# Neural computation of visual motion in macaque area MT

by

Andrew D. Zaharia

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Center for Neural Science

New York University

September 2016

---

J. Anthony Movshon

---

Eero P. Simoncelli

*Every man takes the limits of his own field of vision for the limits of the world.*

Arthur Schopenhauer, Studies in Pessimism: The Essays

## Dedication

In memory of my father, Vlad, and grandmother, Odette.

## Acknowledgements

I would like to thank my advisors, colleagues, friends, and family for their steadfast support throughout my PhD training. First, I thank my two advisors, Tony and Eero. Reading their papers as an undergrad in 2005 (especially Eero's model of MT, which is a primary focus of this thesis) was a memorable and formative experience. Working in their labs has been in many ways a dream come true. Their rigor, attention to detail, and excellence in both science and its communication has been both instructive and inspiring. I am grateful for the freedom and the opportunities they have given me.

I thank Lynne Kiorpes for the discipline and focus she brought to my committee, her advice on monkey behavior, and for her constant encouragement throughout my studies. I thank my other committee members, Nathaniel Daw and Greg DeAngelis, for their deep and insightful feedback.

Robbe and Brett were my two closest collaborators, and were both extremely generous as mentors. I thank Brett for showing me the ropes of the anesthetized prep and physiology, data analysis and model fitting, and cycling in New York City. I am immensely grateful to Robbe, for all the late nights he spent helping me with model fitting, creating slides, and for exposing me to his ways of thinking about problems in general.

I want to thank the members of Tony's and Eero's labs. Thanks to Romesh and Najib, for their help getting me up and running on not just one, but two awake recording rigs, and for training me in surgical technique. Thanks to Michael Gorman, who was a constant source of training and support in all aspects of monkey care and behavior. Thanks to Rob and Umesh, for teaching me the dark

arts of software version control and pomodoros. Thanks to Neil for advice on optimization, both of models and of life. Thanks to all the other members of the labs: Christopher, Luke, Yan, Chaitu, Deep, Jeremy, Alex, Olivier, Johannes, Elad, Yifei, Pascal, and James. Last but not least, I want to thank Corey for all the illuminating conversations we've had on science, film, and music, and his continued support and friendship throughout our journeys through grad school.

I want to thank my subjects, Albert and LW, from whom I recorded all the awake data in this thesis. There were portions of my graduate career during which I spent more time with them than any other primates, human or otherwise.

Finally, thanks to all my friends, NYU and non-NYU alike, who helped me keep (some) perspective. And thanks to my family, Marinela and Alan, and to my wife, Viktorya. Your love and support kept me going all these years.

## Preface

Chapters 2 and 3 are the result of a close collaboration with Robbe Goris in the labs of Eero Simoncelli and Tony Movshon. Robbe and Eero designed the stimuli, and Robbe collected most of the anesthetized dataset. Portions of chapter 2 have been presented previously [59, 182, 183], and is in preparation for publication. Chapter 3 represents the current state of ongoing work.

## Abstract

How does the visual system determine the direction and speed of moving objects? In the primate brain, visual motion is processed at several stages. Neurons in primary visual cortex (V1) filter incoming signals to extract the motion of oriented edges at a fine spatial scale. V1 neurons send these measurements to the extrastriate visual area MT, where neurons are selective for direction and speed in a way that is invariant to simple or complex patterns.

Previous theoretical work proposed that MT neurons achieve selectivity to pattern motion by combining V1 inputs consistent with a common velocity. Here, we performed two sets of experiments to test this hypothesis. In the first experiment, we recorded single-unit V1 and MT responses to drifting sinusoidal gratings and plaids (two gratings superimposed). These stimuli either had jointly varying direction and drift rate (consistent with a constant velocity) or independently varying direction and drift rate. In the second experiment, we presented arbitrary, randomly chosen combinations of gratings in rapid succession in order to sample, as widely as possible, the space of stimuli that could excite or suppress neural responses.

Responses to single gratings alone were insufficient to uniquely identify the organization of MT selectivity. To account for MT responses to both simple and compound stimuli, we developed new versions of an existing cascaded linear-nonlinear model in which each MT neuron pools inputs from V1. We fit these models to our data. By comparing the performance of the different model variants and examining the parameters that best accounted for the data, we showed that MT responses are best described when selectivity is organized along a common ve-

locity. This confirms previous predictions that MT neurons are selective for the arbitrary motion of objects, independent of object shape or texture. We explored new model variants of MT computation that capture this behavior. Our studies show that in order to characterize sensory computation, stimuli must be complex enough to engage the nonlinear aspects of neural selectivity. By exploring different linear-nonlinear model architectures, we identified the essential components of MT computation. Together, these provide an effective framework for characterizing changes in selectivity between connected sensory areas.

Supplementary materials: figures 3.4(a-e), 3.10(a-e), and 3.14(a-e) are rendered as movies.

# Contents

# List of Figures

# Chapter 1

# Introduction

A critical skill for an organism's survival is the detection of resources and threats in its environment. Humans and many other animals depend on vision for this. Predator and prey alike use vision to detect moving objects in their environment, assess their direction and speed, and use this information to make decisions. How does the brain detect and measure visual motion?

Primates dedicate upwards of 55% of cortex [48] to purely visual and visual association areas. Of these areas, there are three in particular in which sensitivity to motion is prominent and known to change in quality: primary visual cortex (V1), area V5/MT, and area MST. Selectivity and invariance for motion both increase [139] in each of these strongly connected areas [166, 48]. Neurons in other areas of the visual system are also selective for the direction of motion, such as V3 and V6 [8], FST [97, 37], and VIP [166, 28], but their precise selectivities and invariances have not been as extensively characterized as in V1, MT, and MST.

Direction-selective cells in V1 respond to the motion of oriented edges [74, 76] (figure 1.1). In MT, neurons are selective for direction of motion in a way that is invariant to other stimulus properties [12, 47, 4, 105, 6, 117]. MST neurons

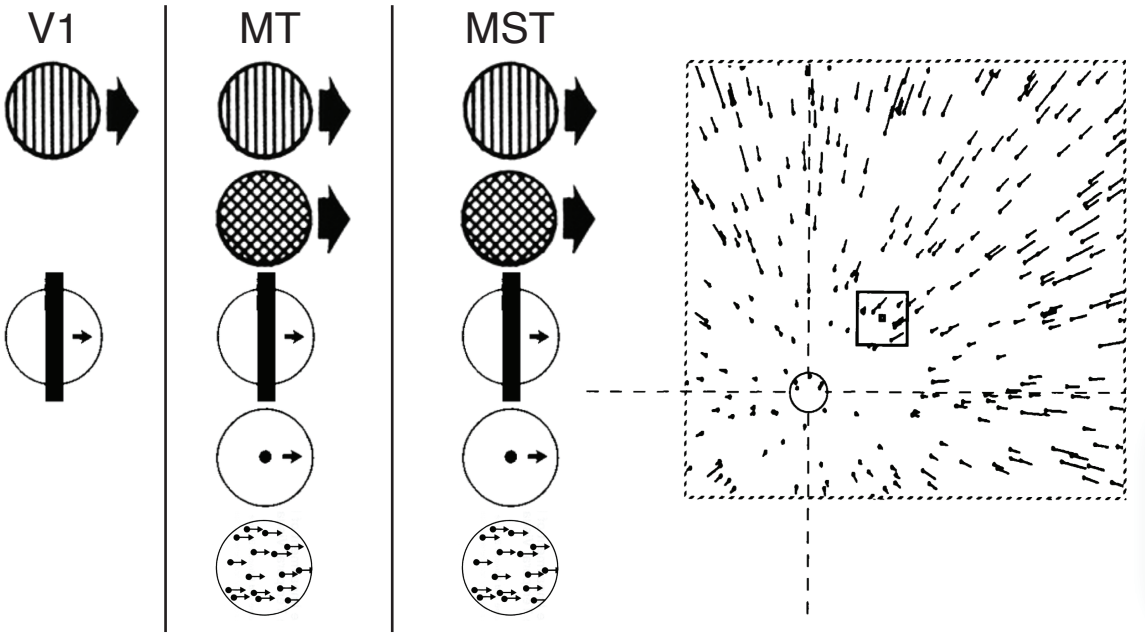


Figure 1.1: Different motion stimuli for which V1, MT, and MST are selective. V1 direction selective neurons exhibit direction tuning when presented with drifting oriented edges and sinusoids. MT neurons are additionally direction tuned for pattern motion. MST neurons are additionally tuned for optic flow. In the optic flow diagram, the small black filled square, larger open black square, and largest dashed square correspond to relative receptive field sizes for V1, MT, and MST neurons. Stimulus diagrams from [5, 105, 109, 181].

respond to stimuli containing just a one (global) velocity in a manner similar to MT neurons. In the dorsal aspect of MST (MSTd), neurons are additionally selective for more complex "optical flow" motion [143, 181, 43], which is motion consistent with the self-motion of an observer moving through an environment [43], such as a field of dots moving away from a central point.

While selectivity for the direction of motion is common in V1, many neurons are more strongly modulated by other stimulus dimensions, such as retinal position, orientation, size, color, and disparity [76]. Neurons in MT and MST, however, are particularly dedicated to visual motion processing in a way that V1 is not: nearly all cells are tuned for direction of motion [41, 184, 37, 143]. Furthermore, they

each represent the entire visual field in cortical areas roughly one twentieth the size of V1 [166, 48], suggesting that they are encoding fewer features at each position in the visual field.

This thesis focuses on the first site at which motion signals are transformed from edge motion to "general" motion—area MT. Specifically, we explore the question: how do MT neurons integrate input signals from V1 and transform them? How does MT represent motion, and what are the computations underlying the emergence of this representation?

By studying how the primate brain computes motion, we hope to better understand the calculations that individual neurons perform, and how these hint at fundamental computations that may underpin sensory processing in general [95, 40, 68, 23, 102]. To begin to address this question, we need to understand how the visual system is organized in the context of motion processing: what the inputs to the system are, how signals carrying information about motion are extracted and manipulated, and what the eventual outputs are.

## 1.1 The cortical representation of visual motion

The visual system has historically been separated into two parallel "streams" of processing: the ventral "what" stream, concerned with the identification of objects, and the dorsal "where" stream, localizing where objects are and may be moving [146, 165]. These streams have alternately been described as the "what" and "how" pathways [58], focusing instead on the assumed outputs of the pathways. Under this framework, the dorsal stream's function is to guide motor action based on objects identified by the ventral stream. Areas MT and MST, both dedicated to motion, are firmly in the dorsal/"where" pathway (figure 1.2; but see [57] for a

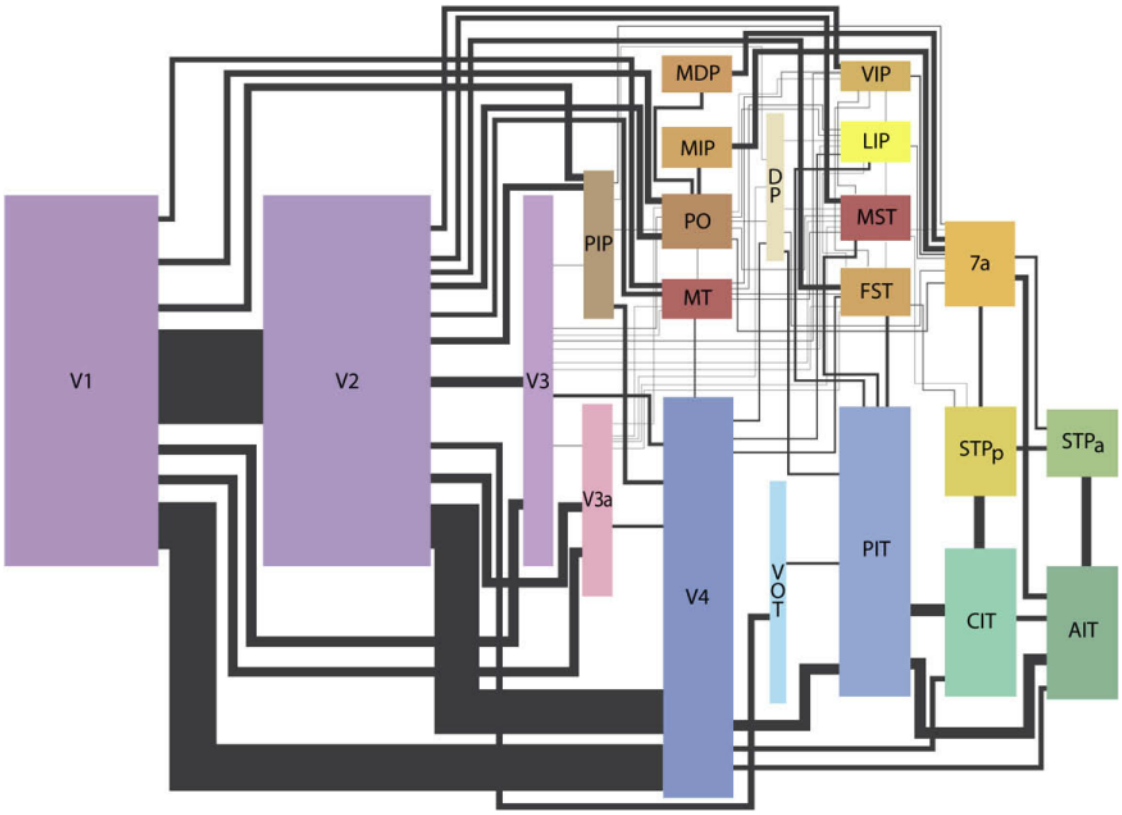


Figure 1.2: Macaque cortical visual areas.
The areas of the boxes and thicknesses of the lines correspond to the surface areas of each cortical visual area and the strength of the connections between them. The areas in the top half (reds and browns) are in the dorsal stream, and those in the bottom half (blues and greens) are in the ventral stream. From [173], modified from [48].

recent, critical examination of MT's placement in the dorsal pathway).

The visual system is organized hierarchically, with receptive fields increasing in size as one moves up the hierarchy [48], as well as in selectivity and invariance [139]. Vision begins with light entering the eye and hitting the retina. Changes in light intensity (i.e., contrast) across adjacent photoreceptors are detected by retinal ganglion cells and signaled to the lateral geniculate nucleus (LGN). In the LGN, the contrast measurements collected from the same visual hemifields by the two eyes are brought into register. V1 simple cells combine several inputs from the LGN, aligned along a specific orientation; they represent local orientation in terms of increments and decrements of light at a specific retinal location within their receptive fields [74, 130]. Complex cells, whose receptive fields are selective for orientation but invariant to the precise contrast polarity and location within the receptive field, combine spatially-offset simple cells with the same orientation preference [74, 104, 107]. V1 neurons appear to span a continuum of selectivities from the two extremes of simple and complex cell behavior [142]. In addition, both simple and complex V1 neurons can also be tuned for absolute disparity [74, 76, 121, 30, 126].

### 1.1.1 Motion selectivity begins in V1

A subset of both simple and complex cells in V1 are selective for the direction of motion [74, 76]. In the case of simple cells, changes over time in receptive field subregion selectivity for light and dark produce selectivity for motion [3, 34, 35, 142, 170].

The ways in which selectivity can change over time span two extremes: space-time separable and inseparable responses [3, 33] (figure 1.3). Separable responses

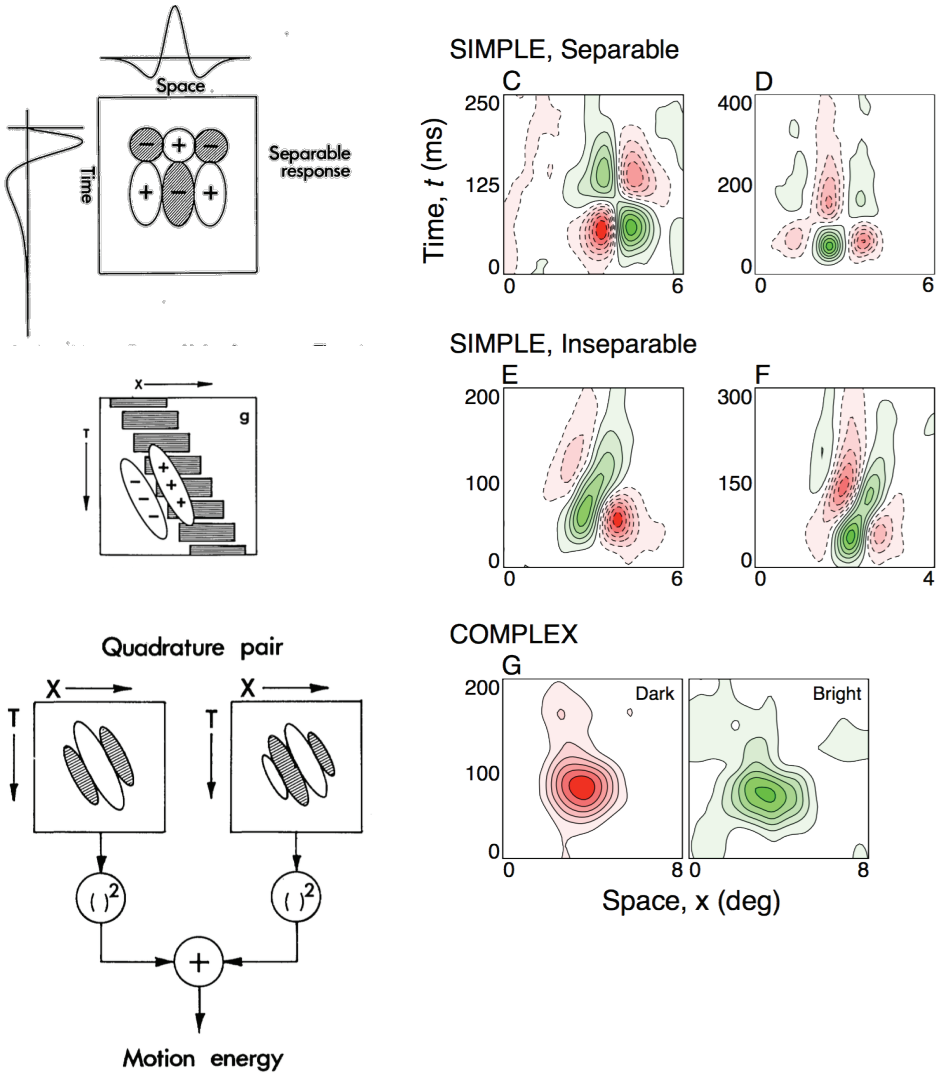


Figure 1.3: Space-time profiles of idealized motion detectors and measured V1 receptive fields.
Idealized separable and inseparable motion detectors on the left (from [3]). A phase-insensitive motion detector can be constructed by combining input from two filters with offset phases. On the right, measured spatiotemporal receptive field profiles from example V1 neurons (from [33]).

can be fully described by the separable product of one-dimensional tuning functions—in this case, for space and time. Separable V1 simple cell responses can be described as having their light- and dark-preferring subregions flip polarity after a delay. Inseparable responses cannot be described as a separable product; the two dimensions in question are jointly represented. Inseparable simple cells exhibit tuning in which the light- and dark-preferring subregions shift their spatial phase over time. Adelson & Bergen (1985) predicted that both separable and inseparable space-time tuning could produce sensitivity to motion, but that a separable representation cannot distinguish opposing directions of motion. An inseparable representation, however, can.

Since complex cells' phase-invariance makes their light- and dark-preferring subregions spatially overlap [74, 104, 3, 33], and thus cancel in a purely first-order, phase-dependent analysis, their space-time separability can only be assessed by examining their second-order response properties in terms of combinations of receptive field subunits [72, 107, 104, 33, 163, 142, 171, 170]. Subunits recovered from complex cells appear to be separable in space-time if they are not direction-selective [142] and inseparable in space-time if they are [142, 169, 170].

V1 directly contributes to motion selectivity in MT. MT receives a large proportion of its input from V1 [166, 97, 37], which is composed of predominantly spiny stellate neurons [108] from layer 4B, but also from layers 5/6 [90, 158, 161, 149]. Movshon & Newsome (1996) showed that neurons projecting from V1 to MT, identified through antidromic activation, are direction-selective, tend to be complex, and are broadly tuned to spatial and temporal frequency. By removing the inputs from V1, either through lesioning or reversibly cooling V1, Rodman et al. (1989) showed that MT neural responses are greatly diminished, but that direction

and disparity selectivity persists. In doing so, they showed that information about motion and disparity can reach MT through paths other than V1.

### 1.1.2 The mystery of the second visual area (V2)

The second visual area (V2) has a similar number of direction selective cells as in V1 (but many fewer than MT) [184, 41, 99]. In fact, three times as many neurons, retrogradely labeled from MT, were found in V2 as in V1 or V3, making V2 the strongest source of input to MT [97]. Disparity- and orientation-selective cells, primarily from the thick stripes [38, 148, 75, 86, 150] in layers 2/3 [97], project to MT. Reversible cooling of V2 and V3 significantly reduced disparity tuning in MT, without a corresponding reduction in direction selectivity [123].

While it is clear that V2 contributes disparity information (and serves as an indirect pathway for motion information [134]) to MT, it is not known how else it may be contributing. This may be because there are highly variable accounts of the stimulus features V2 encodes, which include illusory contours [71, 120], border ownership [186], complex shape characteristics such as curvature [**Essen2000b**], and angles [11]. Perhaps the most successful descriptions of V2 feature representation so far, in differentiating neural responses in V2 from those in V1 [187], are in terms of texture [52, 51, 53, 188, 187] and relative disparity [121, 122, 160, 15]. Both of these features are highly relevant for the interpretation of visual motion, yet precisely how their representation in V2 contributes to motion processing in MT is less well-known. Because of our fuller understanding of direction-selective neurons in V1, we will focus on their contribution to the interpretation of two-dimensional image motion in MT.

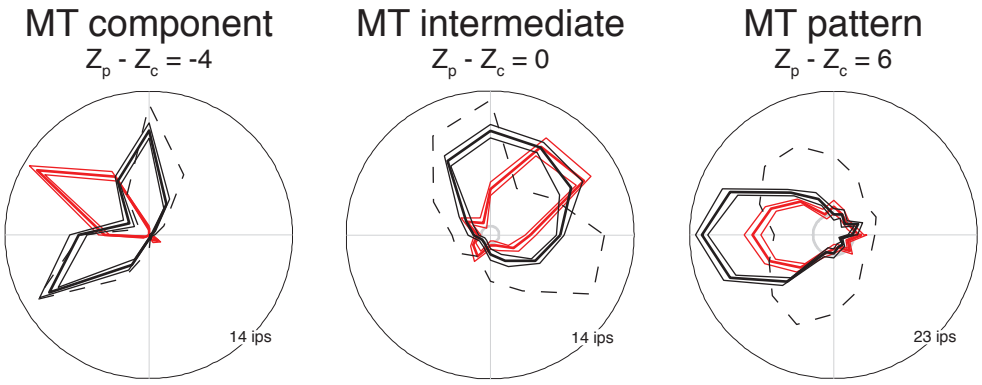


Figure 1.4: Example MT tuning direction tuning curves.
Responses to single gratings and plaids are in red and black, respectively. Thick traces indicate the mean and thin traces ±1 s.e.m. Dashed black lines indicate the component prediction.

### 1.1.3 Motion tuning becomes more invariant in MT

Neurons in MT represent motion differently from their V1 inputs. Like V1, MT receptive fields are localized at a specific retinal position in the visual field [9, 41, 166]. However, MT receptive fields span areas of the visual field 5-10 times larger than those in V1 (and about 2-3 times larger than those in V2) at the same eccentricity [166, 56, 37, 25, 52]. MT neurons must therefore be integrating inputs from a number of neurons to represent motion at all locations within their receptive fields.

Even in early experiments, direction selectivity appeared to be different in MT than in V1. MT neurons were tuned for the direction of motion of different types of moving stimuli, including bars [41, 166, 99], single dots [5], and random dot fields [5, 111].

Wallach (1935) observed that in order to identify the true direction of motion of a drifting pattern, viewed through an aperture, more than one orientation needs to be present. The motion of a single drifting orientation can be interpreted as having

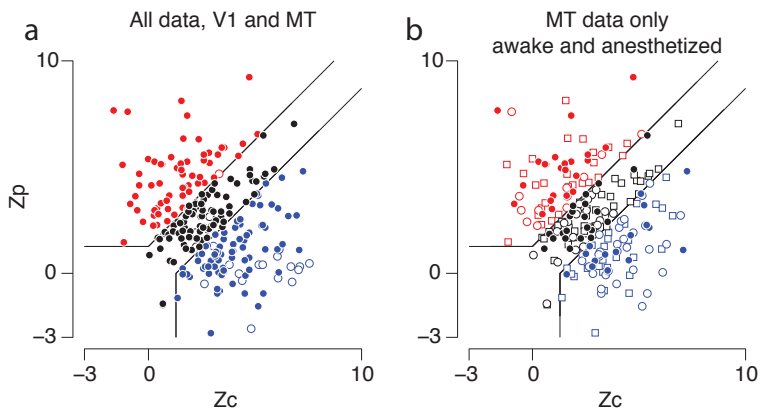


Figure 1.5: Pattern classification of all recorded V1 and MT neurons.
The Z-scored, normalized correlation between recorded plaid tuning to the component and pattern predictions (Zc and Zp, respectively). Pattern cells are in red, intermediate in black, and component in blue. (a) All the recorded neurons featured in this thesis are shown ($n = 180$). Filled circles are MT neurons, open are V1 neurons. (b) Only MT data is shown ($n = 160$). Filled circles are anesthetized data, open circles are awake data from monkey LW, open squares are awake data from monkey A.

any non-parallel direction [172, 180, 94], a phenomenon later termed the "aperture problem" [94]. Building on this, Movshon et al. (1985) showed, using sinusoidal grating and plaid (two gratings superimposed) stimuli, that about a third of MT neurons are selective for the direction of coherent pattern motion (example cell tuning curves in figure 1.4). Just as V1 neurons span a spectrum of simple and complex selectivity, neurons in MT span a continuum of pattern selectivity [105, 141] (figure 1.5). Pattern direction selectivity, however, is exceedingly rare in V1 [105]. Direction tuning appears to be relatively constant at different locations in the receptive field [128] (but see [29, 131]). Spatial and temporal frequency tuning is relatively broad in MT, compared to V1 as a whole, but similar to the subset of direction-selective neurons in V1 [106, 66, 174].

MT neurons are tuned for several other stimulus attributes other than direction of motion. One is speed, whether it be to moving bars [99, 5], single dots [5, 135],

gratings and plaids [133], or random dot fields [5]. MT appears to have Gaussian-shaped tuning for speed on a logarithmic axis [115]. Another is disparity [185, 96, 36, 164]. Tuning for disparity appears to be separable with respect to direction tuning [157]. MT responses are also modulated by stimulus size [166, 7, 98, 147, 70, 91].

Aside from MT's coarse retinotopic structure [166, 37], it may have a columnar organization based on direction [99, 5] and disparity [32].

A number of perceptual studies have implicated the importance of MT in motion perception. Lesioning MT led to decreased motion sensitivity to random dot fields [109]. Electrical microstimulation in MT led to biased performance in direction discrimination tasks [144, 145] and depth discrimination tasks [36]. Neurons in MT have also been shown to be weakly correlated with decisions made in direction discrimination tasks [110, 17, 18, 127].

### 1.1.4 MST: more selective, more invariant

Aside from feedback connections to V1 and V2, a major destination of MT projections is MST [97, 14]. Neurons in MST are selective for the direction of motion [143, 83] in a way that reflects and builds upon the input received from MT. First, the majority of neurons in MST are pattern selective [83]. Second, MSTd neurons are selective for more complex types of motion that include more than one direction. These types of motion include expansion and contraction, rotation [143, 42], and shear (the combination of translation and rotation) [103].

In addition to direction of motion, MST is selective for relative disparity (depth) [45]. Depth and optic flow provide information about an individual's own motion through an environment [42], and there is compelling evidence that MST is special-

ized for this purpose. MSTd neurons integrate signals from the vestibular system to compute estimates of heading direction [63, 62].

## 1.2 Computational models of motion

### 1.2.1 Computer vision approaches

Computer vision models of motion have focused on the problem of estimating motion in moving images from the perspective of optical flow. Horn & Schunck (1981) defined optical flow as "the distribution of apparent velocities of movement of brightness patterns in an image... [arising] from relative motion of objects and the viewer." They recognized that only measuring local changes in pixel values between frames and nearby spatial locations does not uniquely constrain object motion [172, 49, 73]. A previous approach had been to calculate local pixel gradients and apply a clustering algorithm afterwards to smooth the resulting optical flow fields [49]. Instead, Horn & Schunck (1981) built a model which formulated optic flow as a series of partial derivative equations to be numerically solved. Importantly, they introduced two constraints on flow: global brightness constancy and (global) spatial smoothness of optic flow fields.

Lucas & Kanade (1981) used a similar, but more extreme version of the spatial smoothness constraint: they assumed optic flow was approximately constant in local patches of pixels. This allowed them to reduce the optic flow calculation to a (much more computationally tractable) least squares problem [88]. As a consequence of its relatively simple formulation and calculation, this algorithm underpins many contemporary optic flow estimation procedures [19, 136].

It can be argued that signals reaching MT already have these two constraints

imposed: local contrast is subject to gain control in the retina, and the larger receptive fields in MT combine a number of more local measurements of motion from their V1 inputs, likely with the same or similar direction preferences [128] (but see [131]).

### 1.2.2 Computer vision meets biology

From the biological perspective, models of motion also began with few constraints. The Reichardt (1961) model (originally of fly vision) detects motion by computing the correlation between the luminance at a given location and at an adjacent spatial location, with a temporal delay between the two. The result is a model that detects motion at a particular direction and speed, but its output is modulated by the spatial phase and luminance of its input. Inverting the contrast of a moving image, for example, will also invert the response of such a detector, even if the direction of motion has not changed.

In a highly influential model, Adelson & Bergen (1985) extended this framework to create a phase-invariant motion detector, inspired by existing models of V1 neurons [74, 162, 107]. Building on recent observations that objects moving at a constant velocity have constant spatiotemporal slope in the frequency domain [46, 176], they introduced the idea of "motion as orientation" in space-time (figure 1.1). In their model, an ideal motion detector is oriented, and thus inseparable, in space-time. The full model measures motion "energy" by summing the squared output of two filters oriented in space-time, selected specifically to be 90 degrees out of phase (i.e., in "quadrature", see figure 1.1). These filters are tuned for both direction and speed, but in a manner insensitive to phase. The filters themselves are implemented as Gabor functions (a 2D sinusoid windowed spatially with a

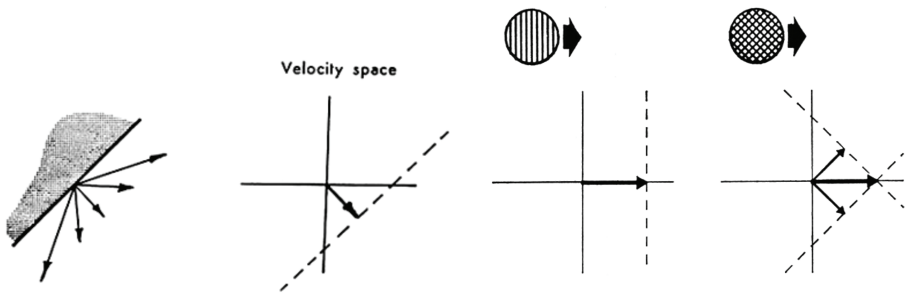


Figure 1.6: Intersection of constraints.
From [105].

Gaussian [54]). These had been proposed as ideal motion filters [176] and as a model of V1 simple cells [93, 107], later verified [80]. As discussed in the previous section, both inseparable Gabor-like filter and separable filter responses have been observed in V1 [34, 163, 142, 169, 170] (figure 1.3).

The motion energy model alone is only selective for edge motion—it will not produce pattern selectivity. The "intersection of constraints," formulated by Adelson & Movshon (1982), is a geometric solution to isolate the unique velocity of a rigidly moving pattern. A single drifting edge, viewed through an aperture, can be explained by the set of velocities lying on a constraint line in velocity space [172, 4, 105] (figure 1.6). The velocities of two orientations in a drifting plaid uniquely identify the velocity of the plaid where the constraint lines intersect in velocity space (figure 1.6).

### 1.2.3 Cascade models of MT motion processing

Combining Gabor filters with the intersection of constraints formulation, Heeger (1987) built a motion of optical flow estimation and MT function. The model was able to extract optic flow on arbitrary images and produced (not entirely biologically realistic) pattern-selective responses to grating and plaids. Recognizing

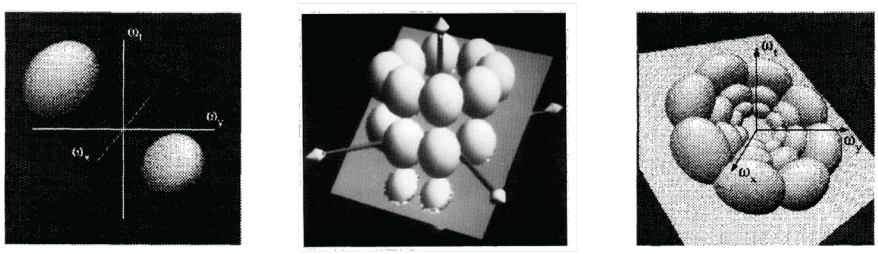

Figure 1.7: V1 and MT spatiotemporal frequency filter responses in frequency space.
(left) The frequency response of a single spatiotemporal filter in the V1 stage. (middle) The cylindrical lattice of Gabor filter responses, used in the Heeger (1987) model. The plane corresponding to the preferred velocity of the pattern cell is also shown. (right) A pattern cell in the Simoncelli & Heeger (1998) model. Figures from [67, 155].

that a rigidly moving texture corresponds to energy on a plane in frequency space [176, 177], the model calculated velocity by fitting a plane to Gabor filter responses on local image patches. In principle, this is the same constant local flow assumption used by Lucas & Kanade (1981); its implementation, however, is global and uses non-biologically plausible operations.

The Heeger (1987) model used a set of Gabor filters whose spatiotemporal frequency response were arranged in a cylindrical lattice (figure 1.7). Then, a population of velocity-selective (MT pattern) cells were simulated, each taking the difference of the sum of the squared filter responses computed on the image and the cell's preferred velocity plane. The velocity preference of the cell with the highest activation was then read out as the velocity predicted by the model.

