit and optimized through a maximum-likelihood approach as described in detail (McFarland et al., 2013). This approach maximizes the log-likelihood of the model, which is given by the equation

$$LL = \sum_t (R_{\text{obs}}(t) \log r(t) - r(t)) \tag{2.2}$$

where *r(t)* is the predicted firing rate and $R_{\text{obs}}(t)$ is the observed spike count, assuming a Poisson noise model. The goodness of fit of the models is assessed using 5-fold cross-validation: with model parameters fit using 80% of the data, and performance assessed using the *LL* of the models on the remaining 20% of the data.

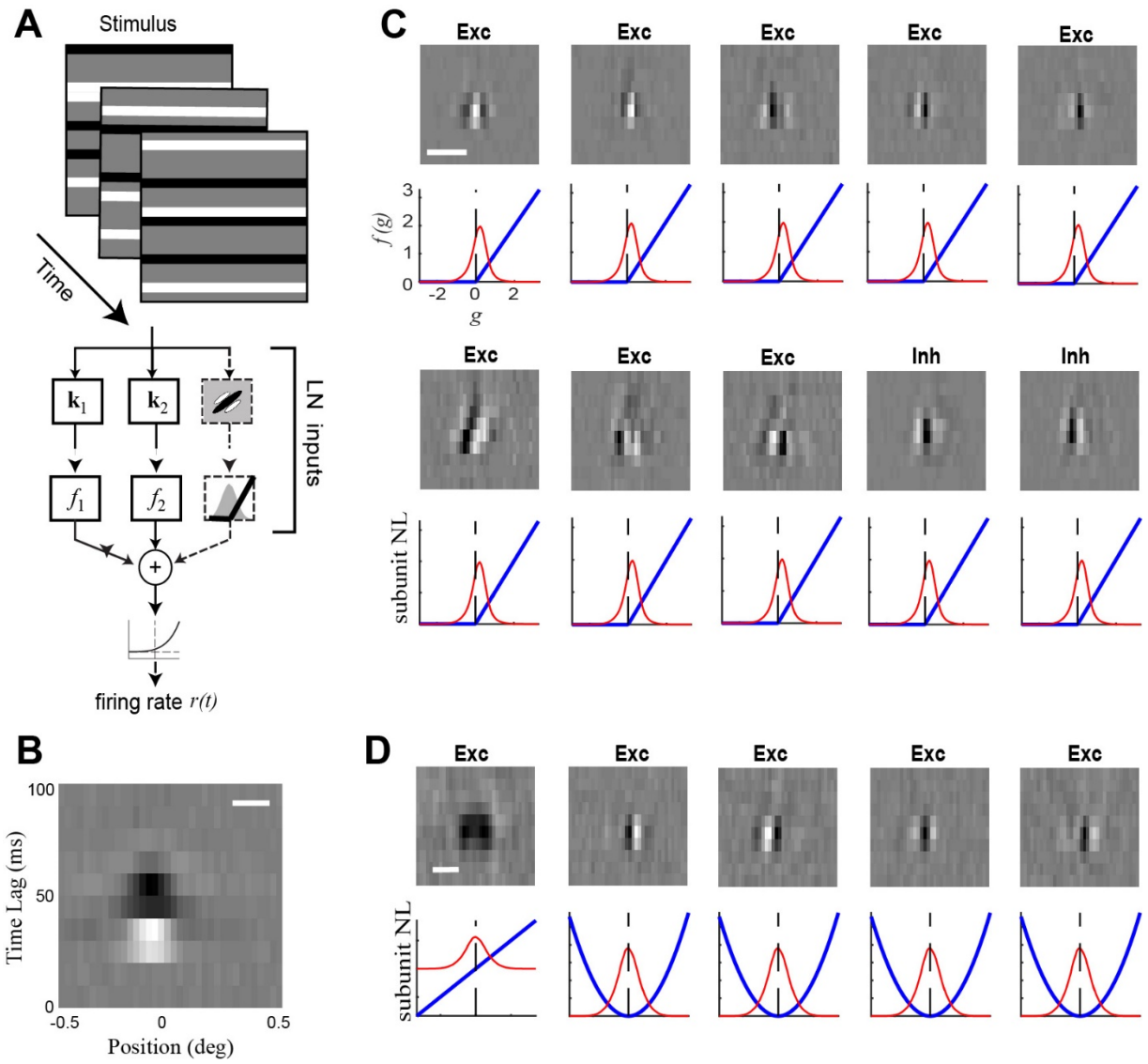

Figure 2.1: Extracting RF components using stimulus models.

A, Schematic of the nonlinear LNLN cascade stimulus processing models. Stimulus frames (top) are passed through spatiotemporal filters **k**, the outputs of which are then passed through a rectifying nonlinearity. The outputs of these "LN subunits" are then summed to give the 'generating signal' $g(t)$, which is transformed into a firing rate by the spiking nonlinearity

(bottom). In the case of the PCNIM, filters $\mathbf{k}_i$ were not fit using the full stimulus space, but rather its low-dimensional projection into the space outlined by the principal components. After fitting, they were retranslated into full space for interpretation. ***B,*** Example spatiotemporal filter. ***C,*** Example of filters fit using the PCNIM with 8 excitatory and 2 suppressive filters. The nonlinearity associated with each filter is shown below (blue), as is the distribution of stimulus projections onto each filter (red). ***D,*** Example of filters fit by a GQM architecture with 4 excitatory quadratic filters. Nonlinearities are again displayed below each filter in blue. The scale bars in panels B-D indicate 0.1 degrees of visual angle.

2.2.7 Model hyper-parameter optimization

In addition to the parameters corresponding to the spatiotemporal filters, each model had several "hyper-parameters" corresponding to other aspects of the model architecture, including the number of filters used and the ratio of excitatory to inhibitory filters (E-I ratio). The number of filters and E-I ratio were determined through nested cross-validation across a range of possible combinations, refitting model parameters for each new combination and choosing the combination yielding the best cross-validated performance. Note that because subunits were fit in a subspace, they inherited the regularization of the models that were used to generate the filter subspace (see above). When fitting quadratic models, regularization hyper-parameters were optimized through nested cross-validation using the same process used to generate the models used for the NIM models that filter subspace was based on.

2.2.8 Estimating spatial frequency tuning using forward correlation.

To compare model-based measures of spatial frequency (SF) tuning to a model-free measure, we used forward correlation. We first extracted spatial frequency spectra of each stimulus frame via fast-Fourier transform (FFT) and ranked frames by power in each resulting frequency bin. For each frequency, we then extract two peri-stimulus time histograms (PSTHs) corresponding to frames with power at that frequency in the top 30th percentile and in the bottom 30th percentile. The difference between these two PSTHs provides a model-free estimate of the neuron's SF tuning directly comparable to those obtained using classical measures of SF tuning. We defined the cell's preferred frequency as the peak of this curve.

2.2.9 Estimating nonlinearity across space using forward correlation.

We also used forward correlation to validate measures of nonlinearity, computing the average response following the presentation of black, white, or gray bars at a given spatial position: $\bar{r}_{black}(x)$, $\bar{r}_{white}(x)$, and $\bar{r}_{gray}(x)$, respectively. As described in the text, the phase-reversal correlation as a function of spatial position is then given by:

$$PRFC(x) = \frac{|\bar{r}_{black}(x) - \bar{r}_{white}(x)|}{2 \max [\bar{r}_{black}(x) - \bar{r}_{gray}(x), \bar{r}_{white}(x) - \bar{r}_{gray}(x)]} \quad (2.3)$$

This measure can also be summed across space, weighted by the spatial power (cumulative variance of model filters across time, as described in the results), to result in a single PRFC measure for a given neuron.

2.3 Results

2.3.1 Estimation of a neuron's spatiotemporal feature space

In order to characterize the stimulus processing of V1 neurons, we fit the Nonlinear Input Model (NIM) (McFarland et al., 2013) to a database of 122 V1 neuron responses recorded in the context of noise stimuli consisting of ternary bars aligned to preferred orientation (Fig. 2.1A) (McFarland et al., 2013, 2014) (see Methods). The resulting subunit model of a given V1 neuron provides a list of spatiotemporal stimulus filters that modulate the neuron's response, as well as a positive or negative weight corresponding to whether the feature was excitatory or suppressive (Fig. 2.1B). As described above, these model components (Fig. 2.1C) contain rich information about the selectivity of the neuron being modelled, and here we are interested in analyzing the properties of these filters to extract such information.

We first analyzed the spatial profile of the filters: computing the amount of power (i.e., variance across time lag) at each spatial position, which defines the spatial profile for each spatiotemporal filter (Fig. 2.2A, B). The overall spatial profile of the neuron is then defined as the average of its filter profiles, weighted by the relative contribution of each filter's output to the overall predicted firing rate (Fig. 2.2B, bottom panel). To associate a single "width" to a given spatial profile, we defined the width of a given profile using the spatial extent between 30% and 70% of the cumulative spatial power are under the profile curve. The same procedure was applied to assess each neuron's temporal profile by computing power across spatial positions at each time

lag; we define temporal onset and offset of each temporal profile using the same cumulative power thresholds as were used for the spatial profiles.

This analysis reveals that, in a large fraction of neurons, there was a significant difference between extent of each neuron's summed spatial profile and those of its individual subunits. Across the population, 31/122 cell RFs were composed of filters with a mean width less than 75% of the total RF width (Fig. 2.2C). This discrepancy can be explained by a scatter of the locations of a given neuron's filters (Fig. 2.2B), such that the full profile of the neuron results from the combination of the spatial extent of its component subunits and the spatial scatter of these elements. This model-based observation is not readily accessible using direct experimental measurements of RF spatial potential and scale, which produce a single scale and location or preferred phase (Carandini et al., 1997a; Kagan et al., 2002; Mante and Carandini, 2005; Touryan et al., 2005).

This observed scatter is unlikely to be caused by our eye-tracking procedure. First, the amount of scatter is neuron-dependent: neurons with small RFs display less scatter, even though inaccurate eye tracking should make scattering of the filters in such highly localized RFs more salient and disproportionately broaden them. Second, we have already corrected for eye-position estimates in the data we use using our model-based eye tracking procedure as previously reported (McFarland et al., 2014), and the observed scatter is at a much larger scale than the arcminute accuracy of this procedure. Finally, the spatial scatter of subunits is observed in our quadratic model fits (Fig. 2.2F), as well as in previous studies using other datasets and modeling approaches

(e.g., spike-triggered covariance) (Rust et al., 2005; Lochmann et al., 2013), suggesting that it is not dependent particular choice of model architecture.

Because individual subunits in the NIM are rectified, subunits are implicitly either excitatory (only generating positive or zero output) or suppressive (with a negative weight so its output can only be less than or equal to zero) (McFarland et al., 2013). Such computational distinctions are difficult to assess using classical characterization methods, and reveal systematic differences in the selectivity of putative excitatory and inhibitory elements. For example, suppressive subunits typically had an increased width compared with excitatory subunits (Fig. 2.2G; *sign-rank test, p*=0.0021). This finding is consistent with intracellular recordings that found that the tuning of inhibitory V1 neurons typically have broader spatial profiles than excitatory V1 neurons (Haider et al., 2012).

To verify that our observations were not dependent on the specific form of the model, we applied the same analysis to models fit using quadratic subunits (see Fig. 2.1D for an example model) (Park and Pillow, 2011; McFarland et al., 2013, 2014). Measurements we applied to the quadratic models in general displayed the same overall trends for filter width (Fig. 2.2E) and scatter in subunit location (Fig. 2.2F) as the PCNIM, indicating that these results do not depend solely on our model architecture. The fact that filter scatter appears more pronounced in the NIM can be explained by the quadratic model's arrangement of filters into quadrature pairs that approximates spatial invariance of V1 neuron selectivity (Lochmann et al., 2013; Sharpee, 2013; Almasi et al., 2020). Because the quadratic models also underperformed the NIM at predicting neural responses (Supplemental Fig. 2A1D), we take this as indication that one aspect of the

increased flexibility with the NIM as a general function approximator (Kriegeskorte, 2015; Butts, 2019) is its ability to identify more spatially localized feature selectivity, and that this more localized feature selectivity allows us to more clearly see trends present in the quadratic models.

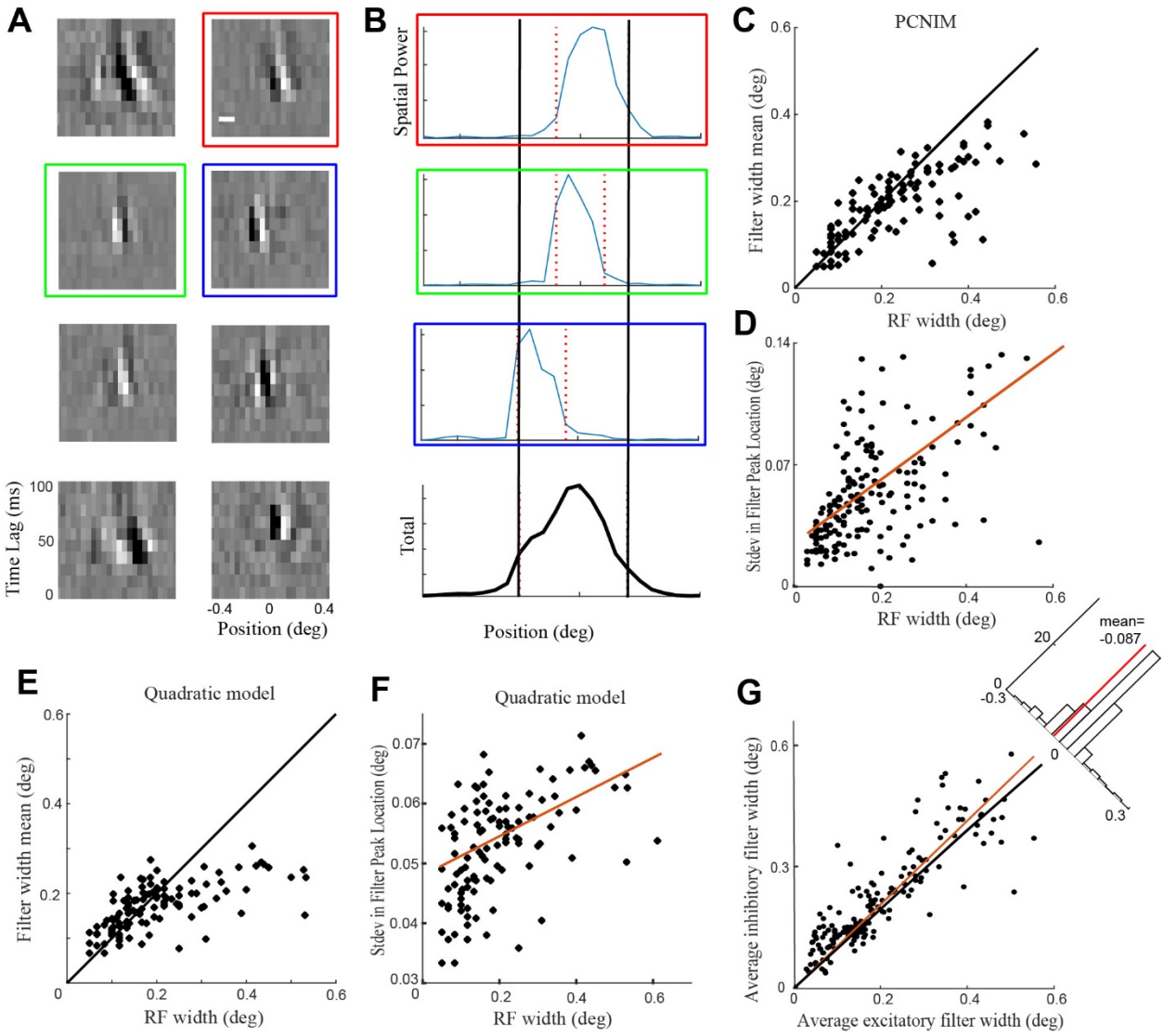

Figure 2.2: The PCNIM reveals high-resolution RF components.

A, The 8 most highly weighted subunit filters for an example neuron, plotted as a function of spatial position (horizontal axis) and temporal lag (vertical axis). **B**. Three example spatial power profiles (top) for the highlighted filters (colored boxes) in A, compared with the total spatial power profile of the model (black, bottom). Vertical lines indicate the edges of each filter, which is used to determine filter width. **C**, Population results comparing the RF width (i.e. width of the total spatial power profile of each neuron) with it mean subunit filter width. **D**, Population results comparing RF width and standard deviation of filter locations in excitatory filters, demonstrating that neurons with larger RFs typically have more spatial scatter in the locations of individual filters. **E, F**, Same as C and D, respectively, showing mean and standard deviation of filter widths in quadratic model fits. **G**, Average width of each model's excitatory and inhibitory filters. Inhibitory filters are overall wider than excitatory filters.

2.3.2 Measuring spatiotemporal selectivity

A dominant conceptual model of V1 neuron selectivity is that it extracts local information about the spatiotemporal content of the stimulus (Adelson & Bergen, 1985; Mante & Carandini, 2005), and a main reason classical tests of V1 tuning involve using measurements of responses to gratings. Here, we can instead extract information about such tuning by Fourier analysis of the model subunits.

Applying two-dimensional Fourier transforms to each subunit (Ringach and Shapley, 2004) yields a two-dimensional power frequency spectrum (Fig. 2.3A). Summing this spectrum along either axis generates a tuning curve representing the subunit's selectivity to either spatial or

temporal frequency. We define the neuron's preferred frequency by the peaks of the relevant subunit spectra. To then determine the model's overall frequency tuning, we average the tuning spectra across subunits (again weighted by each subunit's contribution to the model output) (Fig. 2.3B). The resulting tuning curves as well as peak tuning measures can be directly compared to classical measures, while also allowing them to be decomposed into each subunit's contribution.

To validate the model-based measures with experimental data in the absence of direct measurements of the classical tuning curves using gratings, we used a model-free estimation of spatial frequency tuning using forward-correlation of the spike train of each cell (Spratling, 2012; Tanabe and Cumming, 2014), where we extracted the spatial frequency components of each stimulus frame and identified the frequency components most correlated with the observed neural response (see methods). In applying this procedure to both the observed and the model-predicted firing rates, we found that the two agree well (Fig. 2.3D) as statistical tests did not detect any difference in distributions (*paired t-test*, p=0.8577), indicating that the models are successfully capturing the cell's SF tuning.

Model-based characterization also can address how SF tuning relates to other properties in our models. For example, inhibitory subunits generally have selectivity to lower SFs (Fig. 2.3C, *sign-rank test*, p=0.0061), in line with previous results indicating that inhibitory inputs into V1 tend to act on broader spatial scales (Ringach et al., 2003; Haider et al., 2012; Taylor et al., 2021). Additionally, we found a positive relationship between subunit SF selectivity and latency (Fig. 2.3E, Pearson's correlation, r=0.3407, 95% CI=[0.1656, 0.4950], p=0.0002), consistent with

(Mazer et al., 2002), indicating that low-frequency features have lower onset latencies and thus may be processed faster.

Using a similar approach applied to the full spatiotemporal spectra, we inferred V1 neuron direction selectivity (DS) from our models by extracting the direction selectivity index (DSI) (Priebe and Ferster, 2005) from each filter's 2DFFT (see Fig. 2.3A, red lines in bottom panels). We found that subunits of the same model exhibit a variety of spatial frequency sensitivities and DS strengths (Fig. 2.3G), suggesting that some direction-selective cells will be modulated by stationary inputs. We additionally tested for relationships between DS strength and other tuning properties including SF selectivity and RF size but did not find significant relationships.

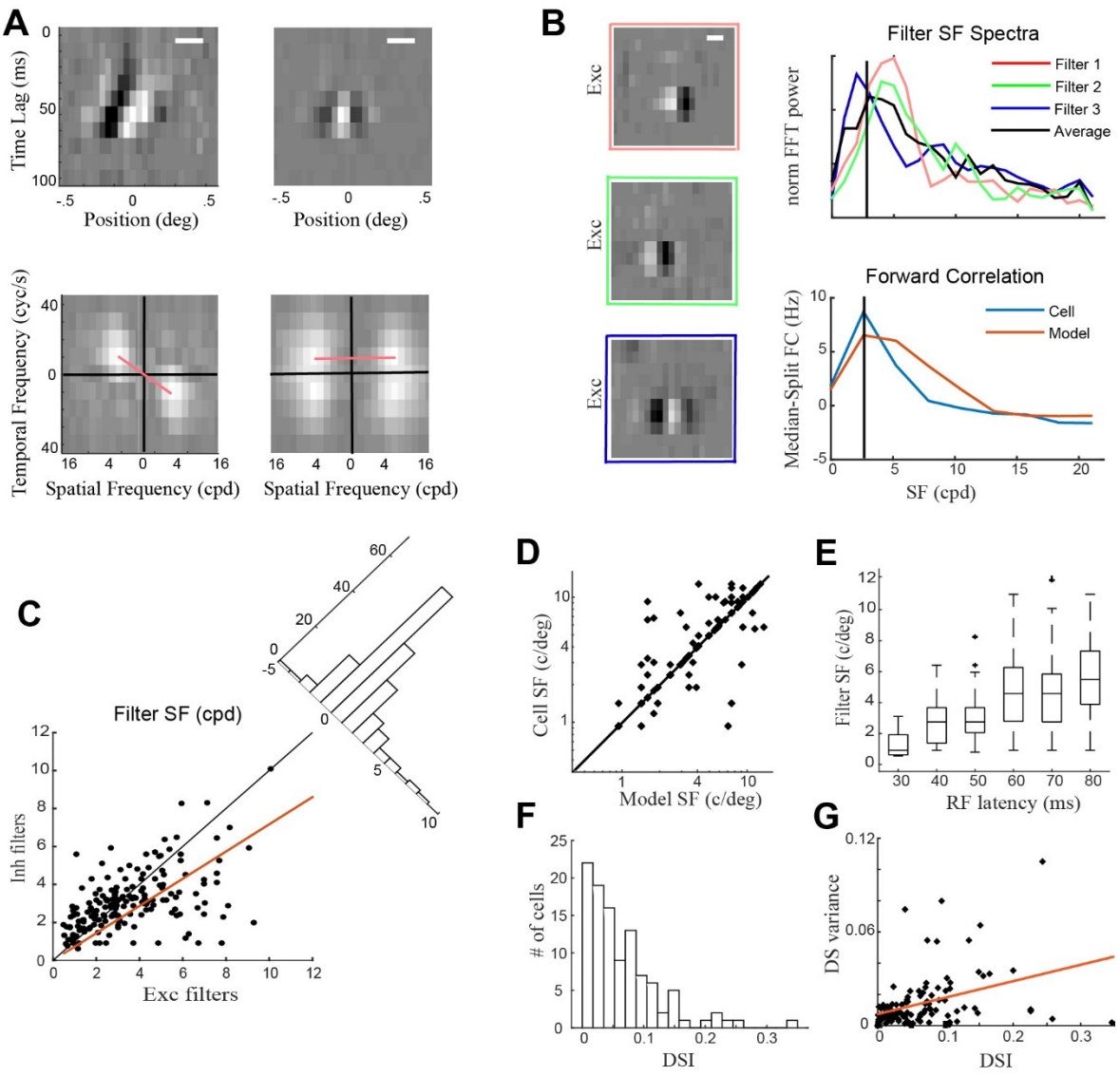

Figure 2.3: Extracting spatial, temporal, and spatiotemporal feature selectivity from models.

A*, Top*: Example filters demonstrating direction selectivity (left) and a non-direction selective filter (right). Scale bars indicate .2 degrees of visual angle. *Bottom*: 2DFFT spectra of the above filters. Peaks in the spatial and temporal dimensions are extracted as a measure of frequency

tuning preference of the filter. A slope can be fit to the peak locations of the 2DFFT spectrum (red line); this slope is perpendicular to the direction a filter is selective to and can thus be used to estimate direction and velocity tuning, while a 0 slope indicates a filter tuned to static feature (right). Additionally, the difference in total power in the top left and top right quadrants of the 2DFFT spectrum can be used to quantify the DSI: this measure yields both the direction the cell is selective to, and the strength of its direction selectivity. ***B****, Left Column*: Example filters. *Right Column, Top*: SF spectra extracted from the 2DFFT of each filter shown to the left. Each filter tuning curve corresponds to the filter in the left column outlined with the same color. The weighted mean across all filters (weighted by the filter contribution to the model output) is shown in black. *Right Column, Bottom*: Frequency spectrum of the same cell estimated using forward correlation. Black vertical lines indicate the peak frequency. ***C****,* Differences in spatial frequency tuning of excitatory and inhibitory filters from each model across our population. Excitatory filters are on average tuned to higher spatial frequencies. ***D****,* Preferred spatial frequency as estimated by the model-free forward correlation using observed firing rate and the model prediction. The two measures highly agree, further validating our model approach. ***E****,* Filter spatial frequency preference binned by the filter's onset latency. There is a positive correlation between filter onset delay and spatial frequency tuning. ***F***, histogram of DSI values across our sample. ***G***, variability in DSIs of each filter of a given model VS the model's overall DSI.

2.3.3 Model-based characterization of SF tuning

Having established model-based measures of spatiotemporal selectivity, we can now compare to more classical ways of estimating SF tuning using drifting grating stimuli (Victor et al., 1994; Ninomiya et al., 2012) to learn how selectivity to SF might be generated. One advantage of model-based approaches is that the models are "image-computable": meaning that we can simulate the responses of the model to arbitrary stimuli. While our dataset did not include recordings of responses to drifting gratings, we used the models of our sample cells to simulate their firing rates in response to static and drifting gratings and measured the F0 (mean) and F1 (modulation depth) components of the response to each grating frequency (Fig. 2.4B, solid lines show the responses of the model displayed in Fig. 2.4A, left) and compared those to our purely model-based estimates of SF preference. This led to a surprise: for some models (27/122), the predicted preferred SF tuning in the context of gratings of the full simulated model was a much lower frequency than that predicted by the model-based measures that were derived from the filters themselves, as described above. Closer inspection of these cells' SF tuning curves revealed two peaks both in their F0 and F1 spectra: with the low-frequency peak in the F1 spectrum indicating the highest response. Normally, the low-SF F1 peak being the largest response would be interpreted as indicating a simple cell tuned to a low SF. However, the F1 outputs of individual filters showed that all were preferentially tuned to higher frequencies (Fig. 2.4C), indicative of a more complex-like cell that could detect high-frequency inputs. Plotting the outputs in response to different gratings over time revealed that for high-frequency gratings: individual filters were being driven most strongly, but that the output of each subunit did not coincide across time due to their

different spatial locations. In contrast, to low-frequency gratings, the relatively small outputs of each subunit did overlap in time and summed (Fig. 2.4D). Therefore, the apparent selectivity to a low-SF grating was in fact generated as a result of a low-SF "envelope" driving multiple high-SF features suboptimally, but simultaneously. Thus, this type of cell appears to be simple-like for low SFs, but complex for high SFs.

