Modeling Spatial Integration and Normalization Underlying Motion Processing in Receptive Fields of MT Neurons

Zhaojie Yao

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Wyeth Bair, Chair

Amy Orsborn

Eric Chudler

Greg Horwitz

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Modeling Spatial Integration and Normalization Underlying Motion Processing in Receptive Fields of MT Neurons

Zhaojie Yao

Chair of the Supervisory Committee:
Wyeth Bair
Biological Structure

Information processing in sensory neural systems relies on a hierarchical architecture. The peripheral sensory organs encode physical inputs as neural signals, which propagate through a multi-stage network. Progression in the network is associated with increasingly abstract representations, more complex functionality, and larger receptive fields. A classic example is the visual motion processing pathway in primates. The abstraction of motion information initiates when the spatiotemporal characteristics of 1D contours are extracted locally in the primary visual cortex (V1). The resulting neural encoding goes through complex processing to generate a cohesive representation of motion directions of surfaces and objects in the mid-level cortical motion area (MT) through integration of convergent direct and indirect V1 projections. Such integration is spatial – signals from many spatially offset V1 receptive fields combine to form

MT receptive fields that have diameters up to ten-fold larger. Strikingly, most modeling studies of MT signal processing have overlooked the profound impact of spatial integration in favor of explaining motion computation related to simple, isolated, and spatially uniform visual stimuli, leaving a gap in our understanding of the implications and mechanisms of spatial integration in mid-level motion processing. To address this issue, I have implemented the first image-computable MT model that includes realistic spatial integration of V1 complex units in a flexible framework. The response of the model to a series of dynamic, compound motion stimuli has led to insights into the spatial processing of motion signals in MT. (1) A random wiring paradigm for V1-MT connectivity can mostly account for the heterogeneity of the sensitivity profile of MT receptive fields observed *in vivo*, but falls short in explaining the diversity of motion direction selectivity reported in experimental data. (2) The well-established non-linear responses of MT neurons to two spatially offset motion stimuli within their receptive fields can be largely explicated by a cascade of mechanisms spread across model stages, including V1 surround suppression, spatially dependent signal weighting and nonlinear signal integration of inputs arriving in MT, whereas MT population-level normalization is less likely to play a role. (3) The spatial scale of the MT neuron's ability to integrate across local component motions to resolve the true direction of motion of a stimuli is not limited by surround suppression, as proposed in the literature, but is more likely to depend on opponent direction interactions at the V1 level. These findings highlight the intricate interplay between spatial and motion integration in the MT receptive field, and shed some light on the computational processes and the hierarchical architecture of the visual cortex.

Modeling Spatial Integration and Normalization Underlying Motion Processing in Receptive Fields of MT Neurons

Zhaojie Yao

Abstract

Information processing in sensory neural systems relies on a hierarchical architecture. The peripheral sensory organs encode physical inputs as neural signals, which propagate through a multi-stage network. Progression in the network is associated with increasingly abstract representations, more complex functionality, and larger receptive fields. A classic example is the visual motion processing pathway in primates. The abstraction of motion information initiates when the spatiotemporal characteristics of 1D contours are extracted locally in the primary visual cortex (V1). The resulting neural encoding goes through complex processing to generate a cohesive representation of motion directions of surfaces and objects in the mid-level cortical motion area (MT) through integration of convergent direct and indirect V1 projections. Such integration is spatial – signals from many spatially offset V1 receptive fields combine to form MT receptive fields that have diameters up to ten-fold larger. Strikingly, most modeling studies of MT signal processing have overlooked the profound impact of spatial integration in favor of explaining motion computation related to simple, isolated, and spatially uniform visual stimuli, leaving a gap in our understanding of the implications and mechanisms of spatial integration in mid-level motion processing. To address this issue, I have implemented the first image-computable MT model that includes realistic spatial integration of V1 complex units in a flexible framework. The response of the model to a series of dynamic, compound motion stimuli has led to insights into the spatial processing of motion signals in MT. (1) A random wiring paradigm for V1-MT connectivity can mostly account for the heterogeneity of the sensitivity profile of MT receptive fields observed *in vivo*, but falls short in explaining the diversity of motion direction selectivity reported in experimental data. (2) The well-established non-linear responses of MT neurons to two spatially offset motion stimuli within their receptive fields can be largely explicated by a cascade of mechanisms spread across model stages, including V1 surround suppression, spatially dependent signal weighting and nonlinear signal integration of inputs arriving in MT, whereas MT population-level normalization is less likely to play a role. (3) The spatial scale of the MT neuron's ability to integrate across local component motions to resolve the true direction of motion of a stimuli is not limited by surround suppression, as proposed in the literature, but is more likely to depend on opponent direction interactions at the V1 level. These findings highlight the intricate interplay between spatial and motion integration in the MT receptive field, and shed some light on the computational processes and the hierarchical architecture of the visual cortex.

Table of Contents

Chapter 1.

Introduction

A fundamental principle of visual cortical processing entails a hierarchy in which the neuronal receptive fields (RFs) progressively increase in size and functional complexity with depth in the network (Felleman & Van Essen, 1991). Despite this well-established principle, there is limited understanding of the mechanisms behind it, particularly at stages beyond the early integration of lateral geniculate nucleus inputs to form the RFs of simple cells in the primary visual cortex, V1. For instance, the direction-selective (DS) response of neurons in the cortical visual motion area, MT, is thought to arise largely from the integration of DS signals that originate in V1 (Dubner & Zeki, 1971; Movshon & Newsome, 1996; Nassi & Callaway, 2007). These V1 inputs are narrowly tuned for oriented 1D features (e.g., luminance edges) with specific spatial and temporal frequencies (SFs and TFs), and serve as spatially localized preliminary motion detectors (Movshon & Newsome, 1996). The subsequent spatial integration process creates MT RFs that are up to ten times as large in diameter as those of the V1 inputs (Albright & Desimone, 1987; Wang & Movshon, 2016).

Many past modeling studies of the V1-to-MT circuitry have set out to explain MT cells' sensitivity to visual motion, particularly, how *pattern motion* sensitivity, which arises in MT (Movshon et al., 1985), differs from *component motion* sensitivity in V1. Among them, the models of Simoncelli & Heeger (1998) and Rust et al. (2006) have set the standard for explaining how selectivity for the overall motion of a 2D pattern is computed from integrating the V1 signals that encode only the local motion of oriented 1D components within a moving target. However, like most modeling studies, they overlooked how V1-to-MT integration happens spatially, thus leaving an outstanding gap in our knowledge of the relationship between *motion integration* and *spatial integration*. However, several neurophysiological studies that used compound, spatially distributed stimuli to probe the MT RF have produced data that can only be accounted for by MT models that take on the important problem of spatial integration.

For instance, Richert and colleagues (2013), using stimuli of spatially scattered random dots traveling in various directions, raised questions about the spatial structure of V1-to-MT signal integration when they demonstrated that a considerable fraction of macaque MT neurons have heterogenous, multi-regioned RFs. This challenged the conventional view assumed by most MT models – that spatial summation of V1 outputs produces relatively homogeneous MT RFs, such that the MT cell prefers the same direction throughout its spatial RF (Livingstone et al., 2001), and that sensitivity is relatively high in the center but drops off toward the edge (Lagae et al., 1994; Raiguel et al., 1995). Their results suggest that V1-to-MT integration may involve spatially diverse and patchy inputs, and a more complex paradigm than the standard models have presumed.

Other studies have tried to investigate spatial integration in the MT RF by presenting multiple drifting sinusoidal grating stimuli simultaneously and analyzing the interaction between them. Britten and Heuer

(1999) found that there are normalizing and suppressive effects between two spatially separated gratings with identical motion in the MT RF. Majaj and colleagues (2007) observed that pattern motion sensitivity in MT cells breaks down when the component gratings of pseudo-plaids do not colocalize in the RF. These results cannot be reproduced by the standard MT models lacking realistic substructures within an overall RF. Thus, addressing spatial integration is crucial to studies that aim to create MT models that can process complex stimuli where the motion signal varies on a local level within the RF, and more generally, it is important to bridge the significant gap in our understanding of the role of spatial integration in the visual hierarchy.

In this thesis, I will present the first image-computable MT model with spatial integration of convergent V1 inputs. By controlling the topographic map of the V1 units, and the corresponding signal integration architecture, I will demonstrate the potential RF substructures and computational mechanisms that can account for MT responses observed *in vivo* to a series of stimuli where the distribution of motion signals is spatially nonuniform.

A brief summary of the following chapters and the main results is as follows:

- In **Chapter 2**, I review the pertinent literature and provide the general background to motivate the investigations conducted for this thesis.
- In **Chapter 3**, I introduce the modeling methods employed, including the processing of video inputs, and the architecture of the integrative network and the corresponding computation. Methods that are involved in specific investigations are documented in detail in the relevant individual chapters.
- In **Chapter 4**, I construct an MT model to examine the extent to which a random wiring hypothesis can explain the irregularity observed in the MT RF. I found that while random wiring of inputs can account for spatial heterogeneity at low V1 input numbers, it falls short in explaining the diversity in direction tuning observed across the MT RF reported in the literature. Additionally, I discuss the shortcomings of the random stimulus mapping method that was used in the original *in vivo* study, and propose alternative strategies for mapping RF homogeneity.
- In **Chapter 5**, I explore a variety of spatial normalization schemes to test if they can account for the nonlinear interaction within the MT RF. I found that three classical computations in visual processing: surround suppression, nonlinear input integration and divisive population normalization can all contribute to the nonlinearity of responses to spatially offset concurrent stimuli in the MT RF. However, each of these potential mechanisms endows the MT response with particular characteristics that bear on its plausibility to account for *in vivo* data. I conclude that nonlinear input integration is the best single candidate mechanism, but a combination of mechanisms is likely at play. In doing so, I reveal a novel, non-trivial relationship between spatial normalization and the RF size, from which one may reach paradoxical conclusions based on different classical measuring techniques.
- In **Chapter 6**, I examine the spatial integration underlying pattern motion sensitivity in MT, focusing on resolving the open question in the literature as to whether the spatial scale of motion integration is determined in V1, particularly by V1 surround suppression, or whether it is set in MT by specialized local processors of V1 inputs. I provide strong evidence that suggests that V1 surround suppression is not

involved, but rather V1 motion opponency is a likely and more parsimonious mechanism at play than specialized MT motion integration subunits.

- In **Chapter 7**, I discuss the implications of my results on spatial and motion integration, and their interrelationship. I also cover the limitations of the model and advise on future experiments.

Chapter 2.

Background: physiology and models of MT neurons

The visual system is frequently conceptualized by neuroscientists as a hierarchical flow chart (see Figure 4 in Felleman & Van Essen, 1991). Environmental inputs are converted to neural signals at the front end of the system – photoreceptors. Such signals then travel sequentially through layers in the retina and subcortical structures, and then from one cortical area to another that specializes in analyzing different features of increasing complexity. Although significant progress has been made in identifying the physiological specialization of various cortical regions within the major streams of this processing hierarchy, works that investigate the circuitry mediating such information flow are relatively lacking, particularly in the dorsal visual stream.

The dorsal pathway has long been known to specialize in the processing of visual motion. The network architecture for the elaboration of motion computations along this pathway remains largely unknown. The current consensus holds that the initial step of cortical motion processing starts at V1, where the cells extract preliminary local motion signals from oriented 1D components in dynamic visual scenes. V1 afferents then send the outputs to MT, which consequently calculates the overall pattern motion of the 2D scene within the constraints of the MT RF.

Two types of signal integration are essential for performing such computation. First, in many cases, the true motion of a complex object differs significantly from that of its oriented local features. Thus, the MT neuron must combine a range of V1 signals representing the motion of individual components across a multi-dimensional parameter space: directions, SFs and TFs – I will refer to this as *motion integration*. Second, because different components of a complex object sometimes do not colocalize, the MT neuron also must merge the outputs originating from multiple spatially dispersed V1 RFs – *spatial integration*. Working together in estimating the true motion of objects and surfaces in visual scenes, these two kinds of signal integration are essentially two sides of the same coin. Many researchers have debated the computational scheme for motion integration, while spatial integration has received much less attention, though, being equally important, if not more. In this chapter, I will delineate several questions my modeling research aims to address in light of evidence and implications from past neurophysiological and computational studies.

2.1. Motion processing in MT

Located in the posterior bank of the superior temporal sulcus, a direct V1 axonal projection zone was first identified in the macaque (Zeki, 1969; Dubner & Zeki, 1971). Around the same time, a similar area was also identified in the posterior third of the middle temporal gyrus of the owl monkey, hence the name MT (Allman & Kaas, 1971). The cells in MT are relatively insensitive to form and color but show strong selectivity for the

direction and speed of moving stimuli (Dubner & Zeki, 1971; Zeki, 1974b; Maunsell & Van Essen, 1983a; Albright, 1984; Zeki, 1987; Albright, 1992). Thus, MT was assigned the functional specialty of visual motion processing (Dubner & Zeki, 1971; Zeki, 1974b).

Since the inputs MT receives from V1 are also highly direction and speed selective (Orban et al., 1986; Movshon & Newsome, 1996), it is conceivable that MT neurons may simply inherit these properties from V1 cells. Then, what does MT add in the dorsal pathway to the motion processing ability that already exists in V1?

Because the MT neuron's RF is much larger compared to that of the V1 cell, one might speculate that MT's role is to compute motion on a larger spatial scale than V1 does. Seemingly parsimonious, many investigators were attracted to this idea. Mikami and colleagues (1986) measured the spatial limit of directional interaction of V1 and MT neurons using bar stimuli with apparent motion. They found that the maximum spatial interval of interaction for MT neurons seems to be about three times as long as that for V1 cells. On the other hand, when Churchland and colleagues (2005) revisited this issue, they concluded that the upper limit for spatial interaction is very similar in MT and V1 neurons. In fact, using dot stimuli, Livingstone and colleagues (2001) demonstrated that the spatial interaction limit for MT neurons is extremely small, only less than 1°, which is comparable to that of V1 cells (Livingstone & Conway, 2003). However, none of these studies took into account that the spatial limit of interaction measured using certain stimuli may depend on the spatial scale of such stimuli. Nevertheless, there has been no sufficient evidence to suggest that MT cells process visual motion in a larger spatial range than V1 neurons do. At the same time, there is compelling evidence that MT does not process higher-level motion (Hedges et al., 2011), but primarily functions as short-range motion processors (Braddick, 1974).

Another view of MT neurons' function, being generally accepted, is that many of them are more concerned with signaling the true motion of a visual pattern rather than the local motion of 1D components making up that pattern. Indeed, the enlarged RFs in MT should allow them to use the motion cues of different 1D components across space to compute a coherent velocity of the object as a whole.

In order to discuss how MT processes pattern motion, I will first define the computational challenge at the heart of this problem. For any rigid moving object, all parts within it should share a common physical velocity, but for the visual system, the perceived velocity of different components may not be consistent due to what is termed the *aperture problem* (Fennema & Thompson, 1979). When a moving straight line is viewed through a small aperture, regardless of the overall velocity of the motion, only the velocity component orthogonal to the line is detectable to the observer. **Fig. 2-1** illustrates such visual motion ambiguity of a square translating towards the bottom right when only partial views of the outline are seen through small windows. The actual motion can be determined if the window encompasses the corner. However, if the window is positioned on the edge, only the velocity component perpendicular to the edge can be resolved. Therefore, the perceived local velocity of different parts of an object depends on the orientation of their local contour. This creates a problem for the visual system – as visual motion signals are first extracted by V1 neurons, their small RFs restrict them from accurately estimating the real velocity of the object.

Despite being called the aperture problem, the true challenge of this issue lies in the 1D nature of such visual inputs, such as a straight line, or any infinitely extending contour whose spatial intensity function only varies in a single dimension. For example, the motion direction of a vertically oriented grating is also ambiguous as motion in any direction will appear as horizontal motion. When multiple 1D contours moving in their orthogonal directions form a pattern, instead of isolating the motion of individual components, the visual system can register the overall motion direction of the pattern. For instance, if a vertical grating drifting to the right overlaps with a horizontal grating translating upwards at the same speed, forming a plaid, the perceived pattern motion will be diagonal towards the upper right. If such a translating plaid is presented to a DS V1 cell, the direction tuning curve of that cell will have two peaks (Movshon et al., 1985), each produced by the instance when the motion direction of one of the component gratings aligns with the preferred direction (PD) of the cell (**Fig. 2-2**). This is because, computationally, V1 cells approximately function as linear band-pass spatiotemporal filters (Reid et al., 1987; McLean & Palmer, 1989). Owing to the linear nature of V1-level computation, the response of V1 cells to plaids will approximate the sum of the responses to its individual components, hence rendering two peaks in the direction tuning curve.

When tested with a moving plaid, some MT neurons cannot overcome the aperture problem and respond similarly as V1 cells. Movshon and colleagues (1985) found that MT neurons display a spectrum of plaid direction tuning curve shapes. At one extreme, some MT neurons' turning curves are bimodal, similar to those of V1 cells. On the other end, some MT neurons' tuning curves have only one peak, corresponding to the instance when the overall direction of the moving plaid aligns with the MT cell's PD. They called the first type of neurons *component direction selective* (CDS) cells, and the latter *pattern direction selective* (PDS) cells. They reported 25% of the MT neurons are significantly PDS, and 40% are CDS.

2.2. The hierarchical contribution of V1 inputs to MT

First enunciated by Hubel and Wiesel (1962, 1965), the principle of hierarchy is reflected in all aspects of the organization of the visual system – the anatomy, physiology and functions; the projections from V1 to MT are no exception.

2.2.1. Motion integration

MT receives inputs from many of the DS and orientation selective neurons in Layer 4B and upper Layer 6 of V1 (Lund et al., 1975; Ungerleider and Desimone, 1986; Shipp and Zeki, 1989; Nassi & Callaway, 2007), where motion information of 1D features is extracted. To correctly compute the pattern motion, the MT neuron must recruit component signals from a population of V1 neurons. Although making up a small fraction of the 5,000 ~ 10,000 synapses that form on a typical MT neuron, the number of the V1 synapses are nevertheless on the order of hundreds, and these V1 connections are likely significantly stronger than the inputs from other cell types (Anderson et al., 1998).

2.2.1.1. Models of the V1 cell

Since the 1980s, many investigators have proposed different architectures of V1-to-MT circuitry for motion processing. Although, the mechanisms they suggest accounting for the integration of V1 signals by the MT cell often differ, they all simulated V1 neurons as DS filters tuned for particular SFs and TFs. Because 2D motion can be characterized as an oriented structure in the *x*-*y*-*t* space, motion information can be captured by spatiotemporally oriented filters (Adelson & Bergen, 1985). Arguing that a physiologically realistic motion detector should only be built upon units that are spatiotemporally separable, Adelson and Bergen (1985) noted such filters can be constructed from spatial and temporal filters in the following manner:

$$h_{\mathrm{st}}(x, y, t) = h_{\mathrm{s}}(x, y) h_{\mathrm{t}}(t) \qquad \textbf{Equation 2-1}$$

where, for the 1D case, they chose $h_{\mathrm{s}}(x, y) = h_{\mathrm{s}}(z)$ to be the negative 2nd derivative or the reflected 3rd derivatives of a Gaussian function, $g(z)$, here denoted by $h_{\mathrm{s}}(z) = g_2(z) = -\ddot{g}(z)$ or $h_{\mathrm{s}}(z) = g_3(z) = \dddot{g}(-z)$, respectively (**Fig. 2-3A**). z is an intermediary spatial variable; given α as the angle of orientation,

$$z = x \cos\alpha + y \sin\alpha \qquad \textbf{Equation 2-2}$$

From **Eq. 2-1**, $h_{\mathrm{t}}(t)$ takes the form of the following:

$$f(t) = \left(\frac{1}{n!} - \frac{(kt)^2}{(n+2)!}\right) (kt)^n \exp(-kt) \qquad \textbf{Equation 2-3}$$

where n is 3 or 5. In either case, the filter is denoted by $f_3(t)$ or $f_5(t)$, respectively (**Fig. 2-3B**). By selecting one from each of the two spatial and temporal filters, the multiplicative combination of the filters rendered four separable spatiotemporal filters:

$$\begin{aligned} h_{2,3}(x, y, t) &= g_2(z) f_3(t) \\ h_{2,5}(x, y, t) &= g_2(z) f_5(t) \\ h_{3,3}(x, y, t) &= g_3(z) f_3(t) \\ h_{3,5}(x, y, t) &= g_3(z) f_5(t) \end{aligned} \qquad \textbf{Equation 2-4}$$

By taking appropriate sums and differences of the outputs of the above linear filters, they then constructed two pairs of motion detectors tilted in either direction orthogonal to α (I use H to represent the response of filter h):

$$H_{1-1}(x,y,t) = +H_{3,3}(x,y,t) + H_{2,5}(x,y,t)$$
$$H_{1-2}(x,y,t) = -H_{3,5}(x,y,t) + H_{2,3}(x,y,t)$$
$$H_{2-1}(x,y,t) = -H_{3,3}(x,y,t) + H_{2,5}(x,y,t)$$
$$H_{2-2}(x,y,t) = +H_{3,5}(x,y,t) + H_{2,3}(x,y,t)$$

Equation 2-5

where $H_{i-1}(x,y,t)$ and $H_{i-2}(x,y,t)$ are a pair of impulse responses tilted in the same direction and approximately in quadrature (2D representations in the z-t plane are shown in **Fig. 2-3C**), whose importance will be explained later.

Acknowledging that the spatial and temporal responses of the V1 neuron are largely separable (Ikeda & Wright 1975; Tolhurst & Movshon 1975; Holub & Morton-Gibson 1981), Adelson and Bergen's (1985) approach to construct inseparable impulse responses using separable filters is very elegant. However, the link between the parameters in their formulation, and the central SFs and TFs of the unit is not straightforward. Thus, while some MT models adopted their spatial or temporal filters (Nowlan & Sejnowski, 1994, 1995; Tsui et al., 2010), most opted for Gabor filters as their V1 modules (Heeger, 1987, 1988; Grzywacz & Yuille, 1990; Simoncelli & Heeger, 1998; Rust et al., 2006; Baker & Bair, 2016, 2017).

A typical odd-phased 2D Gabor filter is defined as follows:

$$g(x,y,t) = \frac{(2\pi)^{-\frac{3}{2}}}{\sigma_s^2 \sigma_t} \exp\left(-\frac{x^2+y^2}{2\sigma_s^2}\right) \exp\left(-\frac{t^2}{2\sigma_t}\right) \sin(2\pi\omega_s(x\cos\theta + y\sin\theta) + 2\pi\omega_t t)$$

Equation 2-6

which is a sine wave multiplied by a pair of spatial and temporal Gaussian windows (Daugman, 1980, 1985). Here, θ indicates the direction of preferred motion; σ_s and σ_t are the spreads of the spatial and temporal Gaussian windows, respectively; ω_s and ω_t are the central SF and TF, respectively, giving the SFs in the x-y space as follows:

$$\omega_x = \omega_s \cos\theta$$
$$\omega_y = \omega_s \sin\theta$$

Equation 2-7

Similarly, an even-phased Gabor filter can be defined using a cosine wave as follows:

$$g(x,y,t) = \frac{(2\pi)^{-\frac{3}{2}}}{\sigma_s^2 \sigma_t} \exp\left(-\frac{x^2+y^2}{2\sigma_s^2}\right) \exp\left(-\frac{t^2}{2\sigma_t}\right) \cos(2\pi\omega_s(x\cos\theta + y\sin\theta) + 2\pi\omega_t t)$$

Equation 2-8

Together, they form a pair of filters in quadrature (the 2D representations in the z-t plane are shown in the left and middle panels of **Fig. 2-3D**). Compared to Adelson and Bergen's (1985) filters, the Gabor filters' central SF and TF are directly defined by the parameters in the formulae, making them easier to interpret. In the meantime, the separability of SF and TF tuning curves of the response are preserved through separable Gaussian envelopes in space and time.

Both Adelson and Bergen's (1985) filters and the Gabor filters are constructed as a pair in quadrature, since the phase sensitivity of a single filter can cause its response to a translating pattern to depend on the pattern's instantaneous line-up to the RF. Thus, as the pattern shifts in and out of phase with the filter, the time course of the output is not stable even though the stimulus is a stationary (constant velocity motion) signal. This property resembles the phase-modulated response of a simple cell to drifting gradings (Ibbotson et al., 2005). Simple cells have segregated subfields that detect either brightness increments (ON) or decrements (OFF) (Hubel & Wiesel, 1962; Henry, 1977). Therefore, as a grating is moving across the RF of the simple cell, which can be modeled as a single Gabor filter, the response will show temporal modulation as the grating matches and mismatches with the ON-OFF substructures of the RF, which is not consistent with human experience of constant motion. Moreover, the sign of the response also depends on the polarity of the stimulus contrast, so that a black stimulus and a white stimulus with the same motion will give inverted responses.

On the other hand, complex cells are far less phase-sensitive than simple cells (Hubel & Wiesel, 1962; Henry, 1977; Hietanen, 2013), and the RFs of the V1 cells that project to MT are dominantly complex (Movshon & Newsome, 1996). Such a feature can be modeled by summing the squared outputs of a pair of (even and odd-phased) filters in quadrature. In the case of Gabor filters, one can square the outputs of each filter to take the advantage of $\sin^2\phi + \cos^2\phi = 1$, and such a unit's response to a constant motion input will be positive and stable, i.e., relatively unmodulated by the phase of the drifting grating.

Adelson and Bergen (1985) referred to the models of such complex cell-like motion detectors as the *motion energy filters*. The spectrum of the filters in such a motion detector lies diagonally across the origin in Fourier space, as a pair of blobs centered at certain spatiotemporal frequency, which in the case of the Gabor motion energy filter, is determined by $(\omega_s \sin\theta\,, \omega_s \cos\theta\,, \omega_t)$ (**Eq. 2-7**), corresponding to the central SFs on the x and y axes, and the central TF of the filters. A representation in the Fourier z-t plane is shown in **Fig.2-3D** (right panel): the spectral energy along the line connecting the two blobs corresponds to the energy of motion at a given velocity across different SF and TF. The motion energy filter can extract the energy of the input signal at the specified spatiotemporal frequency band. Compared to Adelson and Bergen's (1985) formulation, the Gabor energy filter's spectrum is cleaner due to its relatively more ideal mathematical properties.

2.2.1.2. Models of the V1-MT connection

Given that CDS MT cells' motion direction tuning curves are generally similar to those of V1 cells' in shape, their behavior can typically be explained by aggregating the outputs of a group of motion energy filters with identical direction selectivity. Thus, the primary distinction among most MT models in the literature is how they

construct the V1-MT circuitry that gives rise to pattern direction selectivity, specifically, the selection and combination of distinctively parameterized V1 afferents that are fed into a PDS MT unit. Here, I will discuss the V1-MT connection and computation in a group of models (Heeger, 1987, 1988; Grzywacz & Yuille, 1990; Nowlan & Sejnowski, 1994, 1995; Simoncelli & Heeger, 1998) grounded in what Born and Bradley (2005) termed the *F-plane* principle. The underlying concept of such models can be better understood when viewed through the lens of visual motion in the Fourier domain (Watson & Ahumada, 1983, 1985).

For a translating sinusoidal grating, its spatiotemporal dynamics can be adequately defined by its SF-wavelength (ω_s, λ), and TF-period (ω_t, T). The velocity, $\mathbf{v}$, of such travelling waveform can be inferred as the product of TF and wavelength,

$$|\mathbf{v}| = \lambda\omega_t = \frac{\omega_t}{\omega_s} \qquad \textbf{Equation 2-9}$$

giving us the ratio between TF and SF. In the frequency domain, this corresponds to the slope of the line connecting the pair of points that the spectrum of the grating resides in and are symmetric about the origin.

Therefore, for a visual pattern, which consists of many frequency components, translating at velocity $\mathbf{v}$, there is a one-to-one correspondence between its SF and TF components:

$$\omega_t = |\mathbf{v}|\omega_s \qquad \textbf{Equation 2-10}$$

This means, such a visual pattern's spectrum lies within a single line passing through the origin with a slope equal to $|\mathbf{v}|$. If we extend the analysis of velocity to the 2D case, $\mathbf{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix}$, the spectrum of such a pattern lies within a single plane defined by the following equation (Watson & Ahumada, 1985; Heeger, 1987, 1988):

$$\omega_t = v_x\omega_x + v_y\omega_y \qquad \textbf{Equation 2-11}$$

Based upon on such inference, Heeger (1987, 1988) suggested that one can estimate the velocity of a moving image by finding the plane where its spectrum resides. The azimuth, $\text{atan2}(v_y, v_x)$, and the elevation, $\text{atan2}\left(1, \sqrt{v_x^2 + v_y^2}\right)$, of the plane reflect the speed and direction of the image motion, respectively (Bradley & Goyal, 2008; Nishimoto & Gallant, 2011). Similarly, given a population of motion energy filters – models of complex V1 units, if we compare the observed normalized outputs to the input visual pattern, and the theoretical prediction of those to an ideal spatially-white random pattern moving in a particular velocity, we can build a model PDS unit tuned for that velocity by establishing a negative correlation between the unit's response and the difference between the observed and predicted outputs of the motion energy filters.

Heeger (1987, 1988) built a PDS model based on such principles using Gabor filters, as their responses to the above-mentioned ideal random moving pattern can be easily calculated due to the pattern's energy being

constant on the velocity-specified plane in the Fourier domain, and zero everywhere else. The group of Gabor filters in the model are tuned for identical SF and TF magnitudes but different directions. The spectral representations of these filters form a pair of rings. Each point on the ring corresponds to a filter of a specific direction, with the radius of the ring equal to the magnitude of the central SF of the filters, while the distance to the zero-TF plane equal to the magnitude of the central TF. He also included an extra group of filters, with the same SF and direction preferences as the former group but tuned for stationary images (i.e., the central TF is zero), forming a ring on the zero-TF plane. The spectrum of all the filters of the model creates a cylinder in the Fourier domain (see Figure 12 in Heeger, 1988). Normalization can be done by dividing the response of a filter by the sum of responses of the filters tuned for the identical SF, i.e., three filters tuned for the same SF, but for positive, zero and negative TFs, making up a vertical column on the lateral surface of the cylinder.

Some critics have pointed out, however elegantly formulated, Heeger's (1987, 1988) model assumed that the PDS MT neuron is capable of comparing the motion energies of a given input to that of a spatially-white random pattern translating at the neuron's preferred velocity, which involves mathematical operations that may not be easy to implement biologically (Grzywacz & Yuille, 1990). Alternative MT models use a population of complex V1 units to sufficiently tile the frequency domain, and the outputs of these units are linearly summed with weights that are inversely proportional to the V1 units' distance to the plane of the MT unit's preferred velocity, so that the MT unit responds most vigorously when motion energy is captured by V1 detectors located on that plane (Grzywacz & Yuille, 1990; Nowlan & Sejnowski, 1994, 1995). Other MT models simply integrate signals exclusively from motion energy filters that reside on that velocity plane (Simoncelli & Heeger, 1998), and the findings of more recent animal studies have lent credibility to such models. For instance, Simoncelli and Heeger's (1998) model successfully explains why, unlike for plaids, CDS cells respond to the average direction of superimposed dot fields, whereas PDS cells can signal the constituent motions (McDonald et al., 2014). Using enhanced motion stimuli, Nishimoto and Gallant (2011) also confirmed that the excitatory spectral RFs of some PDS cells indeed reside on a velocity plane, partially occupying a tilted iso-SF ring.

The MT models discussed above aimed to build a PDS unit that are selective for a particular visual motion velocity. Hence, their construction centers around the idea that PDS MT neurons' computational task is to either sample or infer the motion energy on a particular preferred-velocity plane. However, this is by no means the only viable approach. Rust and colleagues (2006) have suggested that signal integration of V1 units with diverse direction preferences and strong opponent motion suppression, along with fine-tuned normalization, may be the key to PDS circuitry. Building upon this idea, Baker and Bair (2016) demonstrated that PDS direction tuning can be obtained through a model that samples motion energy from a group of V1 channels with the same central frequency range but in different directions; their spectrum forms a single ring in the Fourier domain; the PD of the model is determined by the weights applied to the direction channels. Because these channels do not reside on an iso-velocity plane, although these models can sufficiently represent the direction selectivity of a PDS cell, they lack speed selectivity. However, speed selectivity may not be as prevalent in MT as previously assumed. When tested with sinusoidal gratings, about 75% of the MT cells do not maintain

consistent speed tuning as the spatial frequency of the stimuli varies, although speed selectivity does become more common when tested with plaid stimuli (Priebe et al., 2003).

Others have suggested that the integration of V1 channels selected for a range of directions may not be necessary at all for building a PDS model. The V1 units in Tsui and colleagues' (2010) model all share the same PD; their results indicate that nonlinear mechanisms, such as surround-suppression (end-stopping) at the V1 level and softmax pooling at the MT level, may play a crucial role in the MT cell's ability to overcome the aperture problem. Their model can resolve the true motion of a finite-length bar moving in a tilted (not perpendicular to the orientation of the bar) direction. Nevertheless, such pattern motion sensitivity falls apart when their model is tested with moving plaids.

2.2.2. Spatial integration

Besides summing inputs from V1 neurons to generate a DS output, the MT cell also integrates signals from a group of V1 neurons with spatially dispersed RFs, endowing the MT neuron with an RF that is much larger than those of the V1 cells (Albright & Desimone, 1987). This is the result of converging anatomical inputs (Zeki, 1971). Here, I will first focus on the analysis of the spatial substructure of the MT RF in the literature. Then, I will review studies on the interaction of responses elicited by concurrent presentation of multiple stimuli at different locations within the MT RF, and the corresponding implications about the circuitry underlying V1-to-MT spatial integration.

2.2.2.1. MT receptive field substructure: homogeneous or heterogeneous

There has been a commonly held belief that spatial summation within the MT RF should exhibit homogeneity. A stereotypical RF are thought to have an approximately Gaussian shape. The sensitivity is high at the center of the RF and rolls off uniformly towards the edge, and direction selectivity should remain consistent across the RF. Although observations of such MT RFs have been reported in some literature, they were not always warranted thorough quantitative examination.

While the size of the MT RF has always remained the focus of the investigators who mapped the spatial RF, past surveys of the shape of the RF, including both the 2D geometry of the outline and the response sensitivity profile, have been mostly very crude. When Baker and colleagues (1981) manually mapped the RFs of the cells in the extrastriate areas, including MT, they made the observation that most RFs are elliptical, with some of them being rarely odd-shaped, such as very long and narrow rectangles or kidney-like shapes. However, their assessment is only qualitative. Ever since, most studies have assumed a 2D Gaussian model of the sensitivity profile of the MT RF and rarely contemplated more complex structures (Maunsell & Van Essen, 1983a, 1983b; Felleman & Kaas, 1984; Takana et al., 1986; Saito et al., 1989; Lagae et al., 1989, 1993). Lagae and colleagues (1994), as well as Raiguel and colleagues (1995), are the first groups of investigators that tried to examine the spatial sensitivity profile of the MT RF on a sub-RF scale and study its geometry. They presented patches of moving dots on a 5×5 grid that spans the classical (excitatory) RF (CRF), and drew the contour profile according

to the 25 response data points. They reported, albeit qualitatively, that the MT CRFs are generally elliptical, and the response is strongest in the center and gradually diminishes toward the edges. However, their stimuli were all moving in the global PD of the MT cell. It is possible that when measured with stimuli moving in directions other than the preferred, the response profile might take different shapes. In fact, Lagae and colleagues (1994) noticed that the peak location of the sensitivity profile varies as stimulus speed changes. On the other hand, evidence has suggested that the inhibitory surround that flanks the MT CRF often have irregular shapes (Xiao et al., 1995).

Most studies have not tried to examine direction selectivity in MT at the sub-RF level, as they usually use single-direction motion stimuli that span the entire RF, based on the seemingly parsimonious, yet not rigorously verified, presumption that direction selectivity should remain consistent throughout the RF. Some investigators have attempted to confirm the consistency of direction selectivity in MT RFs, but often only provided incidental findings. Livingstone and colleagues (2001) used sparse 2D white noise to map the MT RF substructure. They concluded that the direction preference is similar across the RF. Howbeit, they examined merely five MT cells, and only sampled a very limited number of locations within the RFs.

To some extent, the homogeneous view of MT RFs is a structural motif accepted without proper investigation until a meticulous analysis of the direction selectivity and spatial sensitivity profiles in MT at the sub-RF scale was conducted by Richert and colleagues (2013). They used the random dot stimulus to map the substructure of the MT RF. They found that nearly half of their recorded cells deviate from the classical view of MT RFs: direction preference depends the location within the RF, and the spatial response profile often has multiple peaks with apparent gaps in between. This indicates that the MT neuron is not simply tiling its RF evenly with subunits consisting of V1 cells and circuits sharing identical substructures. The mechanisms that contribute to the heterogeneity of such MT RFs have remained unclear, but some have speculated that the discrepancy of direction selectivity between the excitatory and inhibitory subfields, and their irregular spatial profiles, might give rise to heterogeneous local direction tuning in the MT RF (Cui et al., 2013).

2.2.2.2. Spatial interaction of multiple stimuli in the MT receptive field

As the convergence of V1-to-MT propagation enlarges the RF, the MT RF of primates will often encompass more than one moving object in a natural environment. This raises questions about how the MT cell processes multiple concurrent stimuli presented at different locations within the RF, and how the corresponding response relates to those elicited by the same stimuli presented individually. When Recanzone and colleagues (1997) recorded the responses of MT neurons to two dots moving in different directions within the RF, they found that the response is best approximated as the average of the responses to each dot presented alone. This indicates that spatial summation in MT is not a simple addition of signals across different locations within the RF. This calls for more sophisticated models to fully explain the complex mechanisms of spatial summation in MT.

Britten and Heuer (1999) compared the responses to a pair of Gabor stimuli presented simultaneously and separately at different locations within the MT RF. They found that on average, the observed response to the paired stimuli is less than the unscaled linear addition of single stimulus responses. They then tried to fit their results to the *power-law summation model*, formulated as follows:

$$r_{12} = a(r_1^n + r_2^n)^{\frac{1}{n}} + b \qquad \textbf{Equation 2-12}$$

where r_1 and r_2 are the responses to the individual stimuli, and r_{12} is the response to the pair of stimuli presented simultaneously. The scaling factor, a, the intercept, b, and the exponent, n, are all free parameters. **Fig. 2-4** shows **Eq. 2-12** graphed in the r_{12}-r_1-r_2 space. When $n = 1$, the summation is linear such that the equation lies on a slanted plane whose steepness is controlled by a (**Fig. 2-4B**). When $n < 1$, the plane of the equation starts to bulge, forming a convex surface (**Fig. 2-A**). When $n > 1$, the plane of the equation starts to curve inward, generating a concave surface (**Fig. 2-4C**). If n is extremely large, the model degenerates to $r_{12} \approx a \max(r_1, r_2) + b$ (**Fig. 2-4D**). Thus, the larger n gets, a stronger normalization effect will be observed, meaning that the response to a pair of stimuli is dominated by the stimulus that elicits the stronger response when presented individually, rather than a cumulative linear effect. In the unit square domain (the values of r_1 and r_2 are limited within $[0, 1]$), the lowest point of the summation function is b, which stands for the estimated maintained activity, and the highest point is $2^{\frac{1}{n}}a + b$. In the case of unscaled addition, $a = 1.0$, and $a = 0.5$ in the case of balanced averaging. For the Britten & Heuer (1999) data, the scaling factor is almost exactly halfway between averaging and addition, and the summation is on average modestly nonlinear, characterized by an exponent of 2.72.

They also examined how the response varies as a function of the stimulus locations (Britten & Heuer, 1999). They found that spatial wise, the summation is highly nonlinear in the sense that the response to the paired stimuli depends heavily on the stimulus more adjacent to the center of the RF, and moving the far stimulus even further from the center has minimum influence.