Simoncelli and Heeger [154, 155] used this model as a foundation for a new model of MT pattern selectivity by adopting a cascaded computation architecture [68] and incorporating more biologically plausible computations. It improved accuracy by using derivative filters [151] on a spherical lattice [154] (figure 1.7). The model also gives improved predictions to grating and plaid responses, as well as

other observed MT response properties, such as speed tuning, as measured with drifting bars [135].

The cascade model repeated the same sequence of operations in the simulated V1 and MT stages. In the initial V1 stage, the image is filtered, then the filter responses are summed, half-wave rectified and squared, and subject to normalization. The normalization step leaves tuning intact, but accounts for other nonlinear behaviors such as contrast response saturation and cross-orientation suppression [69, 22, 24]. The result is the simple cell response. Summing over simple cells with spatially offset receptive fields yields the complex cell responses. These serve as input to the MT stage, which linearly weights complex cell responses on a preferred velocity plane. These responses are summed, half-squared, and normalized.

Since the weights in the MT stage (off the preferred velocity plane) take negative values, there is some suppression to motion in the opposite direction. Opponent suppression is also part of both the Adelson & Bergen (1985) and Heeger (1987) models and has been observed in MT [99, 47, 135], although it could be inherited from V1 [140, 170].

The cascade models introduced predictions about the construction of MT receptive fields that had not yet been verified—specifically, the notion that MT receptive fields are organized along a preferred velocity plane in frequency space, and that normalization in MT plays a role in shaping direction selectivity. The experiments described in the chapters in this thesis aim to test these predictions on macaque MT neural responses, building on further simplifications that Mante (2000) and Rust et al. (2006) made to enable the cascade model to be fit to data.

## 1.3 Automated and model-based approaches to characterizing visual receptive fields

In order to characterize receptive fields along more than one stimulus dimension in the limited time constraints of a typical physiological experiment, experimenters have turned to automated techniques.

In V1, space-time receptive fields have been mapped through the rapid presentation of light and dark spots at random spatial locations [81, 34]. By computing a sum of these stimuli, weighted by the spikes occurring in response, one obtains a spike-triggered average (STA) [27, 153]. Other randomized stimuli have been used to characterize V1 neurons as well, such as 2D sinusoids [132], light and dark bars [142], 2D Gaussian noise [113], and natural images [163, 26].

STA methods had early success characterizing phase-sensitive simple cells [81, 34, 132], but second-order methods were required to resolve phase-insensitive complex cell receptive fields. Complex cell receptive field elements were observed using spike-triggered covariance (STC). In STC, the eigenvectors of the covariance matrix of the spike-triggered stimuli (with the STA projected out) represent overlapping elements of the receptive field [153, 142, 163]. An alternative approach, local spectral reverse correlation, calculated the spike-triggered average on the frequency spectra of local image patches [113].

To reduce bias in recovered receptive field elements from, and increase their predictive power to, randomized stimuli, researchers have begun directly fitting receptive field models to these data [171, 87, 170].

In MT, sparse flashed bar and dot stimuli have been used to generate difference maps, representing the apparent motion generated by the change in relative

position of the bar or dot from one frame to the next [85, 118]. These maps could accurately predict the preferred direction of the neurons [118], but their predictive power for pattern direction, speed, and spatiotemporal frequency tuning was not directly verified.

Rust et al. (2006) used hyperplaids (six gratings superimposed) to stimulate MT neurons and then fit a version of the Simoncelli & Heeger (1998) cascade model to the responses. Their model provided accurate predictions of pattern selectivity, and they identified opponent suppression and V1 normalization as crucial features mediating it. Given that their stimuli did not vary in spatial or temporal frequency, they could not make any statement about the organization of MT receptive fields in those dimensions.

Nishimoto & Gallant (2011) used "motion-enhanced" natural movies (movies of natural scenes with rapidly-moving computer-generated shapes superimposed) to stimulate MT neurons (figure 1.8). They characterized MT receptive fields in 3D frequency space by fitting a version of the cascade model. They recovered "partial ring"-shaped linear weights confined to the preferred velocity plane (figure 1.8). These partial rings have narrower bandwidth in the direction domain than was predicted by the original cascade models. Using hyperplaids to calculate STAs in the frequency domain, Inagaki et al. (2016) recovered excitatory weights with a similar, but somewhat less planar, structure and weaker suppressive weights (figure 1.8). The difference in the shapes of the recovered weights is likely due to the stimuli used by Nishimoto & Gallant (2011) being biased towards having a planar structure, whereas no such bias exists in the Inagaki et al. (2016) stimuli. In both of these studies, pattern selectivity was not directly verified, leaving the 3D frequency structure of pattern-selective MT receptive fields unresolved.

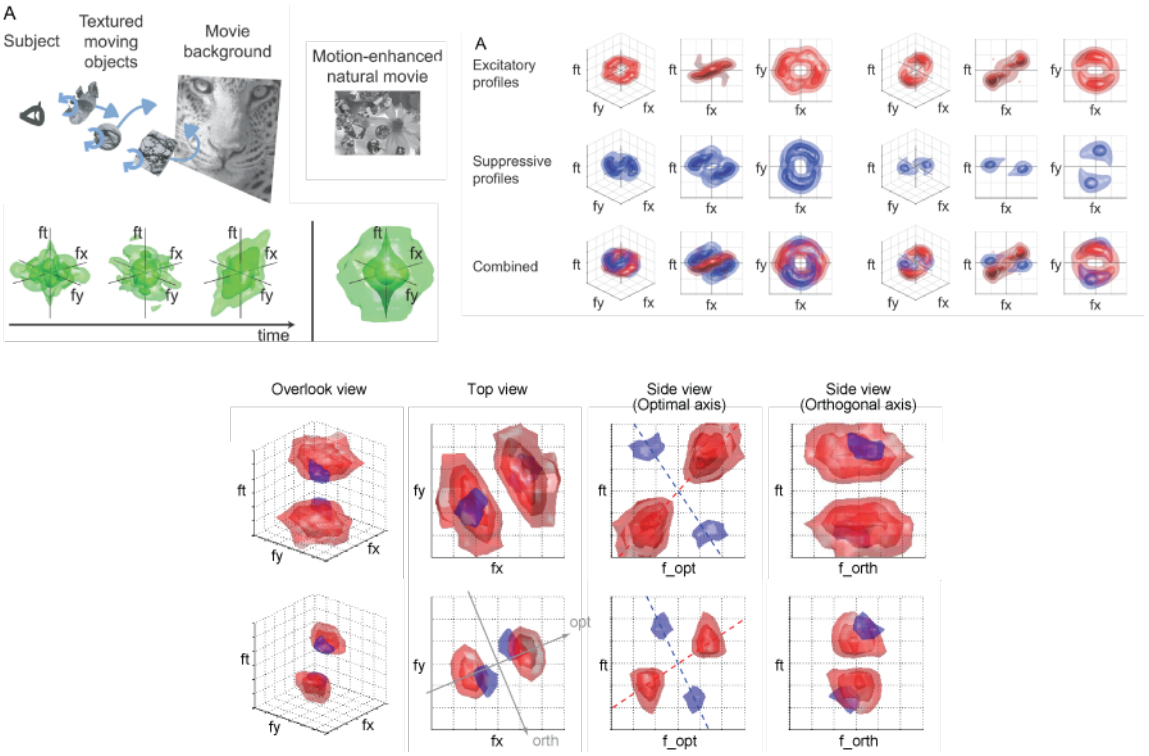


Figure 1.8: Previous characterizations of MT spatiotemporal selectivity in 3D. (top left) The motion-enhanced natural movie construction. (top right) Example receptive fields recovered by Nishimoto & Gallant (2011). (bottom) Example receptive fields recovered by Inagaki et al. (2016). Figures from [112, 78].

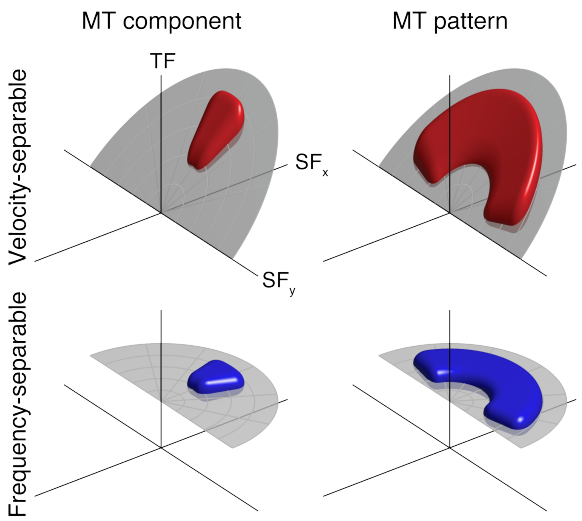


Figure 1.9: Example idealized separable MT receptive fields.
Idealized component (left) and pattern selectivity (right), organized separably along preferred velocity and preferred temporal frequency planes (top and bottom, respectively).

We set out to assess spatiotemporal frequency selectivity in MT neurons, specifically to ascertain if they are organized along a preferred velocity plane. In chapter 2, we address this question by introducing a simplifying assumption to the cascade model: that MT receptive fields are separable in 3D frequency space.

Separability of tuning has been examined throughout the dorsal stream. In V1, separability has been assessed in terms of spatial and temporal frequency [162, 16, 125], orientation and spatial frequency [31, 178, 81, 65, 168, 100], and direction and disparity [10, 61]. Direction and disparity tuning has been shown to be separable in MT [157] and inseparable in a subset ("almost 50%" [138]) of MSTd neurons [137, 138]. In MT, separable tuning for spatial and temporal frequency tuning has been proposed [67, 155] and tested in two dimensions [124, 125, 89], but not in three.

We compared two versions of the cascade model, each separable in three dimen-

sions of frequency space. One was separable about the preferred velocity plane, the other about the preferred temporal frequency plane (figure 1.9). The latter model, representing our null hypothesis, was chosen because tuning to spatial and temporal frequency at earlier stages of visual processing (e.g., photoreceptors and V1 simple cells) is separable [125].

In chapter 3, we presented hyperplaids and performed reverse correlation and model fitting to resolve MT receptive fields in 3D frequency space, without any assumption of separability. In both these chapters and in the final one, we examine the relationship between the recovered receptive field structures and pattern selectivity.

# Chapter 2

# A planar-separable model of MT selectivity

## 2.1 Introduction

Visual motion is processed in multiple stages in the primate brain [4, 76, 41]. In the first stage, incoming visual signals are filtered in space and time. This operation reveals the motion of local oriented elements in visual scenes and is represented by the activity of neurons that are selective for direction of motion in the primary visual cortex (V1). These neurons, considered in isolation, cannot uniquely identify the coherent motion of a complex pattern containing multiple oriented components, since its components could be moving independently and in different directions. V1 neurons provide input to the extrastriate area MT (V5), where the second stage of motion processing occurs. MT neurons compute complex pattern motion, becoming more invariant to component orientation [105].

An influential model of MT computation proposes that selectivity for pattern motion emerges by pooling responses of V1 neurons whose preferred stimulus direction, spatial frequency, and temporal frequency are consistent with a common

velocity [155]. This type of receptive field organization is ideally suited to analyze rigidly moving objects, and there is some empirical evidence in support of this idea [112, 124]. This conversion of information from V1 into a velocity-based representation in MT is the fundamental computation in the model that underlies MT's solution to the motion ambiguity problem. However, this is not the only kind of receptive field organization consistent with known MT properties. An alternative possibility is that MT neurons pool V1 responses with a common temporal frequency preference determined independently from direction of motion. This type of organization is consistent with neural selectivity observed earlier in the visual processing hierarchy, such as V1 simple cell selectivity [125]. An MT neuron with this type of organization would still be direction-selective, but in a manner that is less invariant to pattern.

We compared these two models of MT computation by measuring the responses of neurons in areas V1 and MT of anesthetized and awake macaques to a large collection of sinusoidal gratings and plaids (superimposed gratings with different orientations). We fit these responses with a linear-nonlinear model of MT computation, in which the linear receptive field was constructed by either a joint or an independent representation of motion direction and speed. We refer to the former model as the velocity-based model and to the latter as the frequency-based model. V1 neurons tended to be better described by the frequency-based model. When probed with single sinusoids, MT responses were similarly well described by both models. However, when probed with more complex plaid stimuli, the velocity-based model systematically outperformed the frequency-based model for pattern-selective neurons. Our results clarify how receptive field organization changes throughout the visual hierarchy and demonstrate that stimulus complexity

determines the effective complexity of neural computation.

## 2.2 Methods

### 2.2.1 Anesthetized recording procedures

We recorded from 7 anesthetized, paralyzed, adult male macaque monkeys (*M. fascicularis*) and one adult female macaque (*M. mulatta*) using our standard procedures for surgical preparation and single-unit recording, as described previously [25]. We maintained anesthesia and paralysis by intravenously infusing sufentanil citrate (6-30 $\mu$g kg$^{-1}$ h$^{-1}$), and vecuronium bromide (Norcuron, 0.1 mg kg$^{-1}$ h$^{-1}$), respectively, in isotonic dextrose-Normosol solution (4-10 mL kg$^{-1}$ h$^{-1}$). We continuously monitored vital signs (heart rate, lung pressure, electroencephalogram (EEG), electrocardiogram (ECG), body temperature, urine flow and osmolarity, and end-tidal $CO_2$ partial pressure ($pCO_2$)) and maintained them within appropriate physiological ranges. We applied atropine topically to dilate the pupils. Gas-permeable contact lenses protected the eyes. We refracted them with supplementary lenses chosen by direct ophthalmoscopy. Experiments typically lasted 5-7 days at the end of which the monkey was killed with an overdose of sodium pentobarbital. We conducted all experiments in compliance with the US National Institutes of Health Guide for the Care and Use of Laboratory Animals and with the approval of the New York University Animal Welfare Committee.

The monkey was positioned so his eyes were 57-114 cm from the display. Grating and plaid stimuli each lasted for 1,000 ms and were presented in randomly interleaved blocks. We used quartz-platinum-tungsten microelectrodes (Thomas Recording) to make extracellular recordings in the brain through a craniotomy

and small durotomy. For each isolated unit, we determined eye dominance and occluded the non-preferred eye.

### 2.2.2 Awake recording procedures

We also recorded from 2 awake, actively fixating, adult male macaques (one *M. mulatta* and one *M. nemestrina*). We surgically implanted a headpost for head stabilization using the design and methods described in [1]. In a second surgical procedure, we implanted a chamber for chronic electrode recording over the superior temporal sulcus (STS) of the left hemisphere, using the techniques and a variant of the design described in [2]. Prior to surgery, we used structural MRI and Brainsight software (Rogue Research, Canada) to design a chamber with legs matched to the curvature of the monkey's skull [79] above the STS.

We acclimated each monkey to his recording chair and experimental surroundings. After this initial period, he was head-restrained and rewarded for looking at the fixation target with dilute juice or water. Meanwhile, we used an infrared eye tracker (EyeLink 1000; SR Research, Canada) to monitor eye position at 1000Hz via reflections of infrared light on the cornea and pupil. The monkey sat 57 cm from the display.

The monkey initiated a trial by fixating on a small white spot (diameter 0.1°), after which he was required to maintain fixation for a random time interval between 2,350 and 4,350 ms. A grating or plaid stimulus would appear 100 ms after fixation began and last for 250 ms. Stimulus conditions were presented in randomly interleaved blocks. The monkey was rewarded if he maintained fixation within 1-1.75° from the fixation point for the entire duration of the stimulus. No stimuli were presented during the 300 to 600 ms in which the reward was being delivered. If

the monkey broke fixation prematurely, the trial was aborted, a timeout of 2,000 ms occurred, and no reward was given.

We used tungsten microelectrodes (FHC, Bowdoin, ME) to make extracellular recordings. We identified area MT from gray matter-white matter transitions and isolated neurons' brisk, direction-selective responses.

### 2.2.3 Visual stimulation

We presented visual stimuli on a gamma-corrected CRT monitor (Eizo T966 (anesthetized experiments), HP P1230 (awake experiments); mean luminance, 33 cd/m$^2$) at a resolution of 1,280 $\times$ 960 with a refresh rate of 120 Hz. Stimuli were generated and presented on an Apple Mac Pro using Expo software (`http://corevision.cns.nyu.edu`).

For each isolated unit, we presented vignetted sinusoidal grating stimuli to map each cell's receptive field and determine its preferred size. We then characterized neuronal selectivity for direction, spatial frequency, and drift rate. Thereafter, stimuli were presented in a window of the preferred size at specific spatiotemporal frequencies relative to the optimal spatiotemporal frequency. All receptive fields were centered between 2° and 30° from the fovea.

For the single component study, 17 unique tuning curves were measured. All featured single gratings presented at 100% contrast. Two direction tuning curves from -90° to 90° relative to the preferred direction, in 15° intervals, were collected along the optimal frequency-based path (keeping the optimal spatial and temporal frequencies constant) and along the optimal velocity-based path (keeping the optimal velocity constant). Four direction tuning curves were collected at 18° intervals from -90° to 90° relative to the preferred direction: one at a higher and

one at a lower than optimal temporal frequency while fixing the optimal spatial frequency, and two more at a high and a low spatial frequency while fixing the optimal temporal frequency.

Two spatial frequency tuning curves, at 13 log-spaced values from 0.1 cycles/degree to 10 cycles/degree, were collected along the optimal frequency- and velocity-separable paths. Four spatial frequency tuning curves, at 11 log-spaced values from 0.1 cycles/degree to 10 cycles/degree, were collected at a high and low temporal frequency while maintaining the optimal direction. Two more were collected at suboptimal directions, while maintaining the optimal temporal frequency.

One temporal frequency tuning curve, at 13 log-spaced values from 0.1 cycles/second to 60 cycles/second, was collected at the optimal direction and spatial frequency. Four temporal frequency tuning curves, at 11 log-spaced values from 0.5 cycles/second to 60 cycles/second, were collected at a high and low spatial frequency while maintaining the optimal direction. Two more were collected at suboptimal directions, while maintaining the optimal spatial frequency. The "high" and "low" non-preferred spatiotemporal frequencies used in suboptimal tuning curves were chosen to maximally distinguish the frequency- and velocity-separable models.

For the two-component study, four unique direction tuning curves at the optimal spatial frequency were measured (see figure 2.4(a,b)). The first two were single, 50% contrast gratings, one with temporal frequency held constant at the optimal value (frequency-based) at all directions in 30° intervals, and one with constant, optimal velocity (velocity-based) from -90° to 90° relative to the preferred direction, in 15° intervals. Since velocity represents both direction and speed, and is uniquely represented as a tilted plane in frequency space, the velocity-based

gratings vary in temporal frequency with the cosine of the direction. This ensures that the velocity-based gratings presented are always on the optimal velocity plane. The two main differences between these and the first two direction tuning curves of the single component study were: (1) the gratings were at 50% contrast instead of 100%, and (2) the constant temporal frequency gratings spanned the whole range of directions rather than just the semicircle of directions centered at the preferred one.

The last two tuning curves consisted of 120° "plaids", or two superimposed gratings with orientations 120° apart. Their component gratings each had 50% contrast, making each plaid's contrast 100%. The plaids were presented with the mean angle of their two gratings matched to the directions of the single grating tuning curves.

In response to the single component study, we recorded single-unit responses of 12 V1 neurons and 39 MT neurons (all anesthetized). For the two-component study, we recorded 20 V1 neurons (all anesthetized) and 111 MT neurons (54 anesthetized, 30 from awake monkey A, and 27 from awake monkey LW). For 29 of the 54 anesthetized MT neurons in the two-component study, the single component study was also run. All of the 12 V1 neurons from the simgle component study are in the set of 20 V1 neurons in the two-component study.

### 2.2.4 Analysis of neuronal response

Following stimulus onset, we counted spikes in either a 1,000 ms window (anesthetized experiments) or a 250 ms window (awake experiments). We estimated the latency of each cell by maximizing the sum of variances of its responses for all stimulus conditions [25] and shifted the spike count window accordingly.

To characterize MT cell selectivity for pattern motion, we used standard methods to compute each cell's "pattern index" [105, 156]. First, we computed partial correlations between the actual response to (constant temporal frequency) plaids with idealized predictions of pattern and component direction selectivity ($r_p$ and $r_c$, respectively). We then converted these values to Z-scores to stabilize the variances of the correlations ($Z_p$ and $Z_c$). Finally, the pattern index is the difference of these two quantities: $Z_p - Z_c$. Cells were classified as pattern selective if $Z_p - Z_c > 1.28$, or component-selective if $Z_c - Z_p > 1.28$. Both thresholds correspond to a significance of $P = 0.90$. Confidence intervals on pattern index were computed from the standard deviation of 100 bootstrapped estimates [44, 141].

### 2.2.5 The frequency- and velocity-based models

The MT linear weighting functions for both the frequency- and velocity- based models are defined in terms of a separable product of tuning functions $w_d$, $w_s$, and $w_t$ in the direction, spatial frequency, and temporal frequency dimensions. These functions operate independently on the $i$th stimulus component's direction ($d_i$), spatial frequency ($s_i$), or temporal frequency ($t_i$) There is either one component for a grating or two for a plaid. The frequency-separable linear weighting on the $i$th component, $F_i$, is defined as follows:

$$F_i\left(d_i, s_i, t_i\right) = w_d\left(d_i\right) \cdot w_s\left(s_i\right) \cdot w_t\left(t_i\right) \tag{2.1}$$

Direction tuning is represented above by a von Mises function:

$$w_d\left(d\right) = \frac{e^{\sigma_d cos(d-\mu_d)}}{2\pi I_0(\sigma_d)} \tag{2.2}$$

where $\mu_d$ and $\sigma_d$ represent the direction preference and bandwidth, respectively, and $I_0()$ is the modified Bessel function of order 0 (which normalizes the integral of the von Mises). Spatial frequency is represented by a Gaussian function, $\mathcal{N}(x \mid \mu, \sigma)$, in log2 coordinates, with spatial frequency preference $\mu_s$ and bandwidth $\sigma_s$:

$$w_s(s) = \mathcal{N}(log_2(s) \mid log_2(\mu_s), \sigma_s) = \frac{1}{\sigma_s\sqrt{2\pi}} e^{-(log_2(s)-log_2(\mu_s))^2/2\sigma_s^2} \tag{2.3}$$

Temporal frequency is represented by a Gaussian in coordinates which are linear at low frequencies and logarithmic at higher ones, determined by the function $g(t)$:

$$w_t(t) = \mathcal{N}(g(t) \mid g(\mu_t), \sigma_t) = \frac{1}{\sigma_t\sqrt{2\pi}} e^{-(g(t)-g(\mu_t))^2/2\sigma_t^2} \tag{2.4}$$

where $g(t)$ is

$$g(t) = \mathrm{sgn}(t) \log_2\left(\frac{|t|}{\tau} + 1\right) \tag{2.5}$$

where sgn() is the sign function. Using this functional form for temporal frequency tuning allows for the function to be logarithmic at high temporal frequencies, but also be defined as zero-valued and continuous at zero temporal frequency. Here $\tau$ determines the temporal frequency at which the function transitions from linear to logarithmic, and $\mu_t$ and $\sigma_t$ are the temporal frequency preference and bandwidth, respectively.

The velocity-separable linear weighting function, $V_i$, is defined as follows:

$$V_i(d_i, s_i, t_i) = w_d(d_i) \cdot w_s(s_i) \cdot v_t(d_i, s_i, t_i) \tag{2.6}$$

where the velocity-separable temporal frequency function, $v_t$, is defined as a Gaussian, again linear at low frequencies and logarithmic at higher ones:

$$\begin{aligned} v_t\left(d,s,t\right) &= \mathcal{N}\left(g\left(t\right) \mid g\left(P\left(d,s\right)\right) \ , \ \sigma_t\right) \\ &= \frac{1}{\sigma_t\sqrt{2\pi}} e^{-(g(t)-g(P(d,s)))^2/2\sigma_t^2} \end{aligned} \tag{2.7}$$

The only difference between $w_t(t)$ (equation 2.4) and $v_t(d,s,t)$ (equation 2.8) is that in the latter, temporal frequency tuning is separable about the preferred speed plane $P(d,s)$:

$$P\left(d,s\right) = s\frac{\mu_t}{\mu_s} cos\left(d-\mu_d\right) \tag{2.8}$$

Since the components of the plaid stimuli are always 120° apart, the component orientations are too far apart for cross-orientation effects in V1 to significantly modulate responses. Additionally, since component contrasts are all 50%, we assume the effects of V1 normalization are negligible. Therefore, the MT neurons in this model sum the V1 responses to each stimulus component separately. In the following equation, the $F_i$ and $V_i$ terms represent the frequency- and velocity-separable MT linear weightings of V1 responses. The nonlinear response of the MT neuron is the weighted V1 responses passed through the MT nonlinearity, to yield the predicted firing rates for the frequency- and velocity-based models ($R_f$ and $R_v$, respectively):

$$R_f\left(d_i, s_i, t_i, t_{max}\right) = \alpha_0 + \frac{\alpha_1 n^{(1-2\beta)/3}\left(\sum_i^n F_i\left(d_i, s_i, t_i\right)\right)^\beta}{\alpha_2 + \sum_i^n N_i\left(t_i, t_{max}\right)} \tag{2.9}$$

$$R_v\left(d_i, s_i, t_i, t_{max}\right) = \alpha_0 + \frac{\alpha_1 n^{(1-2\beta)/3}\left(\sum_i^n V_i\left(d_i, s_i, t_i\right)\right)^\beta}{\alpha_2 + \sum_i^n N_i\left(t_i, t_{max}\right)}$$

The two MT nonlinearities, $R_f$ and $R_v$, are identical except for their linear weighting functions ($F_i$ and $V_i$, respectively). The $\alpha_0$ and $\alpha_1$ parameters represent the spontaneous and maximum discharge rates of the cell.

The MT nonlinearity consists of an exponentiation and divisive normalization, conceptually following prior versions of the cascade model [155, 141]. In the original Simoncelli & Heeger (1998) cascade model, the MT normalization stage was implemented by simulating a population of MT neurons. Simulating an entire MT population in the context of fitting the cascade model to data would be prohibitively computationally intensive. Therefore, we approximated the effects of tuned normalization in MT with a simple functional form, based on the assumption that a normalization in MT would be strongest at lowest temporal frequencies.

To understand the rationale for this assumption, let us consider the distribution of overlap of tuning for a population of MT neurons which fills frequency space. We will examine this separately for three types of selectivity in the context of the separable models: component selectivity, frequency-based pattern selectivity, and velocity-based pattern selectivity. Component neurons are simulated as having narrow tuning. Therefore, a population of component-selective neurons will have overlap of tuning only among neurons with adjacent spatiotemporal tuning preferences. As a consequence, the tuning overlap will be distributed evenly across frequency space.

Frequency-based pattern selective neurons have broad direction tuning, so their overlap will occur most strongly in direction. The overlap, however, will be separable in spatial and temporal frequency, so for any subpopulation with the same spatial and temporal frequency tuning at all directions, the overlap will be confined to a donut-shaped region centered on those spatial and temporal frequencies. Since we assume the population of frequency-based pattern selective neurons are evenly distributed all preferred spatial and temporal frequencies, the tuning overlap will also be evenly distributed.

Finally, velocity-based pattern selective MT neurons will have strong overlap at zero and low temporal frequencies, at all directions. This is due to the fact that they are organized along tilted planes which pass through the origin. For any given direction, neurons tuned to high and low speeds will overlap at zero and low temporal frequencies.

Intermediate cells under both models will feature the same overlap as their pattern-selective counterparts, but to a lesser extent. Since the overlap at low temporal frequencies is the only overlap of tuning across all cell types, it is this configuration which we used to approximate tuned normalization.

The effects of divisive normalization for both models are approximated by $N_i$ in equation 2.9. Suppression for the $i$th grating is modeled by a power function dependent on temporal frequency. The pool is maximally active at zero temporal frequency, with a value of 1, and minimally active at the cell's (experimentally determined) preferred temporal frequency, $t_{max}$, with a value of $\gamma_0$. The exponent of the power function is $\gamma_1$:

$$N_i\left(t_i, t_{max}\right) = \left(1 - \gamma_0\right)\left(1 - \frac{t_i}{t_{max}}\right)^{\gamma_1} + \gamma_0 \tag{2.10}$$

We chose this functional form because it is a simple parameterization that: (1) ensures there is no suppression at the preferred temporal frequency, (2) can be completely disabled by setting $\gamma_0 = 1$, and (3) can be sub-linear, linear, or super-linear.

The relative gains of responses to grating and plaid are controlled by the $n^{(1-2\beta)/3}$ term in the numerator in equation (2.9), where $n$ is the number of components in the stimulus.

Since there was only one component present at any given moment during the single-component study and all gratings were presented at full contrast, the full nonlinearity in the model is unconstrained. This is because the exponent in the MT stage governs how plaid components interact to create pattern tuning. For single gratings, the MT exponent $\beta$ forms a degeneracy in the model with all three separable tuning widths $\sigma_d$, $\sigma_s$, and $\sigma_t$ because they are exponents within $F_i$ and $V_i$ (see equations 2.2, 2.3, 2.4, and 2.8). Therefore, a reduced version of the model is fit with only a fixed quadratic nonlinearity:

$$\begin{aligned} R'_f\left(d_i, s_i, t_i\right) &= \alpha_0 + \alpha_1 F_i\left(d_i, s_i, t_i\right)^2 \\ R'_v\left(d_i, s_i, t_i\right) &= \alpha_0 + \alpha_1 V_i\left(d_i, s_i, t_i\right)^2 \end{aligned} \tag{2.11}$$

### 2.2.6 Estimating model parameters for individual cells

In total, the model has 9 free parameters for the single-grating study and 10 for the two-component study. For the former, they are: the direction preference and bandwidth ($\mu_d$ and $\sigma_d$), spatial frequency preference and bandwidth ($\mu_s$ and $\sigma_s$), temporal frequency preference, bandwidth, and log-linear transition ($\mu_t$, $\sigma_t$, and $\tau$), and the spontaneous and maximum firing rate ($\alpha_0$ and $\alpha_1$). For the latter

experiment, $\mu_s$, $\sigma_s$, and $\mu_t$ are unconstrained by the data and are therefore held fixed at experimentally determined values, but the exponent ($\beta$), semi-saturation constant ($\alpha_2$), and normalization parameters ($\gamma_0$ and $\gamma_1$) are free. To avoid model fits producing spuriously wide temporal frequency tuning, we included temporal frequency tuning data in the fitting of the two-component dataset. That temporal frequency tuning data, along with the two-component stimuli which sample different directions, constrain $\mu_d$, $\sigma_d$, and $\sigma_t$. In each study, the frequency- and velocity-based models have the same parameters, and only differ in the coordinate system of their linear weighting functions, $F_i$ and $V_i$ (see equations (2.1) and (2.6)).

For each cell, we optimized the model parameters by minimizing the negative log-likelihood (NLL) over the observed data, assuming spike counts arise from a modulated Poisson model. An additional parameter, $\sigma_G$, describes across-trial fluctuations in neural response gain [60] and was optimized to the data independently from the frequency- and velocity-based models and held constant during model fitting. We performed the optimization in successive steps, using optimal values from one step as initialization values for the next. First, we fit $\tau$, then added the rest of the MT linear weighting parameters, and then in the case of the two-component experiment, the MT parameters controlling the MT nonlinearity. For the two-component experiment, we also included data from a temporal frequency tuning experiment collected immediately prior to constrain the parameter search to realistic temporal frequency bandwidth values. We used a simplex algorithm (the Matlab function 'fmincon') to do the (constrained) parameter search. In order to avoid overfitting and obtain estimates of parameter stability (i.e., the error bars in figures 2.6(a,b) and 2.7), we fit the model on 100 bootstraps of the data. Bootstrapping was done on a per stimulus-condition basis–that is, trials within

each stimulus condition were sampled with replacement. This was to ensure that there were no stimulus conditions without data.

## 2.3 Results

### 2.3.1 A separable model of direction selectivity in the Fourier domain

Any set of moving images can be completely described as a combination of the vertical and horizontal spatial frequencies within each image and the temporal frequencies present across images. The presence of frequencies within these three dimensions can be measured by applying the Fourier transform. Together, they constitute the "Fourier domain," which can alternatively be represented in polar coordinates as orientation, spatial frequency, and temporal frequency. Thus, the Fourier domain is a natural way to represent the sets of moving images for which individual V1 and MT neurons are selective.

A single point in the Fourier domain represents, in the image domain, a drifting sinusoid with a unique orientation, spatial frequency, and temporal frequency (figure 2.1(a)). V1 neurons tend to be sharply selective for only a small set of frequencies near the preferred stimulus, so V1 selectivity is best approximated by a ball in the Fourier domain.

A tilted plane in the Fourier domain, going through the origin, represents in the image domain a set of dots moving at the same velocity (figure 2.1(b)). Furthermore, any rigidly moving object or texture can be represented by all of its spatial frequencies projected onto a tilted plane, the slope of which is equal to the object's velocity. A previous model of MT computation proposed that MT neurons are specialized for analyzing rigid motion, and therefore are organized along

just such a plane with slope equal to a preferred velocity (figure 2.1(c), "velocity model", in red) [155, 177]. Alternatively, MT direction selectivity could treat spatial and temporal frequency independently, leading to organization along a plane with constant temporal frequency (figure 2.1(c), "frequency model", in blue).

To examine MT receptive field organization in the Fourier domain, we fit two modified versions of a previously published model of MT direction selectivity to the responses of individual neurons. Both models have the same structure: they have two stages, each with a linear weighting followed by a nonlinearity (figure 2.1(d)). The first (V1) stage consists of narrowly-tuned direction-selective complex cells, simulated with a linear weighting of a narrow band of frequencies, followed by a squaring point nonlinearity. The second (MT) stage also contains linear weighting on its V1 inputs, followed by squaring. This second linear weighting in the MT stage represents the computation crucial to computing pattern motion. It is parameterized by the separable product of three tuning curves. The first two, direction and spatial frequency tuning, are common to both models. In the frequency model, the third separable tuning function is temporal frequency tuning, independent of the other two dimensions. In the velocity model, temporal frequency tuning co-varies with spatial frequency tuning such that their ratio is held constant at the preferred velocity of the neuron. This difference in temporal frequency tuning parameterization is the only difference between the two models.