Indeed, this apparent discrepancy between subunit tuning and the tuning to gratings can be explained by the previously observed spatial scatter between filters. To demonstrate this, we simulated two models for each neuron: the original model of the data, along with a second model constructed with identical subunits, except that the filters of each subunit were spatially aligned to have the same position, eliminating their spatial scatter (Fig. 2.4A, *right*). As expected, these peak-aligned models display greatly enhanced tuning to the carrier frequency (Fig. 2.4B, dotted lines) relative to the original models. Across our sample, peak-alignment resulted in the model completely eliminating the striking discrepancy between model-based and simulated-grating-based estimates of SF tuning (Fig. 2.4F, *paired-t-test,* p=0.0086), resulting in the two measures closely in register with one another (Fig. 2.4H, *paired-t-test,* p=0.4831). In the context of classical F0/F1 measures, these results indicate that some cells may show large modulation in their F1 component at a lower frequency, while still having significant tuning to a higher SF that is obscured by the choice of gratings used, similar to some previously observed properties of complex cell responses (Movshon et al., 1978b). Thus, we interpret tuning at the envelope frequency as not being caused by the presence of an explicit envelope filter, but rather as arising from the interactions between individual, rectified, high-resolution filters.

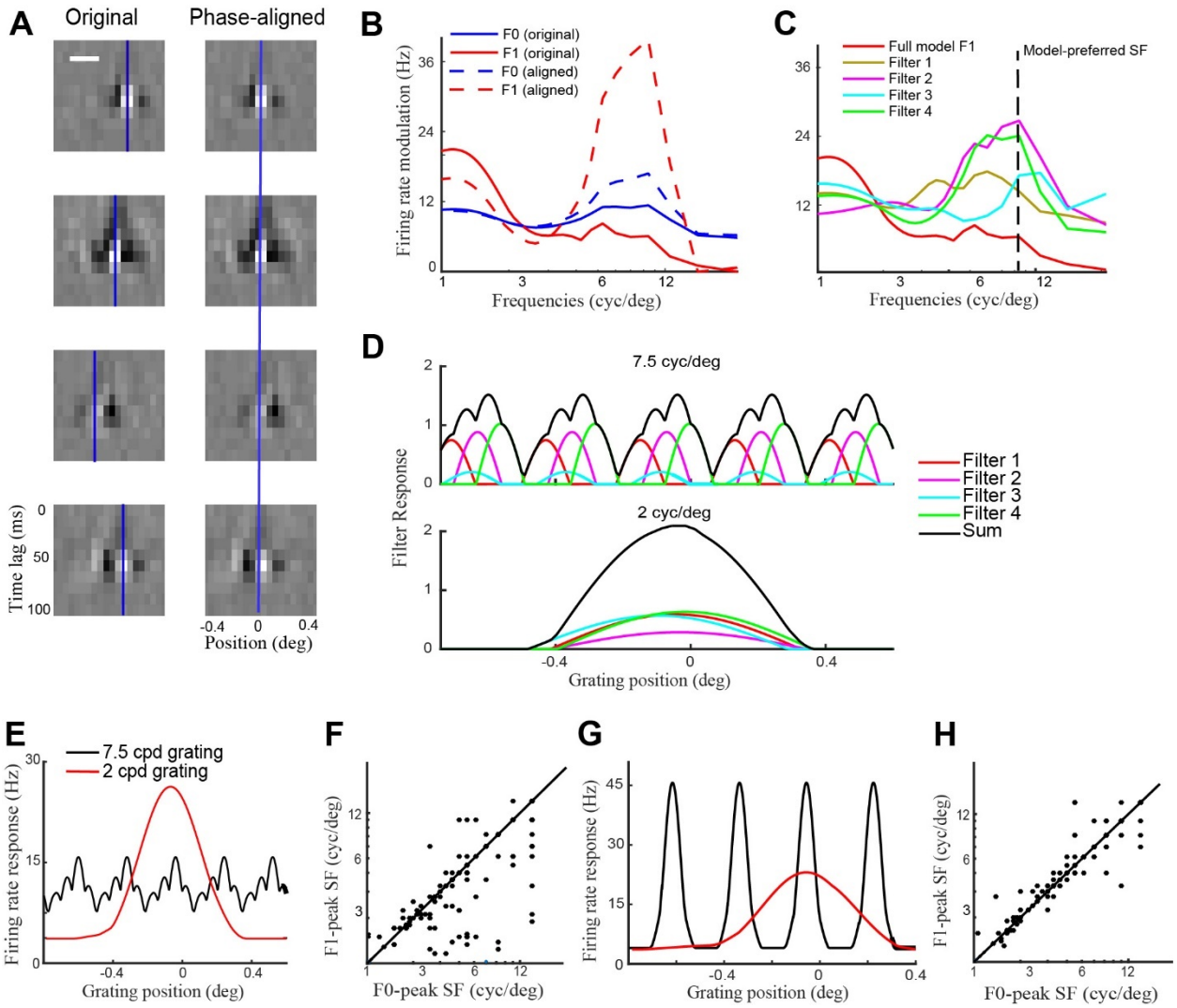

Figure 2.4: Responses to gratings by individual filters and the whole model.

A, Example model of a cell displaying envelope tuning. The 4 spatiotemporal filters with the highest model contribution are displayed on the left, while the same filters following phase alignment are displayed on the right. Blue vertical lines indicate the positive phase peak of the Gabor function used to fit the filter. B, F0 (blue) and F1 (red) responses to simulated gratings for

the original and aligned model's simulated responses to grating stimuli. ***C,*** F1-based SF tuning curves of individual filters and the whole cell estimated using simulated responses of the model to standard grating stimuli. The vertical line indicates the model-estimated SF tuning. ***D,*** Responses over time to a high-frequency (HF) 6 cpd grating (top) and a low-frequency (LF) 1.5 cpd grating (bottom) by individual filters (colored lines) and their sum (black line). The HF grating drives individual filters well, but the cell's overall phase invariance means that the individual responses do not sum, whereas the LF grating at the envelope frequency drives all filters simultaneously. ***E,*** Original model output in response to LF and HF gratings over time. The sum of weaker modulation by the LF grating generates a larger response than the individual responses of each filter to the HF grating. ***F,*** SF measurements based on F1 and F0. Measuring the grating responses using F0 and F1 reveals that F0 predicts tuning to higher frequencies, with 27 units being more than 2 SF steps lower in their F1 tuning. ***G,*** Original model output in response to LF and HF gratings over time for the phase-aligned model. HF responses sum temporally, the HF response is now again much larger than the LF envelope response. ***H,*** SF measurements based on F1 and F0 after phase-alignment. Across the population, this also aligns the preferred SF as estimated using F0 and F1 measures.

2.3.4 Quantifying the degree of nonlinearity across cortical layer

Model-based measures can also be useful to identify the degree that a given V1 neuron is nonlinear in its spatiotemporal processing. Traditionally, nonlinearity in V1 processing is referenced to the distinction between simple (relatively linear) and complex cells (nonlinear)

(Hubel and Wiesel, 1962; Skottun et al., 1991; Mechler and Ringach, 2002), usually estimated from the F1/F0 ratios described above. In such characterizations, however, the degree of nonlinearity does not necessarily generalize across stimulus contexts (Ringach and Shapley, 2004; Yeh et al., 2009; Butts, 2019; Almasi et al., 2020), and thus we propose an alternative model-based measures that can be easily adjusted for use with arbitrary spatiotemporally-varying stimuli.

We consider two time-varying spatiotemporal stimuli $\mathbf{s}(t)$ and $\tilde{\mathbf{s}}_x(t)$, which are different only because the stimulus at position x is inverted (i.e., black instead of white and vice versa). By comparing the response of the neuron to both stimuli, we can gauge whether it has a linear (inverted) or nonlinear (same sign) response, analogously to how the neuron responds to opposite phases of a drifting grating. This motivates our definition of the spatial phase-reversal modulation *sPRM*(*x*):

$$sPRM(x) = \frac{1}{T}\sum_t \frac{|r[\mathbf{s}(t)] - r[\tilde{\mathbf{s}}_x(t)]|}{\bar{r}} \tag{2.4}$$

where $\bar{r}$ is the firing response rate, and *T* is the total experiment duration. Notably, this is analogous to our previously proposed PRM measure (McFarland et al., 2014), but provides spatial resolution across the receptive field, allowing for direct model-free validation using forward correlation.

Specifically, to validate the nonlinearity measurements provided by the *sPRM*(*x*), we measure the average response of the neuron across all instances in the experiment where a black bar was at position *x*, compared with when it was a white bar. The resulting phase-reversal forward correlation differences (PRFC, Fig. 2.5C) closely mirror the sPRM (Fig. 2.5B). In simple cells,

presentation of white and black bars at the same spatial location is associated with opposite-sign modulation of firing rate relative to the firing rates elicited by gray bars, yielding overall PRFC values close to 1; In complex cells, either change in luminance elicits the same response, causing the firing rates to sum and leaving the PRFC near 0 (Fig. 2.5C). Likewise, the distribution of PRFC values (Fig. 2.5D) shows a high prevalence of complex cells and reflects a bimodal distribution often seen in previous V1 cell classifications (Skottun et al., 1991; Mechler and Ringach, 2002; Priebe et al., 2004).

Like the PRM, the average PRFC (average weighted across the receptive field of the neuron) can also be used to summarize the overall degree of nonlinearity of the neuron. As expected, these two overall measures of nonlinearity also closely corresponded: we found significant agreement between the two measures (Fig. 2.5E), further validating the PRM and sPRM measures.

To demonstrate the utility of such model-based measures of nonlinearity, we computed the degree of nonlinearity across cortical layers, expecting increasing degrees of nonlinearity outside the input layers of V1 (layer IV). The PRFC measure of phase-invariance displays the expected distribution across cortical layers, with cells in layer IV having large values on this indicator of phase-invariance, indicating more simple-like cells, while extragranular layers show significantly lower values more associated with complex-like cells (Fig. 2.6A) (Yeh et al., 2009; Sharpee et al., 2011; Atencio et al., 2012). Notably, both simple and complex cells can have a large number of subunits, indicating that simple cells still typically have nonlinear processing (Rust et al., 2005; Fournier et al., 2014). Finally, we also compared model features across layer. Notably, our models

also featured an additional aspect of infra-granular cells: their temporal kernels display longer onset latencies on average compared to Layer IV cells (Fig 2.6B) as previously described (Nowak et al., 1995; McFarland et al., 2015).

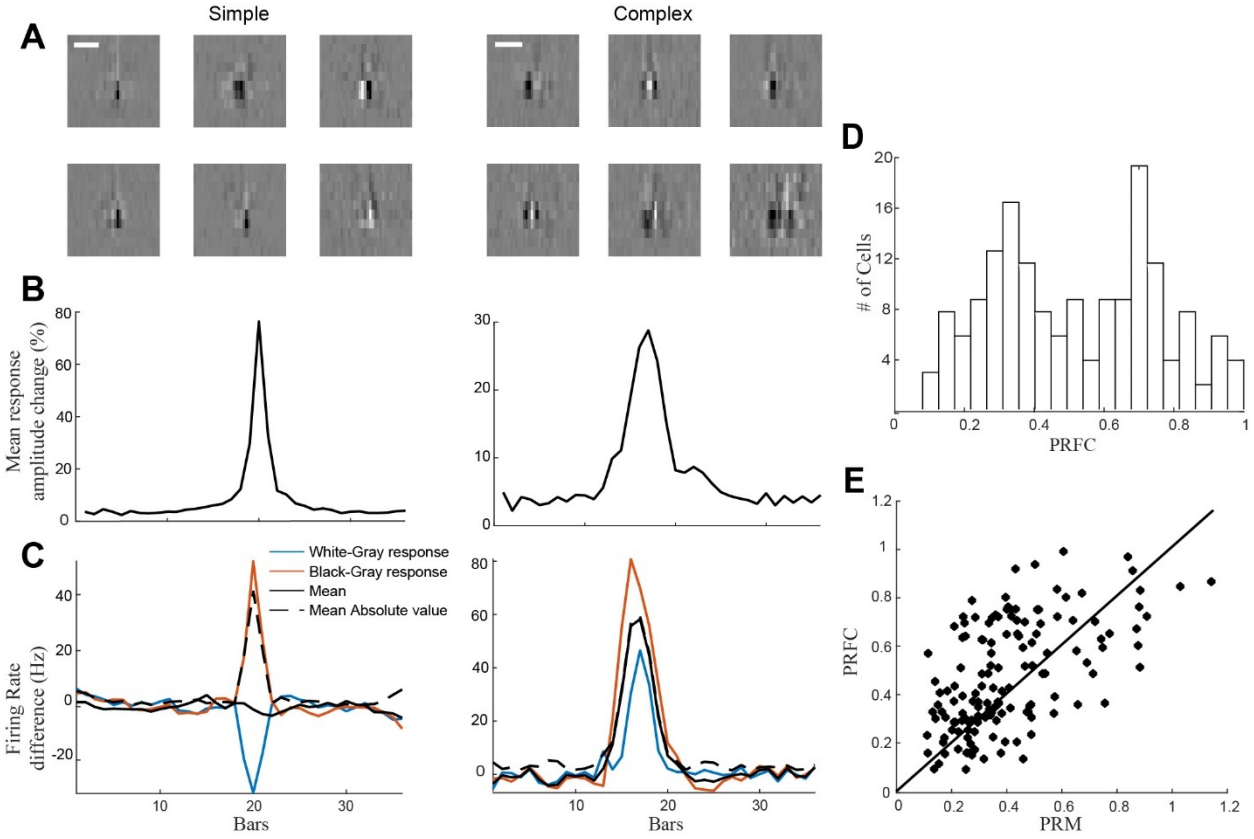

Figure 2.5: Measures of cell complexity and nonlinearity.

A, Example filters of a simple cell (left) and a complex cell (right). ***B,*** Spatial PRM profiles for the above cells. These spatial profiles closely match those established by the model filters in A. ***C,*** The spatial profiles of the forward-correlation based measure of phase insensitivity (PRFC) corresponding to the above example cells. Triggering on each instance of a bar at one spatial position being non-gray yields a measure of firing rate modulation across space. For a simple cell

(left), responses to bars of the opposite sign are also opposite, leading them to average out relative to the average of their absolute values. In a complex cell (right), any non-gray bar increases the firing rate, leaving the average modulation identical to the average of its absolute values. ***D*****,** Histogram of PRFC values, showing the broadly bimodal distribution normally associated with the simple/complex spectrum. ***E,*** Agreement between model-based and model-free measures of cell complexity.

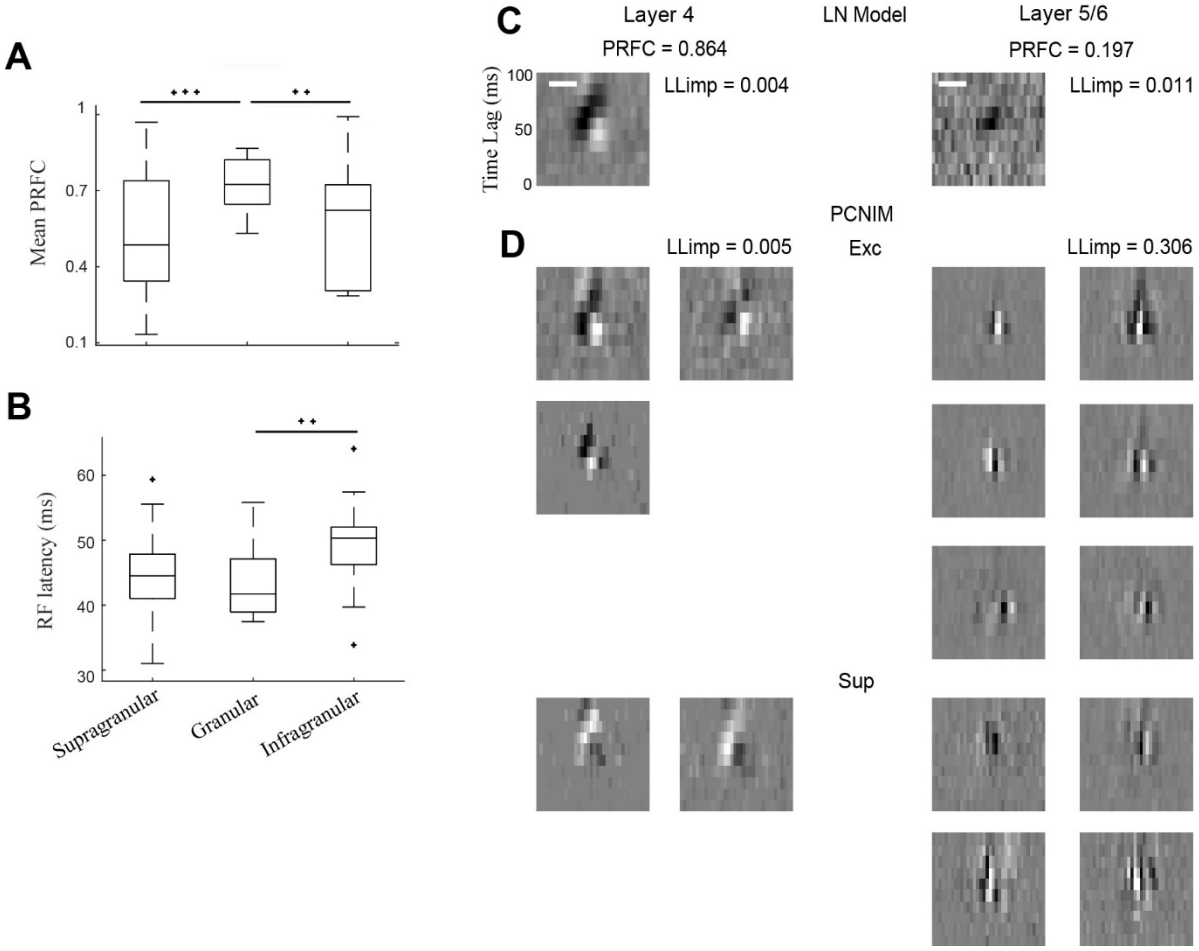

Figure 2.6: Degree of nonlinearity across cortical layer.

A, PRFC across the recorded population, revealing that neurons in Layer 4 are less nonlinear (have large PRFC values), while infra-and supra granular cells are more nonlinear (small PRFC values), indicating a much higher proportion of complex cells in those layers. ***B,*** Mean latency of excitatory filters for each model. Cells from infragranular layers have slower onsets, indicating that our models capture a delay in processing onset. ***C and D,*** Example LN (in **C**) and PCNIM (in **D**) fits for a layer 4 simple cell (left) and a layer 5/6 complex cell (right). The PCNIM provides only a modest benefit in log-likelihood improvement for the simple cell, but a massive improvement for the complex cell. Beyond having a greater number of filters and being able to fit the phase-invariant complex cell, the PCNIM reveals additional high-resolution RF structure in both cases.

2.4 Discussion

We have demonstrated how model-based characterizations of V1 neurons offer a more sophisticated description of their visual processing that is consistent with – but goes beyond – classical measures of V1 selectivity. Indeed, the underlying motivation is that nonlinear statistical models that are fit to more complex spatiotemporal stimuli capture aspects of V1 selectivity that is not evident in their responses to simpler stimuli, and thus measures based on such models will necessarily provide more information. Similarly, classical measures may not be possible where the stimulus cannot be precisely controlled, as is the case in an increasing number of experiments performed in awake animals in more natural contexts (Michaiel et al., 2020; Yates et al., 2021; Miller et al., 2022). Model-based characterization bridges the gap between recent advances in

statistical modeling of neuron responses to complex stimuli (Butts, 2019) and the descriptions of V1 tuning established by classical stimuli. Thus, it can fill an important gap in relating what can be relatively uninterpretable, complex models to the fundamental stimulus selectivity of the modeled neurons. As such, we propose this approach as a new benchmark for describing V1 neuron selectivity that is applicable to its processing of complex stimuli.

We have illustrated such measures, and their implications, using an established nonlinear modeling framework (McFarland et al., 2013). Much of our analysis relies on our model's ability to identify a given neuron's selectivity as a function of multiple subunits, as opposed to a single receptive field filter. Based on our analyses, most V1 neurons exhibit a diversity of subunit selectivity, in that individual cells are well described by many subunits that are often selective to multiple distinct spatiotemporal elements. For example, some V1 neurons might combine direction-selective subunits with those tuned to non-moving features (such as the example model in Fig. 2.1C, where only subunit 6 is direction-selective). Our results thus not only highlight the diversity of selectivity known to be exhibited by V1 neurons, but also emphasize the need for model-based characterization to understand this diversity.

Furthermore, the approaches presented here can be applied to any model of V1 whose first stage consists of multiple spatiotemporal processing elements and otherwise is image-computable (e.g., generates a predicted response to stimulus input), and thus in principle can be applied to the range of models from spike-triggered neural characterization (e.g., (Touryan et al., 2002, 2005; Rust et al., 2005)) to more recent models based on deep neural networks (e.g., (Kindel et al., 2017; Klindt et al., 2017; Moskovitz et al., 2018; Cadena et al., 2019; Ustyuzhaninov et al., 2020; Lurz

et al., 2021)), as the first stage of many such networks likely approximates a similar dimensionality reduction process as our filter subspace approach. However, we expect that such characterizations will be most accurate in models that accurately capture the neural computations (i.e., have high predictive power) and spatiotemporal processing attributable to individual V1 neurons, and thus use the NIM (McFarland et al., 2013; Almasi et al., 2020) in this paper for the basis of these characterizations.

A key finding of our analyses is that multiple subunits conferred tuning to a broader range of stimulus properties than would have been observable under classical contexts using traditional analyses. In the case of spatial frequency tuning, we find that some cells, while capable of resolving high-resolution inputs as indicted by the SF tuning of their subunits, are also tuned to broader stimulation by low-resolution inputs that spread across the cell's entire RF. This also potentially offers an explanation of how complex cells can still show some phase selective response patterns when presented with low-frequency gratings moving across the receptive field, while stationary gratings failed to elicit the same response, and explains the previously demonstrated dependence of the classical F1/F0 measure on spatial frequency (Movshon et al., 1978b). These findings are also consistent with findings of envelope tuning (Zavitz and Baker, 2014; Gharat and Baker, 2017), in which V1 processing for boundary segmentation is described as a theoretical LNLN cascade-style model with high-frequency filters in the first LN step and a coarse-scale boundary determined by the second LN step. They similarly observed responses to high-SF stimuli when presenting contrast-reversing luminance gratings (Gharat and Baker, 2017), implying that the ternary bar stimuli used in this study likely drives V1 activity in similar ways. However, unlike their model,

we did not observe a single filter outlining the envelope, but instead find that the response to the low-frequency envelope emerges out of the combined contribution of many more localized filters that are spatially scattered around the cell's RF.

Additionally, our model filters capture the high-resolution components of the cell RF as predicted by that study and thus provide a way of quantifying the differences between tuning to envelope and carrier frequencies of a visual signal. Because related theoretical work (Li et al., 2010) found that scattered high-frequency inputs may also be indicative of cells with broad orientation tuning, the interaction between subunit scatter and orientation tuning will need to be investigated in the future using stimuli with more dimensions than those used in our study, such as bar stimuli presented at multiple orientations or 2-dimensional noise stimuli. Such an approach might also yield clearer insight into other tuning properties where we observed significant variability between subunits, such as direction selectivity.

Our model-based approach also allows us to investigate the degree of nonlinearity needed to capture the responses of different V1 populations despite the well-established diversity of nonlinearity and tuning properties of primary visual cortex (Mechler and Ringach, 2002; Goris et al., 2015; Talebi and Baker, 2016) at much higher spatial resolution than was previous possible. Our model-based measures demonstrate a bimodal distribution of the degree of nonlinearity (Chen et al., 2009; Mechler & Ringach, 2002), and allows for additional insight into the construction of phase-invariant RFs by showing the spatial extent of individual subunits and their overlap. Because of this increased spatial resolution, the utility of our models as a measure of phase-invariance thus also goes beyond what is possible with classical analyses, such as using relative modulation

analysis after presenting grating stimuli (Kagan et al., 2002). A key point to emphasize is that it is only possible to study the effects of nonlinear interactions across space in the context of sufficiently complex stimuli. Our analyses of the resulting models could focus on interactions between filters at the same spatial position, which appear to be the real mark of whether the cell exhibits linear 'simple-cell-like' stimulus processing or spatial-phase invariant 'complex-cell-like' processing.

Likewise, simpler measures such as the number of subunits are not a reliable indicator of cell complexity (consistent with previous modeling studies (Rust et al., 2005; Park and Pillow, 2011; McFarland et al., 2013; Fournier et al., 2014; Cadena et al., 2019)). Indeed, we see that the smaller scale of spatial selectivity relative to the full RF of the neuron can require multiple filters regardless of whether neurons are phase-invariant. However, how this scale difference manifests depends on the assumptions of modeling: techniques such as spike-triggered covariance will often explain the "most variance" of the subunit ensemble with large quadrature pair filters (Rust et al., 2005; Touryan et al., 2005; Park et al., 2013) and leave finer elements of spatial selectivity to later filters – but such characterizations are consistent with small, localized filters (Lochmann et al., 2013). Indeed, statistical models that have flexibility in their nonlinear processing, which for example is afforded by rectified subunits (McFarland et al., 2013; Kriegeskorte, 2015; Butts, 2019) not only can better capture neural responses but offer a finer description of such nonlinear interactions.

We propose that our approach for model-based measures provides a new benchmark for quantitative estimation and comparison of tuning properties. The investigation of differences

between neurons across cortical lamina is just one application of our approach – future experiments may make comparisons between other neural populations, such as across eccentricity or between CO blobs, to probe the functional organization of the visual cortex by identifying which computational subunits change in different contexts. Furthermore, while we utilized relatively low-dimensional bar stimuli, the approach is applicable to responses to any complex stimuli, including more natural stimuli that vary in additional dimensions such as color and stimulus contrast. Indeed, while probing more stimulus dimensions requires more data to support the resulting models (Butts, 2019), models fit using those stimuli could reveal how such properties integrate with the selectivity measures described in this paper. Likewise, model-based characterizations might similarly be applied to "deeper" models with additional stages of nonlinear processing than the two LN stages we used here (Kriegeskorte, 2015; Moskovitz et al., 2018; Butts, 2019; Cadena et al., 2019; Richards et al., 2019; Lurz et al., 2021), either to describe later regions in the visual system or to further explore the additional nonlinearities present in V1. Such deeper networks would also be a more powerful tool to accurately identify shared computation in the context of more complex, and thus more high-dimensional, stimuli. Finally, due to the structural similarities among sensory cortical areas, model-based characterization might be further adapted to auditory or somatosensory modalities, potentially providing insight into general processing patterns in cortex by allowing for more direct comparison of shared or distinct computational subunits.