Britten and Heuer (1999) discussed how the parameters of the power-law summation model relate to MT circuitry models regarding spatial integration with power nonlinearity. For a MT cell that integrates the outputs from V1 units that are distributed across different locations within the RF, a family of nonlinear summation can be expressed as follows:

$$P = \left(\sum_i C_i\right)^{\frac{1}{n}} \qquad \textbf{Equation 2-13}$$

where C represents the output of a local V1 unit, and i parameterizes the location of the units. The exponent, n in **Eq. 2-12**, and the power, $\frac{1}{n}$ in **Eq. 2-13**, have an inverse relationship, corresponding to the "square of a sum"

in Simoncelli and Heeger's model (1998, Equation 5). Their "underlying linear response" (equivalent to P^n, $n = 0.5$, in the formulation here) depends on the weighted linear summation of V1 complex cell outputs, followed by half-wave rectification and squaring (Simoncelli & Heeger, 1998). This corresponds to $r_{12} = a(\sqrt{r_1} + \sqrt{r_2})^2 + b, n = 0.5$ for **Eq. 2-12**: the square root operation applied to each term in the sum recovers the underlying linear response, and the summed quantity is then squared. In the subsequent step, a divisive population normalization down-scales the rectified and squared underlying linear response as described as follows (Simoncelli & Heeger, 1998, Equation 6):

$$P'_j = \frac{P_j}{\beta_1 \sum_i P_i + \beta_2} \qquad \textbf{Equation 2-14}$$

where the normalized response of the *j*th MT cell, P'_j, is the result of the divisive suppression from a pool of MT cells, parameterized by i, with close-by RFs (β_1 is the parameter that controls the strength of the divisive normalization, and β_2 determines the semi-saturation level and avoids division by zero). Thus, a in **Eq. 2-12** should be less than 1.0, representing the lateral inhibitory effect of the normalizing pool of MT neurons, and reflects the strength of divisive normalization.

Simoncelli and Heeger's (1998) "square of a sum" and divisive population normalization is not the only spatial summation scheme that has been proposed in MT modeling research. Tsui and collogues (2010) have suggested that softmax pooling followed by sigmoid nonlinearity should be able to approximate the operation in MT indicated by past studies (Snowden et al., 1992; Pack et al., 2004). They chose a moderate level of nonlinearity for softmax pooling, corresponding to linear averaging with an amplification of the stronger V1 inputs relative to the weaker ones. They claimed that this is empirically similar to the kind of nonlinearity in Simoncelli and Heeger's (1998) model.

Although Britten and Heuer (1999) addressed spatial summation in MT, to some extent, their work lacks the sense of specificity to MT. Their experiment set a generic paradigm for examining spatial summation in visual neurons. However, the central function of MT, motion sensitivity, was overlooked, as the parameterization of visual motion was absent in their study. They did not consider the close tie between spatial summation and motion sensitivity in MT, which is also a common oversight in many studies that concentrate solely on motion integration in MT. The typical stimuli used to probe MT cells and assess their pattern motion processing ability are plaids of overlapping gratings drifting in different directions, unlike which, objects in the real environment often have moving contours separated in space. Like the rectangle in **Fig. 2-1**, if the encircled corner falls in the RF of a PDS MT cell, because the left and the bottom edges of the corner appear to move in different directions, signal integration has to be performed across space to compute the overall motion. As conventional stimuli used in early works on pattern direction selectivity in MT were often homogenous, full-field patterns, they did not reveal any property of the underlying spatial integration mechanisms, which can be made possible by stimuli with motion signals that vary on a local sub-RF scale.

To test if motion signals are integrated globally across the entire RF in MT, Majaj and colleagues (2007) measured the pattern direction selectivity of MT cells under two conditions: stimulation with conventional plaids (**Fig. 2-5A**) and with pseudo-plaids, where the two components of the plaid are segregated spatially (**Fig. 2-5B**). They discovered that pattern motion sensitivity breaks down with pseudo-plaids and the PDS cells became CDS. Therefore, they hypothesized that motion integration in the MT RF is not global, and it must operate on a limited local scale that is much smaller than the RF of MT, potentially facilitated by localized motion processing subunits.

To verify the hypothesis of Majaj et al. (2007), Perrone and Krauzlis (2008, Figure 2B) replicated their observation by using a model with spatial integration of local pattern motion detectors. The model consists of a 2D array of localized subunits distributed throughout the RF, each containing ten direction channels. They integrate local signals and compute pattern motion independently of subunits positioned elsewhere. Each subunit receives excitatory inputs from five channels within the 180° range flanking the PD of the subunit. The speed preferences of these channels are determined by the cosine of the difference between the PD of the subunit and the direction of the channel. The subunit also receives inhibitory inputs from five direction channels within the 180° of the anti-PD of the subunit, and the speed selectivity of these channels also follow a cosine relationship with respect to the anti-PD. The combined signals from the excitatory and inhibitory channels are half-wave rectified, and the signals from all subunits are summed. Essentially, these subunits are velocity detectors that endow the model with local motion opponency. Because suppressive motion opponency is one of the key circuit elements for pattern motion sensitivity (Rust et al., 2006), the breakdown of pattern motion sensitivity to pseudo-plaids can be explained by the silenced inhibitory inputs. For a normal plaid stimulus, because the two gratings of different directions colocalize, the activations of the excitatory and inhibitory channels are balanced within the local subunit. When the two components are spatially separated, and fall within the detectable ranges of different subunits, the inhibitory contribution of one of the components will be filtered out by half-wave rectification.

On the other hand, local motion processing subunits may not be the only possible circuitry that can account for Majaj and colleagues' (2007) observation. Their results may also be related to the limited spatial extent of V1 surround suppression, which is also critical to pattern motion processing (Tsui et al., 2010). Kumbhani and colleagues (2011) examined the spatial limit of pattern motion integration in MT using multi-patch pseudo-plaids on a grid. The spatial limit they found is somewhat larger than the V1 CRF, but comparable to the size of V1 surround (Angelucci et al., 2002; Cavanaugh et al., 2002). Such surrounds possibly provide the source for tuned normalization in V1, which is believed to be critical to the PDS circuit in MT (Rust et al., 2006; Tsui et al., 2010). Since patches of different motions alternates in the pseudo-plaid, the CRF and the surround of the V1 units may not be activated in a balanced manner, leading to weakened or abolished tuned normalization and a decline in pattern motion sensitivity.

2.3. Summary

The review of the systems neuroscience of spatial and motion integration in MT curated here is only a selective representation of the vast body of important research. Historically, efforts to study the MT functionality and the corresponding neural circuits have primarily focused on how the MT cell can integrate motion signals that originate from V1 neurons across different features, e.g., frequencies and directions, and the work on the spatial aspect of the V1-MT connection has more or less stayed out of the spotlight. Thus, I tried to examine the relatively limited literature pertaining to the spatial substructure of the MT RF and signal integration. Although spatial integration in MT may seem like a more trivial aspect compared to motion integration, it lies in the very basis of pattern motion sensitivity: the MT cell has to integrate signals from multiple V1 channels across the RF to accurately detect the true motion of an object, and pattern motion processing only operates within a limited spatial range. Given the lack of work in this area, substantial future research is needed for us to have a more comprehensive understanding of the V1-MT circuitry.

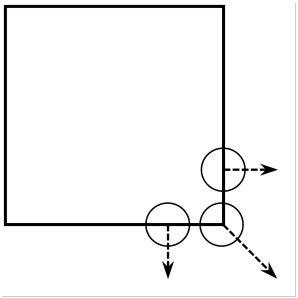

Figure 2-1 The aperture problem

The global velocity of a moving pattern cannot always be detected through an aperture. The true motion of the square is towards the bottom right, which can be detected if viewed through an aperture that includes the corner. When viewed through an aperture located on the edge, only the velocity component perpendicular to the edge can be detected.

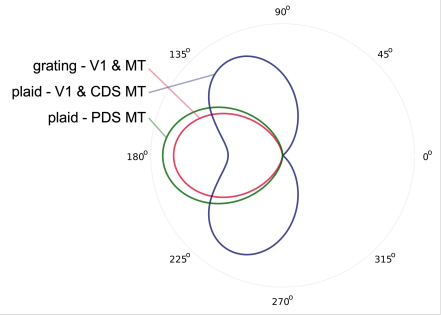

Figure 2-2 Direction tuning curves of direction selective neurons

Conceptualized direction tuning curves of direction selective visual neurons are shown. The responses of V1 and MT cells to drifting gratings should resemble the red curve, and has a single optimal direction; the responses of V1 and component MT cells to drifting plaids should resemble the purple curve, and does not have a single optimal direction; the responses of pattern MT cells to drifting plaids should resemble the green curve, and has a single optimal direction.

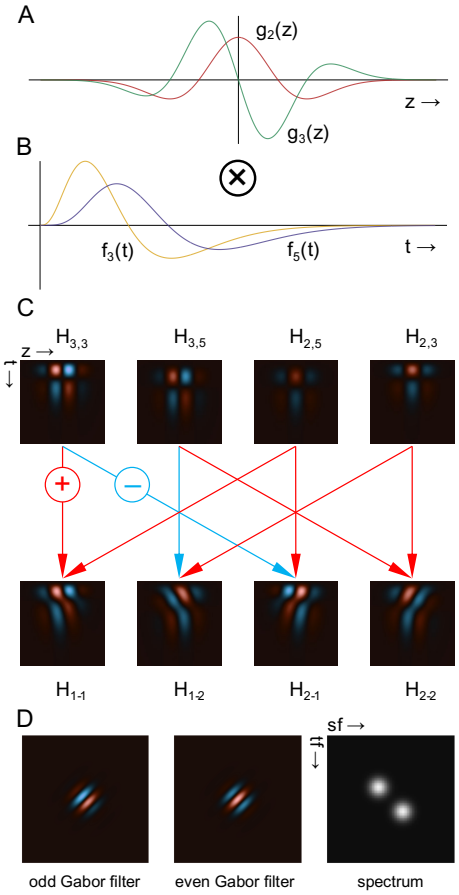

Figure 2-3 Oriented spatiotemporal filters of V1 models

(A) Two spatial filters that can be used to construct oriented spatiotemporal filters. $g_2(z)$ (red curve) and $g_3(z)$ (green curve) are the negative second and reflected third derivatives of the Gaussian function. **(B)** Two temporal filters that can be used to construct oriented spatiotemporal filters. The formulation of $f_3(t)$ (yellow curve) and $f_5(t)$ (purple curve) is described in ***2.2.1.1. Models of the V1 cell.*** **(C)** The construction of oriented spatiotemporal filters. The first row shows the impulse responses of the filters constructed as the product of the spatial and temporal filters shown in **(A)** and **(B)**, respectively. The horizontal axis corresponds to the spatial axis; the vertical axis corresponds to the temporal axis. The label, H_{ij}, indicates the choice of the spatial filter *i* and the temporal filter *j*. These filters are then combined through addition and subtraction to generate a pair of rightward-selective impulse responses (H_{1-1} and H_{1-2}) and a pair of leftward-selective responses (H_{2-1} and H_{2-2}). The members of a pair are approximately in quadrature (adapted from Adelson & Bergen, 1985). **(D)** Some V1 models adopt the motion energy filter, which are constructed through the sum of squares of a pair of odd- and even-phased Gabor filters shown in the two panels on the left. The horizontal axis corresponds to the spatial axis; the vertical axis corresponds to the temporal axis. The spectrum of the motion energy filter is shown in the panel on the right.

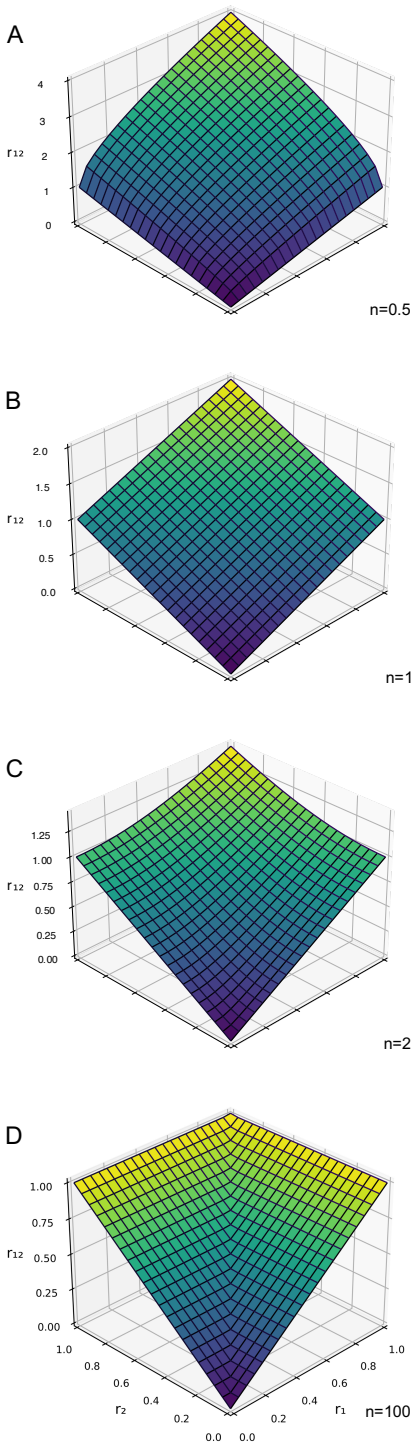

Figure 2-4 The power-law summation model

Power-law summation models of varying exponents, *n*. **(A)** *n=0.5*; **(B)** *n=1*; **(C)** *n=2*; **(D)** *n=100*. The response to simultaneous presentation of a pair of stimuli is plotted as a function of the response to each individual stimulus presented alone. The r_1 and r_2 axes correspond to the responses to the stimuli if they have appeared individually; the r_{12} axis corresponds to the response to the stimuli if they have appeared in a pair simultaneously.

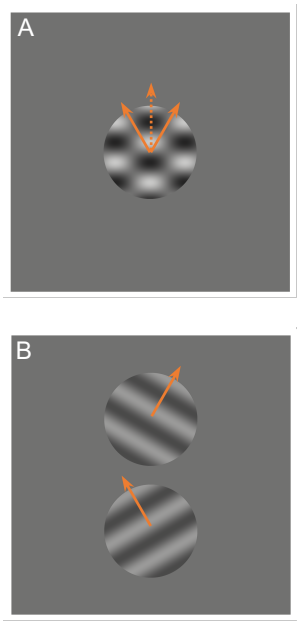

Figure 2-5 Conventional and pseudo-plaids

(A) A conventional plaid of two overlapping sinusoidal gratings drifting in different directions. **(B)** A pseudo-plaid where the component sinusoidal gratings drifting in different directions are spatially isolated in different windows. The solid arrows represent the component motions of the gratings. The dashed arrow represents the pattern motion of the conventional plaid.

Chapter 3.

Simulation methods: the MT model framework

It has long been the goal of modeling studies on MT to identify the potential circuits and algorithms associated with pattern motion sensitivity. Over the years, the principles of the architecture and computation involved in the modeling studies have shifted from complexity to simplicity and biological feasibility. For several decades, the predominant theoretical basis of MT modelling has been the *F-plane* motif: the architecture of the V1-MT network adopts certain topographies in the frequency domain such that the model can resolve the spectral energy within or relative to a constant velocity plane (Born & Bradley, 2005). Despite the ingenious computational solutions, such networks often appear highly idealized and curated, and a biological implementation like that is highly unrealistic. In response to such concerns, Rust and colleagues (2006) blazed a new trail, and proposed a model, which captures the full extent of pattern direction selectivity in MT, by adopting convergent inputs from V1 units with a wide range of PDs, strong opponent motion suppression and tuned normalization. However, intuitively simple and elegant, their model is not image computable. Baker and Bair (2016) later implemented an image-computable MT model based on similar principles, capable of being tested with any achromatic dynamic visual stimulus.

While it appears that the state-of-the-art models can fully account for MT motion sensitivity without failing the challenge of simplicity and biological feasibility, they are far from being physiologically realistic. The model of Baker & Bair (2016) does not provide a precise representation of the spatial RFs of the V1 and MT cells. Like any models that overlook the spatial aspect of V1-to-MT signal integration (Heeger, 1987, 1988; Grzywacz & Yuille, 1990; Bowns, 2002; Perrone, 2004), the entire visual field that encompasses the stimulus is visible to all V1 units, which therefore lack structured RFs. This makes the response of the model independent of the spatial distribution of the motion energy in the stimulus, and any stimulus that is not spatially uniform is not ideal for testing such models with. On the other hand, in *in vivo* studies, the adoption of such stimuli, where multiple moving components are spatially offset, has allowed researchers to probe the finer spatial composition of V1-MT circuitry (Britten & Heuer, 1999; Heuer & Britten, 2002; Majaj et al, 2007; Kumbhani et al., 2015; Wiesner et al., 2020), which has remained under the radar in the modeling studies of MT.

Building on the work of Baker and Bair (2016), a MT model previously developed in our laboratory, I have constructed the first image-computable MT model with spatial integration circuits of V1 units. By presenting stimuli with complex spatial layout to models with various architectural setups, I will showcase how such circuits shape the sensitivity profile of the MT RF and the way MT units respond to multiple spatially segregated motion signals in the RF. Here, I will outline the architecture of the V1-MT network (**Fig. 3-1A**), and explain the mathematical formulation of the computational modules.

3.1. Direction-selective V1 units

The front end of the MT model is made up of complex V1 units modeled as motion energy filters (Adelson & Bergen, 1985). These filters are parameterized by their spatial locations in the visual field and their PDs. The computation of the V1 stage involves (1) the extraction of motion signals through filters, (2) V1 normalization, and (3) the subtraction of the opponent motion signal.

3.1.1. Motion energy filters

The input layer of my model is a visual field of 128×128 pixels at a resolution of 0.1°/pixel. At each pixel sits the RF center of a stack of twelve V1 units selective for twelve directions of motion covering the 360° range. The motion energy filter extracts motion signals in a dynamic image sequence (video) by filtering the input with a pair of Gabor filters in quadrature. The odd-phased filter is defined as follows:

$$f_i^{\mathrm{o}}(x,y,t) = \exp\left(-\frac{x^2+y^2}{2\sigma_{\mathrm{s}}^2}\right)\exp\left(-\frac{t^2}{2\sigma_{\mathrm{t}}}\right) \sin(2\pi\omega_{\mathrm{s}}(x\cos\theta_i + y\sin\theta_i) + 2\pi\omega_{\mathrm{t}}t) \qquad \textbf{Equation 3-1}$$

where i corresponds to one of the twelve directions, i.e., θ_i. The spatial and temporal spreads of the Gaussian envelopes of the Gabor functions are determined by σ_{s} and σ_{t}, respectively, which control the width of the pass band of the filter in space and time, and in turn, the width of direction tuning, which has an inverse relationship with the spatiotemporal bandwidth. The central SF and TF are ω_{s} and ω_{t}, respectively. Similarly, the even-phased filter is defined as follows:

$$f_i^{\mathrm{e}}(x,y,t) = \exp\left(-\frac{x^2+y^2}{2\sigma_{\mathrm{s}}^2}\right)\exp\left(-\frac{t^2}{2\sigma_{\mathrm{t}}}\right) \cos(2\pi\omega_{\mathrm{s}}(x\cos\theta_i + y\sin\theta_i) + 2\pi\omega_{\mathrm{t}}t) \qquad \textbf{Equation 3-2}$$

The L^2-norm of the outputs of the pair of filters is the motion energy readout:

$$m_i(x,y,t) = \sqrt{\left(f_i^{\mathrm{o}}(x,y,t) * v(x,y,t)\right)^2 + \left(f_i^{\mathrm{e}}(x,y,t) * v(x,y,t)\right)^2} \qquad \textbf{Equation 3-3}$$

where $v(x,y,t)$ represents the input video stream, and $*$ the convolution operation. In the notation of the output, $m_i(x,y,t)$, (x,y) indicates that the output is associated with the V1 unit sitting at such a spatial location.

3.1.2. V1 normalization

The output of the motion energy filter is normalized divisively in the model, using a scheme modified from Rust et al. (2006):

$$n_i(x,y,t) = \frac{m_i(x,y,t)}{\alpha_1 m_i(x,y,t) + \alpha_2 \frac{\sum_k m_k(x,y,t)}{M} + \alpha_3}$$

Equation 3-4

where α_1 controls the strength of the tuned normalization (normalization against the individual direction channel), and α_2 controls the untuned normalization (normalization against the entire range of directions), and α_3 determines the half-saturation level and avoids division by zero. $M = 12$ is the number of direction channels.

Alternatively, V1 surround suppression can be substituted for the above normalization:

$$n_i(x,y,t) = \frac{m_i(x,y,t)}{\alpha_1 s_i(x,y,t) + \alpha_3}$$

Equation 3-5

Compared to **Eq. 3-4**, here, the surround signal, $s_i(x,y,t)$, serves as the source of tuned normalization here, and the untuned normalization is omitted. The surround signal is computed as the convolution of the motion energy outputs and the Gaussian surround fields, $g(x,y)$:

$$s_i(x,y,t) = \left(m_i(x,y,t) + m_{\left(i+\frac{M}{2}\right) \bmod M}(x,y,t)\right) * g(x,y)$$

Equation 3-6

where i and $\left(i + \frac{M}{2}\right) \bmod M$ represent a pair of channels of opposite directions, given that there are M total direction channels. The surround signal involves the summation of motion signals from two opposite direction channels because the V1 surround is presumed to be orientation-selective instead of being DS (Jones et al., 2001; Cavanaugh et al., 2002). In the base model, compared to the spatial SD of the Gabor filters being 0.36°, the SD of the surround Gaussian filed is 1.26°. Thus, the spatial scale of the suppressive surround extends about 3.5 times the diameter of the CRF.

3.1.3. Motion opponency

The half-wave rectified difference between the signals of a pair of opponent direction channels is then amplified as defined below:

$$o_i(x,y,t) = \kappa_1 \left[n_i(x,y,t) - n_{\left(i+\frac{M}{2}\right) \bmod M}(x,y,t)\right]$$

Equation 3-7

where $\lfloor \quad \rfloor$ represents the half-wave rectification, and κ_1 is the gain of the amplification. This kind of motion opponency is a local mechanism operating at the V1 level between direction channels that are spatially concurrent.

3.2. V1-to-MT integration

The previous MT model (Baker & Bair, 2016) does not include spatial integration in the V1-MT wiring. Here, I introduce spatial integration by allowing the MT unit to sum, linearly or nonlinearly, the V1 outputs across various spatial locations with weights determined by the PD of the V1 unit.

The weighting function is symmetric with the peak positioned at the channel of the MT unit's PD. The CDS units adopt a weighting function with very narrow direction bandwidth (**Fig. 3-1B**, red curve) and is positive only for the PD channel of the MT model and very slightly negative for the anti-PD channels. For the PDS units, the shape of the weighting function is considerably broader, and the weights are moderately negative for the anti-PD channels (**Fig. 3-1B**, teal curve), as proposed in Rust et al. (2006). The opposite signs of the weights applied to the PD and anti-PD channels define the excitatory and inhibitory direction channels and effectively establish motion opponency at the MT level.

3.2.1. Spatial configurations of V1 channels

The topography of the spatial locations of V1 channels are parameterized in a three-dimensional domain – (1) the number of V1 inputs (**Fig. 3-2A,E,F&G**), (2) whether the V1 inputs are arranged in "stacks", or randomly, across direction channels – *stacked* vs. *unstacked* configurations (**Fig. 3-2A&B** vs. **C&D**), and (3) whether the V1 inputs are evenly or randomly distributed in space – *even* vs. *uneven* configurations (**Fig. 3-2A&C** vs. **B&D**). The latter two parameters are binary, and the product of them results in four composite configurations: the *stacked-even* (**Fig. 3-2A**), *stacked-uneven* (**Fig. 3-2B**), *unstacked-even* (**Fig. 3-2C**) and *unstacked-uneven* configurations (**Fig. 3-2D**).

3.2.1.1. The unstacked-uneven configuration

In the unstacked-uneven configuration, as shown in **Fig. 3-2D**, the spatial locations of the V1 inputs are chosen randomly from a uniform spatial distribution, and such locations are not identical across direction channels.

3.2.1.2. The unstacked-even configuration

In the unstacked-even configuration, although the spatial layouts of the V1 locations do not align across direction channels, within each direction channel, it is not chosen completely randomly, as shown in **Fig. 3-2C**.

For the even setups, I used a quasi-random procedure derived from Roberts (2018) to generate V1 input locations that are evenly spaced in a circular area within the boundary of the MT RF. Below, I demonstrate this

procedure mathematically. It involves (1) generating evenly spaced points in a square area, and (2) mapping these points to a circular area.

Given a 2.0-by-2.0 square domain centered at the origin, one can generate a quasi-random 2D sequence that corresponds to evenly spaced coordinates within the domain using the following equations:

$$x[i] = 2\left\{x_0 + \frac{1}{\phi_1}\right\} - 1$$
$$y[i] = 2\left\{y_0 + \frac{1}{\phi_2}\right\} - 1$$

Equation 3-8

where $i = K, K+1, K+2 \ldots K+N-2, K+N-1$ with N being the number of spatial points and K being a random integer, and ϕ_1 and ϕ_2 are the unique positive roots of $z^2 = z + 1$ and $z^3 = z + 1$, respectively, and $\{\ \}$ represents the fractional part function. x_0 and y_0 can be any real numbers, and, along with K, can be chosen randomly for a specific direction channel.

These coordinates were then mapped to a unit circle using the following equation (Shirley & Chiu, 1997):

$$(u, v) = \begin{cases} \left(x\cos\left(\frac{\pi}{4}\frac{y}{x}\right), x\sin\left(\frac{\pi}{4}\frac{y}{x}\right)\right) & |x| \geq |y| \\ \left(y\sin\left(\frac{\pi}{4}\frac{x}{y}\right), y\cos\left(\frac{\pi}{4}\frac{x}{y}\right)\right) & |x| < |y| \end{cases}$$

Equation 3-9

These (u, v) coordinates within a unit circle are then transformed proportionally to a circular area with respect to the size of the MT RF.

3.2.1.3. The stacked-uneven configuration

In the stacked-uneven configuration, the locations of the V1 channels are randomly selected but align across all direction channels, therefore, forming "stacks" of inputs that sample the complete range of directions at each location (**Fig. 3-2B**).

3.2.1.4. The stacked-even configuration

In the stacked-even configuration, the V1 inputs are evenly spaced and aligned across direction channels (**Fig. 3-2A**).

3.2.1.5. Centralization of V1 channels

The V1 channels sit within a circular area with a radius of 3.75°, which marks the boundary of the model MT RF. In the spatial layouts mentioned above, especially in the even cases, the density of the V1 inputs do not systematically vary depending on the relative location within the MT RF. This corresponds to a flat spatial

sensitivity profile. However, in some cases, a spatial gradient of V1 input density is preferred such that the sensitivity is high at the center of the RF and rolls off towards the boundary, corresponding to a more bell-shaped sensitivity profile.

To achieve such a centralization effect, while being able to parametrically control it, the following transform:

$$\psi = 1 - (1 - \phi^{\kappa})^{\frac{1}{\kappa}} \qquad \textbf{Equation 3-10}$$

is used to remap the V1 locations towards the center of the RF. The variables, ϕ and ψ, whose values are within [0.0,1.0], are the relative locations of the V1 unit with respect to the center of the MT RF, as the ratio to the radius of the MT RF boundary, before and after the transform, respectively. The extent of the centralization process is controlled by $\kappa = \frac{1}{1-c}$, with c functioning as the centralization factor taking values between 0.0 and 1.0, and larger values corresponding to stronger centralization. This transform has no impact on the V1 units that sit at the center ($\phi = 0.0$) or the boundary ($\phi = 1.0$) of the MT RF, but locations in between the extremes will be pulled towards the center. **Fig. 3-3A** shows the relative density of V1 units as a function of the relative distance from the center of the model RF while the centralization factor varies. The shape of the density function is concave, and the roll-off is steeper for models with stronger centralization. **Fig. 3-3B-E** provides the example spatial arrangements of V1 locations at different centralization levels.

3.2.2. Signal integration schemes

In the MT models of Rust et al. (2006) and Baker & Bair (2016), the outputs of the V1 channels are integrated using linear summation. Here, I extend the V1-to-MT integration scheme with spatial summation, and demonstrate how variations of the scheme in combination with certain spatial layouts of V1 channels can construct complex motion processing circuits and MT RF substructures.

Some evidence has suggested that spatial integration in MT may not be completely linear, as the response of a MT cell to an ensemble of sub-RF stimuli presented simultaneously with spatial offsets is not always the linear sum of the responses to such stimuli presented individually (Britten & Heuer, 1999; Heuer & Britten, 2002). Therefore, alternatively, I have also implemented nonlinear V1-to-MT integration in my models.

3.2.2.1. Linear integration

In the linear V1-to-MT integration scheme, the outputs of the V1 channels are weighted and summed, and the weights depend on the PD of the V1 unit relative to the designated PD the MT model, with the V1 unit whose PD aligns with the PD of the MT model receiving the strongest weight (**Fig. 3-1B**). Such an integration scheme can be described with following equation:

$$p(t) = \frac{1}{MN}\sum_{j=1}^{N}\sum_{i=1}^{M} w_i o_i\left(x_{h(i,j)}, y_{h(i,j)}, t\right)$$

Equation 3-11

where $o_i\left(x_{h(i,j)}, y_{h(i,j)}, t\right)$ is the output of the jth V1 unit of the ith direction channel located at $\left(x_{h(i,j)}, y_{h(i,j)}\right)$; $h(i,j)$ indicates that the location of the V1 unit may vary within and across direction channels in an unstacked configuration. N is the number of V1 locations per direction channel; $M = 12$ is the number of direction channels. w_i represents the weight applied to the units of the ith direction channel.

In such a setup, the summation of V1 outputs happens across spatial and direction channels at the same time, and the motion signal is integrated globally across the MT RF. When combined with the unstacked layout of V1 locations, this paradigm is referred to as the *no-subunit* architecture as there is no processing module that operates on a local scale.

In a stacked configuration, **Eq. 3-11** degenerates to a special case where the spatial arrangement of the V1 units is independent of the direction of the channel, as defined in the following equation:

$$p(t) = \frac{1}{MN}\sum_{j=1}^{N}\sum_{i=1}^{M} w_i o_i\left(x_{h(j)}, y_{h(j)}, t\right)$$

Equation 3-12

where the spatial location of the V1 unit, $(x_{h(j)}, y_{h(j)})$, only depends on j – the layout of the V1 locations is identical across direction channels, and $\sum_{i=1}^{M} w_i o_i\left(x_{h(j)}, y_{h(j)}, t\right)$ represents the jth stack of V1 channels that samples the complete direction range of 360° at that location.

In such a setup, the summation of V1 outputs still happens globally, but each stack of V1 channels can potentially support localized signal processing subunits, albeit not in this case due to the lack of any localized nonlinear computation. Therefore, I refer to this paradigm as the *false-subunit* architecture.

In the *true-subunit* architecture where the V1 channels are stacked, signal integration happens locally within each stack before being summed across all locations. Such an operation can be described by the following equation:

$$p(t) = \frac{1}{MN}\sum_{j=1}^{N}\left\lfloor\sum_{i=1}^{M} w_i o_i\left(x_{h(j)}, y_{h(j)}, t\right)\right\rfloor$$

Equation 3-13

where $\left\lfloor\sum_{i=1}^{M} w_i o_i\left(x_{h(j)}, y_{h(j)}, t\right)\right\rfloor$ represents localized summation, followed by half-wave rectification, of the outputs of the V1 units within a processing subunit (stack). In this case, motion integration first happens locally within a subunit where both input pooling and computation operate independently of signals from V1 units located elsewhere. In such models, localized motion processing occurs both structurally and computationally, therefore forming true subunits.

3.2.2.2. Nonlinear integration

Alternatively, the weighted summation of V1 outputs can be nonlinear, as defined as follows:

$$p(t) = \mathrm{P}^{\left(\frac{1}{\gamma}\right)}\left(\frac{1}{MN}\sum_{j=1}^{N}\sum_{i=1}^{M} w_i \, \mathrm{P}^{(\gamma)}\left(o_i\left(x_{h(i,j)}, y_{h(i,j)}, t\right)\right)\right)$$

Equation 3-14

where $\mathrm{P}^{(\xi)}(z)$ represents a sign-preserving power operation, equivalent to $|z|^{\xi} \operatorname{sgn} z$. This summation process is essentially a weighted average of the V1 outputs raised, while preserving the signs, to the power of γ, followed by the inversion of the power, also in a sign preserving manner. This treatment draws inspiration from Britten and Heuer's (1999) *power-law summation model*. Here, γ, the power, controls the degree of the nonlinearity of the integration. When $\gamma = 1$, this integration is a simple linear averaging process. For $\gamma > 1$, as the value becomes larger, the integration will be more prominently dominated by the strongest input, leading to a winner-take-all operation, reducing the contributions of the weaker, however possibly more prevalent, inputs.

3.3. Divisive MT population normalization

Optionally, the MT signal resulted from integration can be normalized in a divisive process against the average signal from a population of MT units, whose RFs are off-centered relative to the MT unit of interest and cover the surrounding area. Their direction preferences can be different than that of the interested MT unit, and span the complete 360° range. The mathematical formulation of normalization is described in the following equation:

$$q(t) = \frac{p(t)}{\beta_1 \frac{\sum_{k=1}^{K} |p_k(t)|}{K} + \beta_2}$$

Equation 3-15

where $\frac{\sum_{k=1}^{K} |p_k(t)|}{K}$ is the average half-wave rectified MT signal over K units, and β_1 determines the strength of normalization, and β_2 avoids division by zero as well as controls the half-saturation level. The normalization operation is more effective when $\beta_1 \frac{\sum_{k=1}^{K} |p_k(t)|}{K} \gg \beta_2$.

3.4. Poisson spiking

The resulting MT output is then half-wave rectified and amplified as follows:

$$r(t) = \kappa_2 \lfloor q(t) \rfloor$$

Equation 3-16

where κ_2 is the gain of amplification. Such analogue signals are then fed to a Poisson spike generator which produces the final output of the model.

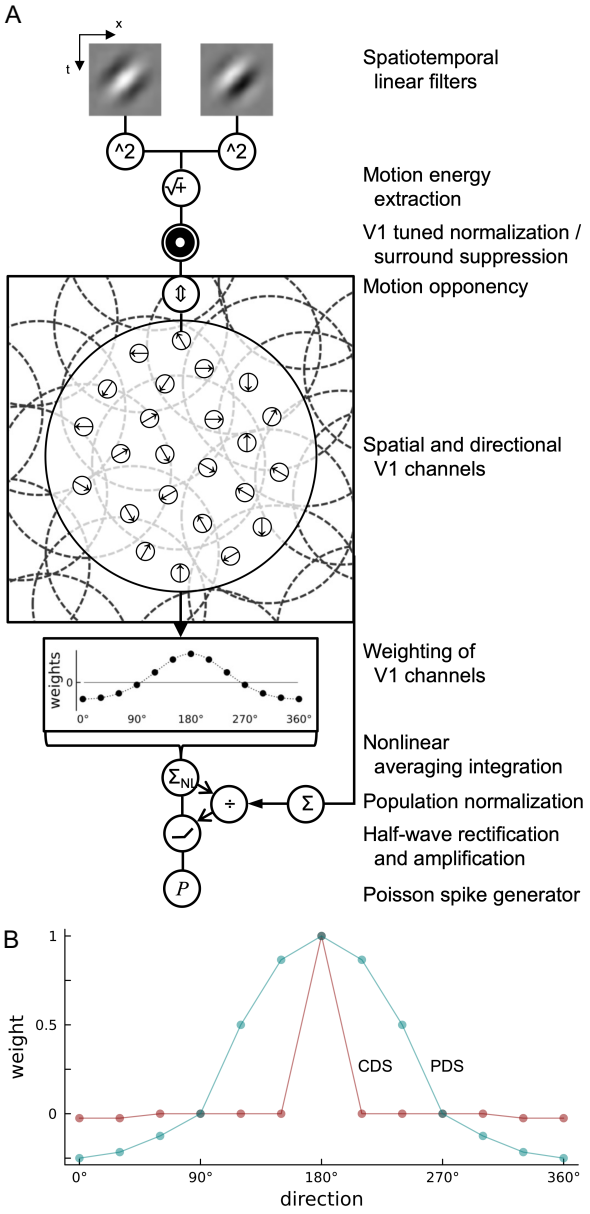

Figure 3-1 MT model

(A) The framework of the MT model. Inside the spatial receptive field of the MT unit, sit the smaller receptive fields of V1 units of various direction preferences at different locations. The V1 units are modeled as complex cells using motion energy filters. The output of the motion energy filter is normalized, and the signal from the opponent direction channel is subtracted. The outputs of all V1 channels are then integrated with direction-dependent weights and generate an analogue signal. The integration process can be optionally nonlinear, and the formulation of the process is described in ***3.2.2.2. Nonlinear integration***. An optional population normalization stage nonlinearly scales the MT output. After half-wave rectification and amplification, the analogue signal is eventually fed to a Poisson spike generator. **(B)** The direction weighting function. The weighting function for direction channel integration is wider for the pattern MT units.

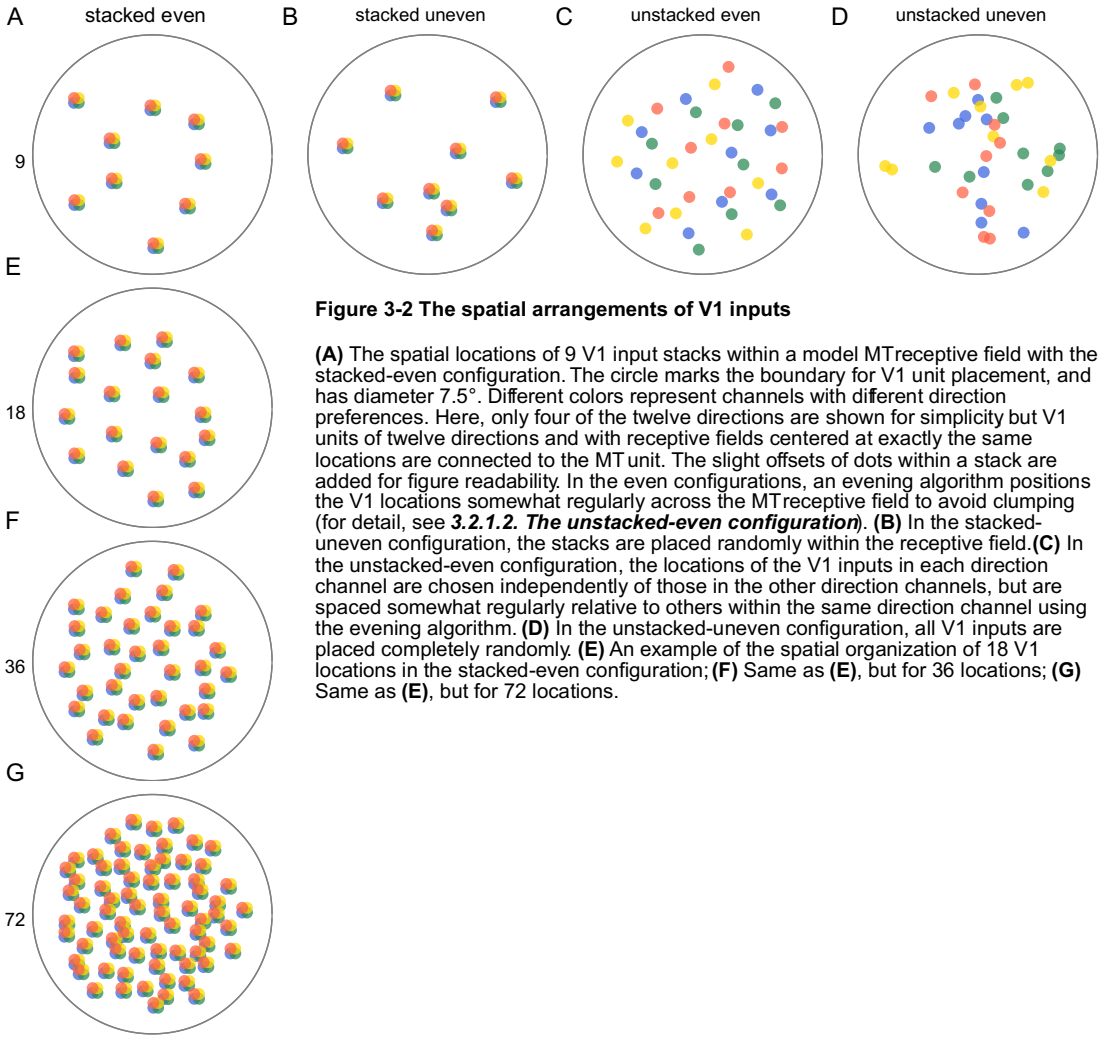

Figure 3-2 The spatial arrangements of V1 inputs

(A) The spatial locations of 9 V1 input stacks within a model MTreceptive field with the stacked-even configuration. The circle marks the boundary for V1 unit placement, and has diameter 7.5°. Different colors represent channels with different direction preferences. Here, only four of the twelve directions are shown for simplicity but V1 units of twelve directions and with receptive fields centered at exactly the same locations are connected to the MT unit. The slight offsets of dots within a stack are added for figure readability. In the even configurations, an evening algorithm positions the V1 locations somewhat regularly across the MTreceptive field to avoid clumping (for detail, see ***3.2.1.2. The unstacked-even configuration***). **(B)** In the stacked-uneven configuration, the stacks are placed randomly within the receptive field.**(C)** In the unstacked-even configuration, the locations of the V1 inputs in each direction channel are chosen independently of those in the other direction channels, but are spaced somewhat regularly relative to others within the same direction channel using the evening algorithm. **(D)** In the unstacked-uneven configuration, all V1 inputs are placed completely randomly. **(E)** An example of the spatial organization of 18 V1 locations in the stacked-even configuration; **(F)** Same as **(E)**, but for 36 locations; **(G)** Same as **(E)**, but for 72 locations.