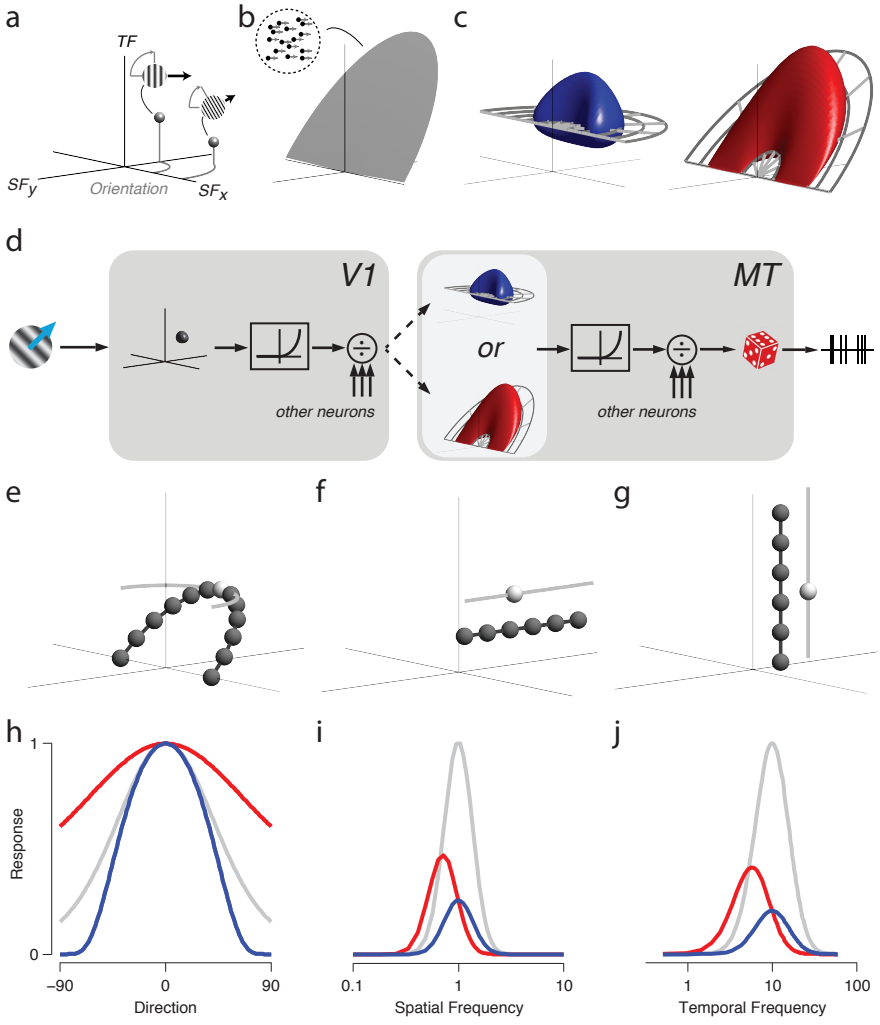


Figure 2.1 (*continued on next page*)

Figure 2.1 (*previous page*): Frequency and velocity model predictions in Fourier space.

(a) The Fourier domain is a three dimensional representation of moving images, with two spatial frequency axes and one temporal frequency axis. It can alternatively be expressed in terms of orientation, spatial frequency, and temporal frequency. A single point in the Fourier domain represents a single, unique drifting grating. (b) A plane in the Fourier domain corresponds to a set of dots drifting together with identical direction and speed. (c) Two possible hypotheses for MT selectivity in the Fourier domain. The velocity-based model (red) predicts spatial and temporal frequency tuning are jointly determined along a tilted, preferred velocity plane. The frequency-based model (blue) predicts spatial and temporal frequency tuning are independent. (d) The separable model. A stimulus is passed through a narrowly tuned V1 linear weighting, then squared. V1 output is then passed to the MT neuron, which applies either a frequency- or velocity-based linear weighting, then squares the output. Finally, spikes are generated by a modulated Poisson process. (e-g) Stimuli for three "classical" tuning experiments (light gray lines) containing the optimal stimulus (light gray ball) and suboptimal stimuli (dark gray): constant-frequency direction and constant-velocity direction tuning (e), optimal and low-temporal frequency spatial frequency tuning (f), and optimal and low-spatial frequency temporal frequency tuning (g). (h-j) The two models are matched to give identical predictions for "classical" stimuli (light gray). The frequency model (blue) has narrower constant-velocity direction tuning (h) and unchanging spatial (i) and temporal frequency (j) tuning preferences. The peaks of the velocity model's (red) spatial and temporal frequency tuning for suboptimal stimuli are shifted away from the peak tuning for optimal stimuli. Specifically, the peaks will be lower than optimal when tuning is measured at frequencies lower than optimal ((i) and (j)), and higher for measurements done at higher frequencies.

### 2.3.2 Single gratings do not differentiate model predictions

To distinguish the two different models, we presented a sequence of full contrast sinusoidal gratings at different orientations and spatial and temporal frequencies designed to sample the Fourier domain as efficiently and meaningfully as possible. The sampling was tailored to each individual neuron based on its tuning preferences and consisted of three tuning experiments containing the optimal grating, as well

as twelve suboptimal tuning experiments chosen to maximally distinguish the two models (see methods for details).

The two example models (figure 2.1(c)) have the same "classical" tuning properties: direction, spatial frequency, and temporal frequency tuning stimuli (stimuli shown in figure 2.1(e-g), light gray line) centered around the cell's preferred frequency (light gray ball) produce the same responses (figure 2.1(h-j), gray). Suboptimal tuning stimuli (figure 2.1(e-g), dark gray) yield quite different predictions (red and blue, figure 2.1(h-j)). Constant velocity direction tuning (left) is narrower for the frequency model (blue). Spatial and temporal frequency tuning preferences do not change between the optimal and suboptimal tuning experiments. The velocity model (red) predicts that the preferred spatial frequency, indicated by the peak of the tuning curve, increases when stimuli are presented at a higher temporal frequency (figure 2.1(i)), compared to the preference when measured at the optimal temporal frequency (gray). This same increase in tuning preference holds true when evaluating temporal frequency tuning evaluated at a higher than optimal spatial frequency (figure 2.1(j)).

We fit the frequency and velocity models to data from all 17 tuning experiments simultaneously. We used the optimized models to predict responses to each trial and generated predicted tuning curves for each tuning experiment. Figure 2.2(a-d) shows tuning curves for two example MT component cells 2.2(a, b) and two MT pattern cells 2.2(c, d). For clarity, only four of the 17 tuning curves are shown, corresponding to the four which exhibited the greatest difference between the frequency and velocity model predictions on average across the population. Overall, both models can account for the tuning curve shapes and, for most stimulus conditions, changes in relative response gain.

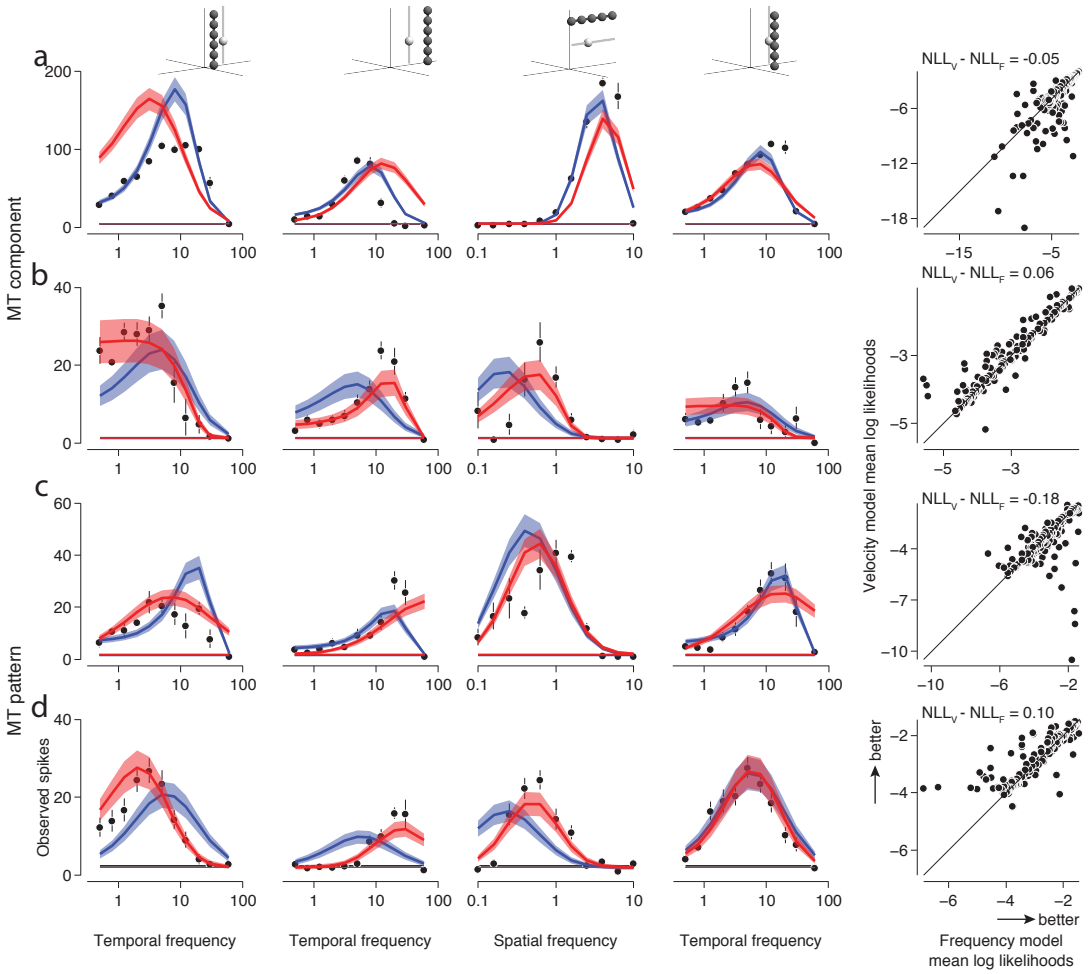


Figure 2.2: Comparison of actual and model-predicted responses to gratings for four example cells.
(a,b) Two example component cells, one better fit by the frequency model (a) and one better fit by the velocity model (b). (c,d) Two example pattern cells, one better fit by the frequency model (c) and one better fit by the velocity model (d). Measured spike rate mean and s.e.m. are shown in black. Velocity model predicted spike rates are shown in red, frequency model predictions in blue. Means are indicated by the dark lines, s.e.m. by the lighter shaded areas. In the rightmost column, each point in the scatter plots represents how well the frequency and velocity models predict the mean firing rate for a given stimulus condition ($NLL_F$ and $NLL_V$, respectively). Goodness of fit is expressed in terms of log likelihood under the modulated Poisson process, where values closer to zero indicate a better fit.

Each point in the scatter plots in figure 2.2 corresponds to the goodness of fit of the two models for a single stimulus condition from all 17 tuning experiments. For example, tuning peaks in figure 2.2(a) appear to remain constant, consistent with the frequency model prediction. In figure 2.2(b) and 2.2(d), tuning peaks shift in a manner consistent with the velocity prediction—peaks are higher when measured at higher frequencies (second and third columns). The peaks shift for some, but not all tuning experiments in figure 2.2(c). This range of behavior was observed across the population. In general, V1 cells were slightly better fit by the frequency model (the mean difference in negative log-likelihoods between the velocity and frequency models was $-0.08$, $P = 0.016$ Wilcoxon signed rank test). Some MT neurons were clearly better fit by one model or the other, but overall, neither model was significantly better.

Neither model consistently performed better than the other, despite the fact that spatiotemporal frequency space was sampled with single gratings that should have distinguished the two separable models. This was a surprising finding because the tuning of speed to gratings had previously been reported [119, 152, 124, 125], and the responses overall are well fit by a simple model. This led us to wonder whether examining the subset of gratings presented at the preferred direction would reveal velocity separability. We simply asked whether temporal frequency preference did or did not change when changing grating spatial frequency (figure 2.3a). Normalizing spatial and temporal frequency to the optimal spatiotemporal frequency for each neuron, we fit lines, constrained to go through that optimal value, to the measured preferences at non-optimal spatial frequencies. Most MT neurons (red) had a non-zero slope ($0.36 \pm 0.03$ s.e.m., red shaded area), consistent with the predictions of the velocity-based model (figure 2.3c). However,

V1 neurons (blue) had slopes close to zero ($0.06 \pm 0.04$ s.e.m., blue shaded area), consistent with the frequency-based model. Performing the same analysis for the spatial frequency preferences as a function of stimulus temporal frequency (figure 2.3b,d) showed even higher slopes in MT ($0.50 \pm 0.07$ s.e.m.) and more variable, but similar on average, slopes in V1 ($-0.02 \pm 0.26$ s.e.m.).

When confining the analysis to stimuli at the preferred direction, MT neurons appear to be velocity-based and V1 neurons appear to be frequency-based. When including stimuli at non-preferred directions, neither model better explains the data in either area. We wondered whether single grating stimuli adequately constrain the models. The primary feature that distinguishes direction selectivity between V1 and MT is selectivity to pattern motion, which cannot be fully assessed using single gratings. In other words, single gratings do not fully exercise MT neurons' nonlinearities. To address this, we designed a second set of stimuli to characterize pattern motion selectivity in both constant frequency and constant velocity coordinates.

### 2.3.3 Compound stimuli reveal velocity-based organization in MT

We ran a second study, in which gratings and 120° plaids were presented at a given neuron's optimal spatial frequency and its optimal temporal frequency or optimal velocity (figure 2.4(a,b)). These stimuli can be equivalently described as gratings and 120° plaids drifting either along the circular mean component direction orthogonal to their orientation(s) (constant frequency, figure 2.4(a)) or in the preferred direction of the cell (constant velocity, figure 2.4(b)).

The two models required additional nonlinear elements to be able to simultaneously account for MT neural responses to all four stimulus conditions. First,

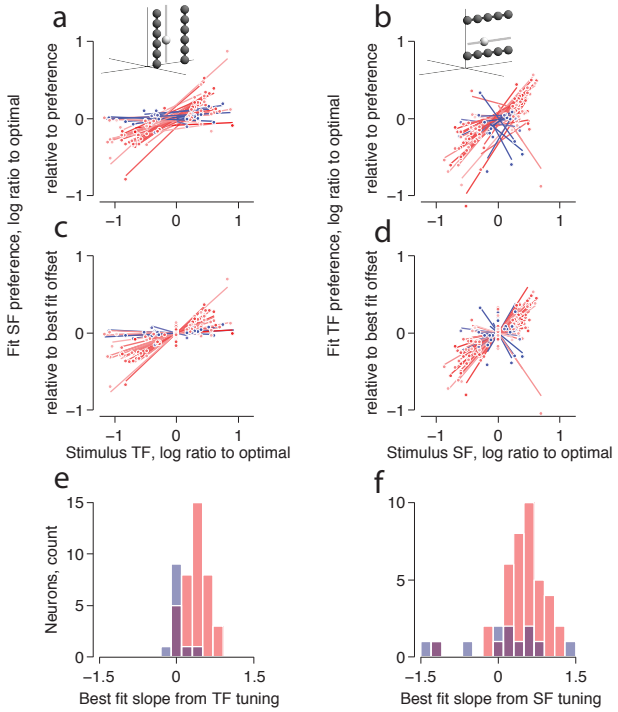


Figure 2.3: For single gratings moving in the optimal direction, V1 is frequency-based and MT is velocity-based.
(a) Data from temporal frequency tuning curves at optimal and non-optimal spatial frequencies. The abscissa represent the log10 ratio of the spatial frequencies of the TF tuning curves and the preferred spatial frequency. The ordinate represents the best fit temporal frequency for each tuning curve. Each point is for a suboptimal tuning curve for one cell. Each line is the best fit line to the data, including the optimal spatial and temporal frequency (the origin), to which all other points are aligned. Lines and points are shaded by the pattern index corresponding to each individual cell. Red corresponds to MT neurons, with darker shades corresponding to higher pattern index, and blue corresponds to V1 neurons, with darker shades corresponding to lower pattern index. (b) Same as (a), but based on spatial frequency tuning curves at optimal and suboptimal temporal frequencies. (c) Same as (a), but data and best fit lines are aligned to the offsets of the best fit lines. (d) Same as (c), but based on spatial frequency tuning curves at optimal and suboptimal temporal frequencies. (e) Histograms of the slopes of the best fit lines in (a) and (c), for V1 (blue) and MT (red). (f) Same as (e), but based on spatial frequency tuning curves at optimal and suboptimal temporal frequencies.

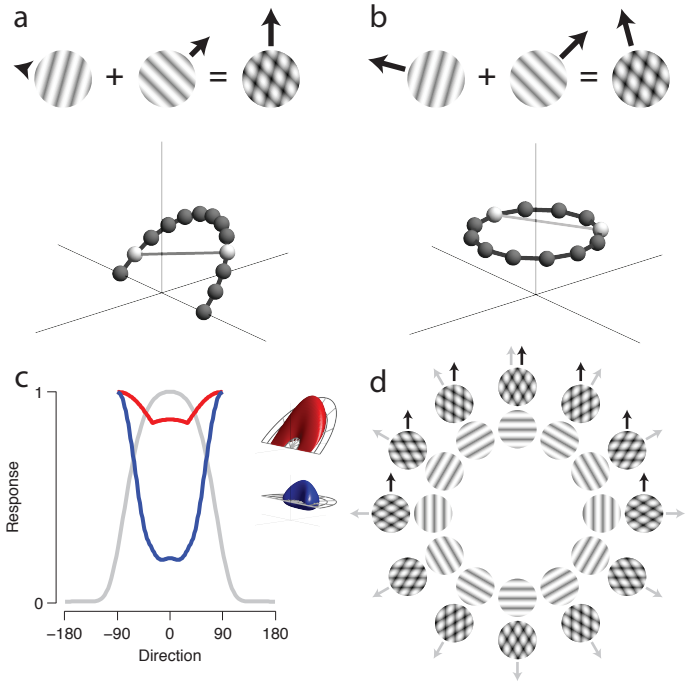


Figure 2.4: Two-component "planar plaid" experiment design.
Constant-velocity and constant-frequency direction tuning experiments were done with gratings (see figure 2.1(e)) and plaids (a,b). Constant-velocity plaids (a) were constructed by superimposing two gratings 120° apart and drifting at a temporal frequency determined by the optimal velocity plane. Constant-frequency plaids (b) were two gratings 120° apart superimposed and drifting at the optimal temporal frequency. The example plaids shown contain the same orientations, but have different perceived drift directions. (c) For the two models matched in constant-frequency plaid direction tuning (light gray), the velocity model (red) predicts a high response rate to all constant-velocity plaids. The frequency model (blue) is more narrowly tuned. (d) Constant-frequency plaids drift (gray arrows) in the direction of the circular mean of the component orientations. Constant-velocity plaids drift (black arrows) in the preferred direction of the neuron.

the exponent on the responses of MT linear weights became a free parameter. An additional divisive suppression step followed, which included both stimulus-independent and temporal frequency-dependent suppression terms (see Methods). The spatial frequency preference and bandwidth and temporal frequency preference parameters were unconstrained by this data, and therefore fixed to values determined in tuning measurements done prior. In total, there was one additional free parameter fit compared to the previous dataset.

Qualitatively, the models predict that direction tuning bandwidth should be wider when the coordinate system of the model matches that of the stimuli. Figure A.1 shows measured and predicted responses for four example cells. Two features of the data stand out. First, constant frequency and constant velocity direction tuning curves to gratings are nearly indistinguishable for all cells. Second, the pattern selective MT neuron (figure A.1(d)) exhibits much wider direction tuning bandwidth for constant velocity plaids as opposed to constant frequency plaids, while the other cells show more similar tuning. For all cells, both models capture grating data well. However, the frequency model cannot account for the pattern selective neuron's responses to both types of plaids simultaneously (fig A.1(d)). The best it can do is pick a compromise direction tuning bandwidth that is too wide for constant frequency plaids and too narrow for constant velocity plaids. The velocity model, on the other hand, is able to account for all the data simultaneously, including the different plaid tuning bandwidths. This pattern cell is the only one of the four example cells that has large differences in the frequency- and velocity-based model predictions. It is clearly better fit by the velocity model, as can be seen in both the predicted tuning curves and scatter plots.

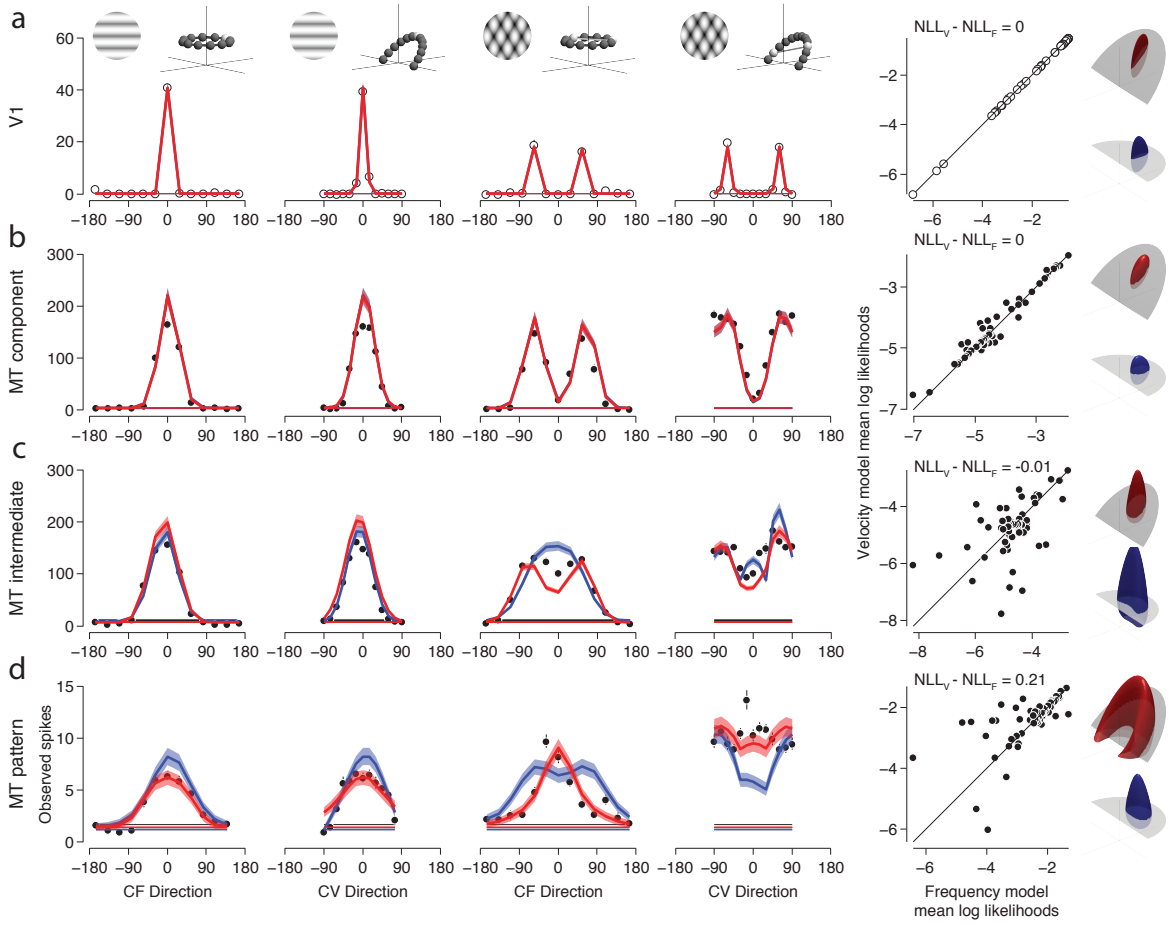


Figure 2.5: Comparison of actual and model-predicted responses to gratings and plaids for four example cells.

Neurons are ordered, from top to bottom, with increasing pattern index: (a) an example V1 component-selective neuron, (b) an MT component neuron, (c) an MT intermediate neuron, and (d) an MT pattern-selective neuron. Scatter plots of goodness of fit per stimulus condition. See figure 2.2 details. Renderings of the frequency and velocity model linear weightings for each example cell (right-most column). Differences in between the two models become more apparent with increasing pattern selectivity.

### 2.3.4 Model comparison across the population

In the first study, both models described responses to single gratings well. For most cells in the population, single gratings yielded spatial and temporal tuning preferences that changed subtly across experiments and in ways that were not exclusively consistent with either model. We assessed overall fit quality on a cell-by-cell basis by normalizing the log likelihoods of the models to null and oracle models. The null model assumes the cell has two possible response rates: one when a stimulus is present and another when there is no stimulus. These are fixed to the measured-mean spontaneous and stimulus-driven response rates, respectively. The oracle model serves as an upper bound for the models' performance. It is a lookup table that uses the measured mean responses to each stimulus condition to predict the neuron's response to any future repeat of that stimulus. "Velocity superiority" is the difference of the normalized log likelihoods of the velocity-based model and the frequency-based model (figure 2.6(a,b)). For the single grating study, there is no significant correlation between the difference of the two models' normalized log likelihoods for each cell and their pattern indices (figure 2.6(a)), for all cells (Pearson's $r = 0.02$, $P = 0.91$) or MT alone ($r = 0.03$, $P = 0.85$). There is, however, a significant negative correlation for V1 ($r = -0.76$, $P = 0.004$). The frequency-based model, on average, fit V1 data better ($P = 0.016$ Wilcoxon signed rank test on "velocity superiority"). Neither model was better for MT neurons ($P = 0.16$).

This trend is stable across the population, as shown by the running mean lines (dark gray for MT, light gray for V1). The frequency-based model has a slight advantage for V1 cells (62.4±5.6% of tuning curves were better fit by the frequency

model), but not MT ($53.4 \pm 2.6\%$ of tuning curves were better fit by the frequency model). There was no relationship between pattern index and the number of tuning curves per cell better fit by one model or the other ($P = 0.57$, Student's t-test).

The population trend for predictions to the compound stimulus dataset is quite different (figure 2.6(b)). First, there is little difference between the two models for component cells, so neither model fits better. As the tuning curves (figure A.1, left four columns) linear weighting renderings illustrate (figure A.1, rightmost column), the models cannot be distinguished by narrowly tuned cells. As tuning bandwidth increases in the intermediate cells, the models begin to diverge slightly, but there is still no significant difference between the two models' predictions ($P = 0.76$, Wilcoxon signed rank test). Nearly every pattern cell, however, is better fit by the velocity model ($P < 0.00001$, Wilcoxon signed rank test).

As a further evaluation of the models' ability to account for the responses to compound stimuli, we asked how well they could predict each cell's pattern selectivity. The velocity model accounts for the full range of pattern selectivity across the population (figure 2.6(c), Pearson's $r = 0.80$). The frequency model, however, fails to produce any cells with pattern tuning (figure 2.6(d), Pearson's $r = 0.70$), due to the compromises it must make when fitting both constant frequency and constant velocity plaid responses simultaneously.

How does the velocity-separable model account for pattern selectivity? First, direction tuning bandwidth is strongly correlated with pattern index (Pearson's $r = 0.75$, figure 2.7(a)) and the exponent in the nonlinearity (Pearson's $r = 0.71$, figure 2.7(b)). There is also a weak correlation between pattern index and the exponent in the temporal frequency-dependent suppression term (Pearson's $r = 0.33$, $P = 0.0001$, figure 2.7(c)).

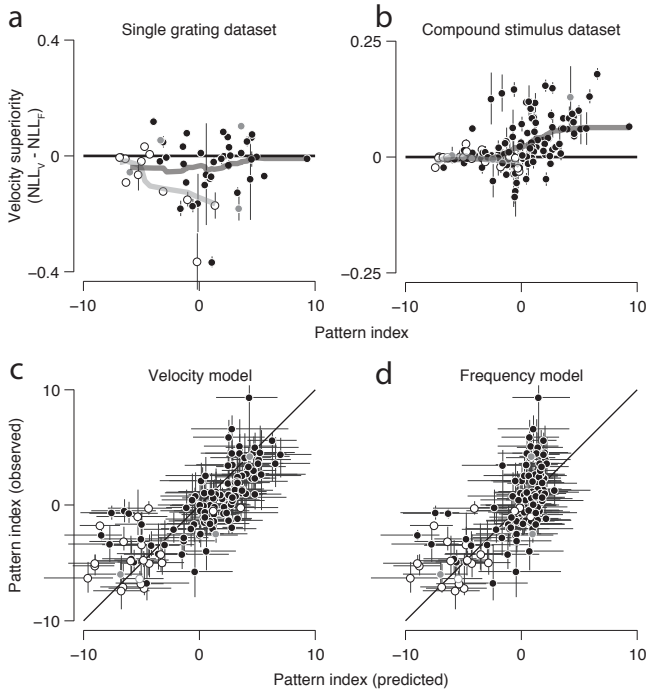


Figure 2.6: Compound stimuli reveal velocity-based organization for pattern cells. (a,b) Velocity superiority, or the difference of normalized log likelihoods between the velocity and frequency models, per cell as a function of pattern index. V1 cells ($n = 12$) appear as open circles, MT closed ($n = 39$). Example cells featured in figures 2.2 and A.1 are highlighted in gray. Light and dark lines indicate the running mean, with a window of $\pm 1/3$ of cells in each population. Error bars indicate $\pm 1$ standard deviation, calculated from model fits to bootstrapped data. On average, neither model better explains the single grating MT data (a) for any class of cells. The frequency model explains the V1 single grating data better ($n = 20$). Pattern cell responses to the compound stimulus dataset (b) are clearly better explained by the velocity model ($n = 111$). Observed and predicted pattern indices for each cell, derived from the compound stimulus dataset, for the velocity model (c) and frequency model (d). The velocity model can account for pattern index across all cell types, whereas the frequency model fails to predict the pattern selectivity (Pearson's $r = -0.01$, $P = 0.97$) of neurons classified as pattern-selective based on measured repsonses. Error bars indicate $\pm 1$ standard error, generated from pattern indices calculated by bootstrapping measured and predicted spike trains.

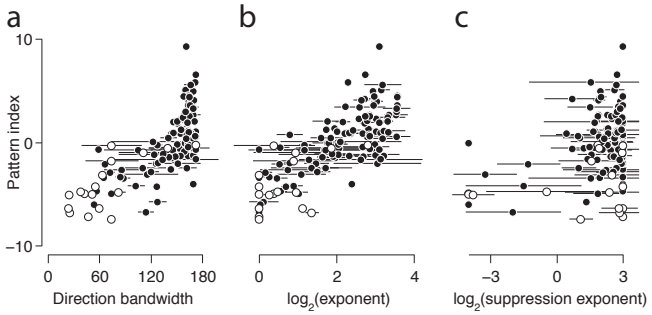


Figure 2.7: Relationship between velocity-separable model parameters and pattern index.
Pattern index is strongly correlated with direction tuning bandwidth (a) and the log of the MT nonlinearity exponent (b). (c) Pattern index is weakly correlated with the log of the exponent in the temporal frequency-dependent suppression term.

Taken together, the two datasets reveal important insights about MT computation. First, simple stimuli reveal a velocity-based organization in spatial and temporal frequency, but are not sufficient to reveal the nonlinear behaviors that distinguish direction selectivity observed in MT from that observed in V1. Second, compound stimuli reveal that MT receptive fields are organized along a preferred velocity plane.

## 2.4 Discussion

The representation of motion changes as one moves up the hierarchy of visual areas. V1 complex cells are narrowly tuned detectors of moving edges. MT neurons are jointly selective for the direction and speed of moving complex visual patterns. This makes them ideal detectors of the velocity of rigidly moving objects, while being invariant to the texture or shape of the object. A previous theoretical model [155] predicted that this conversion from independent to joint selectivity is the crucial computation that gives rise to motion direction selectivity and pattern invariance

in MT. For the first time, we directly verified this prediction by fitting modified versions of that model to V1 and MT responses to drifting gratings and plaids. The models were parameterized in the same manner, with the only difference being that one represented direction and speed jointly, while the other represented direction and speed independently.

Here we have shown that compound stimuli (such as plaids) are necessary to reveal the structure of MT receptive fields. By presenting plaids and gratings on frequency- and velocity-based rings (figure 2.4), pattern-selective neural responses are unambiguously velocity-based (figure 2.6). Due to their narrower tuning, it is not surprising that component and intermediate cells are not revealed to be purely velocity-based, since the predicted receptive field structures of the velocity- and frequency-based models become increasingly similar as tuning narrows. It was initially puzzling, however, that pattern cells showed no change in direction tuning bandwidth between frequency- and velocity-based gratings (figure A.1)—an apparent violation of the assumptions of both the frequency- and velocity-based models (figure 2.1). This led us to the conclusion that single gratings alone do not constrain the nonlinear behavior of MT neurons in the way that plaids do. Incorporating plaid stimuli allowed us to fit the model nonlinearity. We found its exponent to be strongly correlated with pattern index (figure 2.7).

A higher exponent, followed by normalization, allows for shallower tuning to single components and sharpened tuning for conjunctions of components. In the case of pattern cells, these conjunctions are components consistent with a preferred velocity. This could explain why pattern cell responses to single gratings, taken in their entirety, do not favor either model (figures 2.2 and 2.6). Their selectivity to single gratings appears more difficult to precisely resolve in the Fourier domain.

Limiting the analysis to the tuning curves measured at the preferred direction (figure 2.3), however, reveals a velocity-based representation for the population of MT neurons. This apparent inconsistency in predicted receptive field structure could arise if the receptive fields are not perfectly separable in either frequency-based or velocity-based coordinates—one of the simplifying assumptions we made to be able to fit these models. Alternatively, the inconsistency could be the result of the MT neuron selectively weighting a limited set of inputs that are strongly overlapping on a preferred velocity plane without perfectly aligning to it. As a result, MT receptive fields could display a diversity of volume shapes in the Fourier domain when probed with single gratings; when multiple gratings are present, however, the nonlinearity amplifies overlapping inputs, which are only on the preferred velocity plane.

### 2.4.1 Relationship to previous models

The separable models we developed and tested are modified from previous models [155, 141]. The original cascade model [155], while not directly fit to data from a population of single unit recordings, attempted to comprehensively simulate MT computation. The model simulated an entire population of V1 neurons that linearly filtered incoming images and passed their outputs through rectifying nonlinearities and divisive normalization. The second (MT) stage of the model linearly weighted these incoming signals from V1 (along a plane) and again passed the responses through a rectifying nonlinearity and divisive normalization. An entire population of MT neurons was simulated for the final normalization step.

In order to fit this model to data, it was not feasible to simulate entire populations of V1 and MT neurons, so Rust et al. (2006) simplified the cascade model

to focus on direction selectivity and pattern integration. First, stimuli were confined to a frequency-based ring. Likewise, the model itself was effectively one-dimensional—it focused solely on direction tuning to compound stimuli. The V1 stage was simplified to evaluate tuning based on a population of direction tuning functions, the responses of which could be further modulated by tuned and untuned normalization. The MT stage then linearly weighted the V1 inputs and passed them through a nonlinearity. Rust et al. (2006) showed that opponent suppression and tuned normalization shaped pattern selectivity.