2.5 *Supplement*

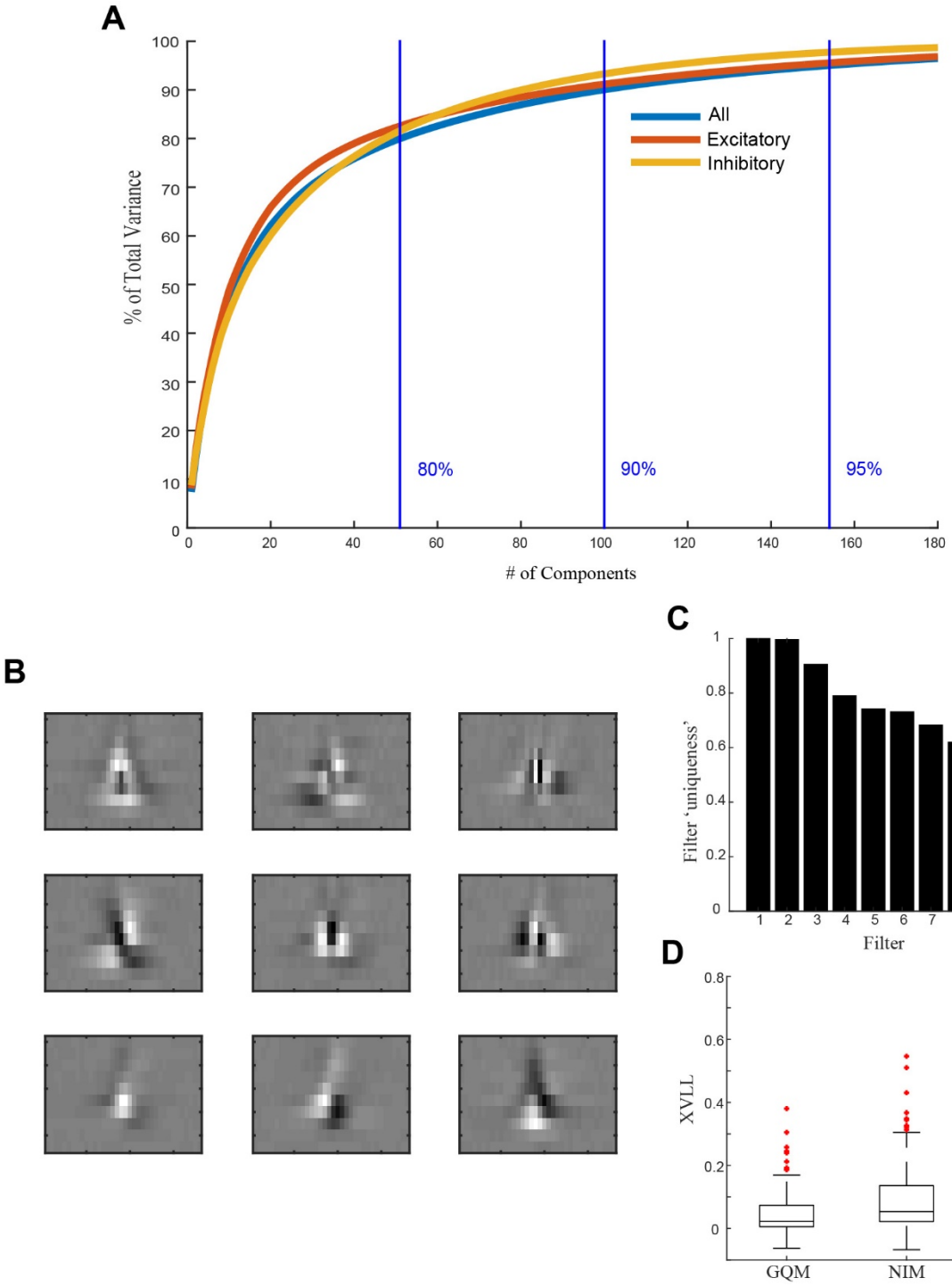

Figure 2.7 A: The Principal Component Nonlinear Input Model (PCNIM)

A, Variance explained by each Principal Component of a filter bank consisting of 1528 excitatory and 487 inhibitory filters. 80% of the total variance is explained by 51 PCs, 90% is explained by 100 PCs, and 95% is explained by 154 PCs. ***B,*** First 9 Principal Components used for dimensionality reduction during modeling. ***C,*** 'Uniqueness" of each filter of the example model shown in Figure 2.1C, quantified by the filter's vector magnitude after orthogonalizing it from the previous filters with the Gram-Schmidt process. Most filters make significant contributions to outlining a different area of stimulus space the cell responds to. ***D,*** Log-Likelihood improvement of the models using the PC-space for dimensionality reduction relative to full-space models. The PCNIM clearly outperforms other models in describing the observed data.

3 Amplification of disparity selectivity by spatial convolutions in primary visual cortex

This chapter is a manuscript currently being prepared for submission, and was co-written by Dan Butts.

3.1 Introduction

The functional role of primary visual cortex (V1) has been characterized as processing its inputs to generate outputs that have selectivity to specific, more complex features (Olshausen and Field, 1997, 2004). However, as modern visual neurophysiology is increasingly studying processing in more natural visual environments, the increased complexity in stimuli is likewise revealing increasingly complex processing (Olshausen and Field, 1996). It is clear that in more complex environments, canonical models of V1 neurons can only explain a fraction of V1 responses (David and Gallant, 2005), raising the question of whether the canonical picture of cortical processing is not relevant to this context, or if existing models are a valid basis of our understanding that simply need further adjustment. Identifying such adjustments to existing theoretical models to account for responses to more complex stimuli is a challenging process since V1 neurons generally receive a large number of functionally diverse inputs (Basole et al., 2003), and many "natural" visual stimuli encompass a wide range of possible visual features that may or may not be relevant to V1 computation (Olshausen and Field, 1996; David et al., 2004; David and Gallant, 2005); thus, failures of V1 models to reproduce them are difficult to attribute to any particular deficiencies of our understanding of cortical processing.

A demonstration of this challenge can be found in the processing of binocular disparity in V1. Binocular disparity is the difference in the retinal position of an image between the left and right eyes, and generated by objects in depth, whose disparity will depend on their depth relative to the vergence depth (Fig. 3.1A). Classical explanations of V1 disparity selectivity can explain their response to drifting gratings for both simple and complex cells (Ohzawa et al., 1997; Anzai et al., 1999). However, the model of disparity-selective complex cells, the Binocular Energy Model (BEM), has been shown to have dramatic failures when considered in response to more complex visual stimuli (Read and Cumming, 2003a). A number of theoretical expansions upon the BEM have been described that could potentially account for these failures by building upon the existing model (Read et al., 2002; Tanabe and Cumming, 2008; Doi et al., 2013; Henriksen et al., 2016; Goncalves and Welchman, 2017). However, these proposed additions to the BEM became increasingly complex and biologically implausible in their attempts to explain complex disparity selectivity, and none of these studies demonstrated that V1 neurons use their proposed mechanisms.

Therefore, identifying a more parsimonious solution to describing V1 disparity tuning would be a clear demonstration of a better approach to updating canonical models of V1 for more complex, natural stimuli.

Here, we propose a novel model for V1 disparity processing that is based on canonical principles of cortical processing but includes an additional mechanism to incorporate shared computations inherent to cortical processing. The critical component of this model is based on convolutional processing: the original principle suggested by Hubel and Wiesel to explain phase-

invariance of complex cells (Hubel and Wiesel, 1962). To demonstrate this, we present new methods of statistical modeling, combined with a novel binocular stimulation paradigm that allows us to specifically study the ability of our models to reproduce disparity processing. Our approach reveals a more general role of convolutional processing in amplifying V1 feature selectivity, and thus generalized this understanding in a way that can be applied to both modern neuroscience, and machine learning, which is based on similar processing.

3.2 Results

The premise underlying our stimulus design is that any explanation of disparity selectivity in general visual contexts will necessarily require a model with many more parameters than can be constrained by simple stimuli such as binocular drifting gratings. To characterize disparity selectivity in V1 neurons, we therefore used a spatiotemporal noise stimulus based on "ternary bar" patterns. This ternary bar stimulus is composed of randomly placed black or white bars on a gray background, oriented to match the recorded neurons' preferred orientation, and updated at either 30 or 100 Hz. The 30ms duration avoids other confounds of long-duration stimuli while allowing for enough integration time in V1 for accurate disparity processing to occur (Tanabe and Cumming, 2014). Each bar pattern was presented to both eyes using a haploscope (Fig. 3.1A) and the binocular disparity of each bar pattern was randomly chosen. Additionally, on each frame, there was a 10% chance that the stimuli presented to each eye were uncorrelated, and a 45% chance that the right-eye stimulus was inverted (changing black to white and white to black) to produce

an "anti-correlated" disparity stimulus (Cumming and Parker, 1997) that offers a further test of models of V1 disparity processing (Cumming et al., 1998; Read et al., 2002).

The stimulus design also has two built-in advantages for characterizing disparity selectivity. First, each neuron's tuning to disparity can be measured by its average firing rate in response to a given disparity, averaged across all patterns. Second, because the disparity of the stimulus and its pattern are uncorrelated on each frame, the fraction of the neuron's response driven by disparity can be mathematically separated from that driven by bar pattern via the Law of Total Variance (Fig. 3.1C and see Methods). We will use the fraction of each neuron's response that is driven by disparity, its *disparity variance fraction* or DVF, as a measure for how strongly disparity-tuned it is (Fig. 3.1D). We will demonstrate below that by understanding the factors that influence the DVF provide a key to understanding the processing of disparity by V1 neurons. Namely, we will show that large DVFs result from amplification via cooperative interactions in disparity processing of individual processing units, combined with canceling interactions in the processing of matters.

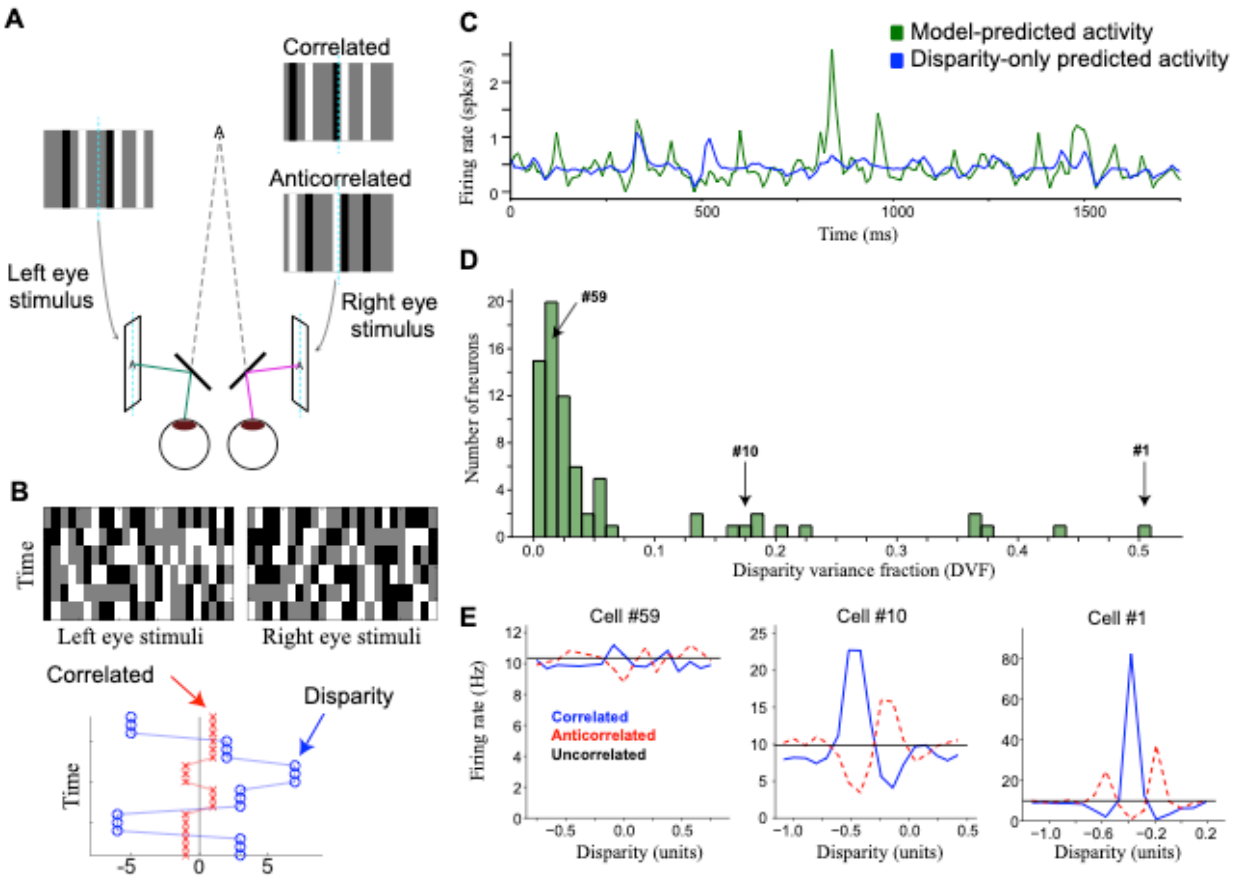

Figure 3.1: Exploring V1 disparity selectivity.

A: Stimuli were presented via a haploscope: Two screens were placed next to the animal, while mirrors reflected the screens to create an emulation of binocular inputs. Stimuli consisted of bars oriented at the cell's preferred orientation, which were either binocularly correlated (and thus eliciting a perception of depth) or anticorrelated (not associated with depth and instead driving false matches). **B**: Schematic of stimuli over time. Stimuli at a given frame were randomly shifted by some number of pixels (disparity) and randomly correlated or anticorrelated. Additionally, stimuli changed either every frame of the 100Hz monitor, or only every third frame – this was

done to ensure that slower binocular integration processes could still be observed. The frame rate of stimuli was held constant within each stimulus block. **C**: Demonstration of firing rate variance over time. At each time point, we can make a prediction of the expected firing rate using both a full model considering the whole stimulus, and a reduced model only accessing the metadata of each frame's disparity and correlation. The variance in firing rate explained by the full model is the total variance, while the fraction of the total variance explained by just the disparity information is the Disparity Variance Fraction (DVF). **D**: Distribution of DVFs across our recorded population. Many cells have very little disparity tuning, while a small subset have strong selectivity to disparity. **E**: Examples of disparity tuning curves in our sample, including an untuned cell (left), a "odd-tuned" cell (middle), and an "even-tuned" cell (right).

3.2.1 Failure of the data-driven Binocular Energy Model in explaining V1 disparity processing

To begin, we consider the canonical model of disparity processing in V1, the Binocular Energy Model (BEM) (Ohzawa and Freeman, 1986; Ohzawa et al., 1990). The classical BEM is based on two binocular spatiotemporal filters (Fig. 3.2A), whose output is a sum of linear spatiotemporal processing performed by filters processing the stimulus in each eye, with the filters identical but with one shifted relative to each other by the neuron's preferred disparity. The two binocular filters of the BEM are typically Gabors that comprise a "quadrature" pair, such that one is shifted in phase by 90 degrees relative to the other. Like the standard Energy Model of V1 cells (Adelson and Bergen, 1985), the linear outputs of each filter are squared and then summed to predict the neurons response. Thus, like the standard Energy Model, the BEM's binocular filters

confer selectivity to the orientation and spatial frequency of the individual filters, as well as phase invariance of the output (Lochmann et al., 2013; Bartsch et al., 2022). Furthermore, because the filters of the BEM are binocular and shifted by the same preferred disparity, the resulting model is additionally disparity-selective (Ohzawa et al., 1997).

Here, we cast the BEM in a data-driven framework using an LNLN cascade model (Park and Pillow, 2011; McFarland et al., 2013; Vintch et al., 2015; Henriksen et al., 2016), adding a linear filter to the model to complement the two binocular filters to fit phase-selective responses (for example produced by simple cells), and a spiking nonlinearity that flexibly scales the response but otherwise does not affect stimulus processing. We use *maximum a posteriori* (MAP) estimation to fit this statistical model to each recorded neuron, and thus can empirically measure the binocular filters that best reproduce V1 neuron responses independent on any assumptions other than the mathematical structure of the model itself (Butts, 2019). For disparity-selective V1 neurons, the resulting models demonstrate the features of in the binocular filters that were described above (Fig. 2.2A), recapitulating earlier work using spike-triggered covariance on a similar dataset (Tanabe and Cumming, 2014).

We can go further with this modeling framework because it produces models that are "image-computable", allowing for the direct comparison of the BEM's predicted disparity tuning curves (Fig. 3.2B) — given by the average firing rate response to a given disparity averaged across all patterns — with that measured from neural responses directly (Fig. 3.1E). This demonstrates, unfortunately, that data-driven BEMs are unable to reproduce each neuron's disparity tuning (Fig. 3.2B): while they can reproduce the core features of the observed disparity tuning curve, they are

neither able to match the magnitude of disparity tuning (Henriksen et al., 2018), nor the attenuation of response to anti-correlated stimuli (Cumming et al., 1998; Read et al., 2002).

This failure in predicting the disparity tuning curve was consistent across all strongly disparity tuned neurons in our dataset (Fig. 3.2C), although the BEM could perform better on neurons that are weakly disparity tuned and still predicted the overall responses to stimuli well (Fig. 3.2D). This has suggested to many that strongly disparity-tuned V1 neurons find some way to amplify their responses to certain disparities that goes beyond the mechanisms of the BEM (Read and Cumming, 2003b, 2007; Tanabe and Cumming, 2014), and many different ideas have been presented. However, here we suggest a new explanation that depends on the interaction between pattern and disparity selectivity, which — paradoxically — is already present in the two filters of the BEM, but implicitly limited by this small number of filters.

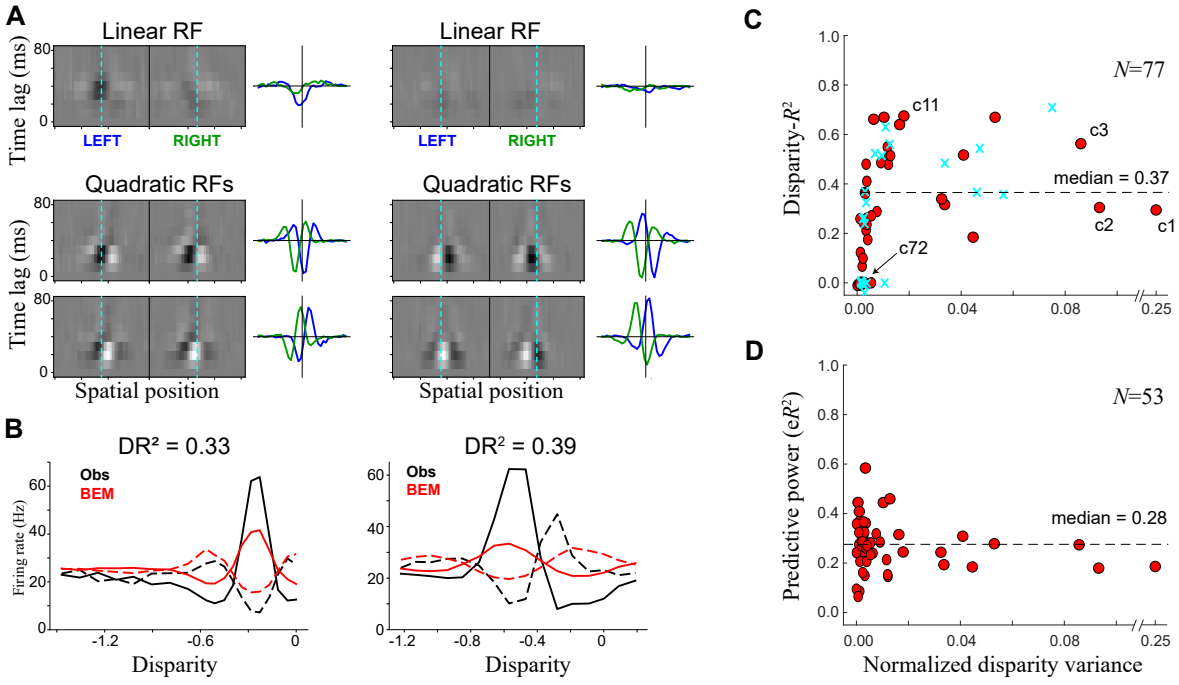

Figure 3.2: Directly fitting V1 responses using the BEM framework.

A: Examples of two BEM filter sets (one linear and two quadratic filters) for two strongly disparity-tuned cells. Left and right eye filters are selective to the same pattern, but at different spatial positions, corresponding to the preferred disparity. Panels on the right show a cross-section of the corresponding filters at the best lag. **B**: Disparity tuning curves for the corresponding cells, with the true disparity tuning in black, and BEM prediction in red. DR2, or disparity R-squared, quantifies that the BEM fails to adequately explain disparity tuning in these cells. **C**: Disparity R-squared values across the population show that BEMs consistently fail to fully explain disparity tuning. **D**: Likewise, predictive power for disparity-tuned cells shows that the BEM architecture does not fully account for V1 binocular integration processes.

3.2.2 Disparity amplification in the BEM

To understand disparity amplification, we first consider individual binocular filters. Each filter of the BEM implicitly links disparity and pattern processing: the sensitivity to disparity of a given binocular filter intimately depends how closely the stimulus matches their spatiotemporal filter. Thus, disparity selectivity of the filter requires the stimulus to match the spatiotemporal pattern of the filter, and the pattern selectivity will ultimately bound the degree of disparity selectivity — at least for a given filter.

We can take advantage of our experimental design to formally establish this bound, because disparity and pattern are uncorrelated in the stimulus. As a result, the Law of Total Variance allows us to partition the variance of a filter's output into a component driven by

spatiotemporal patterns, and a component driven by disparity directly (Fig. 3.1C). Because the response to disparity depends on how well the pattern matches the spatiotemporal aspect of the filter, the ratio of disparity variance to its total variance (i.e., its DVF, defined above) is bounded for a single filter (Fig. 3.3D, E), with the exact bound depending experimental details relating to the scale of the noise patterns (in space and time) relative to the filters themselves. This bound is illustrated by the DVF for each binocular filter across all the models in our study (Fig. 3.3A): both for filters found with the BEM and those of our novel BCS model described below. Comparing to the DVF measured across the V1 neurons in our study (Fig. 3.1C), this demonstrates that the DVF for the strongly disparity-tuned neurons is larger than any single filter can produce. Instead, any such model with linear filtering as its first stage necessarily requires nonlinear processing to amplify its DVF, regardless of binocular filter design.

In fact, the partitioning of each filter's output into that driven by disparity and that driven by pattern provides a means to understand how the BEM can have a larger DVF than the bound on each individual filter. This disparity amplification is possible because of the relationships between the outputs of each filter. Because the two filters comprising the BEM are quadrature (90 degrees out of phase), their responses to patterns will be uncorrelated, and their pattern variances simply add when summed to generate the model's output. However, because — by design — both filters have the same disparity tuning, the disparity signal in their outputs will be completely correlated, and thus combine superlinearly to result in a larger combined disparity variance than their independent sums. This will increase the fraction of disparity variance (i.e., DVF) in the

output of the BEM, demonstrating how disparity amplification can occur via the correlation of their disparity processing.

Based on this analysis, it follows that models can gain further amplification by adding further filters. However, there is no clear way to extend the BEM in its quadratic form, because 2 quadratic filters in quadrature already account for a significant fraction of the stimulus space (McFarland et al., 2013), and thus additional quadratic filters will interfere with existing filter projections, at best adding marginal explanatory power and at worst confounding the fitting procedure and making filters less interpretable. Therefore, different nonlinear approaches are needed to capitalize on disparity amplification through multiple filters.

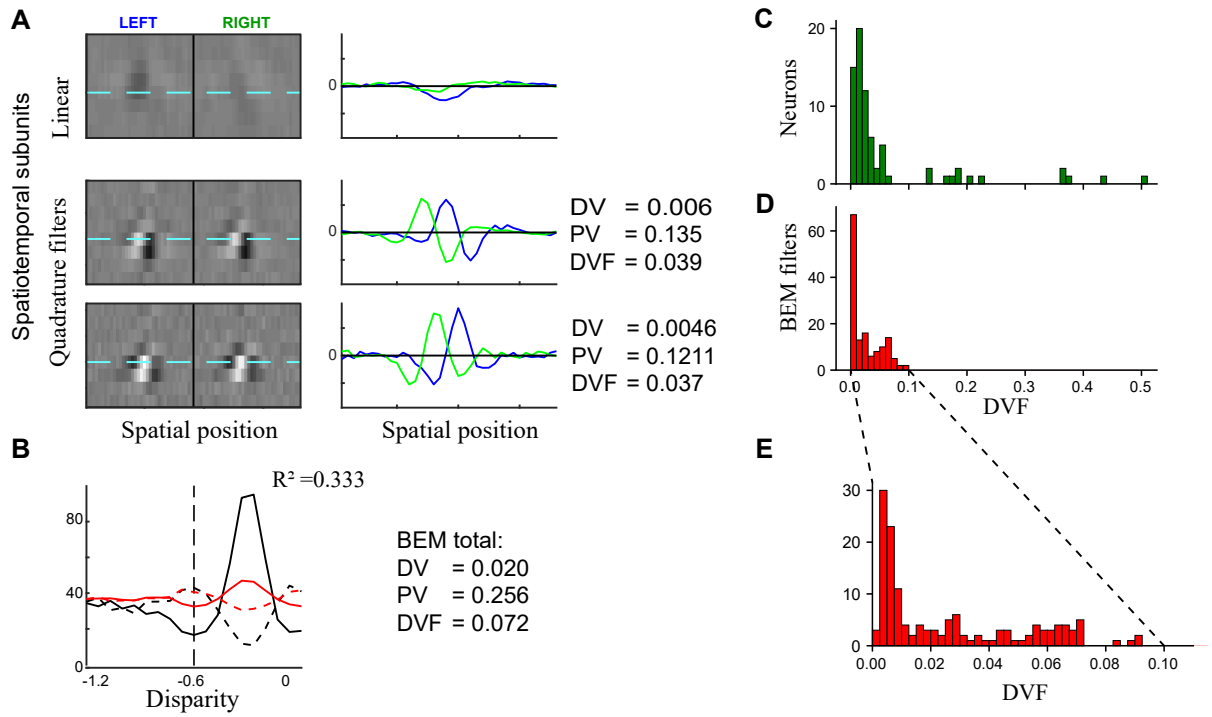

Figure 3.3: Disparity amplification in the BEM.

A: Example BEM from Figure 3.2, with Disparity variance (DV) and Pattern variance (PV) shown for each of the quadratic filters. Despite being clearly tuned to a particular disparity, most of the contribution of the two filters lies in their selectivity to stimulus pattern. **B**: DV and PV for the whole BEM output. Because the pattern selectivities of the filters are independent but disparity selectivity is shared between the filters, the DVF of the BEM is enhanced by a factor of 2. **C**: DVFs of the cells in our sample, reproduced from Fig. 3.1D. **D**: DVFs of individual BEM filters across our population. No filter achieves a DVF above 0.1, whereas DVFs observed in cells can be as great as 0.5. This demonstrates the ceiling in DVF inherent to individual BEM filters and proves that traditional BEMs cannot capture the full range of disparity selectivity in V1. **E**: same data as D, but at smaller bin size. Most BEM filters have low or nonexistent disparity selectivity.

3.2.3 Disparity amplification through spatial convolutions

Thus, we consider a more general, flexible, subunit model framework, and adapt the structure of the Nonlinear Input Model (McFarland et al., 2013, 2015, 2016; Vintch et al., 2015; Almasi et al., 2020) to extend to our binocular stimulation paradigm (Henriksen et al., 2018). This binocular convolutional (BC) model processes visual stimuli using a collection of nonlinear subunits (Fig. 3.4A): stimuli are first passed through a set of temporal basis filters which reduce dimensionality of temporal signals (Fig. 3.4B) and are then combined with spatial information to create a set of monocular filters (Fig. 3.4C). The next layer can then project weights to these monocular filters and combine any combination of them as left and right eye filters to create a set of binocular spatiotemporal filters (Fig. 3.4D). Then, these binocular filters are convolved with

stimuli at multiple spatial positions, thus generating an output at each such spatial position, which is then rectified and constrained to be either excitatory or inhibitory (Fig. 3.4E). In principle, this means that each binocular subunit can be at every spatial position. The underlying binocular filters thus more generally represent features that modulate the neuron's response, and the subunit output is summed and passed through a spiking nonlinearity to generate a predicted firing rate. Because this approach effectively just adds convolutional subunits to the NIM (Vintch et al., 2015), both the filters and their weights across spatial position can be simultaneously fit using the same procedures.

