Systems/Circuits

A Model of Binocular Motion Integration in MT Neurons

Pamela M. Baker and Wyeth Bair

Department of Biological Structure, Washington National Primate Research Center, University of Washington, Seattle, Washington 98195

Primate cortical area MT plays a central role in visual motion perception, but models of this area have largely overlooked the binocular integration of motion signals. Recent electrophysiological studies tested binocular integration in MT and found surprisingly that MT neurons lose their hallmark "pattern motion" selectivity when stimuli are presented dichoptically and that many neurons are selective for motion-in-depth (MID). By unifying these novel observations with insights from monocular, frontoparallel motion studies concurrently in a binocular MT motion model, we generated clear, testable predictions about the circuitry and mechanisms underlying visual motion processing. We built binocular models in which signals from left- and right-eye streams could be integrated at various stages from V1 to MT, attempting to create the simplest plausible circuits that accounted for the physiological range of pattern motion selectivity, that explained changes across this range for dichoptic stimulus presentation, and that spanned the spectrum of MID selectivity observed in MT. Our successful models predict that motion-opponent suppression is the key mechanism to account for the striking loss of pattern motion sensitivity with dichoptic plaids, that opponent suppression precedes binocular integration, and that opponent suppression will be stronger in inputs to pattern cells than to component cells. We also found an unexpected connection between circuits for pattern motion selectivity and MID selectivity, suggesting that these two separately studied phenomena could be related. These results also hold in models that include binocular disparity computations, providing a platform for future exploration of binocular response properties in MT.

Key words: binocular; model; motion-in-depth; MT; pattern motion

---

Significance Statement

The neural pathways underlying our sense of visual motion are among the most studied and well-understood parts of the primate cerebral cortex. Nevertheless, our understanding is incomplete because electrophysiological research has focused mainly on motion in the 2D frontoparallel plane, even though real-world motion often occurs in three dimensions, involving a change in distance from the viewer. Recent studies have revealed a specialization for sensing 3D motion in area MT, the cortical area most tightly linked to the processing and perception of visual motion. Our study provides the first model to explain how 3D motion sensitivity can arise in MT neurons and predicts how essential features of 2D motion integration may relate to 3D motion processing.

Introduction

Cortical area MT is a critical stage in the visual motion processing pathway and is known for the emergence of "pattern motion" sensitivity (Movshon et al., 1985; Rodman and Albright, 1989; Stoner and Albright, 1992), which refers to the ability of some MT neurons, known as pattern cells, to signal the direction of overall motion of a visual target. This ability is thought to arise by appro-

priate integration of direction-selective (DS) component motion signals that originate in V1, where "component" refers to local, narrowly oriented image features. Although many models have been proposed to explain how such integration takes place (Heeger, 1987; Grzywacz and Yuille, 1990; Nowlan and Sejnowski, 1995; Simoncelli and Heeger, 1998; Bowns, 2002; Perrone and Thiele, 2002; Pack et al., 2004; Perrone, 2006; Rust et al., 2006; Tsui et al., 2010), these models of motion processing in MT are not binocular, despite the fact that MT neurons are strongly driven when stimuli are presented to either eye (Zeki, 1974; Maunsell and Van Essen, 1983; Felleman and Kaas, 1984; DeAngelis and Uka, 2003).

The integration of motion signals across eyes has received far less attention than monocular motion tuning; however, several recent experimental studies have specifically examined integration of binocular motion signals in MT using simple sinusoidal grating stimuli. Tailby et al. (2010) found a striking reduction in

Table 1. Key parameter values for the component and pattern models presented in this study[a]

Parameter in V5_ME model	Component cell (Figs. 3, 4)	Pattern cell (Figs. 3, 4)	Component cell (Figs. 5, 7, 8, 10, 12)	Pattern cell (Figs. 5, 7–9, 11, 12)
Gabor filter SF f_r (cyc/deg)	3.6	2.4	2.4	2.4
Gabor filter TF f_t (cyc/s)	10	10	10	10
a1	0.5	1	0.5	1
a2	0	0	0	0
a3	0.4	0	0.4	0
rmsm_A	1	1	1	1
rmsm_B	0.7	0.5	0	0
W (left eye)	[0, −0.2, 0, 0.05, 0.05, −0.05, 1, 0, 0, 0, −0.05, −0.2]	[−1, −1, −1.1, −0.1, 0.25, 1, 0.9, 0.8, 1, 0.6, 0, −0.9]	[−0.1, −0.1, 0, 0, 0, 0, 1, 0, 0, 0, 0, −0.1]	[−1, −0.87, −0.5, 0, 0.5, 0.87, 1, 0.87, 0.5, 0, −0.5, −0.87]