We sought to characterize MT receptive field structure in all three dimensions of the Fourier domain, but in order to do so, we had to further streamline the model. The first study contained stimuli sampling all three dimensions, while the second had stimuli varying in just direction and temporal frequency. Since we presented only gratings or 120° plaids, and kept all grating components at a constant contrast, it was not necessary to simulate V1 tuned normalization. Therefore, our model did not explicitly simulate the V1 stage; rather, it evaluated tuning directly based on the separable product of tuning curves along three dimensions in the Fourier domain. Since all three tuning curves are exponential functions, the separable tuning volume and exponent approximately accounts for both the linear weighting stages and exponential nonlinearities of V1 and MT. Temporal frequency-dependent divisive suppression was a final addition to the model necessary to simultaneously account for all direction tuning bandwidths in the second study. This type of suppression is an approximation of the effects of MT normalization (see methods for a detailed justificcation). Suppression for low temporal frequencies has been observed previously [99].

More recent studies [112, 78] have explored MT selectivity in all three dimen-

sions of the Fourier domain. Nishimoto & Gallant (2011) used "motion-enhanced" natural movies to visualize 3D spectral receptive fields for the first time. They fit a model in which stimuli were linearly filtered and passed through a compressive nonlinearity and divisive normalization. They visualized the MT neuron's linear weights on those outputs, showing excitation organized along a partial ring on the a plane that avoided low temporal frequencies. Suppression also appeared as partial rings off the preferred velocity plane. Inagaki et al. (2016) performed linear regression directly on the frequences of the stimulus, which was comprised of multiple superimposed gratings, each lasting for 237ms, beginning at 39ms offsets. They observed 3D partial ring receptive fields in two pattern cells and observed diffuse suppression off the preferred velocity plane. These provide indirect support for our use of suppression at low temporal frequencies.

Both studies used stimuli that make interpretation of the receptive field structures more difficult. The spectral content of "motion-enhanced" natural movies may not follow a Gaussian distribution, meaning a regression-based analysis will yield biased results. Neither of these studies directly confirmed that their models could produce pattern tuning, making the connection between the receptive field structure they observed and pattern selectivity harder to interpret. Our model is able to reproduce pattern tuning in both frequency- and velocity-based coordinates, while making slightly different predictions of receptive field structure. Pattern cells have excitation on a full ring on the preferred velocity plane, with partially overlapping suppression at low temporal frequencies.

In earlier attempts to fit a version of the cascade model to our data, we tried different configurations of the model and its free parameters. We found a partial excitatory ring with subtractive suppression is not sufficient to simultaneously

account for both the narrow and the wide tuning bandwidths we observed in response to constant frequency and velocity plaids, respectively. Incorporating a V1 normalization stage added more complexity but with very little gain in model performance.

We also tried models which included subtractive suppression, defined as a separable volume in 3D Fourier space. In one instance, the volume's shape was constrained in the same manner as the excitatory volume. In another instantiation of the model with subtractive suppression, the shape of suppression was constrained to be an exact copy of the excitatory volume, but rotated $180^{circ}$. In that case, only the relative gain of excitation and suppression was fit. For both model versions, subtractive suppression improved fits of both models to constant frequency plaids, but was not sufficient for either model to explain the constant velocity plaid data. Fitting the shape of a subtractive volume was also poorly constrained by the data. The model we presented here is a minimal, parameterized model that, when used with compound stimuli, exposes the core computation of MT direction selectivity.

We have shown that MT neurons perform a transformation on the stimulus representation from local oriented edge motion in V1 to the rigid motion of patterned objects. This change of basis represents a shift from sharper simpler selectivity to a more complex selectivity that is also more invariant. In general, as one moves up the hierarchy of a sensory system, invariance and complexity of selectivity increases [139]. How does the separable model account for this? It is constructed based on the idea of cascaded computation: the sequential repetition of similar computational structures that extract different stimulus features in each brain area. Our separable model implements this in the form of a linear-nonlinear (LN) cascade, in which inputs to each stage undergo a linear coordinate transformation, followed

by a nonlinear rescaling of the input distribution in the new coordinates. Both the linear and nonlinear components selectively emphasize and discard information. In the case of our separable model, the MT linear weights selectively emphasize any stimuli that have at least one component roughly consistent with a preferred velocity plane; the nonlinearity discards compound stimuli that do not have all components on the plane.

The cascaded LN framework is a powerful tool for testing targeted hypotheses of computation in individual sensory areas. In order to most effectively explore the space of possible models, stimuli must be complex enough to explore both the selectivity and invariance of the area in question, but simple enough to quantify and interpret without additional confounds or biases.

# Chapter 3

# A non-parametric model of MT selectivity

## 3.1 Introduction

We previously showed that MT receptive fields are organized along a preferred velocity plane in 3D frequency space. This conclusion depended on the assumption that MT receptive fields are separable along three frequency dimensions. We compared two different hypotheses: that the receptive fields are separable along either the temporal frequency or velocity axis.

Our stimuli were chosen to distinguish these two specific types of receptive field organization. Probing the receptive fields with a limited set of single gratings does reveal velocity separability (figure 2.3 and [124, 125]). However, model fits using the entire set of single gratings in our stimulus set did not reveal any distinction between the two separable models (figure 2.6) for any class of MT cells. This was despite the fact that, for any given neuron, each separable model could fit nearly all tuning curves well (figure 2.2).

This suggests that MT receptive fields, when probed with single gratings, are

not strictly separable along either of the coordinate systems tested. While presenting plaids revealed velocity separability in pattern-selective neurons, many other neurons were not distinguished by either model (figure 2.6). This could be a natural consequence of the fact the model predictions converge as tuning becomes narrower, or it could be another indication that the receptive fields are not perfectly separable.

Another factor that could influence receptive field structure is the role of suppression in the receptive field. Suppression in MT is poorly understood, and it is not clear what the ideal configuration of suppression should be in 3D frequency space.

We wondered, if allowed to take any arbitrary shape, what would MT receptive fields look like? Will they be organized along a plane? What form will suppression take?

Ideally, one would present every possible combination of gratings at all directions and spatial and temporal frequencies. Since single unit recording time is limited, typically to approximately one hour, there isn't enough time to fully sample all frequencies. Because stimuli in so much of the 3D space do not excite the neuron, it is difficult to record enough spikes during the experiment to gain sufficient statistical power to describe the neuron's selectivity.

To overcome this obstacle, others have used stimuli rich in frequency content, including natural movies [112] or hyperplaid stimuli limited to a sparsely sampled lattice [78], to characterize MT receptive fields in 3D frequency space. Both studies employing these stimuli had weaknesses, including a reliance on regularization to smooth recovered receptive fields and a dearth of recorded pattern-selective neurons. We sought to improve upon these methods to gain a more accurate

description of excitatory and suppressive receptive field structure in 3D and link 3D structure more directly to pattern selectivity.

We presented hyperplaid stimuli that were tailored to each cell recorded from and consisted of combinations of gratings continuously sampling 3D frequency space. To more efficiently sample the space and maximize the dynamic range of neural response, we concentrated the stimuli so that at least half of the hyperplaid components occurred near the neuron's preferred stimulus frequencies.

We show that single grating tuning is well captured by both a STA and a nonlinear model. The nonlinear model improves fit quality and gives cleaner weight estimates. Suppression is weaker in the nonlinear fits. Nonlinear neural behavior, such as pattern selectivity, is not well described by either the STA or the nonlinear model when trained on hyperplaids. Training the nonlinear model to the "planar plaid" dataset from the previous chapter, however, captures pattern selectivity well, but not responses to hyperplaid stimuli.

## 3.2 Methods

### 3.2.1 Recording procedures

We recorded from two anesthetized and paralyzed adult male macaque monkeys (*M. fascicularis*), as well as the same two awake and actively fixating adult male macaques (one *M. mulatta* and one *M. nemestrina*) referenced in the previous chapter. We used the same standard procedures for surgical preparation and single-unit recording as described in the previous chapter. The behavioral paradigm was the same, with one exception; stimuli lasted for 133ms in both preparations, rather than the 1,000ms and 250ms used in the anesthetized and awake preparations,

respectively. This was done to maximize the number of spatiotemporal frequencies we could present in a single experiment and is close to the 160ms duration of the hyperplaid stimulus used by Rust et al. (2006).

### 3.2.2 Visual stimulation

For each isolated unit, we presented vignetted sinusoidal grating stimuli to map each cell's receptive field and determine its preferred size. We then characterized neuronal tuning preferences and bandwidths for direction, spatial frequency, and drift rate. Finally, "hyperplaid" stimuli were presented in a window of the preferred size.

Each hyperplaid consisted of up to four simultaneous and superimposed sinusoidal gratings. The spatiotemporal frequency of each grating was drawn randomly from within a hollow cylinder in 3D frequency space. The cylinder spanned all directions, was bounded in spatial frequency between 0.1 and 10 cycles/degree, and was bounded in temporal frequency between 0.1 and 50 cycles/sec.

The stimuli were distributed unevenly so that preferred spatiotemporal frequencies would occur more frequently to ensure that the neuron would maintain a higher level of excitation above baseline than would be expected if it was responding to truly random stimuli. Up to two gratings were drawn from a wedge-shaped "excitatory" region near the preferred velocity plane, as determined from the single-grating "basic characterization" experiments performed immediately prior (figure 3.1(a)). The directions, spatial frequencies, and temporal frequencies at which the neural responses reached half the maximum response served as the bounds of the excitatory wedge. For neurons with the half-maximum response occurring between single grating stimulus samples, bound values were either linearly interpolated from

raw tuning curves or extrapolated using descriptive fits to the data (a von Mises function for direction tuning, a log-Gaussian for spatial frequency, and a difference of exponentials for temporal frequency tuning [66]). The other two gratings were drawn from the remaining "inhibitory" region inside the hollow cylinder excluding the excitatory wedge (figure 3.1(b)). Within the bounds of these two regions, samples were drawn uniformly across direction and uniformly on a base-2 log scale across spatial frequency and speed (figure 3.1(c-e)).

The contrast for each grating component was assigned pseudo-randomly, according to the following criteria. First, the maximum summed contrast for the excitatory and inhibitory gratings was chosen. The total excitatory contrast had a uniform probability between 15% and 65%. The total inhibitory contrast had a uniform probability between 15% and the total excitatory contrast. This ensured that the total contrast would not exceed 100%. Next, the total excitatory contrast was divided among the two excitatory gratings. The first grating could take, with uniform probability, zero contrast, the total excitatory contrast, or any fraction inbetween of the total excitatory contrast. The second grating's contrast was set to the remainder of the total excitatory contrast. The same process was used to assign the inhibitory grating contrasts. Lastly, to vary the number of components present during any given trial, some individual components' contrasts were randomly forced to be zero. This resulted in lower contrast stimuli with fewer components, or blanks. The following criteria determined how many components were shown: (1) there was a 10% chance neither excitatory component appeared, a 50% chance one excitatory contrast appeared, and a 40% chance that both appeared, and (2) a 33% chance that neither inhibitory contrasts appeared, a 22% chance that one inhibitory contrast appeared, and a 45% chance that both appeared. The

distributions of total hyperplaid contrast and the contrasts for each component are shown in figure 3.1(f-j).

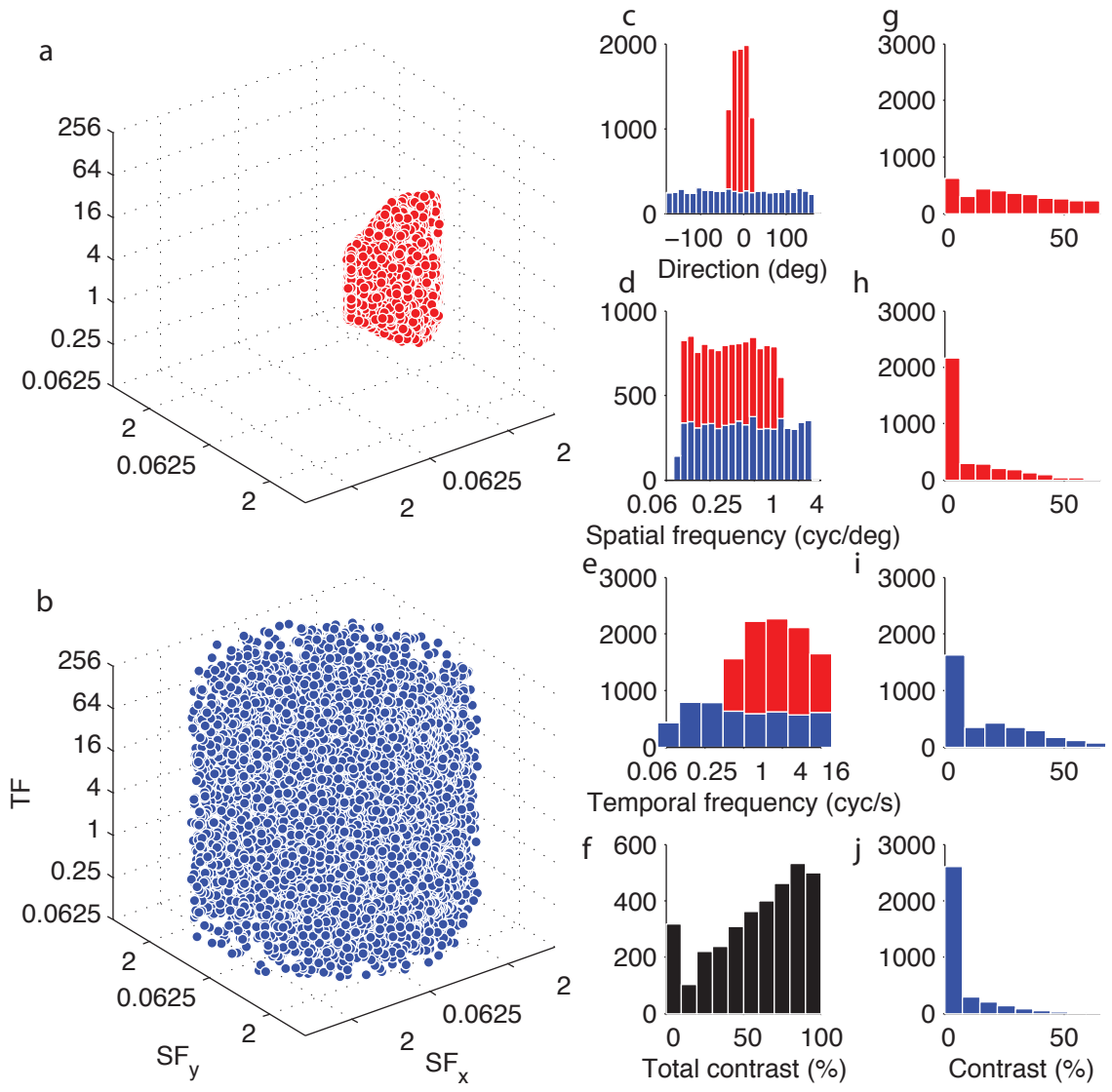


Figure 3.1 (*continued on next page*)

Figure 3.1 (*previous page*): The hyperplaid stimulus.

In (a) and (b), each point represents a single spatiotemporal frequency that can make up one of the (up to four) possible components in the hyperplaid stimulus. (a) Randomly sampling 10,000 hyperplaid stimuli and plotting the excitatory component spatiotemporal frequencies reveal the "excitatory wedge." The wedge is bounded by the directions, spatial frequencies, and speeds at which neural responses were at at least half their maximum response in preceding single grating experiments. (b) The "inhibitory" components are in a hollow cylinder with the excitatory wedge removed. (c-e) The distribution of directions (c), spatial frequencies (d), and temporal frequencies (e) for all excitatory (red) and inhibitory components (blue). (f) The distribution of the summed contrast of all components in a given hyperplaid. (g-j) The distributions of contrast for the four individual components in the hyperplaids.

We ran the stimulus with the goal of presenting 10,000 hyperplaid stimuli (133ms each in duration) for each neuron; however, recording isolation did not always permit this. Due to the difference in stimulus presentation timing between the anesthetized and awake preparations, 10,000 hyperplaid stimulus conditions corresponded to 22 minutes of anesthetized recording time and roughly 40 minutes of awake recording time.

### 3.2.3 Analysis of neural responses

Spikes counted in a 133ms window, as well as latency and pattern index, were calculated in the same way as described in the previous chapter. We also ran the velocity- and frequency-separable "planar plaid" experiment presented in the previous chapter on a subset of the neurons presented in this chapter ($n = 43$ out of 135). These data are a subset of the dataset presented in the previous chapter. The basic characterization data for these neurons are also the same.

### 3.2.4 The V1-MT cascade model

*The V1 stage*

We fit a cascade model of MT motion computation similar in architecture to the one presented in the previous chapter, but with crucial differences. Because the hyperplaid stimuli had instances where components with similar orientations appeared simultaneously, a V1 stage had to be explicitly simulated to account for effects of normalization in V1. Each V1 complex neuron was simulated as the squared sum of raised cosine filters in phase quadrature. The cosine filter response function $f$ is

$$f\left(x \,|\, \bar{x}, \hat{x}\right) = \mathbf{1}_{[\bar{x}-\hat{x}, \bar{x}+\hat{x}]}(x) \cos\left(\frac{\pi\left(x-\bar{x}\right)}{2\hat{x}}\right) \tag{3.1}$$

where $\bar{x}$ and $\hat{x}$ are the filter preference and bandwidth, respectively, and $\mathbf{1}_{\{-\hat{x}, \hat{x}\}}(x)$ is the indicator function, taking a value of 1 when $x$ is in the interval $[\bar{x}-\hat{x}, \bar{x}+\hat{x}]$, and 0 elsewhere. This form of cosine filter allowed for nearly perfect tiling of frequency space.

In order to have 3D spectrotemporal selectivity, the filter response of the $v$th V1 neuron to the $i$th component, $S_{vi}$, was the separable product of cosine filters, modulated by each component's stimulus contrast ($c_i$), direction ($d_i$), and spatial and temporal frequency ($s_i$ and $t_i$, respectively):

$$S_{vi}\left(c_i, d_i, s_i, t_i\right) = c_i f\left(d_i \,|\, \bar{d}_v, \hat{d}\right) f\left(\log_2(s_i) \,|\, \log_2(\bar{s}_v), \hat{s}\right) f\left(z(t_i) \,|\, z(\bar{t}_v), \hat{t}\right) \tag{3.2}$$

where the V1 neurons are arranged in a cylindrical lattice at 16 directions, 9 spatial frequencies, and 7 temporal frequencies. They are distributed such that they tile

the frequency space, linearly in direction, and on a $log_2$ scale in spatial frequency (from 0.1 to 10cpd). $z(x)$ is a warping function that allows the temporal frequency response of the V1 population to perfectly tile [55] from 0 to 50cps by smoothly transitioning from linear at 0 to logarithmic at higher values:

$$z\left(x\right) = 15\ \mathrm{sgn}(x)\ \left(\log_2(|x|+1)\right) \tag{3.3}$$

As a consequence, the original temporal frequency preferences ($\bar{t}^\star$), uniformly distributed on a linear scale before warping, will effectively be at the inverse of the warping function applied to those values:

$$\bar{t} = z^{-1}\left(\bar{t}^\star\right) = \mathrm{sgn}(\bar{t}^\star)\ \left(2^{\mathrm{sgn}(\bar{t}^\star)\,\bar{t}^\star/15} - 1\right) \tag{3.4}$$

The tuning amplitudes of all filters are equal along any given dimension—the result being that they always sum to 1 at any given part of the frequency space (i.e., they tile the space evenly).

In order to create phase-invariant complex cell responses from phase-dependent raised cosine filters, we evaluated four filters offset in phase by 90° (in quadrature) at every frequency, for every frame of the stimulus. For the $T$th frame, if the $i$th component has phase $p_i$, the $q$th filter response is:

$$F_{viq}\left(c_i, d_i, s_i, t_i, p_i, T\right) = \cos\left(p_i + T + q\pi/2\right) S_{vi}\left(c_i, d_i, s_i, t_i\right) \tag{3.5}$$

These filters sum the responses of each component, are half-squared, and then

summed over the filter phase, which gives the instantaneous complex cell response:

$$C_v\left(c,d,s,t,p,T\right)=\sum_q\left\lfloor\sum_i F_{viq}\left(c_i,d_i,s_i,t_i,p_i,T\right)\right\rfloor^2 \tag{3.6}$$

where $\lfloor\cdot\rfloor=\max(\cdot,0)$. Finally, the complex cells accumulate their responses over all $N_T$ frames during the whole trial (133ms) and undergo untuned contrast normalization:

$$X_v\left(c,d,s,t,p,T\,|\,\sigma\right)=\frac{\sum_{T=1}^{N_T} C_v\left(c,d,s,t,p,T\right)}{\sigma^2+(\sum_i c_i)^2} \tag{3.7}$$

where $\sigma$ is the semi-saturation contrast.

*The MT stage*

The output of the V1 stage is a $V$ by $N$ matrix, $X$, where $V=1008$ is the number of simulated V1 neurons (16 directions times 9 spatial frequencies times 7 temporal frequencies), and $N$ is the number of trials run. The MT neuron linearly weights the V1 responses with a $V$-vector, $\mathbf{w}$, relative to a baseline term, $\alpha$, which is then half-rectified and raised to a power, to yield its spike rate $\hat{\mathbf{r}}$:

$$\hat{\mathbf{r}}\left(X\,|\,\mathbf{w},\alpha,\beta\right)=g(\alpha+X\mathbf{w})^{\circ\beta} \tag{3.8}$$

where $\mathbf{x}^{\circ\beta}$ is the element-wise power.

When fitting the model to the "planar plaid" stimulus set (§2.2.3), the simulated V1 neurons which are not stimulated at all are excluded from fitting and analysis—the result is a smaller $V$.

Normalization in the MT stage was implemented in the original versions of the

cascade model [67, 155] by simulating an entire population of MT neurons. Due to the large number of parameters we are using to fit a single MT neuron, simulating a whole population of simulated neurons at every iteration of the optimization procedure is too computationally intensive for now. Furthermore, previous versions of the cascade model were fit to data without including an MT normalization stage [92, 141]. Implementing a simpler form of MT normalization amenable to optimization is however, a goal for future versions of the model.

### 3.2.5 Estimating model parameters for individual cells

*Spike-triggered average analysis*

For the spike-triggered average (STA) analysis, $g(x)$ in equation 3.8 is again $\max(x, 0)$, $\alpha$ is fixed to the mean spike rate to all stimuli, $\beta = 1$, and $\sigma = 0.05$. In order to correct for biases in the stimulus (e.g., the uneven distribution of hyperplaid components in both frequency and contrast—see §3.2.2), the responses of each V1 neuron to all stimuli ($\hat{\mathbf{X}}_v$) were first divided by $u_v$, which represents the probability that a given simulated V1 neuron was activated by a stimulus, then "whitened":

$$X'_{vn} = \frac{X_{vn}}{u_v} \tag{3.9}$$

$$\hat{\mathbf{X}}_v = \frac{N\left(X'_{vn} - \frac{1}{N}\sum_{n=1}^{N} X'_{vn}\right)}{\sum_{n=1}^{N}\left(X'_{vn} - \frac{1}{N}\sum_{n=1}^{N} X'_{vn}\right)^2} \tag{3.10}$$

The result is that each of the $V$ columns of $\hat{X}$ has zero mean and unit variance. Applying a whitening transformation on V1 responses is biologically realistic inso-

far as it normalizes, across the population, the maximum and minimum response rates, and their variability, across all frequencies represented, based on the statistics of the stimuli presented. In other words, it could be argued that this type of whitening is analogous to assuming that over the course of the entire experiment, neurons have adapted to the statistics of the stimuli. The STA was calculated by averaging the whitened V1 responses, weighted by the measured spikes $\mathbf{r}$:

$$\mathbf{w}_{STA} = \hat{X}^T \mathbf{r} \tag{3.11}$$

Predicted spike rates were generated from the STA by substituting $\mathbf{w}_{STA}$ for $\mathbf{w}$ in equation 3.8.

*Nonlinear model*

For the full model, $\alpha$ and $\beta$ in equation 3.8 are free parameters, and a "softplus" rectifier, a smooth approximation of $\max(x, 0)$, was used instead to avoid numerical errors during optimization:

$$g\,(x) = \log(e^x + 1) \tag{3.12}$$

The V1 semi-saturation contrast parameter, $\sigma$ in equation 3.2, was also a free parameter fit in the full model.

We minimized the summed squared error between the measured and predicted spike rates for each trial:

$$E = (\hat{\mathbf{r}} - \mathbf{r})^T (\hat{\mathbf{r}} - \mathbf{r}) \tag{3.13}$$

We used the 'interior-point' algorithm as part of MATLAB's fmincon function to minimize the summed squared error. Because all the weights and the three parameters of the nonlinearity had to be simultaneously estimated (up to a total of 1011 parameters), analytic gradients needed to be calculated so that the optimization procedure would converge in a reasonable amount of time. The partial derivatives comprising the gradient are:

$$\mathbf{n} = \alpha + \hat{X}\mathbf{w} \tag{3.14}$$

$$\frac{\partial E}{\partial \hat{\mathbf{r}}} = 2\left(\hat{\mathbf{r}} - \mathbf{r}\right) \tag{3.15}$$

$$\frac{\partial \hat{\mathbf{r}}}{\partial \alpha} = \beta\, g(\mathbf{n})^{\beta-1} \odot \frac{e^{\mathbf{n}}}{1+e^{\mathbf{n}}} \tag{3.16}$$

$$\frac{\partial E}{\partial \alpha} = \sum\left(\frac{\partial E}{\partial \hat{\mathbf{r}}} \odot \frac{\partial \hat{\mathbf{r}}}{\partial \alpha}\right) \tag{3.17}$$

$$\frac{\partial E}{\partial \beta} = \sum\left(\frac{\partial E}{\partial \hat{\mathbf{r}}} \odot \hat{\mathbf{r}} \odot \log\left(g(\mathbf{n})\right)\right) \tag{3.18}$$

$$\frac{\partial E}{\partial \sigma^2} = -\sum\left(\frac{\partial E}{\partial \hat{\mathbf{r}}} \odot \frac{\partial \hat{\mathbf{r}}}{\partial \alpha} \odot \frac{\hat{X}\mathbf{w}}{\left(\sigma^2 + \sum_i \mathbf{c}_i\right)^{\circ 2}}\right) \tag{3.19}$$

$$\frac{\partial E}{\partial \mathbf{w}} = \hat{X}^T\left(\frac{\partial E}{\partial \hat{\mathbf{r}}} \odot \frac{\partial \hat{\mathbf{r}}}{\partial \alpha}\right) \tag{3.20}$$

where the numerator in equation 3.19 is element-wise divided by the denominator. The weights were initialized to $\mathbf{w}_{STA}$, computed from all trials.

Convergence of the fitting procedure was determined when any of the following conditions were met: the objective function was evaluated 200,000 times, 20,000 iterations had passed, the per-iteration change in the objective function value was less than 1e-4 (on average about 0.00005%), or the change in the estimate of $\mathbf{w}$ was less than 1e-5 times the norm of $\mathbf{w}$. Typically, convergence was achieved by

one of the latter two criteria.

We ran several diagnostics on the optimization procedure to make sure it wasn't falling into local minima or overfitting. We used the DERIVEST function in the DERIVESTsuite ( https://github.com/samuellab/MAGATAnalyzer-Matlab-Analysis/tree/master/utility%20functions/DERIVESTsuite ) to numerically verify the analytic gradients. During optimization, we observed the objective function to decrease smoothly and monotonically. Furthermore, the optimization was robust to different initial parameter values. For the results we report here, we used the STA, $\beta = 1$, and $\alpha$ equal to the spontaneous firing rate of the neuron as initial conditions, but using constant weights or different $\beta$ values produced similar results (that took longer to converge).

*Assessing STA and nonlinear model prediction performance*

Only simulated V1 neurons that were modulated by the stimulus in a given experiment, and their corresponding weights, were included in the optimization. We performed 12-fold cross-validation for both STAs and full model fits: the model was divided randomly into twelve parts. For each "fold," the model was trained on 11 of the 12 parts, and tested on the last, held-out part. This was done 12 times, so that the model was tested (separately) on every twelfth of the data. The correlation ($r$) values reported are the averages of the correlations between measured and predicted spike rates for each held out data set across all 12 folds. This value was meant to match those values reported in [112].

Due to differences in response gain across the basic characterization and hyperplaid recordings, which could occur as much as 40 minutes apart for a single neuron, a separate gain and offset term were fit to the model's spike rate pre-

dictions for the held out basic characterization data. STA and nonlinear model performance on these datasets is reported in terms of "goodness of fit" ($r^2$), calculated on the tuning curve values, to match those reported in [78]. However, our $r^2$ values include direction, spatial frequency, and temporal frequency tuning experiments, while theirs only include direction tuning.

### 3.2.6 Interpreting estimated spatiotemporal frequency weights

*Plane tuning maps*

In order to visualize the interaction of direction and speed tuning, we calculated the sum of weights inside slabs centered at planes of all directions and speeds. The minimum thickness of the slabs used here was $\pm 1$ cycle/second, so we could visualize planes going through weights on neurons preferring zero temporal frequency. Thicker slabs smeared out weights centered at higher speed planes. The resulting "plane tuning" maps indicate the relative excitation or suppression elicited by pattern motion at a given velocity, such as from drifting random dots.

For comparison, we generated plane tuning predictions for idealized pattern and component cells. The pattern prediction was generated by calculating plane tuning to points on a plane and its upper and lower slab bounds. The component prediction was the result of assessing plane tuning to points on a sphere.

*Plane fitting*

We fit planes to all weights, both positive and negative simultaneously, as well as positive and negative separately. In the former case, the planes were attracted to positive weights and repelled by negative weights.

Some neurons' direction tuning was narrow enough that the optimal plane

would have its width axis in the spatial frequency rather than direction dimension. Put another way, the optimal direction represented by the optimal plane was rotated 90° away from the true direction preference of the neuron. To avoid this, we used the same symmetry constraint during plane fitting as was used previously [112, 78]. Briefly, we chose a direction for the best fit plane that minimized the difference between two halves (mirror-reflected over direction) of the plane tuning map.

Whereas previously the optimal speed represented by the plane was chosen by hand [112, 78], we did so algorithmically. Given the plane tuning map, $M$, the bounds on speed ($B_s$) were determined by the squared mean weight response for the portion of the map where directions are ±90 degrees from the experimentally determined preferred direction:

$$m_s = \frac{1}{D} \sum_{d=-90:90} M_{sd}^2 \tag{3.21}$$

where $D$ is the number of directions in the plane map between ±90 degrees from preferred. The bounds on the speeds are the smallest and largest speeds where $m_s$ is greater than half of its maximum value:

$$b(s) = \begin{cases} 1, & \text{if } m_s > \max(m_s)/2 \\ 0, & \text{otherwise} \end{cases} \tag{3.22}$$

$$B_s = \{\arg\min_s b(s) = 1, \arg\max_s b(s) = 1\} \tag{3.23}$$

The direction bounds were further narrowed (from the ±90 degrees direction bound in equation 3.21) in the same manner, starting with the speed bounds just calcu-

lated above.

Finally, to find the optimal plane, the weighted total least squares of the weights was minimized, subject to the constraints above:

$$\min_{\mathbf{a}} ||\mathbf{w}^T \hat{X} \mathbf{a}||^2 \tag{3.24}$$

where $\mathbf{a}$ is the optimal plane normal vector.

*On-plane ratio*

In order to quantify how planar the weights are, we calculated the on-plane ratio, as used previously [112, 78]. It expresses how many weights, as a proportion of all weights, reside inside a "slab" centered on the plane. The slab was $\pm 1$ octave thick, or $\pm 5$ cycle/second, whichever was greater. We calculated the on-plane ratio for all weights (positive and negative) together, as well as separately.

## 3.3 Results

We presented hyperplaid stimuli to awake and anesthetized macaques and recorded single-unit responses in MT. Each hyperplaid lasted for 133ms and had 1-4 superimposed sinusoidal gratings (17-24 per trial; see figure 3.2(a)), which pseudo-randomly sampled 3D frequency space in a manner designed to exercise the MT neuron's dynamic range as much as possible (see methods and figure 3.1).

We characterized MT responses in terms of a V1-MT cascade model (figure 3.2(b)). Briefly, each stimulus trial is filtered by a population of narrowly tuned, complex, direction selective V1 neurons, whose responses are half-wave rectified, squared, and subject to contrast normalization. The V1 responses form the input

to the MT neuron, which linearly weights (figure 3.2(c)) and sums them, half-wave rectifies the result, and raises it to a power to generate a predicted spike rate (figure 3.2(e)). Because the simulated V1 neurons tile spatial and temporal frequency in the log domain (see §3.2.4), while the predicted organization of the weights on these neurons is planar in the linear domain, we visualized the weights in terms of isosurface level sets in the linear frequency domain (figure 3.2(d)).

### 3.3.1 Spike-triggered averages predict single grating tuning

A special case of the V1-MT cascade model assumes that the MT neuron's responses are a (rectified) linear function of the simulated V1 responses. As the simplest version of the model, it served as a natural starting point for our analysis. We estimated the MT linear weights by computing the spike-triggered average (STA) V1 response to hyperplaid stimuli.

In order to evaluate the predictive power of the STAs, we generated predicted spike rates by evaluating the linear version of the cascade model with the weights set to the STAs. We assessed spatiotemporal frequency selectivity by generating spike rates to the basic (single-grating) characterization experiments that we had run prior to the hyperplaid experiment. Since the STAs were not calculated on these data, their predictions are not guaranteed to match the measured tuning.