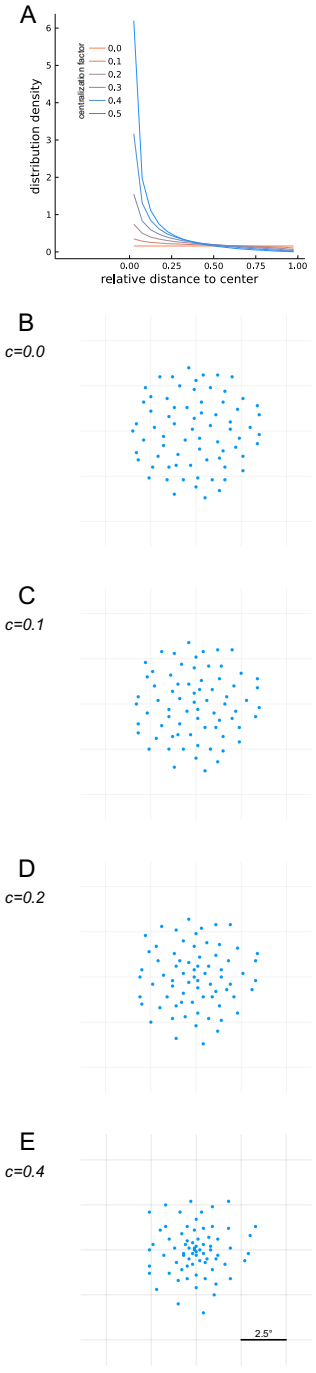

Figure 3-3 Centralized distribution of V1 units in the model MT receptive field

The V1 inputs can be arranged to be denser in the center. The level of centralization is controlled by the centralization factor, *c*. **(A)** The relative distribution density of V1 inputs as a function of the relative distance from the center of the receptive field at different centralization level. **(B)** An example illustration of the V1 input arrangement when *c=0.0* (no centralization); **(C)** *c=0.1*; **(D)** *c=0.2*; **(E)** *c=0.4*.

Chapter 4.

Modeling of spatial inhomogeneity in the MT receptive field

A foundational principle of visual cortical processing entails a hierarchical organization in which neuronal RFs increase in size and elevate in functional complexity as one traverses progressively deeper in the network; however, a limited understanding currently exists regarding the mechanisms underlying this phenomenon beyond the primary visual cortex. For example, the DS response of neurons in MT is thought to arise largely from the integration of DS signals that originate in V1 (Dubner & Zeki, 1971). These inputs are spatially localized direction channels that are tuned for orientation and SF and TF (Movshon and Newsome, 1996). The spatial integration process creates MT RFs that are several to ten times larger in diameter than those of the V1 inputs (Albright & Desimone, 1987; Wang & Movshon, 2016). Many past computational studies of MT models focused on integration across direction channels, and have mostly disregarded the spatial mechanisms of signal integration and the accuracy of the MT RF (Heeger, 1987; Grzywacz & Yuille, 1990; Bowns, 2002; Perrone, 2004; Rust et al., 2006; Nishimoto & Gallant, 2011; Baker & Bair, 2016). For studies that modeled the spatial profile of MT RFs, the spatial structure of the RF has largely been overlooked and assumed to be relatively homogeneous, such that the MT cell prefers the same direction throughout the spatial RF and the sensitivity is high at one location, but drops off toward the edge of the RF (Raiguel et al., 1995).

Although past evidence aligns with the view of homogeneous direction selectivity across MT RFs (Livingstone et al., 2001), a more recent study by Richert and colleagues (2013) raised the question about the spatial structure of V1-to-MT signal integration when they demonstrated that a considerable fraction of macaque MT neurons have heterogenous, multi-regioned RFs. Besides being homogeneous (**Fig. 4-1A**), the MT RF may be multi-parted (**Fig. 4-1B**) or have different direction preferences across space (**Fig. 4-1C**). By incorporating a simple V1-MT random wiring process into an adapted MT model of Baker & Bair (2016), here, I tested if such a process could achieve realistic spatial integration of V1 inputs in MT, and give rise to a RF heterogeneity level comparable to that reported in Richert et al. (2013).

To allow for a direct comparison to the results of Richert et al. (2013), I mapped the sensitivity profiles and direction preferences of the model RFs using random dot stimuli similar to theirs. I found that the randomization of the spatial structure of V1 connections is able to produce multi-peaked RFs with segregated subfields; however, these subregions within the RF generally shares similar PDs. Hence, I conclude that spatial randomness in V1-MT wiring alone cannot account for the heterogeneous direction preference within a single MT RF recorded by Richert and colleagues (2013). This indicates that the level of RF heterogeneity observed experimentally most likely arises from concerted, non-random developmental processes.

4.1. Methods

4.1.1. The stimulus

To probe the models, I used a sparse, dynamic random dot stimulus (**Fig. 4-2A**) modeled after that of Richert et al. (2013). Independently moving dots with a density of 0.25 dot/deg^2 were presented in a field modeled to be 12.8° × 12.8° (128×128 pixels). The dots have a 2D Gaussian luminance profile with the σ being 0.2° and are truncated at 2.5σ. The dots move in random directions at 8°/s, while extinguishing and reappearing at random locations every 200 ms with newly assigned directions sampled from a uniform distribution with an increment of 15°. My random dot stimulus differs from that of Richert et al. (2013) only in that their dot trajectories are spatially continuous – dots merely change direction periodically. Pixel luminance ranges from 0.0 to 1.0, the background is held at 0.1 to simulate the grey background used by Richert and colleagues (2013), and the dots have a Weber contrast of 3.0 measured at the peak of the Gaussian profile. When two or more dots overlap, the luminance is summed.

4.1.2. Simulation

I presented the stimuli to the models in 1,000 trials, each lasting for 1.5 s, with no repetition, providing at least 40,000 spikes for computing spike-triggered average (STA) (Chichilnisky, 2001; Schwartz et al., 2006). For comparison, the median number of spikes in the STAs of Richert et al. (2013) is 8,156.

4.1.3. Computation of spike-triggered average and the spatial receptive field map

STA was used to generate spatial maps of direction preference within the RF of the model unit. It resolves the temporal evolution of direction tuning for spatially defined unit areas in the RF, which is broken into a 12×12 grid of square blocks of 1° dimensions. First, I correlate the spike data and the stimuli by selecting the frames within a symmetric time window of 500 ms around each spike (the window is symmetric due to the non-causal nature of the filters in the model). At each frame, I count the instances of dots appearing in each block and record their direction of motion. I then aggregate the counts across all spikes at the same latency and compute a histogram of direction distribution of the dot motion, which can be interpreted as the direction tuning for a point in the spatiotemporal RF (**Fig. 4-2B&C**). Therefore, the STA is a 3D spatiotemporal array of circular histograms representing the direction distribution of the stimulus motion signals associated with output spikes.

Given the STA, for each latency of a 1 ms bin, a 2D map is generated to survey the direction selectivity across the spatial RF. For each 1°-by-1° block at that latency, an average vector is generated by treating the histogram in the STA as a circular distribution and calculating the sample mean vector – the PD vector. The angle of the vector represents the PD of the corresponding block, and I refer to the magnitude of the vector as the strength of tuning. This metric of direction tuning strength has been argued by Mazurek and colleagues (2014) to be more robust to noise than the traditional direction index metric, as the latter depends on

responses at only two directions, and my method is similar to the tuning strength metric of Richert et al. (2014), except that they used the vector sum, not average.

These 2D maps of PD vectors can be visualized in pseudo-color, with the hue coding for the direction and the luminance indicating the tuning strength, i.e., local sensitivity within the RF (**Fig. 4-2D**). An optimal latency and its corresponding 2D map for the RF are selected based on the criterion of maximizing the sum of tuning strengths across all blocks within the map.

4.1.4. Define significant receptive field blocks

A block in the observed optimal RF map is considered significant if its tuning strength is above the threshold determined by a randomization procedure. A reference STA was computed after a random shuffle of the trials (n = 1000) disassociates the correspondence between the spike data and the stimuli. I then generate 500 reference optimal RF maps by resampling the histograms in the reference STA. These randomly generated maps provide the distribution of tuning strength in the RF maps if it arose by chance. For each block, the upper boundary of the 95% confidence interval (adjusted for multiple comparison using the Holm–Bonferroni method) determined for tuning strength using the reference maps is set as the threshold for significant tuning.

4.1.5. Local spatial sensitivity heterogeneity of the receptive field envelope

I calculated the amplitude variance of STA within the RF convex hull – the smallest convex set that covered all significant blocks (**Fig. 4-3**) – to quantify the spatial heterogeneity of the tuning strength. I first identify all the blocks within the convex hull. Given block $p_i^{(0)}$ in the convex hull C with a tuning strength of $s_i^{(0)}$, and its eight neighboring blocks $p_i^{(k)}(k = 1,2,3 \dots 8)$ with a tuning strength of $s_i^{(k)}$, its local heterogeneity is computed as:

$$h_i = \frac{\sum_{k|p_i^{(k)} \in C} \sqrt{\left(s_i^{(k)} - s_i^{(0)}\right)^2}}{\left|\left\{k|p_i^{(k)} \in C\right\}\right|}$$

Equation 4-1

where $\left|\left\{k|p_i^{(k)} \in C\right\}\right|$ is the number of neighboring blocks that are within C. The heterogeneity of the overall RF is computed as the mean of the local heterogeneity for all blocks within the convex hull.

4.1.6. Define receptive field subregions

Individual significant blocks were clustered into RF subregions based on the connectivity of spatial locations and the similarity in direction preference. Two blocks are considered spatially connected if they are adjacent by edges or corners (8-connectivity). Similarity in direction preference is quantified by the PD difference between the two blocks. The direction tuning of each block is represented by a circular histogram with bin-widths of 15°. I rotate the histogram of one of the blocks by steps of 15° and calculate the Bhattacharyya

coefficient, which ranges from 0.0 to 1.0, with 0.0 indicating no overlap of the distributions and 1.0 indicating identical distributions. The rotation angle that generates the largest Bhattacharyya coefficient is chosen as the direction difference between the two blocks, and if it is within 15°, the two blocks are considered tuned for the same direction. Blocks belong to the same subregion only when they are spatially connected, and each block shares the same direction preference with at least one of its neighbors within the subregion.

4.1.7. The valley-over-peak ratio

I computed the valley-over-peak ratio to quantify the drop of sensitivity between the most prominent response peaks within the RF convex hull. I look for the primary and secondary tuning strength peaks using the following procedure. I order all blocks enclosed by the convex hull according to their tuning strength values. I choose the block with the maximum value as the summit of the primary peak. I then iterate through the rest of the blocks in descending order. The block being inspected is considered to belong to the primary peak only if it is connected to any of the blocks already identified to belong to the primary peak. The iteration is to stop when a block not associated with the primary peak has been discovered, and that block will be treated as the summit of the secondary peak. Finally, I determine the lowest tuning strength along the line connecting the summits of the two peaks. Because of the raster nature of the RF maps, such lines are approximated using the straight line algorithm by Bresenham (1965). I define the valley-over-peak ratio as the lowest tuning strength on the line connecting the primary and secondary peaks divided by the tuning strength of the summit of the secondary peak.

4.1.8. A model of the MT neuronal receptive field

The MT models construct PDS and CDS units from the outputs of DS V1 inputs. Here, I briefly describe the shared architecture of the models (for details, see **Chapter 3. Simulation methods: an MT model framework**) before discussing the topographic variants of the MT RF.

The input layer of the model is a 2D array of pixels that match the spatial dimension and resolution of the stimulus. The V1 units are modeled using Gabor motion energy filters selective for one of the twelve directions spanning the 360° range. The spatial SD of the Gabor function, which determines the size of the V1 RF, is 0.36° (the spatial scale of the V1 filter and the stimulus dot is shown in **Fig. 4-4**). To generate a PDS model, I adopt tuned normalization at the V1 level (Baker & Bair, 2016). A motion opponent V1 output is generated by subtracting from each channel the signal in the oppositely tuned channel, followed by half-wave rectification.

Spatial integration in the wiring of V1 to MT is achieved by allowing the MT unit to linearly sum the V1 outputs across various spatial locations with weights determined by the PD of the V1 unit relative to that of the MT model. Compared to that of the CDS units, the weighting function of the PDS units is significantly broader. The linear summation signal is half-wave rectified and amplified before being fed to a Poisson spike generator.

To study the spatial integration pattern at the MT stage, I varied three properties that influence the spatial distribution of the V1 inputs connected to the MT unit: (1) the number of V1 inputs, (2) whether the V1 inputs

are arranged in "stacks", or randomly, across direction channels (*stacked* vs. *unstacked* configuration), and (3) whether the V1 inputs are evenly or randomly distributed in space (*even* vs. *uneven* configuration, see **3.2.1. Spatial configurations of V1 channels**). In the MT models, there are 9, 18, 36 or 72 V1 inputs spread over space in each direction channel (**Fig. 3-2A&E,F,G**). In the stacked configuration (**Fig. 3-2A&B**), the spatial locations of V1 inputs are aligned across all direction channels; in the unstacked configuration (**Fig. 3-2C&D**), such locations are independent across direction channels. For the even configuration (**Fig. 3-2A&C**), I use a quasi-random procedure to generate V1 input locations that are evenly separated from each other while covering a circular area within the RF (Roberts, 2018); in the uneven configuration (**Fig. 3-2B&D**), such locations are sampled completely randomly in the circular area. The diameter of the circular area is 7.5°, and limited centralization with a factor 0.05 is applied (for detail, see **3.2.1.5. Centralization of V1 channels**).

4.2. Results

To understand how RFs in MT might be constructed spatially, I presented the random dot stimuli adapted from Richert et al. (2013; see **Fig. 4-2A** and **4.1.1. The stimulus**) to a set of models of MT neurons that account for both CDS and PDS cells. The major variables that I considered were: (1) the density of the V1 inputs that are integrated by each MT unit, (2) the spatial uniformity of the coverage of those inputs, (3) the organization of the V1 inputs into coherent local spatial subunits (false subunits, see **3.2.2.1. Linear integration**), and (4) the direction bandwidth of the weighting function of the V1 inputs, which essentially controls whether the unit was PDS or CDS.

Fig. 4-5A:I&II show an example of mapping a model MT RF with the dynamic random dot stimulus. The left panel (**Col. I**) shows the STA map, where brighter pixels correspond to stronger sensitivity and the color indicates the PD (**Fig. 4-2E** shows the color key), while the right panel (**Col. II**) shows the locations of the centers of the RFs of the V1 inputs that were connected to, thus driving, the model MT unit. This example unit received V1 inputs from only 9 locations (**Fig. 4-5A:II**), but I varied this parameter, testing values up to 72, as shown in **Fig. 4-5A:III&IV**. In both of these examples in **Fig. 4-5A**, the inputs are placed somewhat evenly in the RF to avoid clumping (see **3.2.1.2. The unstacked-even configuration**). I also allowed that the inputs can be placed entirely randomly (unevenly), as shown in **Fig. 4-5B** for 9 and 72 input locations per direction channel. To understand how these connectivity parameters influenced the homogeneity of the RFs, I computed a metric of the spatial heterogeneity of the RF sensitivity/tuning strength profile based on the variance in the STA amplitude between neighboring blocks in a region of the field that was determined to be within a convex hull that included all significant blocks (see **4.1.5. Local spatial sensitivity heterogeneity of the receptive field envelope**).

I found that the spatial sensitivity profiles of the model RFs for the CDS units (**Fig. 4-5A-D**) became substantially less heterogeneous (i.e., smoother) on average as the number, and thus the density, of V1 inputs increased (**Fig. 4-6A**). This trend held whether the inputs were placed randomly (uneven, green curves) or in a more regular fashion (even, blue curves). However, the heterogeneity was significantly higher for unevenly

placed inputs compared to the even case. In the same models, I also measured the heterogeneity of direction tuning across the RF, having only included the significant blocks because blocks with weak, non-significant responses had no certainty in their direction preference. **Fig. 4-6E** shows that the SD of direction preference across all significant blocks was low (not above 15°), and did not change substantially with the number of inputs or even/uneven spatial sampling. These same trends held for model PDS units (**Fig. 4-6C&G**). In particular, the RF sensitivity profile became less heterogenous as the number of inputs locations increased (**Fig. 4-6C**), and the variation in preferred direction remained low (SD not above 12° on average) across the RF regardless of the number of input locations per direction channel (**Fig. 4-6G**).

In the models discussed so far, I assumed the existence of V1 spatial substructures whereby MT cells recruit V1 inputs having different direction preferences but with RFs coincident at the same spatial locations, effectively forming local "stacks" of V1 channels that pooled motion signals from all directions (see **4.1.8. A model of the MT neuronal receptive field** and **Fig. 3-2A&B**). As having observed that, for such models, neither even nor uneven spatial sampling of V1 inputs was able to produce MT RFs with high variability of direction preference, I sought to boost the randomness of V1-MT wiring in the models by introducing the unstacked configuration, which disrupts the stacked local substructures by spatially randomizing V1 inputs independently across direction channels (**Fig. 3-2C&D**).

For the CDS units, I found that unstacking the V1 inputs did not substantially change the heterogeneity of RF tuning strength nor the variability in direction tuning (**Fig. 4-6B&F** vs. **A&E**), compared to those for the PDS units (**Fig. 4-6D&H** vs. **C&G**), but this should be expected due to the very narrow bandwidth of the direction weighting function for CDS units (**Fig. 3-1B**, red curve), which effectively reduces the unstacked models to having similar spatial variation as the stacked ones. On the other hand, the PDS units with unstacked V1 inputs showed significantly lower heterogeneity in tuning strength (**Fig. 4-6D**, compared to the stacked cases in **C**), as the broader direction weighting function (**Fig. 3-1B**, green curve) effectively increases the sampling density by recruiting V1 channels of different PDs at more locations, making the RF more homogeneous. Unstacking the V1 inputs had the opposite effect on the variability of direction tuning in PDS units: the wider array of heavily weighted directions in the V1 inputs translated into higher variability in direction preference (**Fig. 4-6H**) compared to the stacked models (**Fig. 4-6G**). However, such variability still fell short of what was required to produce clear, multi-direction RFs, like that shown in **Fig. 4-1C**. For example, typical unstacked PDS model units (**Fig. 4-5G:I&H:I**) displayed a profile where direction tuning was relatively consistent throughout the RF (color varies only in the blue to green range).

4.2.1. Comparison to the electrophysiology data

I have demonstrated that changing the key model parameters, i.e., the number of inputs, the regularity of the spatial distribution of the inputs, the local stacking of inputs across direction channels, and the shape of the direction weighting function (CDS vs. PDS), could influence the tuning strength heterogeneity of the modeled MT RFs, yet the impact to the variability of tuning direction was very limited. To directly address the question

as to whether the level of RF heterogeneity I observed could sufficiently account for the results reported in Richert et al. (2013), I carries out some analyses similar to theirs, including the subregion analysis, the valley-over-peak analysis and the convex-hull analysis.

4.2.1.1. Analysis of subregions

Richert and colleagues (2013) found that 38% of the MT neurons they studied have multiple subregions with different PDs, and 24% of the neurons' RFs encompass more than one sensitivity peak. I conducted a similar analysis where I defined subregions in the RF maps of the modeled units based on the contour of the change in sensitivity profile and the variability in direction preference. A clump of significant RF blocks with consistent direction preferences are considered to form a subregion when they are separated from other subregions by gaps of non-significant blocks. Thus, two neighboring subregions can emerge because they are not directly connected, or they have different direction preferences (see **4.1.6. Define receptive field subregions**).

I first tested the models having stacked spatial substructures and found that unevenly positioned spatial inputs of lower density led to more fragmented MT RFs, i.e., more subregions (**Fig. 4-7A**, the green curve lies above the blue curve, two-way ANOVA, $p < 0.05$), reaffirming what the RF tuning strength heterogeneity metric revealed above. For the CDS models with spatially uneven inputs, about 75% of the RFs were multi-regioned (about 65% for the even case) when the number of inputs was low (9 locations per direction channel), and this dropped to about 35% (about 5% in for the even case) for units with more than 36 input locations (see data presented as dots in **Fig. 4-8A&B**). Similar trends held for the PDS models (**Fig. 4-7C**). On the other hand, unstacking the V1 inputs significantly reduced the number of subregions for the PDS units, and most such units had only one uniform region (**Fig. 4-7D**). This was consistent with a substantial increase of spatial coverage in the unstacked PDS models, facilitated by the wide range of V1 channels, with no spatial alignment across directions, being recruited due to the broader direction weighting function. When I tried to compare these results to the subregion analysis of Richert et al. (2013), I had to note, while among the group of MT RFs recorded by them, many were multi-direction, almost none of the multi-regioned RFs generated by my models contained subregions with various PDs (**Fig. 4-8**). Given that my definition of the RF subregion is associated with both the adjacency of RF blocks with significant sensitivity intensity, and the consistency of direction preference across them, my results reflected that denser and more uniform (even) spatial sampling covered the RF more effectively, eliminating potential localized sensitivity roll-off within the RF.

For MT neurons with multi-regioned RFs recorded by Richert and colleagues (2013), the smaller subregion's size is typically two thirds of that of the largest region. Correspondingly, I evaluated the size ratio of the secondary subregion to the primary subregion for the multi-regioned model RFs (**Fig. 4-9A&B**, data only shown for CDS units as most PDS units had only single-regioned RFs). I identified the primary (secondary) subregion as the one with the largest (second largest) sum of tuning strength. I found that the secondary subregion was usually less than half the size of the primary subregion. I also noticed that, for CDS models, this ratio was affected by the uniformity of spatial sampling: on average, it was higher for the uneven configuration

than for the even configuration (two-way ANOVA, $p < 0.05$). In the uneven models, because the locations of the V1 inputs are completely random, clumping often occurred. A gap could appear in any part of the RF, resulting in subregions with a wide range of sizes. **Fig. 4-5F:I** demonstrates such an RF of an uneven model, where a gap in the middle separated it into two subregions of comparable sizes. In the even models, the regular spacing of the V1 locations allowed fewer large gaps, often leading to a single uniform RF region. In cases where segregation did occur, the secondary subregion usually appeared as an island near the edge of the primary region and likely much smaller. **Fig. 4-5E:I** shows a typical multi-regioned RF of an even model, where, compared to the primary region, a considerably smaller subregion sat on the side.

Despite the size difference between subregions, Richert and colleagues (2013) reported that the difference in mean tuning strength of subregions to be significantly less substantial, which my findings of the modeled RFs agreed with. **Fig. 4-9C&D** show that the mean tuning strength ratio of the secondary subregion to the primary subregion was generally above 0.85 (data only shown for CDS models). This result is consistent with the fact that all the V1 inputs with positive weights to the CDS model have the same tuning function because the CDS model has a high weight in only one direction channel (**Fig. 3-1B**, red curve). To generate differences in tuning strength across subregions in this type of CDS models would require allowing V1 inputs within each direction channel to vary their tuning functions in a relevant manner (e.g., differences in weight, bandwidth or additive noise).

Although I was able to create multi-regioned model MT RFs with the appropriate parameter setup, the subregions analyzed above almost always originated from heterogeneity in the sensitivity magnitude of the RF maps. While Richert and colleagues (2013) reported that many MT neurons have subregions with PDs that differ by more than 30°, my MT models failed to produce RFs with any large direction selectivity variance. The difference of direction preference between the primary and secondary subregions for the CDS models is shown in **Fig. 4-8**. The angle of difference was small for all conditions, and rarely exceeded 15°.

4.2.1.2. Topographic irregularity of the receptive field

Similar to the multi-peaked MT RFs recorded by Richert and colleagues (2013), the multi-regioned RF maps generated by my MT models were often characterized by undulated sensitivity profiles and jagged RF outlines. To analyze such topographic irregularity, I computed two metrics reported in Richert et al. (2013) – the missing proportion of the RF envelope and the valley-over-peak ratio (see **4.1.7. The valley-over-peak ratio**). These metrics survey the convex hull (the smallest convex polygon that envelopes the raster centers of all significant blocks of the RF, see **Fig. 4-3**) formed by the significant blocks of the RF. The former captures the extent of oddity of the RF shape; the latter describes the fluctuation of the spatial sensitivity function within the RF envelope.

I calculated the missing proportion as the fraction of the enveloped non-significant blocks in the convex hull of the RF. My analysis indicated that its value depended on several parameters of the models. **Fig. 4-10A-D** show that two trends held for both CDS and PDS models. First, the missing proportion decreased as a function

of the density of V1 inputs; second, its value was lower for models with even spatial sampling than those with uneven spatial sampling (two-way ANOVA, $p < 0.05$). On the other hand, unstacking the V1 inputs distinguished the CDS and PDS models. Although, for the CDS units, losing such spatial substructures did not affect this metric (**Fig. 4-10A** vs. **B**), for the PDS units, the unstacked models delivered substantially lower values (**Fig. 4-10C** vs. **D**, two-way ANOVA, $p < 0.05$). These effects are consistent with the findings of Richert et al. (2013) – they reported, on average, the missing proportion is significantly different between single-regioned (0.09) and multi-regioned RFs (between 0.36 and 0.70). For the models, as a higher number of V1 inputs and a uniform distribution of them both led to fewer subregions, a lower missing proportion for those models should be expected. Because the unstacked PDS models almost always formed single-regioned RFs, their missing proportion was extremely low (between 0.07 and 0.14, **Fig. 4-10D**).

To calculate the valley-over-peak ratio, I first identify two isolated peaks within the RF convex hull. I then find the lowest sensitivity point on the line connecting the two peaks. The ratio is the tuning strength of the lowest point divided by that of the secondary peak. Thus, a lower ratio corresponds to a deeper drop of sensitivity amongst peaks, manifesting the rough surface of the RF sensitivity profile. The median valley-over-peak ratio reported by Richert et al. (2013) is 0.21 for multi-regioned RFs, which was at the same level (0.25) for my most multi-region-like models (CDS models with 9 V1 input locations per direction, **Fig. 4-10E&F**).

Fig. 4-10E-H present the results of the valley-over-peak analysis, which resonated with what the missing proportion analysis revealed. For both CDS and PDS models, the valley-over-peak ratio increased as a function of the number of V1 inputs, and its value was higher for the even models than for the uneven ones (two-way ANOVA, $p < 0.05$). I also noticed, for the PDS models, the unstacked configuration produced higher peak-over-valley ratios than the stacked configuration (two-way ANOVA, $p < 0.05$). These findings suggested that the density and the uniformity of spatial sampling, and the presumptive existence or not of localized spatial substructures impacted the spatial undulation of the RF sensitivity level in the same way as they did that of the RF boundary shape.

Together, these measurements of the model MT units confirmed that my simulation, with appropriate parameters, could account for the degree of topographic irregularity of MT RFs delineated in Richert et al. (2013).

4.2.2. The quality of receptive field mapping with random dots

In the RF maps shown above in **Fig. 4-5**, I noticed that for the stacked models with 9-input locations, the RF map appeared to mirror the input map (e.g., **A:I** reflects **A:II**, and **B:I** reflects **B:II**). However, this correspondence broke down for the dense input models. For example, the RF maps in **A:III** and **C:III** did not seem to relate with the clustering in the **A:IV** and **C:IV** input maps, respectively. Thus, I hypothesized that the limited ability to average out the spatiotemporal correlations in the finite random dot stimulus introduced stimulus-dependent artifact in the STA RF maps. This effect would be more apparent for models with more

sufficient spatial coverage. To understand this, I carried out two tests described below: mapping with a different set of random dot stimuli, and also with non-random stimuli.

First, I repeated the RF mapping process using a second set of independently seeded random dot stimuli, which again consisted of 1000 trials, each lasting 1.5 sec duration. **Fig. 4-11A&B** display the two STA RF maps for the same unit (CDS, stacked-even, 9 input locations) generated using two distinct stimulus sets. The maps clearly suggested a consistent spatial structure of the RF, with both reflecting the 9 input locations (the linear Gaussian representation of the input location map is shown in **Fig. 4-11D**). Conducting the same test on a 72-input location model (CDS, stacked-even), however, yielded a strikingly different result. **Fig. 4-11E&F** depict the RF maps from the two sets of random dot stimuli, which exhibited very different local variations in tuning strength and moderate inconsistency in PD.

I quantified the similarity of the two STA maps by taking the correlation coefficient of the blocks encompassed in the convex hull (see **Fig. 4-3**) of the union set of the significant blocks of the two maps. **Fig. 4-12** reveals that the *r*-value is high for the 9-input maps, but is substantially lower for the 72-input maps (mean 0.906 for the 9-input maps, pink histogram; mean 0.563 for the 72-input maps, cyan histogram; Fisher's *Z*-transformed *t*-test, $p < 0.05$, $n = 64$).This comparison yielded two conclusions: (1) a significant fraction of the variance in the STA RF maps could be attributed to the random composition of the stimulus (e.g., on average, only 34.4% of the variance is shared between the two maps in the 72-input case); and (2) the density of the anatomical connections of the model strongly impacted the level of variation across the two maps, indicating that the random dot mapping method might be less reliable for higher spatial sampling density.

Considering the potential limitation of the random dot stimulus, I tested whether mapping with non-random stimuli could provide more precise RF maps of the model MT units. I employed small patches (1.3° diameter) of drifting sinusoidal gratings at twelve directions (30° increments) placed at 1° offsets on a 12×12 grid. The resulting RF maps (calculated using the same PD-vector method as for STA, but with vector summation instead of averaging, for detail, see **4.1.3. Computation of spike-triggered average and the spatial receptive field map**) are shown in **Fig. 4-11C&G**, respectively for the models with 9 and 72 inputs. The resemblance was evident between the grating-patch map (**Fig. 4-11C**) and both random-dot maps (**Fig. 4-11A&B**) for the 9-input unit, whereas, the grating-patch map (**Fig. 4-11G**) of the 72-input model appeared more homogeneous than either of its random-dot maps (**Fig. 4-11 E&F**), and thus was potentially a better reflection of the true underlying spatial structure of the RF.

To validate this, I synthesized the linear estimate of the RF based on the positions of the V1 inputs and the Gaussian spread of the V1 Gabor filters as a "ground-truth" map. The linear estimates for the 9-input and 72-input MT model examples are displaced in **Fig. 4-11D&H**, respectively, and they bore a clear similarity to the grating-patch maps (**Fig 4-11C&G**). I quantified how well the random-dot maps and the grating-patch maps matched the linear estimates using, again, the correlation metric for the pixels within the RF convex hull determined by the random-dot maps. For the 9-input models, the *r*-value was high for both the random-dot and grating-patch maps (**Fig. 4-13A**, brown and indigo histograms, respectively), but was slightly higher for the latter (mean 0.832 for the random-dot maps; 0.866 for the grating maps; Fisher's *Z*-transformed *t*-test, $p <$

0.05, n = 64). For the 72-input models, however, the random-dot maps were only modestly correlated with the linear estimates (mean 0.483, n = 64), whereas the grating-patch maps were strongly correlated (mean 0.816, n = 64) with the linear estimates (**Fig 4-13B**; Fisher's *Z*-transformed *t*-test, $p < 0.05$, n = 64). Therefore, mapping with the grating patches provided a better approximation of the spatial structure of the neuronal RF than mapping with the random dots.

Fundamentally, the inherent randomness of the random dot stimulus has the potential to impart a substantial level of noise to the RF maps derived from the STA. For model units with spatially sparse V1 inputs, whose RFs are hence more likely to have multiple subregions, the RF maps surveyed with the random dots appear to authentically reflect the spatial topography of the V1 inputs. However, for model units with denser spatial coverage of inputs, resulting in more uniform RFs, the dot RF maps present strong local variation patterns that appear inconsistent with the linear RF profiles derived from the underlying spatial structure of V1 connections.

These observations are consistent with the fact that the RF map obtained with random stimuli of finite lengths should depend on both the underlying spatial structure of the RF and the spatial composition of the random motion signals in the stimuli. If the stimuli are infinitely long, the spatial distribution of the motion signals will be uniform on average, and the spatial structure of the RF will be the sole determinant of the resulting map. Given a finite set of stimuli, when the model RF is highly heterogeneous (high "signal"), the robust variation of local sensitivity is able to overcome the noisy nature of the stimulus, and the sparse spatial pattern of the V1 arrangement should dominate the resulting map. However, for a model with dense spatial sampling and a less heterogeneous RF (low "signal"), the undesired random correlation inherent to the stimuli will dominate the map, potentially leading to over-estimated heterogeneity. In fact, for the 72-input models I examined, each set of random dot stimuli imposed its own characteristic spatial noise pattern across the RFs of all model units, in spite of the independent random connectivity pattern underlying each unit (not shown).

The extent of the noise in the random-dot RF maps only becomes obvious because I had a simple ground-truth visual comparison – in our case, a highly uniform spatial RF architecture. No such comparison is available when recording *in vivo*, and therefore it is important to measure the uncertainty in the RF mapping technique. I used the simple *split-half* test on my models by comparing the maps from two independent stimulus sets, which can be applied *in vivo*. My MT models allow me to go further and test whether a more direct, non-random stimulus could be more efficient for assessing heterogeneity. I found that small patches of gratings generate RF maps that are a better match to the model RF estimated from the known underlying architecture of the V1 inputs (**Fig. 4-13**). Thus, when choosing whether to use random or non-random stimuli to map the RFs of visual neurons, neurophysiologists should carefully consider the quality of the outcome, as well as the efficiency of the method. The grating-patch maps made use of about 8,000 spikes and a total stimulus duration of 1770 sec. In comparison, the random-dot maps resulted from 40,000 spikes on average (a median of 9,000 spikes were used in Richert et al., 2013), and a total run of 1,500 sec of stimuli. If the goal is to determine the homogeneity of the MT RF profile and the variation of PD across space, I would recommend using small patches of deterministic stimuli, and placing them in some, but not all, locations in the RF. For example, the grating

paradigm could be used on a checkerboard grid to reduce the running time by half, or to double the amount of data collected for each location, to estimate the RF parameters more accurately, albeit more sparsely. The random dot stimuli, which can place motion signals at any location, test each location only a few times, and lose the ability to average away the spatiotemporal correlations in the stimulus.

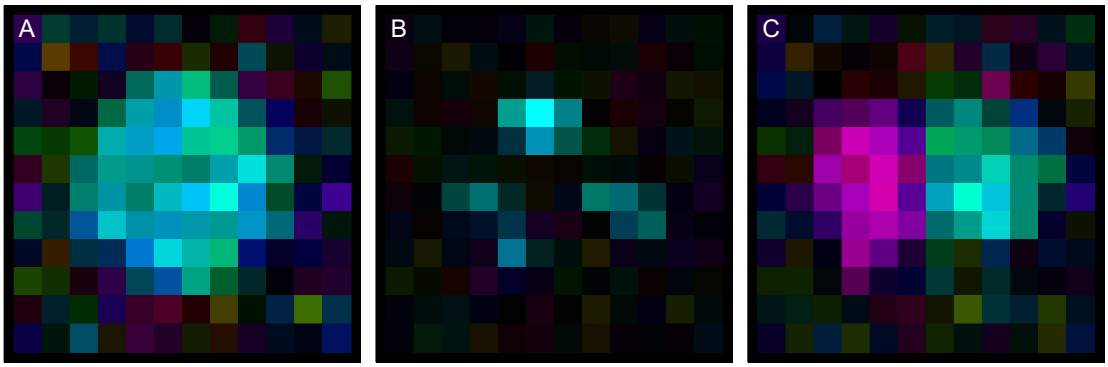

Figure 4-1 Alternative spatial structures of MT receptive fields

(A) A single-region receptive field with homogeneous direction selectivity **(B)** A receptive field with separate regions of similar direction selectivity. **(C)** A receptive field with two regions of different preferred directions. The luminance indicates the tuning strength; the hue indicates the direction preference (see ***Fig. 4-2E***).

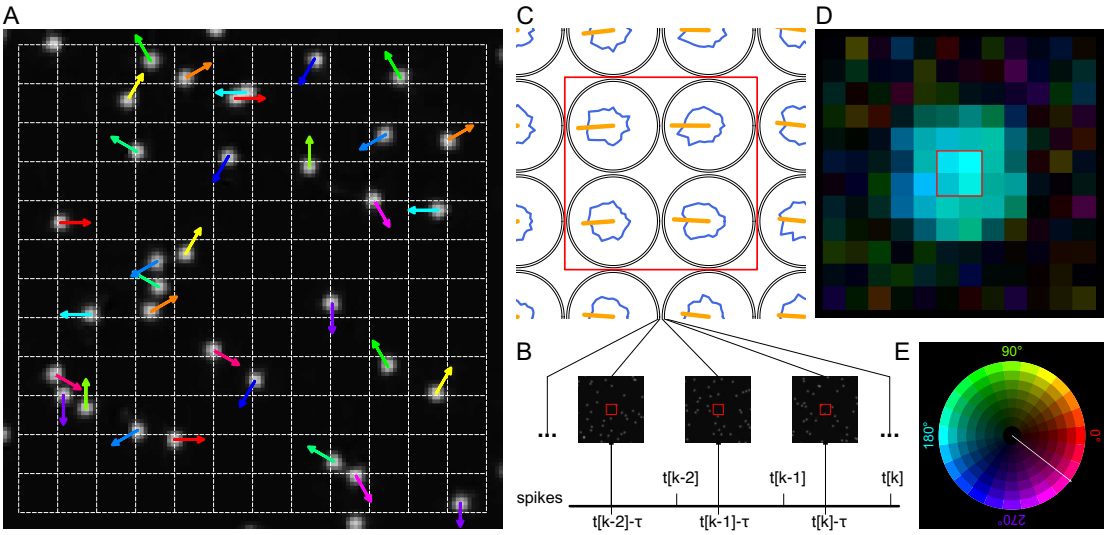

Figure 4-2 Generation of receptive field maps

(A) The random dot stimulus consists of randomly placed dots moving in randomly chosen directions that appear and last for 200 ms. For analysis, the locations of the dots are binned in 1°-by-1° blocks.The arrow indicates the direction of dot motion and matches the color key in **(E)**. **(B)** Spike-triggered average. I compute the spike triggered average by selecting the stimulus frames within a symmetric time window of 500 ms around each spike,*t[k]*. For each stimulus frame of a specified latency, which are denoted by *τ*, I aggregate the motion signals in the stimulus across spikes, and generate a circular histogram of the motion directions of the dots in each block.**(C)** Local direction tuning of receptive field blocks. The histogram for each block can be interpreted as a direction tuning curve (blue polar plots). By treating the histogram as a circular distribution, I compute a mean vector (orange lines).The angle of the vector corresponds to the preferred direction, and the magnitude of the vector is the tuning strength.**(D)** The spatial receptive field map. An optimal latency is chosen to maximize the summed tuning strength across all blocks in a single latency frame.A receptive field map is generated by representing the mean vector for each block with pseudo-color The hue corresponds to the angle of the vector – the preferred direction, and the luminance indicates the magnitude of the vector – the tuning strength.The red squares in **(B)**, **(C)** and **(D)** correspond to the same 2°-by-2° area. **(E)** Pseudo-color representation of the preferred direction and tuning strength. The hue of the motion arrows in **(A)** and the blocks in **(D)** can be mapped to directions using this color wheel. The gradient of luminance corresponds to the variation of tuning strength in**(D)**.