By applying these spatial convolutions to a small set of binocular filters, the disparity-driven output of each filter will be completely correlated because the same disparity is present across spatial positions, just like the two binocular filters of the BEM. Adding the outputs of such spatially shifted subunits will thus amplify the disparity variance. At the same time, depending on the details of the spatiotemporal filter and its spacing across space, the pattern-drive output can easily be uncorrelated — or even anti-correlated — resulting in a relative suppression of the pattern variance.

The resulting BC model was able to explain nearly twice as much of each V1 neuron's response across the population of recorded neurons on average (Fig. 3.5D, compare to Fig 3.2C), independent of whether the neurons were tuned for disparity or not. More strikingly, for the strongly disparity-tuned neurons (N=13), the model was able to reproduce the response to disparity with high fidelity: explaining a median above 90% of the disparity variance. The model also does an unprecedently good job predicting the overall response, with an average performance across

the entire dataset of 53% (Fig. 3.5E, compare to Fig. 3.2D). This means that it is also accurately predicting the response of non-disparity tuned cells.

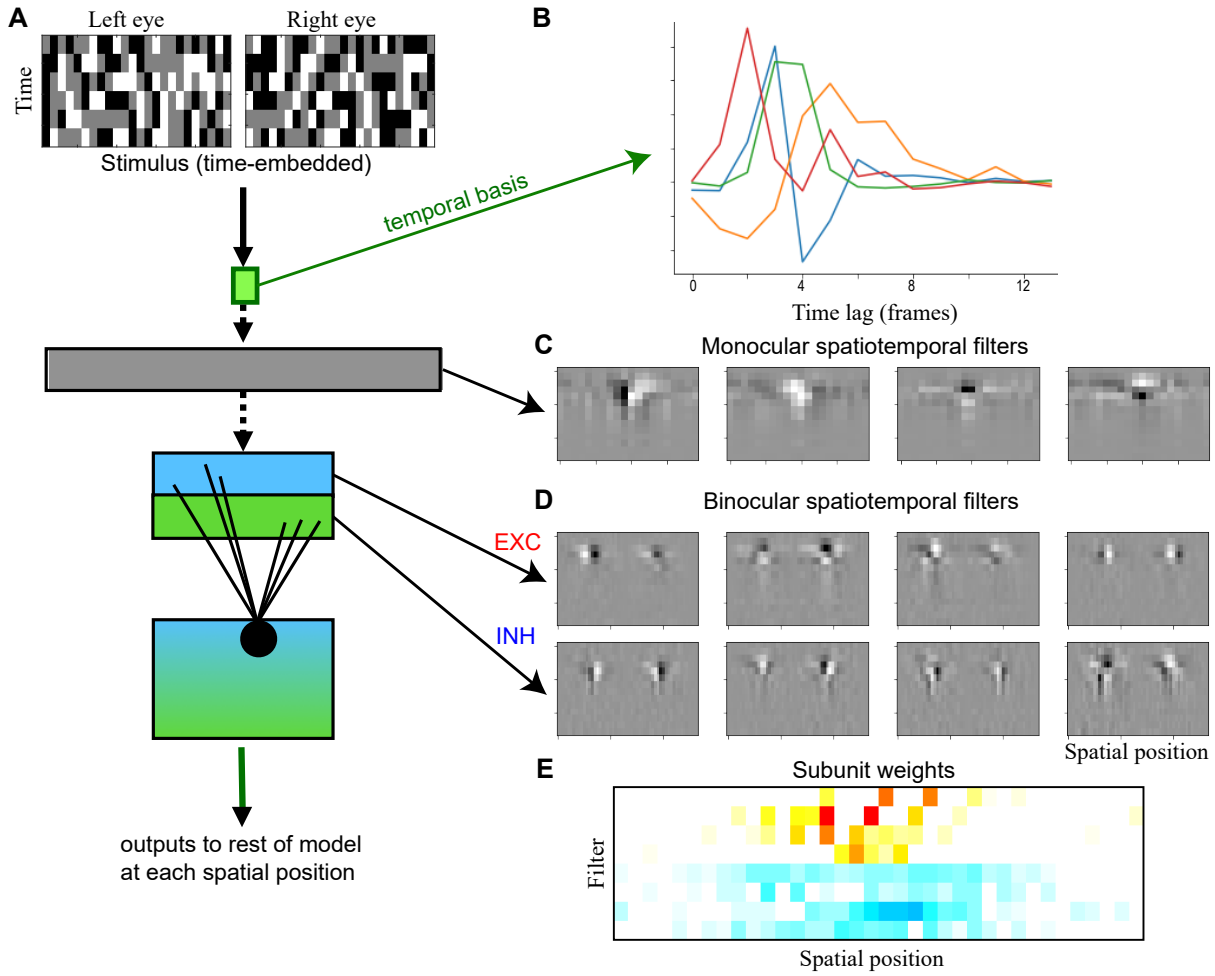

Figure 3.4: The Binocular Convolution Model.

A: Schematic of the BC model. Left and right eye stimuli are fed into the model, passed through multiple stages of processing, and finally used to generate a firing rate prediction. All processing stages can be fit to data using maximum likelihood approaches. **B**: Temporal basis. The first step of the model is to convolve the stimulus timecourses with a set of 4 temporal kernels,

which reduce the dimensionality of the temporal domain to 4 dimensions. **C**: Monocular spatiotemporal filters. The outputs of the temporal convolution are then processed across space to generate selectivity to spatiotemporal features. **D**: The monocular filters are used in the next layer to construct binocular spatiotemporal filters by selecting a combination of monocular filters and a single disparity between those filters. The total number of filters can be changed, but in this example consists of 4 excitatory (top) and 4 inhibitory (bottom) filters. **E**: The binocular filters from the previous layer are finally convolved across space and each spatial position is assigned a weight to determine the strength of its selectivity at that position.

3.2.4 Excitation and suppression in the BC model

Because of rectification of the subunits, the filters themselves can only either excite or inhibit the neuron's response. We enforce this identification of convolutional subunits as excitatory or inhibition explicitly by constraining the spatial weights to be only positive or only negative when fitting the BC model. Indeed, not doing so results in the optimization failing to find models with close to the same performance.

Analyzing the contributions of individual filters, we again observe only small DVF values among both excitatory and inhibitory filters, at the same scale as those observed in single BEM filters. However, we also find that spatial convolutions indeed amplify disparity tuning as expected: in example cells, we regularly observe amplification an order of magnitude above the contributions of individual filters, even if the filter was only represented across a few spatial positions (Fig. 3.5B). Looking at the contributions to DV across all excitatory filters, we further

observe that DV values increase superlinearly, exceeding the effects of disparity amplification we would see if DVs merely added together (Fig. 3.5C), thus demonstrating disparity amplification via spatial convolutions.

As a further step of amplification, we find that although excitatory and inhibitory subunits have similar tuning to patterns, they have opposite tuning to disparity. This means that inhibition causes pattern variances to cancel out, while still amplifying the disparity variance by suppressing the presence of the wrong disparities – in some cases, even amplifying disparity selectivity more than excitatory filters (Fig. 3.5B). Like the DV amplification in excitatory filters, we find that the contributions of inhibitory subunits have nonlinear effects on PVs, but in this case, act to generate sublinear effects (Fig. 3.5C), effectively suppressing selectivity to certain patterns as part of a mechanism to disregard false matches in the stimuli.

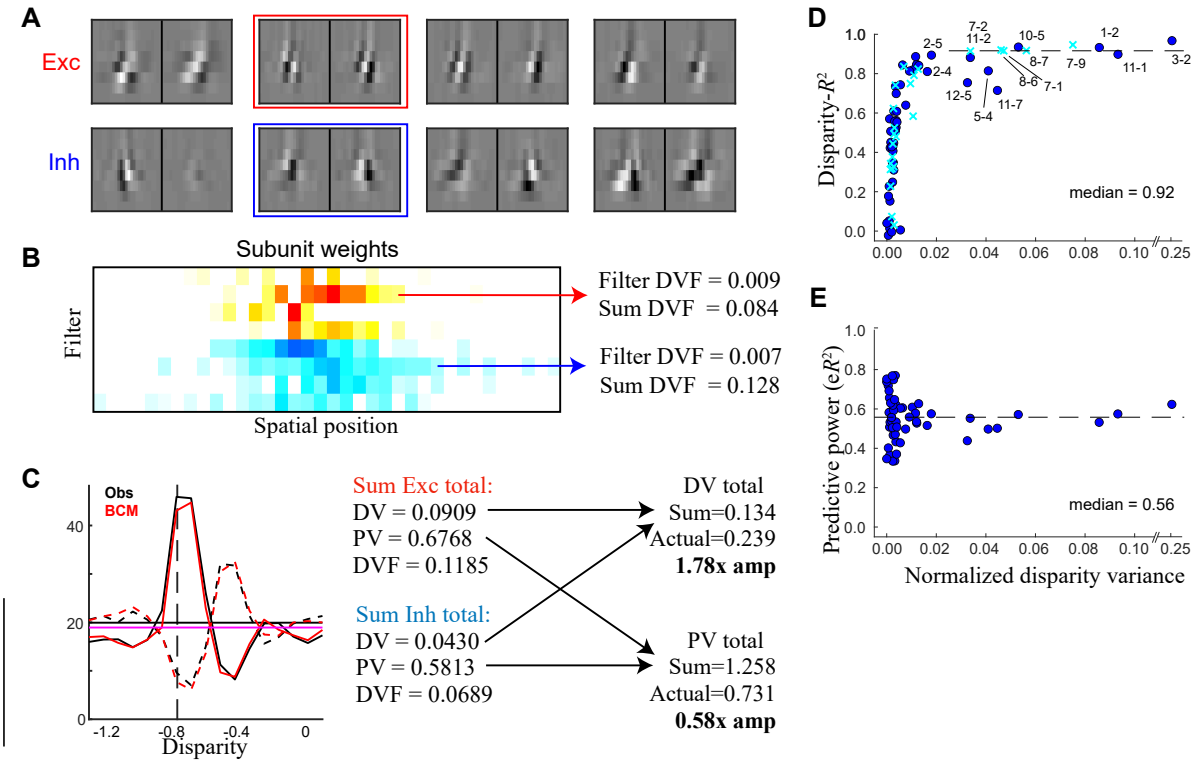

Figure 3.5: Disparity amplification in the BCM.

A: Example set of binocular filters for one cell's BCM. **B**: Subunits weights for the same model. Arrows indicate the DVF information for filters 2 and 6, which are marked by the red and blue boxes in A. Although the individual filters have low DVFs, measuring the DVF of the combined outputs of the filter across spatial positions shows a 9-fold enhancement in the excitatory filter and an 18-fold enhancement in the inhibitory filter. **C**: Disparity tuning curve and DVF measures for the whole model. The BCM fully captures this disparity tuning curve, including its attenuation, and shows how excitatory and inhibitory filters interact to amplify tuning to disparity. **D**: Disparity R-squared values across the population show that BCMs capture disparity tuning

well, explaining over 90% of disparity variance in disparity-tuned cells. **E**: Likewise, predictive power for disparity-tuned cells shows that the BCM significantly improves upon the BEM (compare Fig. 3.2D).

3.2.5 Attenuation in the BC model

A core requirement for any model of disparity detection is a mechanism to identify false positives (Cumming and Parker, 1997), a process which reveals itself via attenuated modulation by anti-correlated stimuli (Read and Cumming, 2007). Anti-correlated stimuli activate disparity detectors such as the BEM but contain no global solution of the correspondence problem (Henriksen et al., 2016), and elicit no sensation of depth (Cumming and Parker, 1997; Cumming et al., 1998). Such stimuli are thus a simple test to assess models of binocular integration. Several theoretical expansions to the BEM that would produce attenuated responses to anti-correlated stimuli have been proposed, but none have been validated on real data (Samonds et al., 2013; Goncalves and Welchman, 2017). Analyzing the BC model, we find that it perfectly captures attenuation by leveraging the same interplay of excitation and inhibition mechanisms used to enhance disparity selectivity in general.

These results provide experimental validation for the proposal that attenuation can thus be characterized not as a failure of suppressing the negative responses to anticorrelated stimuli, but rather a failure to accurately capture the relative amplification of the responses to matching stimuli (Henriksen et al., 2018) which is provided by our model.

3.3 Discussion

In conclusion, here we presented a new model for how disparity tuning is generated by V1 neurons. This model allowed us to cast the computations underlying the tuning of the BEM in a new light and put it in the context of the full convolutional model. We demonstrated our approach by using data-driven modeling in more complex stimuli, where it predicted the disparity tuning – and the response in general – with unprecedented accuracy. Furthermore, the methods of analyses we developed, disparity and pattern variance, both provided a means to see how the model components interact, as well as reveal the subset of neurons in V1 that are specialized for disparity.

It is important to note that previous work has shown that it is possible to hand-tune a BEM to produce a similarly accurate fit to the tuning curves of non-attenuated disparity-tuned cells by focusing on the disparity response (Read and Cumming, 2003a). However, while such a hand-tuned model might be able to capture the disparity tuning curve, it would do so at the cost of describing the responses to particular stimuli and thus do much worse than the fitted BEM in predicting the response of the cell to actual bar stimuli. The BCM, however, did not have to make a tradeoff between capturing disparity selectivity and pattern selectivity and can indeed capture the full range of monocular, binocular, and temporal dynamics of the cell's stimulus selectivity.

The fact that the BCM performed similarly well at describing the responses of non-disparity-tuned V1 cells, most of which were still binocular to some degree, is another testament to its versatility. Further analysis into the binocular computations revealed by the model may compare the particular selectivities of disparity-tuned cells with those that are not, and thus reveal more fundamental mechanisms that facilitate the specialization for disparity tuning.

Another benefit of the BC model is that it is notably simpler than other recent attempts to build upon the BEM, such as neural networks using a diverse set of overlapping stimulus filters to detect matches and identify false matches simultaneously, based on local image statistics related to depth in natural images (Goncalves and Welchman, 2017). While our model outlines an optimal coding strategy for binocular integration in natural scenes, it lacks constraints linking it to biologically plausible mechanisms. The computations implemented by the BC model, on the other hand, could be easily explained at a neurophysiological level through simple experience-driven wiring of V1 neurons to multiple afferent receptive fields with matching disparities – a hypothesis first put forward by Hubel & Wiesel (Hubel and Wiesel, 1962; Menz and Freeman, 2004).

Prior research has shown that much binocular processing occurs in higher-order visual areas; the first brain area in the visual hierarchy to solve the correspondence problem is likely V2 (Parker, 2007; Tanabe and Cumming, 2008; Henriksen and Read, 2016). However, studying these regions has proven difficult without knowing the true format of the information they receive from V1. The BCM thus provides a potential steppingstone for better studies of these brain regions: Because the BCM can account for the pre-processing of depth cues in V1, it presents a potential front-end for future models of binocular integration.

Finally, the absolute disparities computed in V1 have been implicated as important inputs into the dorsal stream regions planning vergence eye movements (Parker, 2007), which bring left and right eye images into registration. Understanding the nature of V1's disparity processing is therefore also important to understand the nature of eye movements. in higher-order visual areas, and binocular vision in general.

3.4 Methods

The methods used to collect the data used in this study have been published previously (Henriksen et al., 2018), but are restated here:

3.4.1 Electrophysiology:

A laminar probe was lowered through the dura mater on each recording session. Voltage signals were amplified and band-pass filtered (0.1 – 10 kHz) before being stored. Spike waveforms were detected online using Blackrock software and recorded at 32 kHz sampling rate. Following online isolation of a single unit, the unit's classical tuning components such as orientation, phase, direction selectivity, and spatial frequency as well as ocular dominance were characterized by presenting drifting sinusoidal gratings. Disparity tuning was assessed with a random-line stereogram (RLS) centered on the cell's minimum response field (Prince et al., 2002b, 2002a; Read and Cumming, 2003b). If the cell's firing rate was below 10Hz to the stereogram, or did not show significant disparity selectivity, it was discarded from the dataset.

3.4.2 Stimulus:

Stimuli were presented on a 100Hz CRT monitor. All stimuli were centered on the average MRF center of the recorded population. Recording sessions were divided into several blocks of 100 trials consisting of 300 frames each. The total number of blocks on each recording day varied from 3 to 16. The monkey's eye position was tracked throughout the experiment; if the monkey deviated from the fixation point during a trial, it was abandoned and restarted with a new set of stimuli. The same set of stimulus-generation seeds was used for each recording day, in randomized

order. Thus, each recording generated data using the same stimulus sequences within trials, allowing for easier pooling of data across recording session.

To assess disparity tuning, stimuli were presented via haploscope, and stimuli in the right eye were shifted by randomly varying degrees compared to the left eye. Right eye stimuli have an equal probability (p=.45) of being either the same as the left eye's or being its inverse (anti-correlated stimuli, where each black bar is turned white and each white bar is turned black), with the remaining 10% of stimuli being uncorrelated.

3.4.3 Disparity tuning curves:

To compute disparity tuning curves, we performed a forward correlation analysis. Simply put, we computed the response in a 30ms window around the peak response for each neuron. For each disparity and correlation, we first identified all the patterns which were presented with the given stimulus parameters. For each pattern, we computed the number of spikes observed in bins tmax – 1, tmax, and tmax + 1, where tmax is the time bin with the largest variance across disparities. The mean spike count was then computed for every disparity/correlation, giving a mean spike count for each disparity-correlation combination. The same exact procedure was performed for both the real cells and their model counterparts, with the only difference being whether the sequence of spike counts was predicted or observed.

3.4.4 Disparity Variance Fraction (DVF):

Because the experimental design independently assigned disparities and patterns (as noise stimuli), disparity and pattern in any given frame are independent. This allows us to apply the law

of total variance to divide the explainable variance of the neural response into the disparity variance, and pattern variance. We calculated the explainable variance in a model-free way, by assessing the variance between trials where stimulus sequences were repeated (Cooper Roddey et al., 2000). The disparity variance is defined as the portion of the neural response we can predict using only the disparity information of a frame, not its actual contents. We identify disparity variance by taking the true disparity information for each frame (disparity in pixels, and anticorrelated/correlated) and encode the information into a 1-hot vector. We then fit a linear model with temporal regularization to this vector to get a model prediction of expected activity based solely on the disparity information. The variance of this prediction is the Disparity Variance (DV). We then calculate Pattern Variance (PV) by subtracting the DV from the explainable variance of the neural response. Finally, DVF is calculated as the ratio of DV over total variance.

3.4.5 BEM fitting procedure:

The BEM can be treated as an instance of the class of 'linear–nonlinear' (LN) models that have been widely used to describe responses in early visual cortex and can be fit to data (Rust et al., 2004). These models linearly fit a number of spatiotemporal filters *k(i)* to the high-dimensional stimulus *s(t)*; each *k* is convolved with the stimulus, then passed through a nonlinearity, marking the first Linear-Nonlinear step. The nonlinearity in this case is simple quadratic function, which marks the model as an energy model, as opposed to rectified linear units (ReLU) used in the BCM. Each filter output is then summed linearly, and the sum is again passed through a spiking nonlinearity in the second LN step (Park and Pillow, 2011).

4 V1 mechanisms at the center of gaze

This chapter is a draft of an upcoming publication and was written with significant inputs from Bevil Conway and Daniel Butts.

4.1 Introduction

High-acuity vision depends upon the integration of spatial and chromatic signals. The three photoreceptor types (cones) in the retina that translate light into neural signals implicitly subserve all aspects of photopic vision, including so-called achromatic processes of object recognition and motion perception, as well as color perception. The highest visual acuity is achieved at the center-of-gaze, corresponding to the fovea, where the receptive-field center of each retinal ganglion cell is supplied by as few as a single cone (Wässle et al., 1990; Bringmann et al., 2018). Despite the importance of foveal retinal signals for vision, the mechanisms in primary visual cortex (V1) that process them remain almost entirely unknown. This is both due to the inherent challenges in studying color processing (Weller and Horwitz, 2017) as well as the ability to control the stimulus meant to probe neuron's at sufficiently high resolution, due to the challenge of precisely tracking the small eye movements that occur during normal vision at the level of cone spacing (Martinez-Conde and Macknik, 2008; Rucci and Poletti, 2015).

The high resolution of foveal processing presents a challenge for the processing of color, which requires a comparison of the activity of cones with different spectral tuning (cone opponency, meaning a comparison between long (L), medium (M), or short (S) wavelength-preferring cones) that receive the same signal. Namely, pooling a visual stimulus across

neighboring cones to extract chromatic information necessarily means sacrificing spatial acuity at single-cone resolution. The finest chromatic spatial acuity could, in theory, be achieved by a retina consisting of a regular alternating array of cone types, but the random cone mosaic in foveal retina (Schein, 1988) as well as the rarity of S cones (Roorda and Williams, 1999) makes it likely that chromatic acuity would be many factors lower than cone spacing. High-acuity spatial processing, meanwhile, benefits by a comparison of the activity of neighboring cones, but also would be compromised by having cones of different spectral tuning. It thus is an open question whether foveal V1 neurons somehow maintain high-acuity color signals, sacrifice their high-acuity to subserve color, or separately process color and high-acuity luminance information and color (which is generated via cone-opponent processing).

Here, we address this question by recording from foveal V1 in macaque during presentation of spatiotemporal-chromatic noise which explores the full space of possible combinations of cone inputs. We are able to measure foveal RFs with high spatial resolution by combining large-scale multi-electrode recordings of V1 and expanding upon a model-based eye tracking procedure that can infer eye position based on observed neural activity (McFarland et al., 2014).

When fitting linear and nonlinear models to this noise data, we indeed find a clear dichotomy between high-acuity spatial processing and broader-scaled color processing. The most high-acuity spatial processing, with receptive field components on the scale of 2-3 arcminutes, is performed by cells not modulated by color. We additionally find that RF subregions are several times smaller than predicted by previous studies in V1 with no clear dependence on eccentricity within the fovea. The fine scale of RF subregions may allow cells to represent different

arrangements of fine-scale inputs throughout the overall RF, which would facilitate efficient representation of foveal inputs at multiple scales. By comparison, a smaller subset of foveal V1 neurons were cone opponent and had significantly larger RFs and distinct patterns of spatiotemporal processing, including a general lack of orientation selectivity and nonlinear processing. In between these two extremes were a small subset of "double-opponent" (D-O) neurons responding to both high-acuity inputs and color. However, unlike the D-O cells reported in the parafovea which respond to the same spatial scale for both color and spatial opponency (Conway, 2001), our population of foveal D-O cells responded to high-acuity spatial processing in the luminance channel but had larger cone-opponent RFs; this finding indicates that in foveal cells, cone computations across space are likely limited by cone sampling. In total, we conclude that foveal V1 RFs are extremely fine and apply different computations to inputs from overlapping spatial positions, but cone sampling prevents cone-opponent processes from using the same levels of spatial acuity.

4.2 Results

We recorded from one male macaque using two multielectrode arrays chronically implanted in foveal V1 in one hemisphere and a laminar probe inserted acutely in a chamber over foveal V1 in the other hemisphere. We recorded 424 well-isolated single and multi-units, with 154 from the arrays and 270 from laminar probes across 20 recordings.

We designed a color cloud (CC) stimulus in DKL space (Fig. 4.1A and 4.1C, also see methods) to optimally explore spatiochromatic combinations while still strongly driving V1

neurons. In addition, we presented a battery of other stimuli, including rapidly flashed Hartley series gratings varying in orientation, spatial frequency (SF), phase, and color (Fig 4.1B, similar to (Johnson et al., 2008), see methods) to get more classical tuning characterizations for the same cells.

We then applied an eye-tracking (ET) algorithm to determine the precise location of the eye during each fixation, defined by microsaccades. Briefly, the algorithm first fit models assuming perfect fixation to identify a first-pass estimate of RF locations and structure. It then used those models to predict the activity for a population of cells during a given fixation for any possible eye position. By then comparing the observed activity during that fixation with the predicted activities, it generated a likelihood surface predicting the most likely true eye position for each fixation (Fig 4.1D, bottom). By only including cells recorded from the chronic arrays in this procedure, the laminar probe cells act as cross-validation avoid overfitting – if the models for laminar cells improved by applying ET correction despite not being used in the algorithm, then algorithm must have captured the real eye position. The ET algorithm significantly refined and sharpened RF maps (Fig. 4.1E). Across the population, eye-tracking was associated with a 1.8-fold increase in log-likelihood scores (Fig. 4S1A) and a 20% reduction in RF size (Fig. 4S1B).

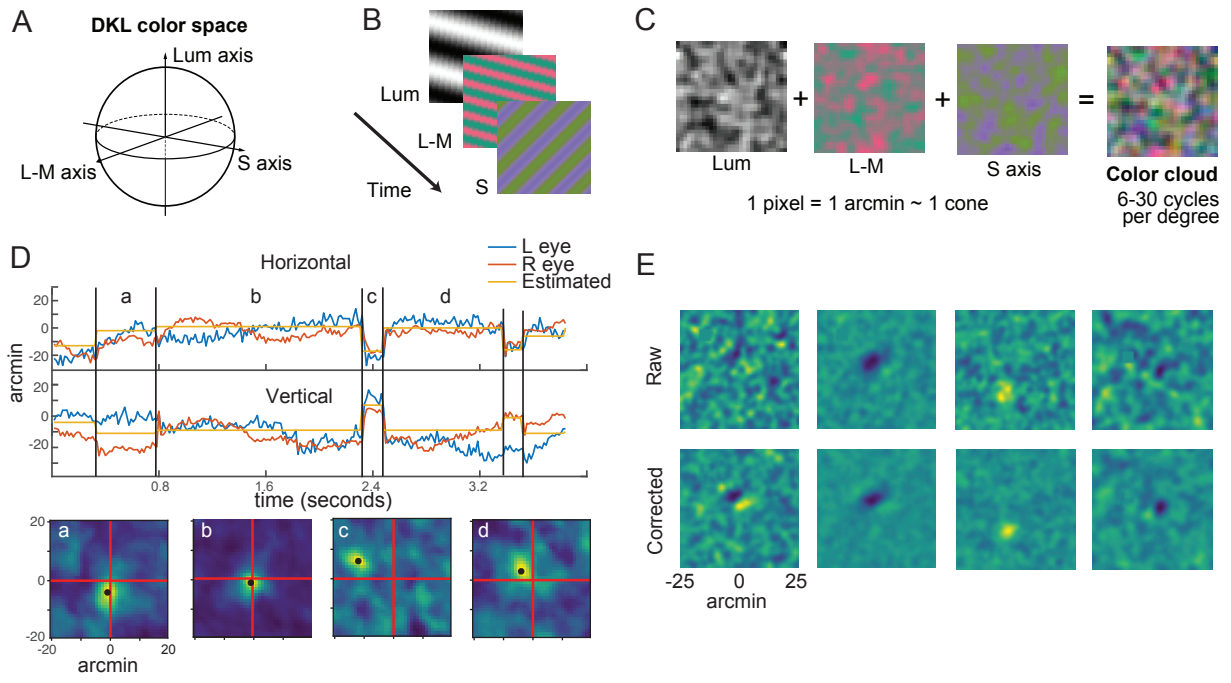

Figure 4.1: Foveal receptive fields using spatiotemporochromatic noise and model-based eye tracking.