The STAs capture spatiotemporal selectivity to single gratings well (figure 3.3(a-c)). Across the population, single grating goodness of fit was high for the majority of cells ($r^2 = 0.82 \pm 0.14$, figure 3.3(d)). This performance is slightly higher than was reported for single grating direction tuning [78]: 70% of cells have a goodness of fit greater than 0.8. We also evaluated the STA predictions for hyperplaid data by measuring the correlation coefficient between predicted and

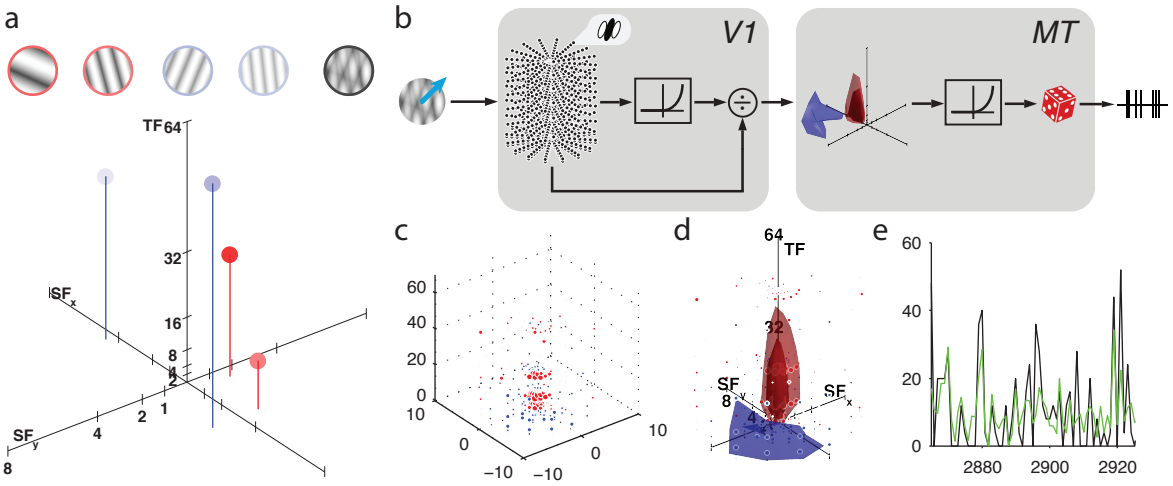


Figure 3.2: Hyperplaid stimuli and the V1-MT cascade model.
(a) An example hyperplaid. The excitatory and inhibitory components (and their spatiotemporal frequencies) are show in red and blue, respectively. The hue intensity corresponds to the contrast of each component. (b) The V1-MT cascade model. The stimulus is filtered by direction-selective V1 complex cells, narrowly tuned in the spatiotemporal frequency, and distributed on a cylindrical lattice. The filter responses are half-squared, and subject to contrast normalization. The output of the V1 stage forms the input to the MT neuron, which linearly weights the V1 responses, half-wave rectifies the result and raises it to a power, which in turn forms the predicted spike rate. (c) The linear weights from an example MT neuron on its V1 inputs. Red indicates the neuron is excited by stimulus energy captured by the corresponding weight, the selectivity of which is centered at that spatiotemporal frequency. Blue indicates a suppressive influence. The area of each point corresponds to the magnitude of the weight. The MT weights (and V1 selectivities) themselves are logarithmically spaced in the frequency domain, but can be interpreted as volumes in the same domain, with isosurfaces shown in (d). (e) A 60-trial sample of the measured spike rates (black) and those predicted by the cascade model (green).

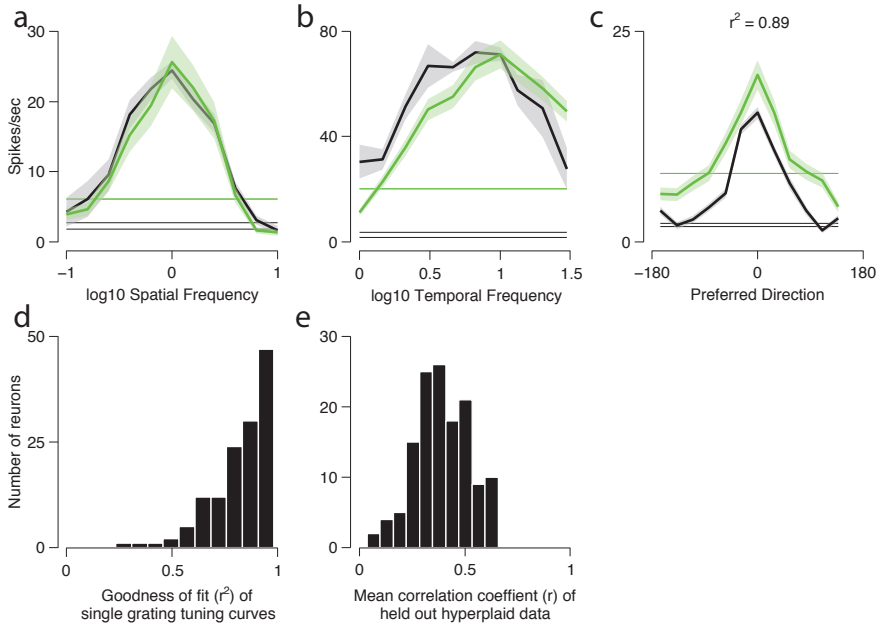


Figure 3.3: STA performance on single grating tuning and hyperplaids.
Single grating tuning curves for spatial frequency (a), temporal frequency (b), and direction (c) are shown for the measured responses (black) and those predicted by the spike triggered average. None of these data were used to calculate the STA. The shaded areas denote $\pm 1$ s.e.m. The goodness of fit for all three tuning curves together is reported in terms of $r^2$. (d) The distribution of goodness of fit to (held out) single grating tuning curves for all cells in the population. (e) The distribution of mean correlation coefficients (r) for all held out hyperplaid datasets for all cells.

measured spike rates on held out hyperplaid data ($r = 0.40 \pm 0.13$, figure 3.3(e)).

The weights for five example cells are shown in figure 3.4, ordered from the most component (a) to the most pattern selective (e). A feature common to all cells is that they are slightly elongated in the temporal frequency domain; this is not predicted explicitly by the Simoncelli-Heeger model [155], but does match recorded temporal frequency tuning (e.g., see figure 3.3(b)) and is consistent with the results reported in the previous chapter (c.f., figure A.1). Another feature common to all cells is that suppression is weak relative to excitation. Where peaks in suppression occur, they are localized in spatiotemporal frequency, at directions

near the preference of the cell.

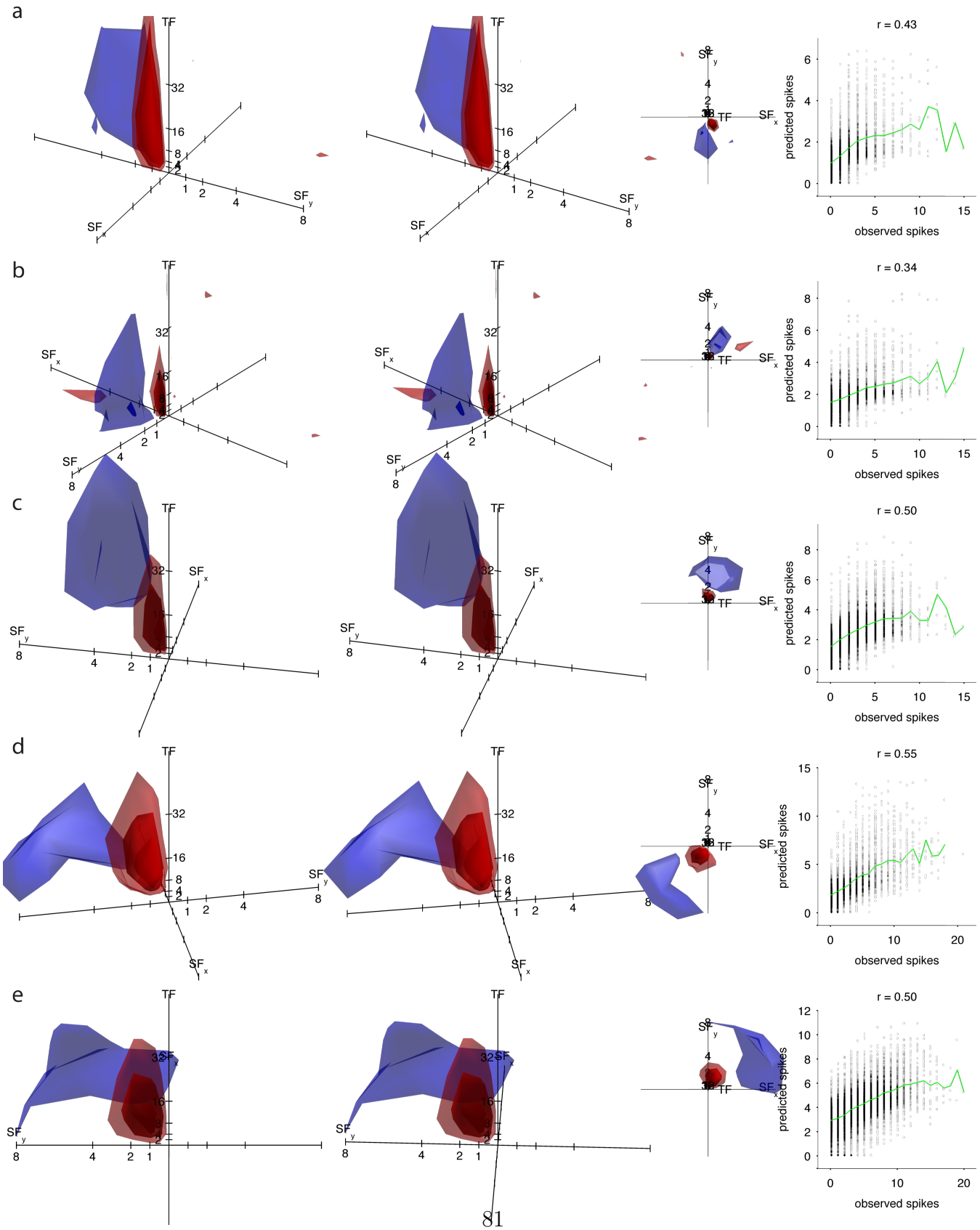
a
b
c
d
e
TF
SF
r = 0.43
r = 0.34
r = 0.50
r = 0.55
r = 0.50
predicted spikes
observed spikes

Figure 3.4 (*previous page*): STA predictions for five example cells.

(a–e) Neurons are ordered by their pattern index, from the most component (a) to the most pattern (e). The 3D linear weights are shown on the left. The first two columns are side-by-side stereoscopic renderings of the weights. The third column shows a "bird's eye view" of the weights. The right-most column shows scatter plots of measured and predicted spike counts for each trial in the hyperplaid experiment. The overlaid green traces correspond to the mean predicted spike count for a given observed spike count.

For the example cells, as pattern selectivity increases, the direction bandwidth of the weights increases, albeit very subtly. Given the difference in pattern index in the example cells, the difference in direction bandwidths is much lower than would be expected. Consequently, the weights appear less planar than the Simoncelli-Heeger model predicts they should be.

We sought to quantify the degree to which weights were confined to a common velocity plane by calculating the on-plane ratio. This is the number of nonzero weights within a "slab" $\pm 1$ octave or $\pm 5$ cycles/second, whichever is greater, over the total number of nonzero weights (figure 3.5(a), as previously described in [112, 78]). We saw no relationship between on-plane ratio and pattern index (figure 3.5(b)), calculated on the positive or negative weights in isolation. Positive weights tended to be larger than negative weights (figure 3.5(b) and (c)), and were very strongly anti-correlated ($r = -0.96$, figure 3.5(c)). This relationship held if the on-plane ratio was evaluated with planes fit separately to just the positive or negative weights ($r = -0.93$, figure 3.5(d)).

We wondered how the cells might sum rigidly moving stimuli, based on the weights given by their STAs. To generate a "map" of "plane tuning", we summed the weights in slabs centered on planes of all directions and speeds. An idealized pattern cell, simulated with weights on a slab, has a plane tuning map with a peak

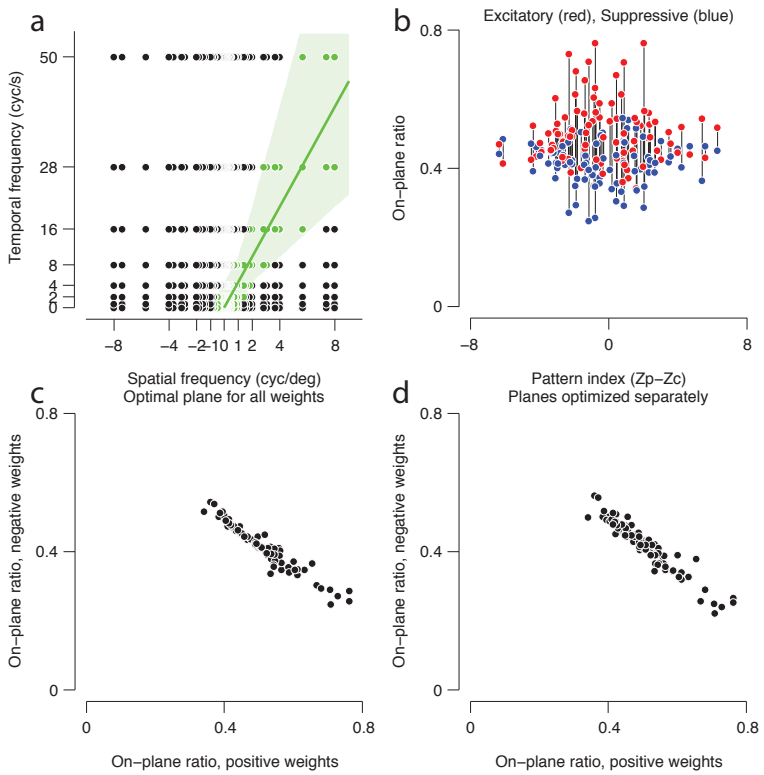


Figure 3.5: On-plane ratio.
(a) The on-plane ratio is calculated by defining a "slab" (green shaded area), which is $\pm 1$ octave or $\pm 5$ cycles/s, whichever is greater, relative to a given plane (green line). A cross-section of the slab is shown. The ratio is the number of weights within the slab (green points) divided by the total number of weights. (b) On-plane ratio calculated on the positive weights (red points) and negative weights (blue), along the best-fit plane to all weights, as a function of pattern index. (c) The same on-plane ratios as in (b), but plotted against each other. (d) The on-plane ratio for positive and negative weights, when calculated relative to planes separately optimized for just the positive or just the negative weights, respectively.

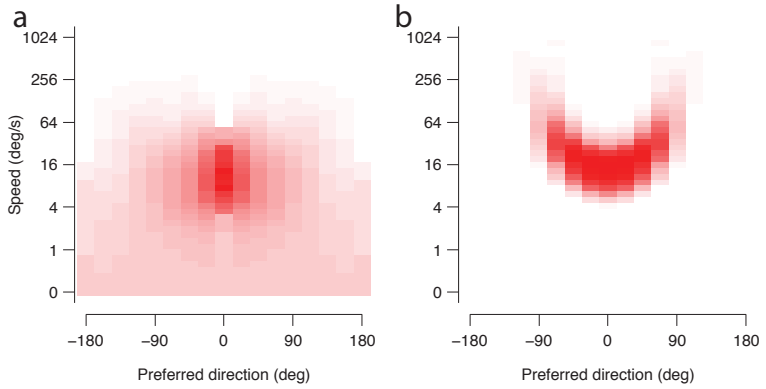


Figure 3.6: Predicted plane tuning for idealized pattern and component cells. (a) In addition to having the strongest weighting at the optimal velocity plane, an idealized pattern cell, with weights in a slab, has weight along low-speed planes at all directions. (b) For an idealized component cell, with weights in a sphere, only planes near the optimal velocity can capture its weights.

at its preferred direction and speed, which slowly tapers off as the direction represented by the plane moves away from that preferred (figure 3.6(a)). This tapering occurs more sharply at and above the preferred speed; the sustained response to planes of all directions at low speeds is a characteristic that distinguishes a pattern and a component cell in this domain. An idealized component cell (figure 3.6(b)), with its weights on a sphere, has a more localized, "U"-shaped plane tuning map. The further a plane's direction is from that preferred, the higher its speed must be to be near the component cell's weights, giving the tuning map a characteristic 'U' shape. Since it is not clear what specific volume, if any, suppression should take, we make no attempt to simulate it here.

The plane tuning maps for the same cells as in figure 3.4(a) are shown in figure 3.7. The plane tuning maps are generally smoothly structured (with the exception of the second cell) and become less "U"-shaped as pattern index increases. All the cells are suppressed by planes away from the preferred one. The peak suppression is often at a direction within ±90 degrees of the preferred direction, but with a

nonoptimal speed. Noticeably absent in pattern cell plane tuning maps is the positive response to all low speed planes (figure 3.6(a)).

Figure 3.7(b) shows the value of each weight, as a function of its distance from the optimal plane. The running means (green lines) show a similar trend as was the case in the plane tuning maps: the closest weights to the plane are those with the largest positive values; those with the largest negative values also occur relatively close to the optimal plane, before the weights eventually taper off around zero at the furthest distances.

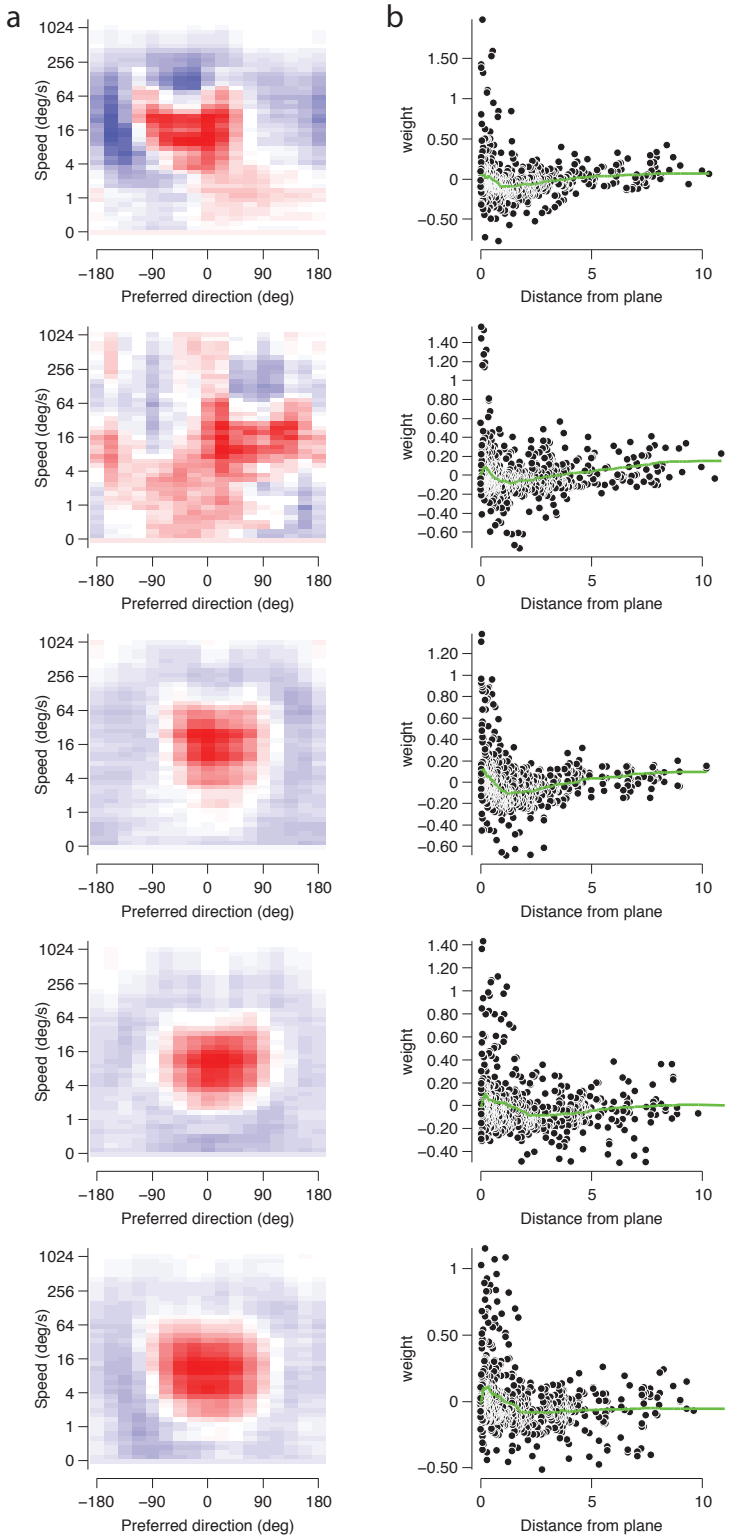
a
b
Speed (deg/s)
1024
256
64
16
4
1
0
−180
−90
0
90
180
Preferred direction (deg)
weight
Distance from plane
5
10

Figure 3.7 (*previous page*): Plane tuning for five example cells, predicted by their STAs.

The example cells are the same as whose shown in figure 3.4. (a) The tuning maps become less 'U' shaped as pattern selectivity increases. All neurons exhibit suppression for all planes away from preferred. (b) The value of each weight is plotted as a function of its (orthogonal) distance from the optimal plane (black points). The green line corresponds to the running mean of the weights, over a window of 1/10th of the weights. Weights in the range of ±0.0001% of the maximum are not plotted.

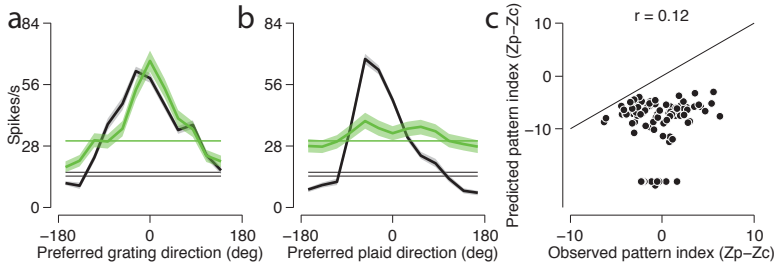


Figure 3.8: STA predictions fail to account for pattern selectivity.
Direction tuning curves to gratings (a) and plaids (b). Measured responses are shown in black, predicted in green. (c) Observed and predicted pattern index, for each cell in the population.

So far, evidence that weights obtained by calculating the STA are planar is weak. They appear to be somewhat planar, but are not nearly broad enough in the direction domain or close enough to low speed planes to conform to the predictions made by Simoncelli & Heeger [155]. Thus, we wondered how well these predictions reproduced pattern motion selectivity. Direction tuning to plaids typically appeared as it does in figure 3.8(b): bimodal, and thus consistent with component selectivity. Across the population, all cells' pattern selectivity was consistently underestimated, with no cells being predicted as pattern-selective.

### 3.3.2 Nonlinear model fits have weaker suppression

While the STAs and linear model could predict single grating tuning well, they lacked the ability to generate pattern selectivity. This led us to wonder whether the weights were indeed capturing the linear aspect of MT selectivity while a rectifying nonlinearity was insufficient to account for the nonlinear aspect of MT selectivity (i.e., pattern selectivity). We hoped to address this by allowing the linear responses to be raised to a power to generate better predictions. We fit the weights, MT exponent, and V1 semi-saturation contrast simultaneously and separately for each cell.

Performance improves for nearly every cell when testing on held out hyperplaid data (figure 3.9(a)). The mean validation performance was $r = 0.54 \pm 0.13$, similar to the nonlinear model performance reported in [112] (mean $r = 0.52$). The nonlinear model goodness of fit for single grating tuning is mixed when compared to the STAs, but they are on average slightly, but significantly, better than the STA predictions ($P = 0.00022$, Wilcoxon signed rank test, figure 3.9).

Figure 3.10 shows the weights recovered from fitting the nonlinear model to the same example cells as shown before (figures 3.4 & 3.7). Overall, the positive weights are more saturated in intensity (the isosurfaces are closer together) than those given by the STA. The negative weights are weaker when compared to the STAs (see also figure 3.11(a)). In the nonlinear fits, the suppression ratio is strongly correlated with the maximum suppressive weight, normalized to the maximum excitatory weight (figure 3.11(b)).

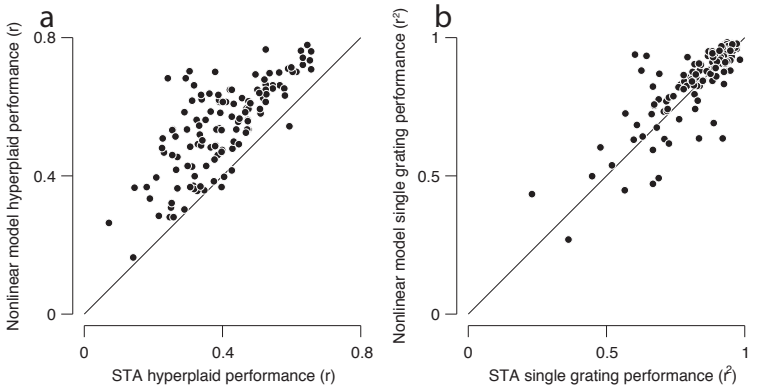


Figure 3.9: Relative performance of the STA linear model and the nonlinear model fits.
(a) Performance is in terms of the mean correlation coefficients ($r$) between the measured and predicted spikes, calculated on held out hyperplaid data and averaged across validation folds. (b) Goodness of fit of single grating tuning curves, in terms of $r^2$.

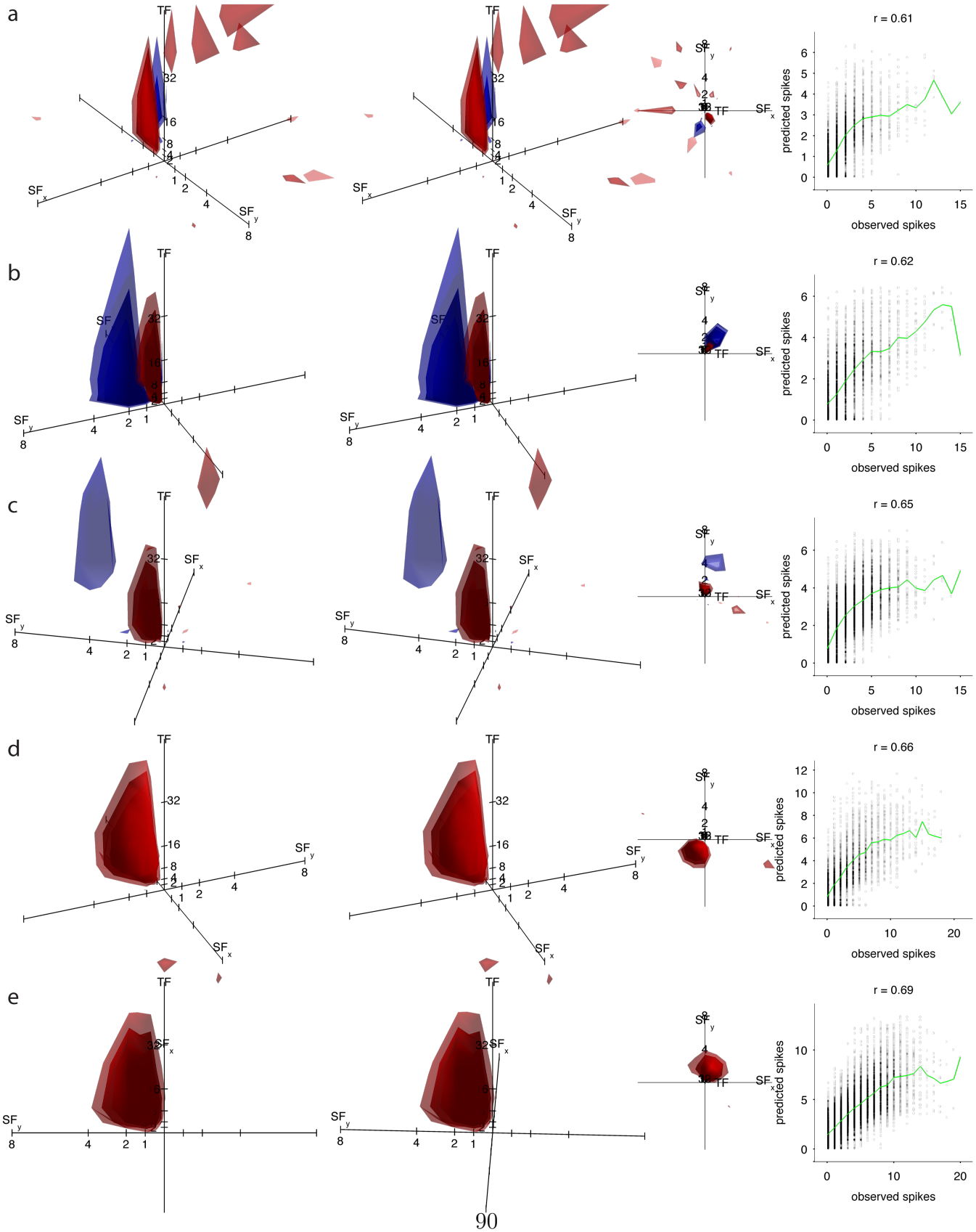
a
b
c
d
e
TF
SF x
SF y
r = 0.61
r = 0.62
r = 0.65
r = 0.66
r = 0.69
predicted spikes
observed spikes

Figure 3.10 (*previous page*): Example nonlinear model predictions for five example cells.

Weights and spike counts predicted by nonlinear model fits to hyperplaids. See figure 3.4.

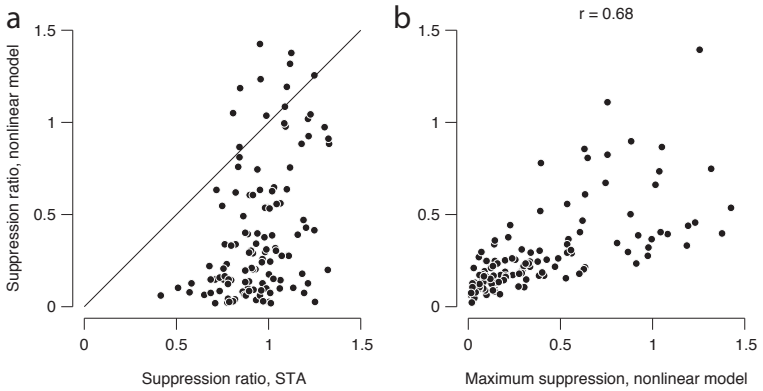


Figure 3.11: Suppression in the nonlinear model fits is weaker.
The suppression ratio is the negative sum of all negative weights over the sum of all positive weights. (a) Suppression ratio is lower in nonlinear model fits than in the STA for most cells. (b) In the nonlinear model fits, as suppression ratio increases, so does the absolute value of the maximum suppressive weight, normalized by the maximum positive weight value.

The shapes of the weights predicted by the linear and nonlinear models appear very similar overall. However, because the V1 selectivities and their corresponding MT weights are spaced in the log-domain, differences in their weights at low spatiotemporal frequencies are difficult to resolve. The plane tuning maps (figure 3.12(a)) provide insight into effects of tuning in that region of frequency space.

Component cells retain their characteristic "U" shape, and the second example cell's map is much cleaner than that generated from its STA (c.f. figure 3.7). Most cells also exhibit more positive weightings on lower speed planes, with the most pattern-selective cell's map closely resembling the idealized pattern cell prediction (3.6(a)). Intriguingly, the most component-selective cell exhibits both the "U" shape and positive weights on low speed planes.

The weaker level of suppression in the nonlinear fits, relative to the STAs, is also evident in weight value vs distance from the optimal plane scatter plots (figure 3.12(b)): there are fewer stongly negative points.

Despite performing better on held-out data, the nonlinear model still does not generate pattern behavior (figure 3.13). This was unexpected, since we thought stronger performance on the more complex stimulus set would translate to a more substantial improvement on the relatively simple plaid stimuli. If the hyperplaid stimulus is too complex, or complex in a manner that is not well described by the cascade model, perhaps a simpler stimulus set could provide insights into the cascade model's behavior.

### 3.3.3 Nonlinear model fits to the planar plaid dataset predict pattern selectivity

We fit the same nonlinear cascade model as described above (§3.3.2) to the velocity- and frequency-based gratings and plaids as described in the previous chapter (§2.2.3 and §2.3.3). Aside from being simpler than the hyperplaids, this dataset has the advantage of containing two static gratings and two unikinetic plaids, where the preferred velocity plane intersects the zero temporal frequency plane. In principle, this should allow weights on V1 neurons that are selective for zero temporal frequency to have nonzero values, which was not possible before.

Figure 3.14 shows the weights predicted by the nonlinear model fits to the plaid data. The plaids are only presented at a single (preferred) spatial frequency. Since simulated V1 neurons not engaged by the plaid data at all are excluded from model fitting, the weights are tightly localized to those spatial frequencies. For that reason, the suppression that was broad in spatial frequency seen in previous

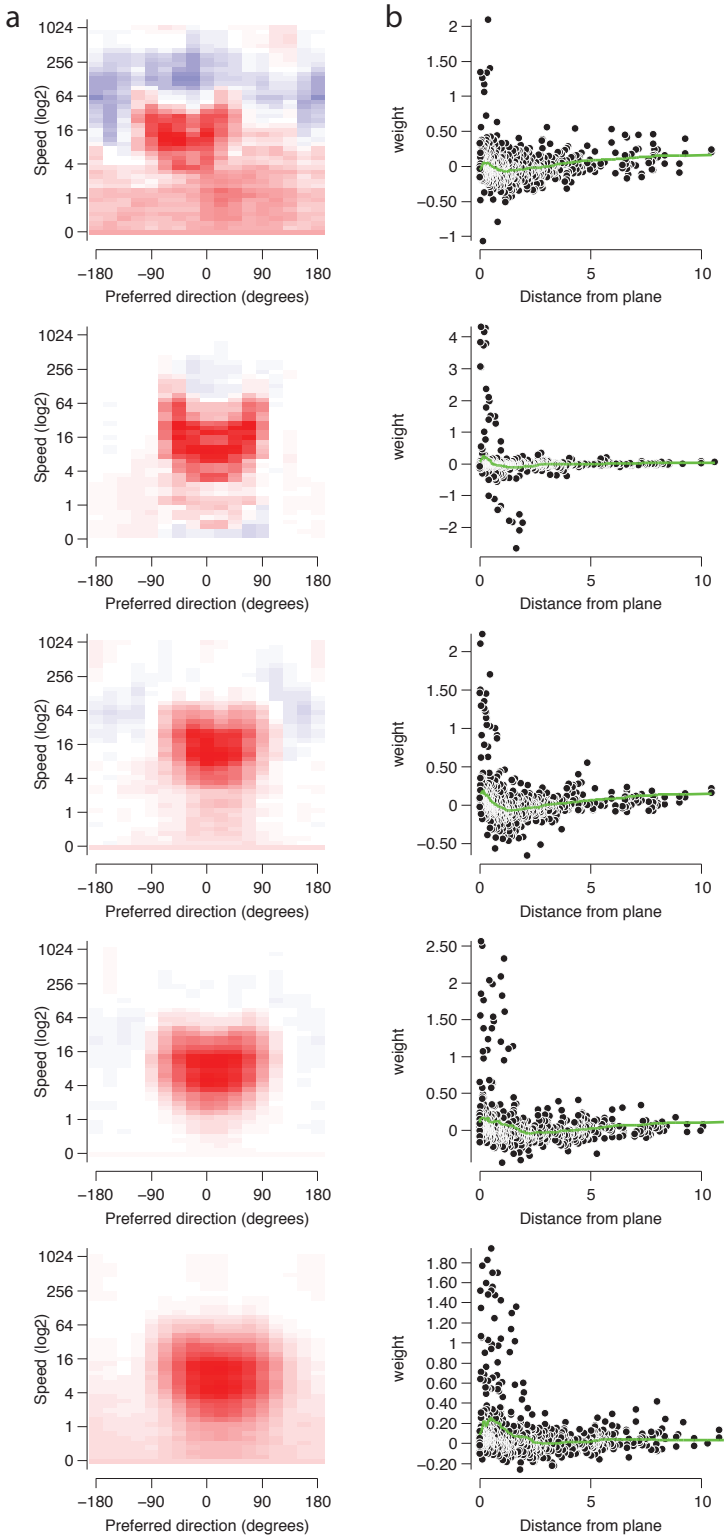


Figure 3.12: Plane tuning predicted by the nonlinear model for five example cells. See figure 3.7.