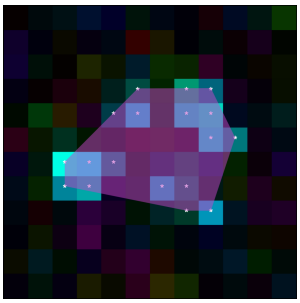

Figure 4-3 An example of the convex hull of a receptive field map

This MT receptive field consists of 4 subregions formed by connected significant blocks (indicated by the white asterisks) in the map. The purple-shaded convex polygon corresponds to the convex hull of the receptive field.

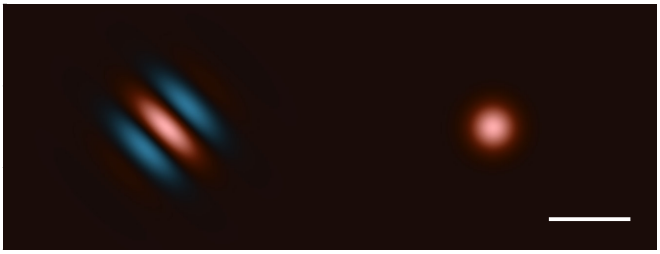

Figure 4-4 The spatial scale of the V1 Gabor filter and the stimulus dot

A Gabor filter is shown on the left, and a stimulus dot on the right.The bar represents 1° of visual angle.

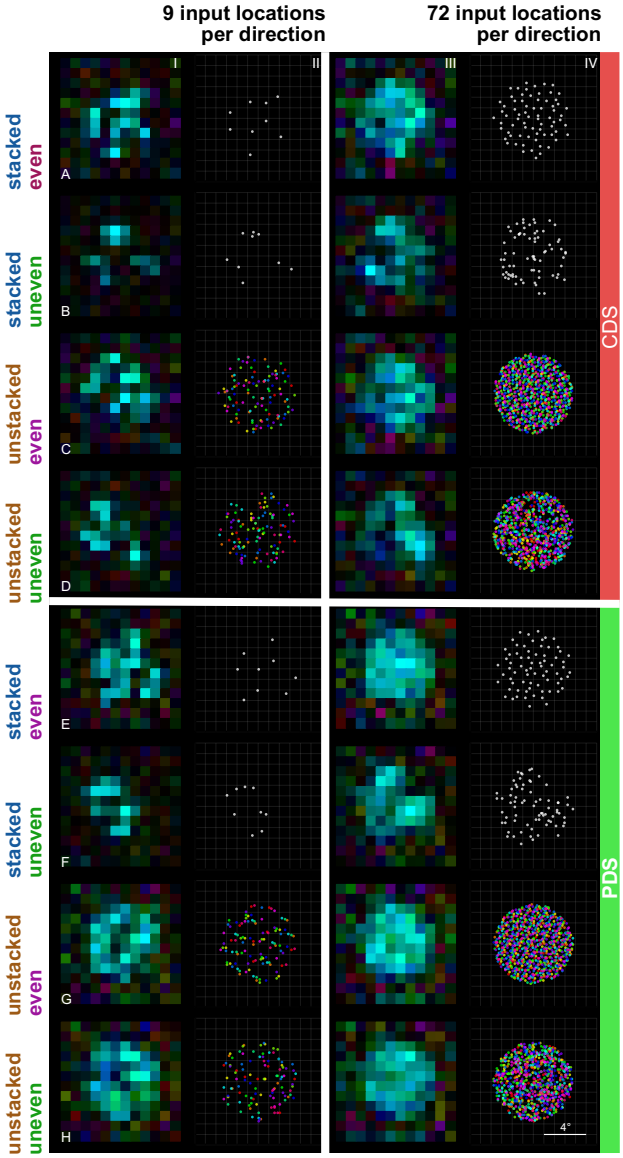

Figure 4-5 Receptive field maps and the organization of V1 inputs for example MT units

The measured receptive field maps of example MT models are shown in **Columns I** and **III**, and the corresponding V1 input locations are shown in **Columns II** and **IV**, where the color of the dots indicates the V1 direction preference. **Columns I** and **II** show the data for the models with 9 V1 inputs per direction channel, and **Columns III** and **IV** for those with 72 inputs. The upper four rows correspond to component MT units with different spatial configurations of V1 channels: **(A)** stacked-even; **(B)** stacked-uneven; **(C)** unstacked-even; **(D)** unstacked-uneven. The lower four rows, **(E)** – **(H)**, show the same cases for pattern MT units.

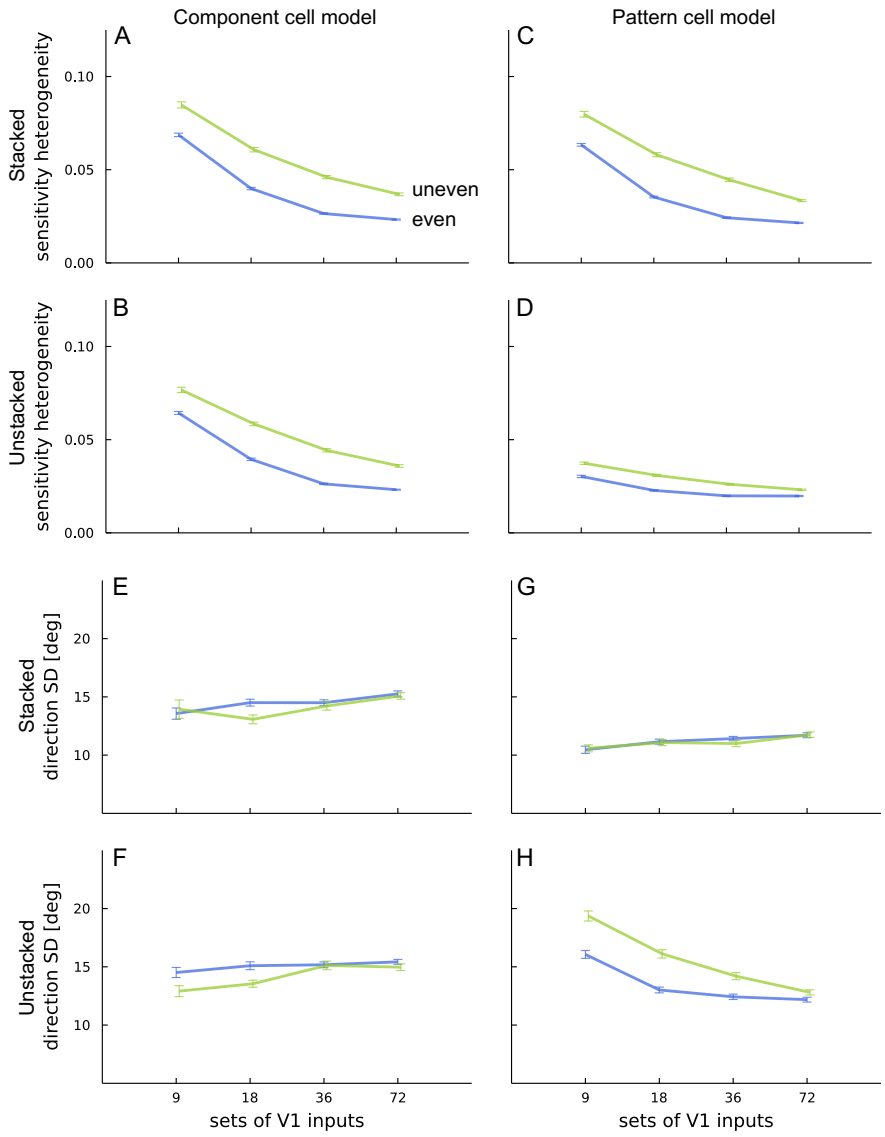

Figure 4-6 Heterogeneity of the spatial profile of sensitivity and the variance of direction preference in model MT receptive fields

(A) The average sensitivity heterogeneity of the receptive field profile for n=64 stacked component models is plotted as a function of the number of V1 inputs per direction channel for the even (blue) and uneven (green) cases.**(B)** Same as **(A)**, but for the unstacked configuration. **(C)** Same as **(A)** but for the pattern models. **(D)** Same as **(C)** but for the unstacked configuration. **(E)** The average, over 64 units, of the standard deviation of the preferred directions across the significant blocks in the receptive field map for the stacked component models.**(F)** Same as **(E)**, but for the unstacked configuration. **(G)** Same as **(E)**, but for the pattern models. **(H)** Same as **(G)**, but for the unstacked configuration. Error bars show SEM.

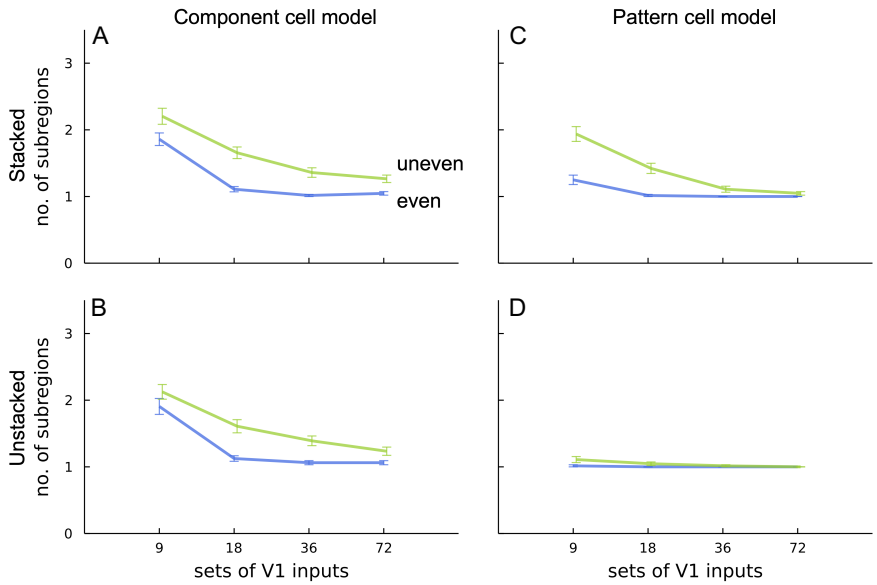

Figure 4-7 The number of subregions in model MT receptive fields

(A) The number of distinct subregions detected in the receptive field for the stacked component models in the even (blue) and uneven (green) cases is plotted as a function of the number of V1 inputs per direction channel.**(B)** Same as **(A)**, but for the unstacked configuration. **(C)** Same as **(A)**, but for the pattern models. **(D)** Same as **(C)**, but for the unstacked configuration. Each data point is the sample mean of n=64 model units. Error bars show SEM.

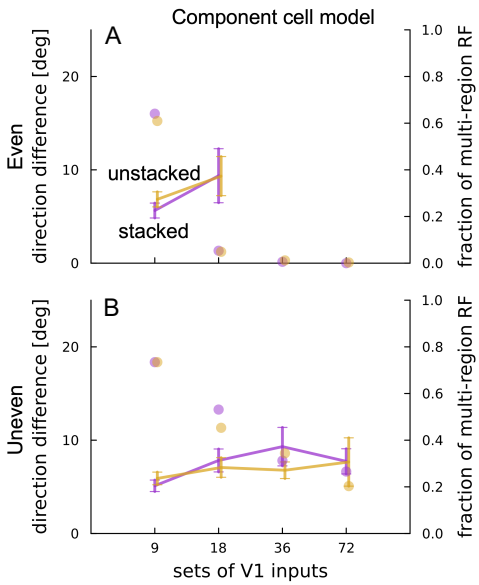

Figure 4-8 Difference of direction preference between the primary and secondary subregions in model MT receptive fields

(A) The curves show the difference between the preferred directions of the primary and secondary receptive field subregions for the component models with the even configuration in the stacked (purple) and unstacked (orange) cases. The dots show the fraction of units with multi-region receptive fields out of the population of simulated units. **(B)** Same as **(A)**, but for the uneven configuration. Each data point in the curves plots the sample mean. Error bars show SEM. The number of data points used to generate the curves for the models with 9, 18, 36 and 72 inputs per direction channel are, respectively: n=41, 36, NA, NA (stacked-even), n=39, 33, NA, NA (unstacked-even), n=47, 34, 20, 17 (stacked-uneven), n=47, 29, 22,13 (unstacked-uneven). Some data points are omitted due to the corresponding models generating predominantly single-region receptive fields.

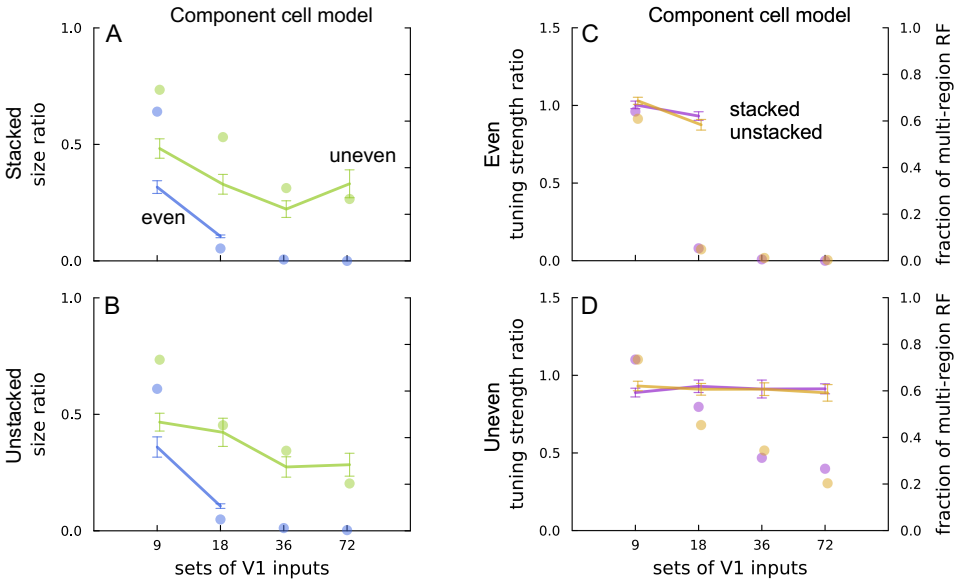

Figure 4-9 Comparison of the spatial profile between the primary and secondary subregions in model MT receptive fields

(A) The curves show the mean ratio of size between the primary and secondary receptive field subregions of the stacked component models in the even (blue) and uneven (green) cases.The dots show the fraction of units with multi-region receptive fields out of the population of simulated units. **(B)** Same as **(A)**, but for the unstacked configuration. **(C)** The curves show the mean ratio of tuning strength of the primary and secondary receptive field subregions for the even component models in the stacked (purple) and unstacked (orange) cases.**(D)** Same as **(C)**, but for the uneven configuration. The number of data points used to generate the curves for the models with 9, 18, 36 and 72 inputs per direction channel is respectively: n=41, 36, NA, NA(stacked-even), n=47, 34, 20, 17 (stacked-uneven), n=39, 33, NA, NA, (unstacked-even), n=47, 29, 22, 13 (unstacked-uneven). Some data points are omitted due to the corresponding models generating predominantly single-region receptive fields.

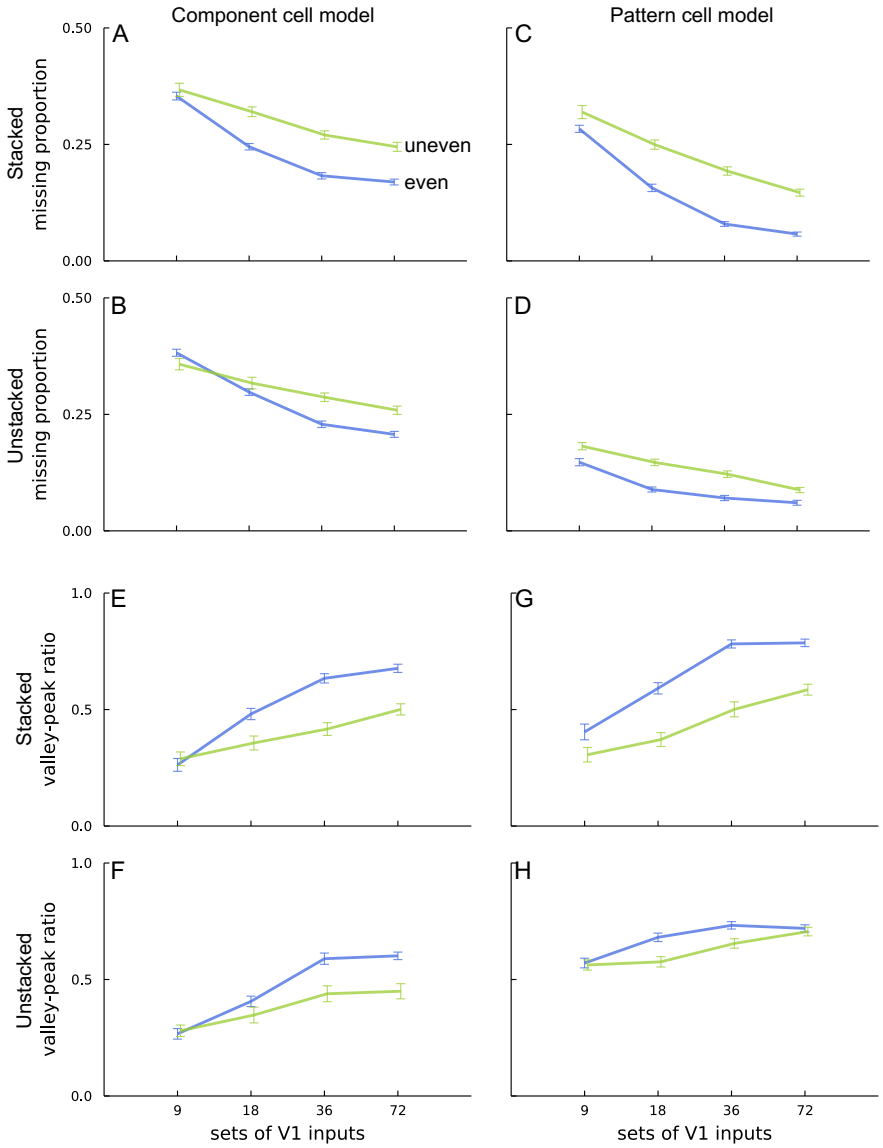

Figure 4-10 Topographic irregularity in model MT receptive fields

(A) The proportion of missing area in the convex hull of the receptive field of n=64 stacked component models is plotted as a function of the number of V1 inputs per direction channel for the even (blue) and uneven (green) cases.**(B)** Same as **(A)**, but for the unstacked configuration. **(C)** Same as **(A)**, but for the pattern models. **(D)** Same as **(C)**, but for the unstacked configuration. **(E)** The average, over 64 units, of the ratio between the tuning strengths of the valley and the second highest peak of the receptive field profile of the stacked component models.**(F)** Same as **(E)**, but for the unstacked configuration. **(G)** Same as **(E)**, but for the pattern models. **(H)** Same as **(G)**, but for the unstacked configuration. Error bars show SEM.

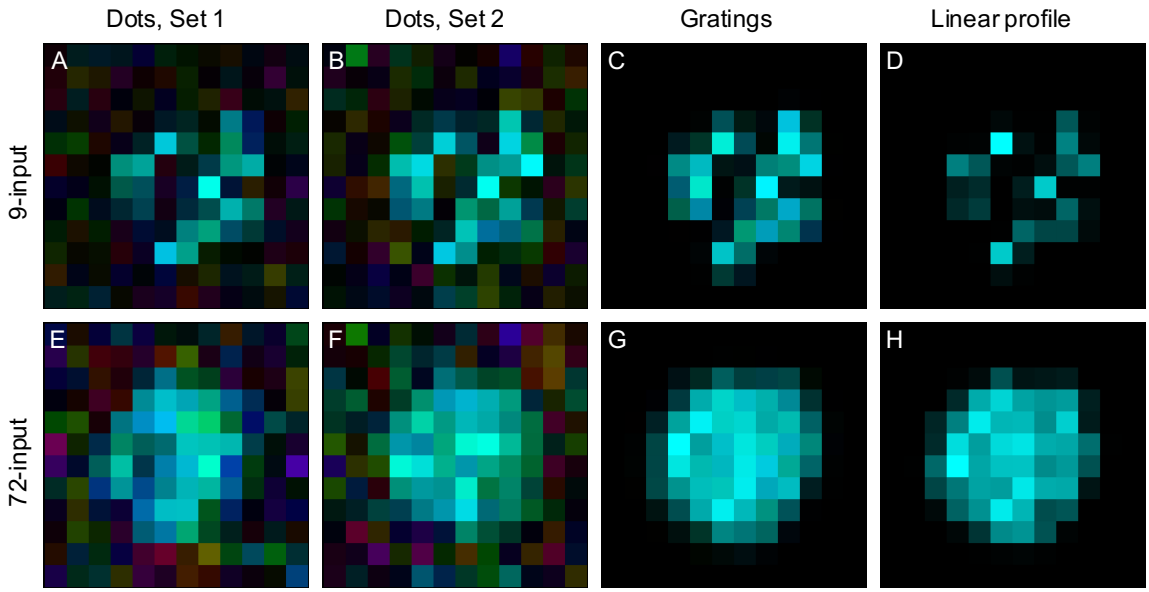

Figure 4-11 Comparison of receptive field maps surveyed with different stimuli

(A) A receptive field map of a stacked-even unit with 9 inputs per direction channel surveyed using a set of random dot stimuli. **(B)** Same as **(A)**, but surveyed using a different set of random dot stimuli. **(C)** Same as **(A)**, but surveyed using patches of small gratings. **(D)** A synthesized linear representation of the V1 input field of the unit in **(A)**. **(E)** A receptive field map of a stacked-even unit with 72 inputs per direction channel surveyed using the same set of random dot stimuli used for **(A)**. **(F)** Same as **(E)**, but surveyed with the same set of random dot stimuli used for **(B)**. **(G)** Same as **(E)**, but surveyed using patches of small gratings. **(H)** A synthesized linear representation of the V1 input field of the unit in **(E)**.

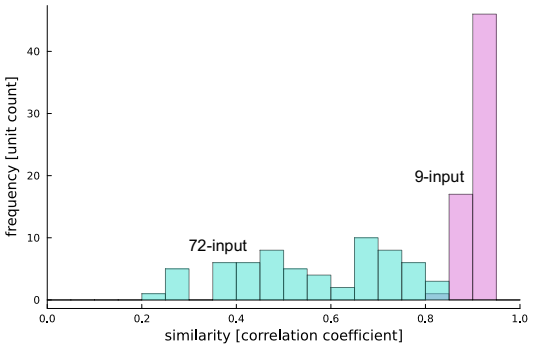

Figure 4-12 Comparison of the consistency of the receptive field maps surveyed with two independently seeded sets of random dot stimuli

The histograms show the distributions of the Pearson correlation coefficients between the maps surveyed with different sets of random dot stimuli. The cyan bars correspond to the stacked-even MT models with 9 V1 inputs per direction channel, and the pink bars correspond to the models with 72 V1 inputs per direction channel.

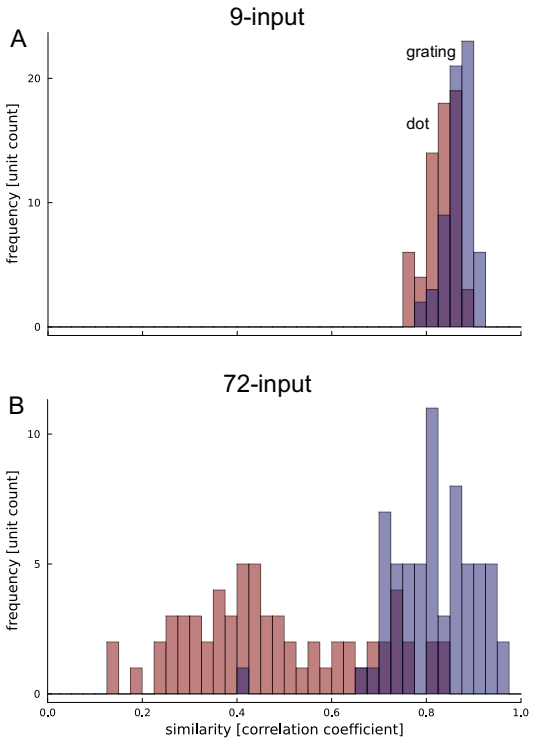

Figure 4-13 Comparison of the faithfulness of the receptive field maps surveyed with different stimuli

(A) The brown histogram shows the distribution of the Pearson correlation coefficients between the receptive field maps surveyed with random dot stimuli and the linear representations of the V1 input fields for the stacked-even models with 9 inputs per direction channel. The indigo histogram shows the distribution of the coefficients between the receptive field maps surveyed with small patches of gratings and the linear representations of the V1 input fields for the same models. **(B)** Same as **(A)**, but for the models with 72 inputs per direction channel.

Chapter 5.

Modeling of spatial normalization in the MT receptive field

MT specializes in processing visual motion: its neurons receive convergent inputs from V1 neurons and integrate them spatially, enlarging the MT RFs up to ten times the diameter of the corresponding V1 afferents (Zeki, 1971; Gattass & Gross, 1981; Albright & Desimone, 1987; Wang and Movshon, 2016). Such spatial integration is essential to motion processing, as it allows MT cells to analyze the overall motion of visual objects, achieved by pooling motion signals extracted at a more local scale in V1 (Movshon et al., 1985). Despite the numerous models that have been developed to understand the properties of the DS cells in MT, they have often overlooked the spatial aspect of V1-to-MT signal integration (Heeger, 1987; Grzywacz & Yuille, 1990; Simoncelli & Heeger, 1998; Bowns, 2002; Perrone, 2004; Rust et al., 2006; Nishimoto & Gallant, 2011; Baker & Bair, 2016).

The front-end of such models typically features motion-sensitive linear filters, which are then followed by a variety of subsequent nonlinear processing steps. The linear-nonlinear cascade has been a widely adopted motif in visual neurophysiology research (Chichilnisky, 2001; Pillow & Simoncelli). With MT sitting in the intermediary stage of the motion processing pathway, several modeling studies have suggested that the integration of V1 signals in MT should be nonlinear (Simoncelli & Heeger, 1998; Tsui et al., 2010). Such an integration paradigm and the associated nonlinearity are usually interpreted in the framework of normalization, and the corresponding spatial mechanisms may help to explain many response properties of MT neurons observed during the interaction of multiple stimuli within the RF (Britten & Heuer, 1999; Heuer & Britten, 2002; Majaj et al., 2007; Kumbhani et al., 2015; Weisner et al., 2020).

Although some established models, such as those of Simoncelli & Heeger (1998) and Rust et al. (2006), have been considered the classic studies of circuits that can account for MT motion processing, there have been very limited efforts to encompass spatial integration in such frameworks, or to understand its connection to motion integration. Using a model that incorporates V1-to-MT spatial integration with classical direction integration circuitry (Rust et al., 2006; Baker & Bair, 2016), this study is the first to investigate how computational schemes implemented at various stages of the V1-MT pathway can give rise to the nonlinear interaction, such as normalization and suppression, as observed in the response to multiple spatially separated stimuli in the MT RF (Britten & Heuer, 1999). I will demonstrate how such a response is a combinatory function of the responses to individual stimuli presented alone, and how potential normalization mechanisms at various levels of V1-MT processing shapes such functions. I found that the nonlinear interaction in the MT RF can be accounted for by either V1-to-MT integrative input normalization, or divisive MT population normalization. However, as the latter paradigm resulted in unrealistic spatial sensitivity profiles of the model MT units, I conclude that integrative input normalization is likely the critical source of spatial normalization in the MT RF.

5.1. Methods

5.1.1. An MT model with spatial integration

The model of Baker & Bair (2016) integrates signals across twelve direction channels through motion energy filters, each hosting a single V1 unit sitting at the center of the model visual field, thus, resulting in an unrealistic MT RF of the same size as the V1 RFs. I introduced spatial integration to the model by allowing the MT unit to sum V1 outputs across 72 spatial locations per direction channel. Such locations are restricted within a radius of 3.75° from the nominal center of the MT RF. These V1 units are organized in the stacked-even spatial configuration (see **3.2.1.4. The stacked-even configuration**), where the V1 locations are aligned across the direction channels, and are chosen through a quasi-random procedure to ensure consistent spacing between them. The outputs of the V1 channels are then weighted based on the channel's PD and the type of direction selectivity of the MT model, i.e., CDS or PDS (for simplicity, only the PDS models are discussed in this chapter, and the CDS models did not render qualitatively different results), regardless of their spatial locations, and summed (for detail, see **Chapter 3. Simulation methods: an MT model framework**).

5.1.2. Nonlinear signal normalization

I implemented three types of nonlinear normalization schemes that operate before, during, and after V1-to-MT integration: (1) V1 tuned normalization in the form of surround suppression, (2) V1-to-MT integrative input normalization, and (3) divisive MT population normalization, respectively. The corresponding formulation is described, respectively in detail in the previous sections: **3.1.2. V1 normalization**, **3.2.2.2. Nonlinear integration**, and **3.3. Divisive MT population normalization**.

Signal normalization at the V1 level (**Eq. 3-4**) may underlie several nonlinear behaviors of the V1 physiology, including response saturation and cross-orientation inhibition (Bonds, 1989; Albrecht & Geisler, 1991; Heeger, 1991, 1992a, 1992b, 1993; DeAngelis et al., 1992; Carandini & Heeger, 1994; Carandini et al., 1997; Tolhurst & Heeger, 1997a, 1997b; Nestares & Heeger, 1997). The model of Simoncelli & Heeger (1998) featured untuned normalization, which may reflect the suppression that occurs within the V1 CRF (Carandini et al., 1997). On the other hand, Rust and colleagues (2006) introduced tuned normalization in their model; this kind of normalization may be linked to V1 surround suppression (Lennie & Movshon, 2005), and is associated with the PDS behavior of their MT model. Provided that untuned normalization does not alter the relative response of the V1 units, and merely limits the dynamic range of the response given the stimulus contrast (Simoncelli & Heeger, 1998), it is omitted in the MT models being inspected in this chapter; the focus is on the effect of tuned normalization, for which surround suppression serves as the source (**Eq. 3-5-6**), and hence the choice to analyze PDS MT models exclusively.

The normalization at the stage of V1-to-MT signal integration is achieved by a nonlinear summation process (**Eq. 3-14**). The formulation of the process takes inspiration from Britten and Heuer's (1999) power-law summation model (**Eq. 2-12**) of the nonlinear interaction of multiple stimuli in the MT RF. The power, γ in

Eq. 3-14, of the nonlinear integration process of my MT model controls the degree of nonlinearity, and functions similarly to the exponent, n in **Eq. 2-12**, in their analysis (for clarity of reference, I will refer to the n of Britten and Heuer's (1999) power-law summation as the *exponent*, which is consistent with their choice of words, and the γ in my linear integration process as the *power*). When $\gamma = 1.0$, this integration is a simple linear averaging process. For $\gamma > 1.0$, as the value becomes larger, the integration will be more prominently dominated by the strongest input, leading to a winner-take-all operation, reducing the contribution of the weaker, however, possibly more prevalent inputs.

The normalization process performed by the MT population involves dividing the output of a single MT unit by the average signal of a group of neighboring MT units with diverse direction preferences and noncoincident RFs (**Eq. 3-15**). This type of normalization has been previously implemented by Simoncelli and Heeger (1998).

5.1.3. The stimulus

I employed the same motion stimuli used by Britten and Heuer (1999) to probe the RFs of the model MT units. Small "generalized" Gabor stimuli (Watson & Turano, 1995) are presented at the nodes of a 5×5 grid that spans 8° (**Fig. 5-1**, see Equation 1 of Britten & Heuer, 1999, for the formulation of the function of the stimulus), while the RFs of the V1 units in the model are distributed within a 7.5° diameter circle. These Gabor stimuli are "generalized" because they are static sinusoidal gratings behind a translating Gaussian window. The SF of the grating, 1.2 cycle/deg, matches the central SF of the V1 units. The spatial SD of the Gaussian window is 0.2° and 0.4°, respectively, for the axes orthogonal and parallel to the orientation of the underlying grating. In comparison, the spatial SD of the Gabor filters of the V1 units is 0.36°. The window travels in the direction orthogonal to the orientation of the grating, consistent with the PD of the MT model, at a speed associated with an effective TF of 10.0 Hz, which matches the central TF of the V1 units in the model. Each stimulus runs for 0.07 s at full contrast, covering a path of 0.58° visual angle.

The stimuli were presented in two types of trial blocks. In the single stimulus block, a single grating appears at one of the nodes of the grid. In the double stimulus block, a pair of stimuli appear simultaneously at two different nodes of the grid.

5.1.4. Gaussian mapping of the MT receptive field

The responses of the MT model in the single stimulus block are used to fit a 2D Gaussian sensitivity function for the spatial RF, which are defined as follows:

$$r(x,y) = A\exp(-\alpha(x-x_c)^2 - \eta(x-x_c)(y-y_c) - \beta(y-y_c)^2) + B \qquad \textbf{Equation 5-1}$$

where $r(x,y)$ corresponds to the response to a single stimulus at spatial location (x,y), and the free parameter pair (x_c, y_c) is the center of the Gaussian function. A and B are also free parameters: A controls the amplitude

of the function, and B accounts for the baseline activity. α, β and η are synthetic free parameters that depend on another set of free parameters, as defined in the following equations:

$$\alpha = \frac{\cos^2\theta}{2\sigma_x^2} + \frac{\sin^2\theta}{2\sigma_y^2}$$
$$\beta = \frac{\sin^2\theta}{2\sigma_x^2} + \frac{\cos^2\theta}{2\sigma_y^2}$$
$$\eta = \frac{\cos 2\theta}{2\sigma_y^2} - \frac{\sin 2\theta}{2\sigma_x^2}$$

Equation 5-2

where σ_x and σ_y are the SDs of the two principal axes of the Gaussian function, and θ represents the angle of the first principal axis with respect to the *x*-axis of the coordinate system.

The raw responses of the model unit are scaled by the height of the corresponding Gaussian profile, $A + B$. Such normalization allows for the fair comparison between units with responses of large amplitude difference.

5.1.5. The power-law summation model

To quantitatively examine the interaction between the pair of concurrent stimuli within the MT RF, I fitted the response (average spike rate) of the models to such stimuli as a *power-law summation function* of the responses to the individual stimuli when they appear alone (Britten & Heuer, 1999). The power-law summation function is described as follows:

$$r_{12} = a(r_1^n + r_2^n)^{\frac{1}{n}} + b$$

Equation 5-3

Here, r_1 and r_2 are the responses to the individual stimuli presented alone, and r_{12} is the response to the pair of same stimuli presented together. The scaling factor, a, the intercept, b, and the exponent, n, are free parameters to be determined through fitting. Among these three parameters, the scaling factor and the exponent are the focus of the analysis, as they reflect the nonlinear interactions between the paired stimuli. **Fig. 5-2** shows variants of the function (**Eq. 5-3**) graphed in the $x(r_1)$-$y(r_2)$-$z(r_{12})$ coordinate system with the values of r_1 and r_2 being limited between 0.0 and 1.0. When $n = 1.0$, the summation is linear, and the function lies on a slanted plane, whose slope is controlled by a (**Fig. 5-2A&B**). In this case, when $a = 1.0$, **Eq. 5-3** is a complete summation (**Fig. 5-2A**); when $a = 0.5$, **Eq. 5-3** is an averaging operation (**Fig. 5-2B**). A smaller a value reflects a stronger suppressive interaction between the paired stimuli. When $n < 1.0$, the plane of the function will start to bulge, forming a convex surface (**Fig. 5-2D**); when $n > 1.0$, the plane of the function will begin to curve in, generating a concave surface (**Fig. 5-2C**). When n is extremely large, the function degenerates to a winner-take-all situation such that $r_{12} \approx a \max(r_1, r_2) + b$ (**Fig. 5-2E**). Therefore, a larger n indicates a more powerful

normalization effect in the sense that the response to a pair of stimuli is dominated by the stimulus that, by itself, elicits the stronger response when presented individually, rather than reflecting a cumulative process.

5.2. Results

5.2.1. Linear spatial integration is insufficient

Signal integration in the MT neurons cannot be fully accounted for by linear summation. The response of an MT neuron to two stimuli presented simultaneously within its RF is typically substantially less than the simple sum of the responses elicited by the same stimuli presented alone (Britten & Heuer, 1999). This spatial normalization operation can be captured by the power-law summation model (**Eq. 5-3**). Britten and Heuer (1999) found that for macaque MT cells, the scaling factor (a) typically ranges from 0.5 to 1.0, and on average is almost exactly halfway between averaging (0.5) and summation (1.0), reflecting a suppressive interaction between the two stimuli; they also observed a normalizing effect between the two stimuli given that the summation is on average modestly nonlinear, characterized by an average exponent (n) value of 2.72, ranging widely across neurons from less than 0.5 to greater than 5.0 (see Figure 10 in Britten & Heuer, 1999).

To verify that linear spatial integration could not achieve such normalization, I configured my MT models in a linear regime, and tested them with Britten and Heuer's (1999) stimulus paradigm. These models lack any normalization operation at either the V1 stage or the MT population level, and the V1-to-MT integration is linear ($\gamma = 1.0$). **Fig. 5-3B** displays the fitted power-law summation model (green meshed surface) for the responses to all pairs of simultaneously presented and spatially separated stimuli (orange stems) within the model MT RF as a function of the responses to the same stimuli when presented individually (I will refer to this function as the *response summation function* from here on). The shape of the fitted function's surface was largely a slanted plane ($n = 1.03$), indicating nearly linear signal summation and the absence of apparent normalization, as expected. The stimuli also did not interact suppressively ($a = 0.96$) with each other, as the value of the fitted function at (1.0,1.0) reached almost 2.0. Additionally, I calculated the population average of the response summation functions for a group of 100 model MT units by binning the data with respect to the responses to the individual stimuli, and taking the geometric mean of the corresponding responses to the paired stimuli. **Fig 5-3A** shows such a population average of the response summation function. It had a relatively smooth and linear transition from the bottom left corner to the top right corner, indicating signal integration was linear within the model MT RFs across the population.

Past modeling research has suggested that tuned normalization at the V1 level is particularly important to pattern motion integration in MT neurons (Rust et al., 2006), and surround suppression is a possible source for this operation (Tsui el al., 2010; Kumbhani et al., 2015). To investigate the potential role of V1 surround suppression in spatial normalization in MT, I added such computation to the linear integration models. The strength of surround suppression is controlled by α_1 in **Eq. 3-5**, and can be quantified by a physiologically meaningful metric – the suppression index (SI) of the V1 unit in the model. I first measure the size tuning curve of the V1 unit by recording its response to a sinusoidal grating that drifts behind an aperture whose diameter

expands in a series. The SI is calculated as the difference between the peak of the size tuning curve and the value at the largest stimulus size, divided by the latter (Cavanaugh et al., 2002).