A: All stimuli were defined using the DKL space (Derrington et al., 1984), which defines inputs along three axes: luminance, L-cone vs M-cone activation, and S-cone vs L+M cone activation. This color space is physiologically defined to approximate the cone-opponent processes occurring in LGN, and thus is an approximation of the structure of LGN outputs to V1. **B**: We also presented rapidly flashed Hartley grating stimuli allow for characterization of classical stimulus tuning properties: orientation, spatial frequency, phase selectivity, and color preference. **C**: Color clouds stimuli were created by using a gaussian filter applied to independent white noise along each of the three DKL axes, to optimally drive V1 cells in their preferred range of spatial frequencies. **D**: Top: examples of horizontal and vertical eye traces from a video-eye tracker during

one 4-second trial. Vertical black lines indicate the times that microsaccades were detected; blue and red traces show the eye traces from the camera eye tracker; orange traces show the final estimate of eye position based on the ET algorithm. Bottom: Examples of 2-dimensional likelihood surfaces for corresponding fixations. Intensity at each position indicates the likelihood that the eye was positioned at that spatial position during that fixation. The black dots correspond to the peaks in these likelihood surfaces. Long and relatively stable fixations, such as b, yield the clearest surfaces, Shorter fixations, such as c, have less data to integrate evidence from, and thus yield noisier decision surfaces, but even these usually still generate a single peak, meaning one clearly best estimate of eye position. **E**: Eye correction significantly improves receptive field estimates and enables measurement of foveal RF properties. The procedure refines RF estimates, and in many cases heavily sharpens features to yield RF maps more accurately representing the high spatial resolution of the fovea. In some examples, the procedure reveals RFs that were otherwise undetectable.

Using the inferred eye-position, we shifted the stimulus to correspond to what was seen in retinal coordinates, and then fit LNLN-cascade style models with one linear and two quadratic subunits (General Quadratic Models (GQMs), see Fig 2A) to recorded V1 neurons using statistical modeling approaches (Mante and Carandini, 2005; McFarland et al., 2013). The GQM yields spatiotemporal-chromatic receptive field maps that can be used to describe any kind of V1 tuning well: simple cells are described purely by the linear filter (example in Fig 2B), and the nonlinearity of complex cells can be captured by the quadratic filters (example in Fig 2L). The linear filter

approximates any linear receptive field that could be mapped using spike-triggered averaging but has the added benefit of regularization. The quadratic filters can capture most V1 nonlinear processes by arranging intro "quadrature pairs", 2 gabor-like filters that are 90^{o} out of phase with each other – because the outputs of these filters are squared, they can respond equally to themselves and their inverse, meaning that 2 quadratic filters in quadrature can describe an oriented receptive field at any phase, thus capturing phase invariance Thus, by combining these filters, any combination of processes along the simple-complex spectrum (Mechler and Ringach, 2002; Carandini, 2006) can be described, and the contribution of each filter to the model's overall firing rate prediction can be used as a reliable estimator of the "complexness" of the cell.

We then extracted basic tuning properties from these RF maps, including RF width (see Figure 3), orientation tuning, color modulation, spatial frequency preference, and direction selectivity (see methods, and (Bartsch et al., 2022)). We validated these model-based metrics using the observed responses to Hartley gratings, which can also capture orientation, SF, and color mmodulation. We found that model-based measures for orientation and SF replicate the measures extract using Hartley measures (Figure 2H-K, R-U).

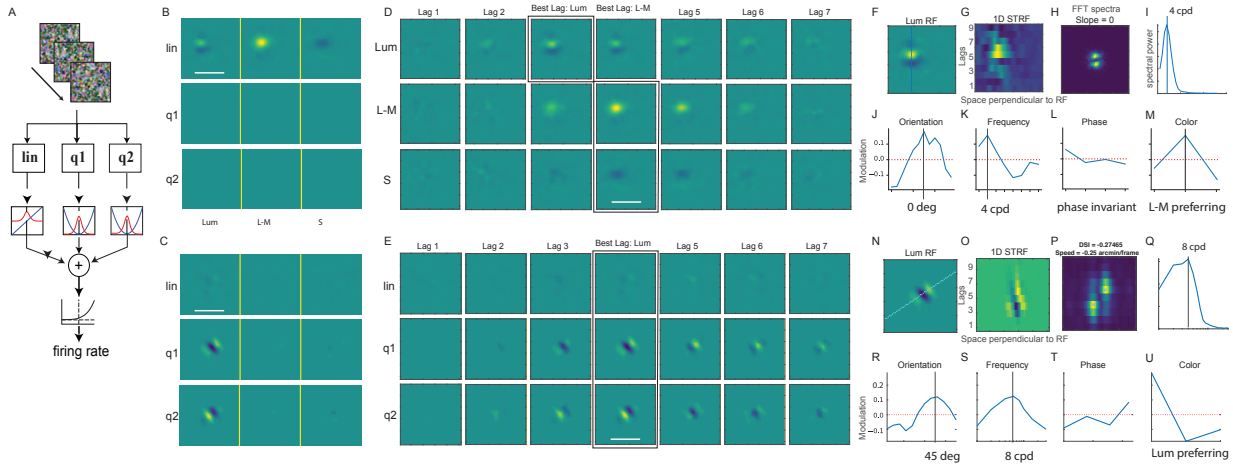

Figure 4.2: Detailed RF characterization reveals linear and nonlinear spatio-chromatic properties.

A: Schematic of our model. We applied a simple LNLN cascade with one linear and two quadratic subunits, an implementation of the classical energy model. **B**: An example of a simple color selective cell. Despite having access to nonlinear computations, the model did not use them, indicating that this is a simple cell. The filter's total projection into each color axis was used to calculate the cell's overall color modulation. The white scale bar indicates 30 arcmin. **C**: An example of a complex luminance-only cell. Although a linear approximation of this RF is possible, it much better explained by adding the quadrature pair of filters. **D**: Filter evolution over time for the cell in B. We can separate out the spatial dimension of each filter and identify selectivity to stimulus changes in time, such as direction selectivity. **E**: Same as D, for cell C. F-I: Extracting measures of orientation and direction selectivity from filters in B. **F**: We use the 2DFFT of the filter to identify the axis of the cell's preferred orientation. **G**: Bisecting the RF along its preferred

axis extracts its 1D spatiotemporal RF. **H**: Taking the 2DFFT of the 1D STRF and comparing the strengths of Q1 and Q2 yields the disparity selectivity index. **I**: Extracting the preferred axis of the 2DFFT yields the model's SF preference curve, in this case with a peak at 4 cycles per degree. **J-M**: Independent Hartley analysis confirms the model-based orientation (J), SF (K), and color tuning (M), as well as lack of phase selectivity (L). **N-Q**: Same as F-I, for the example cell in C. Note that this cell demonstrated clear direction selectivity. **R-U**: Same as J-M, but for the example cell in C.

4.2.1 Scales of processing in foveal V1

The most obvious question in studying foveal processing is the resolution of computations in foveal receptive fields. To asses RF size in a way comparable to previous studies (e.g. (De and Horwitz, 2021)), we first identify the spatial power at each pixel by squaring the weights of each filter at each spatial position, summing across time and color dimensions. We then binarize the resulting spatial power spectrum, setting a threshold to capture 90% of the total filter power. Finally, we identify the boundaries of the resulting RF contour using a contour function (Python, Scikit-learn toolbox) (Fig. 4.3A). Across the measured eccentricities, we find a wide range of RF sizes, centered at 0.1 degrees (Stdv 0.12^{o}) for cells only modulated by luminance (luminance weight >0.9, N=185), with no significant effect of eccentricity on RF size (Fig. 4.3C). Given the prevalence of cortical magnification across the visual field, and the cortical overrepresentation of the fovea providing ample cortical area to extend magnification effects across the fovea, we

expected to see some effect of eccentricity on RF sizes, but find no evidence of this here. This result indicates that cortical magnification does not extend to the center of gaze, and instead foveal receptive fields operate on the same range of spatial scales whether they are right at the foveal center or at 1 degree. Furthermore, the significant overlap between receptive fields (Fig. S3) means that many different kinds of processing can be applied to inputs from the same retinal location.

We compared this to neurons that had any degree of color modulation (luminance weight <0.9, N=239), which had RFs are significantly wider (Fig. 43D). For these neurons, we do observe an effect of increasing eccentricity within the fovea being associated with larger RFs. These results highlight a qualitative difference in how luminance contrast is used to compute shape information at the highest possible resolution, while cone-opponent processes, while still at quite fine resolution in the fovea, still have to integrate over larger retinal regions to compute color.

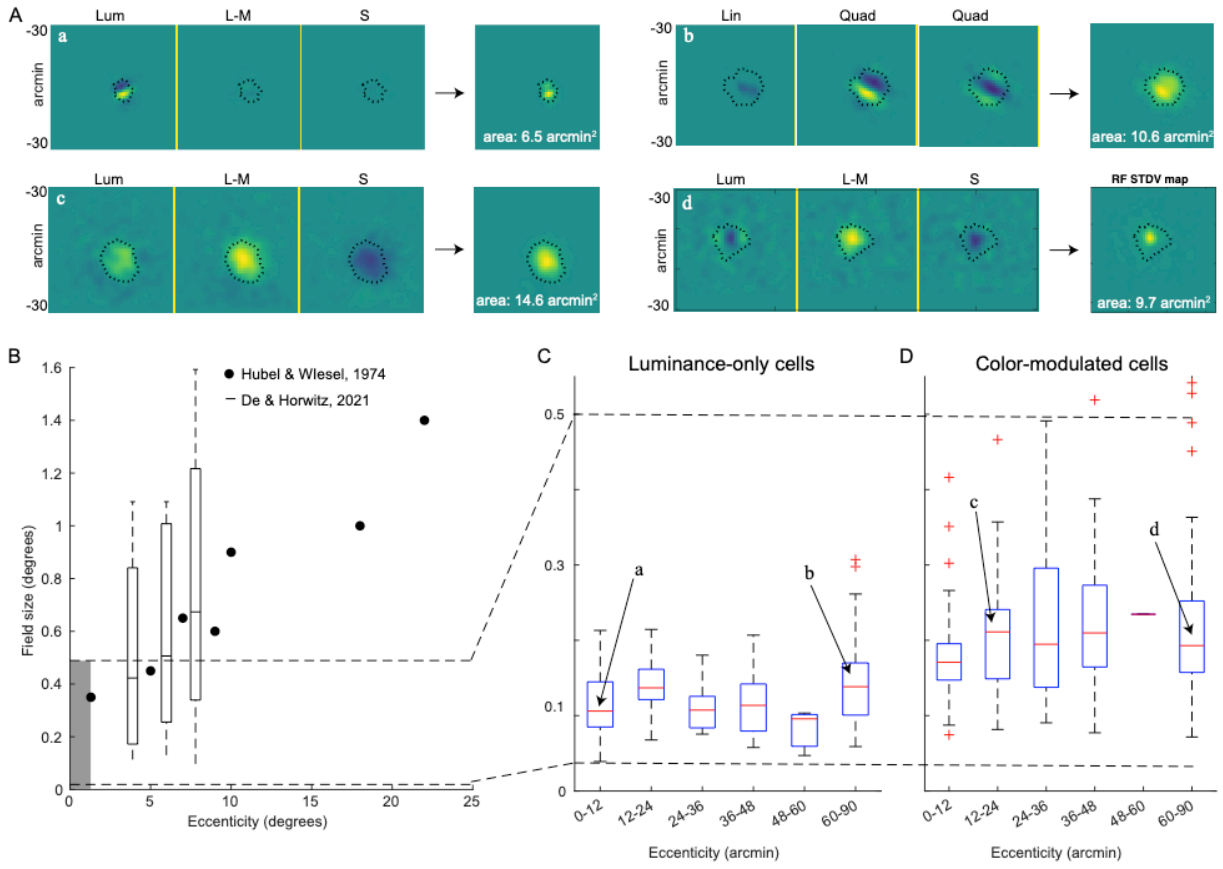

Figure 4.3: Spatial scales of foveal processing.

We extract the size of foveal RFs and show that luminance signals are privileged for spatial processing and remain fine throughout the fovea, establishing a dichotomy between low-spatial-resolution cone-opponent and high-spatial resolution luminance cells. **A**: Examples of RF area estimation. To determine the cell's RF area, we calculated the total filter power for each spatial position across all lags and colors, then identified the contour of the resulting RF spatial power map. Examples: a) a simple-like, luminance only cell recorded at 8 arcmin. b) a complex luminance-only cell recorded at 68 arcmin. c) a single-opponent color cell recorded at 20 arcmin.

d) a color cell recorded at 65 arcmin. **B**: RF area across eccentricity based on previous studies (Hubel & Wiesel 1974, De&Horwitz2021). For reference, the gray bar represents the scale and extent of our measurements. **C**: Our measures of RF area in foveal luminance-only neurons. We find that throughout the central 1 degree of vision, luminance-only cells have sizes less than 12 arcmin and can be as small as 2-3 arcmin. We also do not find a significant effect of eccentricity on RF size, indicating that V1 RFs operate at the same range of spatial scales throughout the fovea. **D**: Our measures of RF area in foveal cone-opponent neurons. While they are still fine, cone-opponent cells have significantly larger RFs, and their width also increases with increasing eccentricity even within the fovea.

We found that the distinction between luminance-only and cone-opponent cells extends beyond their different RF sizes. We selected cells with purely luminance tuning (lum weight >0.9, N=215) and cells with significant color weight (either L-M or S weight >0.2, N=133) to ensure we could extract correct information for these measurements from the RFs. Extracting the orientation selectivity and SF tuning of these populations using the Hartley stimuli (estimated as modulation by orientation or SF relative to total modulation across all gratings), we find that luminance-only cells are more orientation selective (Fig. 4.4A, one-tailed Wilcoxon rank-sum test, p=.0001) and tuned to higher spatial frequencies (Fig 4.4B, one-tailed Wilcoxon rank-sum test, p=.01) compared to color-modulated cells. Additionally, we find that color-tuned cells in our sample extremely rarely show any direction selectivity (Fig 4.4C). This result makes sense given that cone opponent

processes require spatial integration of multiple cone types, and computing cone opponency alongside movement across the cone mosaic would be computationally difficult.

However, we do find a small population of double-opponent (D-O) cells, which respond both to orientation and spatial structure while still being strongly modulated by color (an example is shown in Fig 4.3D). Just like D-O cells reported previously, these cells were rare and represent less than 5% of our sample. However, even among these cells, we did not observe any direction selectivity, indicating that cone-opponent processing and direction selectivity are indeed mutually exclusive.

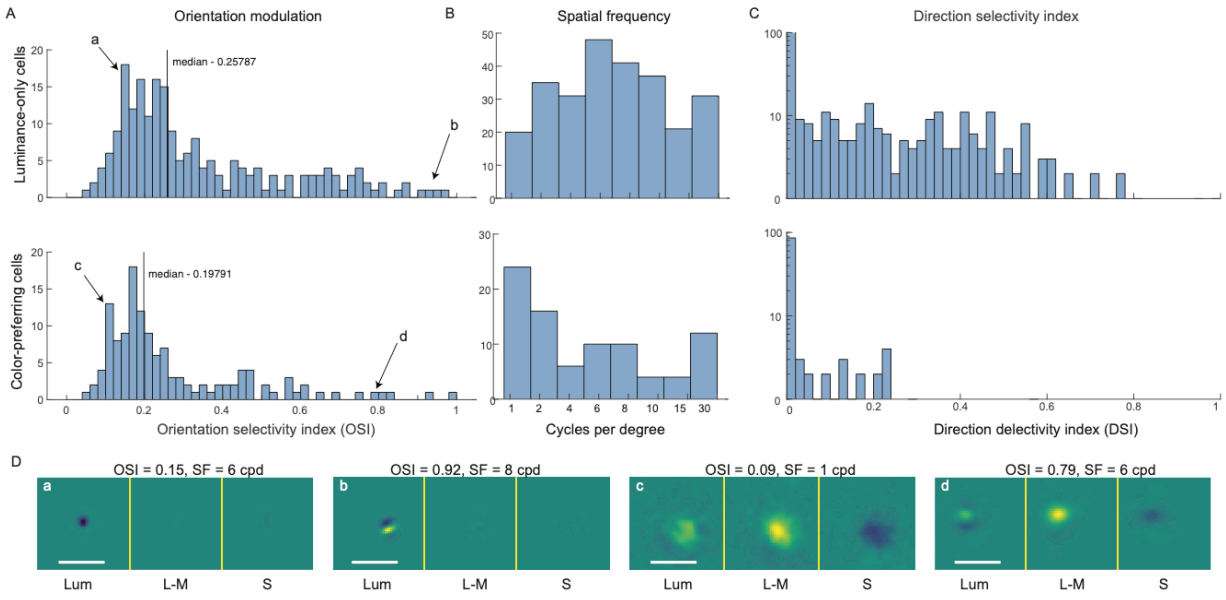

Figure 4.4: The dichotomy between cone-opponent and luminance processing is present in classical tuning measures.

While some previous studies using parametric stimuli in parafovea found that spatial acuity is the same across color channels (according to Johnson, Shapley, Hawken), our data shows that

cone-opponent processes operate at lower spatial scales. **A**: Orientation selectivity index (OSI) for luminance (top) and color (bottom) cells. Cells with color modulation tend to show less preference for orientation on average. The color-modulated cells that do respond to orientation (ex. d) fall into the category of D-O cells. **B**: Preferred spatial frequency for the same population of luminance and color cells. Cells with stronger cone opponency are also usually tuned to lower SFs, further demonstrating that color is processed at a lower spatial scale than luminance. **C**: Direction selectivity index (DSI) for the same populations. Direction selectivity (a property necessitating high spatial acuity over time) and cone opponency are effectively mutually exclusive tuning properties, as hypothesized by Conway & Livingstone, 2001. **D**: Example cells demonstrating the dichotomy between high-acuity luminance and coarser color processing. a: A fine, monophasic RF lacking orientation selectivity. b: a similarly fine, but highly orientation selective luminance-only cell. c: a single-opponent color cell with a typical, large, unstructured, monophasic RF. d: A D-O cell with high orientation selectivity as well as strong color selectivity. These cells are exceedingly rare and only represent about 5% of our sample.

4.2.2 Joint representation of color and shape in foveal V1

Prior to our experiments, it was not clear whether foveal cells could exhibit double-opponency, given the density of cones and random cone mosaic. In the parafovea, large RFs have an abundance of cones to sample from and can use gain control mechanisms to generate cone-opponency alongside spatial opponency (Fig 4.6A). However, in the fovea, cone sampling may prevent arrangements suitable for D-O cells in many circumstances (Fig 4.6B).

We did find clear evidence of double-opponency in our sample, but also observed some striking differences compared to previously reported D-O cell properties. Whereas parafoveal D-O cells showed the same spatial scale for both color and shape selectivity, our cells display an arrangement where orientation selectivity is only present in the luminance dimension, while the RF components in the color channels are monophasic and tend to cover most of the cell's overall RF (Fig 4.6C, left).

To test how these cells sample cone inputs to generate this selectivity we translate the filter projection in DKL space into cone-increments (Fig 4.5C, middle; also see Brainard). This analysis revealed that these RFs are generated by cone-opponent processes, with centers integrating L or M ON signals from one cone type, and surrounds integrating OFF signals from the opponent cone type (Fig 4.5C, right). This arrangement yields clear shape selectivity due to the ON and OFF regions, but because the inputs come from opponent cones, the full field of the RF is selective to the same direction in DKL space. This kind of RF construction makes sense in the context of an RF sampling from few cones which are not arranged in a way that would facilitate orientation selectivity (for example, Fig 4.5B), and may be a way for foveal V1 to jointly represent color and shape in an efficient way despite limited cone sampling options.

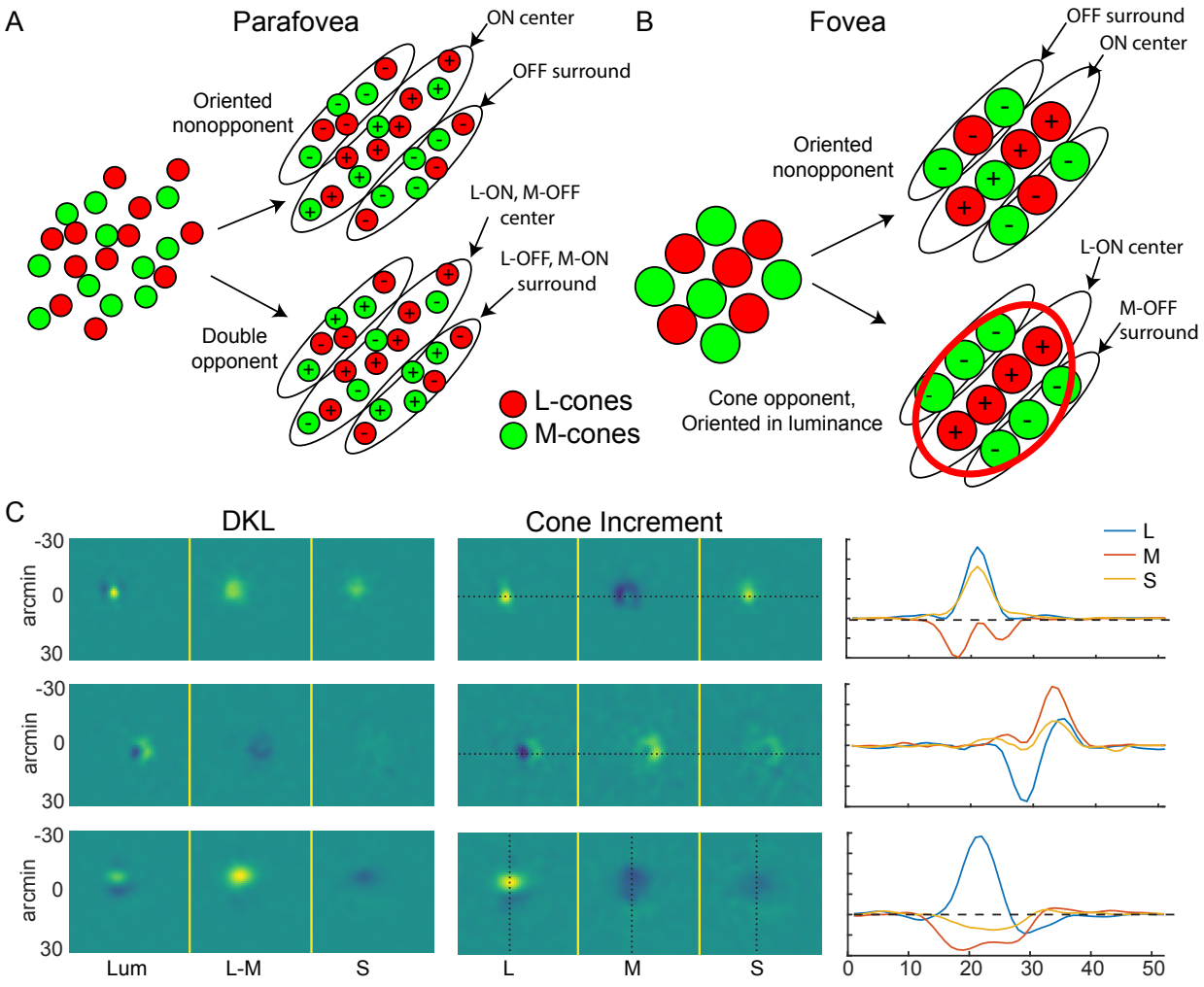

Figure 4.5: Combined representation of shape and color in foveal V1.

A: Example of cone sampling leading to either an oriented non opponent or a double-opponent cell. **B**: Example of foveal cone sampling. Because of the high resolution of foveal RGCs, foveal V1 neurons likewise sample from few cones. This means that traditional wiring of double-opponent cells is difficult to achieve but can lead to a wiring leading to orientation in ON vs OFF responses, but the same tuning in cone-opponent terms. **C**: Example of a receptive field of

two such D-O cells, showing a cone opponent, oriented in luminance (COOL) arrangement. RFs on the left show the original filter fit in DKL space, while the middle column shows the same cell fit in Cone-Increment space, revealing how their selectivity is constructed. The Right column shows a 1D representation of the cone inputs sampled to construct these RFs.

4.2.3 Increased spatial resolution of luminance inputs is facilitated by subregion structure.

Our finding of certain cells that jointly represent both luminance and color prompts the question of whether the tradeoff between spatial resolution and cone-opponent processing manifests as different classes of cells, or different computations performed on the DKL channels. To test this, we identified the RF subregions of each channel independently and quantified the size of the dominant subregion relative to the cell's overall RF (Fig 4.6. A,B). We do this by splitting each receptive field into positive and negative components, thus separating it into On and OFF activation regions; we then apply the same thresholding and contour detection approach used to detect the full RF sizes. For each cell and channel, we then determine the activation region with the great weight on the channel as determined by the sum of the weights at each pixel, and take this to be "dominant" subregion. In this analysis, for each channel, we only include measurements from cells where at least 20% of the filter is represented in that channel. This ensures that our measurements quantify real RF components and are not capturing noise.

Assessing the subregions in the luminance channel, we find that RF subregions become increasingly finer within larger RFs (Fig 4.6. C), meaning that even in foveal cells that integrate over larger retinal fields, high-resolution information can be represented by maintaining fine

subregions. We find much more linear increases in subregion size with RF size among the L-M (Fig 4.6. D) and S channels (Fig 4.6. E), indicating that the preference for high-resolution processing indeed is specific to the luminance dimension.

We further find that subregions in all three channels (Fig 4.6. H-J) are finer than the total RF sizes (Fig 4.6. F-G). However, the effect was significantly more pronounced in the luminance channel, further lending credence to our interpretation that lummiance is privileged for high spatial acuity processing.