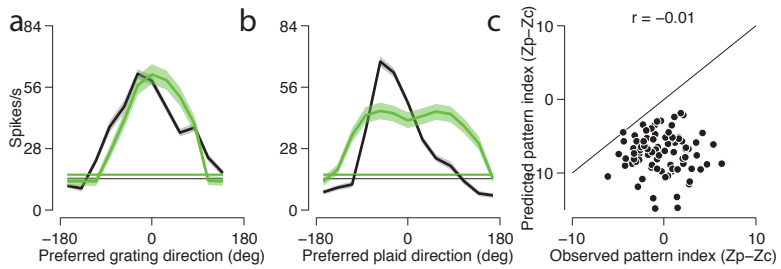


Figure 3.13: Nonlinear model predictions fail to account for pattern selectivity. Direction tuning curves to gratings (a) and plaids (b). Measured responses are shown in black, predicted in green. (c) Observed and predicted pattern index, for each cell in the population. See figure 3.8.

fits to hyperplaid data (figures 3.4 and 3.10) cannot be resolved by these plaids.

An important feature common to all but one example cell is the presence of opponent suppression—indicated by the negative weights centered at directions opposite to the directions at which the positive weights are centered. Not only is the opponent suppression well localized and centered at the opponent direction, but it is stronger than the suppression observed in previous fits.

Furthermore, the positive weights of the intermediate and pattern cells appear to be on the velocity plane. The scatter plots of observed and predicted spike counts for hyperplaid data (rightmost column in 3.14) show that these fits do not perform as well on the hyperplaid dataset.

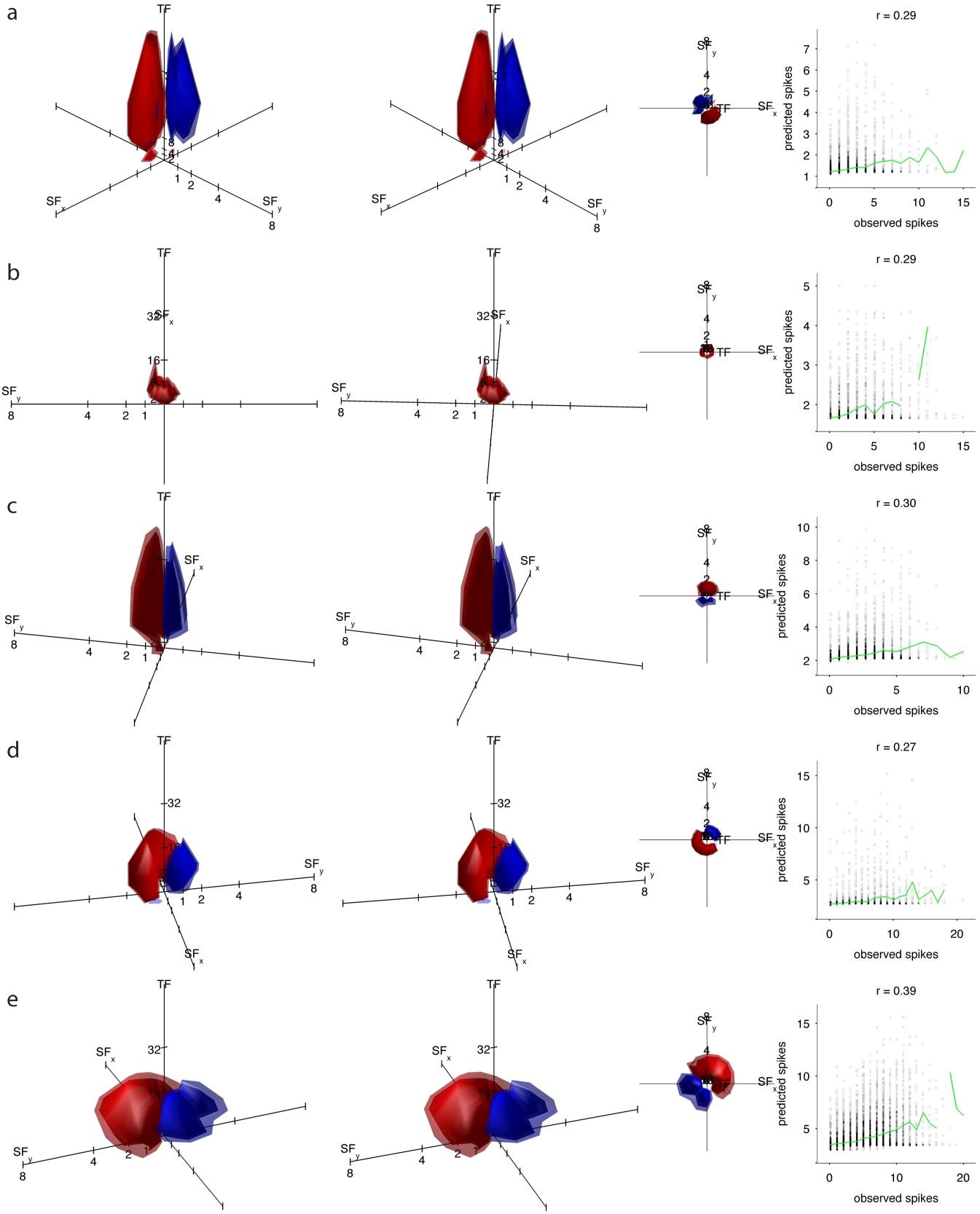
a
b
c
d
e
TF
SF x
SF y
r = 0.29
r = 0.29
r = 0.30
r = 0.27
r = 0.39
predicted spikes
observed spikes

Figure 3.14 (*previous page*): Example nonlinear model predictions, for five example cells, trained on plaids.

See figures 3.4 and 3.10.

Plane tuning for these weights (figure 3.15(a)) is substantially different from those observed in the STAs and nonlinear model fits to hyperplaids. The dark red region in the maps correspond to the range of planes that have a large positive sum of weights near them. The region is smaller for all cells, most likely due to the narrowed spatial and temporal frequency profile of these weights. Almost every cell exhibits some opponent suppression. The transition from "U"-shaped tuning in component cells to diffuse and low-speed tuning at all directions in the pattern cells is still present and appears more consistent.

Since the suppression is predominantly in the opponent direction, the positive and negative weights are well segregated in terms of the planes that intersect them (figure 3.15(a)). This segregation is also visible in (figure 3.15(b)), where positive weights and negative weights occur at nearly distinct distances from the optimal plane. Furthermore, the strongest negative weights are much closer in magnitude to the strongest positive weights than was seen in previous fits.

The nonlinear model fits the frequency- and velocity-based plaid and grating data well (figure 3.16), not only predicting pattern selectivity, but also successfully predicting the flat tuning to velocity-based plaids characteristic to pattern cells (figure 3.16(d)). The fits predict pattern index well for the majority of cells (figure 3.16(e)).

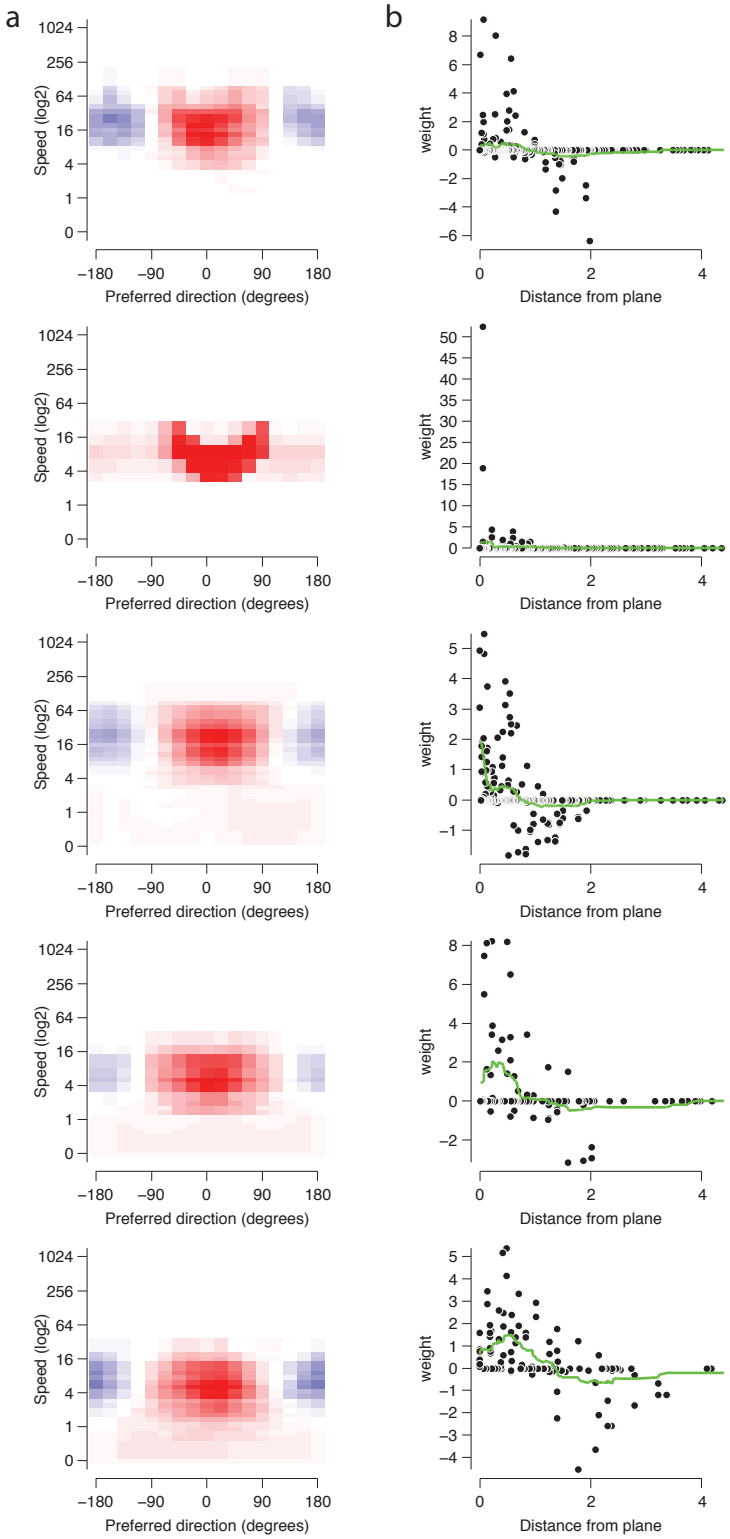


Figure 3.15: Plane tuning predicted by the nonlinear model, for five example cells, trained on plaids.
See figures 3.7 and 3.12.

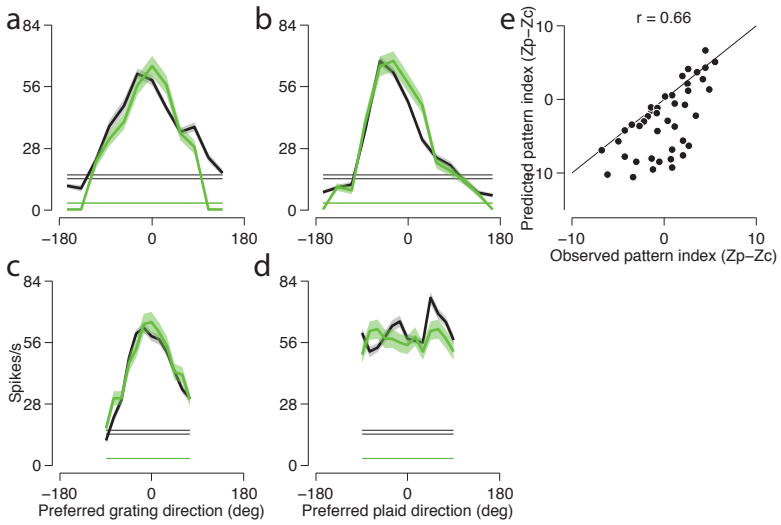


Figure 3.16: Nonlinear model fits to planar plaids predict pattern selectivity. Direction tuning curves to frequency-based gratings (a) and plaids (b), and velocity-based gratings (c) and plaids (d). Measured responses are shown in black, predicted in green. (e) Observed and predicted pattern index, for each cell in the population. See figures 3.8 and 3.13.

### 3.3.4 Comparing recovered elements of model fits

To better understand how the different model fits make different predictions of grating and plaid direction selectivity, we examined the structure of the weights from each model, collapsed onto the direction dimension. To do this, we summed the weights for each neuron over spatial and temporal frequency and normalized the resulting direction tuning curves by their maxima (figure 3.17, light traces). By aligning those tuning curves by their peaks, and averaging them, we obtained population tuning curves (figure 3.17, dark traces). The population was separated into component, intermediate, and pattern cells (the top, middle, and bottom rows, respectively).

There are a few noticeable differences between the weights given by the STA and the nonlinear model fits to hyperplaid and plaid data (figure 3.17 left, middle,

and right columns, respectively). First, the direction bandwidth is larger when the nonlinear model is fit to the planar plaid dataset, with bandwidth increasing with increasing pattern selectivity. There does not appear to be a relationship between pattern selectivity and bandwidth. Another difference is the consistency of tuning. The STAs have noisier weight profiles than the nonlinear fits. Finally, a feature that seems to be common to the different model fits is that the weights at the opponent direction become less positive as pattern selectivity increases.

We examined the relationship between the nonlinear model's performance and its predicted exponent, separately for the fits to the two different datasets (figure 3.18). When trained on hyperplaids, the nonlinear model predictions to hyperplaids are best for exponents lower than 1 (figure 3.18(a), black points). These same fits, however, yield high errors in predicted pattern index (figure 3.18(b)). Conversely, the best fits to the planar plaid dataset yield the lowest pattern index errors when the fit exponent is greater than 1 (figure 3.18(b), green points). Correspondingly, the fits with the highest exponents above 1 tend to be the ones that perform the worst on the hyperplaid dataset (figure 3.18(a)).

The exponent in the cascade model typically determines the interactions of the individual components in a plaid, with exponents higher than 1 for pattern neurons. This explains why it predicts pattern selectivity well when trained on the planar plaid dataset. The exponent is clearly being used for a different purpose when fit to hyperplaids: it is most likely being used to capture changes in contrast. Since MT neurons' contrast sensitivity typically saturates quickly, they are best captured with an exponent lower than 1. This discrepancy in fit exponent values reveals a fundamental computation of the nonlinear model's being used in a different manner based on the stimulus used to train it. It indicates that the current formulation

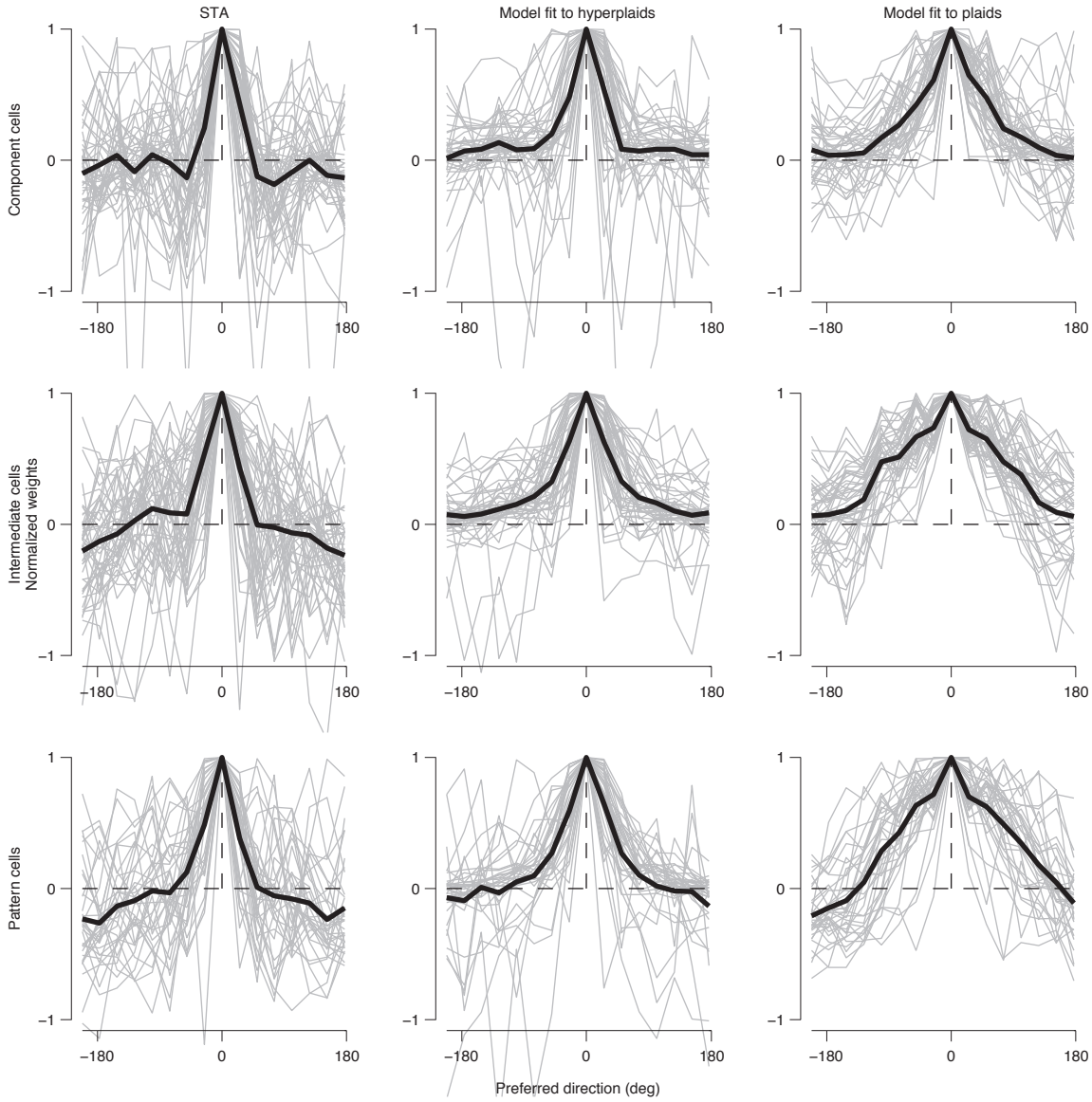


Figure 3.17: Normalized linear weight direction selectivity compared.
The first column shows the direction selectivity obtained by summing the weights predicted by the STAs over spatial and temporal frequency. The second and third columns are from nonlinear model fits, fit to hyperplaids and plaids, respectively. Each thin gray trace corresponds to this direction tuning for an individual neuron, normalized so the peak response is 1. All direction tuning curves are aligned to their preferred direction. The thick black line is the mean response across all cells. Cells are split by their pattern classification, with component cells on the top row, intermediate cells in the middle row, and pattern cells on the bottom row.

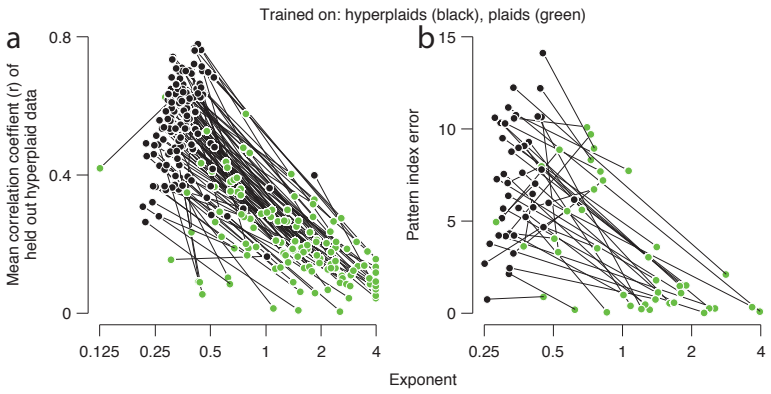


Figure 3.18: Relationship between model performance and fit exponent.
(a) Nonlinear model performance on held out hyperplaid data, as a function of the fit exponent, when trained on hyperplaids (black) and plaids (green). (b) Nonlinear model performance on held out plaid data, as a function of the fit exponent, when trained on plaids and hyperplaids. Model performance is quantified by the correlation coefficient ($r$) in (a), and absolute difference between the observed and predicted pattern index in (b).

of the model is likely insufficient to capture all the behaviors in both datasets simultaneously.

Finally, we wondered how differences in the directionality of the positive and negative weights could account for pattern selectivity. We quantified this in terms of the absolute difference in the direction of the center of mass of the positive and the negative weights. By using the center of mass, we make no assumptions about the degree to which weights adhere to a plane. There is no relationship between pattern index and the direction difference for the STAs and nonlinear model fits to the hyperplaids (figure 3.19, (a) and (b), respectively). The direction differences are relatively distributed across cells for these fits. There is a weak relationship between direction difference and pattern index for nonlinear model fits to the planar plaid dataset (figure 3.19(c)). This relationship is weaker when comparing to the predicted pattern indices predicted by the model (Pearson's $r = 0.22$, data not

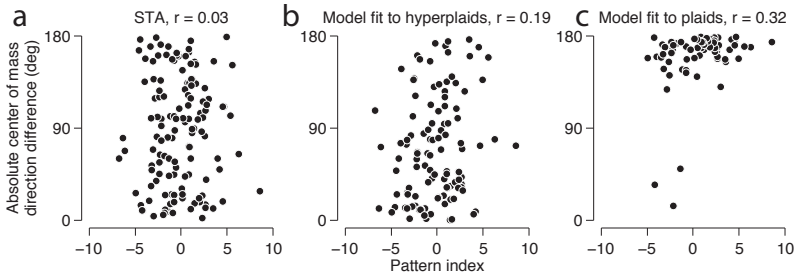


Figure 3.19: Relative excitatory and inhibitory direction tuning and pattern index. The relative difference in direction tuning for positive and negative weights is shown in terms of the absolute difference in direction coordinates for the centers of mass of the positive and negative weights. Only cells with a suppression ratio of at least 5% are included. Values of 180° correspond to opponent suppression. Direction difference for weights generated from the STAs (a), and nonlinear model fits to hyperplaids (b) and plaids (c), are plotted against pattern index.

plotted). Furthermore, the vast majority of cells exhibit opponent suppression, as is demonstrated by the preponderance of cells with a nearly 180° direction difference. All pattern cells have suppression at or near the opponent direction.

## 3.4 Discussion

We used a complex hyperplaid stimulus to characterize MT receptive fields in the 3D frequency domain. By tailoring the stimulus to contain both excitatory and suppressive elements simultaneously, and fitting a cascade model to the data, we were able to visualize excitation and suppression in 3D. We did this without imposing regularization [112] or smoothing the estimated weights [78].

We found that both linear and nonlinear variants of the cascade model could capture tuning to single gratings along three cardinal dimensions in frequency space. We found that excitatory weights tended to be more planar than the suppressive weights for the majority of cells, consistent with previous findings [112, 78]. There was a very strong inverse relationship between these two quantities.

The linear and nonlinear variants of the cascade model gave conflicting accounts of the shape of suppression, depending on which variant was used, and the dataset used to train the model. These conflicts reflect the sketchy accounts in the literature as well. Nishimoto & Gallant (2011) reported suppression was off the preferred velocity plane, and for 3 example cells, show suppression on velocity planes of the opponent direction. They did not characterize the shape of the suppression further. Inagaki et al. (2016) also reported suppression away from the preferred velocity plane, of varying strengths. A third of their cells showed suppression that exceeded excitation. Their example cells also showed opponent suppression.

We show suppression in STAs that tended to be localized to a region of frequency space near the excitatory region, but weaker than the excitation. Across the population, the suppression could be centered at any direction relative to preferred. In the nonlinear model fits to hyperplaids, suppression was weaker and more diffuse. When fitting the nonlinear model to the planar plaid dataset, suppression was strong and opponent for nearly all the cells, or nonexistent for the remainder.

The recurring observation of opponent suppression suggests that this may be a critical mechanism underlying MT selectivity. Indeed, it is a fundamental component of the Rust et al. (2006) model of pattern selectivity. We observed an increasing role of opponent suppression on direction tuning within the weights in increasingly pattern selective cells (figure 3.17), which tracks what Rust et al. (2006) observed (their figure 6).

The shape of the excitation also varied depending on the model. Weights from the nonlinear model fits to the planar plaid dataset had wider direction

tuning bandwidth, and the bandwidth had a more direct relationship to pattern selectivity than was observed in the other fits. As a consequence of having wide direction bandwidth and having weights organized along the preferred velocity plane, some pattern cells had positive weights at zero temporal frequency. This is in opposition to the other studies [112, 78], which showed no excitation at low temporal frequencies.

The connection between the structure of the weights recovered by the models and pattern selectivity remains unclear, given the different predictions the various models and stimuli make in this study and in previous ones [112, 78]. The previous studies could not make a strong link to pattern selectivity because they either did not measure it directly [112] or had only two pattern cells in their population [78]. Since we recorded from a substantial number of pattern cells, and recorded their responses to velocity plaids, these datasets have the potential to complement each other in future efforts to probe MT selectivity and understand how the models operate.

The following chapter will explore further the successes and failures of the models we have presented, how the behavior of one model can inform interpretation of the other, and how taken together, the models can point to future avenues of exploration and model development.

# Chapter 4

# Successes and failures of the two models

We have examined motion processing in area MT through the lens of Fourier analysis. We recorded single unit V1 and MT responses to drifting sinusoidal gratings, in isolation and in combinations spatially superimposed. We characterized neural selectivity to these moving stimuli in terms of weights placed on volumes within 3D spatiotemporal frequency space.

In chapter 2, we began with a targeted question: are MT receptive fields best described as having separable tuning with respect to a preferred velocity plane? With this question in mind, we designed stimuli to maximally distinguish two specific models of separable tuning in MT.

In chapter 3, we sought to characterize MT spatiotemporal selectivity more generally, with excitation and suppression allowed to take any shape, without any constraints on separability or even compactness. In particular, we hoped to resolve the suppressive elements of the receptive field. Consequently, the hyperplaid stimuli were designed to sample 3D frequency space more broadly. In order to gain more power in estimating the shape of suppression, we designed the hyperplaids to

have, on average, half of their components at frequencies near the cell's preferred stimuli, and the other half at any other frequencies (away from the preferred stimuli).

To analyze and predict the responses of MT neurons to the different stimuli, we fit variations of a V1-MT cascade model [67, 68, 155, 92, 141, 112], with modifications that reflected the nature of the stimuli used and the questions asked. To address the question of which separable receptive field organization is best, we used two model variants that were parameterized to have either velocity- or frequency-separable MT linear weighting functions. For the more general question of the shape of excitation and suppression, we used a nonparametric version of the model, where spatiotemporal frequency tuning could be represented by any arbitrary combinations of weights of V1 inputs.

Using the parametric, separable models (chapter 2), we showed that a velocity-based model better explained MT selectivity than a frequency-based model, which treated spatial and temporal frequency independently. In order to properly constrain these models to distinguish their predictions and account for critical aspects of MT behavior, such as pattern selectivity, we had to present compound stimuli (part of the "planar plaid" dataset, see figure 2.4) to exercise the neurons' nonlinearities. In other words, the complexity of the stimuli and models had to be appropriately matched.

The nonparametric, general cascade model (chapter 3) could account for pattern selectivity, but only when trained on the same planar plaid dataset used with the parametric model. When trained on hyperplaids, degree of pattern selectivity was associated with direction-independent tuning to low speeds, but ultimately did not predict pattern behavior. Furthermore, the better the nonparametric model

performed when fit to one dataset, the worse its predicted responses to the other dataset were. This could be an indication that the complexity of the model and stimuli are not appropriately matched.

This chapter aims to answer lingering questions from the previous chapters. What have we learned from these parametric and nonparametric variations of the V1-MT cascade model, and the different predictions they make on different datasets? What mechanisms exist in each model to generate complex, nonlinear behaviors?

## 4.1 The parametric and nonparametric model architectures

The separable models did not explicitly simulate the V1 stage, and as such did not include any normalization of V1 responses. The models directly operated on the spatiotemporal frequency energy of the stimuli, applying parameterized linear weighting functions to them. The responses were summed and run through a point nonlinearity, a half-wave rectified power function. The result was subject to a temporal frequency-dependent divisive suppression term which simulated the effects of MT normalization.

The nonparametric model, on the other hand, included a V1 stage in which a population of simulated V1 complex cells with narrow spatiotemporal frequency tuning tiled 3D frequency space. Their responses were subject to contrast normalization before serving as inputs to the MT neuron. The MT neuron could then apply any arbitrary configuration of weights, positive or negative, to these inputs. As in the separable models, the responses are summed and run through a half-wave rectified power function point nonlinearity. No MT normalization was applied.

The separable and nonparametric models operated under different noise mod-

els [179] of spike variability. The separable models used the modulated Poisson framework [60] to make more accurate predictions of spike rates. For computational tractability, the nonparametric model assumed a quadratic loss function.

## 4.2 Single gratings on one-dimensional paths through frequency space are weak model constraints

One of the surprises in our studies of separable tuning was that responses to constant frequency and constant velocity single grating direction tuning experiments were nearly identical for the vast majority of neurons. This is an apparent violation of the linear predictions of both the frequency- and velocity-based separable model variants, since direction tuning should be broader when the plane of the stimuli and model match (figure 2.1(e) and (h)). However, when the (nonlinear) separable models were fit to the grating responses, both models could account for the different tuning widths, making their performance on these grating responses indistinguishable (the two leftmost columns in figure 4.1).

Both spike-triggered averages and nonparametric model fits trained on hyperplaid data accurately predict single grating tuning along the principal dimensions of 3D frequency space sampled in the "basic characterization" experiments done immediately prior to the hyperplaid experiment (see §2.2.3 and figures 3.3 and 3.9). Inagaki et al. (2016) also reported high goodness-of-fit for single grating direction tuning curves.

These single grating tuning experiments all encompass excursions, along one dimensional paths, from the stimulus with optimal spatiotemporal frequency. The paths themselves are all along principal dimensions (direction, spatial frequency,

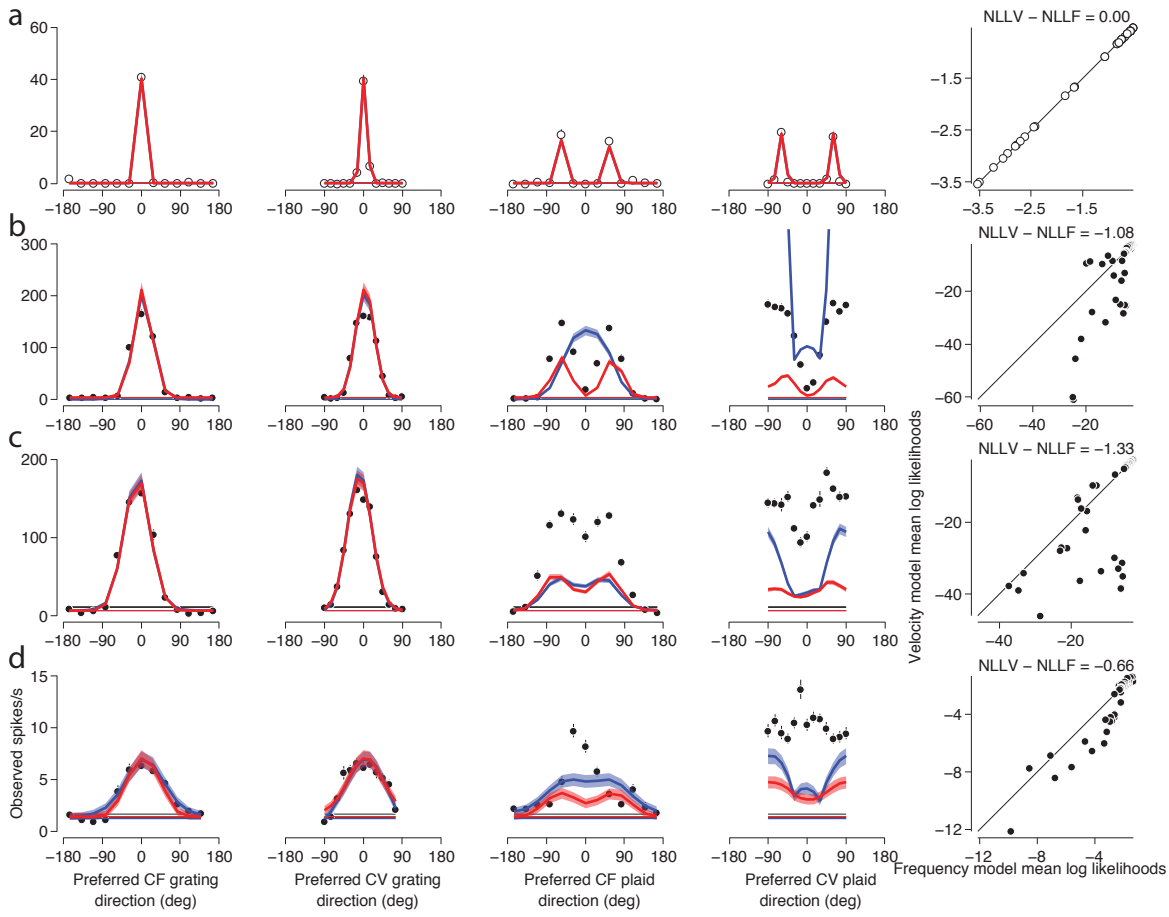


Figure 4.1: Separable model fits, trained on gratings, for four example cells. Two example component cells (a,b) and two example pattern cells (c,d). The frequency- and velocity-separable models (blue and red, respectively) were trained on constant frequency and constant velocity gratings (left two columns), as well as data from the temporal frequency tuning experiment performed immediately prior. The example cells are the same as shown in figure A.1; refer to it for more details.

temporal frequency). As a consequence, the dataset does not provide any constraint on whether these dimensions are jointly or independently represented [124, 125, 89]. MT responses along these paths through frequency space appear to be roughly linear. As a result, there are many classes of models that can capture responses to such stimulus sets.