Overall, the inclusion of V1 surround suppression contributed very little to spatial normalization in the models with linear integration. **Fig. 5-3C&D** illustrate the population-averaged response summation function for a group of 100 model MT units with V1 surround suppression, and the fitted power-law summation model for an example unit, respectively. The V1 SI of these models is 0.74. For the example unit, the fitted function's surface was slightly concave, reflecting very limited sub-linearity ($n = 1.11$), and compared to that of the model without V1 surround suppression, the there was no apparent change in the scaling factor ($a = 1.02$). On the population level, the response summation function was similar in shape to that of the models without V1 surround suppression (**Fig. 5-3C** vs. **A**). To determine if the strength of the surround suppression had any systematic effect on the response summation function, I plotted the population average of the scaling factor and the exponent of the fitted power-law summation model as a function of V1 SI (**Fig. 5-3E&F**, respectively). I found that increasing the strength of V1 surround suppression introduced small increments in the exponent, but it did not affect the scaling factor. Thus, even with V1 surround suppression, an otherwise linear integration regime in my MT model could not account for the degree of nonlinear interaction observed *in vivo* (Britten & Heuer, 1999).

5.2.2. Nonlinear spatial integration and normalization

I speculated about the possible nonlinear neural computation that could provide the mechanism for spatial normalization in the MT RF. A parsimonious candidate that could give rise to Britten and Heuer's (1999) power-law summation nonlinearity could be a normalizing nonlinear stage of spatial integration similar in formulation to the power-law summation model. Such an integrative input normalization scheme is described in **Eq. 3-14**, where the inputs from multiple spatially localized V1 inputs are weighted and combined through a sign-preserving power-law averaging process, and the level of nonlinearity is controlled by the parameter of power, γ.

I created a series of MT models with integrative input normalization/nonlinear integration of various power values, and inspected how the power correlated with the observed level of normalization in the responses of these models. **Fig. 5-4A-G** show the population-averaged response summation functions for MT models with the power equal to ¼, ⅓, ½, 1, 2, 3, and 4. When the power was less than 1.0, the models demonstrated substantial super-linear behavior: either of the paired stimuli could generate strong responses when presented alone; concurrent presentation led to a response many fold stronger than the simple summation of the responses to single stimuli. This kind of super-linear dynamics did not align with the data from the majority of the MT cells recorded by Britten and Heurer (1999), which exhibited *sub-linear normalization* that limits the dynamic range of signal summation (for simplicity, from here on, I will omit the adjective, "sub-linear", and simply refer to this kind of nonlinearity as *normalization)*. For these models, a weak stimulus provided effective suppression such that the response to a strong stimulus drastically decreased when

accompanied by a weaker stimulus (**Fig. 5-4A-C**). On the other hand, the models with a power more than 1.0 exhibited effective normalization: the surface of the response summation function had an apparent concave shape, and the response to a pair of equally strong stimuli was only slightly larger than the response to a strong stimulus paired with a weaker one (**Fig. 5-4E-G**).

The same trend also manifested quantitatively for individual MT units. For the example unit with a power of ½ shown in **Fig. 5-4H**, the surface of the fitted response summation function was convex ($n = 0.51$) with considerable downscaling ($a = 0.85$). When the power is 2.0 (**Fig. 5-4J**), the surface of the response summation function became concave ($n = 2.08$) with no visible scaling ($a = 0.98$).

I then plotted the population average of the fitted scaling factor and exponent against the power of the integrative input normalization operation. I found that there was a nearly exact linear correspondence between the power of nonlinear integration and the fitted power-law summation exponent of the response summation function ($n \approx \gamma$, **Fig. 5-4L**). Therefore, by adjusting the power of the nonlinear integration in the MT models, I could precisely control the level of spatial normalization observed in the output.

The relationship between the power of the MT model and the scaling factor of the fitted response summation function was two-phased (**Fig. 5-4K**). When the power was less than 1.0, the scaling factor was less than 1.0, and reduced as the power decreased further. When the power was more than 1.0, the scaling factor remained slightly below 1.0 and did not vary substantially, reflecting the absence of a suppressive interaction between the paired stimuli. Thus, for MT models with spatial normalization operating at the physiological level ($\bar{n} = 2.72$, Britten & Heuer, 1999), a simple, direct scheme of nonlinear V1-to-MT integration could not account for the suppressive effect observed in Britten and Heuer's (1999) data.

Although V1 surround suppression did not influence the scaling factor of the fitted response summation function for the linear MT models, I explored the possibility of it bringing forth suppressive interactions between paired stimuli in nonlinear MT models. For nonlinear ($\gamma = 2.72$) integration models with V1 surround suppression that varies in strength, **Fig. 5-5** plots the average scaling factor and the exponent of the fitted response summation functions as a function of V1 SI. Stronger V1 surround suppression led to slightly lower scaling factor values, indicating that V1 surround suppression induced a mild suppressive interaction between paired stimuli for the nonlinear models (**Fig. 5-5A**). The exponent also rose moderately as V1 SI increased (**Fig. 5-5B**). Therefore, although V1 surround suppression did not introduce effective spatial normalization in the models with linear integration, it could be a contributing factor that enhanced the existing spatial normalization mechanisms in MT models that demonstrated nonlinear summation.

5.2.3. Suppression from the MT population

Many MT cells have a center-surround structure beyond the CRF: the cell does not respond to stimuli in these surround regions; however, such stimuli can suppressively modulate the cell's response to stimuli within the CRF (Allman et al., 1985; Takana et al., 1986; Born & Tootell, 1992). To determine whether such center-surround interaction could provide the mechanism to produce suppressive spatial interaction in my MT models,

I introduced an MT population normalization stage in the linear integration models with V1 surround suppression (SI=0.74). The normalization is divisive and pools signals from a population of MT units with offset RFs that extend beyond the boundaries of the RF of the unit of interest. To control the behavior of population normalization, I systematically varied the parameters that govern the strength and the threshold of the operation, β_1 and β_2 in **Eq. 3-15**, respectively.

Fig. 5-6 depicts how the values of β_1 and β_2 affected the scaling factor and the exponent of these models' fitted response summation function. The scaling factor dropped, and the exponent rose, as β_1 increased and β_2 decreased. The attainable levels of the two metrics suggested that MT population normalization could achieve spatial normalization and suppression to an extent comparable to that observed by Britten and Heuer (1999) in macaques. These effects were significant when the population signal had a large gain, β_1, to overcome the threshold set by β_2. It is also shown in **Fig. 5-6** that the scaling factor and the exponent are negatively correlated when I varied β_1 and β_2. However, Britten and Heuer's (1999) electrophysiology data did not reflect such a trend, therefore, indicating population normalization might not be the major source for the normalizing and suppressive forces in the MT RF.

These results suggested that spatial summation, being nonlinear in the MT RF, might not be the direct outcome of a nonlinear V1-to-MT integration scheme per se: other nonlinear mechanisms operating at alternative processing stages could lead to similar effects. Given that V1-to-MT integrative input normalization and MT population normalization could both explain spatial normalization in the model MT RF, one might ask which operation resembles more closely the neural mechanism in the visual cortex.

5.2.4. Spatial weighting and sampling density of V1 inputs

The analysis above is based on an implicit assumption that the signal summation scheme is invariant across the MT RF, which is consistent with the MT model setup: in the spatial integration stage of the model (**Eq. 3-11**, **Eq. 3-12**, **Eq.3-13** and **Eq. 3-14**), the weights assigned to the inputs depend only on the PD of the V1 channel but are unaffected by the V1 unit's location in the RF. However, this assumption is not compatible with Britten and Heuer's (1999) data. They plotted the responses to their spatially offset pairs of moving stimuli as a function of the locations of the individual stimuli (from here on, I will refer to this function as the *areal summation function*). The typical geometry of such a function in MT is illustrated in **Fig. 5-7B**. The shape of the function is concave, and the contour lines lie mostly parallel to the axes of the graph. The green point in **Fig. 5-7B** represents a pair of stimuli where Stimulus II is much closer to the center of the MT RF than Stimulus I. If the location of Stimulus II changes slightly, corresponding to the green point shifting vertically, the value of the areal summation function will adjust substantially. On the other hand, if the stimulus further from the RF center (Stimulus I) is to be perturbed, as the green point moves horizontally, the response to the stimulus pair will remain largely the same in value. This kind of behavior in Britten and Heuer's (1999) data suggests that the weighting of the inputs in the MT RF is high at the center, while dropping steeply in the close vicinity of the

center, and is lower and flatter towards the edge of the RF, and if the spatial weighting profile is plotted as a function of the distance from the RF center, the shape should be concave (e.g., like the red curve in **Fig. 5-7A**).

Below, I will demonstrate how the shape of the spatial weighting function affects the geometry of the areal summation function using a simple 1D model. I first specify an RF with a Gaussian-shaped sensitivity profile (solid blue curve in **Fig. 5-7A**), described as follows:

$$R(\varrho) = \exp\left(\frac{-\varrho^2}{2s^2}\right) \qquad \textbf{Equation 5-4}$$

where ϱ is the distance from the center of the RF, ranging from -1.0 to 1.0, with 0.0 representing the center, and $s = 0.5$ is the SD of the Gaussian function. I generated two types of spatial weighting function: Type I has a concave shape (red curve in **Fig. 5-7A**), and is defined as follows:

$$W_{\mathrm{I}}(\varrho) = 1 - \sqrt{1 - (1 - \varrho)^2} \qquad \textbf{Equation 5-5}$$

Type II a convex shape (green curve in **Fig. 5-7A**), and is defined as follows:

$$W_{\mathrm{II}}(\varrho) = \sqrt{1 - \varrho^2} \qquad \textbf{Equation 5-6}$$

where the weights, $W(\varrho)$, applied to V1 inputs depend on the locations, ϱ , of such units.

I then inspected the response to a pair of spatially separated stimuli in this model RF which has a weighted nonlinear ($\gamma = 2.72$) summation scheme similar to that of the full-blown MT model, as follows:

$$R_{12}(\varrho_1, \varrho_2) = \mathrm{P}^{\left(\frac{1}{\gamma}\right)}\left(\frac{1}{2}\left(W(\varrho_1)\mathrm{P}^{(\gamma)}\big(R(\varrho_1)\big) + W(\varrho_2)\mathrm{P}^{(\gamma)}\big(R(\varrho_2)\big)\right)\right) \qquad \textbf{Equation 5-7}$$

where ϱ_1 and ϱ_2 specify the locations of the two stimuli, and $W(\varrho)$ can be either $W_{\mathrm{I}}(\varrho)$ or $W_{\mathrm{II}}(\varrho)$ (see **3.2.2.2. Nonlinear integration**, and **Eq. 3-14** for the definition of the sign-preserving power operation, $\mathrm{P}^{(\xi)}(z)$). **Fig. 5-7B&C** plot the response of the models as a function of the locations of the stimuli, i.e., the areal summation function, when the weighting function is Type I and II, respectively. For the concave-shaped weighting function, Type I (**Fig. 5-7B**), the contour lines of the model response function aligned with the axes of the graph, and the concave shape of the function matched that of the areal summation function reported in Britten & Heuer (1999). When the weighting function's shape was convex, Type II (**Fig. 5-7C**), the model response function demonstrated a different geometry, where the lower left corner started to bulge. The upper left corner was able to maintain a concave surface largely due to the particular curvature of the Gaussian sensitivity profile of the RF. When a flat sensitivity profile (dashed blue curve in **Fig. 5-7A**) was adopted, the entire domain became a convex surface (**Fig. 5-7E**).

Instead of testing how the shape of the spatial weighting profile would affect the geometry of the areal summation function of the full-blown MT models, for the convenience of implementation, I examined, equivalently, how the variation of the spatial density function of the V1 inputs shaped the model response. I manipulated the density of V1 channels across the RF by introducing a centralization process (see **3.2.1.5. Centralization of V1 channels**), by utilizing a nonlinear transform (**Eq. 3-10**) to remap evenly distributed V1 locations towards the center. The extent of the centralization process is controlled by the centralization factor, c. **Fig. 5-8A** shows the spatial profile of the V1 distribution density as a function of the distance to the center of the model MT RF on a relative scale, as the centralization factor varies in increments. The shape of the density functions is concave, and the roll-off is steeper for larger c values.

I built a group of nonlinear integration ($\gamma = 2.72$) MT models with V1 surround suppression (SI = 0.74), while varying the strength of centralization. **Fig. 5-8E,G,I&K** show examples of the arrangement of the V1 locations for the centralization factor ranging from 0.0 to 0.4. The population-averaged areal summation functions for the corresponding MT models are shown in **Fig. 5-8D,F,H&J**. The location of the stimulus was scaled as a fraction relative to the spread of the fitted Gaussian RF profile on the axis connecting the position of the stimulus and the center of the Gaussian function (see **5.1.4. Gaussian mapping of the MT receptive field**). The data were binned by stimulus locations, and the geometric mean was taken within each bin. As the centralization factor increased, the shape of the areal summation function became more concave, and resembled more closely Britten and Heuer's (1999) results (see the shape of the function illustrated in **Fig. 5-7B**). Thus, centralized spatial mapping/weighting of V1 units should be able to account for the kind of spatial dependency of stimulus interaction in the macaque MT RF captured in Britten & Heuer (1999).

Such a mechanism could potentially play a role in the normalization effect in the model MT RF. As the centralization factor rose in value, the exponent of the model's fitted response summation function also increased moderately (**Fig. 5-8C**).

5.2.5. Neural implementation of nonlinear V1-to-MT integration

Before attempting to answer whether V1-to-MT integrative input normalization or divisive MT population normalization bears a closer resemblance to the neural normalization mechanism in MT processing, it is important to first take into account the biological feasibility of implementing such computations. While MT population normalization can leverage the lateral inhibition architecture of MT surround suppression, on the surface, integrative input normalization – applying nonlinear power operations to V1 outputs – does not seem easy to implement biologically. Here, I can show that the response generated by the power-law nonlinear integration MT models can be explained by a *softmax summation model* (Bridle, 1989). The corresponding integration scheme has been suggested in past MT modeling studies, and can be achieved by a lateral, possibly recurrent, network of V1 connections (Reichardt et al., 1983; Nowlan & Sejnowski, 1995; Riesenhuber & Poggio, 1999; Tsui et al., 2010).

The softmax summation model fits the response of an MT unit to the paired stimuli as a weighted sum of the responses to the corresponding individual stimuli (**Eq. 5-9**), and the weights are determined through a softmax function:

$$\omega_i = \frac{\epsilon^{r_i}}{\sum_k \epsilon^{r_k}} \qquad \textbf{Equation 5-8}$$

$$r_{12} = u(\omega_1 r_1 + \omega_2 r_2) + v \qquad \textbf{Equation 5-9}$$

where ϵ, the base of the exponential operation, u and v, the scaling factor and the intercept of the summation, respectively, are the parameters to be determined by the fitting process. r_1 and r_2 are respectively the responses to either of the two stimuli presented alone, and r_{12} is the response to the same pair of stimuli when presented together. **Eq. 5-8** and **Eq. 5-9** are a nonlinear summation process, and, similar to how n modulates the degree of nonlinearity of the power-law summation model (**Eq. 5-3**), here, ϵ controls the level of nonlinearity. When $\epsilon = 1.0$, the summation is linear, and in the $x(r_1)$-$y(r_2)$-$z(r_{12})$ space, the function (**Eq. 5-9**) lies on a slanted plane whose slope is controlled by u (**Fig. 5-9A**). When $\epsilon < 1.0$, the plane of the function will start to bulge, forming a convex surface (**Fig. 5-9C**); when $\epsilon > 1.0$, the plane of the function will begin to curve in, generating a concave surface (**Fig. 5-9B**). When ϵ is extremely large, the function degenerates to a winner-take-all situation such that $r_{12} \approx a \max(r_1, r_2) + b$. Therefore, a larger ϵ value is associated with a more powerful normalization effect in the data, where the response to a pair of stimuli is dominated by the stimulus that elicits the stronger response when presented alone.

Comparing the power-law summation model and the softmax summation model fitted to the data set generated by an MT model with moderately nonlinear integration ($\gamma = 2.72$), V1 surround suppression (SI = 0.74) and centralization ($c = 0.4$), there was no apparent distinction between the shapes of the fitted power-law and softmax summation models (compare **Fig. 5-9D**, power-law summation model, $n = 3.55$, to **Fig. 5-9E**, softmax summation model, $\epsilon = 16.86$). For a group of model MT units, I computed the variance explained to evaluate the goodness of fit for the two summation models, and they accounted for the data equally well: the mean variance explained was above 0.95 for both models (**Fig. 5-9F**). The nonlinear interaction within the RF of the model MT units with power-law nonlinear integration could be captured by a softmax response summation function, indicating that the behaviors of the power-law nonlinearity and the softmax nonlinearity were similar. Therefore, the integrative input normalization process adopted by my MT models could potentially be implemented through a lateral network operating on the V1 inputs to MT through softmax-like mechanisms.

5.2.6. Spatial and motion sensitivity of normalization models

To gain a deeper understanding of the differences between the two normalization schemes – V1-to-MT integrative input normalization and MT population normalization, I analyzed their impact on the spatial RF

profile and the direction selectivity of the model MT units as measured by standard physiological characterization techniques.

I generated two groups of PDS MT models with either normalization scheme, both having V1 surround suppression (SI = 0.74) and centralization ($c = 0.4$). The strengths of the normalization ($\gamma = 2.72$ for integrative input normalization, and $\beta_1 = 10.0$ and $\beta_2 = 0.05$ for population normalization) were chosen such that the response and areal summation functions demonstrated a level of nonlinear interaction comparable to the data of Britten & Heuer (1999). The population response and areal summation functions, respectively, of the models are shown in **Fig. 5-10A&C** for integrative input normalization, and in **Fig. 5-10B&D** for population normalization.

I first probed the models with drifting gratings and plaids to survey their direction selectivity. The direction tuning curves of the example model unit with population normalization shown in **Fig. 5-10G** (pink curves) demonstrated clear pattern direction selectivity, and varying the strength of normalization did not impact the pattern index (PI, which quantifies the level of pattern direction selectivity; for the calculation method, see **6.1.3. The pattern index**) of such models (**Fig. 5-10I**). On the other hand, for the model unit with integrative input normalization shown in **Fig. 5-10G** (teal curves), the direction tuning curves for gratings are excessively wide, and the plaid direction tuning curves did not demonstrate typical pattern direction selectivity. In fact, a higher level of nonlinearity in such models was associated with impaired pattern motion sensitivity, as **Fig. 5-10H** illustrates a drastic reduction of PI when the power of nonlinear integration was raised from 1.0 to 2.0. Such results suggested the potential anti-correlation between spatial normalization and pattern motion sensitivity in MT.

To characterize how the strength of normalization affected the spatial profile of the model MT RFs, I compared two approaches for mapping the RF. First, I measured the size tuning curves of the MT models using the standard paradigm in which a series of circular grating patches is presented at varying diameters. **Fig. 5-10E&F** show such curves for the MT models with either integrative input normalization or population normalization, respectively, with varying levels of nonlinearity. A common trend shared by the two sets of models was that the increase of normalization strength led to a left shift of the peak of the size tuning curve. For the models with population normalization, this is because more intense normalization created a stronger suppressive field.

Second, I quantified the RF size of these models as the half-height width of the Gaussian RF spatial response profile mapped using single-stimulus presentation (see **5.1.4. Gaussian mapping of the MT receptive field**). The resulting size values are indicated as vertical lines superimposed on the size tuning curves in **Fig. 5E&F**. For both models, the peaks of the size tuning curves were to the left of the RF half-width values (except in the cases of no normalization: $\gamma = 1.0$ and $\beta_1 = 0.0$). However, what set the models apart was that the RF half-width was not affected by the strength of population normalization, whereas it in fact expanded as the level of nonlinearity increased in the models with integrative input normalization. It appeared that, for the MT models with nonlinear integration, different spatial tests might lead to contradicting conclusions about the relationship between the spatial scale of the MT RF and the degree of the nonlinearity: as the power of nonlinear integration

increased, the size tuning curve of the MT unit narrowed, while the mapped spatial response profile widened. In **5.2.6.1. A conceptualized 1D model of nonlinear V1-to-MT receptive field integration**, I will use a conceptualized 1D V1-MT RF model to illustrate an intuitive explanation for how these seemingly conflicting results are compatible under the context of power-law integration.

When comparing the models with integrative input normalization ($\gamma = 2, \bar{n} = 2.67$) and MT population normalization ($\beta_1 = 5.0, \beta_2 = 0.005, \bar{n} = 2.57$) at a similar level of nonlinearity, the shape of the size tuning curve of the integrative input normalization model was significantly less contracted (**Fig. 5-10E** vs. **F**) and more realistic to the size tuning curve of macaque MT neurons (Tsui & Pack, 2011), suggesting that the former normalization scheme might share more similarity to the neural computation in primate MT.

5.2.6.1. A conceptualized 1D model of nonlinear V1-to-MT receptive field integration

To understand how RF spatial mapping and the size tunning protocol could generate seemingly paradoxical observations regarding the effect of the nonlinearity of signal integration on the RF size estimation for my complex, full-blown MT model, and analyze it in terms of a simpler, conceptual system, here, I demonstrate, using a minimalistic 1D model, how nonlinear integration of the V1 RFs can shape the spatial characteristics of the MT RF.

I first set up an MT cell with connections from an array of V1 neurons that are evenly spaced, and have Gaussian-shaped CRFs and divisive suppressive surrounds. The dashed blue curves in **Fig. 5-11A** illustrate the organization of such V1 RF profiles mapped as the spatial response to a small stimulus. The MT cell integrates V1 signals with power-law nonlinear summation as described as follows:

$$l(x) = \left(\sum_i h_i^{\gamma}(x) \right)^{\frac{1}{\gamma}}$$

Equation 5-10

where $h_i(x)$ is the spatial response of the ith V1 cell, and $l(x)$ is the spatial response of the MT neuron. The solid red curves in **Fig. 5-11A** show how the width of such an MT spatial response, i.e., the RF profile, expanded as a function of the power (γ) of the integration. The MT RF profile grew wider for higher power values. This was because a higher power caused the MT RF profile to be dominated by the peak values of the V1 cells' RF profile curves. As the power grew, a stronger winner-take-all effect caused the MT RF profile to converge to the trapezoidal envelope of the V1 spatial response curves, thus explaining the expansion of the RF size in the full blown MT models with integrative input normalization when the level of nonlinearity increased.

Fig. 5-11B displays the contraction of the size tuning curve (solid red curves, measured as the response to rectangular pulses of varying widths) of the 1D model as a function of the power of integration, which is similar to the behavior of the full-blown MT models with integrative input normalization (compared to **Fig. 5-10E**). Such a trend could be explained as follows: in the 1D model, because each V1 cell has a unique distance to the

center of the MT RF, when the probing signal, a centered rectangle signal, expanded in size, the resulting "size tuning curves" – the responses to the rectangle signal as a function of its width – of these V1 cells exhibited interleaved peaks of unequal heights (**Fig. 5-11B**, dashed blue curves). The taller blue curves with peaks on the left side of the figure correspond to the V1 units sitting at the central area of the MT RF; the flatter, more right-shifted blue curves correspond to the V1 units located in the progressively more peripheral regions. When the power of integration was high, due to the power-law nonlinearity, the shape of the MT cell's size tuning curve was biased towards the V1 units at the center with tall curves, hence a peak on the left side of the graph. Conversely, if the power was closer to 1.0, the shape of the MT size tuning curve would approach the average over the V1 "size tuning curves", causing the peak to shift to the right (**Fig. 5-11B**, solid red curves).

The fact that a simple numerical exercise like this is able to utilize the conceptualized principles from a complex system and produce simulation results that can assist the interpretation of findings, which would otherwise have been considerably more challenging to analyze, is the precise testament to the values that modelling studies bring to neuroscientific research. Through modelling of the sensory system, one can study the interaction of a multitude of stimuli with a myriad of neural circuit variants, and streamline the process of hypothesis testing as well as generate insightful conjectures that can be tested in future animal experiments.

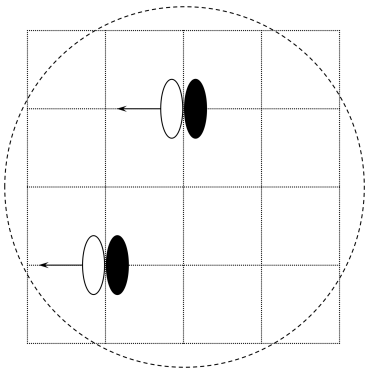

Figure 5-1 Generalized Gabor stimuli

The stimulus consists of generalized Gabor patches that can appear either individually or in a pair at the nodes of a 5×5 grid. The grid covers the receptive field of the MT unit and spans 8° in each dimension.

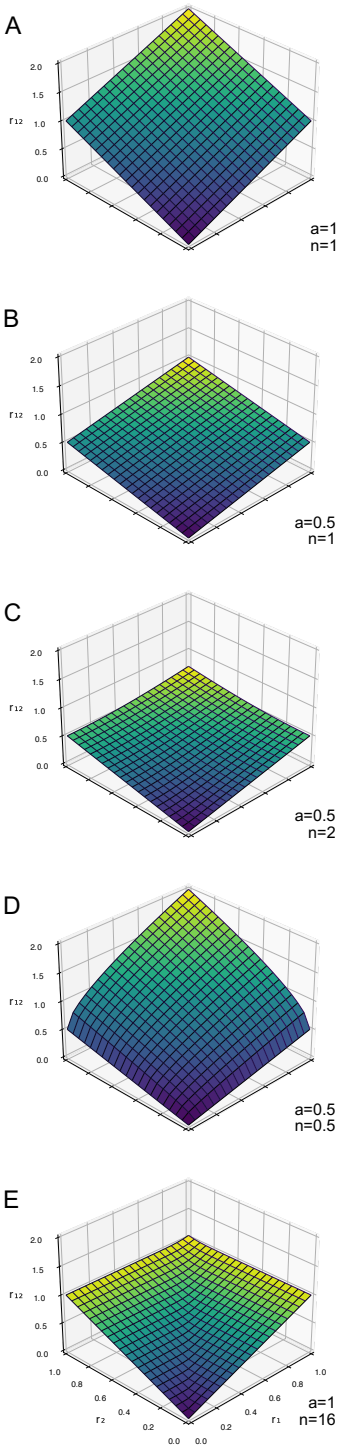

Figure 5-2 The power-law summation model

The shape of the surface of the power-law summation model, which is controlled by the scaling factor (*a*) and the exponent (*n*): **(A)** *a=1*, *n=1*; **(B)** *a=0.5*, *n=1*; **(C)** *a=0.5*, *n=2*; **(D)** *a=0.5*, *n=0.5*; **(E)** *a=1*, *n=16*. The r_1 and r_2 axes correspond to the responses to the stimuli if they have appeared individually; the r_{12} axis corresponds to the response to the stimuli if they have appeared in a pair simultaneously.

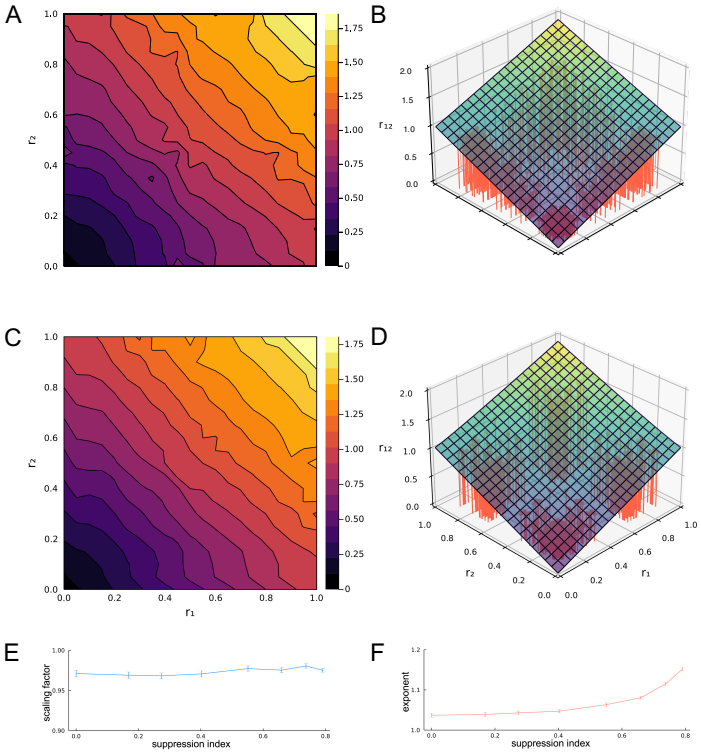

Figure 5-3 Response summation functions of models with linear V1-to-MT integration

(A) The population-averaged response summation function of 100 models with linear V1-to-MTintegration and without V1 surround suppression. The r_1 and r_2 axes correspond to the responses to the stimuli if they have appeared individually; the pseudo-color of the heatmap corresponds to the response to the stimuli if they have appeared in a pair simultaneously **(B)** The response summation function of an example model unit with linear V1-to-MTintegration and without V1 surround suppression. The r_{12} axis corresponds to the response to the stimuli if they have appeared in a pair simultaneously The orange vertical stems indicate the measured response data points for the example unit.The curved surface represents the fitted power-law summation model. **(C)** Same as **(A)**, but for linear integration models with V1 surround suppression. **(D)** Same as **(B)**, but for an example linear integration model unit with V1 surround suppression.The V1 suppression index of the models in **(C)** and **(D)** is 0.74. **(E)** The change of the scaling factor of the response summation function fitted to the power-law summation model as the strength of V1 surround suppression, represented by the V1 suppression index, varies. Each data point is the sample mean of n=100. Error bars show SEM.**(F)** Same as **(E)**, but for the exponent.

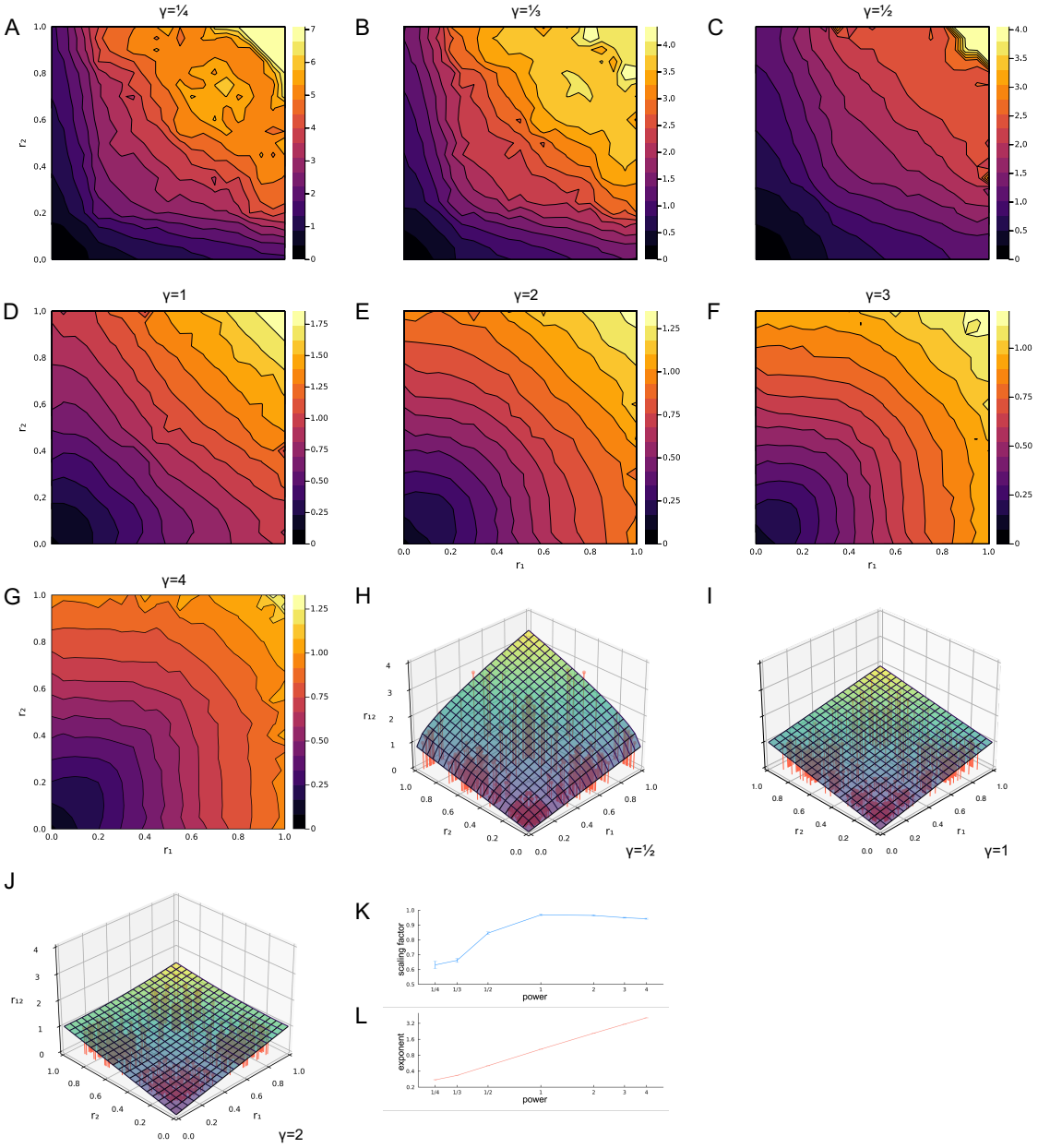

Figure 5-4 Response summation functions of models with nonlinear V1-to-MT integration

(A) The population-averaged response summation function of 100 models with nonlinear V1-to-MTintegration, with a power of *γ=¼*; **(B)** *γ=⅓*; **(C)** *γ=½*; **(D)** *γ=1*; **(E)** *γ=2*; **(F)** *γ=3*; **(G)** *γ=4*. The r_1 and r_2 axes correspond to the response to the stimuli if they have appeared individually; the pseudo-color of the heatmap corresponds to the responses to the stimuli if they have appeared in a pair simultaneously. **(H)** The response summation function of an example model unit with nonlinear V1-to-MT integration, with a power of *γ=½*; **(I)** *γ=1*; **(J)** *γ=2*. The r_{12} axis corresponds to the response to the stimuli if they have appeared in a pair simultaneously. The orange vertical stems represent the measured response data points for the example unit. The curved surface represents the fitted power-law summation model.**(K)** The change of the scaling factor of the response summation function fitted to the power-law summation model as the power of the nonlinear integration varies. Each data point is the sample mean of n=100. Error bars show SEM.**(L)** Same as **(K)**, but for the exponent.

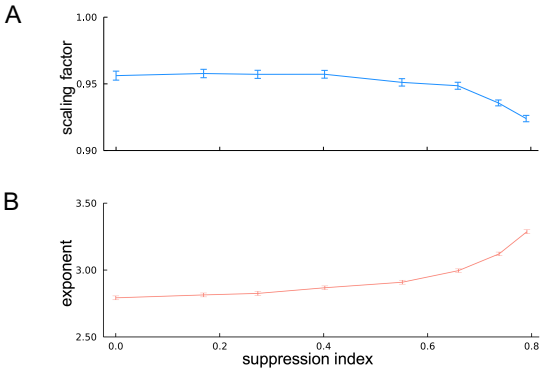

Figure 5-5 Response summation function characteristics of nonlinear V1-to-MT integration models with V1 surround suppression

(A) The change of the scaling factor of the response summation function of models with nonlinear V1-to-MT integration (γ=2.72) fitted to the power-law summation model as the strength of V1 surround suppression, represented by the V1 suppression index, varies. Each data point is the sample mean of n=100. Error bars show SEM. **(B)** Same as **(A)**, but for the exponent.

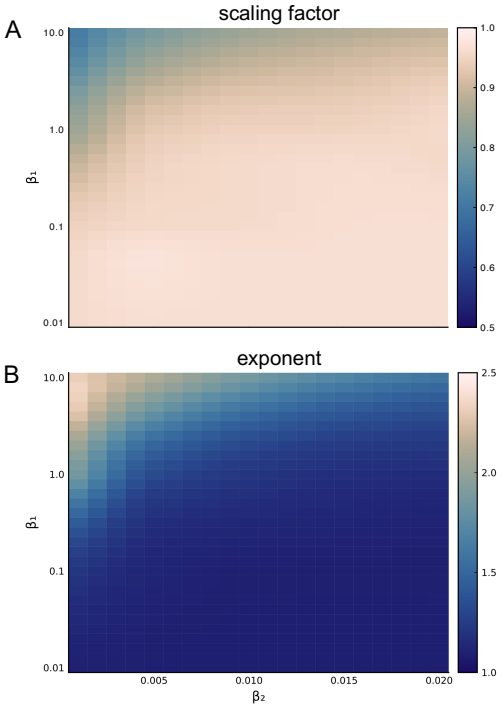

Figure 5-6 Response summation function characteristics of linear V1-to-MT integration models with divisive MT population normalization

(A) The change of the scaling factor, represented by the pseudo-color of the heatmap, of the response summation function fitted to the power-law summation model as the parameters of the population normalization vary. The data are the sample mean of n=100. **(B)** Same as **(A)**, but for the exponent.

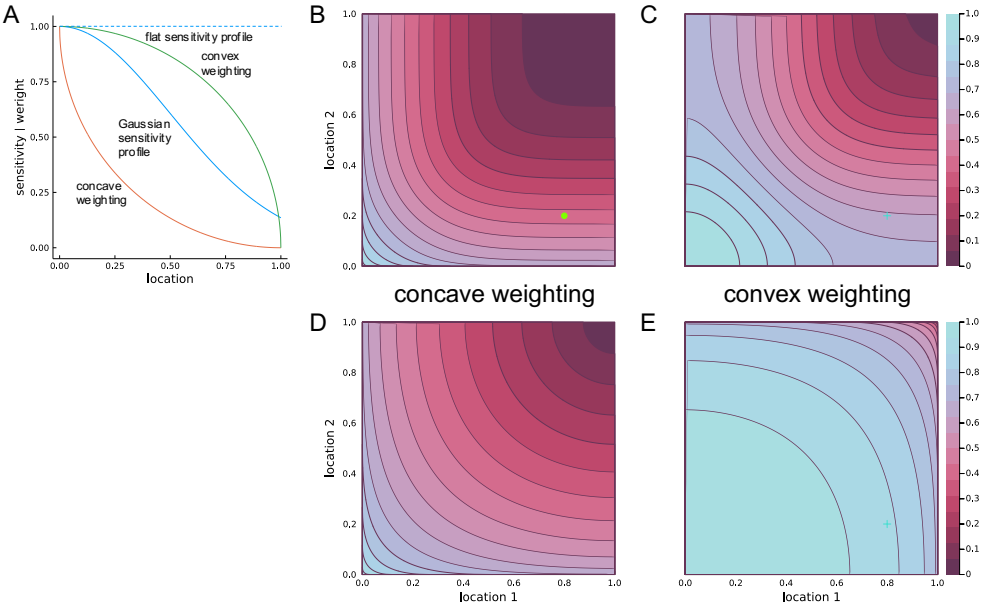

Figure 5-7 Simulation of spatially weighted integration in a conceptualized 1D receptive field model

A conceptualized 1D receptive field model with spatially weighted integration accounts for the shape of the areal summation function. **(A)** The solid blue curve demonstrates a Gaussian spatial sensitivity profile of the model receptive field. The dashed blue line depicts a flat profile. For any point on these curves, the ordinate value represents the strength of the input generated by a conceptual point stimulus placed at the location corresponding to the abscissa value, the relative distance from the center of the receptive field. When multiple inputs are present, the model unit integrates the signals using weighted power-law integration (γ=2.72). The weight depends on the location of the input. Two types of spatial weighting functions were tested: Type I – concave weighting, shown as the red curve; Type II – convex weighting, shown as the green curve. **(B)** The areal summation function of the model with a Gaussian spatial sensitivity profile and Type I spatial weighting. The Location 1 and Location 2 axes correspond to the locations of two inputs appearing simultaneously in the receptive field. The pseudo-color of the heatmap corresponds to the response elicited by the two inputs. **(C)** Same as **(B)**, but for the model with Type II spatial weighting. **(D)** Same as **(B)**, but for the model with a flat spatial sensitivity profile. **(E)** Same as **(C)**, but for the model with a flat spatial sensitivity profile.