[a]The parameters are as follows: f_r (spatial frequency) and f_t (temporal frequency) of the Gabor filter RFs for the V1 direction channels as in Equations 1 and 2 (see Materials and Methods); a1 (tuned), a2 (untuned), and a3 (non–response-dependent), constants that control the normalization of the raw V1 signals from Equations 6 and 9; rmsm_A and rmsm_B, which set the output nonlinearity for the raw MT output signal (A and B, respectively, in Eq. 15); W, the vector of full-strength MT input weights for the 12 V1 channels (Eq. 14).

pattern motion sensitivity in virtually all cells they recorded when the sinusoidal components of a plaid pattern (Adelson and Movshon, 1982) were presented to different eyes. Czuba et al. (2014) used a similar dichoptic grating protocol to test sensitivity to 3D motion, or motion-in-depth (MID) and found MT neurons that responded well to MID, including many cells tuned for opposite directions in each eye, as previously reported (Zeki, 1974; Albright et al., 1984; but see Maunsell and Van Essen, 1983). Another recent study of MT used a stimulus protocol that isolated motion and disparity cues and found that dichoptic motion signals alone can produce responses tuned for MID (Sanada and DeAngelis, 2014), underscoring the importance of understanding binocular integration of motion in MT. Pattern motion selectivity and tuning for MID have so far been studied separately; thus, binocular MT models are needed to allow constraints from both phenomena to be incorporated concurrently, leading to more general models and more powerful predictions.

To address this need, we built binocular, motion-sensitive models of V1 and MT processing stages within an image-computable framework that facilitated the testing of candidate circuits with diverse monocular and binocular stimulus paradigms. Here, we focus on two key issues: identifying possible circuit mechanisms underlying the loss of pattern motion sensitivity when plaid stimuli are presented dichoptically, and investigating possible links between the phenomena of component versus pattern-DS tuning and frontoparallel (FP) versus 3D motion tuning (3DT). We created component and pattern motion-sensitive MT units within our framework that successfully fit monocular MT data reported in published literature. We found that responses to dichoptic plaids can be explained by including monocular opponent suppression in the circuit, and we found clear differences in the ability of component and pattern models to fit the responses of the FP and 3DT cells. Our models identify mechanisms that can account for these dichoptic response properties of MT cells, and they offer several novel testable predictions regarding V1 to MT circuitry. Specifically, the models predict the following: (1) there is motion opponent suppression in the inputs to MT; (2) this opponent computation occurs before binocular integration; (3) there is a relationship between FP/3DT and component/pattern-DS selectivity in single MT neurons; and (4) 3D-biased motion tuning in MT can be derived from imbalanced ocular dominance and FP direction tuning preference. We also found that our major results hold when binocular disparity tuning is introduced in the model, providing a basis for future studies of how joint selectivity for motion and disparity might arise and shape responses in MT.

Table 2. Range of parameter values for the V1 and MT level ocular dominances and motion opponent suppression that were used in the figures

Figure	Panel	V1 motion opponency c_{opp}	V1 ocular dominance b	MT inhibitory weight strength k_{inh}	MT ocular imbalance A_R
5	A	0.5	1	0	1
	B	1	1	0.25	1
	C–H	0–1	1	0–1	1
6	B–G	0–1	0.5	0–1	1
8	A–D	0.5	1	0	1
	E–H	1	1	0.25	1
	I–L	0.5	1	0	1
	M–P	1	1	0.25	1
	Q–T: component	0.5	1	1	0.5
	Q–T: pattern	1	1	0.75	0.5
9	A	0.5	1	0–1	0.5–0.9
	B	1	1	0–1	0.5–0.9
	C	1	0.5–0.9	0–1	1
	D	1	0.7	0–1	0.5–0.9
11	B	0.5	1	0	0.7
12	A	0.5	0.5	0	1
	B	1	0.5	1	1
	C–H	0–1	0.5	0–1	1
	I	1	0.5–0.9	0–1	1
	J	1	0.7	0–1	0.5–0.9