## 4.3 Separability in 3D frequency space

The model proposed by Simoncelli & Heeger (1998) operated under the assumption that pattern cells sum spatiotemporal frequency energy along a preferred velocity plane. Rust et al. (2006) aimed to characterize pattern selectivity solely in the direction domain and accordingly used stimuli with components confined to a single, constant spatial and temporal frequency ring. By limiting the stimuli and model to the direction domain, they showed that pattern selectivity could be simulated without explicitly building a planar structure into the cascade model. Because no assumptions were made about spatial and temporal frequency selectivity, they showed that a frequency-separable model could also generate pattern tuning.

The planar plaid dataset we presented contained gratings and plaids on velocity- and frequency-separable rings. Our initial attempts to fit the separable models to this dataset gave a curious result: the velocity and frequency-separable models, with a saturating nonlinearity, could both explain the data equally well. We realized, upon closer inspection, that each model achieved this by predicting pathologically wide temporal frequency tuning. The higher the temporal frequency bandwidth, the more similar the two separable models became. Our solution was to include the data from the "basic characterization" temporal frequency tuning experiment, performed immediately prior to the hyperplaid experiment, in the fit-

ting procedure. The resulting fits were more functionally accurate and informative in distinguishing the two models.

By more widely sampling frequency space, we could more directly address questions of 3D tuning structure and separability. Previous studies [124, 125] have shown that the majority of MT neurons are tuned independently for speed and spatial frequency, consistent with a velocity-separable model (but see [89]). We presented single grating stimuli along multiple (optimal and suboptimal) paths to assess how temporal frequency tuning preference changes as a function of spatial frequency (figure 2.3(a,c,e)), and vice versa (figure 2.3(b,d,f)). Taken in isolation, these data show that tuning preferences change in a manner consistent with velocity-separable tuning in MT and frequency-separable tuning in V1.

Neither the frequency- nor the velocity-separable model consistently outperformed the other in MT when all the single grating tuning curves (such as direction tuning at optimal and suboptimal frequencies) in the dataset were included (figures 2.2 and 2.6(a)). This is most likely because MT neurons are not perfectly separable in either coordinate system. Figure 2.2(c) shows one such neuron—the velocity model prediction is better in the leftmost plot, but worse in the other three and overall (rightmost scatter plot).

## 4.4 Linear suppression in MT

Suppressed responses to directions opposite the neuron's preference have been consistently observed in a subset of MT neurons [99, 47, 135, 140, 141]. These characterizations of suppression in MT, however, were limited to the single dimension of direction tuning. How suppression interacts with other stimulus dimensions is less well understood. In response to complex optic flow stimuli, Cui et al. (2013)

characterized MT receptive fields in terms of spatial subregions of excitatory and suppressive direction-selective subunits, which could overlap to varying degrees. In their model, the change in direction between the excitatory and inhibitory subunits varied across the population. Most cells had direction-selective suppression centered at either the opponent or the preferred directions, but a smaller subset exhibited suppression at orthogonal directions.

Similarly, suppression peaked at a range of directions in our STAs and nonparametric model fits to hyperplaid stimuli (figure 3.19). When trained on the planar plaid dataset, suppression in the nonparametric model was overwhelmingly opponent. Given that these fits predicted pattern index more accurately (figure 3.16), it is likely that opponent suppression is a critical part of how the nonparametric model achieves pattern selectivity.

By design, the separable models could not predict localized opponent suppression. This was because the planar plaid stimuli were limited to rings and therefore did not constrain a separable suppressive volume. It was unclear what shape, if any, should be assumed for suppression, other than perhaps a shape similar (or identical) to the excitation. One way to incorporate suppression without making assumptions about its shape is to assume it is untuned. This could be implemented by subtracting the mean of the weights from the rest of the weights, as in the original cascade model (see appendix in [155]). Our nonparametric model includes a baseline term that performs a similar function in a less constrained manner (equation 3.8). Such a model would, however, introduce additional local minima to the error surface, making optimization difficult.

## 4.5 Nonlinear suppression in MT

Pattern selectivity is a nonlinear response behavior because it represents a failure of superposition [105] of multiple components. In the absence of compound stimuli, the nonlinear elements of the cascade model are unconstrained. This is evident in the dismal predictions to plaid tuning generated by the separable models trained on gratings (figure 4.1, third and fourth columns).

There are several mechanisms in the cascade model that can give rise to pattern selectivity. In the original cascade model [155], half-squaring and subtractive and divisive suppression acted at the MT stage to shape pattern selectivity. Rust et al. (2006) realized a form of the cascade model that could produce pattern selectivity. They attributed its ability to do so to the opponent suppression they observed in the weights on simulated V1 inputs in their model, in conjunction with super-linear (exponential) nonlinearities.

These are precisely the mechanisms that allow the nonparametric cascade model to fit the planar plaid dataset well. Figure 4.2 shows how the nonparametric model compares with the velocity-separable model for four example pattern cells. The nonparametric model can account for the interesting behaviors (flat, elevated velocity plaid responses and identical constant frequency and velocity grating responses) present in the data for some, but not all, cells. While the nonparametric model also better captures the variety of shapes of constant velocity plaid tuning (fourth column), its predictions to the other tuning curves are often noisier and less smooth than the separable model predictions.

Since the frequency-separable model is a special case of the Rust et al. (2006) model, we trained both separable models on the constant frequency stimuli in

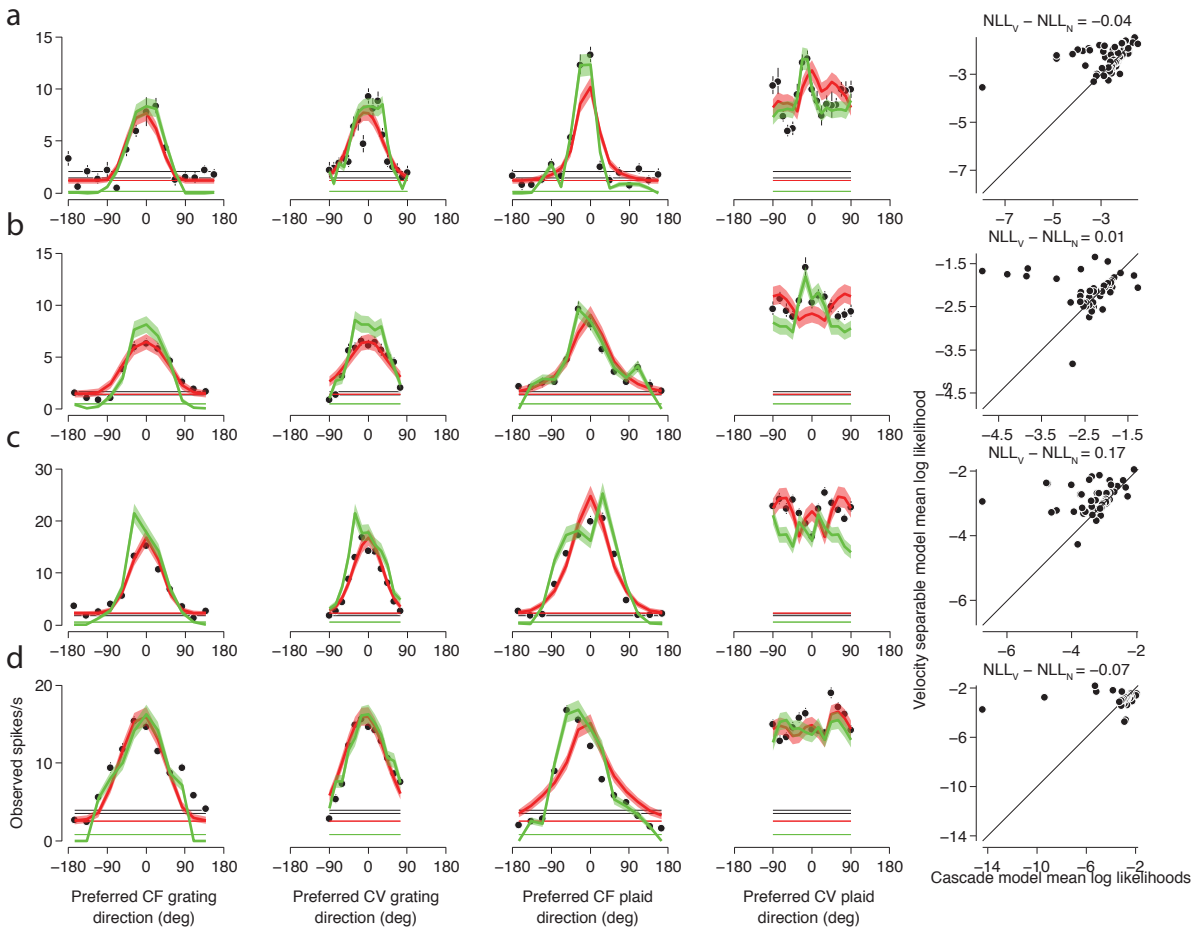


Figure 4.2: Velocity-separable and nonparametric model fits, trained on the planar plaid dataset, for four example pattern cells.
The velocity-separable and nonparametric model predictions are shown in red and green, respectively. See figure A.1 for details.

the planar plaid dataset (confined to the same ring that they used) and temporal frequency tuning data (figure 4.3). The frequency-separable model does produce pattern tuning, even without opponent suppression. The velocity-separable model, however, consistently outperforms the frequency-separable model because it better predicts constant velocity plaid responses (fourth column).

In these example cells, the limitations of the models, as well as the stimuli on which they were trained, are evident. First, the predicted constant velocity grating responses (second column from the left) are the wrong bandwidth, consistent with the linear predictions of both models (figure 2.1). Second, the predicted responses of the frequency-separable model to constant frequency plaids (third column from the left) tend to be overly broad. Third, the frequency-separable model cannot produce a flat response to constant velocity plaids (fourth column), while maintaining the correct temporal frequency tuning bandwidth. This behavior was again predicted from idealized neurons (figure 2.4(c)). Including temporal frequency tuning data was, once again, critical in keeping the two separable models realistic and distinguishable.

How does the frequency-separable model achieve (an albeit limited form of) pattern tuning without opponent suppression? The functional effects of opponent suppression may be achieved as a result of the temporal frequency-dependent divisive suppression.

The divisive suppression term was added to the separable model to approximate the effects of normalization in MT. The assumption is that a population of component and intermediate cells with tuning spanning all directions and speeds provide a negligible tuned normalization signal. Their spatiotemporal frequency tuning would be too narrow to systematically favor any particular region of fre-

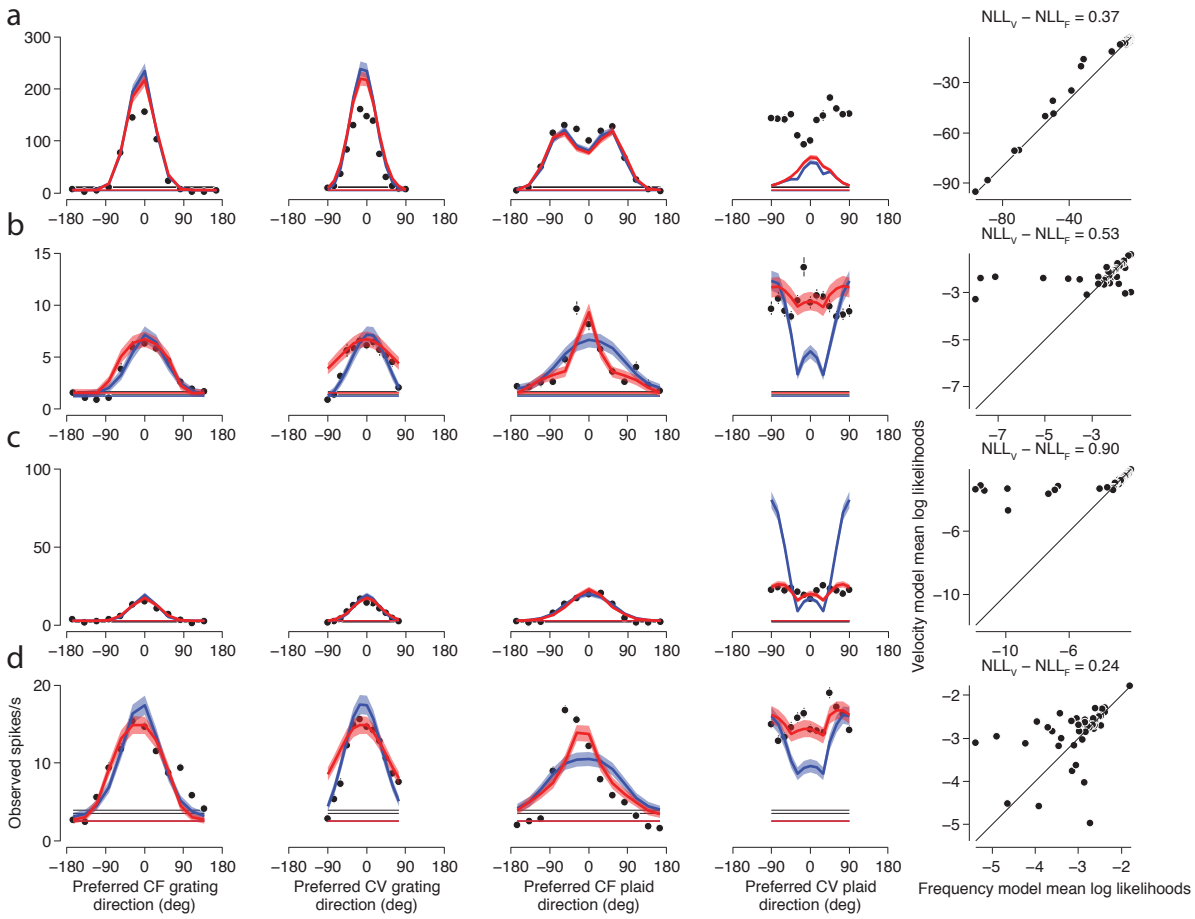


Figure 4.3: Separable model fits, trained on frequency-based gratings and plaids, for four example pattern cells.

The frequency- and velocity-separable models (blue and red, respectively) were trained on constant frequency gratings and plaids (first and third columns from the left), as well as data from the temporal frequency tuning experiment performed immediately prior. See figure A.1 for details.

quency space; thus, they fill it in an unbiased manner.

A population of frequency-separable pattern cells would also have a negligible impact on tuning because the effect of any pattern cell preferring one direction and temporal frequency will be canceled out by a pattern cell preferring the opposite direction (and same temporal frequency). Velocity-separable pattern cells, however, would represent frequency space in a biased manner, since they would all have energy at zero and low temporal frequencies, where the slabs centered on their preferred velocity planes would overlap. This is precisely what we simulated when generating the predicted "plane tuning" of an idealized pattern cell (figure 3.6(a)), which exhibits excitation for all directions at low speeds. The temporal frequency-dependent suppression term has the added benefit of being consistent with both separable models, since its effect on tuning is organized around the only plane that is constant in both temporal frequency and velocity: the zero velocity/temporal frequency plane.

Temporal frequency-dependent suppression is flexible enough to give rise to opponent suppression under both separable models. At preferred directions and speeds, excitation and suppression cancel, leaving suppression at opponent direction. This form of suppression, rather than the subtractive suppression used in the Rust et al. (2006) model, was necessary to simultaneously account for the entire planar plaid dataset and temporal frequency tuning. The crucial difference between the two types of suppression is that the (divisive) temporal frequency-dependent suppression is applied after the linear weight responses are raised to a power. For the velocity-separable model, this means that suppressive effects are stronger for single-component stimuli, which allows for narrow constant velocity grating tuning while maintaining excitatory tuning weights wide enough to support

pattern tuning. For an equivalent frequency-separable model, tuning to constant velocity gratings will already be narrower than tuning to constant frequency gratings (figure 2.1(h)), so additional suppression at low temporal frequencies would only exacerbate the discrepancy between predicted and actual direction tuning bandwidth.

In the velocity-separable model, the temporal frequency-dependent suppression is a more concise and accurate mechanism for shaping nonlinear neural selectivity. It combines the linear, subtractive suppression terms and the nonlinear, divisive normalization terms from previous versions of the model, and in doing so, makes more specific predictions about neural behavior. To date, this is the first case in which a normalization signal specifically arising in MT has been shown to be necessary to explain the full manifestations of pattern selectivity.

Is temporal frequency-dependent suppression evident in neural responses to hyperplaids? Because this suppressive mechanism is nonlinear, it is difficult to say definitively without explicitly including it in the model. The separable model allows excitation and suppression to overlap in frequency space, which is not possible in the nonparametric model.

We do observe, in the STAs, suppression at low temporal frequencies for several cells (figure 3.4(a,b,d,e)), but the suppression tends to be direction-tuned. Suppression is also common at low speed planes (figure 3.7). Suppression is weaker in the nonlinear nonparametric model fits, possibly because excitation and suppression are canceling each other out or because some of the effects of temporal frequency-dependent suppression may be achieved by the baseline term (equation 3.8). A further difficulty in assessing suppression in this dataset is that the hyperplaid stimuli did not include components with zero temporal frequency.

Suppression at low speed moving bars has been observed [99], in one case in as much as 82% of cells recorded [101].

## 4.6 Gain control

The effects of contrast on MT responses were not explored because the grating components in the planar plaid dataset were all presented at the same (50%) contrast. This was a major reason why MT normalization was simulated in terms of its effect on spatiotemporal frequency selectivity alone. The separable models included an external gain factor, based on the MT exponent, which set the relative gain between the half- and full-contrast stimuli (i.e., the gratings and plaids, see equation 2.9).

The hyperplaid stimulus had a much more complex distribution of contrasts and directions. We therefore had to simulate a V1 stage explicitly that included contrast normalization. Mante (2000) and Rust et al. (2006) showed that including self-normalization at the V1 stage enables the model to produce tuning which is stronger to plaids than gratings (e.g., the rightmost cell in figure 1.4). We also observed this phenomenon when fitting the nonparametric model, which included V1 normalization, to the planar plaid dataset (figure 4.2). If the optimization is feasible, this approach to gain control could replace the external gain factor in the separable model. Doing so, however, would add an additional free parameter.

Accounts of gain control in MT differ, depending on the stimuli used. Using gratings matched to the receptive field size, Sclar et al. (1990) reported contrast sensitivity with similar slopes in V1 and MT (exponents of 2.4 and 3.0, respectively), but much earlier saturation in MT (semi-saturation constants of $c_{50} = 7\%$, versus 20-33% in V1). Others [91, 175, 70] have found higher $c_{50}$ values of 11-20%,

often in response to smaller stimuli on the scale of V1 receptive fields [91, 70]. Since all our stimuli are matched to the classical receptive field size [25], we expect gain control to manifest with mid-to-low semi-saturation constants. Regardless, the semi-saturation constants in MT are lower than those in in V1 [147, 70, 91].

Nonparametric model fits to hyperplaids yield exponents in the MT nonlinearity almost entirely below 0.5 (figure 3.18(a)). Conversely, when fit to the planar plaid dataset, the majority of exponents were above 1. It is probably this discrepancy that underlies the poor ability of the model, when trained on one dataset, to predict responses in the other dataset. Contrast tuning is likely to be quickly saturating, corresponding to exponents below 1, whereas pattern selectivity generally demands exponents higher than 1 [155, 140, 92]. A power function as the sole nonlinear mechanism in the nonparametric model MT stage cannot handle these simultaneous, conflicting behaviors. In this instance, the complexity of the nonparametric model does not seem to be appropriately matched to that of the hyperplaid stimulus. Resolving this mismatch will require new experiments with different stimuli, changes to the model, or possibly both.

## 4.7 Proposed experiments

There are several changes that could be made to the hyperplaid stimulus to improve model fits. The most important change, in light of the separable and nonparametric model fits to the planar plaid dataset, is to include zero temporal frequency stimuli. Further, covariation in contrast and spatiotemporal frequency in the stimulus should be reduced as much as possible. The effects of contrast gain control mechanisms may be minimized by forcing all components to have equal contrast, as some previous studies have done [141, 78].

More can be done to reduce bias in the stimuli. The most direct approach is to construct the hyperplaid stimulus to be spherically symmetric [27, 159, 153]. There are several variations that could meet this criterion. In all of them, the spatiotemporal frequencies presented must be distributed in a spherically symmetric arrangement, rather than the cylindrical one we used.

In the first candidate replacement stimulus, the frequencies lie on a spherical lattice. Keeping constant the sum of all the squared component contrasts would satisfy the spherical symmetry criterion.

In the second possibility, component frequencies and contrasts are independently Gaussian distributed. This stimulus has the advantage of being amendable to STC techniques, but its main problem is that they would most likely not elicit enough spikes during an experiment to discern signals from noise. Furthermore, hyperplaids with Gaussian distributed spatial frequencies would be dominated by speeds too low to excite most MT neurons. Similarly, Gaussian distributed contrasts would yield many hyperplaids with very low contrasts.

While these stimuli would be highly informative in terms of distinguishing pattern and component cell behaviors [92], they may make it difficult to characterize the entire volume of MT spatiotemporal selectivity.

The hyperplaid stimulus we used had half of its components within the excitatory region of spatiotemporal frequency space. This stimulus structure could be kept to increase neural responsivity, but in that situation, tighter control of contrast, such as keeping summed squared contrast constant, is recommended. In a third stimulus variant, a higher level of excitation could maintained using an elliptically symmetric sampling of spatiotemporal frequency space. This sampling could readily be transformed to a spherically symmetric sampling, allowing unbi-

ased estimation of the STA [153]. The width and height of the ellipsoid, in spatial and temporal frequency, would be matched to the preferences of the neuron. The radii of the spheres could also be customized for the neuron's preferences, more densely sampling near the neuron's preferred spatial frequency.

A final advantage to all three of these stimulus variants, all spherically symmetric in frequency space, is that zero temporal frequency components would be included.

We observed elevated responses to constant velocity plaids, showing that pattern cells prefer compound stimuli on the preferred velocity plane. Therefore, a fourth hyperplaid variant, forced to have all components be consistent with a velocity plane, may be able to increase neural spiking to hyperplaids.

A final experiment, complementing these spatiotemporal frequency characterization experiments, could analyze natural movies and quantify the incidence of local motions (in both space and time) in terms of their spatiotemporal frequency content. Specifically, does motion in natural movies tend to have spatiotemporal energy on a plane? Furthermore, would a sparse coding model of motion [116, 84, 82] generate planar spatiotemporal filters? Some attempts have been made to characterize local motion statistics [50, 39, 77, 136, 20, 21, 64, 114], but none have done so in the context of planar spatiotemporal frequency.

## 4.8 Proposed changes to the model

The nonlinear cascade model can be improved even if no new data is collected with different stimuli. The most important issue to resolve is the discrepancy in predicted exponents in the MT nonlinearity when fit to different datasets. Perhaps the simplest modification to the model is to change the form of the MT nonlinearity

to be a sigmoid, implemented with divisive self-normalization, rather than a power function. This would have two advantages: (1) closer conformity to the original Simoncelli & Heeger (1998) model, and (2) the model would become a true cascade model in that the same general operations would be repeated in V1 and MT [68]. An initial attempt to implement this modification introduced local minima into the error surface, which suggests that significant optimization challenges would need to be overcome.

Alternatively, there may be formulations of the MT nonlinearity that treat (contrast) gain control via a mechanism separate from the exponent that governs spatiotemporal frequency selectivity. Wang & Movshon (2016) reported that there was no relationship between pattern index and semi-saturation or slope of the contrast response, supporting the idea that contrast and direction selectivity are handled by separate mechanisms.

A third approach involves building opponent suppression explicitly into the model. This could be done by creating a subtractive term that is identical to the weights, but rotated 180 degrees and scaled down (to avoid canceling out the excitation). A more elaborate version of this approach would be to impose constraints on the shapes of the excitatory and suppressive weights. We observed a tight inverse relationship between excitatory and suppressive weights, in terms of the degree to which they were confined to a plane (figure 3.5). A built-in opponent suppression constraint would force suppressive weights to be in a cone centered at the opponent velocity, the width of which is inversely related to the cone containing the excitatory weights (figure 4.4, top row). A variant of this constraint, combines opponent and temporal frequency-dependent suppression by aligning the suppression cone with the zero-temporal frequency plane (figure 4.4,

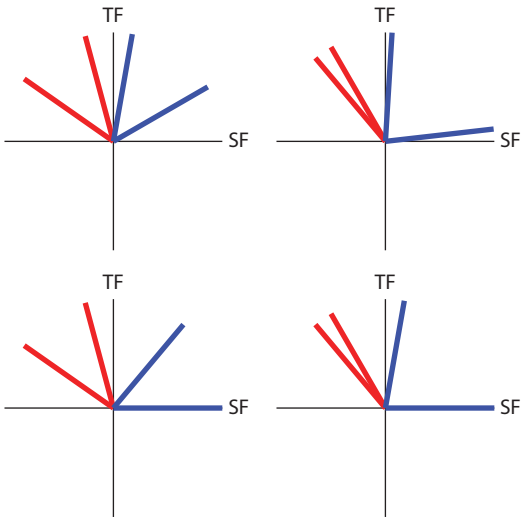


Figure 4.4: Hypothetical constraint on excitatory and inhibitory weights.
The excitatory and inhibitory weights are constrained to be within cones of with inversely related widths. The top row shows suppression centered at the opponent velocity. The bottom row shows suppression bounded at zero temporal frequency.

bottom row).

Finally, to account for the lack of perfect separability observed in the single grating study in chapter 2, a subunit model framework [167, 163, 142, 26, 171, 87, 170] may be adopted. In the simplest instantiation of this idea, a non-homogeneous population of V1 neurons would serve as inputs to the MT neuron. Each would have part of its spatiotemporal receptive field going through the MT neuron's preferred speed plane, but would be poorly aligned to it. Given that antidromically identified direction-selective inputs from V1 to MT were found to be broadly tuned to spatial and temporal frequency [106], it is likely that V1 inputs contribute frequency sensitivity far off the preferred plane. As a consequence, their receptive fields would overlap in the frequency domain, but only near the preferred velocity plane. If stimulating a single one of these inputs with a single grating would be

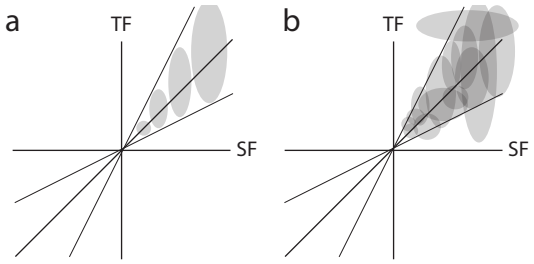


Figure 4.5: Overlapping spatiotemporal subunits.
(a) A side view of V1 neurons tiled along a preferred velocity plane in spatiotemporal frequency space according to the Simoncelli & Heeger (1998) model. (b) Proposed model in which V1 neurons with overlapping spatiotemporal frequency preferences provide input to the MT neuron. All V1 inputs have some portion of their receptive fields intersecting the preferred velocity plane, but they are poorly aligned to it. Their overlap occurs only within a narrow band around the preferred plane.

sufficient to excite the MT neuron, then its responses to single gratings would probably not be separable in either a frequency- or velocity-based coordinate system. If the MT neuron only responded to conjunctions of V1 inputs when presented with compound stimuli, then its responses would be velocity-separable.

A further extension of the subunit idea would create the equivalent of a complex cell in MT. Under such a "LN-LN" model [171, 169, 170], multiple sets of overlapping weights, representing multiple inputs from other neurons in MT, would be nonlinearly transformed, separately, before being linearly combined and passed through a second nonlinearity. In principle, this could allow overlapping weightings, such as planar excitation and temporal frequency-dependent suppression, to emerge from nonparametric model fits.

## 4.9 Conclusion

We observed that pattern cells in MT have narrow direction tuning to both constant frequency and constant velocity gratings, as well as extremely broad tuning to constant velocity plaids (figure A.1). This means that these cells are specialized for detecting the rigid motion of objects and textures, as would be predicted for detectors organized along a tilted plane in spatiotemporal frequency space [4, 177, 105, 155, 124, 125]. If they were simply summing energy along a preferred velocity plane, however, these cells would be "fooled" by the constant velocity gratings and would have shown extremely broad direction tuning. Instead, they treat the constant frequency and constant velocity gratings in the same manner. Pattern cells are therefore performing a computation that is both more complex than previously thought and more biologically relevant. MT pattern cells, in response to constant velocity gratings, signal their direction of motion correctly, despite their tuning to the pattern motion of compound stimuli. In short, they correctly solve the aperture problem for both simple and compound stimuli.

In the experiments presented in this thesis, we employed several different stimuli (single gratings, planar plaids, hyperplaids), some of which were presented to the same cells. The overlap of these datasets proved invaluable in validating and refining our models. The different levels of complexity within each stimulus set allowed us to appropriately tune the complexity of the models. Employing stimuli of varying levels of complexity is a useful strategy that can be applied generally to aid in performing systems identification of sensory systems.

Area MT has been the subject of extensive study since its original functional characterization 45 years ago [41]. Its study was highly influential in developing our

understanding of the visual system, both in terms of functional architecture and computation [13]. Because area MT has been so well characterized, it is tempting to dismiss its continued study as "scraping the bottom of a barrel," to quote a colleague. It is precisely because of our depth of knowledge about MT, however, that its study can further reveal to us the computations our brains use to process our sensory world.

# Appendix A

# Awake and anesthetized recordings

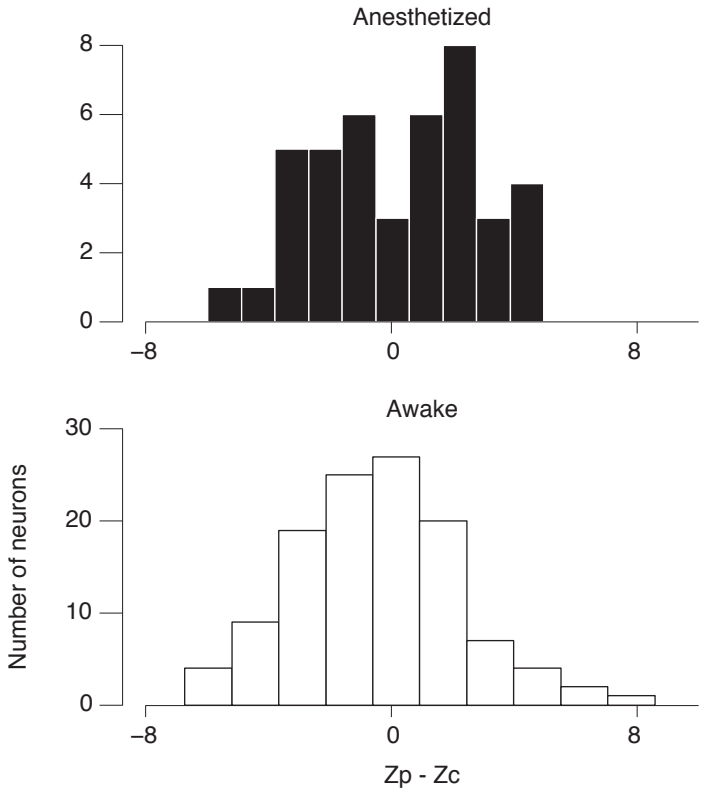


Figure A.1: Histograms of pattern indices for awake and anesthetized MT neurons. The upper plot is a histogram of pattern indices (Zp - Zc) for all anesthetized MT neurons recorded ($n = 42$). The upper plot is a histogram of pattern indices (Zp - Zc) for all awake MT neurons recorded ($n = 118$). For anesthetized MT neurons, the mean pattern index is $0.27 \pm 2.7$. For awake MT neurons, the mean pattern index is $-0.38 \pm 2.7$.

# Bibliography


[1] Daniel L Adams, John R Economides, Cristina M Jocson & Jonathan C Horton. "A biocompatible titanium headpost for stabilizing behaving monkeys." In: *Journal of neurophysiology* 98.2 (Aug. 2007), pp. 993–1001.

[2] Daniel L Adams, John R Economides, Cristina M Jocson, John M Parker & Jonathan C Horton. "A watertight acrylic-free titanium recording chamber for electrophysiology in behaving monkeys." In: *Journal of neurophysiology* 106.3 (2011), pp. 1581–1590.

[3] E H Adelson & J R Bergen. "Spatiotemporal energy models for the perception of motion." In: *Journal of the Optical Society of America. A, Optics and image science* 2.2 (Feb. 1985), pp. 284–99.

[4] Edward H Adelson & J Anthony Movshon. "Phenomenal coherence of moving visual patterns." In: *Nature* 300.5892 (Dec. 1982), pp. 523–5.

[5] T D Albright. "Direction and orientation selectivity of neurons in visual area MT of the macaque Direction and Orientation Selectivity of Neurons in Visual Area MT of the Macaque". In: *Journal of neurophysiology* 52.6 (1984), pp. 1106–1130.

[6] T D Albright. *Form-cue invariant motion processing in primate visual cortex.* 1992.

[7] T. D. Albright & R. Desimone. "Local precision of visuotopic organization in the middle temporal area (MT) of the macaque". In: *Experimental Brain Research* 65.3 (1987), pp. 582–592.

[8] J. M. Allman & J. H. Kass. "The dorsomedial cortical visual area: A third tier area in the occipital lobe of the owl monkey (aotus trivirgatus)". In: *Brain Research* 100.3 (1975), pp. 473–487.

[9] John M Allman & Jon H Kaas. “A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey”. In: *Brain research* 31 (1971), pp. 85–105.

[10] Akiyuki Anzai, I Ohzawa & R D Freeman. “Joint-encoding of motion and depth by visual cortical neurons: neural basis of the Pulfrich effect.” In: *Nature neuroscience* 4.5 (May 2001), pp. 513–8.

[11] Akiyuki Anzai, Xinmiao Peng & David C Van Essen. “Neurons in monkey visual area V2 encode combinations of orientations.” In: *Nature neuroscience* 10.10 (2007), pp. 1313–21.