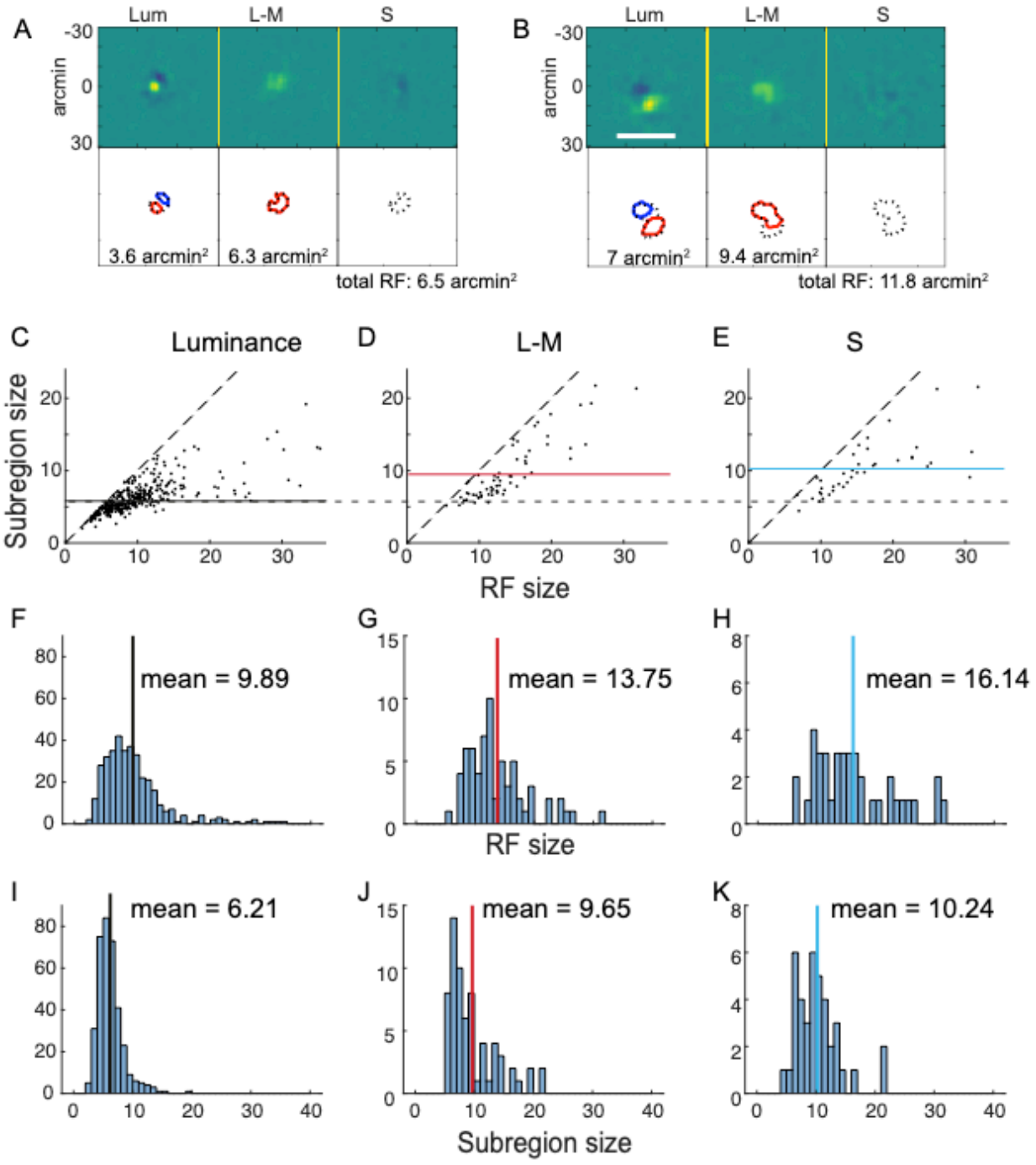
A
Lum
L-M
S
-30
0
30
arcmin
3.6 arcmin²
6.3 arcmin²
total RF: 6.5 arcmin²
B
7 arcmin²
9.4 arcmin²
total RF: 11.8 arcmin²
C
Luminance
D
L-M
E
S
Subregion size
RF size
F
mean = 9.89
G
mean = 13.75
H
mean = 16.14
I
mean = 6.21
J
mean = 9.65
K
mean = 10.24
Subregion size

Figure 4.6: Receptive field substructure in the luminance domain allows for finer resolution even in larger RFs.

A-B: Example of estimation of RF subregions, with RFs on top and corresponding contours on the bottom. RF areas are computed via the same contour detection as full RFs (see Figure 4.3). **C-E**: RF size plotted against the size of RF subregions. Color subregions tend to increase alongside the overall RF size, but luminance subregions remain small even as overall RF size gets larger, meaning that even large RFs can represent high acuities. **F-H**: RF size distributions for luminance, L-M, and S channels respectively, plotted for cells with at least 30% modulation in the respective channel. **I-J**: RF subregion distributions for the same cells as in the above plots. Subregions are by definition finer than the full RFs, but the effect is especially pronounced in the luminance dimension.

4.3 Discussion

Here, we present the first cone-resolution spatiotemporal-chromatic maps of foveal V1 neurons. We find that individual receptive field features can be as small as 3 arcmin (0.05 degrees) and that larger foveal RFs may still maintain that resolution when integrating over larger numbers of these features. Furthermore, we find that these foveal RFs have the highest spatial resolution in the luminance dimension, while cone-opponent processes appear to operate at lower spatial scales. We further demonstrate double-opponency in the fovea, but note an important difference from parafoveal cells in that in foveal D-O cells, the spatial opponency is primarily observed in the luminance dimension. This result further highlights that luminance contrast is the primary signal

used for shape processing in the fovea and follow the proposal that the parvocellular pathway is designed for high-acuity processes, while color "gets a free ride" (Martin et al., 2011).

It is also noteworthy that among luminance-preferring cells, we find no increase of RF size with eccentricity, and even among color cells find only a minor effect. This implies that cortical magnification, which is well documented across the visual field, may not hold within the fovea itself, and thus has significant implications for retinotopy and hyperacuity.

A particular strength of our approach here is the use of laminar probes which cover the full cortical depth, compared to recent studies using 2-photon imaging and only observing supragranular color responses (Garg et al., 2019; Li et al., 2022). However, even when comparing processes across laminae, few coherent patterns of processing have emerged beyond categorical observations that more complex color processing seems to occur in supragranular cells as described previously (Lennie et al., 1990; Skottun et al., 1991; Li et al., 2015).

While our stimulus design does include higher luminance contrast on average, these results could not be explained by the lower color contrast in the stimuli: because our model-fitting procedure is probabilistic, it can identify even subthreshold inputs into cells. The fact that we observe many cells with equal ratios of luminance and cone-opponent inputs indicates that luminance contrast was not dominating our findings. Additionally, because our stimuli ranged the full gamut of the monitor's capabilities, our stimulus set reflects the fact that luminance contrast is similarly dominant in the statistics of the natural viewing conditions our visual systems are adapted for. Future work may investigate whether the cone-opponent, oriented in luminance

(COOL) cells we describe here still show their asymmetry for resolution in luminance when presented with gratings matching luminance and cone contrast.

Our results of foveal D-O cells manifesting as COOL cells are surprising, as other studies have described parafoveal double-opponent (DO) cells demonstrating spatial and color opponency at the same scales (Conway, 2001; Johnson et al., 2008; Shapley and Hawken, 2011). Thus, it's likely that COOL cells are a specifically foveal phenomenon, and future work will aim to verify this by expanding upon this dataset to measure more parafoveal receptive fields with the same stimuli and eye-tracked models. If our assertion is true, then RF maps of more parafoveal cells generated using the same stimuli and modeling approaches should closely resemble classical double-opponent cells. If this is not the case, it would indicate that our cloud stimuli drive D-O cells in a specific way that causes them to manifest double-opponency as COOL cells. However, our interpretation that COOL cells arise due to limitations in the cone mosaic and do not require very specific circuitry is well in line with recent work showing that DO cells may use similar circuitry as nonopponent simple cells (De and Horwitz, 2021). Clearly, more advanced nonlinear processing will be necessary to accurately describe the computations performed across chromatic inputs. Horwitz showed that quadratic computations capture color processing in V1 better than simple linear or difference-of-gaussian operations (De and Horwitz, 2021) but the truth is likely significantly more complex and will require likewise complex analysis using nonlinear models.

Our finding that S-cone selectivity appears to be present throughout the fovea also has significant implications about the nature of foveal processing. Adult monkey foveas have an overall S-cone foveal density of 10%, with several areas lacking a few S cones that are not

coincident with the area of highest cone density. In contrast, adult humans have a distinct but variable central zone about 100 μm wide that lacks S cones and is surrounded by a ring in which the S-cone density is 8% (Bumsted and Hendrickson, 1999). Based on the best available estimates of foveal cone density (Wells-Gray et al., 2016), fine foveal RFs like the one shown in Figure 4.6A should be constructed with about ~50 cones, or about 25 cones for each subregion. Assuming a 10% S-cone ratio, this means that each subregion has a 92.02% probability of containing at least one S-cone, meaning that even highly foveal cells should have a high chance of representing S-wavelength information. Naturally, RF maps based purely on S-cone inputs would operate at lower spatial resolutions; however, because S-cone signals are compared to L and M signals, V1 cells can integrate those signals with S-cone opponent processes to generate smooth RF maps. Given the possibility that chromatic aberration reduces the spatial resolution of S-cone signals somewhat, this may even be the most efficient way to encode such opponency. This calculation should remain true for humans outside the ~100 μm S-cone hole at the center of gaze. The reported 8% S-cone ratio ring surrounding the central zone should be similarly capable of using L and M signals to fill in information conveyed via S-cones. Indeed, it may even be possible for cells to calculate S-(L+M) opponency using an S-cone in the ring and L+M signals from the foveal center to fill in what S-cone opponent information there is available at the center of gaze.

There is also a valid concern that chromatic aberration may be suppressing the ability of our stimuli to drive S-cone responses. Chromatic aberration refers to the phenomenon of different wavelengths of light being refracted by different amounts when passing through a lens, such as the lens of the eye, causing them to focus on different points. Shorter wavelengths (corresponding to

blue light) are more strongly refracted, and thus are focused closer to the lens. There are two main types of chromatic aberration: transverse (or lateral) chromatic aberration (TCA), and longitudinal (or axial) chromatic aberration (LCA). The two types of chromatic aberration have different characteristics and may occur together. TCA occurs when different wavelengths of light are focused at different positions in the image plane. In the case of the eye, this means that chromatic errors will reduce the contrast of the retinal image at lower wavelengths more than high wavelengths, and the reduction will generally increase with eccentricity in the peripheral field. Because we presented stimuli at the center-of-gaze, the effects of TCA should be considered minimal - the most eccentric receptive fields (at 1.5 degrees eccentricity) presented would still have an expected TCA of less than ~0.3 arcmin, much lower than the resolution of the retinal mosaic (Winter et al., 2016). If we did observe TCA effects, they should manifest as an offset between the L-M and the S components of a given color RF; however, we consistently observe those components in overlapping spatial positions, with any differences in substructure much more easily attributable to differences in processing of the same cone mosaic. LCA occurs when different wavelengths of light do not converge at the same point along the optical axis, causing a variation in focus along the axis. Thus, different colors within the image plane may not be in sharp focus simultaneously, resulting in a lack of color convergence at a single depth. In other words, LCA renders the eye to be relatively more myopic for shorter wavelengths than for longer ones (Wang et al., 2008). This effectively reduces the relative contrast of S-cone activity and may be the evolutionary reason for the lower density of S-cones in the fovea, as the S wavelength signal may simply lack high spatial fidelity relative to L and M wavelengths. However, there are two

reasons to believe that LCA did not significantly skew our results: 1) our detection of cells modulated only by S-cone inputs, which we should not be able to find if S wavelength rays were being significantly suppressed; and 2) the chromatic cloud stimuli we used were generated with spatial smoothing to cut off high spatial frequencies, and this smoothing would dilute the effects of LCA. as the slow changes in S wavelength power inherent to the stimulus minimize the effects of the greater scattering of S wavelength rays.

4.4 Methods

4.4.1 Subjects:

One male rhesus macaque (Macaca mulatta), weighing 14 kg, designated as monkey J. The animal was implanted with a plastic (Delrin) chamber and headpost. Surgical implantation protocol has been described previously (Lafer-Sousa and Conway, 2013). J was implanted with a chamber over the left hemisphere and two chronic arrays in the right hemisphere. All procedures were approved by the Animal Care and Use Committee of the National Eye Institute and complied with the regulations of the National Institutes of Health.

4.4.2 Electrophysiology:

In monkey J, we implanted two chronic multi-electrode arrays (Utah Array, Blackrock microsystems; 96 contacts, 100-µm spacing, and Plexon N-form Array, Plexon; 100 contacts) in the right hemisphere, while laminar probes (V-probes, Plexon, 24 channels, 50-µm spacing) were

inserted through a chamber implanted over the left hemisphere using a Narishige hydraulic microdrive (Tritech Research). Voltage traces were digitized and saved with a Plexon MAP system (Plexon Inc.). Spike waveforms were sorted offline using Kilosort 2.5, and single units were defined on the basis of waveform and interspike interval.

4.4.3 Stimuli:

Stimuli were generated in MATLAB (MathWorks, Natick, MA, USA) using the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007) and displayed using a Display++ LCD monitor (Cambridge Research Systems, Kent, UK) with a resolution of 1920 × 1080 pixels, a mean luminance of 58 cd/m2 and chromaticity x,y = {0.30, 0.33}. Animals were seated at a distance of 1.27 m from the screen, at which it subtended 32 × 18 degrees of visual angle with a resolution of 60 pixels per degree. The monitor was calibrated using a PR-655 SpectraScan Spectroradiometer (Photograph Research Inc.); we used the Stockman-Sharpe cone fundamentals to calibrate the transformation from monitor RGB values to DKL space (Kleiner et al., 2007). Our color stimuli assume that monkeys and humans have the same cone fundamentals, which is justified because the cone absorption spectra are very similar in the two species.

4.4.4 Color clouds:

The color clouds used here are analogous to cloud stimuli used previously (Niel 1999), which offer the ability optimally stimulate V1 neurons at their optimal spatial and temporal frequencies while maintaining pixel- (i.e., cone-) resolution. Each cloud stimulus frame is generated by spatially band-passing a white noise stimulus. For each frame, three such

uncorrelated images were generated, and each was assigned to a channel in the DKL color space, (i.e., the luminance, L-M, S-cone axes). The spatial band-pass was chosen to optimally drive the recorded neurons, frequencies and the frequencies were determined based on the on-line measurements using the Hartley gratings.

4.4.5 Hartley gratings:

To get classical tuning measures for cells in our sample, we generated Hartley stimuli (Ringach et al., 1997) defined along the main axes of the DKL color space as described previously (Johnson et al., 2008). Hartley gratings were defined using 12 orientations (angles evenly spaced distributed with 15 degree spacing), 8 spatial frequencies (1,2,4,6,8,10,15,30 cpd), 4 phases, and the 3 DKL axes (luminance, L-M, and S). Gratings were presented at the same arcminute resolution and 60 Hz temporal resolution as the clouds.

4.4.6 Procedure:

Recordings were performed in a light-controlled room, with the animals perched in an upright position. Animals were acclimatized to head restraint to minimize head movement during recordings and trained to fixate on a white fixation dot (3-pixel radius) at the center of the screen. Eye position was monitored throughout the experiments using an infrared eye-tracker (ISCAN for experiments J220201 to J220611, Eyelink for experiments J220616 and onward).

One experimental session started by inserting an electrode and advancing until neural activity observed along the entire probe was maximized. Units were then preliminarily sorted

online using the waveforms, and the receptive fields of the isolated units were first mapped by hand. Stimuli were then centered on these mapped receptive fields and the recording session began. Cloud stimuli and the Hartley gratings were presented in 4-second blocks with a 500ms break between trials. If fixation was broken for more than 100ms (a duration chosen to account for eyeblink artefacts) the trial was aborted. Once a probe was placed recording continued until the animal stopped working, thus yielding a variable number of trials per session. Hartley stimuli were presented first for 100-200 trials, at which point this data was used in a quasi-online analysis to determine orientation and spatial frequency selectivity of the recorded cells. These parameters were then used to select which cloud stimulus parameters would optimally drive these cells.

4.4.7 Model optimization:

The model parameters were fit on 4/5 of randomly selected trials while the remaining 1/5 of the data were set aside for meta-parameter optimization, including regularization. A grid search was performed for each regularization parameter, with models being refit each iteration until an optimal set of model parameters was determined.

4.4.8 Extracting feature selectivity from models:

Once receptive field maps were generated for each neuron, these maps were used to quantify certain tuning parameters. A 2DFFT was applied to each filter to identify its 2D Fourier spectrum. The peaks of this spectrum were used to determine the optimal orientation of each filter (see Fig. 4.2D). The plane passing through the center of the receptive field was also used to determine the SF spectrum, the peak of which was determined to be the model-predicted preferred

SF. The pixels on this plane were also extracted for each time lag in the filter to generate a map of the RF's spatiotemporal evolution perpendicular to the RF orientation. Applying another 2DFFT transform to this spatiotemporal map yielded a spectrum of temporal evolution, which was used to determine strength and speed of direction selectivity (Fig 4.2D).

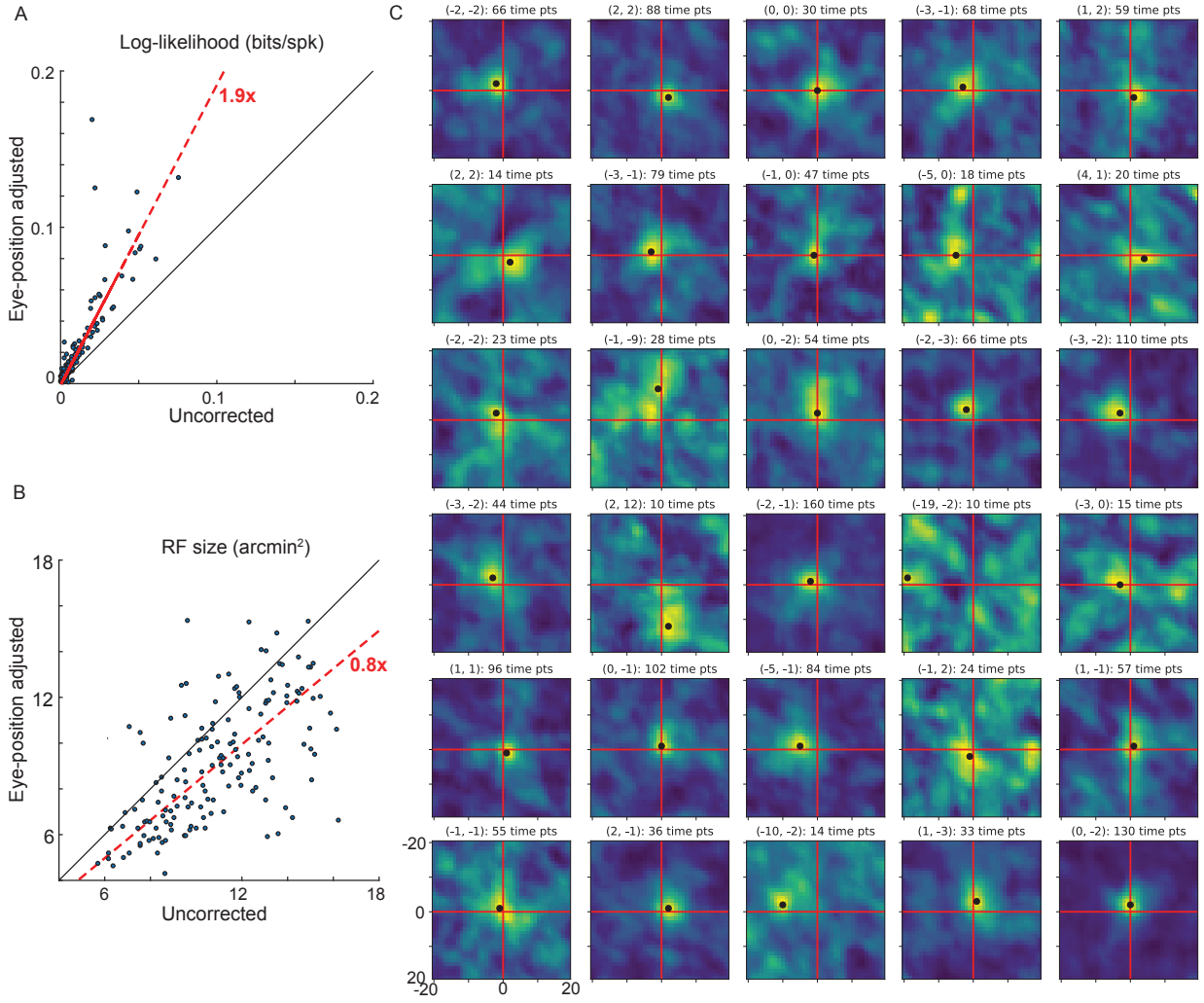

Figure 4.7: Algorithmic eye-position inference.

A: Effect of eye correction across the population. RF sizes shrink an average of 20% following the procedure. **B**: Effect of eye correction on model fits. The procedure has a massive effect on model performance, improving cross-validated likelihood scores almost twofold. **C**: Additional examples of 2-dimensional likelihood surfaces for various fixations.

5 Discussion

5.1 Summary

I have presented three studies that advance our understanding of V1 by applying sophisticated statistical and machine learning techniques to electrophysiological data, constructing computational models that illuminate the intricate computations performed in increasingly complex stimulus contexts within this critical region of the visual hierarchy.

In Chapter 2, I described the model-based characterization of V1 receptive fields by analyzing responses to 1-dimensional luminance stimuli. Statistical models allowed me to identify high-resolution components of these receptive fields, providing new insights into the computations conducted by V1 neurons. Particularly, I demonstrated how V1 neurons can simultaneously represent high-resolution information and respond to full-field inputs through the integration of multiple overlapping inputs, an insight not attainable using classical stimuli. Moreover, I introduced model-based measures of feature selectivity, revealing their potential to surpass classical measures in explaining the actual computations carried out by V1.

Chapter 3 builds upon this foundation by expanding the same nonlinear modeling techniques to explore how V1 neurons combine binocular inputs when presented with 1-D luminance stimuli over multiple binocular disparities. By fitting single-unit and shared nonlinear models, I identified the computations underpinning disparity tuning and elucidated how spatial convolutions enhance selectivity to disparity. The interactions between excitatory and inhibitory signals further refine disparity selectivity without compromising selectivity to specific spatial

patterns, providing an efficient implementation of classical hypotheses regarding disparity selectivity in V1.

In Chapter 4, I analyzed novel data recorded during the presentation of 2-D color stimuli at the fovea, again extending the modeling approaches to identify the computations by which V1 neurons represent chromatic inputs. This study generated the first high-resolution spatiochromatic measurements of V1 processing at the center of gaze, revealing the presence of diverse receptive fields within the same spatial regions. Characterizing the computations employed by V1 color-sensitive neurons to combine cone inputs for the representation of spatial and chromatic stimuli revealed that luminance inputs are represented at a much finer scale than color processes. I also illustrated how the subregion structure of receptive fields allows foveal cells with large receptive fields to maintain high acuity, and that these computations are not unique, as some cells simultaneously represent color and spatial opponency.

In summary, this dissertation offers an extensive exploration of V1 computations and serves as a significant contribution to our understanding of the intricacies of the primary visual cortex.

5.2 Using model subunits to comprehend neural computations

As mentioned in the introduction, subunit models have been a significant tool to better understand neural computations, particularly in V1. It is difficult to overstate how important the application of subunit models to neural data was to the projects presented here. In chapter 2, I describe how multiple, overlapping subunits can allow cells to represent high-acuity inputs

alongside responses to the RF envelope, and that aligning these subunits revealed the mechanism for this joint tuning. When I first presented these results at SfN 2018, multiple senior vision scientists looked at my poster and said "Oh so *that's* what those cells are doing!" - because for decades, they had observed cells responding like simple cells to low-frequency gratings and like complex cells to high-frequency gratings, but lacked a clear, interpretable explanation or classification for them. But simply describing these cells in terms of overlapping subunits, each easily interpretable both as individual receptive fields and as an aggregate, provided a clear explanation for this behavior.

This anecdote highlights the strength of the subunit model approach - 1) they are data-driven and capture neural computations with relatively little bias, 2) they are easily interpretable because they map directly onto specific stimulus dimensions, and 3) they can be easily manipulated to probe just how they achieve their computations. This is in stark contrast with deep neural networks, which outperform these models in point 1) but are notoriously difficult to interpret or use to identify how they capture certain computations. Of course, in some ways, that is the point - if we could describe neural computations in simple terms we could understand, we wouldn't need complicated deep networks. However, I propose that the work presented here lays the foundation for an alternative - using neural data to construct computational subspaces and tools that remain interpretable, and adding complexity step-by-step, stimulus feature by stimulus feature, identifying the computations used to represent each addition, until models have achieved the complexity necessary to explain arbitrary stimuli. While this approach is undoubtedly painstaking, the possible benefits of an interpretable model of V1 would certainly be worth it.

5.3 Accounting for eye movements

Another fundamental component of the studies presented here is the importance of eye tracking. The data presented in chapters 2 and 3 included eye tracking using implanted eye coils in addition to the further eye position correction provided by the ET algorithm, while the study presented in chapter 4 used camera-based eye trackers and would not have been possible without the ET algorithm achieving the resolution of foveal computations. The early figures in chapter 4 show just how blurry receptive field estimates without correction would be. Across all three chapters, the ET algorithm significantly improved results far beyond what would have been possible using hardware-based tracking alone. However, as beneficial as correcting for eye position has been in the studies presented, this is by no means the best solution.

Rather than simply "correcting" for eye position, it is important to consider vision as an active process, and eye movements an integral part of the visual process. I briefly reviewed the concepts of saccadic modulation in V1 in chapter 1, but barely scratched the surface - entire careers have been built on researching the interaction of eye movements and visual processing. It is a cyclic process: eye movements change the inputs into the visual system, which then uses the new information to determine the next eye movements. It is self-evident that saccades are the most significant tool of overt attention as we move our eyes to fixate on objects that we want more information on, but more recent work provides strong evidence that even fixational eye movements are more than just oculomotor noise. Indeed, they seem to actively increase the information our visual system has access to by moving receptive fields over particularly fine visual features to maximize information (Rucci et al., 2007; Rucci and Poletti, 2015). As such, these targeted

fixational eye movements may underlie more recent findings on hyperacuity (Poletti et al., 2013). Studying the modulation of visual processing in those contexts will necessitate not just models beyond those presented here, but also even more accurate eye tracking.

Luckily, a significant new technique, digital dual-Purkinje eye tracking, can provide that necessary resolution to eye position estimates, by capturing the reflections of a single-photon laser aimed through the cornea and lens of the eye (Yates et al., 2021). This technique is currently being used in ongoing work following up on the study presented in chapter 4, and will undoubtedly yield major breakthroughs in the future.

5.4 Assessing laminar differences

These studies have the ability to make comparisons across layers, thus revealing more information about organizational principles of V1. However, this is predicated on models that can identify and categorize computations reliably, and those computations really being localized to particular layers. Indeed, the same computations could be implemented in multiple layers with different physiological mechanisms, in which case a model categorizing the two different mechanisms as different computations would obscure the result. This is one of the reasons that the concern of columnar organization in cortex has seen some criticism recently: many different organizational principles are being described as columns in the literature, but they do not necessarily correspond to each other and may not actually relate to the same fundamental principles of V1 structural organization (Haueis, 2021).

As a result, while the process of identifying organizational principles of cortex using functional measures is a complex endeavor, it is also appealing – if we do find specific computations that are reliably localized, it could yield immense insight into the inner workings of V1. Unfortunately, I am not able to report such a true breakthrough here, instead only finding some examples of computations more common in certain layers than others. Chapter 2 is an example of this in the description of cell's exhibiting more complexity in extragranular layers. Chapters 3 and 4 also recorded with laminar probes and did not reveal particular patterns across layer in the analyses I applied thus far.

5.5 Implications for understanding the role of single neurons

The results presented here also speak to a much broader question in neuroscience: what exactly is the role of individual neurons to perception? The question has been discussed for decades (Barlow, 1985; Lennie, 1998) and continues to inform much of the study of V1, but the discussion surrounding it has changed since the advent of large-scale neurophysiological recordings and corresponding population analyses. Many studies interpreted recordings of populations in V1 as doing processing that allows higher-order areas to decode the visual scene (Victor et al., 1994; Benucci et al., 2009). In this interpretation, the activity of a single neuron is largely meaningless, and can only be understood in the context of its population. In contrast, other studies maintained the interpretation by Hubel & Wiesel that V1 neurons individually serve as feature detectors (Hubel and Wiesel, 1968), meaning a single neuron carries meaningful information about the presence of a particular, localized visual feature. This approach would allow

V1 to represent sensory information using a small number of active neurons, in other words, a sparse representation (Olshausen and Field, 1997, 2004).