Materials and Methods

We first describe the basic components of our binocular MT models and then describe the visual stimuli and data analyses below. All models were implemented within our custom software Working Model (WM) system (written in C) that is freely available online from www.iModel.org. All parameter files for the models, stimuli, and analyses to allow the simulations to be rerun are available at www.iModel.org. Key parameters for the versions of the binocular models presented in this paper are shown in Table 1. The values of parameters that were varied for the plots in the figures are shown in Table 2.

Model overview

The main features of the framework, as shown schematically in Figure 1, are as follows: Figure 1A, spatiotemporal filters representing DS V1 channels at the front end that process visual inputs for the left and right eyes; Figure 1B, response normalization of the V1 linear outputs; Figure 1C, V1-level opponent motion suppression; Figure 1D, a variable amount of binocular mixing of normalized V1 signals within a direction channel; Figure 1E, MT-level weighting and integration of the V1 direction channels within the left and right streams; Figure 1F, MT-level integration across the left and right streams; and Figure 1G, nonlinear transformation of the integrated signal into spiking output. The space-time motion energy filters at the front of each V1 channel (Fig. 1A) make the frame-
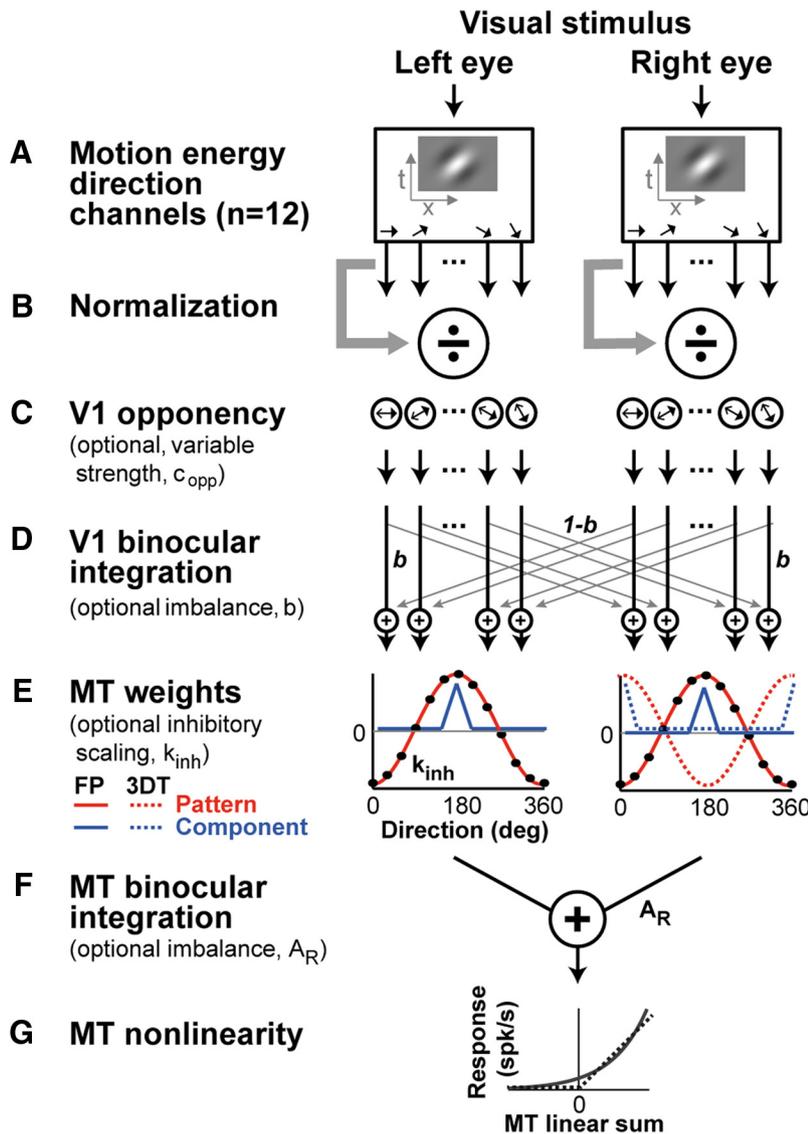
Figure 1. Schematic of the binocular MT model. For complete details, see Materials and Methods. ***A***, The front end of the model is composed of 24 spatiotemporal motion energy filters (12 in each eye) in 30° steps of preferred direction. The visual stimulus is convolved with these filters to generate input to the model. ***B***, Tuned and untuned normalization is applied to scale the raw V1 filter outputs. ***C***, Variable strength opponent motion suppression may be computed by subtracting a proportion (c_{opp}) of the signal from the antipreferred direction V1 channel. This step can occur before or after binocular integration. ***D***, Variable strength V1 binocular integration may be implemented by combining a weighted sum of signals from the left and right streams within a direction channel. For example, $b = 0.7$ produces left stream signals composed of 70% monocular left eye signal and 30% monocular right eye signal, and vice versa for the corresponding right stream. ***E***, Weighting of V1 inputs to MT. Pattern cell weights (red traces) are broadly tuned and have significant inhibitory weights, whereas component cells typically have only one strong positive weight and negligible inhibitory weights (blue traces). Weights are shifted 180° between FP motion tuned and 3DT models (see Fig. 7). ***F***, After weighting, signals are summed across both eyes and all direction channels. There is an optional additional weighting by a scale factor A_R that reduces the strength of all right streams equally before summation. ***G***, An output nonlinearity is applied to the summed signal, before optional scaling and offsetting, to produce spiking MT output. The nonlinearity can be an exponential (solid trace, after Rust et al., 2006) or simple rectification (dashed trace).