[12] James F Baker, Steven E Petersen, William T Newsome & John M Allman. “Visual Response Properties of Neurons in Four Extrastriate Visual Areas of the Owl Monkey (Aotus trivirgatus):” in: *Journal of neurophysiology* 45.3 (1981), pp. 397–416.

[13] Richard T Born & David C Bradley. “Structure and Function of Visual Area Mt”. In: *Annual review of neuroscience* 28.March (2005), pp. 157–89.

[14] D. Boussaoud, L. G. Ungerleider & R. Desimone. “Pathways for motion analysis: Cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque”. In: *Journal of Comparative Neurology* 296.3 (1990), pp. 462–495.

[15] C. E. Bredfeldt, J. C. A. Read & B. G. Cumming. “A Quantitative Explanation of Responses to Disparity-Defined Edges in Macaque V2”. In: *Journal of Neurophysiology* 101.2 (2008), pp. 701–713.

[16] C E Bredfeldt & D L Ringach. “Dynamics of spatial frequency tuning in macaque V1.” In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 22.5 (2002), pp. 1976–1984.

[17] K H Britten, M N Shadlen, William T Newsome & J Anthony Movshon. “The analysis of visual motion: a comparison of neuronal and psychophysical performance.” In: *The Journal of Neuroscience* 12.12 (1992), pp. 4745–4765.

[18] Kenneth H Britten, William T Newsome, Michael N Shadlen, S Celebrini & J Anthony Movshon. “A relationship between behavioral choice and the visual responses of neurons in macaque MT”. In: *Visual Neuroscience* 13 (1996), pp. 87–100.

[19] Andrés Bruhn, Joachim Weickert & Christoph Schnörr. “Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods”. In: *International Journal of Computer Vision* 61.3 (2005), pp. 1–21.

[20] C. F. Cadieu & B. a. Olshausen. “Learning Transformational Invariants from Natural Movies”. In: *Advances in Neural Information Processing Systems 21* (2008), pp. 292–292.

[21] Charles F. Cadieu & Bruno a. Olshausen. “Learning Intermediate-Level Representations of Form and Motion from Natural Movies”. In: *Neural Computation* 24.4 (2012), pp. 827–866.

[22] Matteo Carandini & D J Heeger. “Summation and division by neurons in primate visual cortex.” In: *Science* 264.5163 (1994), pp. 1333–1336.

[23] Matteo Carandini & David J Heeger. “Normalization as a canonical neural computation.” In: *Nature Reviews Neuroscience* November (2012), pp. 1–12.

[24] Matteo Carandini, David J Heeger & J Anthony Movshon. “Linearity and normalization in simple cells of the macaque primary visual cortex.” In: *Journal of Neuroscience* 17.21 (Nov. 1997), pp. 8621–44.

[25] James R Cavanaugh, Wyeth Bair & J Anthony Movshon. “Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons.” In: *Journal of neurophysiology* 88.5 (Nov. 2002), pp. 2530–46.

[26] Xiaodong Chen, Feng Han, Mu-Ming Poo & Yang Dan. “Excitatory and suppressive receptive field subunits in awake monkey primary visual cortex (V1).” In: *Proceedings of the National Academy of Sciences of the United States of America* 104.48 (2007), pp. 19120–19125.

[27] E J Chichilnisky. “A simple white noise analysis of neuronal light responses.” In: *Network (Bristol, England)* 12.2 (2001), pp. 199–213.

[28] C L Colby, J R Duhamel & M E Goldberg. “Ventral intraparietal area of the macaque: anatomic location and visual response properties”. In: *Journal of Neurophysiology* 69.3 (1993), pp. 902–914.

[29] Yuwei Cui, Liu D Liu, Farhan a Khawaja, Christopher C Pack & Daniel a Butts. “Diverse suppressive influences in area MT and selectivity to complex motion features.” In: *The Journal of Neuroscience* 33.42 (2013), pp. 16715–28.

[30] B G Cumming & a J Parker. “Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the Macaque.” In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 20.12 (2000), pp. 4758–4767.

[31] Russell L. De Valois, E. William Yund & Norva Hepler. *The orientation and direction selectivity of cells in macaque visual cortex*. 1982.

[32] G C DeAngelis & William T Newsome. “Organization of disparity-selective neurons in macaque area MT.” In: *J Neurosci* 19.4 (1999), pp. 1398–1415.

[33] G C DeAngelis, I Ohzawa & R D Freeman. “Receptive-field dynamics in the central visual pathways.” In: *Trends in Neurosciences* 18.10 (Oct. 1995), pp. 451–8.

[34] G C Deangelis, I Ohzawa & R D Freeman. “Spatiotemporal organization of simple-cell receptive fields in the cat’s striate cortex. I. General characteristics and postnatal development”. In: *Journal of Neurophysiology* 69.4 (1993), pp. 1091–1117.

[35] G C Deangelis, I Ohzawa & R D Freeman. “Spatiotemporal Organization of Simple-Cell Receptive Fields in the Cat’s Striate Cortex. II. Linearity of Temporal and Spatial Summation”. In: *Journal of Neurophysiology* 69.4 (1993), pp. 1118–1135.

[36] Gregory C DeAngelis, B G Cumming & W T Newsome. “Cortical area MT and the perception of stereoscopic depth.” In: *Nature* 394.August (1998), pp. 677–680.

[37] R Desimone & L G Ungerleider. “Multiple visual areas in the caudal superior temporal sulcus of the macaque.” In: *Journal of Comparative Neurobiology* 248.2 (June 1986), pp. 164–89.

[38] E a DeYoe & D C Van Essen. “Segregation of efferent connections and receptive field properties in visual area V2 of the macaque.” In: *Nature* 317.6032 (1985), pp. 58–61.

[39] Daiwei W Dong & Joseph J Atick. “Statistics of Natural Time-Varying Images”. In: *Network: Computation in Neural Systems* 6.3 (1995), pp. 345–358.

[40] Rodney J. Douglas, Kevan a.C. Martin & David Whitteridge. “A Canonical Microcircuit for Neocortex”. In: *Neural Comput* 1.4 (1989), pp. 480–488.

[41] R Dubner & S M Zeki. “Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey.” In: *Brain research* 35.2 (Dec. 1971), pp. 528–32.

[42] C J Duffy & R H Wurtz. “Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli.” In: *Journal of neurophysiology* 65.6 (1991), pp. 1346–1359.

[43] C J Duffy & R H Wurtz. “Response of monkey MST neurons to optic flow stimuli with shifted centers of motion.” In: *Journal of Neuroscience* 15.7 Pt 2 (1995), pp. 5192–5208.

[44] B Efron & RJ Tibshirani. *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC, 1993.

[45] S Eifuku & R H Wurtz. “Response to motion in extrastriate area MSTl: disparity sensitivity”. In: *Journal of Neurophysiology* 82.1 (1999), pp. 2462–2475.

[46] M. Fahle & T. Poggio. “Visual Hyperacuity: Spatiotemporal Interpolation in Human Vision”. In: *Proceedings of the Royal Society B: Biological Sciences* 213.1193 (1981), pp. 451–477.

[47] D J Felleman & J H Kaas. “Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys.” In: *Journal of neurophysiology* 52.3 (1984), pp. 488–513.

[48] D J Felleman & D C Van Essen. "Distributed hierarchical processing in the primate cerebral cortex." In: *Cerebral cortex (New York, N.Y. : 1991)* 1.1 (1991), pp. 1–47.

[49] Claude L. Fennema & William B. Thompson. *Velocity determination in scenes containing several moving objects*. 1979.

[50] D. J. Field. "Relations between the statistics of natural images and the response properties of cortical cells." In: *Journal of the Optical Society of America. A, Optics and image science* 4.12 (1987), pp. 2379–94.

[51] Jeremy Freeman. "Computation and representation in the primate visual system". PhD thesis. New York University, 2013.

[52] Jeremy Freeman & Eero P Simoncelli. "Metamers of the ventral stream." In: *Nature neuroscience* 14.9 (2011), pp. 1195–1201.

[53] Jeremy Freeman, Corey M Ziemba, David J Heeger, Eero P Simoncelli & J Anthony Movshon. "A functional and perceptual signature of the second visual area in primates." In: *Nature neuroscience* 16.7 (2013), pp. 974–81.

[54] Dennis Gabor. *Theory of Communication*. 1946.

[55] Deep Ganguli & Eero P Simoncelli. "Efficient Sensory Encoding and Bayesian Inference with Heterogeneous Neural Populations". In: *Neural computation* 26.10 (2014), pp. 2103–2134.

[56] R Gattass, C G Gross & J H Sandell. "Visual topography of V2 in the macaque." In: *The Journal of comparative neurology* 201.4 (1981), pp. 519–539.

[57] Sharon Gilaie-Dotan. "Visual motion serves but is not under the purview of the dorsal pathway". In: *Neuropsychologia* 89 (2016), pp. 378–392.

[58] Melvyn A Goodale & A David Milner. "Separate visual pathways for perception and action." In: *Trends in Neurosciences* 15.1 (Jan. 1992), pp. 20–5.

[59] R Goris, EP Simoncelli & JA Movshon. "Separable dimensions for motion selectivity in macaque MT neurons". In: *Annual Meeting, Neuroscience* (2012).

[60] Robbe L T Goris, J Anthony Movshon & Eero P Simoncelli. “Partitioning neuronal variability.” In: *Nature neuroscience* 17.6 (2014), pp. 858–65.

[61] Alexander Grunewald & Evelyn K Skoumbourdis. “The integration of multiple stimulus features by V1 neurons”. In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 24.41 (2004), pp. 9185–9194.

[62] Yong Gu, Dora E Angelaki & Gregory C DeAngelis. “Neural correlates of multisensory cue integration in macaque MSTd”. In: *Nature Neuroscience* 11.10 (2008), pp. 1201–1210.

[63] Yong Gu, Gregory C DeAngelis & Dora E Angelaki. “A functional link between area MSTd and heading perception based on vestibular signals.” In: *Nature neuroscience* 10.8 (2007), pp. 1038–1047.

[64] T. Guthier, V. Willert, A. Schnall, K. Kreuter & J. Eggert. “Non-negative sparse coding for motion extraction”. In: *Proceedings of the International Joint Conference on Neural Networks* (2013).

[65] P Hammond & J N Kim. “Role of suppression in shaping orientation and direction selectivity of complex neurons in cat striate cortex.” In: *Journal of neurophysiology* 75.3 (1996), pp. 1163–76.

[66] M J Hawken, Robert M Shapley & D H Grosof. “Temporal-frequency selectivity in monkey visual cortex.” In: *Visual neuroscience* 13 (1996), pp. 477–492.

[67] D J Heeger. “Model for the extraction of image flow.” In: *Journal of the Optical Society of America. A, Optics and image science* 4.8 (1987), pp. 1455–1471.

[68] D J Heeger, E P Simoncelli & J A Movshon. “Computational models of cortical visual processing”. In: *Proceedings of the National Academy of Sciences of the United States of America* 93.January (1996), pp. 623–627.

[69] Dj Heeger. *Normalization of cell responses in cat striate cortex*. 1992.

[70] Hilary W Heuer & Kenneth H Britten. “Contrast dependence of response normalization in area MT of the rhesus macaque.” In: *Journal of neurophysiology* 88.6 (Dec. 2002), pp. 3398–408.

[71] R.v.d. Heydt, E Peterhans & G Baumgarthner. “Illusory Contours and Cortical Neuron Responses”. In: *Science* 224.4654 (1984), pp. 1260–1262.

[72] S Hochstein & R M Shapley. “Linear and nonlinear spatial subunits in Y cat retinal ganglion cells.” In: *The Journal of physiology* 262.2 (1976), pp. 265–84.

[73] Berthold K.P. Horn & Brian G. Schunck. “Determining optical flow”. In: *Artificial Intelligence* 17.1-3 (Aug. 1981), pp. 185–203.

[74] D. H. Hubel & T. N. Wiesel. “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”. In: *The Journal of Physiology* 160.1 (1962), pp. 106–154.2.

[75] David H Hubel & Margaret S Livingstone. “Segregation of form, color, and stereopsis in primate area 18.” In: *The Journal of neuroscience* 7.11 (1987), pp. 3378–3415.

[76] David H Hubel & T N Wiesel. “Receptive fields and functional architecture of monkey striate cortex”. In: *Journal of Physiology* 195.1 (1968), pp. 215–243.

[77] Aapo Hyvärinen, Jarmo Hurri & Jaakko Väyrynen. “Bubbles: a unifying framework for low-level statistical properties of natural image sequences”. In: *Journal of the Optical Society of America A* 20.7 (2003), p. 1237.

[78] Mikio Inagaki, Kota S. Sasaki, Hajime Hashimoto & Izumi Ohzawa. “Subspace mapping of the three-dimensional spectral receptive field of macaque MT neurons”. In: *Journal of Neurophysiology* (2016), jn.00934.2015.

[79] Jessica M. Johnston, Yale E. Cohen, Harry Shirley, Joji Tsunada, Sharath Bennur, Kate Christison-Lagay & Christin L. Veeder. “Recent refinements to cranial implants for rhesus macaques (Macaca mulatta)”. In: *Lab Animal* 45.5 (2016), pp. 180–186.

[80] J P Jones & L a Palmer. “An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex.” In: *Journal of neurophysiology* 58.6 (1987), pp. 1233–1258.

[81] Judson P. Jones, Aaron Stepnoski & Larry A. Palmer. "The two-dimensional spectral structure of simple receptive fields in cat striate cortex." In: *Journal of Neurophysiology* 58.6 (1987), pp. 1212–1232.

[82] Y Karklin & E P Simoncelli. "Efficient coding of natural images and movies with populations of noisy nonlinear neurons". In: *Computational and Systems Neuroscience (CoSyNe)*. Salt Lake City, Utah, Feb. 2012.

[83] Farhan a Khawaja, James M G Tsui & Christopher C Pack. "Pattern motion selectivity of spiking outputs and local field potentials in macaque visual cortex." In: *Journal of Neuroscience* 29.43 (Oct. 2009), pp. 13702–9.

[84] Taehwan Kim, Gregory Shakhnarovich & Raquel Urtasun. "Sparse Coding for Learning Interpretable Spatio-Temporal Primitives". In: *Neural Information Processing Systems* (2010), pp. 1–9.

[85] M S Livingstone, C C Pack & R T Born. "Two-dimensional substructure of MT receptive fields." In: *Neuron* 30.3 (June 2001), pp. 781–93.

[86] Margaret Livingstone & David Hubel. "Segregation of Depth: Form, Anatomy, Color, Physiology, and Movement, and Perception". In: *Science* 240.4853 (1988), pp. 740–749.

[87] Timm Lochmann, Timothy J. Blanche & Daniel A. Butts. "Construction of Direction Selectivity through Local Energy Computations in Primary Visual Cortex". In: *PLoS ONE* 8.3 (2013).

[88] BD Lucas & T Kanade. "An iterative image registration technique with an application to stereo vision." In: *Proceedings of Imaging Understanding Workshop* 130 (1981), pp. 121–130.

[89] Leo L. Lui, James A. Bourne & Marcello G P Rosa. "Spatial and temporal frequency selectivity of neurons in the middle temporal visual area of new world monkeys (Callithrix jacchus)". In: *European Journal of Neuroscience* 25.6 (2007), pp. 1780–1792.

[90] J S Lund, R D Lund, A E Hendrickson, A H Bunt & A F Fuchs. "The origin of efferent pathways from the primary visual cortex, area 17, of the macaque monkey as shown by retrograde transport of horseradish peroxidase." In: *The Journal of comparative neurology* 164.3 (1975), pp. 287–303.

[91] Najib J Majaj. “Spatial and temporal integration of motion signals in area MT”. PhD thesis. New York University, 2004, pp. 1–238.

[92] Valerio Mante. “Testing models of cortical area MT”. PhD thesis. Institute for Neuroinformatics, ETH, University of Zurich, 2000.

[93] S. Marčelja. “Mathematical description of the responses of simple cortical cells*”. In: *Journal of the Optical Society of America* 70.11 (1980), p. 1297.

[94] D Marr & S Ullman. *Directional selectivity and its use in early visual processing*. 1981.

[95] David Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. 1982, pp. 1–38.

[96] J H Maunsell & D C Van Essen. “Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity.” In: *Journal of neurophysiology* 49.5 (1983), pp. 1148–1167.

[97] J H Maunsell & D C Van Essen. “The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey.” In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 3.12 (1983), pp. 2563–2586.

[98] J H Maunsell & D C Van Essen. “Topographic organization of the middle temporal visual area in the macaque monkey: representational biases and the relationship to callosal connections and myeloarchitectonic boundaries.” In: *The Journal of comparative neurology* 266.4 (1987), pp. 535–55.

[99] John H R Maunsell & David C Van Essen. “Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation.” In: *Journal of neurophysiology* 49.5 (May 1983), pp. 1127–47.

[100] James a Mazer, William E Vinje, Josh McDermott, Peter H Schiller & Jack L Gallant. “Spatial frequency and orientation tuning dynamics in area V1.” In: *Proceedings of the National Academy of Sciences of the United States of America* 99.3 (2002), pp. 1645–1650.

[101] A Mikami, William T Newsome & Robert H Wurtz. "Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT." In: *Journal of neurophysiology* 55.6 (June 1986), pp. 1308–27.

[102] Kenneth D Miller. "Canonical computations of cerebral cortex". In: *Current Opinion in Neurobiology* 37 (2016), pp. 75–84.

[103] P. J. Mineault, F. a. Khawaja, D. a. Butts & C. C. Pack. "Hierarchical processing of complex motion along the primate dorsal visual pathway". In: *Proceedings of the National Academy of Sciences* 109.16 (2012), E972–E980.

[104] J. A. Movshon, I. D. Thompson & D. J. Tolhurst. "Receptive field organization of complex cells in the cat's striate cortex". In: *Journal of Physiology* 283 (1978), pp. 79–99.

[105] J Anthony Movshon, Edward H Adelson, Martin S Gizzi & William T Newsome. "The analysis of moving visual patterns". In: *Pattern Recognition Mechanisms (Pontificiae Academiae Scientiarum Scripta Varia)*. Ed. by Carlos Chagas, Ricardo Gattass & Charles G Gross. Vol. 54. Rome: Vatican Press, 1985, pp. 117–151.

[106] J Anthony Movshon & William T Newsome. "Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys." In: *Journal of Neuroscience* 16.23 (Dec. 1996), pp. 7733–41.

[107] J Anthony Movshon, I D Thompson & D J Tolhurst. "Spatial summation in the receptive fields of simple cells in the cat's striate cortex." In: *The Journal of physiology* 283 (1978), pp. 53–77.

[108] Jonathan J Nassi & Edward M Callaway. "Specialized Circuits from Primary Visual Cortex to V2 and Area MT". In: *Neuron* 55.5 (Sept. 2007), pp. 799–808.

[109] W T Newsome & E B Paré. *A selective impairment of motion perception following lesions of the middle temporal visual area (MT).* 1988.

[110] William T Newsome, Kenneth H Britten & J Anthony Movshon. "Neural correlates of a perceptual decision". In: *Nature* 341.6237 (1989), pp. 52–54.

[111] William T Newsome, Robert H Wurtz & H Komatsu. “Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs.” In: *Journal of neurophysiology* 60.2 (Aug. 1988), pp. 604–20.

[112] Shinji Nishimoto & Jack L Gallant. “A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies.” In: *Journal of Neuroscience* 31.41 (Oct. 2011), pp. 14551–64.

[113] Shinji Nishimoto, Tsugitaka Ishida & Izumi Ohzawa. “Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation.” In: *Journal of Neuroscience* 26.12 (Mar. 2006), pp. 3269–80.

[114] Eyal I Nitzany & Jonathan D Victor. “The statistics of local motion signals in naturalistic movies.” In: *Journal of vision* 14.4 (2014), pp. 1–15.

[115] Harris Nover, Charles H Anderson & Gregory C DeAngelis. “A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance.” In: *Journal of Neuroscience* 25.43 (Oct. 2005), pp. 10049–60.

[116] Bruno A. Olshausen & David J. Field. *Emergence of simple-cell receptive field properties by learning a sparse code for natural images*. 1996.

[117] Guy A Orban. “Visual Processing in Macaque Area MT/V5 and Its Satellites (MSTd and MSTv)”. In: *Extrastriate Cortex in Primates*. Ed. by Kathleen S Rockland, Jon H Kaas & Alan Peters. Vol. 12. Boston, MA: Springer US, 1997, pp. 359–434.

[118] Christopher C Pack, Bevil R Conway, Richard T Born & Margaret S Livingstone. “Spatiotemporal structure of nonlinear subunits in macaque visual cortex.” In: *Journal of Neuroscience* 26.3 (Jan. 2006), pp. 893–907.

[119] J A Perrone & Alexander Thiele. “Speed skills: measuring the visual speed analyzing properties of primate MT neurons.” In: *Nature neuroscience* 4.5 (2001), pp. 526–532.

[120] Rudiger Peterhans, Esther and von der Heydt. “Mechanisms of Contour Perception Contours Bridging Gaps in Monkey Visual Cortex .” In: *Journal of Neuroscience* 9.5 (1989), pp. 1749–1763.

[121] G. F. Poggio & B. Fischer. “Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey”. In: *Journal of neurophysiology* 40.6 (1977), pp. 1392–1405.

[122] Gian F. Poggio, Brad C. Motter, Salvatore Squatrito & Yves Trotter. “Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms”. In: *Vision Research* 25.3 (1985), pp. 397–406.

[123] Carlos R Ponce, Stephen G Lomber & Richard T Born. “Integrating motion and depth via parallel pathways”. In: *Nature Neuroscience* 11.2 (2008), pp. 216–223.

[124] Nicholas J Priebe, Carlos R Cassanello & Stephen G Lisberger. “The neural representation of speed in macaque area MT/V5.” In: *Journal of Neuroscience* 23.13 (July 2003), pp. 5650–61.

[125] Nicholas J Priebe, Stephen G Lisberger & J Anthony Movshon. “Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex.” In: *Journal of Neuroscience* 26.11 (Mar. 2006), pp. 2941–50.

[126] S J Prince, A D Pointon, Bruce G Cumming & A J Parker. “The precision of single neuron responses in cortical area V1 during stereoscopic depth judgments.” In: *Journal of Neuroscience* 20.9 (May 2000), pp. 3387–400.

[127] Gopathy Purushothaman & David C Bradley. “Neural population code for fine perceptual decisions in area MT.” In: *Nature neuroscience* 8.1 (Jan. 2005), pp. 99–106.

[128] S E Raiguel, M M van Hulle, D-K Xiao, V L Marcar & Guy A Orban. “Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque.” In: *The European journal of neuroscience* 7.10 (Oct. 1995), pp. 2064–82.

[129] Werner Reichardt. “Autocorrelation, a principle for the evaluation of sensory information by the central nervous system”. In: *Sensory Communication*. Ed. by WA Rosenblith. MIT Press, 1961, pp. 303–317.

[130] RC Reid & JM Alonso. “Specificity of monosynaptic connections from thalamus to visual cortex”. In: *Nature* 378.6554 (1995), pp. 281–283.

[131] Micah Richert, Thomas D Albright & Bart Krekelberg. “The complex structure of receptive fields in the middle temporal area.” In: *Frontiers in systems neuroscience* 7.March (2013), p. 2.

[132] D Ringach, Sapiro G. & R Shapley. “A subspace reverse correlation technique for the study of visual neurons”. In: *Vision Research* 37.17 (1997), pp. 2455–2464.

[133] H R Rodman & T D Albright. “Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT).” In: *Experimental Brain Research* 75.1 (1989), pp. 53–64.

[134] H R Rodman, C G Gross & T D Albright. “Afferent basis of visual response properties in area MT of the macaque. I. Effects of striate cortex removal.” In: *Journal of Neuroscience* 9.6 (June 1989), pp. 2033–50.

[135] Hillary R Rodman & Thomas D Albright. “Coding of visual stimulus velocity in area MT of the macaque”. In: *Vision Research* 27.12 (1987), pp. 2035–2048.

[136] Stefan Roth & Michael J Black. “On the Spatial Statistics of Optical Flow”. In: *International Journal of Computer Vision* 74.1 (2007), pp. 33–50.

[137] Jean-Pierre P Roy & Robert H Wurtz. *The role of disparity-sensitive cortical neurons in signalling the direction of self-motion.* 1990.

[138] Jean-Pierre Roy, H Komatsu & Robert H Wurtz. “Disparity sensitivity of neurons in monkey extrastriate area MST.” In: *Journal of Neuroscience* 12.7 (1992), pp. 2478–92.

[139] N. C. Rust & J. J. DiCarlo. “Selectivity and Tolerance ("Invariance") Both Increase as Visual Information Propagates from Cortical Area V4 to IT”. In: *Journal of Neuroscience* 30.39 (Sept. 2010), pp. 12978–12995.

[140] Nicole C Rust. “Signal transmission, feature representation and computation in areas V1 and MT of the macaque monkey”. PhD thesis. New York University, 2004, pp. 1–153.

[141] Nicole C Rust, Valerio Mante, Eero P Simoncelli & J Anthony Movshon. "How MT cells analyze the motion of visual patterns." In: *Nature neuroscience* 9.11 (Nov. 2006), pp. 1421–31.

[142] Nicole C Rust, Odelia Schwartz, J Anthony Movshon & Eero P Simoncelli. "Spatiotemporal elements of macaque v1 receptive fields." In: *Neuron* 46.6 (June 2005), pp. 945–56.

[143] H Saito, M Yukie, K Tanaka, K Hikosaka, Y Fukada & E Iwai. "Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey." In: *Journal of Neuroscience* 6.1 (Jan. 1986), pp. 145–57.

[144] C Daniel Salzman, Kenneth H Britten & William T Newsome. "Cortical microstim. influences judgements of motion direction." In: *Nature* 346 (1990), pp. 174–177.

[145] C Daniel Salzman, Chieko M Murasugi, Kenneth H Britten & William T Newsome. "Microstimulation in Visual Area MT: Effects Discrimination Performance on Direction". In: *Journal of Neuroscience* 12.6 (1992), pp. 2331–2355.

[146] Gerald E Schneider. "Two Visual Systems". In: *Science* 163.3870 (1969), pp. 895–902.

[147] Gary Sclar, John H R Maunsell & Peter Lennie. "Coding of image contrast in central visual pathways of the macaque monkey". In: *Vision research* 30.1 (1990), pp. 1–10.

[148] S. Shipp & Semir Zeki. "Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex". In: *Nature* 315.23 (1985), pp. 322–324.

[149] Stewart Shipp & Semir Zeki. "The Organization of Connections between Areas V5 and V1 in Macaque Monkey Visual Cortex." In: *The European journal of neuroscience* 1.4 (1989), pp. 309–32.

[150] Stewart Shipp & Semir Zeki. "The Organization of Connections between Areas V5 and V2 in Macaque Monkey Visual Cortex." In: *The European journal of neuroscience* 1.4 (1989), pp. 333–354.

[151] E P Simoncelli. “Design of multi-dimensional derivative filters”. In: *Proc 1st IEEE Int’l Conf on Image Proc.* Vol. I. Austin, Texas: IEEE Sig Proc Society, Nov. 1994, pp. 790–794.

[152] E P Simoncelli & D J Heeger. “Representing retinal image speed in visual cortex.” In: *Nature neuroscience* 4.5 (2001), pp. 461–462.

[153] E P Simoncelli, L Paninski, J Pillow & O Schwartz. “Characterization of Neural Responses with Stochastic Stimuli”. In: *The Cognitive Neurosciences III.* 2004. Chap. 23, pp. 327–338.

[154] Eero P. Simoncelli. “Distributed Representation and Analysis of Visual Motion”. PhD thesis. 1993, p. 131.

[155] Eero P Simoncelli & DJ Heeger. “A model of neuronal responses in visual area MT”. In: *Vision Research* 38.5 (1998), pp. 743–761.

[156] Matthew A Smith, Najib J Majaj & J Anthony Movshon. “Dynamics of motion signaling by neurons in macaque area MT.” In: *Nature neuroscience* 8.2 (Feb. 2005), pp. 220–8.

[157] Alexandra Smolyanskaya, Douglas a Ruff & Richard T Born. “Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable.” In: *Journal of neurophysiology* 110.12 (2013), pp. 2806–16.

[158] W B Spatz. “Topographically organized reciprocal connections between areas 17 and MT (visual area of superior temporal sulcus) in the marmoset Callithrix jacchus”. In: *Experimenal Brain Research* 27.5 (1977), pp. 559–72.

[159] F. E. Theunissen, S. V. David, N. C. Singh, A. Hsu, W. E. Vinje & J. L. Gallant. “Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli.” In: *Network (Bristol, England)* 12.3 (2001), pp. 289–316.

[160] O M Thomas, B G Cumming & a J Parker. “A specialization for relative disparity in V2.” In: *Nature Neuroscience* 5.5 (2002), pp. 472–478.

[161] J. Tigges, M. Tigges, S. Anschel, NA A Cross, WD D Letbetter & RL L McBride. “Areal and laminar distribution of neurons interconnecting the central visual cortical areas 17, 18, 19, and MT in squirrel monkey (Saimiri)”. In: *The Journal of Comparative Neurology* 202.4 (1981), pp. 539–560.

[162] D J Tolhurst & J a Movshon. “Spatial and temporal contrast sensitivity of striate cortical neurones.” In: *Nature* 257.5528 (1975), pp. 674–675.

[163] Jon Touryan, Gidon Felsen & Yang Dan. “Spatial structure of complex cell receptive fields measured with natural images”. In: *Neuron* 45.5 (2005), pp. 781–791.

[164] Takanori Uka & Gregory C DeAngelis. “Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity.” In: *Journal of Neuroscience* 23.8 (Apr. 2003), pp. 3515–30.

[165] Leslie G Ungerleider & Mortimer Mishkin. *Two cortical visual systems*. 1982.

[166] DC Van Essen, JHR Maunsell & JL Bixby. “The middle temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic organization”. In: *The Journal of comparative neurology* 199.3 (1981), pp. 293–326.

[167] J D Victor & R M Shapley. “The nonlinear pathway of Y ganglion cells in the cat retina.” In: *The Journal of general physiology* 74.6 (1979), pp. 671–689.

[168] Jonathan D Victor, Keith P Purpura, Andrei Belitski, Arthur Gretton, Cesare Magri, Yusuke Murayama, Marcelo A Montemurro, Nikos K Logothetis, Stefano Panzeri, Tomer Shmiel, Rotem Drori, Oren Shmiel, Yoram Benshaul, Zoltan Nadasdy & Moshe Shemesh. “Spatial phase and the temporal structure of the response to gratings in V1”. In: *Journal of neurophysiology* 80.2 (1998), pp. 554–571.

[169] Brett Vintch. “Structured hierarchical models for neurons in the early visual system”. PhD thesis. New York University, 2013.

[170] Brett Vintch, J Anthony Movshon & Eero P. Simoncelli. “A Convolutional Subunit Model for Neuronal Responses in Macaque V1.” In: *The Journal of Neuroscience* 35.44 (2015), pp. 14829–41.

[171] Brett Vintch, Andrew D Zaharia, J Anthony Movshon & Eero P Simoncelli. “Efficient and direct estimation of a neural subunit model for sensory coding”. In: *Neural Information Processing Systems*. 2012, pp. 1–9.

[172] H Wallach. “Ueber visuell wahrgenommene bewegungrichtung.” In: *Psychologische Forschung* 20 (1935), pp. 325–380.

[173] Pascal Wallisch & J Anthony Movshon. “Structure and function come unglued in the visual cortex.” In: *Neuron* 60.2 (Oct. 2008), pp. 195–7.

[174] Helena X Wang & J Anthony Movshon. “Spatial and temporal properties of pattern- and component-direction selective cells in area MT of the macaque”. In: *Society for Neuroscience*. San Diego, CA, 2010, 74.2/OO9.

[175] Helena X Wang & J Anthony Movshon. “Properties of pattern and component direction-selective cells in area MT of the macaque”. In: *Journal of Neurophysiology* 115 (2016), pp. 2705–2720.

[176] A B Watson & A J Ahumada. “A look at motion in the frequency domain”. In: *Motion: Representation and Perception*. Baltimore: ACM, 1983, pp. 1–10.

[177] Andrew B Watson & Albert J Ahumada, Jr. “Model of human visual-motion sensing”. In: *Journal of the Optical Society of America A* 2.2 (1985), pp. 322–342.

[178] Michael A Webster & Russell L De Valois. “Relationship between spatial-frequency and orientation tuning of striate-cortex cells”. In: *Journal of the Optical Society of America A* 2.7 (1985), pp. 1124–1132.

[179] Michael C-K Wu, Stephen V David & Jack L Gallant. “Complete functional characterization of sensory neurons by system identification.” In: *Annual Review of Neuroscience* 29 (2006), pp. 477–505.

[180] Sophie Wuerger, Robert M Shapley & Nava Rubin. “"On the visually perceived direction of motion" by Hans Wallach: 60 years later”. In: *Perception* 25.11 (1996), pp. 1317–1367.

[181] Robert H Wurtz & Charles J. Duffy. “Neuronal correlates of optic flow stimulation.” In: *Annals of the New York Academy of Sciences* 656 (1992), pp. 205–19.

[182] A D Zaharia, R L T Goris, J A Movshon & E P Simoncelli. “Separability of Spatiotemporal Receptive Field Structure in Macaque Area MT”. In: *Annual Meeting, Neuroscience*. Nov. 2014.

[183] Andrew D Zaharia, Robbe L T Goris, J Anthony Movshon & Eero P Simoncelli. “Compound stimuli reveal velocity separability of spatiotemporal receptive fields in macaque area MT”. In: *Annual Meeting, Vision Sciences Society* 15.12 (2015), p. 485.

[184] S M Zeki. “Functional specialisation in the visual cortex of the rhesus monkey.” In: *Nature* 274.5670 (1978), pp. 423–428.

[185] S.M. Zeki. “Functional organization fo avisual area in the posterior bank of the superior temporal sulcus of the rhesus monkey”. In: *Journal of Physiology* 236 (1974), pp. 549–573.

[186] H Zhou, H S Friedman & Rüdiger von der Heydt. “Coding of border ownership in monkey visual cortex.” In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 20.17 (2000), pp. 6594–6611.

[187] Corey M Ziemba. “Neural representation and perception of naturalistic image structure”. PhD thesis. New York University, 2016.

[188] Corey M. Ziemba, Jeremy Freeman, J. Anthony Movshon & Eero P. Simoncelli. “Selectivity and tolerance for visual texture in macaque V2”. In: *Proceedings of the National Academy of Sciences* (2016), p. 201510847.