Within this dichotomy, the studies presented in this dissertation do strongly support the idea that the activity of single neurons is interpretable and meaningful. Although a given neuron may be selective to a range of features, such as orientation, motion, and binocular disparity, the methods presented allow us to read out its selectivity and characterize it both in terms of representing individual stimulus dimensions (for example, a specific orientation) as subunits and representing their aggregate. Given that we can read out selectivity to high-dimensional features, it's likely the brain can do the same. Of course, we find that many neurons represent similar features at the same retinal locations, especially in the fovea; this means that despite a sparse representation, there is still redundancy in the system, and the loss of a given neuron would not lead to a loss of the ability to detect a given feature. Within the scope of small populations, the findings presented here are thus also consistent with a population coding framework that decodes specific features from an overcomplete set of neurons that detect individual features (Olshausen and Field, 1997).

5.6 Towards a comprehensive model of V1

While the studies presented here provide a solid foundation for a comprehensive model of V1, covering basic spatiotemporal feature selectivity, binocular integration, and color processing, there is still much ground to cover. Getting closer to a true V1 model will require both additional model components that capture additional functions of V1 and further analysis of both the data

presented here, as well as additional studies with other experimental controls. By adding components one at a time, interpretability of the model components and computations performed can be maintained. There are of course reasons none of the proposed expansions have been added within this thesis: each comes with its own significant technical challenges, and each might eventually constitute a whole study or even thesis of its own, but the benefits of adding them to create a comprehensive, interpretable model of V1 are evident.

A first significant addition is to allow these models to assess spiking patterns at finer the time scales. The presented studies only considered the timescales inherent to the stimuli, so between 60 and 100 Hz, but the brain, and visual system in particular, has long been known to employ precise time scales to convey information (Butts et al., 2007, 2011). As such, using smaller bin sizes will significantly increase the biological plausibility of these models. However, this addition comes with technical challenges – moving spikes into smaller bins significantly increases the number of parameters that a model needs to represent, meaning that more data will be required to properly constrain the parameters during fitting. Additionally, observing spikes at fine temporal resolution also opens up models to be more vulnerable to the effects of noise in the data, potentially decreasing likelihoods and yielding worse models if proper architecture and regularization are not in place.

A second vital component concerns the formal inclusion of eye movements. I have described the impact of eye movements on neural activity in chapter 1; While eye movements were considered implicitly here via the eye position correction approaches taken, the modulation by saccades was not explicitly included as a model term here. However, other models have included

such terms with great success and thus explained one more source of neural variability (McFarland et al., 2015; Niemeyer and Paradiso, 2018). Since even just the presence or absence of eye movements is a tractable and low-dimensional signal, it will be worthwhile to include moving forward, especially when considering faster timescales as described above.

Feedback inputs represent a more complicated feature to add, but also a much more significant step forward. Many studies of the visual pathway have identified the vital role that top-down feedback plays in modulating visual processing (Briggs and Usrey, 2007; Muckli et al., 2013; Zhaoping, 2017). Of course, feedback is a broad term, that can encompass many concepts including top-down modulation such as attention, or lateral connectivity with temporal delays that facilitates processes such as surround modulation (Smith and Muckli, 2010). Top-down processes are more difficult to incorporate into a model because to do this comprehensively would require a full model not just of V1, but the entire visual hierarchy. However, there are of course plenty of implementations of more specific, and thus tractable, top-down processes. One is implementing a separate term for attention mediated by gain control – giving models a shared gain term to modulate predicted activity is a reductive, but potentially powerful way to improve upon these models without becoming too abstract, especially since divisive gain control has long been the leading theories for how to best describe attention modulation in the visual system (Carandini et al., 1997b). A second possibility is to simply allow for possible feedback signals in the model architecture and infer feedback signals from the observed neural variability not inferred from the stimuli, effectively treat feedback as either a latent variable (Stringer et al., 2019). Finally, while more lateral interactions, such as modulation by extra-classical surrounds, have been described in

many studies of V1 (WEBB et al., 2002; Keller et al., 2020), few have offered model architectures capturing them. The computation of lateral connectivity may seem intuitively easier to approach, it still comes with its own challenges in model architecture – but this area may be a domain where even a multi-layer feedforward network can explain the bulk of the computations, meaning that the real problem lies in the interpretability across multiple layers of nonlinearity, not our ability to fit them.

5.7 Broader applications of V1 models

In summary, while significant progress has been made towards a comprehensive model of V1, including through the work presented in this dissertation, much remains to be done. However, the benefits of such a model make the effort more than worth the effort.

An immediate benefit to basic science consists of the possibility of using a V1 model as a front end to identify processing in higher-order visual regions such as V2 and V3 which remain relatively poorly understood. By representing the actual outputs of V1 in an interpretable way and then mapping higher-order activity not to stimuli directly, but to the expected outputs of V1 feature selective neurons, we could significantly improve our understanding of intermediate visual areas and get closer to identifying representations of visual signals throughout the visual hierarchy.

Having a well-defined, data-based model of V1 affords opportunities for performing *in silico* experiments; in other words, testing how the model performs under other circumstances such as novel stimuli, or manipulating it in was that correspond to physiological impairment. Given the amount of work necessary to conduct electrophysiological studies on primates, having the

opportunity to instead query a model would not only provide direly needed theoretical predictions for neural behavior, but reduce the reliance on animal testing altogether.

Most broadly, revealing the computations performed by V1 has potentially massive clinical applications, as most forms of vision loss spare visual cortex – thus, by identifying the format in which visual signals arrive in cortex, clinical interventions could restore sight by feeding in those signals via engineering solutions.

Bibliography

Adams DL, Horton JC (2003) A Precise Retinotopic Map of Primate Striate Cortex Generated from the Representation of Angioscotomas. J Neurosci 23:3771–3789 Available at: https://www.jneurosci.org/content/23/9/3771 [Accessed October 27, 2023].

Adams DL, Horton JC (2006) Monocular Cells Without Ocular Dominance Columns. J Neurophysiol 96 Available at: http://jn.physiology.org/content/96/5/2253.short [Accessed May 25, 2017].

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2:284 Available at: https://www.osapublishing.org/abstract.cfm?URI=josaa-2-2-284 [Accessed August 28, 2019].

Almasi A, Meffin H, Cloherty SL, Wong Y, Yunzab M, Ibbotson MR (2020) Mechanisms of feature selectivity and invariance in primary visual cortex. Cereb Cortex 30:5067–5087 Available at: https://doi.org/10.1101/2020.02.08.940270 [Accessed December 16, 2020].

Alonso J-M, Martinez LM (1998) Functional connectivity between simple cells and complex cells in cat striate cortex. Nat Neurosci 1:395–403 Available at: http://www.nature.com/articles/nn0998_395 [Accessed September 11, 2018].

Anzai A, Ohzawa I, Freeman RD (1999) Neural Mechanisms for Processing Binocular Information I. Simple Cells. J Neurophysiol 82:891–908 Available at: http://www.physiology.org/doi/10.1152/jn.1999.82.2.891 [Accessed January 2, 2018].

Atencio CA, Sharpee TO, Schreiner CE (2012) Receptive field dimensionality increases from the auditory midbrain to cortex. J Neurophysiol 107:2594–2603 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22323634 [Accessed September 4, 2018].

Barlow HB (1985) The twelfth Bartlett memorial lecture: The role of single neurons in the psychology of perception. Q J Exp Psychol 37:121–145 Available at: https://www.tandfonline.com/doi/abs/10.1080/14640748508400927 [Accessed January 18, 2024].

Bartsch F, Cumming BG, Butts DA (2022) Model-based characterization of the selectivity of neurons in primary visual cortex. J Neurophysiol 128:350–363 Available at: https://journals.physiology.org/doi/10.1152/jn.00416.2021 [Accessed September 12, 2023].

Basole A, White LE, Fitzpatrick D (2003) Mapping multiple features in the population response of visual cortex. Nat 2003 4236943 423:986–990 Available at: https://www.nature.com/articles/nature01721 [Accessed November 13, 2023].

Benucci A, Ringach DL, Carandini M (2009) Coding of stimulus sequences by population responses in visual cortex. Nat Neurosci 2009 1210 12:1317–1324 Available at: https://www.nature.com/articles/nn.2398 [Accessed January 18, 2024].

Blasdel GG, Salama G (1986) Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. Nat 1986 3216070 321:579–585 Available at: https://www.nature.com/articles/321579a0 [Accessed December 15, 2023].

Bosking WH, Crowley JC, Fitzpatrick D (2002) Spatial coding of position and orientation in primary visual cortex. Nat Neurosci 5:874–882 Available at: https://www.nature.com/articles/nn908 [Accessed November 1, 2023].

Bosking WH, Zhang Y, Schofield B, Fitzpatrick D (1997) Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex. J Neurosci 17:2112–2127 Available at: https://www.jneurosci.org/content/17/6/2112 [Accessed October 27, 2023].

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436 Available at: http://color.psych.ucsb.edu/psychtoolbox [Accessed January 18, 2024].

Briggs F, Usrey WM (2007) A Fast, Reciprocal Pathway between the Lateral Geniculate Nucleus and Visual Cortex in the Macaque Monkey. J Neurosci 27:5431 Available at: /pmc/articles/PMC2888515/ [Accessed October 30, 2023].

Bringmann A, Syrbe S, Görner K, Kacza J, Francke M, Wiedemann P, Reichenbach A (2018) The primate fovea: Structure, function and development. Prog Retin Eye Res 66:49–84.

Bumsted K, Hendrickson A (1999) Distribution and Development of Short-Wavelength Cones Differ Between Macaca Monkey and Human Fovea. J Comp Neurol 403:502–516 Available at: https://onlinelibrary.wiley.com/doi/10.1002/ [Accessed January 18, 2024].

Butts DA (2019) Data-Driven Approaches to Understanding Visual Neuron Activity. Annu Rev Vis Sci 5:451–477.

Butts DA, Cui Y, Casti ARR (2016) Nonlinear computations shaping temporal processing of precortical vision. J Neurophysiol 116:1344–1357 Available at: https://journals.physiology.org/doi/10.1152/jn.00878.2015 [Accessed December 21, 2023].

Butts DA, Weng C, Jin J, Alonso JM, Paninski L (2011) Temporal Precision in the Visual Pathway through the Interplay of Excitation and Stimulus-Driven Suppression. J Neurosci

31:11313–11327 Available at: https://www.jneurosci.org/content/31/31/11313 [Accessed October 31, 2023].

Butts DA, Weng C, Jin J, Yeh CI, Lesica NA, Alonso JM, Stanley GB (2007) Temporal precision in the neural code and the timescales of natural vision. Nat 2007 4497158 449:92–95 Available at: https://www.nature.com/articles/nature06105 [Accessed November 8, 2023].

Cadena SA, Denfield GH, Walker EY, Gatys LA, Tolias AS, Bethge M, Ecker AS (2019) Deep convolutional models improve predictions of macaque V1 responses to natural images. PLoS Comput Biol 15.

Callaway EM (1998) Local circuits in primary visual cortex of the macaque monkey. Annu Rev Neurosci 21:47–74 Available at: https://www.annualreviews.org/doi/abs/10.1146/annurev.neuro.21.1.47 [Accessed October 30, 2023].

Callaway EM, Callaway EM (2005) Structure and function of parallel pathways in the primate early visual system. J Physiol 566:13–19 Available at: https://onlinelibrary.wiley.com/doi/full/10.1113/jphysiol.2005.088047 [Accessed October 30, 2023].

Carandini M (2006) What simple and complex cells compute. J Physiol 577:463–466 Available at: /pmc/articles/PMC1890437/ [Accessed October 27, 2023].

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC (2005) Do we know what the early visual system does? In: Journal of Neuroscience, pp 10577–

10597. Society for Neuroscience. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16291931 [Accessed August 30, 2018].

Carandini M, Heeger DJ, Movshon JA (1997a) Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci 17:8621–8644 Available at: http://www.ncbi.nlm.nih.gov/pubmed/9334433 [Accessed September 5, 2018].

Carandini M, Heeger DJ, Movshon JA (1997b) Linearity and Normalization in Simple Cells of the Macaque Primary Visual Cortex. J Neurosci 17:8621–8644 Available at: https://www.jneurosci.org/content/17/21/8621 [Accessed October 27, 2023].

Chen Y, Anand S, Martinez-Conde S, Macknik SL, Bereshpolova Y, Swadlow HA, Alonso JM (2009) The linearity and selectivity of neuronal responses in awake visual cortex. J Vis 9:12–12 Available at: http://jov.arvojournals.org/Article.aspx?doi=10.1167/9.9.12 [Accessed September 4, 2018].

Chen Y, Dhamala M, Bollimunta A, Schroeder CE, Ding M (2011) Current source density analysis of ongoing neural activity: Theory and application. Neuromethods 54:27–40.

Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Netw Comput Neural Syst 12:199–213 Available at: http://www.tandfonline.com/doi/full/10.1080/net.12.2.199.213 [Accessed September 3, 2018].

Coen-Cagli R, Dayan P, Schwartz O (2012) Cortical surround interactions and perceptual salience via natural scene statistics Sporns O, ed. PLoS Comput Biol 8:e1002405 Available at: http://dx.plos.org/10.1371/journal.pcbi.1002405 [Accessed February 15, 2017].

Conway BR (2001) Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (v-1). J Neurosci 21:2768–2783 Available at: https://www.jneurosci.org/content/21/8/2768 [Accessed February 12, 2021].

Cooper Roddey J, Girish B, Miller JP (2000) Assessing the performance of neural encoding models in the presence of noise. J Comput Neurosci 8:95–112 Available at: https://pubmed.ncbi.nlm.nih.gov/10798596/ [Accessed January 19, 2024].

Cumming BG, DeAngelis GC (2001) The physiology of stereopsis. Annu Rev Neurosci 24:203–238 Available at: https://pubmed.ncbi.nlm.nih.gov/11283310/ [Accessed October 27, 2023].

Cumming BG, Parker AJ (1997) Responses of primary visual cortical neurons to binocular disparity without depth perception. Nature 389:280–283 Available at: http://www.nature.com/doifinder/10.1038/38487 [Accessed February 15, 2017].

Cumming BG, Shapiro SE, Parker AJ (1998) Disparity detection in anticorrelated stereograms. Perception 27:1367–1377.

Das A, Gilbert CD (1997) Distortions of visuotopic map match orientation singularities in primary visual cortex. Nature 387:594–598 Available at: https://pubmed.ncbi.nlm.nih.gov/9177346/ [Accessed October 27, 2023].

David S V., Gallant JL (2005) Predicting neuronal responses during natural vision. Netw Comput Neural Syst 16:239–260 Available at: http://www.tandfonline.com/doi/full/10.1080/09548980500464030 [Accessed January 18, 2017].

David S V., Vinje WE, Gallant JL (2004) Natural Stimulus Statistics Alter the Receptive Field Structure of V1 Neurons. J Neurosci 24:6991–7006 Available at: https://www.jneurosci.org/content/24/31/6991 [Accessed November 1, 2023].

De A, Horwitz GD (2021) Spatial receptive field structure of double-opponent cells in macaque V1. J Neurophysiol 125:843–857 Available at: https://journals.physiology.org/doi/10.1152/jn.00547.2020 [Accessed December 15, 2022].

De Ruyter Van Steveninck R, Bialek W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. Proc R Soc London Ser B Biol Sci 234:379–414 Available at: https://royalsocietypublishing.org/doi/10.1098/rspb.1988.0055 [Accessed October 31, 2023].

De Valois RL, Albrecht DG, Thorell LG (1982) Spatial frequency selectivity of cells in macaque visual cortex. Vision Res 22:545–559 Available at: https://pubmed.ncbi.nlm.nih.gov/7112954/ [Accessed January 5, 2021].

De Valois RL, Cottaris NP, Mahon LE, Elfar SD, Wilson JA (2000) Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. Vision Res 40:3685–3702.

DeAngelis GC, Freeman RD, Ohzawa I (1994) Length and width tuning of neurons in the cat's primary visual cortex. J Neurophysiol 71:347–374 Available at: http://www.ncbi.nlm.nih.gov/pubmed/8158236 [Accessed September 19, 2018].

Derrington AM, Krauskopf J, Lennie P (1984) Chromatic mechanisms in lateral geniculate nucleus of macaque. J Physiol 357:241–265 Available at: http://doi.wiley.com/10.1113/jphysiol.1984.sp015499 [Accessed February 11, 2021].

Doi T, Takano M, Fujita I (2013) Temporal channels and disparity representations in stereoscopic depth perception. J Vis 13.

Douglas RJ, Martin KAC (2004) Neuronal circuits of the neocortex. Annu Rev Neurosci 27:419–451 Available at: www.annualreviews.org [Accessed January 5, 2021].

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

Fournier J, Monier C, Levy M, Marre O, Sári K, Kisvárday ZF, Frégnac Y (2014) Hidden complexity of synaptic receptive fields in cat V1. J Neurosci 34:5515–5528 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24741042 [Accessed October 5, 2018].

Garg AK, Li P, Rashid MS, Callaway EM (2019) Color and orientation are jointly coded and spatially organized in primate primary visual cortex. Science (80-) 364:1275–1279 Available at: https://science.sciencemag.org/content/364/6447/1275 [Accessed June 30, 2020].

Gegenfurtner KR, Kiper DC (2003) Color vision. Annu Rev Neurosci 26:181–206 Available at: www.annualreviews.org [Accessed November 1, 2023].

Gharat A, Baker CL (2017) Nonlinear Y-Like Receptive Fields in the Early Visual Cortex: An Intermediate Stage for Building Cue-Invariant Receptive Fields from Subcortical Y Cells. J Neurosci 37:998–1013 Available at: http://www.ncbi.nlm.nih.gov/pubmed/28123031 [Accessed September 6, 2018].

Goncalves NR, Welchman AE (2017) “What Not” Detectors Help the Brain See in Depth. Curr Biol 27:1403-1412.e8.

Goris RLT, Simoncelli EP, Movshon JA (2015) Origin and Function of Tuning Diversity in Macaque Visual Cortex. Neuron 88:819–831 Available at: https://www.sciencedirect.com/science/article/pii/S0896627315008752 [Accessed September 4, 2018].

Gross CG (1998) Brain, vision, memory: Tales in the history of neuroscience. Available at: https://mitpress.mit.edu/9780262571357/brain-vision-memory/ [Accessed July 20, 2023].

Haider B, Häusser M, Carandini M (2012) Inhibition dominates sensory responses in the awake cortex. Nature 493:97–100 Available at: http://www.nature.com/doifinder/10.1038/nature11665 [Accessed September 21, 2018].

Haueis P (2021) The death of the cortical column? Patchwork structure and conceptual retirement in neuroscientific practice. Stud Hist Philos Sci 85:101–113.

Hendry SHC, Clay Reid R (2000) The koniocellular pathway in primate vision. Annu Rev Neurosci 23:127–153 Available at: https://www.annualreviews.org/doi/abs/10.1146/annurev.neuro.23.1.127 [Accessed November 1, 2023].

Henriksen S, Butts DA, Read JCA, Cumming BG (2018) Current models cannot account for V1’s specialisation for binocular natural image statistics. bioRxiv:497008 Available at: https://www.biorxiv.org/content/10.1101/497008v1 [Accessed November 13, 2023].

Henriksen S, Read JCA (2016) Visual Perception: A Novel Difference Channel in Binocular Vision. Curr Biol 26:R500–R503 Available at: http://www.sciencedirect.com/science/article/pii/S0960982216304031 [Accessed May 25, 2017].

Henriksen S, Tanabe S, Cumming B (2016) Disparity processing in primary visual cortex. Philos Trans R Soc B Biol Sci 371 Available at: http://rstb.royalsocietypublishing.org/content/371/1697/20150255 [Accessed May 25, 2017].

Horton JG, Adams DL (2005) The cortical column: A structure without a function. Philos Trans R Soc B Biol Sci 360:837–862 Available at: https://pubmed.ncbi.nlm.nih.gov/15937015/ [Accessed October 27, 2023].

Horwitz GD, Hass CA (2012) Nonlinear analysis of macaque V1 color tuning reveals cardinal directions for cortical color processing. Nat Neurosci 15:913–919 Available at: http://www.nature.com/articles/nn.3105 [Accessed June 20, 2018].

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. J Physiol 148:574–591 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1363130/ [Accessed October 27, 2023].

Hubel DH, Wiesel TN (1963) Shape and arrangement of columns in cat's striate cortex. J Physiol 165:559 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1359325/ [Accessed October 27, 2023].

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195:215–243 Available at: https://pubmed.ncbi.nlm.nih.gov/4966457/ [Accessed October 27, 2023].

Hubel DH, Wiesel TN (1969) Anatomical demonstration of columns in the monkey striate cortex. Nature 221:747–750 Available at: https://www.nature.com/articles/221747a0 [Accessed February 11, 2021].

Hubel DH, Wiesel TN (1972) Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey. J Comp Neurol 146:421–450 Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/cne.901460402 [Accessed October 27, 2023].

Hubel DH, Wiesel TN (1974) Sequence regularity and geometry of orientation columns in the monkey striate cortex. J Comp Neurol 158:267–293 Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/cne.901580304 [Accessed December 15, 2023].

Hubel DH, Wiesel TN, Stryker MP (1978) Anatomical demonstration of orientation columns in macaque monkey. J Comp Neurol 177:361–379 Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/cne.901770302 [Accessed October 27, 2023].

Hubel DN, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154 Available at: http://www.ncbi.nlm.nih.gov/pubmed/14449617 [Accessed September 11, 2018].

Ibbotson M, Krekelberg B (2011) Visual perception and saccadic eye movements. Curr Opin Neurobiol 21:553–558.

Issa NP, Trepel C, Stryker MP (2000) Spatial Frequency Maps in Cat Visual Cortex. J Neurosci 20:8504–8514 Available at: https://www.jneurosci.org/content/20/22/8504 [Accessed October 28, 2023].

Jin J, Wang Y, Swadlow HA, Alonso JM (2011) Population receptive fields of on and off thalamic inputs to an orientation column in visual cortex. In: Nature Neuroscience, pp 232–240. Nature Publishing Group. Available at: https://www.nature.com/articles/nn.2729 [Accessed December 19, 2023].

Jin JZ, Weng C, Yeh CI, Gordon JA, Ruthazer ES, Stryker MP, Swadlow HA, Alonso JM (2008) ON AND OFF DOMAINS OF GENICULATE AFFERENTS IN CAT PRIMARY VISUAL CORTEX. Nat Neurosci 11:88 Available at: /pmc/articles/PMC2556869/ [Accessed December 19, 2023].

Johnson EN, Hawken MJ, Shapley R (2008) The orientation selectivity of color-responsive neurons in macaque V1. J Neurosci 28:8096–8106.

Kagan I, Gur M, Snodderly DM (2002) Spatial Organization of Receptive Fields of V1 Neurons of Alert Monkeys: Comparison With Responses to Gratings. J Neurophysiol 88:2557–2574 Available at: http://www.physiology.org/doi/10.1152/jn.00858.2001 [Accessed September 4, 2018].

Kagan I, Gur M, Snodderly DM (2008) Saccades and drifts differentially modulate neuronal activity in V1: Effects of retinal image motion, position, and extraretinal influences. J Vis 8:19–19 Available at: http://journalofvision. [Accessed November 1, 2023].

Keller AJ, Roth MM, Scanziani M (2020) Feedback generates a second receptive field in neurons of the visual cortex. Nature 582:545–549.

Kindel WF, Christensen ED, Zylberberg J (2017) Using deep learning to reveal the neural code for images in primary visual cortex. J Vis 19:1–12 Available at: https://arxiv.org/abs/1706.06208v1 [Accessed August 27, 2021].

Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C, Cornelissen F (2007) What's new in Psychtoolbox-3? Available at: https://pure.mpg.de/rest/items/item_1790332/component/file_3136265/content [Accessed January 18, 2024].

Klindt DA, Ecker AS, Euler T, Bethge M (2017) Neural system identification for large populations separating "what" and "where." Adv Neural Inf Process Syst 2017-December:3507–3517 Available at: https://arxiv.org/abs/1711.02653v2 [Accessed August 27, 2021].

Kremkow J, Jin J, Wang Y, Alonso JM (2016) PRINCIPLES UNDERLYING SENSORY MAP TOPOGRAPHY IN PRIMARY VISUAL CORTEX. Nature 533:52 Available at: /pmc/articles/PMC4860131/ [Accessed December 19, 2023].

Kriegeskorte N (2015) Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. Annu Rev Vis Sci 1:417–446 Available at: http://www.annualreviews.org/doi/10.1146/annurev-vision-082114-035447 [Accessed December 23, 2020].

Krubitzer L (2007) The magnificent compromise: cortical field evolution in mammals. Neuron 56:201–208 Available at: http://www.cell.com/article/S089662730700757X/fulltext [Accessed July 23, 2023].

Lafer-Sousa R, Conway BR (2013) Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. Nat Neurosci 16:1870–1878 Available at: /pmc/articles/PMC3957328/ [Accessed September 29, 2023].

Lennie P (1998) Single Units and Visual Cortical Organization. http://dx.doi.org/101068/p270889 27:889–935 Available at: https://journals.sagepub.com/doi/abs/10.1068/p270889?casa_token=E5Rwn4o2hCQAAAAA%3AKSUdVsC2x9mf43ShXh4-9iVzEfVdIAJriLKoQw7nKn-XvXOuEcNjX5wkWlJc5DEahHf8lgvdkqk81w [Accessed January 18, 2024].

Lennie P, Krauskopf J, Sclar G (1990) Chromatic mechanisms in striate cortex of macaque. J Neurosci 10:649–669 Available at: https://www.jneurosci.org/content/10/2/649 [Accessed February 11, 2021].

Leopold DA (2012) Primary visual cortex: Awareness and blindsight. Annu Rev Neurosci 35:91–109.

LeVay S, Connolly M, Houde J, Van Essen D (1985) The complete pattern of ocular dominance stripes in the striate cortex and visual field of the macaque monkey. J Neurosci 5 Available at: http://www.jneurosci.org/content/5/2/486.short [Accessed May 25, 2017].

Li G, Talebi V, Yoonessi A, Baker CL (2010) A FPGA real-time model of single and multiple visual cortex neurons. J Neurosci Methods 193:62–66 Available at: https://www.sciencedirect.com/science/article/pii/S0165027010004061 [Accessed December 7, 2018].

Li M, Ju N, Jiang R, Liu F, Jiang H, Macknik S, Martinez-Conde S, Tang S (2022) Perceptual hue, lightness, and chroma are represented in a multidimensional functional anatomical map in macaque V1. Prog Neurobiol 212:102251.

Li X, Chen Y, Lashgari R, Bereshpolova Y, Swadlow HA, Lee BB, Alonso JM (2015) Mixing of Chromatic and Luminance Retinal Signals in Primate Area V1. Cereb Cortex 25:1920–1937 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24464943 [Accessed June 20, 2019].

Livingstone MS, Hubel DH (1982) Thalamic inputs to cytochrome oxidase-rich regions in monkey visual cortex. Proc Natl Acad Sci 79:6098–6101 Available at: https://www.pnas.org/doi/abs/10.1073/pnas.79.19.6098 [Accessed November 1, 2023].

Lochmann T, Blanche TJ, Butts DA (2013) Construction of Direction Selectivity through Local Energy Computations in Primary Visual Cortex Zhang LI, ed. PLoS One 8:e58666 Available at: http://dx.plos.org/10.1371/journal.pone.0058666 [Accessed October 3, 2018].