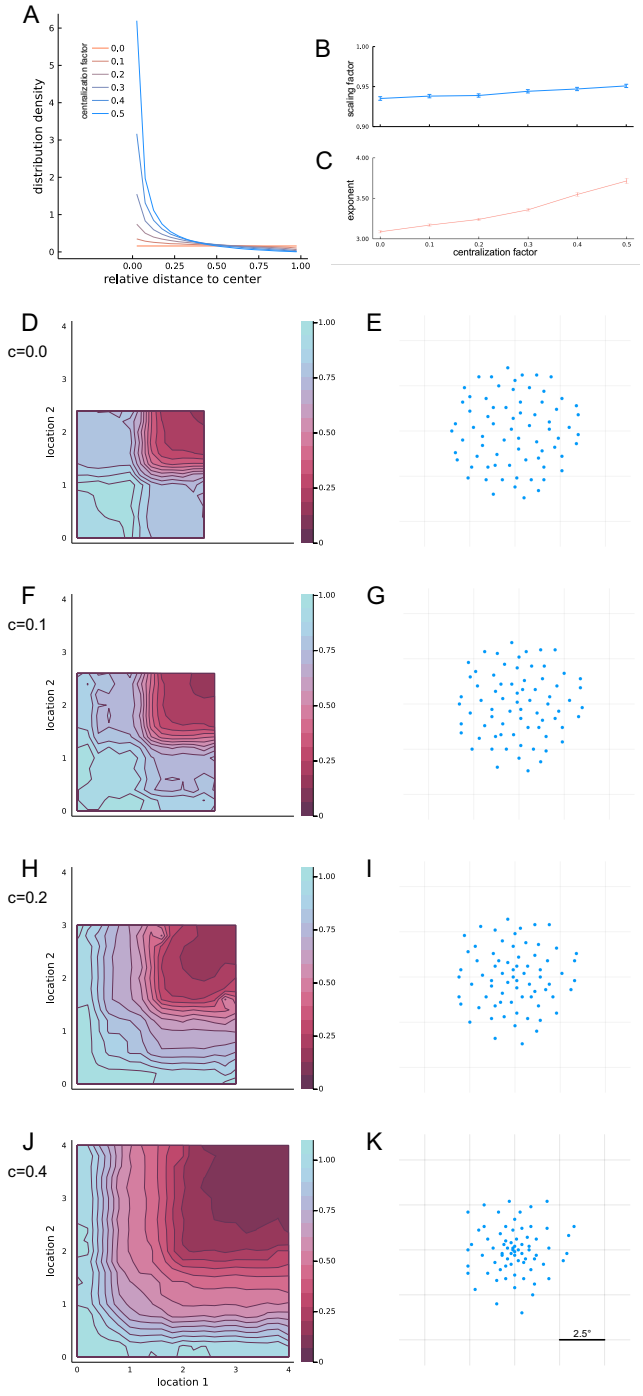

Figure 5-8 Centralized distribution of V1 input locations in model MT receptive fields

Centralized distribution of V1 input locations accounts for the shape of the areal summation function for the MT models with nonlinear V1-to-MT integration (*γ=2.72*). **(A)** The spatial functions of the relative distribution density of V1 inputs at different centralization levels. The centralization level is parameterized by the centralization factor, *c*, of the model. **(B)** The change of the scaling factor of the response summation function fitted to the power-law summation model as the level of centralization varies. Each data point is the sample mean of n=100. Error bars show SEM. **(C)** Same as **(B)**, but for the exponent. **(D)** The population-averaged areal summation function of 100 models with a centralization factor of *c=0.0* (no centralization). The Location 1 and Location 2 axes correspond to the locations of two inputs appearing simultaneously in the receptive field. The pseudo-color of the heatmap corresponds to the response elicited by the two inputs. **(E)** The distribution of example V1 locations when *c=0.0* (no centralization). **(F)** and **(G)** Same as **(D)** and **(E)**, but *c=0.1*. **(H)** and **(I)** Same as **(D)** and **(E)**, but *c=0.2*. **(J)** and **(K)** Same as **(D)** and **(E)**, but *c=0.4*.

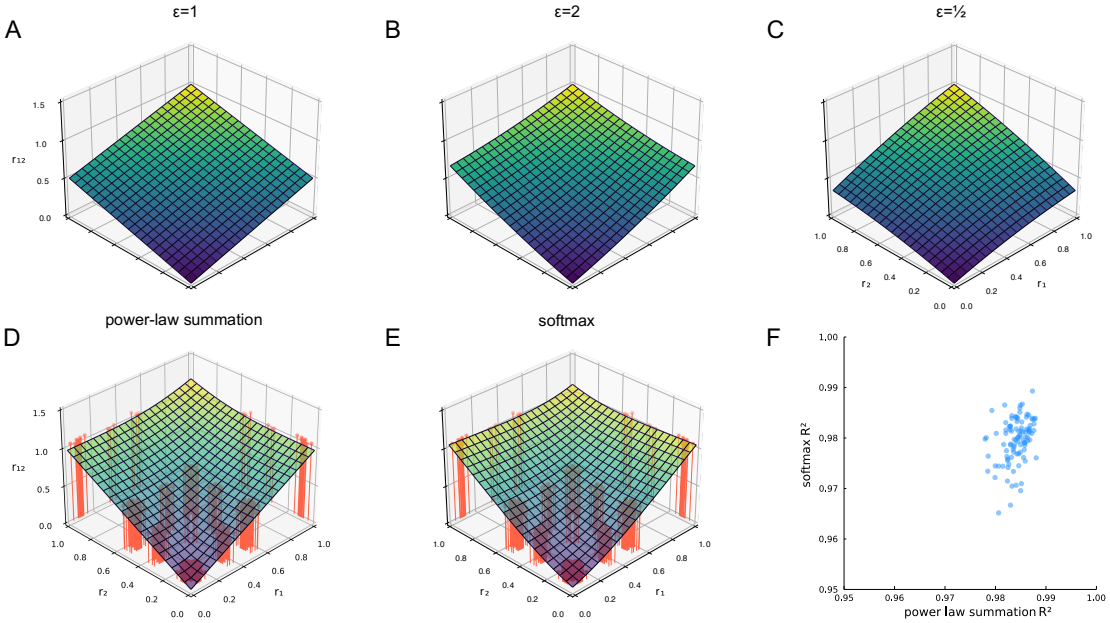

Figure 5-9 The softmax summation model

The softmax summation model accounts for the response summation function for the MTmodels with nonlinear V1-to-MT integration (*γ=2.72*). The curvature of the surface of the softmax summation model is controlled by the base,*ε*: **(A)** *ε=1*; **(B)** *ε=2*; **(C)** *ε =½*. The r_1 and r_2 axes correspond to the responses to the stimuli if they have appeared individually; the r_{12} axis corresponds to the response to the stimuli if they have appeared in a pair simultaneously **(D)** The response summation function of an example model unit fitted to the power-law summation model.The orange vertical stems indicate the measured response data points for the example unit.The curved surface represents the fitted power-law summation model. **(E)** Same as **(D)**, but for the response summation function fitted to the softmax summation model.**(F)** The variance explained of the response summation functions for a group of 100 MTmodel units fitted to the power-law and softmax summation models.

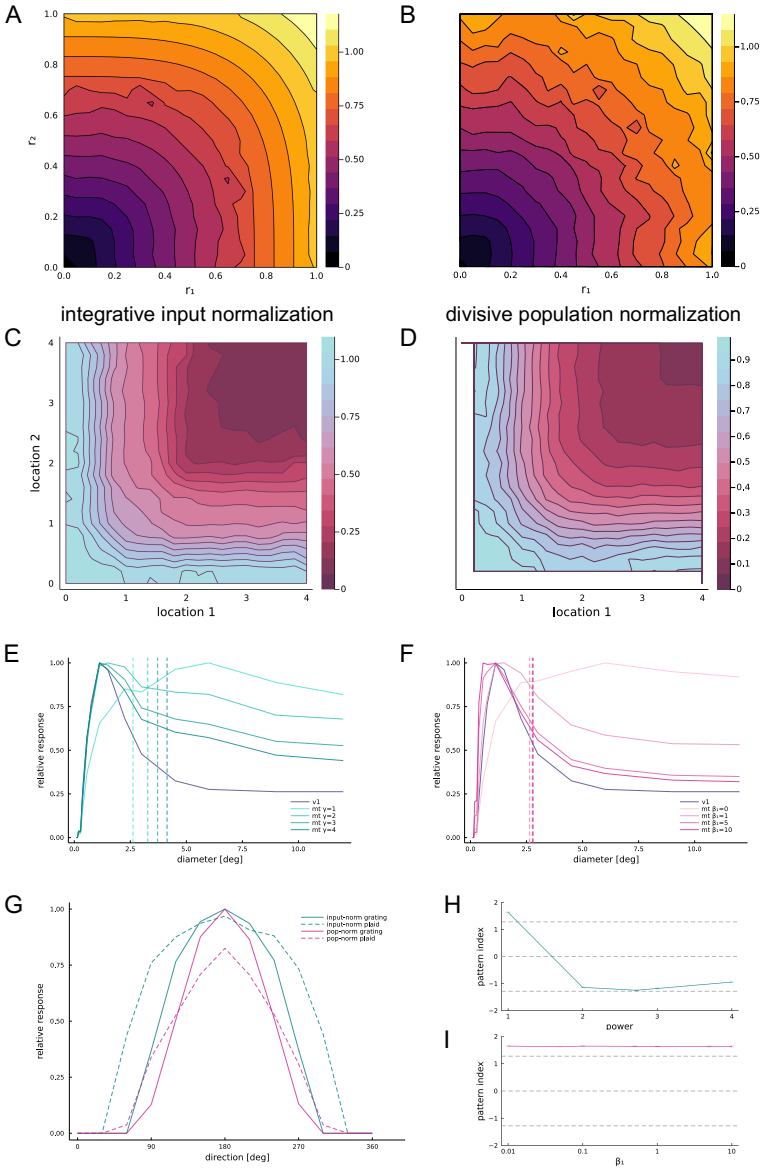

Figure 5-10 Comparison of models with nonlinear integrative input normalization and divisive population normalization

(A) The population-averaged response summation function of 100 models with nonlinear V1-to-MTintegration (γ=*2.72*) and no divisive MT population normalization. **(B)** The population-averaged response summation function of 100 models with linear V1-to-MT integration and divisive MT population normalization (β_1=*10* and β_2=*0.005*). **(C)** The population-averaged areal summation function of the same group of models as in **(A)**. **(D)** The population-averaged areal summation function of the same group of models as in **(B)**. **(E)** The spatial characteristics of the the models with integration of different levels of nonlinearity. The teal curves of various shades are the size tuning curves for different levels of nonlinearity. The purple curve is the size tuning curve of the V1 unit of the models. **(F)** The spatial characteristics of the the models with divisive MT population normalization of different strengths. The magenta curves of various shades are the size tuning curves for different normalization strengths. The purple curve is the size tuning curve of the V1 unit of the models. **(G)** The direction tuning curves of the models with nonlinear V1-to-MTintegration (γ=*2.72*, teal curves) and divisive MT population normalization (β_1=*10* and β_2=*0.005*, magenta curves) to gratings (dashed curves) and plaids (solid curves). **(H)** The change of the pattern index of the models with nonlinear V1-to-MTintegration as the level of nonlinearity varies. Each data point is the sample mean of n=100. Error bars show SEM.The dashed lines at ±1.28 represent the threshold of significance for patter/component direction selectivity. **(I)** Same as **(H)**, but for the models with divisive MT population normalization as the strength of the normalization varies.

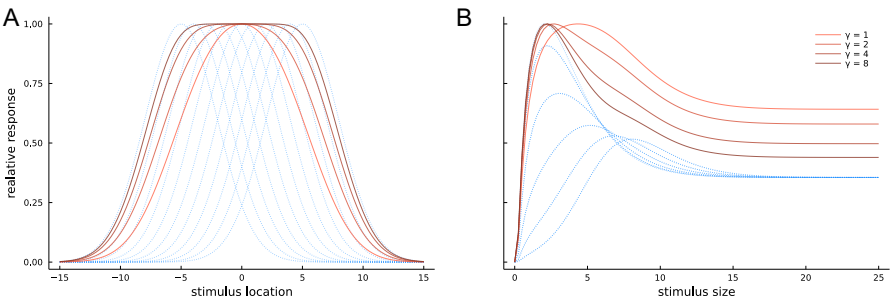

Figure 5-11 Simulation of spatial sensitivity in a conceptualized 1D MT receptive field model with nonlinear spatial integration of V1 inputs

A conceptualized 1D receptive field model with nonlinear integration accounts for the spatial sensitivity of the full-blown MT models with nonlinear V1-to-MT integration. **(A)** The spatial sensitivity profiles of the 1D models with different levels of nonlinearity are represented as solid orange curves of various shades.The dashed blue curves represent the spatial sensitivity profiles and the spatial arrangement of the array of integrated V1 receptive fields of such 1D models.**(B)** The solid orange curves of various shades are the size tuning curves, measured using centered rectangle pulses of different widths, of the same groups of 1D models as in **(A)**. The dashed blue curves are the responses of the array of V1 units of such 1D models to the rectangular pulses used to measure the size tuning of the model.

Chapter 6.

Modeling of the spatial limit of pattern motion processing in the MT receptive field

MT is believed to be the first area in the primate visual motion pathway where a substantial fraction of the cells show similar direction tuning to 1D (e.g., gratings) and 2D (e.g., plaids) stimuli (**Fig. 6-1C**), unlike its afferent V1 cells which, other than signaling the coherent motion of a 2D pattern, respond to the embedded 1D components (**Fig. 6-1B**). Such invariant direction selectivity to 1D and 2D stimuli in MT is termed *pattern direction selectivity*, while other MT cells behave similarly to V1 cells, and their behavior is referred to as *component direction selectivity* (Movshon et al., 1985). Modeling studies have demonstrated that the convergent inputs from DS V1 cells with a broad range of PDs is crucial to pattern motion sensitivity in MT (Rust, et al., 2006). Such signal integration also happens across space where the combination of small V1 RFs generates MT RFs up to ten times larger in diameter (Albright & Desimone, 1987; Wang & Movshon, 2016). The mechanisms underlying such integration may potentially impose certain spatial constraints on the operation of motion integration in MT: Majaj and colleagues (2007) demonstrated in macaques that pattern direction selectivity in MT breaks down when the two corresponding gratings of a plaid are spatially separated and form a pseudo-plaid (**Fig. 6-2D:I**); Kumbhani and colleagues (2014) showed that the spatial limit of pattern motion processing is less than a quarter of the diameter of the MT RF. However, the exact computational mechanism underlying such spatial constraints remains unclear.

Some published studies have speculated that the mechanism of V1-MT processing responsible for localized pattern motion integration is either local processing subunits for direction integration or V1 surround suppression (Majaj et al., 2007; Kumbhani et al., 2015). However, because the modeling efforts in classical studies have mostly overlooked the spatial aspect of signal integration in MT (Rust et al., 2006; Baker & Bair, 2016), neither of the above hypotheses can be tested in an existing computational framework. Using an MT model that includes spatial integration, I examined the effect of varying the spatial scale of various RF substructures on the spatial limit of motion integration, for instance, the size of the suppressive surround field in V1, the distance between adjacent V1 RFs used to form a pooled signal, or the size of the V1 CRF. I found that the breakdown of pattern direction selectivity to pseudo-plaids in MT is linked to the loss of inhibitory signals involved in opponent motion subtraction, the operating range of which correlates with the spatial limit of pattern motion processing observed in my MT models.

6.1. Methods

6.1.1. The stimulus

6.1.1.1. The Double-patch pseudo-plaid

To test if pattern direction selectivity of the MT models breaks down when the component gratings of a plaid were displayed at different locations within the RF, I presented the stimuli, double-patch pseudo-plaids (**Fig. 6-2D:I**), similar to those of Majaj et al. (2007), to the models. I compared the responses of the models to conventional single gratings (**Fig. 6-2A:I**) and true plaids (**Fig. 6-2B:I**) to the responses to double gratings (**Fig. 2C:I**) and pseudo-plaids (**Fig. 2D:I**).

In the single grating stimulus, a sinusoidal grating translates behind a static circular window with a diameter of 4.0° of visual angle. For the base stimulus, the SF of the grating is 1.2 cycle/deg, and the TF is 10.0 Hz. Both frequencies match those of the base models. In the single patch true plaid stimulus, two overlapping gratings are placed behind a single window. The motion directions of the two gratings have a 120° discrepancy. For the double grating, two patches of translating gratings of the same direction are displayed individually behind two windows, which are organized vertically with no overlap, covering a space of 8.0° of visual angle in height. For comparison, the diameter of the MT model RF was 9.0° on average (measured as the width at quarter-height of the mapped Gaussian sensitivity profile of the model RF, see **5.1.4. Gaussian mapping of the MT receptive field**). For the pseudo-plaid, the locations of the two patches are the same as those in the double grating, but the motion directions of the two patches have a 120° discrepancy.

6.1.1.2. The multi-patch pseudo-plaid on a grid

To estimate the spatial limit of pattern motion processing in the MT models, I adopted the multi-patch stimulus on a grid introduced by Kumbhani and colleagues (2015). I presented two kinds of 2D patterns to the models, gratings on a grid (**Fig. 6-3B-E**) and pseudo-plaids on a grid (**Fig. 6-3G-J**), where drifting gratings are displayed behind circular windows of an $n \times n$ grid, and n ranges from 2 to 8 with decreasing levels of spatial dissolution. As the number of patches in the stimulus increases, the size of the individual patch decreases, but the total area covered by the entire arrays of patches remains constant. The entire grid covers an area of 8° × 8° of visual field. For the gratings on a grid, all patches moved in the same direction, while for the pseudo-plaids on a grid, patches drifting in directions of a 120° discrepancy alternate spatially. The SF and TF of the gratings in the base stimuli are 1.2 cycle/deg and 10.0 Hz, respectively.

6.1.2. An MT model

The computational modules and the network architecture of my MT model have been described in detail in **Chapter 3. Simulation methods: an MT model framework**, and concisely reiterated in the previous two

chapters. Here, I will only discuss the model components and parameters that are relevant to the interpretation of the results in this chapter.

The stacked-even and unstacked-even spatial layouts of V1 channels mentioned in **Chapter 4. Modeling of spatial inhomogeneity in the MT receptive field** are utilized here to configure the MT RF with and without local integrative subunits. As discussed in **Chapter 5. Modeling of spatial normalization in the MT receptive field** that the nonlinearity of integrative input normalization could affect the pattern direction selectivity of the MT model (see **5.2.6. Spatial and motion sensitivity of normalization models**), here, linear integration was adopted.

6.1.2.1. The V1 classical receptive field

The first stage of processing in the model is a layer of V1 complex units whose CRFs are modeled as Gabor motion energy filters parameterized by the PD of the unit, being one of the twelve directions that span the 360° range, as well as the central location of the filter window. The motion energy filter sums the squared outputs of two Gabor filters with a 90° phase discrepancy. The spatial SD of the Gabor function in the base models is 0.36°, and the central SF is 1.2 cycle/deg. These V1 RFs are arranged in the even layout (see **3.2.1. Spatial configurations of V1 channels**) within a circular boundary with a diameter of 7.5°.

Two key features of the spatial organization of the V1 RFs were systematically controlled in the simulation, the number of V1 RFs per direction, and the spatial alignment of the V1 inputs across direction channels, which is to be discussed in **6.1.2.3. The spatial structure and computational framework of V1-to-MT integration**. By adjusting the number of V1 units, I was able to change the spatial distance between, i.e., the spatial density of, them.

6.1.2.2. V1 surround suppression and motion opponency

Tuned normalization in V1 is a critical stage in the construction of the PDS MT model of Rust et al., 2006. In my MT model, a V1 surround suppression mechanism operates as the source for tuned normalized. The square root of the output of the motion energy filters is convolved with a suppressive Gaussian field, and the resulting signal is used to divisively normalize the V1 channel in the same and the opposite directions – such a surround is orientation selective (Jones et al., 2001; Cavanaugh et al., 2002). The mathematical formulation is described in detail in **3.1.2. V1 normalization**. As Kumbhani and colleagues (2015) have suggested that motion integration in MT operates within the spatial constraint of V1 surround suppression, to test their conjecture, I controlled the spatial scale of the Gaussian surround field by adjusting its spatial SD in the simulation.

The normalized signal then undergoes the motion opponency stage, where the signal from the opposite direction channel is subtracted (for detail, see **3.1.3. Motion opponency**). This operation is performed locally between V1 units that share the same CRF center location. Thus, it contributes no additional spatial integration beyond that of the upstream CRF and surround signals.

6.1.2.3. The spatial structure and computational framework of V1-to-MT integration

Majaj and colleagues (2007) have hypothesized the existence of local motion processing subunits in V1-MT circuitry. The associated operations may include the integration of V1 signals in MT, the inhibition of opponent motion and tuned normalization in V1 (Rust et al., 2006). To examine the potential mechanism that creates localized V1-to-MT signal integration in pattern motion processing, I constructed PDS MT models with three variations of the linear V1-to-MT spatial integration structure (see **3.2.2.1. Linear integration**), as described next.

6.1.2.3.1. The no-subunit structure

In the *no-subunit* structure, the spatial layout of the V1 units with different PDs is staggered, assuming the unstacked-even configuration (**Fig. 3-2C**). The signal integration that happens when the MT weights are applied to the V1 inputs (**Fig. 3-1**, weighting of V1 channels) occurs globally across all direction and spatial channels in a single stage (**Eq. 3-11**).

6.1.2.3.2. The false-subunit structure

In the *false-subunit* structure, the locations of the V1 units are aligned across direction channels, forming localized V1 unit stacks that each sample the complete range of directions. This is the stacked-even configuration (**Fig. 3-2A**) as studied in **Chapter 4. Modeling of spatial inhomogeneity in the MT receptive field**. In this case, the summation of V1 inputs still happens globally across spatial and direction channels. These stacked structures are hence referred to as "false" subunits, given the lack of localized nonlinear computation (**Eq. 3-12**) that would alter the output if a signal were moved from one spatial subunit to another but remained within the same direction channel.

6.1.2.3.3. The true-subunit structure

In the *true-subunit* structure, the V1 channels are organized in the stacked-even configuration, and signal integration is carried out in two stages. In the first stage, local summation occurs across the V1 units with different PDs sharing the same RF location (within a stack), followed by local half-wave rectification **(Eq. 3-13)**. The half-wave rectification effectively filters out the negative inputs at each location where the inputs from the excitatory direction channels are not strong enough to counteract those from the inhibitory channels. In the second stage, the resulting signals from the V1 stacks are integrated across the entire MT RF **(Eq. 3-13)**. These stacks are referred to as "true" processing subunits, because direction integration first happens locally within a subunit where both input pooling and nonlinear computation operate independently of the signals from V1 units located elsewhere. Local integration is followed by the subsequent summation across all subunits within the model. In such models, localized motion processing exists both structurally and computationally.

6.1.3. The pattern index

To quantitatively describe the degree of the pattern direction selectivity of the MT models, I calculated the PI, with negative values indicating strong component direction selectivity, and positive values indicating strong pattern direction selectivity (Rust et al., 2006). The direction tuning curve of the model MT unit to the gratings, $DT_{\mathrm{gr}}(\theta)$, is measured and then used to generate the predicted direction tuning response, of an ideal CDS unit to the plaid, $PR_{\mathrm{c}}(\theta)$, as follows:

$$PR_{\mathrm{c}}(\theta) = DT_{\mathrm{gr}}(\theta) + DT_{\mathrm{gr}}(\theta + \Delta\theta)$$

Equation 6-1

where $\Delta\theta$ represents the angle of direction discrepancy between the two component gratings of the plaid stimuli, which is 120° in this case. I then compute the component and pattern correlation coefficients, r_{c} and r_{p}, respectively (Movshon et al., 1985), as follows:

$$r_{\mathrm{c}} = \frac{\rho_{\mathrm{c}} - \rho_{\mathrm{p}}\rho_{\mathrm{pc}}}{\sqrt{\left(1-\rho_{\mathrm{p}}^2\right)\left(1-\rho_{\mathrm{pc}}^2\right)}}$$
$$r_{p} = \frac{\rho_{\mathrm{p}} - \rho_{\mathrm{c}}\rho_{\mathrm{pc}}}{\sqrt{\left(1-\rho_{\mathrm{c}}^2\right)\left(1-\rho_{\mathrm{pc}}^2\right)}}$$

Equation 6-2

where ρ_{c} is the Pearson correlation coefficient between the direction tuning of the unit to the plaid stimulus, $DT_{\mathrm{pl}}(\theta)$, and the predicted CDS direction tuning to the plaid stimuli, $PR_{\mathrm{c}}(\theta)$, and ρ_{p} is the Pearson correlation coefficient between $DT_{\mathrm{pl}}(\theta)$ and $DT_{\mathrm{gr}}(\theta)$, which serves as the predicted response of the unit to the plaid if it were fully PDS, and finally, ρ_{pc} is the Pearson correlation coefficient between $PR_{\mathrm{c}}(\theta)$ and $DT_{\mathrm{gr}}(\theta)$. I then apply the Fisher *r*-to-*Z* transform to r_{c} and r_{p} (Smith et al., 2005):

$$Z = \frac{\sqrt{df}}{2}\ln\frac{1+r}{1-r}$$

Equation 6-3

where df is the degree of freedom, the number of data points in the direction tuning curve minus 3 ($df = 9$ in this case). PI is defined as the difference between Z_{p} and Z_{c}. The significance threshold of PI is set to 1.28, which corresponds to α = 0.1 (Smith et al., 2005): a unit demonstrates significant pattern direction selectivity if PI > 1.28, and significant component direction selectivity if PI < -1.28.

The PI of a unit for the *n*×*n*-patch pseudo-plaids is computed by analyzing the direction tuning for such stimuli and that for the corresponding *n*×*n*-patch grating stimuli.

6.2. Results

6.2.1. The response of MT models to double-patch pseudo-plaids and breakdown of pattern direction selectivity

Majaj and colleagues (2007) reported that when presented with double-patch pseudo-plaids, the MT cells, which show PDS-like direction tuning to conventional single-patch plaids, will respond with CDS-like tuning, demonstrating weakened pattern motion integration when the motion signals of different directions in the stimulus do not spatially colocalize. They believed that such a behavior indicated that pattern motion processing in MT did not operate globally across the RF; instead, there might exist spatial subunits with localized motion integration, and the corresponding computation also operated under such spatial constraints.

To test if such subunits performed local input pooling, I built MT models with three levels of localization of the RF substructure and computational architecture (see **6.1.2.3. The spatial structure and computational framework of V1-to-MT integration**). Level 1: in the no-subunit models, the locations of the V1 RFs are independent across direction channels. Therefore, there is no spatial basis for localized input pooling that can facilitate motion processing subunits with independent functionality. Level 2: in the false-subunit models, the spatial layouts of the V1 units are aligned across direction channels. Such a spatial structure can potentially support localized input pooling, which is however not enacted due to the absence of localized computational mechanisms. Level 3: in the true-subunit models, besides the organization of spatially aligned V1 units across direction channels, I implement localized input integration by applying half-wave rectification to the summation of signals within the stack of colocalized V1 units spanning the complete range of directions. These subunits analyze motion locally, and the outputs are then summed globally across the RF.

I presented the conventional single-patch plaid and the double-patch pseudo-plaid stimuli to the three types of MT models. If motion processing in MT had relied exclusively on subunits with localized integration, pattern motion computation would show sub-RF spatial specificity only in models with the true-subunit architecture. However, I found that the spatial specificity of motion integration was a common characteristic shared by the models regardless of their subunit architecture. Therefore, local integration subunits were unlikely to be the predominant source for localized pattern motion processing in MT. For all three types of models, compared to that for the single-patch gratings, the direction tuning curve for the true plaids widened, but only to a limited extent, and the shape maintained single-peaked, demonstrating evident pattern motion sensitivity to the conventional stimuli (**Fig. 6-2A&B**). The models' responses to the double-patch gratings were identical, and remained consistent with those to the single-patch gratings (**Fig. 6-2C**). On the other hand, for all three types of models, the direction tuning for the pseudo-plaids deviated substantially from those for the other three stimuli: it widened extensively (**Fig. 6-2D**), demonstrating that the breakdown of pattern direction selectivity to pseudo-plaids was not exclusive to models with local integrative subunits; however, such morphing of the direction tuning curve for pseudo-plaids was most significant for models with true subunits.

I quantified the spatial specificity of pattern motion sensitivity exposed by pseudo-plaids by calculating the PIs of the models, which dropped, for all three types of models, from identical values in the PDS range to CDS

levels when the stimuli changed from conventional plaids to pseudo-plaids. While the shift of PI was most drastic for the models with true subunits, it was less extreme for models with false or no subunits, and the degree of the shift was identical between the two groups of models (**Fig. 6-4**). Such an effect indicated that, even though the existence of local integrative subunits could not fully account for the breakdown of pattern direction selectivity for pseudo-plaids in MT, it could contribute to the localization of motion processing due to the localized computational architecture. The spatial structure of V1 unit organization is mirrored between the models with true and false subunits. Hence, the extra reduction of PI for pseudo-plaids in the true-subunit models should be attributed to the additional local nonlinear computation – half-wave rectification in such models.

6.2.2. The multi-patch pseudo-plaid and the spatial limit of pattern motion processing

Using multi-patch pseudo-plaids on a grid with different levels of spatial dissolution, i.e., the number of patches in the stimulus, Kumbhani and colleagues (2015) concluded that the spatial limit of pattern motion processing is on par with the scale of V1 surround suppression. They speculated that V1 surround suppression, potentially serving as the source of tuned normalization, was the basis of localized motion processing in MT. Hence, I presented the multi-patch pseudo-plaids to a series of MT models to test their hypothesis and also investigated other potential mechanisms to achieve spatially specific pattern direction selectivity.

6.2.2.1. The spatial scale of V1 surround suppression

If V1 surround suppression is thought to spatially constrain motion integration in MT, then, for the MT models, changes in the spatial extent of the V1 surround field should result in changes in the spatial limit of pattern motion processing. However, when I presented the multi-patch pseudo-plaids to three sets of no-subunit models with V1 surround fields of different sizes, 0.63° (as the SD of the Gaussian field, $g(x, y)$ in **Eq. 3-6**) for small surrounds, 1.26° for intermediate surrounds and 2.52° for large surrounds, the results suggested that the spatial limit of pattern motion processing did not vary noticeably, and certainly not in the manner proposed in the literature.

Fig. 6-5 displays the evolution of PI for models with V1 surrounds of three sizes as the number of patches in the pseudo-plaid increased, and the PI values corresponding to single-patch stimuli represent the baseline level of pattern direction selectivity for true plaids. The baseline PI decreased very little as the size of the V1 surround grew. This was most likely a numerical artifact of computation due to the spatial truncation of the surround field in the model. As the spatial SD of the surround field increased, if the visual field of the model had been infinite, the volume under the Gaussian envelope of the surround field ($g(x, y)$ in **Eq. 3-6**) should remain constant. However, because the visual field was represented with finite truncation in the model, the volume under the Gaussian envelope was effectively reduced when the SD increased, causing the surround signal to be weakened, albeit to a small extent, which was why the decrease of the baseline PI was very limited.

For all three types of models, the PI value dropped into the CDS range in response to the pseudo-plaids on a 2×2 grid, and recovered when the number of patches in the pseudo-plaid increased. If the spatial limit of motion processing had been directly linked to the spatial scale of V1 surround suppression, the curve of the PI evolution in **Fig. 6-5** should shift to the left as the size of the V1 surround grew. However, the curve shifted very slightly to the right for the models with the largest V1 surround, which could be explained by the above-mentioned variation of baseline PI due to spatial truncation. The curves for the three types of models being largely identical indicated that the spatial limit of motion processing was not affected by the size of the suppressive V1 surround.

In all cases, the PI returned to the above-threshold PDS level (PI > 1.28, gray horizontal line) when the grid of the pseudo-plaid was finer than 6×6. Given that the spacings between the nodes of the 6×6 and 8×8 grids were respectively 1.33° and 1.0°, with the average size of the MT RF of the models being 9.0°, the estimated spatial limit of motion processing in my MT models was 10 ~ 15% of the MT RF size. While Kumbhani and colleagues (2015) did not report any data for stimulus grids finer than 4×4 patches, with each patch spanning a quarter of the RF diameter, the fact that at that level PI has not recovered to the PDS range suggested that the spatial limit of pattern motion processing for their MT neurons is less than 25% of the RF size.

6.2.2.2. A descriptive model of the spatial limit of pattern motion processing

Kumbhani and colleagues (2015) proposed a descriptive model to explain the decreasing PI of MT cells for multi-patch pseudo-plaids as a function of the number of patches. The model postulated that the PI was proportional to the degree of overlap between the Gaussian-blurred patches of the two different component gratings of different motions. To further compare my results to theirs, I replicated their analysis and applied it to the responses of my MT models. I will first describe the formulation of their descriptive model, which I hereby refer to as the *blur-overlap model*.

In the multi-patch pseudo-plaid stimulus, the gratings drifting in different directions are restricted within the boundary of interleaved circular windows with no spatial overlap. By filtering the multi-patch spatial masks (a set of flat disks representing the circular windows in **Fig.6-6A&C**) with a Gaussian kernel (**Fig. 6-6E**), the boundaries between the patches of different directions are blurred. This process is described in the following equation:

$$N_i(x,y) = M_i(x,y) * \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \qquad \textbf{Equation 6-4}$$

where $M_i(x,y)$ is the mask for the patches of the grating moving in one of the two directions (the solid red or blue disks in **Fig. 6-6A&C**), and $N_i(x,y)$ is the corresponding Gaussian-blurred mask (the fuzzy red or blue circular patches in **Fig. 6-6B&D**), and i parameterizes the motion direction of the patch. The free parameter, σ, controls the spatial scale of the Gaussian blurring filter. A wider Gaussian kernel boosts the overlap between

the blurred masks of the two gratings. The blur-overlap model was based on the idea that signal processing in the visual cortex prior to MT involved Gaussian-like filters (RFs), and those centered between the gratings in the pseudo-plaid (the purple regions in **Fig. 6-6B&D**) allowed the pattern motion signal to arise from the colocalization of motion energy of different directions within such RFs. Therefore, the increase of MT cell PI for the pseudo-plaids of finer grids should be accounted for by the increasing amount of overlap between the blurred grating patches. Kumbhani and colleagues (2015) assumed a linear relationship between the PI for pseudo-plaids, and the extent of overlap, which is calculated as the congruence coefficient of the blurred masks of the two gratings, as follows:

$$PI \propto \frac{\sum_{x,y} N_1(x,y)N_2(x,y)}{\sqrt{\sum_{x,y} N_1^2(x,y) \sum_{x,y} N_2^2(x,y)}}$$

Equation 6-5

By fitting the responses of the MT cell to the blur-overlap model as a function of the patch number, they could estimate the σ value of the theoretical Gaussian kernel, which they believed to reflect the spatial limit of pattern motion processing.

Although the blur-overlap model provided a speculative account for the data of Kumbhani et al. (2015), it lacked thorough substantiation. When I fitted the simulated response of the MT models with small, intermediate, and large surrounds, I found that the blur-overlap model could not describe the responses of many model units with a unique optimal σ value. For all three groups of models, the fitted σ values congregated into two regions, $0.00° \sim 0.10°$ and $0.30° \sim 0.60°$ (**Fig. 6-7A**). When I compared the σ values of the MT model units with the larger surrounds to the corresponding units (which are connected to the sets of V1 inputs at the same locations with the same direction selectivity) with the smaller surrounds, I found that there was no significant difference between the two groups (**Fig. 6-7B**). This was consistent with the notion that the spatial limit of motion processing in these models was not affected by the size of the surround; however, the observation that the σ values split into two ranges suggested that this analysis was not able to properly represent the spatial scale of motion integration.

I then further inspected, for individual MT units, how the goodness of fit for the blur-overlap model depended on the σ value. For a typical unit shown in **Fig. 6-7C**, the blur-overlap model with a σ value of either $0.40°$ or $0.05°$ could fit the unit's PI evolution equally well. **Fig. 6-7D** depicts the variance of the data explained by the blur-overlap model as a function of the σ value for the same example unit (as in **Fig. 6-7C**): there was no prominent global maximum in the curve, and the explained variance maintained a high level in the σ range of $0.05° \sim 0.40°$. Therefore, for the model MT units, the blur-overlap model and the corresponding metric of σ could not clearly reflect a unique spatial scale of pattern motion processing. This might be due to the assumption of the blur-overlap model about the linear relationship between the PI, and the level of overlap between the patches of gratings with different motion directions. The free scaling and translating factors of the linear regression likely introduced an excessive amount of flexibility for positioning the fitted curve in the

parameter space (**Fig. 6-7C**), and allowed a large range of σ values to produce a good fit, making it difficult for the fitting process to generate a meaningful estimate of the spatial constant of the Gaussian kernel.

In addition, Kumbhani and colleagues (2015) did not provide any validation for the blur-overlap model, and their interpretation relied on the unjustified assumption that the Gaussian blurring filter is functionally equivalent to V1 surround. However, while the blurring filter is a conceptualized model of an integrative sensor, surround suppression is a divisive (or differentiating, depending on the specific implementation in models) process that facilitates end-stopping. Therefore, the blur-overlap model is unlikely a reliable approach for measuring the spatial limit of motion integration in MT.

6.2.2.3. The spatial distance between V1 receptive fields

The rationale behind the blur-overlap model was that the integration of pattern motion was the result of the overlap between the diffused spatial representations of motion signals in the stimulus. As such a model could not account for the simulated responses, I wondered if the extent of another form of spatial overlap, rather than that in the stimuli but that between the V1 RFs, equivalently, the distance between the V1 RFs, could affect the spatial scale of pattern motion processing in my MT models. Hence, I built a series of MT models with different numbers of V1 units in each direction channel: 9, 18, 36 and 72. Because the maximum possible distance of a V1 RF to the center of the MT RF is fixed so that all the MT model RFs have a similar size, the MT models with more V1 units have less distance, i.e., more overlap, between the V1 RFs.

Fig. 6-8 shows the evolution of the PI of these MT models for pseudo-plaids as a function of the number of patches within the stimuli. The curves were identical across the MT models with different numbers of V1 units. The breakdown of pattern direction selectivity to pseudo-plaids recovered for all models when there were more than 6×6 patches in the stimulus. Therefore, the spatial scale of motion processing in the MT models was not affected by the spatial distance between the V1 RFs pooled by the MT unit.

6.2.2.4. The spatial scale of the V1 classical receptive field

Just as important as tuned normalization to the PDS model of Rust et al. (2006), is V1 motion opponency where the signal of the opposite direction channel is subtracted locally from each direction channel (**Eq. 3-7**). The main driving forces of the preferred and opponent signals in the MT models arise from within the spatial extent of the V1 CRF, with generally a weaker (normalizing) contribution from the suppressive surround. Thus, in this case, I wanted to test if the spatial limit of motion processing might be constrained by the size of the V1 CRF, i.e., the operating range of V1 motion opponency.