work image-computable, thereby allowing them to be tested with any achromatic dynamic visual stimulus. Our framework includes normalization (Fig. 1B) and V1 opponent motion suppression (Fig. 1C) stages, which have been proposed as being important for explaining MT neuronal responses in past studies (Rust et al., 2006, Tailby et al., 2010), and several novel features that are important to explore in models that include binocular integration: (1) ocular dominance of V1 inputs (Fig. 1D); (2) MT-level ocular dominance (Fig. 1F); and (3) MT-level dichoptic opponent motion suppression (Fig. 1F). Specific computational details for each of these stages are described in the following sections.

V1 direction channels

For the V1 channels (Fig. 1A), we chose a simple and robust model of direction selectivity based on the motion energy model (Adelson and Bergen, 1985). Our V1 computation consists of the following steps for each direction channel: (1) convolving the stimulus, a sequence of images over time, with two space-time oriented Gabor filters in quadrature; (2) squaring and summing the outputs to produce a directional motion energy signal; and (3) half-wave rectifying this opponent motion energy signal to produce a non-negative response. Specifically, the even and odd Gabor filters for the ith direction channel are defined as follows:

$$f^i_{even}(x, y, t)$$
$$= \cos(2\pi(f_r\boldsymbol{n} \cdot \boldsymbol{r} + f_t t))e^{(-|r|^2/(2s_r^2) - t^2/(2s_i^2))} \quad (1)$$

$$f^i_{odd}(x, y, t)$$
$$= \sin(2\pi(f_r\boldsymbol{n} \cdot \boldsymbol{r} + f_t t))e^{(-|r|^2/(2s_r^2) - t^2/(2s_i^2))} \quad (2)$$

for $i = 1, \ldots, D$, where D is the number of direction channels. For all models tested here, $D = 12$, such that the preferred directions for the filters, $d_i = (i - 1) * (2\pi/D)$, occur in 30° increments across direction channels. For convenience, we define the vectors $r = (x, y)$ and $n = (\cos(d_i), \sin(d_i))$. Other parameters are the filter spatial frequency (SF), f_r, the temporal frequency (TF), f_t, the spread (SD) of the spatial Gaussian, s_r, and temporal Gaussian, s_t. These control the direction (and orientation) bandwidth of the V1 channels. Our model is image-computable, meaning that any pair of time varying stimulus movies, $s_L(x, y, t)$ and $s_R(x, y, t)$ for the left and right eyes, respectively, are operated on as follows:

$$u^i_{L,o}(x, y, t) = f^i_{odd}(x, y, t) * s_L(x, y, t)$$
$$(3)$$

$$u^i_{L,e}(x, y, t) = f^i_{even}(x, y, t) * s_L(x, y, t)$$
$$(4)$$

and similarly for the right-eye stimulus. We then define the monocular motion energy for the ith direction channel as follows:

$$v^i_L(t) = u^i_{L,e}(x_c, y_c, t)^2 + u^i_{L,o}(x_c, y_c, t)^2$$
$$(5)$$

where we have taken the responses for the receptive field at the center (x_c, y_c) of the simulated patch of visual field. A similar definition holds for $v^i_R(t)$, except that the right eye stimulus is applied.

Normalization

The V1 normalization stage (Fig. 1B) in our model is a modified version of the equation for normalizing the raw V1 output, $L_n(t)$, from Rust et al. (2006) as follows:

$$r^i_L(t) = \frac{v^i_L(t)}{a_1 v^i_L(t) + \dfrac{a_2}{D}\displaystyle\sum_{k=1}^{D} v^k_L(t) + a_3} \quad (6)$$

where a_1 controls tuned normalization (i.e., normalization that is limited to the current direction channel), a_2 controls untuned normalization (normalization that spans all direction channels), and a_3 is a constant that we have retained to match output firing rates when fitting to experimental data. In place of our a_3, the Rust et al. (2006) formulation contained a stimulus-contrast-dependent term, $a_3\bar{L}$ (their notation), but this presents a problem for an image-computable model because $\bar{L}$, the mean squared contrast of the stimulus, is both not defined for general image sequences and not knowable in advance (i.e., a causal model cannot know the mean over future time of the contrast of an ongoing stimulus). To allow the normalization to be computed when arbitrary stimuli are presented to the model, without a priori knowledge of the mean squared contrast, we replace this value with a constant term a_3 that we found allowed us to reproduce the fits of the Rust et al. (2006) monocularly obtained data (see Fig. 3). In practice, a_3 can be set to a very small number, or even 0, and still reproduce key properties of MT responses (Jazayeri et al., 2012; Patterson et al., 2014).

V1 motion opponency
We implemented motion opponent suppression in V1 (Fig. 1C) by subtracting a weighted amount, c_{opp}, of the opponent direction V1 channel (after the normalization stage), and we then rectified the resulting signal as follows:

$$o_L^i(t) = \lfloor r_L^i(t) - c_{opp}\, r_L^j(t) \rfloor \qquad (7)$$

where j is the index of the direction channel that is 180° opposite to i. This could occur either before V1 binocular mixing, as shown in Figure 1, or afterward, as explored later in Results. This allows us to directly implement the proposal of Tailby et al. (2010) that there is a monocular component of motion opponency that occurs at the level of V1 (for a review of experimental evidence, see Discussion).

Binocular integration
Our model framework has two possible sites for binocular combination: one in V1 and the other in MT. In V1, an optional, weighted binocular mixing can occur within each direction channel (Fig. 1D). The ocular dominance of binocular V1 channels is set by the weight parameter b, with a value of 0.5 indicating 50/50 left/right eye mixing within a V1 direction channel. The mixed signal, m, is as follows:

$$m_L^i(t) = b\, o_L^i(t) + (1 - b)o_R^i(t) \qquad (8)$$

where b takes values from 0.5 to 1.0. We will continue to refer to the two parallel streams in Figure 1 originating from the left and right eyes as the left and right streams, regardless of the actual ocular dominance (OD) (Hubel and Wiesel, 1962) of the signals within them. In MT, an imbalance in OD can be imposed by setting a parameter, A_R, that scales the signals in the right stream (see Eq. 13 below), as shown in Figure 1F. A value of $A_R = 1.0$ indicates full-strength right stream weights. Unless otherwise stated, models have the default configuration of monocular V1 channels ($b = 1$) and full-strength MT weights in the right stream ($A_R = 1.0$).