Lund JS, Angelucci A, Bressloff CP (2003) Anatomical substrates for functional columns in macaque monkey primary visual cortex. In: Cerebral Cortex, pp 15–24. Oxford University Press. Available at: https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/13.1.15 [Accessed June 8, 2018].

Lurz K-K, Bashiri M, Willeke K, Jagadish AK, Wang E, Walker EY, Cadena SA, Muhammad T, Cobos E, Tolias AS, Ecker AS, Sinz FH (2021) Generalization in data-driven models of primary visual cortex. bioRxiv:2020.10.05.326256 Available at: https://www.biorxiv.org/content/10.1101/2020.10.05.326256v2 [Accessed January 24, 2022].

Mante V, Carandini M (2005) Mapping of stimulus energy in primary visual cortex. J Neurophysiol 94:788–798 Available at: https://pubmed.ncbi.nlm.nih.gov/15758051/ [Accessed October 28, 2023].

Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. Vis a Comput Investig into Hum Represent Process Vis information Available at: https://www.scirp.org/(S(351jmbntvnsjt1aadkozje))/reference/ReferencesPapers.aspx?ReferenceID=816915 [Accessed October 28, 2023].

Martin PR, Blessing EM, Buzás P, Szmajda BA, Forte JD (2011) Transmission of colour and acuity signals by parvocellular cells in marmoset monkeys. J Physiol 589:2795–2812 Available at: http://doi.wiley.com/10.1113/jphysiol.2010.194076 [Accessed July 21, 2023].

Martinez-Conde S, Macknik SL (2008) Fixational eye movements across vertebrates: Comparative dynamics, physiology, and perception. J Vis 8:28–29 Available at: http://journalofvision.org/8/14/28/, [Accessed January 5, 2021].

Martinez-Conde S, Macknik SL, Hubel DH (2000) Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. Nat Neurosci 2000 33 3:251–258 Available at: https://www.nature.com/articles/nn0300_251 [Accessed October 31, 2023].

Martinez-Conde S, Macknik SL, Troncoso XG, Dyar TA (2006) Microsaccades counteract visual fading during fixation. Neuron 49:297–305 Available at: https://pubmed.ncbi.nlm.nih.gov/16423702/ [Accessed October 31, 2023].

Martinez-Conde S, Macknik SL, Troncoso XG, Hubel DH (2009) Microsaccades: a neurophysiological analysis. Trends Neurosci 32:463–475 Available at: http://www.cell.com/article/S0166223609001076/fulltext [Accessed October 31, 2023].

Maunsell JHR, Newsome WT (1987) Visual Processing in Monkey Extrastriate Cortex. Annu Rev Neurosci 10:363–401 Available at: http://www.annualreviews.org/doi/10.1146/annurev.ne.10.030187.002051 [Accessed September 11, 2018].

Mazer JA, Vinje WE, McDermott J, Schiller PH, Gallant JL (2002) Spatial frequency and orientation tuning dynamics in area V1. Proc Natl Acad Sci U S A 99:1645–1650 Available at: http://www.ncbi.nlm.nih.gov/pubmed/11818532 [Accessed September 3, 2018].

McFarland JM, Bondy AG, Cumming BG, Butts DA (2014) High-resolution eye tracking using V1 neuron activity. Nat Commun 5:4605 Available at: http://www.nature.com/articles/ncomms5605 [Accessed September 4, 2018].

McFarland JM, Bondy AG, Saunders RC, Cumming BG, Butts DA (2015) Saccadic modulation of stimulus processing in primary visual cortex. Nat Commun 6:8110 Available at: http://www.nature.com/articles/ncomms9110 [Accessed September 4, 2018].

McFarland JM, Cui Y, Butts DA (2013) Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs. PLoS Comput Biol 9:e1003143 Available at: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003143 [Accessed October 31, 2023].

McFarland JM, Cumming BG, Butts DA (2016) Variability and Correlations in Primary Visual Cortical Neurons Driven by Fixational Eye Movements. J Neurosci 36:6225–6241 Available at: https://www.jneurosci.org/content/36/23/6225 [Accessed October 31, 2023].

Mechler F, Ringach DL (2002) On the classification of simple and complex cells. Vision Res 42:1017–1033.

Menz MD, Freeman RD (2004) Functional Connectivity of Disparity-Tuned Neurons in the Visual Cortex. J Neurophysiol 91:1794–1807 Available at: http://www.physiology.org/doi/pdf/10.1152/jn.00574.2003 [Accessed January 2, 2018].

Michaiel AM, Abe ETT, Niell CM (2020) Dynamics of gaze control during prey capture in freely moving mice. Elife 9:1–27.

Miller CT, Gire D, Hoke K, Huk AC, Kelley D, Leopold DA, Smear MC, Theunissen F, Yartsev M, Niell CM (2022) Natural behavior is the language of the brain. Curr Biol 32:R482–R493 Available at: http://www.cell.com/article/S0960982222004262/fulltext [Accessed June 1, 2022].

Moskovitz TH, Roy NA, Pillow JW (2018) A comparison of deep learning and linear-nonlinear cascade approaches to neural encoding. bioRxiv:463422 Available at: https://doi.org/10.1101/463422 [Accessed December 23, 2020].

Movshon JA, Thompson ID, Tolhurst DJ (1978a) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. J Physiol 283:53–77 Available at: https://physoc.onlinelibrary.wiley.com/doi/full/10.1113/jphysiol.1978.sp012488 [Accessed August 10, 2021].

Movshon JA, Thompson ID, Tolhurst DJ (1978b) Receptive field organization of complex cells in the cat's striate cortex. J Physiol 283:79–99 Available at: https://physoc.onlinelibrary.wiley.com/doi/full/10.1113/jphysiol.1978.sp012489 [Accessed August 10, 2021].

Movshon JA, Thompson ID, Tolhurst DJ (1978c) Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. J Physiol 283:101–120 Available at: http://www.ncbi.nlm.nih.gov/pubmed/722570 [Accessed September 11, 2018].

Muckli L, Petro LS, author C, LarsMuckli L (2013) Network interactions: non-geniculate input to V1. Curr Opin Neurobiol 23:195–201 Available at: http://dx.doi.org/10.1016/j.conb.2013.01.020 [Accessed January 2, 2018].

Najafian S, Koch E, Teh KL, Jin J, Rahimi-Nasrabadi H, Zaidi Q, Kremkow J, Alonso JM (2022) A theory of cortical map formation in the visual brain. Nat Commun 2022 131 13:1–20 Available at: https://www.nature.com/articles/s41467-022-29433-y [Accessed October 27, 2023].

Nassi JJ, Callaway EM (2009) Parallel Processing Strategies of the Primate Visual System. Nat Rev Neurosci 10:360 Available at: /pmc/articles/PMC2771435/ [Accessed October 27, 2023].

Niemeyer JE, Paradiso MA (2018) Saccade-based termination responses in macaque V1 and visual perception. Vis Neurosci 35:E025.

Ninomiya T, Sanada TM, Ohzawa I (2012) Contributions of excitation and suppression in shaping spatial frequency selectivity of V1 neurons as revealed by binocular measurements. J Neurophysiol 107:2220–2231 Available at: http://www.physiology.org/doi/10.1152/jn.00832.2010 [Accessed January 2, 2018].

Nowak LG, Munk MHJ, Girard P, Bullier J (1995) Visual latencies in areas V1 and V2 of the macaque monkey. Vis Neurosci 12:371–384 Available at: http://www.journals.cambridge.org/abstract_S095252380000804X [Accessed April 18, 2019].

Nykamp DQ, Ringach DL (2002) Full identification of a linear-nonlinear system via cross-correlation analysis. J Vis 2:1–1.

Obermayer K, Blasdel GG (1993) Geometry of orientation and ocular dominance columns in monkey striate cortex. J Neurosci 13:4114–4129 Available at: https://www.jneurosci.org/content/13/10/4114 [Accessed October 28, 2023].

Ohki K, Chung S, Ch'ng YH, Kara P, Reid RC (2005) Functional imaging with cellular resolution reveals precise microarchitecture in visual cortex. Nature 433:597–603 Available at: https://europepmc.org/article/MED/15660108 [Accessed October 28, 2023].

Ohki K, Chung S, Kara P, Hübener M, Bonhoeffer T, Reid RC (2006) Highly ordered arrangement of single neurons in orientation pinwheels. Nature 442:925–928 Available at: https://europepmc.org/article/med/16906137 [Accessed October 28, 2023].

Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. Science (80-) 249:1037–1041 Available at: https://science.sciencemag.org/content/249/4972/1037 [Accessed January 5, 2021].

Ohzawa I, DeAngelis GC, Freeman RD (1997) Encoding of binocular disparity by complex cells in the cat's visual cortex. J Neurophysiol 77:2879–2909 Available at: http://jn.physiology.org/content/jn/75/5/1779.full.pdf [Accessed March 10, 2017].

Ohzawa I, Freeman RD (1986) The Binocular Organization of Complex Cells in the Cat's Visual Cortex.

Olshausen BA, Field DJ (1996) Natural image statistics and efficient coding. Netw Comput Neural Syst 7:333 Available at: https://iopscience.iop.org/article/10.1088/0954-898X/7/2/014 [Accessed November 13, 2023].

Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision Res 37:3311–3325.

Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. Curr Opin Neurobiol 14:481–487.

Paninski L, Pillow JW, Simoncelli EP (2004) Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. Neural Comput 16:2533–2561 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15516273 [Accessed September 3, 2018].

Park IM, Archer E, Priebe N, Pillow JW (2013) Spectral methods for neural characterization using generalized quadratic models. Available at: http://papers.nips.cc/paper/4993-spectral-methods-for-neural-characterization-using-generalized-quadratic-models.pdf [Accessed October 3, 2018].

Park M, Pillow JW (2011) Receptive Field Inference with Localized Priors Sporns O, ed. PLoS Comput Biol 7:e1002219 Available at: http://dx.plos.org/10.1371/journal.pcbi.1002219 [Accessed September 14, 2018].

Parker AJ (2007) Binocular depth perception and the cerebral cortex. Nat Rev Neurosci 8:379–391 Available at: http://www.nature.com/doifinder/10.1038/nrn2131 [Accessed July 26, 2017].

Perry VH, Oehler R, Cowey A (1984) Retinal ganglion cells that project to the dorsal lateral geniculate nucleus in the macaque monkey. Neuroscience 12:1101–1123.

Pettersen KH, Devor A, Ulbert I, Dale AM, Einevoll GT (2006) Current-source density estimation based on inversion of electrostatic forward solution: Effects of finite extent of neuronal activity and conductivity discontinuities. J Neurosci Methods 154:116–133 Available at: https://s3.amazonaws.com/academia.edu.documents/45622607/pettersen-2006JNSM-iCSD.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1535688516&Signature=lVcjiElzinse7w1KFGv1bgQpjv8%3D&response-content-disposition=inline%3B filename%3DCurrent-source_density_e [Accessed August 30, 2018].

Pillow JW, Simoncelli EP (2006) Dimensionality reduction in neural models: An information-theoretic generalization of spike-triggered average and covariance analysis. J Vis 6:9–9.

Poletti M, Listorti C, Rucci M (2013) Microscopic Eye Movements Compensate for Nonhomogeneous Vision within the Fovea. Curr Biol 23:1691–1695.

Priebe NJ, Ferster D (2005) Direction Selectivity of Excitation and Inhibition in Simple Cells of the Cat Primary Visual Cortex. Neuron 45:133–145 Available at: https://www.sciencedirect.com/science/article/pii/S0896627304008402 [Accessed September 3, 2018].

Priebe NJ, Mechler F, Carandini M, Ferster D (2004) The contribution of spike threshold to the dichotomy of cortical simple and complex cells. Nat Neurosci 7:1113–1122 Available at: /pmc/articles/PMC2915829/ [Accessed August 18, 2021].

Prince SJD, Cumming BG, Parker AJ (2002a) Range and mechanism of encoding of horizontal disparity in macaque V1. J Neurophysiol 87:209–221.

Prince SJD, Pointon AD, Cumming BG, Parker AJ (2002b) Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. J Neurophysiol 87:191–208 Available at: http://www.physiology.org/doi/10.1152/jn.00465.2000 [Accessed January 2, 2018].

Purves D, Augustine GJ, Fitzpatrick D, Hall WC;, LaMantia A-S, Mooney RD;, Platt ML;, White LE (2017) Neuroscience, 6th ed. Sinauer Associates, New York. Available at: https://global.oup.com/ushe/product/neuroscience-9781605353807?q=9781605353807&lang=en&cc=us [Accessed December 15, 2023].

Rajkai C, Lakatos P, Chen CM, Pincze Z, Karmos G, Schroeder CE (2008) Transient Cortical Excitation at the Onset of Visual Fixation. Cereb Cortex 18:200–209 Available at: https://dx.doi.org/10.1093/cercor/bhm046 [Accessed October 31, 2023].

Read JCA, Cumming BG (2003a) Testing Quantitative Models of Binocular Disparity Selectivity in Primary Visual Cortex. J Neurophysiol 90:2795–2817 Available at: http://www.physiology.org/doi/10.1152/jn.01110.2002 [Accessed January 2, 2018].

Read JCA, Cumming BG (2003b) Measuring V1 Receptive Fields Despite Eye Movements in Awake Monkeys. J Neurophysiol 90:946–960 Available at: https://www.physiology.org/doi/10.1152/jn.01068.2002 [Accessed January 5, 2021].

Read JCA, Cumming BG (2007) Sensors for impossible stimuli may solve the stereo correspondence problem. Nat Neurosci 10:1322–1328 Available at: http://www.nature.com/doifinder/10.1038/nn1951 [Accessed February 15, 2017].

Read JCA, Parker AJ, Cumming BG (2002) A simple model accounts for the response of disparity-tuned V1 neurons to anticorrelated images. Vis Neurosci 19:735–753.

Reid RC, Victor JD, Shapley RM (1997) The use of m-sequences in the analysis of visual neurons: Linear receptive field properties. Vis Neurosci 14:1015–1027 Available at: https://www.cambridge.org/core/journals/visual-neuroscience/article/abs/use-of-msequences-in-the-analysis-of-visual-neurons-linear-receptive-field-properties/4B56A2FD7EDC67AA763CC3B017ABD516 [Accessed December 21, 2023].

Reppas JB, Usrey WM, Reid RC (2002) Saccadic eye movements modulate visual responses in the lateral geniculate nucleus. Neuron 35:961–974 Available at: http://www.cell.com/article/S0896627302008231/fulltext [Accessed October 31, 2023].

Richards BA et al. (2019) A deep learning framework for neuroscience. Nat Neurosci 22:1761 Available at: /pmc/articles/PMC7115933/ [Accessed January 25, 2022].

Ringach D, Shapley R (2004) Reverse correlation in neurophysiology. Cogn Sci 28:147–166 Available at: https://onlinelibrary.wiley.com/doi/full/10.1207/s15516709cog2802_2 [Accessed October 27, 2023].

Ringach DL, Hawken MJ, Shapley R (2003) Dynamics of Orientation Tuning in Macaque V1: The Role of Global and Tuned Suppression. J Neurophysiol 90:342–352 Available at: http://www.physiology.org/doi/10.1152/jn.01018.2002 [Accessed January 2, 2018].

Ringach DL, Sapiro G, Shapley R (1997) A subspace reverse-correlation technique for the study of visual neurons. Vision Res 37:2455–2464.

Roorda A, Williams DR (1999) The arrangement of the three cone classes in the living human eye. Nature 397:520–522 Available at: www.nature.com [Accessed February 11, 2021].

Rucci M, Iovin R, Poletti M, Santini F (2007) Miniature eye movements enhance fine spatial detail. Nat 2007 4477146 447:852–855 Available at: https://www.nature.com/articles/nature05866 [Accessed March 6, 2023].

Rucci M, Poletti M (2015) Control and Functions of Fixational Eye Movements. Annu Rev Vis Sci 1:499–518 Available at: https://www.annualreviews.org/doi/abs/10.1146/annurev-vision-082114-035742 [Accessed March 6, 2023].

Russ BE, Leopold DA (2015) Functional MRI mapping of dynamic visual features during natural viewing in the macaque. Neuroimage 109:84–94.

Rust NC, Schwartz O, Movshon JA, Simoncelli E (2004) Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey V1. Neurocomputing 58–60:793–799 Available at: https://www.sciencedirect.com/science/article/pii/S092523120400133X [Accessed October 5, 2018].

Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2005) Spatiotemporal Elements of Macaque V1 Receptive Fields. Neuron 46:945–956 Available at:

https://www.sciencedirect.com/science/article/pii/S089662730500468X [Accessed September 21, 2018].

Samengo I, Gollisch T (2013) Spike-triggered covariance: geometric proof, symmetry properties, and extension beyond Gaussian stimuli. J Comput Neurosci 34:137 Available at: /pmc/articles/PMC3558678/ [Accessed October 31, 2023].

Samonds JM, Potetz BR, Tyler CW, Lee TS (2013) Recurrent connectivity can account for the dynamics of disparity processing in V1. J Neurosci 33:2934–2946 Available at: http://www.jneurosci.org/content/jneuro/33/7/2934.full.pdf [Accessed May 25, 2017].

Sandler RA, Marmarelis VZ (2015) Understanding spike-triggered covariance using Wiener theory for receptive field identification. J Vis 15:16–16.

Sceniak MP, Ringach DL, Hawken MJ, Shapley R (1999) Contrast's effect on spatial summation by macaque V1 neurons. Nat Neurosci 2:733–739 Available at: http://www.nature.com/articles/nn0899_733 [Accessed September 19, 2018].

Schein SJ (1988) Anatomy of macaque fovea and spatial densities of neurons in foveal representation. J Comp Neurol 269:479–505 Available at: https://pubmed.ncbi.nlm.nih.gov/3372725/ [Accessed January 5, 2021].

Shapley R, Hawken MJ (2011) Color in the cortex: single- and double-opponent cells. Vision Res 51:701–717 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21333672 [Accessed July 10, 2018].

Sharpee T, Rust NC, Bialek W (2004) Analyzing Neural Responses to Natural Signals: Maximally Informative Dimensions. Neural Comput 16:223–250 Available at: https://dx.doi.org/10.1162/089976604322742010 [Accessed October 31, 2023].

Sharpee TO (2013) Computational Identification of Receptive Fields. Annu Rev Neurosci 36:103–120.

Sharpee TO, Atencio CA, Schreiner CE (2011) Hierarchical representations in the auditory cortex. Curr Opin Neurobiol 21:761–767 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21704508 [Accessed August 31, 2018].

Simoncelli EP, Paninski L, Pillow J, Schwartz O (2004) Characterization of Neural Responses with Stochastic Stimuli. MIT Press. Available at: http://pillowlab.princeton.edu/pubs/simoncelli03c-preprint.pdf [Accessed September 14, 2018].

Sincich LC, Horton JC (2005) The circuitry of V1 and V2: Integration of color, form, and motion. Annu Rev Neurosci 28:303–326 Available at: www.annualreviews.org [Accessed November 1, 2023].

Skottun BC, De Valois RL, Grosof DH, Movshon JA, Albrecht DG, Bonds AB (1991) Classifying simple and complex cells on the basis of response modulation. Vision Res 31:1078–1086.

Smith FW, Muckli L (2010) Nonstimulated early visual areas carry information about surrounding context. Proc Natl Acad Sci U S A 107:20099–20103 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21041652 [Accessed February 15, 2017].

Smith GB, Whitney DE, Fitzpatrick D (2015) Modular Representation of Luminance Polarity in the Superficial Layers of Primary Visual Cortex. Neuron 88:805–818 Available at: https://pubmed.ncbi.nlm.nih.gov/26590348/ [Accessed December 19, 2023].

Sommer MA, Wurtz RH (2006) Influence of the thalamus on spatial visual processing in frontal cortex. Nat 2006 4447117 444:374–377 Available at: https://www.nature.com/articles/nature05279 [Accessed October 31, 2023].

Spratling MW (2012) Predictive coding accounts for V1 response properties recorded using reverse correlation. Biol Cybern 106:37–49 Available at: http://link.springer.com/10.1007/s00422-012-0477-7 [Accessed September 4, 2018].

Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD (2019) Spontaneous behaviors drive multidimensional, brainwide activity. Science 364:255 Available at: http://www.ncbi.nlm.nih.gov/pubmed/31000656.

Talebi V, Baker CL (2016) Categorically distinct types of receptive fields in early visual cortex. J Neurophysiol 115:2556–2576 Available at: http://www.physiology.org/doi/10.1152/jn.00659.2015 [Accessed August 31, 2018].

Tanabe S, Cumming BG (2008) Mechanisms underlying the transformation of disparity signals from V1 to V2 in the macaque. J Neurosci 28:11304–11314 Available at: https://www.jneurosci.org/content/28/44/11304 [Accessed November 15, 2023].

Tanabe S, Cumming BG (2014) Delayed suppression shapes disparity selective responses in monkey V1. J Neurophysiol 111:1759–1769.

Taylor MM, Contreras D, Destexhe A, Frégnac Y, Antolik J (2021) An anatomically constrained model of V1 simple cells predicts the coexistence of push-pull and broad inhibition. J Neurosci:JN-RM-0928-20 Available at: https://www.jneurosci.org/content/early/2021/07/28/JNEUROSCI.0928-20.2021 [Accessed August 10, 2021].

Tootell RBH, Silverman MS, Hamilton SL, Switkes E, De Valois RL (1988a) Functional anatomy of macaque striate cortex. V. Spatial frequency. J Neurosci 8:1610–1624 Available at: https://www.jneurosci.org/content/8/5/1610 [Accessed October 27, 2023].

Tootell RBH, Switkes E, Silverman MS, Hamilton SL (1988b) Functional anatomy of macaque striate cortex. II Retinotopic organization. J Neurosci 8:1531–1568 Available at: https://pubmed.ncbi.nlm.nih.gov/3367210/ [Accessed July 20, 2023].

Touryan J, Felsen G, Dan Y (2005) Spatial structure of complex cell receptive fields measured with natural images. Neuron 45:781–791 Available at: http://www.cell.com/article/S0896627305000619/fulltext [Accessed October 29, 2023].

Touryan J, Lau B, Dan Y (2002) Isolation of relevant visual features from random stimuli for cortical complex cells. J Neurosci 22:10811–10818 Available at: /pmc/articles/PMC6758424/ [Accessed October 29, 2023].

Tring E, Duan KK, Ringach DL (2022) ON/OFF domains shape receptive field structure in mouse visual cortex. Nat Commun 2022 131 13:1–11 Available at: https://www.nature.com/articles/s41467-022-29999-7 [Accessed December 19, 2023].

Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. J Neurophysiol 93:1074–1089 Available at: https://journals.physiology.org/doi/10.1152/jn.00697.2004 [Accessed December 21, 2023].

Ustyuzhaninov I, Cadena SA, Froudarakis E, Fahey PG, Walker EY, Cobos E, Reimer J, Sinz FH, Tolias AS, Bethge M, Ecker AS (2020) Rotation-Invariant Clustering of Neuronal Responses in Primary Visual Cortex. Iclr:1–3.

Van Essen DC, Newsome WT, Maunsell JHR (1984) The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability. Vision Res 24:429–448 Available at: https://pubmed.ncbi.nlm.nih.gov/6740964/ [Accessed July 20, 2023].

Victor JD, Purpura K, Katz E, Mao B (1994) Population encoding of spatial frequency, orientation, and color in macaque V1. J Neurophysiol 72:2151–2166 Available at: http://www.ncbi.nlm.nih.gov/pubmed/7884450 [Accessed August 31, 2018].

Vintch B, Movshon JA, Simoncelli EP (2015) A Convolutional Subunit Model for Neuronal Responses in Macaque V1. J Neurosci 35:14829–14841 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26538653 [Accessed September 3, 2018].

Vintch B, Zaharia AD, Movshon JA, Simoncelli EP (2012) Efficient and direct estimation of a neural subunit model for sensory coding. Adv Neural Inf Process Syst 25:3113–3121 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26273181 [Accessed September 3, 2018].

Wandell BA (1995) Foundations of Vision, 1st ed. Sunderland, Mass.: Sinauer Associates. Available at: https://foundationsofvision.stanford.edu/ [Accessed December 19, 2023].

Wang J, Candy TR, Teel DFW, Jacobs RJ (2008) Longitudinal chromatic aberration of the human infant eye. J Opt Soc Am A Opt Image Sci Vis 25:2263 Available at: /pmc/articles/PMC2615001/ [Accessed January 16, 2024].

Wässle H, Grünert U, Röhrenbeck J, Boycott BB (1990) Retinal ganglion cell density and cortical magnification factor in the primate. Vision Res 30:1897–1911.

WEBB BS, TINSLEY CJ, BARRACLOUGH NE, EASTON A, PARKER A, DERRINGTON AM (2002) Feedback from V1 and inhibition from beyond the classical receptive field modulates the responses of neurons in the primate lateral geniculate nucleus. Vis Neurosci 19:583–592 Available at: http://www.journals.cambridge.org/abstract_S0952523802195046 [Accessed September 21, 2018].

Weber AI, Pillow JW (2017) Capturing the Dynamical Repertoire of Single Neurons with Generalized Linear Models. Neural Comput 29:3260–3289 Available at: https://dx.doi.org/10.1162/neco_a_01021 [Accessed December 21, 2023].

Weller JP, Horwitz GD (2017) Measurements of neuronal color tuning: Procedures, pitfalls, and alternatives. Vision Res Available at: https://www.sciencedirect.com/science/article/pii/S0042698917302031 [Accessed July 10, 2018].

Wells-Gray EM, Choi SS, Bries A, Doble N (2016) Variation in rod and cone density from the fovea to the mid-periphery in healthy human retinas using adaptive optics scanning laser

ophthalmoscopy. Eye 30:1135 Available at: /pmc/articles/PMC4985666/ [Accessed January 18, 2024].

Winter S, Sabesan R, Tiruveedhula P, Privitera C, Unsbo P, Lundström L, Roorda A (2016) Transverse chromatic aberration across the visual field of the human eye. J Vis 16:9–9.

Yates JL, Coop SH, Sarch GH, Wu R-J, Butts DA, Rucci M, Mitchell JF (2021) Beyond Fixation: detailed characterization of neural selectivity in free-viewing primates. bioRxiv:2021.11.06.467566 Available at: https://doi.org/10.1101/2021.11.06.467566 [Accessed January 23, 2022].

Yeh CI, Xing D, Williams PE, Shapley RM (2009) Stimulus ensemble and cortical layer determine V1 spatial receptive fields. Proc Natl Acad Sci U S A 106:14652–14657 Available at: www.pnas.org/cgi/doi/10.1073/pnas.0909233106www.pnas.org/cgi/doi/10.1073/pnas.0909457106 [Accessed December 22, 2020].

Zavitz E, Baker CL (2014) Higher order image structure enables boundary segmentation in the absence of luminance or contrast cues. J Vis 14:14–14 Available at: http://jov.arvojournals.org/Article.aspx?doi=10.1167/14.4.14 [Accessed August 31, 2018].

Zhaoping L (2017) Feedback from higher to lower visual areas for visual recognition may be weaker in the periphery: Glimpses from the perception of brief dichoptic stimuli. Available at: https://ac.els-cdn.com/S0042698917300858/1-s2.0-S0042698917300858-main.pdf?_tid=303edc64-efdb-11e7-932e-00000aab0f26&acdnat=1514911215_f8c77cddf52ab7b158602886bf8c8a9e [Accessed January 2, 2018].