I varied the spatial scale of the V1 CRF in the MT models by adopting three levels of spatial SD for the V1 Gabor filters, and presented the same set of multi-patch pseudo-plaids to the models. The small V1 CRFs has a SD of 0.24°, and the intermediate and large V1 CRFs have SDs of 0.36° and 0.48°, respectively. I found that the baseline PI of the models plummeted as the size of the V1 CRFs increased (**Fig. 6-9A**). However, the curve of the PI evolution for the pseudo-plaids shifted in the opposite direction, upwards and to the left, for the models

with the larger V1 CRFs, potentially indicating a dependence of the spatial limit of pattern motion processing on the size of V1 CRFs. The shift of the curves could be explained by the stronger motion opponency signal in the models with larger V1 CRFs. The spatial disassociation of the component directions in the pseudo-plaids weakened V1 motion opponency when the V1 units could not simultaneously sense motion signals of different directions. This effect would be more pronounced in MT models with smaller V1 CRFs, hence the down and right shift of the PI curve in thoas emodels, demonstrating a finer spatial limit of motion processing for such models.

However, the unexplained tremendous variation of the baseline PI among the three groups of models could potentially confound the interpretation of the results. Thus below, I try to account for such variation in the context of V1 motion opponency, and explain how I managed to equalize the baseline PI across the models.

The increased baseline PI of the MT models with smaller V1 CRFs resulted from the flattened frequency response due to the narrow spatial window of the V1 Gabor filter. This was reflected in the widened V1 direction tuning curve of the smaller-CRF model in response to the conventional gratings (dashed blue curve in **Fig. 6-9B**), which generated the stronger inhibitory signal in the opponent motion channel (see the stronger signal of the direction tuning curve for conventional plaids in the opponent channel – negative solid blue curve in **Fig. 6-9B**, corresponding to $n_{\left(i+\frac{M}{2}\right) \bmod M}(x, y, t)$ in **Eq. 3-7** , specifically at the peak direction of the excitatory channel – positive solid blue curve in **Fig. 6-9B**, corresponding to $n_i(x, y, t)$ in **Eq. 3-7**). Such enhanced V1 opponent motion inhibition facilitated pattern motion sensitivity. Due to the commutative nature of convolution, constraining the stimulus to smaller patches achieved a similar effect, thus making the PI of some of the previously discussed MT models in response to the pseudo-plaids with 8×8 patches stronger than the baseline PI (**Fig. 6-5** and **Fig. 6-8**).

The variation of the baseline PI across the MT models with different V1 CRF sizes addressed here could be canceled out if the central SF of the V1 Gabor filters was adjusted according to the spatial SD of the Gaussian envelope. To demonstrate this effect, I increased the central SF to 1.8 cycle/deg for the models with the smaller V1 CRFs, and reduced it to 0.9 cycle/deg for those with the larger V1 CRFs (the SF of the corresponding stimuli for either model was matched accordingly as well); in this case, the SF×SD product remained a constant across the models with different V1 CRF sizes. **Fig. 6-10B** shows that the V1 responses were similar across the models, with no apparent variation of the inhibition strength in the opponent motion channel (negative solid curves) at the peak direction of the tuning curve of the excitatory channel (positive solid curvess) for conventional plaids. In this case, the baseline PI was equalized to a similar level across all models (**Fig. 6-10A**). Such equalization was not perfect, likely due to the discretization (fewer pixels in the simulated stimulus within smaller RFs), as it is shown in **Fig. 6-10B**, that the V1 direction tuning curve for the gratings is slightly wider for the model with larger V1 CRFs (dashed green curve), which explained why the baseline PI was also slightly higher for these models (**Fig. 6-10A**).

Once the baseline PI had been equalized, an apparent correlation between the spatial scale of the V1 CRF, and the spatial limit of motion processing in these MT models emerged. **Fig. 6-10A** shows a left shift and

upward scaling of the pseudo-plaid PI evolution curve for the models with larger V1 CRFs (solid green curve), and the PI returned to the baseline level when there were 4×4 to 6×6 patches in the pseudo-plaid, one order smaller than that for the models with intermediate V1 CRFs (solid red curve). Therefore, the spatial limit of motion processing in my MT models became coarser when the size of the V1 CRF increased as it spatially constrained the operating range of V1 motion opponency.

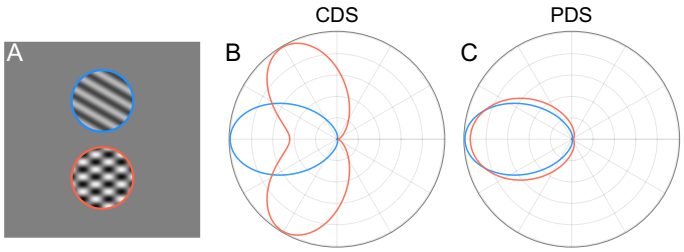

Figure 6-1 Component and pattern direction selectivity

(A) Typical stimuli used to determine the type of direction selectivity of a visual motion neuron: a grating that drifts in the direction orthogonal to the orientation of the contour (blue outline), and a plaid consisting of two overlapping gratings with motion of different directions, and the discrepancy is a constant angle, 120° (red outline). **(B)** Illustration of the direction tuning curves of component direction selective neurons measured as responses to gratings (blue) and plaids (red). **(C)** Same as **(B)**, but for pattern direction selective neurons.

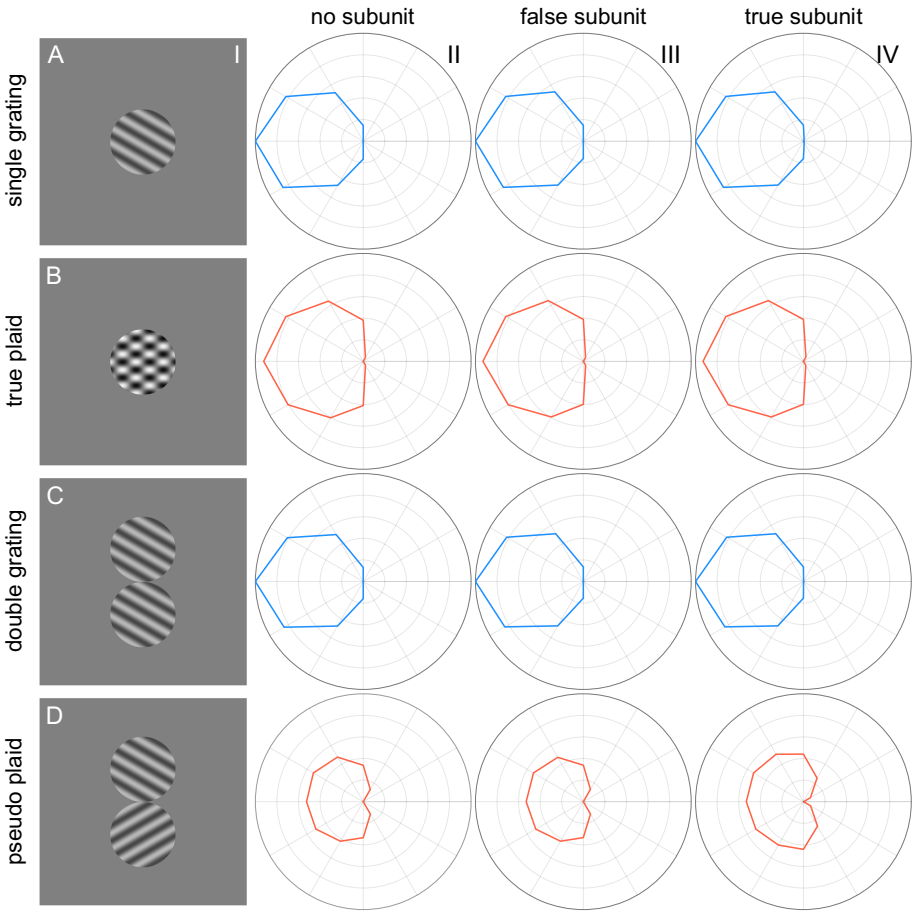

Figure 6-2 Direction tuning of models with different spatial structures

The direction tuning curves of example units are shown for models with three types of spatial structures:**Column II** – no-subunit; **Column III** – false-subunit; **Column IV** – true-subunit. The stimuli used to test the models are shown in **Column I**: **(A)** single-patch gratings; **(B)** single-patch true plaids; **(C)** double-patch gratings where two patches of gratings drift in the same direction; **(D)** double-patch pseudo-plaids where the gratings that should have constitute a conventional single-patch plaid are spatially segregated.

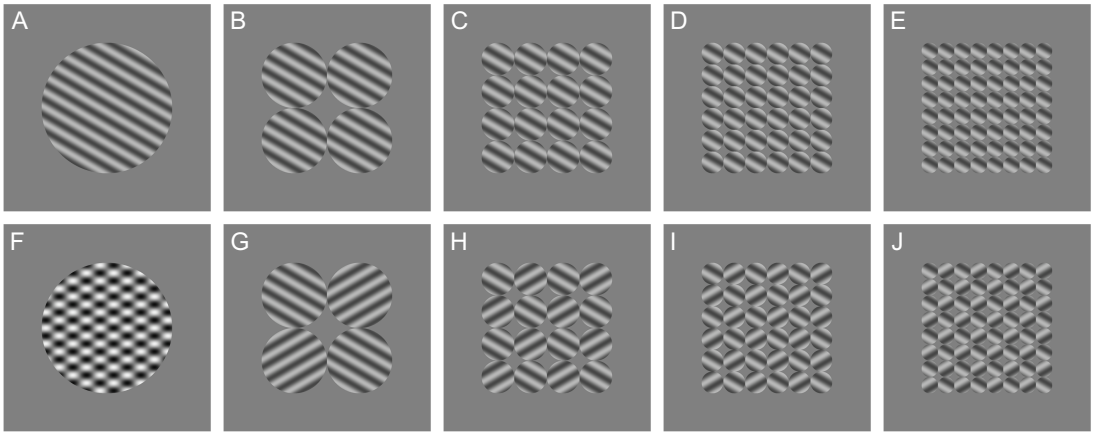

Figure 6-3 Multi-patch stimuli on a grid

(A) Single-patch gratings. **(B)** Gratings on a grid of 2×2 patches; **(C)** 4×4 patches; **(D)** 6x6 patches; **(E)** 8x8 patches. Gratings of all patches drift in the same direction. **(F)** Single-patch true plaids. **(G)** Pseudo-plaids on a grid of 2×2 patches; **(H)** 4×4 patches; **(I)** 6x6 patches; **(J)** 8x8 patches. Patches of gratings drifting in different directions are interleaved. The discrepancy of directions is 120°. As the number of patches increase, the size of each patch decreases, but the summed area of all patches remains constant.

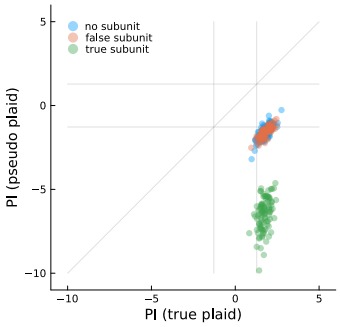

Figure 6-4 Pattern direction selectivity of models with different spatial structures

Pattern index calculated from the responses to the double-patch pseudo-plaids is plotted against that to the single-patch true plaids for models with different spatial structures. n=100, for each group. The grey lines indicate the significance thresholds for direction selectivity.

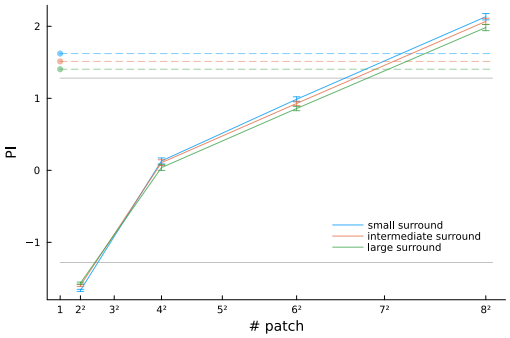

Figure 6-5 The spatial limit of motion integration in models with different sizes of V1 surround

The solid curves plot the pattern index calculated from the responses to multi-patch pseudo-plaids against the number of patches for models with V1 surrounds of various sizes. The dots corresponding to the abscissa value of 1 and the linked dashed horizontal lines represent the pattern index measured with conventional true plaids, the baseline level. The intersection of the solid curve and the dashed line indicates the recovery of pattern direction selectivity as the number of patches in the pseudo-plaid increases, reflecting the spatial limit of motion integration. Each data point on the curve is the sample mean of n=100. Error bars show SEM. The grey lines indicate the significance thresholds for direction selectivity.

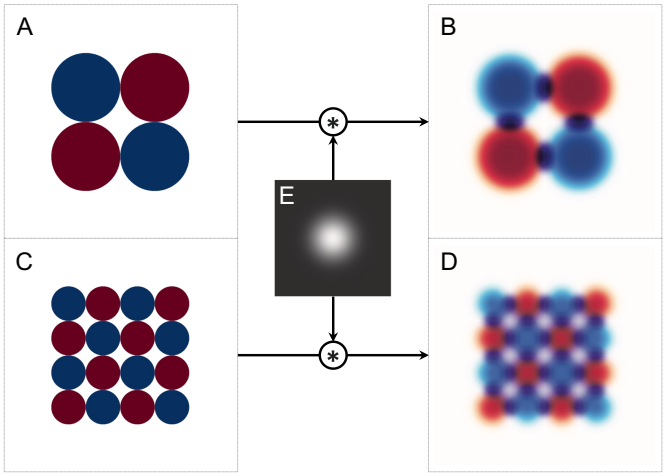

Figure 6-6 The blur-overlap model

The blur-overlap model was used in Kumbhani et al. (2015) to account for the evolution of pattern index of the MT neuron as a function of the number of patches in the pseudo-plaid, and hence the spatial limit of motion integration. **(A)** The masks representing the component gratings of different motions in the 2×2 pseudo-plaids are shown as solid blue and red disks. **(B)** The masks in **(A)** are blurred by the convolution with a Gaussian kernel, **(E)**, and are shown as fuzzy patches. This filtering process results in the overlap between the blurred masks of different colors. The overlapping areas are shown in purple. **(C)** and **(D)** Same as **(A)** and **(B)**, but for the 4×4 pseudo-plaids.

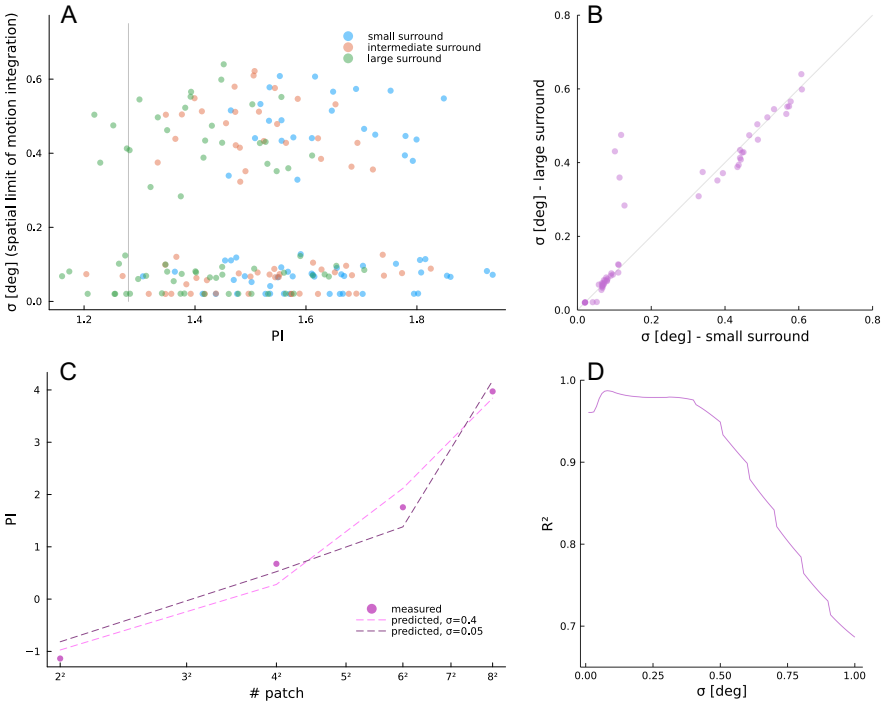

Figure 6-7 Fit of the pattern index evolution for MT models with different V1 surround sizes to the blur-overlap model

(A) The spatial constant, σ, of the fitted blur-overlap model, assumed to be linked to the spatial limit of motion integration, is plotted against the baseline pattern index (calculated from the responses to the conventional true plaids) for models with different V1 surround sizes. n=100, for each group. The grey line indicates the significance threshold for direction selectivity. **(B)** The spatial constant for the models with the large V1 surround is plotted against that for the models with the small V1 surround. Except the size of V1 surround, the two groups of models are otherwise identical. **(C)** Alternative fitting results of the blur-overlap model to an example unit. The dots represent the measured evolution of pattern index as a function of the number pf patches in the pseudo-plaid. The dashed lines in different shades correspond to the predicted evolution of pattern index produced by the blur-overlap models with different spatial constants. **(D)** For the same example model unit as in **(C)**, the coefficient of determination for the predicted evolution of pattern index produced by the blur-overlap model of a varying spatial constant is plotted against the corresponding value of the spatial constant.

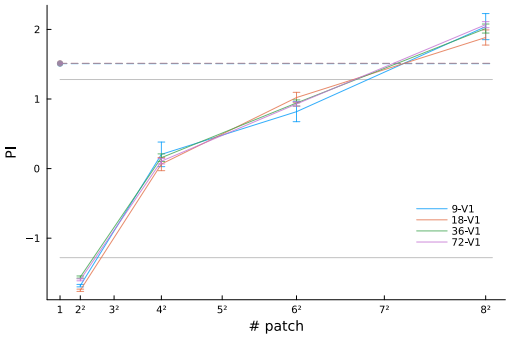

Figure 6-8 The spatial limit of motion integration in models with different V1 spatial sampling density

The solid curves plot the pattern index calculated from the responses to multi-patch pseudo-plaids against the number of patches for models with various numbers of V1 inputs per direction channel. The dots corresponding to the abscissa value of 1 and the linked dashed horizontal lines represent the pattern index measured with conventional true plaids, the baseline level. The intersection of the solid curve and the dashed line indicates the recovery of pattern direction selectivity as the number of patches in the pseudo-plaid increases, reflecting the spatial limit of motion integration. Each data point on the curve is the sample mean of n=100. Error bars show SEM. The grey lines indicate the significance thresholds for direction selectivity.

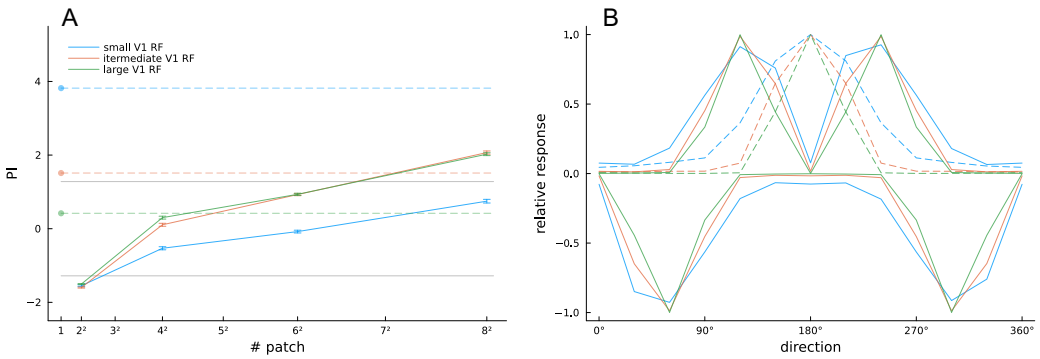

Figure 6-9 The spatial limit of motion integration in models with different sizes of V1 classical receptive field but identical spatial frequencies

(A) The solid curves plot the pattern index calculated from the responses to multi-patch pseudo-plaids against the number of patches for models with V1 classical receptive fields of various sizes but the same spatial frequency. The dots corresponding to the abscissa value of 1 and the linked dashed horizontal lines represent the pattern index measured with conventional true plaids, the baseline level. The intersection of the solid curve and the dashed line indicates the recovery of pattern direction selectivity as the number of patches in the pseudo-plaid increases, reflecting the spatial limit of motion integration. Each data point on the curve is the sample mean of n=100. Error bars show SEM. The grey lines indicate the significance thresholds for direction selectivity. **(B)** Direction tuning of the V1 units in the models. The dashed curves show the direction tuning of the V1 units selective for the direction of 180° measured using the responses to gratings. The solid curves above the abscissa show the direction tuning of the same V1 units measured using the responses to plaids, representing the excitatory channel in V1 motion opponency. The solid curves below the abscissa show the direction tuning to plaids, of the same V1 units selective for the direction of 0°, representing the suppressive channel in V1 motion opponency.

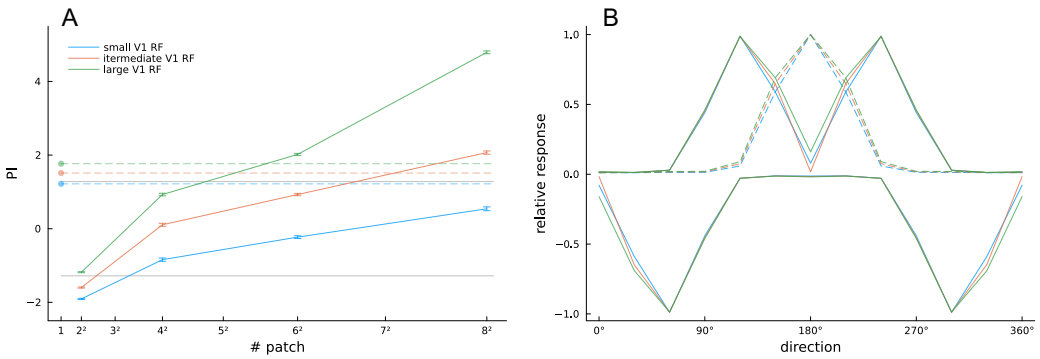

Figure 6-10 The spatial limit of motion integration in models with different sizes of V1 classical receptive field and proportionally matched spatial frequencies

(A) The solid curves plot the pattern index calculated from the responses to multi-patch pseudo-plaids against the number of patches for models with V1 classical receptive fields of various sizes, and the spatial frequencies of these models are proportionally matched to the size of the receptive field, so that the product of the spatial frequency and the standard deviation of the Gabor filter remains constant. The dots corresponding to the abscissa value of 1 and the linked dashed horizontal lines represent the pattern index measured with conventional true plaids, the baseline level. The intersection of the solid curve and the dashed line indicates the recovery of pattern direction selectivity as the number of patches in the pseudo-plaid increases, reflecting the spatial limit of motion integration. Each data point on the curve is the sample mean of n=100. Error bars show SEM. The grey lines indicate the significance thresholds for direction selectivity **(B)** Direction tuning of the V1 units in the models. The dashed curves show the direction tuning of the V1 units selective for the direction of 180° measured using the responses to gratings. The solid curves above the abscissa show the direction tuning of the same V1 units measured using the responses to plaids, representing the excitatory channel in V1 motion opponency. The solid curves below the abscissa show the direction tuning to plaids, of the V1 units selective for the direction of 0°, representing the suppressive channel in V1 motion opponency

Chapter 7.

Discussion

In this thesis, I have presented the first image-computable model of neurons in area MT that accounts for the spatial structure and integration of V1 inputs. Using this model, I have conducted an extensive and systematic study exploring how the spatial structure and network architecture of signal processing and integration can account for salient characteristics of the RFs of MT neurons that are reflected in the responses to a series of dynamic, compound motion stimuli, where multiple localized motion signals are simultaneously present. Understanding the fundamental principles and intricacies of spatial integration in the visual cortex is crucial for comprehending how larger downstream RFs are constructed upon the convergent inputs from smaller upstream RFs, and for ultimately understanding the role of RF size in neural computation in sensory systems.

However, the topic of spatial integration in MT has been largely overlooked in decades of MT modeling research. Many investigators have focused on identifying the computational mechanisms responsible for MT pattern motion processing, but have neglected to model the spatial aspects of the MT RF. Published studies often either excluded spatial integration entirely (Heeger, 1987; Grzywacz & Yuille, 1990; Bowns, 2002; Perrone, 2004; Rust et al., 2006; Nishimoto & Gallant, 2011; Baker & Bair, 2016), or failed to critically analyze the spatial aspect of the integration scheme (Nowlan and Sejnowski's, 1995; Simoncelli & Heeger, 1998; Perrone & Krauzlis, 2008; Tsui et al., 2010). These studies discussed in great detail how signals are integrated across direction and frequency channels to account for the well-known characteristics of MT direction selectivity, but did not address the corresponding spatial mechanisms, even though motion and spatial integration in MT are closely intertwined.

Recently, a line of *in vivo* research has attempted to shed some light on the spatial mechanisms underlying motion integration in the MT RF (Majaj et al., 2007; Kumbhani et al., 2015), but the validity of the resulting conjectures cannot be tested in the existing modeling frameworks that lack spatial integration. Therefore, I constructed the first image-computable MT model with spatial integration, and used the stimuli introduced in past research, which survey the fine detail of the MT RF substructure or probe the mechanism of V1-to-MT integration (Majaj et al., 2007; Richert et al., 2013; Kumbhani et al., 2015), to directly investigate the possible underlying spatial and functional circuits.

In this chapter, I will contextualize my simulation results within the broader scope of past research, delve into the potential processing mechanisms that give rise to several key spatial characteristics of the MT RF, acknowledge the limitations of my MT models, and provide suggestions for future modeling and *in vivo* studies.

7.1. The heterogeneous MT receptive field profile and the topography of V1-MT wiring

The extensive RFs of MT neurons are thought to be derived from the spatial integration of DS signals that originate in V1 (Movshon and Newsome, 1996; Churchland et al., 2005; Nassi and Callaway, 2007), and the combination of such V1 RFs results in MT RFs of up to ten-fold enlargement of diameter (Albright & Desimone, 1987; Wang & Movshon, 2016). The classical view of the MT RF is an idealized, relatively homogeneous structure, with sensitivity that diminishes smoothly from the center towards the elliptical boundary, and with consistent direction selectivity across the field. This view is supported by some of the first sensitivity mapping studies of MT RFs in macaques, although the spatial fidelity of the mapping was rather coarse (Lagae et al., 1994; Raiguel et al., 1995). Most studies fit the RF sensitivity profiles of recorded MT cells with a 2D Gaussian function (Albright & Desimone, 1987; Raiguel et al., 1995; Britten & Heuer, 1999), conforming to the presumption that an MT unit receives connections from V1 inputs whose RFs are spatially uniformly dispersed and tile the MT RF (Allman & Kaas, 1971; Dubner & Zeki, 1971). However, Richert and colleagues (2013) challenged this presumption when they found MT RFs with strikingly heterogeneous sensitivity profiles in the macaque.

7.1.1. Nonuniform spatial sampling in the MT receptive field

The discovery of Richert et al. (2013) led me to ask questions about the uniformity of spatial sampling in V1-MT connections. My simulation shows that the heterogeneous RF profiles of the kind documented by Richert and colleagues (2013) can be attributed to the randomized spatial integration of V1 inputs that are arranged in a nonuniform manner, a *random-wiring* hypothesis, only when the number of inputs is low, i.e., less than around 20 inputs per direction channel (240 across twelve direction channels). I found that increasing the spatial density and uniformity of V1 unit distribution (in terms of their RF locations) promotes more thorough spatial coverage of the MT RF, leading to more homogeneous sensitivity profiles (**Fig. 4-6A-D**) with fewer subregions (**Fig. 4-7**) and more regular outlines (**Fig. 4-10A-D**). In non-human primates, among the 5,000 ~ 10,000 synapses form on a single MT cell, it is estimated that at most a few hundred originate directly from V1 neurons, but these inputs may be significantly more dominant than those from other sources (Anderson et al., 1998). Given such an estimate is rather crude, it is difficult to assess how reasonable it is for an MT cell to have so few V1 inputs. Even if more V1 contacts are involved, it is possible that in MT neurons with more heterogeneous RF profiles, only a fraction of these inputs are significantly more influential than the rest, thus effectively limiting the number of functional V1 connections. It is also possible that direct V1 inputs account for only part of the MT CRF, and that a smaller number of inputs arise from V2 and V3 (Maunsell & Van Essen, 1983a; Felleman & Van Essen, 1991).

7.1.2. Localized V1 spatial substructures in the MT receptive field

I then tested if V1 units form spatial substructures of colocalized direction channels. To the best of my knowledge, no previous study has attempted to determine whether the MT neuron pools V1 inputs by recruiting RFs from completely random locations (the unstacked configuration), or by receiving connections from clusters of V1 channels that share the same spatial position but span the entire range of directions (the stacked configuration). The possible existence of such local substructures at the MT level has been suggested as a potential mechanism for explaining the localized pattern motion processing observed by Majaj and colleagues (2007), and this has been implemented in a model with local integrative subunits by Perrone and Krauzlis (2008).

Nevertheless, it is still possible to achieve localized PDS responses using an unstacked model which means that both such configurations remain plausible for the V1-MT circuit (as my other modeling results demonstrated, discussed below; also see **7.2 Localized pattern motion processing in MT and V1-MT motion opponency**). Here, my results show that unstacking such potential substructures results in a more homogeneous sensitivity profile across the RF map (**Fig. 4-6A-D**). This is because unstacked V1 inputs sample the RF more densely. Such an effect is exclusive to the PDS models as the narrow direction weighting function of the CDS units causes them to effectively receive inputs from only a single direction channel, significantly diminishing the difference between such stacked and unstacked models. I also found that unstacking PDS models increases the variability of direction preference on a global scale across the RF, but only to a very limited extent (**Fig. 4-6E-H**), and not enough to produce subregions with substantially different direction tuning (**Fig. 4-8**). In fact, unstacked PDS models almost always form single-region RFs owing to their highly homogeneous sensitivity profiles.

In essence, the topographic arrangement of V1 inputs directly affects the shape of the model MT RF. Stacked, uneven and sparsely connected V1 inputs are more inclined to produce RFs with irregular sensitivity profiles and multiple subregions. On the other hand, none of the above topographic variations of V1 input locations can accommodate MT RFs with diverse direction selectivity. With the V1 locations being randomly assigned under the random-wiring hypothesis in my models, the overall chance of a location to be covered by any direction channel is equal. Therefore, the level of heterogeneous direction selectivity within MT RFs reported by Richert and colleagues (2013) most likely arises from concerted, non-random developmental processes. Further studies of the properties and projection targets of these types of MT neurons is required to understand what they contribute to visual motion processing.

7.2. Localized pattern motion processing in MT and V1-MT motion opponency

The influential model for pattern motion processing of Rust et al. (2006) suggests that PDS MT cells integrate the signals from V1 units that are selective for a wide range of directions. While they recognized the importance of such integration for pattern motion processing, the integration of these V1 channels across space, presumably global, was not considered a critical issue. Like similar studies, they overlooked the link between

spatial integration and motion integration. However, when Majaj and colleagues (2007) examined the processing of spatially non-overlapping motion signals in the MT RF, they found that pattern motion processing is local in MT. This raised questions about the presumption of global V1-to-MT pooling and emphasized the crucial role of spatial integration to the network architecture of motion processing.

7.2.1. Local integration circuits and motion detectors

Majaj and colleagues (2007) noticed that disengaging spatially the component sinusoidal gratings of a drifting plaid causes PDS MT neurons to respond in a CDS fashion, i.e., the breakdown of pattern motion sensitivity to pseudo-plaids. Such findings imply the potential existence of localized structures in V1-MT wiring, and in the motion processing circuits specifically responsible for integrating across multiple component direction channels to solve the pattern motion problem. Using an MT model with localized pattern motion detectors, Perrone and Krauzlis (2008) reproduced such results in simulation. In their model, each local detection subunit receives excitatory inputs from five velocity-selective channels that fall in the 180° range flanking the PD of the detector, as well as inhibitory inputs from five channels that cover the 180° range surrounding the anti-PD of the detector. The speed preferences of these excitatory/inhibitory channels are scaled by the cosine of the difference angle between the PD/anti-PD of the subunit and the direction of the channel. The output of the local subunit is generated by combining the signals from the excitatory and inhibitory direction channels, followed by half-wave rectification, and the secondary global integration of these outputs follows.

Although the model of Perrone & Krauzlis (2008) seems to confirm the contribution of integrative subunits to the spatial specificity of pattern motion processing in MT, the structure of their motion detectors also localizes another pattern motion computation mechanism – motion opponency (which I will discuss in the following section). Therefore, their results must be interpreted with caution and be considered in the broader context of other motion processing mechanisms present in MT.

7.2.2. MT Motion opponency

The breakdown of pattern direction selectivity to pseudo-plaids in my MT models with local integrative subunits suggests that these subunits are sufficient, but not necessary, for localized pattern motion processing. My simulation shows that pattern direction selectivity to pseudo-plaids also breaks down in MT models without local integrative subunits (**Fig. 6-2**), implying other mechanisms may be at play.

While building the local motion detectors, Perrone and Krauzlis (2008) also implicitly included another localized mechanism deemed critical to pattern direction selectivity by Rust and colleagues (2006) – motion opponency. In such subunits, the localized bipolar inputs of the excitatory and inhibitory channels effectively establish MT motion opponency. Similarly, the strong (compared to those in CDS models) negative V1 inputs from the anti-PD channels in my PDS models also produce motion opponency (see **3.2. V1-to-MT integration**). Although this kind of motion opponency occurs in MT, the stacked architecture of the subunit constrains its

spatial scope to single V1 channels. Thus, either localized component signal integration or motion opponency could potentially explain results of Perrone and Krauzlis (2008).

A comparison of the responses of the MT models with and without local integrative subunits provides further insights on these two potential mechanisms of localized motion processing. The models with true local subunits differ from those without subunits by two distinct aspects of the V1-to-MT signal integration process: (1) the spatial colocalization of the V1 units of different direction channels within the subunit – the stacked configuration (the no-subunit models assume the unstacked configuration), and (2) the local nonlinear computation module – the half-wave rectification nonlinearity following local integration of the V1 signals within a stack. Both types of models (with and without local integrative subunits) show breakdown of pattern direction selectivity to pseudo-plaids, but the extent is more significant for the true-subunit models. To further isolate the mechanism responsible for this effect in the true-subunit models, I constructed false-subunit models, which adopt the stacked configuration, but not the local integration scheme. For these models, pattern direction selectivity also breaks down with pseudo-plaids; nevertheless, the extent is milder and similar to that of the no-subunit models.

For all these models, the presentation of true plaids drives motion signals in units sensitive to the two direction components across the entire RF, so that excitatory (PD) and inhibitory (anti-PD) V1 channels (defined by the positive and negative weights of V1-to-MT integration, see **3.2. V1-to-MT integration**) are activated simultaneously. The polarity of the V1 inputs from the PD and anti-PD channels facilitates PDS responses by allowing the component gratings to sculpt away the shoulders of the plaid direction tuning curve, narrowing it to be a better match to the single grating curve. The degree of the exhibited pattern direction selectivity is identical across all models. In models without true subunits, because signal integration is a single-stage global process, MT motion opponency also operates globally, meaning that an inhibitory V1 signal from any one location can counteract the excitatory inputs from any other locations. Conversely, in the true-subunit models, the nonlinear summation within the subunit causes such opponency to operate locally and requires colocalized stimulus components to activate the excitatory and inhibitory V1 channels concurrently to achieve effective contributions to the PDS behavior.

For the pseudo-plaid stimulus, the two different component directions are spatially segregated in non-overlapping patches. In this situation, a subunit is more likely to receive either excitatory inputs without inhibition, or inhibitory inputs without excitation. In the true-subunit models, the signal from a subunit receiving only inhibitory inputs will be blocked by localized half-wave rectification (**Eq. 3-13**). The loss of this opponent suppression significantly compromises pattern motion sensitivity in such models, similar to how the model of Perrone & Krauzlis (2008) can account for the spatial locality of pattern motion computation. In contrast, in the false-subunit models, even though the activations of the excitatory and inhibitory direction channels driven by the pseudo-plaid lack spatial colocalization, because signal integration happens globally for there is no local half-wave rectification, the inhibitory inputs are able to flow through. As a result, the breakdown of pattern direction selectivity to pseudo-plaids is the strongest in the models with true subunits,

i.e., localized nonlinear integration, whereas the level of breakdown is milder in, but identical between, the models with no subunit and those with false subunits (**Fig. 6-4**).

Therefore, the more extreme breakdown of pattern motion sensitivity in the true-subunit models is not because of the spatial substructures of colocalized V1 units across direction channels, but rather the spatial disassociation of component gratings in pseudo-plaids, which silences the suppressive signal from the opponent motion channel in the localized V1-to-MT signal integration circuit.

7.2.3. V1 motion opponency

Another source of the suppressive signal from opponent motion in the MT model is V1 motion opponency, the loss of which can explain why the models without true subunits also experience the breakdown of pattern direction selectivity to pseudo-plaids, although to a lesser extent. Unlike MT motion opponency, which is localized only in the true-subunit models, the subtraction of the opponent motion signal in V1 motion opponency occurs locally in all three kinds of models, functioning within a single V1 channel. When the colocalization of motion signals of different directions is disrupted in pseudo-plaids, the activations of the excitatory and inhibitory direction channels become spatially exclusive, leading to the decoupling of the excitatory and inhibitory signals and the subsequent discard of the latter during half-wave rectification (**Eq. 3-7**). A similar discovery was made by Bair and Baker (2016) that monocular V1 motion opponency is the key mechanism to account for the breakdown of pattern direction selectivity with dichoptic pseudo-plaids.

Overall, the breakdown of pattern direction selectivity to pseudo-plaids in my MT models is due to localized V1-MT motion opponency. In all three types of MT models, it can be attributed to the depletion of V1 motion opponency, and pattern motion sensitivity is further compromised in the true-subunit models when MT motion opponency is deactivated. The operating range of such motion opponency constrains the special scope of pattern motion processing. When I presented the multi-patch pseudo-plaids to the MT models, I found that the spatial limit of pattern motion processing directly correlates with the size of the V1 CRF (**Fig. 6-10A**), which limits the spatial extent of V1 motion opponency.

7.2.4. Pattern motion processing and the motion of objects

While exploring the spatial localization of pattern motion processing in MT, it is crucial to address the fundamental question: (1) should pattern motion computation only specifically consider image parts that are overlapping each other, as in the case of true plaids, or (2) should it occur more generally even for components that are spatially offset, as in the case of pseudo-plaids? The studies of Majaj et al. (2007) and Kumbhani et al. (2015) avoided this question by implicitly assuming that V1-MT motion processing does not distinguish the two cases. Such a premise effectively aligns with the second hypothesis, which is particularly relevant when considering the motion of an object with complex 2D contours that exhibit spatial locality. For example, in the case of a moving rectangle, the pattern motion problem could only be solved if the MT cell uses the two orientations of the perpendicular edges to integrate the components and compute pattern motion. Such an

integration scheme, independent of the spatial arrangement of the components, would provide a more versatile approach to object motion computation. Nonetheless, this approach runs the risk of integrating textures from separate objects, which challenges object segmentation.

To solve this issue, the V1-MT circuit may leverage additional visual information to group textures beyond the basis of 2D spatial specificity, enabling object segmentation. For instance, the MT cell could utilize the disparity cues (DeAngelis & Newsome, 1999; Prince et al., 2000) to distinguish overlapping or non-overlapping textures of different depths. By resolving the spatial arrangement of the components in 3D, the MT cell could selectively integrate only those textures that belong to the same object. It is also possible that area MT receives inputs from other cortical areas specializing in form processing, such as V4, to facilitate component grouping and coherent motion computation (Zeki, 1978; Desimone & Schein, 1987; Mountcastle et al., 1987; Ferrera et al., 1994; Tolias et al., 2005; Ungerleider et al., 2008).

7.3. Nonlinear interaction in the MT receptive field and nonlinear V1-to-MT signal integration

Electrophysiological studies have clearly demonstrated that spatial integration in the MT RF is not straightforward summation. When presenting multiple moving stimuli in the MT RF, Recanzone and colleagues (1997) discovered that the response to a pair of stimuli can be roughly predicted as the average of single-stimulus responses. Similarly, Britten and Heuer (1999) found that the double-stimulus response is significantly less than the sum of the responses to the individual stimuli. By analyzing the double-stimulus response as a function of the responses to the individual stimuli, which I refer to as the response summation function, they identified types of nonlinear spatial interactions within the MT RF: normalization and suppression.