Binocular disparity
We built a version of the model that had a disparity computation at the V1 level to test whether our results, which relate primarily to the binocular integration of motion signals, also hold when a plausible disparity computation occurs upstream. The details of the disparity model are as follows.

The even and odd linear filter signals, $u_{L,o}^i$ and $u_{L,e}^i$ (Eqs. 3 and 4), in addition to being combined to form a motion energy signal v_L^i (as in Eq. 5 above), are also maintained in separate pathways through the normalization and opponency stages (Fig. 1B,C). This is done because an energy computation is now required at the binocular integration stage in accordance with the binocular disparity energy (BDE) model (Ohzawa et al., 1990). Figure 2A–D shows the circuitry specific to the disparity model, which replaces Stages B–D in Figure 1. In our simulations, the left and right eye filters were set to have 0 phase disparity (tuned excitatory-type disparity tuning). The even and odd pathways are marked in black and green, respectively. In Figure 2A,
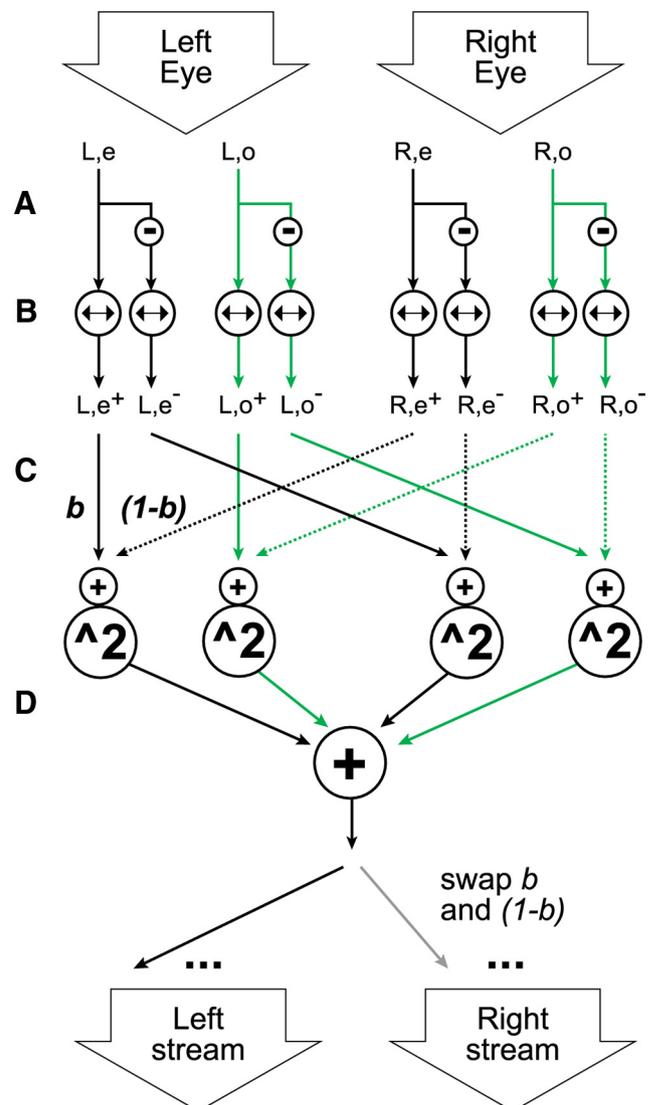
Figure 2. Overview of the circuitry for the binocular disparity computation. *A*, Normalized even and odd linear filter signals from the left and right eyes are duplicated and negated before motion opponency to allow the disparity computation. *B*, The motion opponent signals are subtracted from each subchannel and the resulting signals rectified (Eq. 10). *C*, The rectified signals from left and right eyes are combined in a weighted sum and the result squared (Eq. 11). *D*, The summed, squared outputs are summed in a final step to produce binocular disparity energy signal (Eq. 10).

the even and odd filter outputs are copied and negated, consistent with the BDE model (e.g., Ohzawa et al., 1990, their Fig. 3B; Read et al., 2002, their Figs. 3, 6). In Figure 2B, an opponent signal (described below) is subtracted, the resulting signal is rectified, and then the left and right eye signals are added (Fig. 2C), with the left eye signals given a weight, b, whereas the right eye signals are given a weight $(1 - b)$. Each of these sums is then squared (Fig. 2D) to form the four main terms of the BDE model (Ohzawa et al., 1990; Read et al., 2002). The diagram shows the computation for a single direction channel computed for the left stream (Fig. 2D, bottom left, black arrow). For the right stream direction channels, the computation is the same, except that the weights b and $(1 - b)$ are swapped with respect to the left and right streams.