7.3.1. Normalizing interaction in the MT receptive field

To capture the extent of nonlinearity in the response summation function, Britten and Heuer (1999) proposed a power-law summation model. The observed level of nonlinearity indicates that there is a normalizing operation that regulates the dynamic range of spatial integration in the MT RF, and the exponent readout of their fitted model reflects the degree of normalization. By augmenting the stacked MT models with no local integration (the false-subunit structure) with various nonlinear processing stages, I found two possible mechanisms may be behind such interaction.

First, to reproduce the kind of spatial normalization that can be accounted for by the power-law summation model, I adopted an integrative input normalization paradigm (nonlinear V1-to-MT integration) similar in formulation. By fitting the responses of the MT model to the power-law summation model, I found a nearly exact correspondence between the exponent of the power-law summation fit and the power of nonlinear integration in the MT model (**Fig. 5-4L**), demonstrating that such MT models can fully explain the level of MT spatial normalization observed by Britten and Heuer (1999).

Additionally, I also found that nonlinear spatial normalization can be accounted for by divisive population normalization in MT (Simoncelli & Heeger, 1998). By adjusting the relative scale between the parameters that govern the strength and the threshold of population normalization, I can control the exponent in the power-law summation fit of the response. I replicated the level of spatial normalization effect in the population-normalized linear integration models comparable to that in the models with integrative input normalization (**Fig. 5-6B**). The inclusion of V1 surround suppression and the centralized weighting/positioning of the V1 inputs in these models also reinforces the nonlinear behavior of the model (**Fig. 5-5B** and **Fig. 5-8C**, respectively).

This raises the question as to whether MT population normalization or nonlinear V1-to-MT integration more closely resembles the neural mechanism operating in the primate visual system. Biological implementation of either mechanism is feasible: inhibitory signals originating from neighboring MT cells may provide the force that drives the former, and the latter can be achieved through a softmax lateral network across the V1 neurons (Reichardt et al., 1983; Riesenhuber & Poggio, 1999). Below, I will attempt to address the evidence for either mechanism by discussing the potential suppressive interaction and the spatial sensitivity characteristics of the RFs produced by the MT models with these two schemes.

7.3.2. Suppressive interaction in the MT receptive field

Britten and Heuer (1999) quantified the strength of the suppressive interaction in the MT RF by examining the scaling factor of the fitted power-law summation model to the response summation functions of the MT neurons they recorded. They found, for many neurons, the scaling factor of the fit is less than 1.0, demonstrating substantial nonlinear suppressive interaction in the RF.

My simulation shows that nonlinear V1-to-MT integration alone cannot account for the level of suppressive interaction observed by Britten and Heuer (1999). Fitting the response summation functions of MT models with nonlinear integration operating at the physiological level ($\gamma = 2.72$) produced a scaling factor that is only slightly below 1.0 (**Fig. 5-4K**). However, when V1 surround suppression is incorporated in such nonlinear models, a stronger suppressive effect can be achieved (**Fig. 5-5A**). By comparison, adding divisive MT population normalization in the MT models with linear V1-to-MT integration achieves suppression in the MT RF similar to the level reported in Britten & Heuer (1999) (**Fig. 5-6A**).

The spatial structure of the suppressive field of population normalization in my MT model extends uniformly beyond the MT RF with a flat profile (Simoncelli & Heeger, 1998), which is consistent with Britten and Heuer's (1999) findings that the spatial scale of suppressive interaction in the MT RF surpasses the CRF, and is therefore much larger than the spatial extent of V1 surround suppression. This implies that V1 surround suppression is unlikely to be the dominant suppressive force in the MT RF because, otherwise, such suppression interaction in MT should have operated exclusively within the MT CRF. My simulation also corroborates such inference as the incorporation of V1 surround suppression in the linear integration models does not reduce the scaling factor (**Fig. 5-3E**). On the other hand, although suppression in the models with MT

population normalization can operate both within and beyond the CRF, a flat suppressive field is unlikely to be compatible with other spatial properties typical of non-human primate MT RFs.

7.3.3. Spatial sensitivity

The rationale behind divisive MT population normalization has its roots in non-DS MT surround suppression (Allman et al., 1985). However, the spatial profile of the suppressive filed in my model differs from past models (Raiguel et al., 1995; Liu & Hulle, 1998), and distorts the size tuning of the MT models significantly. Typically, the MT surround in RF models is represented as a Gaussian field, resulting in a reasonably wide size tuning curve, as is demonstrated by *in vivo* data (Tsui & Pack, 2011). Compared to the Gaussian surround, the flat spatial field of MT population normalization in my models generates a stronger suppressive effect, but at the cost of an unrealistic spatial sensitivity profile of the RF.

For the population normalization models with a physiological level of normalizing nonlinearity ($\beta_1 = 5.0$, $\beta_2 = 0.005$, $\bar{n} = 2.57$), the size tuning curves rise rapidly and reach the peak (**Fig. 5-10F**), rendering an excessively contracted shape that deviates from the behavior of MT cells previously reported (Raiguel et al., 1995). Ultimately, a universal suppressive force, which operates on a scale that extends from well within to remarkably beyond the MT CRF, as indicated by Britten and Heuer (1999), inherently contradicts a realistic shape of size tuning curves that rises at a moderate slope with a practical optimal stimulus size.

It is possible that Britten and Heuer (1999) over-estimated the suppressive interaction in the MT CRF due to the interference of spatial attention allocation. Because awake animals were used in the study, when a pair of stimuli appeared, the animal's attention might have been distributed unequally between the stimuli. The unbalanced spatial attention could have modulated the spatial structure of the MT RF, causing it to shrink and shift to the attended stimulus (Womelsdorf et al., 2006; Anton-Erxleben et al., 2009). Therefore, the unattended stimulus could have resided in an area with weaker sensitivity, or in the primarily inhibitory surround. Hence, the estimated scaling factor of power-law summation fit would in turn have been biased towards lower values. To accurately characterize the suppressive interaction in the MT RF, future experiments with anesthetized animals could be carried out to avoid attention shifts, or experimental paradigms that control attention could be used in the awake preparation.

On the other hand, for the models with nonlinear V1-to-MT integration operating at a comparable level of normalization ($\gamma = 2, \bar{n} = 2.67$), the shape of the size tuning curves is more realistic (**Fig. 5-10E**). The level of nonlinearity of spatial integration in the MT model also affects the size of the RF. For these models, the size of the RF estimated through the 2D Gaussian fit of the spatial response is considerably larger than the optimal size of stimulus indicated by the peak of the size tuning curve (**Fig. 5-10E**). A group of non-human primate MT cells with similar spatial sensitivity profiles have been reported in the past, referred to as having *complex* RFs by Born (2000), who also pointed out that such RFs should be less effectively driven by two small patches of motion than one. As I have demonstrated that raising the nonlinearity of the integration process further extends the size of the model MT RF and yet reduces the optimal stimulus size as determined by the peak of

the size tuning curve (**Fig. 5-10E**), future studies are needed to establish whether the predicted correlation between the level of spatial normalization in MT and the properties of the spatial sensitivity profile of the RF hold *in vivo*.

Overall, these considerations suggest that nonlinear interactions in the MT RF are more likely the direct outcome of nonlinear spatial integration rather than MT population normalization, and may potentially be enhanced by other RF substructures, such as V1 surround suppression and centralized V1 unit weighting/positioning (**Fig. 5-5** and **Fig. 5-8B&C**). Further evidence against divisive MT population normalization includes my observation of an anti-correlation between the exponent and the scaling factor of the response summation functions for models with population normalization (**Fig. 5-6**), whereas Britten and Heuer (1999) reported an absence of correlation between these two parameters.

7.4. Model limitations

Despite its ability to simulate certain spatial aspects of the MT RF, my current MT model falls short in capturing the full complexity of motion and spatial processing in MT. Here, I will highlight the limitations of the model and propose potential *in vivo* research to bridge the gap between model predictions and experimental verification.

7.4.1. Simplified MT receptive field substructure

Although my MT model includes the spatial integration of V1 RFs with well-defined substructures, such as the CRF and the surround, the spatial structure of the MT RF is oversimplified. The model only depicts a classical excitatory area and an optional flat suppressive field, ignoring the significant DS suppressive signals present in the MT RF as demonstrated by multiple studies (Allman et al., 1985; Born, 2000; Huang et al., 2007; Hunter & Born, 2011; Tsui & Pack, 2011). For the MT models probed with the dynamic random dot stimulus, it is possible that adding a DS surround field, and allowing some irregularity in its structure, could allow the model to generate RFs with heterogeneous direction preferences as observed for a group of neurons by Richert et al. (2013).

Earlier evidence has supported a classical view of the suppressive field of MT surround that is antagonistic and circularly symmetric (Allman et al., 1985; Tanaka et al., 1986; Lagae et al., 1989), while later reports also suggest that RFs with surround reinforcement of same-direction motion are prevalent in MT, with occasional orthogonal-motion surrounds also present (Born, 2000). Fine-scale mapping of the MT RF in non-human primates has shown that the suppressive surround can be spatially offset to the center, and the modeling of the RF profile has also revealed a wide range of direction preferences for the inhibitory subfield (Xiao et al., 1995; Cui et al., 2013). It is possible that the different direction selectivity and the irregular spatial profiles of the excitatory and inhibitory subfields can give rise to heterogeneous local direction tuning in the RF and produce subregions with diverse direction preferences. Such a hypothesis can be verified through animal studies: if spatially unbalanced surround suppression is the origin of multi-direction preference in the MT RF, given that

surround suppression is contrast dependent in MT (Tsui & Pack, 2011), adjusting the contrast of the random dot stimulus should cause the measured multi-direction RFs to become more unidirectional when surround suppression is impeded.

7.4.2. Monocular processing

Examining monocular models, meaning that V1 units receive inputs from only one image representing a single eye, rather than two images, my exercise overlooks the possible binocular origin of multi-direction preference in the MT RF – the varying direction tuning across the subregions of individual RFs reported by Richert and colleagues (2013) may have reflected the inconsistency in preferred directions across the two eyes (note, they used awake, fixating monkeys). Indeed, MT neurons are well driven by stimuli presented to either eye (Zeki, 1974a; Maunsell and Van Essen, 1983a; Felleman and Kaas, 1984; DeAngelis and Uka, 2003). More recent research has shown that many MT cells encode motion-in-depth signals through interocular velocity difference cues, and the discrepancy of direction preferences between the two eyes, accompanied by imbalanced binocular integration (Zeki, 1974b; Czuba et al., 2014; Sanada & Deangelis, 2014), may form the basis for such processing. Future experiments that map the RF profile and measure the direction tuning of MT cells for each eye separately can provide valuable data for testing this hypothesis.

The binocular MT model of Baker & Bair (2016) may shed some light on how we could build a binocular MT unit with RF subregions that prefer different directions. Having 3D motion sensitivity and binocular direction tuning similar to those reported by Czuba and colleagues (2014), their model features the following: (1) PDS-like weighting of direction channels in the stage of V1-to-MT integration is crucial; (2) the two eyes prefer opposite directions; and (3) the model is ocularly imbalanced. Provided such insights, one could test if MT RFs with subregions of diverse direction tuning could be achieved by building binocular PDS models with opposite direction preferences in the two eyes using a stacked-uneven configuration of sparse V1 inputs (which boosts the sensitivity heterogeneity across the RF spatial profile). However, such a modeling study should be accompanied by *in vivo* data collection to determine if there is any biological basis for a connection between 3D motion and RF heterogeneity in MT.

7.4.3. The Nonlinear V1-to-MT integration scheme

I compared the power-law and softmax fit of the response summation functions of the MT models with nonlinear V1-to-MT integration, and showed that either fitting scheme can well approximate the response summation functions (**Fig. 5-9F**). Given that a straight-forward nonlinear integration circuit with power operations may not be biologically feasible, it is possible that such nonlinear integrative normalization can be achieved through a softmax integration mechanism and has been adopted by modeling studies of the MT (Nowlan & Sejnowski, 1995; Tsui et al., 2010), as well as other visual areas (Riesenhuber & Poggio, 1999). Softmax integration can be implemented in several different types of neural circuits (Feldman & Ballard, 1982; Koch & Ullman, 1985; Grossberg, 1988; Yuille & Grzywacz, 1989; Carandini & Heeger, 1994; Douglas et al., 1995;

Abbot et al., 1997; Chance et al., 1999), and most likely arises from cortical microcircuits of lateral, possibly recurrent, inhibitory activities between neurons within a cortical layer (Riesenhuber & Poggio, 1999). An example is that the proposed *pool* cells in the fly visual system facilitate a relative-motion detection circuit through feedforward (or recurrent) shunting presynaptic (or postsynaptic) inhibition (Reichardt et al., 1983).

The softmax integration scheme has yet to be incorporated into the current version of my MT model, making it difficult to determine its impact on the model's tuning for different visual features. However, changing the level of nonlinearity in the current power-law integration model has been shown to alter the spatial and direction tuning of the model (**Fig. 5-10E&H**). Depending on the metric of choice, stronger nonlinearity leads to the paradoxical expansion or shrinking of the spatial RF, and correlates with compromised pattern motion sensitivity. In order to comprehensively investigate the role of nonlinear mechanisms in shaping motion and spatial sensitivity, incorporating softmax V1-to-MT integration into my current model is a priority for future investigation.

7.4.4. Potentials of artificial neural networks

My research has followed a long tradition of building a hierarchy of linear and non-linear steps that are motivated directly by visual neuroscientific data. However, an increasing number of artificial neural networks are now being trained on time-varying image sequences (videos) to learn to carry out tasks that involve categorizing actions (Bertasius et al., 2021). As these networks begin to perform closer to the human level, it will be interesting to study the heterogeneity of receptive fields and the normalization strategies that are learned during training. This could provide another line of evidence to help us understand the physiological properties observed along the visual motion pathway in the brain.

7.5. Conclusion

In summary, the results from my MT model shows that the structure of the MT RF is shaped by the spatial integration of signals from V1 neurons. The randomness of the topographic arrangement of connections in the V1-MT integrative network can contribute to the heterogeneity of the sensitivity profile of the MT RF, and the nonlinear operations carried out by such a network play a role in the multi-stimulus interaction within the MT RF. Additionally, the spatial limit of motion processing in MT can be influenced by the spatial extent of the integrated V1 CRFs, specifically, the spatial constraint of opponent motion suppression.

While evaluating the spatial influence of tuning the relevant parameters of the integration process on the model RF, I was struck by the interdependence of motion and spatial sensitivity of the model units, which is however not unexpected given them being the different consequential aspects of the same integration process. For instance, although Richert and colleagues (2013) did not comment on the type of direction selectivity, i.e., component or pattern, of the neurons they recorded, my simulation suggests those single-peaked RFs with low geometric irregularity are more likely to belong to PDS cells with neural circuits that do not involve localized spatial substructures. In contrast, the multi-peaked RFs reported by them are more likely to originate from CDS

neurons with low input density. Future experiments should test the potential correlation between the direction selectivity types and the RF sensitivity heterogeneity.

In addition to the V1-to-MT integration process, the other two critical mechanisms of pattern motion processing, as outlined in Rust et al. (2006), namely tuned normalization through V1 surround suppression, and motion opponency, also play a significant role in defining the spatial properties of the MT RF. In this section, I will demonstrate the interconnected nature of spatial and motion processing in MT by examining the impacts of these three mechanisms on the motion and spatial sensitivity of the model units.

7.5.1. V1-to-MT integration

The geometry of the model MT RF is directly linked to the topographic layout of the integrated V1 channels, and the computational properties of the V1-MT connectivity define the sub-RF structure of spatial processing. The level of heterogeneity in the spatial sensitivity profile of the MT RF is influenced by the degree of the nonuniformity and sparsity of the V1 location arrangement (**Fig. 4-6**). These modeled V1 cells are connected to the MT unit through feed-forward synapses and form an integrative circuit, which, in the biological neural system, is likely accompanied by a lateral nonlinear network across the recruited V1 units. This network potentially performs a softmax-like operation, and is the key mechanism to account for the normalizing interactions within the MT RF, and the degree of nonlinearity affects the spatial extent of the MT RF. The strength of the feed-forward connections varies based on the location of the V1 RF, leading to a spatial dependency of the interaction within the MT RF, as observed by Britten and Heuer (1999). The connections originating from V1 units located at the center of the MT RF are likely to carry heavier weights compared to those located further away, and the spatial weighting function of the V1 channels should have a concave shape (**Fig. 5-7**), reflecting the nonlinear spatial processing in the MT RF.

The nonlinear property of the integrative circuit also has implications for motion sensitivity in MT. Heightened nonlinear processing is associated with impaired pattern direction selectivity in my MT models (**Fig. 6-10H**). Future *in vivo* studies may be able to test the potential link between nonlinear integration and component/pattern direction selectivity in MT.

7.5.2. V1 surround suppression

V1 surround suppression impacts nonlinear spatial processing in the MT models in a faciliatory fashion. In MT models with linear V1-to-MT integration, V1 surround suppression does not introduce any effective normalizing interactions in the RF, and its impact on the suppressive interactions is extremely subtle (**Fig. 5-3E&F**). However, in models with a moderate level of nonlinear integration, V1 surround suppression contributes to the nonlinear spatial processing, leading to enhanced normalization and suppression (**Fig. 5-5**). These findings highlight the modulatory role of V1 surround suppression in shaping the spatial processing in the MT RF.

7.5.3. Opponent motion inhibition

Kumbhani and colleagues (2015) speculated that the spatial limit of localized pattern motion processing in the MT RF is determined by the spatial extent of V1 surround suppression. However, they did not provide direct evidence for this hypothesis, and it would be difficult to verify experimentally. Contrary to their speculation, my simulation shows that the spatial limit of motion integration in the MT models is not affected by the changes of the spatial extend of V1 surround (**Fig. 6-5**).

On the other hand, my results indicate that local motion opponency is the major determining factor for localized motion integration in MT. Pattern motion sensitivity of the PDS MT units is disrupted by the pseudo-plaid stimuli due to diminished opponent motion inhibition as the signals from the component gratings of different directions are spatially segregated. Consequently, I found that the spatial limit of pattern motion processing in the modeled MT RFs correlates with the size of the V1 CRF (**Fig. 6-10A**), which constrains the operating range of V1 motion opponency. Pattern direction selectivity recovers when the patches of the pseudo-plaid become small and dense enough to fit the gratings of different directions into individual V1 CRFs, restoring V1 motion opponency. Furthermore, potential integrative subunits formed by stacked V1 direction channels and the corresponding nonlinear computation can potentially further constrict motion integration spatially by localizing MT-level opponent motion suppression.

In conclusion, the findings from my simulation highlight the intricate interplay between spatial and motion processing in the MT RF. By incorporating biologically realistic spatial configurations for the key mechanisms that drive motion processing in MT, I am able to demonstrate how these mechanisms can explain the observed spatial structure of MT RFs, and the interaction of multiple concurrent signals within such, as observed in numerous *in vivo* studies. These results add to our understanding of the complex mechanisms involved in visual processing in MT and contribute to the ongoing research aimed at deciphering its functional roles.

References:

Abbott, L. F., Varela, J. A., Sen, K., & Nelson, S. B. (1997). Synaptic depression and cortical gain control. Science (New York, N.Y.), 275(5297), 220–224.

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America. A, Optics and image science, 2(2), 284–299.

Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. Visual neuroscience, 7(6), 531-546.

Albright, T. D., & Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. Experimental brain research, 65(3), 582-592.

Allman, J. M., & Kaas, J. H. (1971). A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey (Aotus trivirgatus). Brain research, 31(1), 85-105.

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction-and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). Perception, 14(2), 105-126.

Anderson, J. C., Binzegger, T., Martin, K. A., & Rockland, K. S. (1998). The connection from cortical area V1 to V5: a light and electron microscopic study. The Journal of neuroscience : the official journal of the Society for Neuroscience, 18(24), 10525–10540.

Angelucci, A., Levitt, J. B., & Lund, J. S. (2002). Anatomical origins of the classical receptive field and modulatory surround field of single neurons in macaque visual cortical area V1. Progress in brain research, 136, 373-388.

Anton-Erxleben, K., Stephan, V. M., & Treue, S. (2009). Attention reshapes center-surround receptive field structure in macaque cortical area MT. Cerebral cortex (New York, N.Y. : 1991), 19(10), 2466–2478.

Baker, J. F., Petersen, S. E., Newsome, W. T., & Allman, J. M. (1981). Visual response properties of neurons in four extrastriate visual areas of the owl monkey (Aotus trivirgatus): a quantitative comparison of medial, dorsomedial, dorsolateral, and middle temporal areas. Journal of neurophysiology, 45(3), 397-416.

Baker, P. M., & Bair, W. (2016). A model of binocular motion integration in MT neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience, 36(24), 6563-6582.

Baker, P., & Bair, W. (2017). A model for spatial integration of pattern and 3D motion in MT neurons. Journal of vision, 17(10), 410-410.

Bertasius, G., Wang, H., & Torresani, L. (2021). Is space-time attention all you need for video understanding?. In M. Meila & T. Zhang (Eds.), Proceedings of the 38th international conference on machine learning (pp. 813–824).

Bonds, A. B. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. Visual neuroscience, 2(1), 41-55.

Born, R. T., & Tootell, R. B. (1992). Segregation of global and local motion processing in primate middle temporal visual area. Nature, 357(6378), 497-499.

Born, R. T. (2000). Center-surround interactions in the middle temporal visual area of the owl monkey. Journal of neurophysiology, 84(5), 2658-2669.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. Annual review of neuroscience, 28, 157–189.

Bowns, L. (2002). Can spatio-temporal energy models of motion predict feature motion?. Vision research, 42(13), 1671-1681.

Braddick, O. (1974). A short-range process in apparent motion. Vision research, 14(7), 519-527.

Bradley, D. C., & Goyal, M. S. (2008). Velocity computation in the primate visual system. Nature reviews. Neuroscience, 9(9), 686–695.

Bresenham, J. E. (1965). Algorithm for computer control of a digital plotter. IBM Systems journal, 4(1), 25-30.

Bridle, J. (1989). Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. In D. Touretzky (Ed.), Advances in neural information processing Systems (Vol. 2). Morgan-Kaufmann.

Britten, K. H., & Heuer, H. W. (1999). Spatial summation in the receptive fields of MT neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience, 19(12), 5074-5084.

Carandini, M., & Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. Science (New York, N.Y.), 264(5163), 1333–1336.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience, 17(21), 8621-8644.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. Journal of neurophysiology, 88(5), 2530-2546.

Chance, F. S., Nelson, S. B., & Abbott, L. F. (1999). Complex cells as cortically amplified simple cells. Nature neuroscience, 2(3), 277-282.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. Network: computation in neural systems (Bristol, England), 12(2), 199–213.

Churchland, M. M., Priebe, N. J., & Lisberger, S. G. (2005). Comparison of the spatial limits on direction selectivity in visual areas MT and V1. Journal of neurophysiology, 93(3), 1235-1245.

Cui, Y., Liu, L. D., Khawaja, F. A., Pack, C. C., & Butts, D. A. (2013). Diverse suppressive influences in area MT and selectivity to complex motion features. The Journal of neuroscience : the official journal of the Society for Neuroscience, 33(42), 16715-16728.

Czuba, T. B., Huk, A. C., Cormack, L. K., & Kohn, A. (2014). Area MT encodes three-dimensional motion. The Journal of neuroscience : the official journal of the Society for Neuroscience, 34(47), 15522-15533.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A., & Suarez, H. H. (1995). Recurrent excitation in neocortical circuits. Science (New York, N.Y.), 269(5226), 981-985.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. Vision research, 20(10), 847-856.

Daugman J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America. A, Optics and image science, 2(7), 1160–1169.

DeAngelis, G. C., Robson, J. G., Ohzawa, I., & Freeman, R. D. (1992). Organization of suppression in receptive fields of neurons in cat visual cortex. Journal of neurophysiology, 68(1), 144-163.

DeAngelis, G. C., & Newsome, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. The Journal of neuroscience : the official journal of the Society for Neuroscience, 19(4), 1398–1415.

DeAngelis, G. C., & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. Journal of neurophysiology, 89(2), 1094-1111.

Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. Journal of neurophysiology, 57(3), 835–868.

Dubner, R., & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. Brain research, 35(2), 528–532.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. Cognitive science, 6(3), 205-254.

Felleman, D. J., & Kaas, J. H. (1984). Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. Journal of neurophysiology, 52(3), 488-513.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral cortex (New York, N.Y. : 1991), 1(1), 1–47.

Fennema, C. L., & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. Computer graphics and image processing, 9(4), 301-315.

Ferrera, V. P., Rudolph, K. K., & Maunsell, J. H. (1994). Responses of neurons in the parietal and temporal visual pathways during a motion task. The Journal of neuroscience : the official journal of the Society for Neuroscience, 14(10), 6171–6186.

Gattass, R., & Gross, C. G. (1981). Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque. Journal of neurophysiology, 46(3), 621–638.

Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. Neural networks, 1(1), 17-61.

Grzywacz, N. M., & Yuille, A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. Proceedings of the Royal Society of London. Series B, Biological sciences, 239(1295), 129–161.

Heeger D. J. (1987). Model for the extraction of image flow. Journal of the Optical Society of America. A, Optics and image science, 4(8), 1455–1471.

Heeger, D. J. (1988). Optical flow using spatiotemporal filters. International journal of computer vision, 1(4), 279-302.

Heeger, D. J. (1991). Nonlinear model of neural responses in cat visual cortex. In M. S. Landy & J. A. Movshon (Eds.), Computational models of visual processing (pp. 119–133). The MIT Press.

Heeger, D. J. (1992a). Half-squaring in responses of cat striate cells. Visual neuroscience, 9(5), 427-443.

Heeger, D. J. (1992b). Normalization of cell responses in cat striate cortex. Visual neuroscience, 9(2), 181-197.

Heeger, D. J. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. Journal of neurophysiology, 70(5), 1885-1898.

Hedges, J. H., Stocker, A. A., & Simoncelli, E. P. (2011). Optimal inference explains the perceptual coherence of visual motion stimuli. Journal of vision, 11(6), 14-14.

Henry, G. H. (1977). Receptive field classes of cells in the striate cortex of the cat. Brain research, 133(1), 1-28.

Heuer, H. W., & Britten, K. H. (2002). Contrast dependence of response normalization in area MT of the rhesus macaque. Journal of neurophysiology, 88(6), 3398-3408.

Hietanen, M. A., Cloherty, S. L., Van Kleef, J. P., Wang, C., Dreher, B., & Ibbotson, M. R. (2013). Phase sensitivity of complex cells in primary visual cortex. Neuroscience, 237, 19-28.

Holub, R. A., & Morton-Gibson, M. (1981). Response of visual cortical neurons of the cat to moving sinusoidal gratings: response-contrast functions and spatiotemporal interactions. Journal of neurophysiology, 46(6), 1244-1259.

Huang, X., Albright, T. D., & Stoner, G. R. (2007). Adaptive surround modulation in cortical area MT. Neuron, 53(5), 761-770.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of physiology, 160(1), 106-154.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. Journal of neurophysiology, 28(2), 229-289.

Hunter, J. N., & Born, R. T. (2011). Stimulus-dependent modulation of suppressive influences in MT. The Journal of neuroscience : the official journal of the Society for Neuroscience, 31(2), 678-686.

Ibbotson, M. R., Price, N. S. C., & Crowder, N. A. (2005). On the division of cortical cells into simple and complex types: a comparative viewpoint. Journal of neurophysiology, 93(6), 3699-3702.

Ikeda, H., & Wright, M. J. (1975). Spatial and temporal properties of ‘sustained’ and ‘transient’ neurones in area 17 of the cat's visual cortex. Experimental brain research, 22(4), 363-383.

Jones, H. E., Grieve, K. L., Wang, W., & Sillito, A. M. (2001). Surround suppression in primate V1. Journal of neurophysiology, 86(4), 2011-2028.

Koch, C., & Ullman, S. (1987). Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. In L. M. Vaina (Ed.), Matters of Intelligence: Conceptual Structures in Cognitive Neuroscience (pp. 115–141). Springer Netherlands.

Kumbhani, R. D., El-Shamayleh, Y., & Movshon, J. A. (2015). Temporal and spatial limits of pattern motion sensitivity in macaque MT neurons. Journal of neurophysiology, 113(7), 1977-1988.

Lagae, L., Gulyas, B., Raiguel, S., & Orban, G. A. (1989). Laminar analysis of motion information processing in macaque V5. Brain research, 496(1-2), 361-367.

Lagae, L., Raiguel, S., & Orban, G. A. (1993). Speed and direction selectivity of macaque middle temporal neurons. Journal of neurophysiology, 69(1), 19-39.

Lagae, L., Maes, H., Raiguel, S., Xiao, D. K., & Orban, G. A. (1994). Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST. Journal of neurophysiology, 71(5), 1597-1626.

Lennie, P., & Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway (invited review). Journal of the Optical Society of America. A, Optics, image science, and vision, 22(10), 2013–2033.

Liu, L., & Hulle, M. M. V. (1998). Modeling the surround of MT cells and their selectivity for surface orientation in depth specified by motion. Neural computation, 10(2), 295-312.

Livingstone, M. S., & Conway, B. R. (2003). Substructure of direction-selective receptive fields in macaque V1. Journal of neurophysiology, 89(5), 2743-2759.

Livingstone, M. S., Pack, C. C., & Born, R. T. (2001). Two-dimensional substructure of MT receptive fields. Neuron, 30(3), 781-793.

Lund, J. S., Lund, R. D., Hendrickson, A. E., Bunt, A. H., & Fuchs, A. F. (1975). The origin of efferent pathways from the primary visual cortex, area 17, of the macaque monkey as shown by retrograde transport of horseradish peroxidase. The Journal of comparative neurology, 164(3), 287–303.

Majaj, N. J., Carandini, M., & Movshon, J. A. (2007). Motion integration by neurons in macaque MT is local, not global. The Journal of neuroscience : the official journal of the Society for Neuroscience, 27(2), 366-370.

Maunsell, J. H., & Van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. Journal of neurophysiology, 49(5), 1127-1147.

Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. Journal of neurophysiology, 49(5), 1148-1167.

Mazurek, M., Kager, M., & Van Hooser, S. D. (2014). Robust quantification of orientation selectivity and direction selectivity. Frontiers in neural circuits, 8, 92.

McDonald, J. S., Clifford, C. W., Solomon, S. S., Chen, S. C., & Solomon, S. G. (2014). Integration and segregation of multiple motion signals by neurons in area MT of primate. Journal of neurophysiology, *111*(2), 369-378.

McLean, J., & Palmer, L. A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. Vision research, 29(6), 675-679.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. Journal of neurophysiology, 55(6), 1328-1339.

Mountcastle, V. B., Motter, B. C., Steinmetz, M. A., & Sestokas, A. K. (1987). Common and differential effects of attentive fixation on the excitability of parietal and prestriate (V4) cortical visual neurons in the macaque monkey. The Journal of neuroscience : the official journal of the Society for Neuroscience, 7(7), 2239–2255.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. Gross (Eds.), Pattern recognition mechanisms (Vol. 54, Ser. Pontificiae academiae scientiarum scripta varia, pp. 117–151). Vatican Press.

Movshon, J. A., & Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. The Journal of neuroscience : the official journal of the Society for Neuroscience, 16(23), 7733-7741.

Nassi, J. J., & Callaway, E. M. (2007). Specialized circuits from primary visual cortex to V2 and area MT. Neuron, 55(5), 799-808.

Nestares, O., & Heeger, D. J. (1997). Modeling the apparent frequency-specific suppression in simple cell responses. Vision research, 37(11), 1535-1543.

Nishimoto, S., & Gallant, J. L. (2011). A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. The Journal of neuroscience : the official journal of the Society for Neuroscience, 31(41), 14551–14564.

Nowlan, S. J., & Sejnowski, T. J. (1994). Filter selection model for motion segmentation and velocity integration. Journal of the Optical Society of America. A, Optics, image science, and vision, 11(12), 3177-3200.

Nowlan, S. J., & Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates. The Journal of neuroscience : the official journal of the Society for Neuroscience, 15(2), 1195-1214.

Orban, G. A., Kennedy, H., & Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. Journal of neurophysiology, 56(2), 462-480.

Pack, C. C., Gartland, A. J., & Born, R. T. (2004). Integration of contour and terminator signals in visual area MT of alert macaque. The Journal of neuroscience : the official journal of the Society for Neuroscience, 24(13), 3268-3280.

Perrone, J. A. (2004). A visual motion sensor based on the properties of V1 and MT neurons. Vision research, 44(15), 1733-1755.

Perrone, J. A., & Krauzlis, R. J. (2008). Spatial integration by MT pattern neurons: A closer look at pattern-to-component effects and the role of speed tuning. Journal of vision, 8(9), 1-14.

Priebe, N. J., Cassanello, C. R., & Lisberger, S. G. (2003). The neural representation of speed in macaque area MT/V5. The Journal of neuroscience : the official journal of the Society for Neuroscience, 23(13), 5650-5661.

Prince, S. J., Pointon, A. D., Cumming, B. G., & Parker, A. J. (2000). The precision of single neuron responses in cortical area V1 during stereoscopic depth judgments. The Journal of neuroscience : the official journal of the Society for Neuroscience, 20(9), 3387–3400.

Raiguel, S., Van Hulle, M. M., Xiao, D. K., Marcar, V. L., & Orban, G. A. (1995). Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. European journal of neuroscience, 7(10), 2064-2082.

Recanzone, G. H., Wurtz, R. H., & Schwarz, U. (1997). Responses of MT and MST neurons to one and two moving objects in the receptive field. Journal of neurophysiology, 78(6), 2904-2915.

Reichardt, W., Poggio, T., & Hausen, K. (1983). Figure-ground discrimination by relative movement in the visual system of the fly. Biological cybernetics, 35(2), 81-100.

Reid, R. C., Soodak, R. E., & Shapley, R. M. (1987). Linear mechanisms of directional selectivity in simple cells of cat striate cortex. Proceedings of the National Academy of Sciences of the United States of America, 84(23), 8740–8744.

Richert, M., Albright, T. D., & Krekelberg, B. (2013). The complex structure of receptive fields in the middle temporal area. Frontiers in systems neuroscience, 7, 2.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature neuroscience, 2(11), 1019-1025.

Roberts, M. (2018). The unreasonable effectiveness of quasirandom. EXTREME LEARNING. http://extremelearning.com.au/unreasonable-effectiveness-of-quasirandom-sequences/

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. Nature neuroscience, 9(11), 1421-1431.

Saito, H., Tanaka, K., Isono, H., Yasuda, M., & Mikami, A. (1989). Directionally selective response of cells in the middle temporal area (MT) of the macaque monkey to the movement of equiluminous opponent color stimuli. Experimental brain research, 75(1), 1-14.

Sanada, T. M., & DeAngelis, G. C. (2014). Neural representation of motion-in-depth in area MT. The Journal of neuroscience : the official journal of the Society for Neuroscience, 34(47), 15508-15521.

Schwartz, O., Pillow, J. W., Rust, N. C., & Simoncelli, E. P. (2006). Spike-triggered neural characterization. Journal of vision, 6(4), 13-13.

Shipp, S., & Zeki, S. (1989). The organization of connections between areas V5 and V2 in macaque monkey visual cortex. European journal of neuroscience, 1(4), 333-354.

Shirley, P., & Chiu, K. (1997). A low distortion map between disk and square. Journal of graphics tools, 2(3), 45-52.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. Vision research, 38(5), 743-761.

Smith, M. A., Majaj, N. J., & Movshon, J. A. (2005). Dynamics of motion signaling by neurons in macaque area MT. Nature neuroscience, 8(2), 220-228.

Snowden, R. J., Treue, S., & Andersen, R. A. (1992). The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. Experimental brain research, 88(2), 389-400.

Tailby, Chris, Najib J. Majaj, and J. Anthony Movshon. "Binocular integration of pattern motion signals by MT neurons and by human observers." The Journal of neuroscience : the official journal of the Society for Neuroscience 30, no. 21 (2010): 7344-7349.

Tanaka, K., Hikosaka, K., Saito, H. A., Yukie, M., Fukada, Y., & Iwai, E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. The Journal of neuroscience : the official journal of the Society for Neuroscience, 6(1), 134-144.

Tolhurst, D. J., & Movshon, J. A. (1975). Spatial and temporal contrast sensitivity of striate cortical neurones. Nature, 257(5528), 674-675.

Tolhurst, D. J., & Heeger, D. J. (1997a). Comparison of contrast-normalization and threshold models of the responses of simple cells in cat striate cortex. Visual neuroscience, 14(2), 293-309.

Tolhurst, D. J., & Heeger, D. J. (1997b). Contrast normalization and a linear model for the directional selectivity of simple cells in cat striate cortex. Visual neuroscience, 14(1), 19-25.

Tolias, A. S., Keliris, G. A., Smirnakis, S. M., & Logothetis, N. K. (2005). Neurons in macaque area V4 acquire directional tuning after adaptation to motion stimuli. Nature neuroscience, 8(5), 591–593.

Tsui, J. M., Hunter, J. N., Born, R. T., & Pack, C. C. (2010). The role of V1 surround suppression in MT motion integration. Journal of neurophysiology, 103(6), 3123-3138.

Tsui, J. M., & Pack, C. C. (2011). Contrast sensitivity of MT receptive field centers and surrounds. Journal of neurophysiology, 106(4), 1888-1900.

Ungerleider, L. G., & Desimone, R. (1986). Cortical connections of visual area MT in the macaque. The Journal of comparative neurology, 248(2), 190-222.

Ungerleider, L. G., Galkin, T. W., Desimone, R., & Gattass, R. (2008). Cortical connections of area V4 in the macaque. Cerebral cortex (New York, N.Y. : 1991), 18(3), 477–499.

Wang, H. X., & Movshon, J. A. (2016). Properties of pattern and component direction-selective cells in area MT of the macaque. Journal of neurophysiology, 115(6), 2705–2720.

Watson, A. B., & Ahumada Jr, A. J. (1983). A look at motion in the frequency domain (No. A-9304). NASA Technical Memorandum, 84352.

Watson, A. B., & Ahumada, A. J. (1985). Model of human visual-motion sensing. Journal of the Optical Society of America. A, Optics, image science, and vision, 2(2), 322-342.

Watson, A. B., & Turano, K. (1995). The optimal motion stimulus. Vision research, 35(3), 325-336.

Wiesner, S., Baumgart, I. W., & Huang, X. (2020). Spatial arrangement drastically changes the neural representation of multiple visual stimuli that compete in more than one feature domain. The Journal of neuroscience : the official journal of the Society for Neuroscience, 40(9), 1834-1848.

Womelsdorf, T., Anton-Erxleben, K., Pieper, F., & Treue, S. (2006). Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. Nature neuroscience, 9(9), 1156-1160.

Xiao, D. K., Raiguel, S., Marcar, V., Koenderink, J., & Orban, G. A. (1995). Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area. Proceedings of the National Academy of Sciences, 92(24), 11303-11306.

Yuille, A. L., & Grzywacz, N. M. (1989). A winner-take-all mechanism based on presynaptic inhibition feedback. Neural computation, 1(3), 334-347.

Zeki, S. M. (1969). Representation of central visual fields in prestriate cortex of monkey. Brain research, 14(2), 271-291.

Zeki, S. M. (1971). Convergent input from the striate cortex (area 17) to the cortex of the superior temporal sulcus in the rhesus monkey. Brain research, 28(2), 338-340.

Zeki, S. M. (1974a). Cells responding to changing image size and disparity in the cortex of the rhesus monkey. The Journal of physiology, 242(3), 827-841.

Zeki, S. M. (1974b). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. The Journal of physiology, 236(3), 549-573.

Zeki S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. The Journal of physiology, 277, 273–290.