Details of the normalization, opponency, and BDE computation for the disparity model are as follows. The normalized responses (Eq. 6) are computed as follows (equation for the even, left eye signal is shown; the other channels are computed similarly):
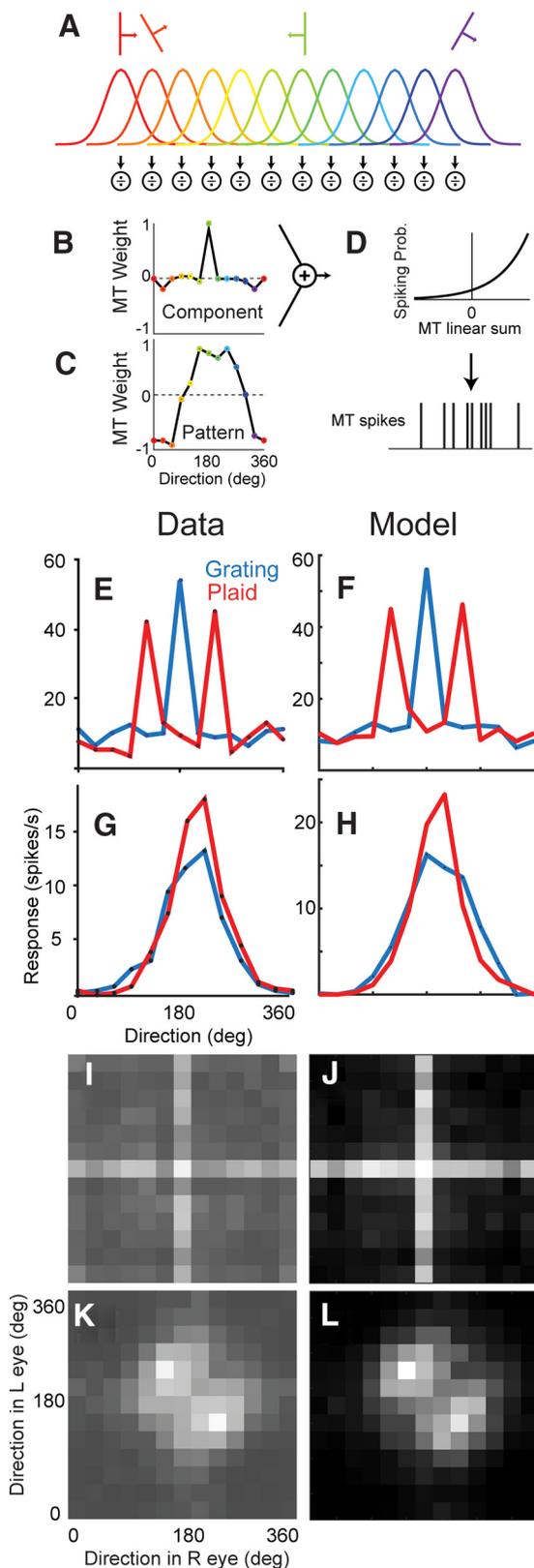
Figure 3. Monocular fits to experimental data on component and pattern cells. Example component and pattern cells in our binocular model using the fits of Rust et al. (2006). Parameters for left and right eyes are identical. Responses shown are for monocular left eye stimulation only. *A*, The monocular model of Rust et al. (2006) used 12 V1 channels modeled as tuning curves, with the output from each channel normalized. These 12 signals were weighted to fit a component cell (*B*, Rust et al., 2006, their Fig. 4*A*) and pattern cell (*C*, Rust et al., 2006, their Fig. 4*E*). The summed output was then passed through a nonlinearity and the signal used to produce spikes (*D*). *E*, Tuning curves for the component cell reproduced from Rust et al. (2006):

$$r^i_{L,e}(t) = \frac{u^i_{L,e}(t)}{a_1 \sqrt{v^i_L(t)} + \frac{a_2}{D} \sum_{k=1}^{D} v^k_L(t) + a_3} \tag{9}$$

The energy-based self-normalization is favorable because it cannot be negative and thus cannot pull the denominator near or across 0. It is also less sensitive to spatial phase, being more complex-cell-like (Adelson and Bergen, 1985), thereby providing a more stable signal. The square root is taken to keep the amplitude of the self-normalization term on the same order as the numerator. The opponency calculation (Eq. 7 above) is modified as follows:

$$o^i_{L,e\pm}(t) = \lfloor \pm r^i_{L,e+}(t) - c_{opp} \sqrt{v^i_L(t)} \rfloor \tag{10}$$

Here positive and negative signals are now computed. The square root of the energy signal is subtracted to keep the two signals being differenced on the same magnitude scale. The resulting matched signals from left and right eyes are then summed and squared, with the optional weighting between left and right eyes applied (here shown for the even, positive channels):

$$g_{e+} = (b\, o^i_{L,e+}(t) + (1-b) o^i_{R,e+}(t))^2 \tag{11}$$

The resulting four squared terms are all summed together for the final BDE computation as follows:

$$m^i_L(t) = g_{e+} + g_{o+} + g_{e-} + g_{o-} \tag{12}$$

Weight distribution of V1 inputs to MT

At the first MT stage, the signals coming from V1 are multiplied by sets of weights, w^i_L and w^i_R, for the left and right eye channels, which we will refer to as "MT weights" (Fig. 1*E*). Thus, the "linear" MT response (i.e., before the output nonlinearity) is as follows:

$$MT(t) = \sum_{i=1}^{D} w^i_L m^i_L(t) + A_R \sum_{i=1}^{D} w^i_R m^i_R(t) \tag{13}$$

Component and pattern motion selective responses are primarily shaped by differences in the distribution of these MT weights over the V1 direction channels. Pattern cells are fit by a broad peak of excitatory (i.e., positive) weights and significant inhibitory (i.e., negative) weights (e.g., Fig. 1*E*, red curve), whereas component cells have a very narrow excitatory peak and negligible inhibitory weights (Fig. 1*E*, blue curve). We implemented the Rust et al. (2006) fits for example pattern and component cells (Figs. 3, 4; Table 1). To reproduce the simulations of Tailby et al. (2010) in Figure 4*C*, *D*, we also implemented a version of the model in which any negative MT weights are reduced by 40% (Tailby et al., 2010) (see Results) when the stimulus is presented dichoptically.

In addition to the models fit to specific example MT cells from the literature, we built canonical models that captured the main features of pattern and component cells using idealized MT weight distributions. The canonical pattern cell model had a cosine weight distribution with weights varying between 1 and −1, and the canonical component model had one strong excitatory weight with amplitude 1, and inhibitory weights in the −30°, 0°, and 30° direction channels set to −0.1 (Table 1). These distributions embody the main features of strongly component and pattern cells as described in the previous paragraph but are not tied to a specific fit to a neuron, making our results more generalizable. The weights shown in Table 1 are the full-strength (100%) weights, W^i, and were varied by scaling all the negative weights by a constant fraction, k_{inh}, to produce the final weights used in Equation 13, as follows:

←

their Fig. 4*A* for single gratings (blue trace) and 120° plaids (red). *F*, Tuning curves for our model component cell using the weights in *B*. *G*, Tuning curves reproduced from the pattern cell in Rust et al. (2006, their Fig. 4*E*). *H*, Tuning curves for our model pattern cell using weights shown in *C*. *I*, 2D map of responses to plaids of all directions (in 30° steps) for the component cell in *E*. *J*, Responses to plaids from our model component cell in *F*. *K*, Plaid matrix responses for the pattern cell in *G*. *L*, Responses to plaids from our model pattern cell shown in *H*.
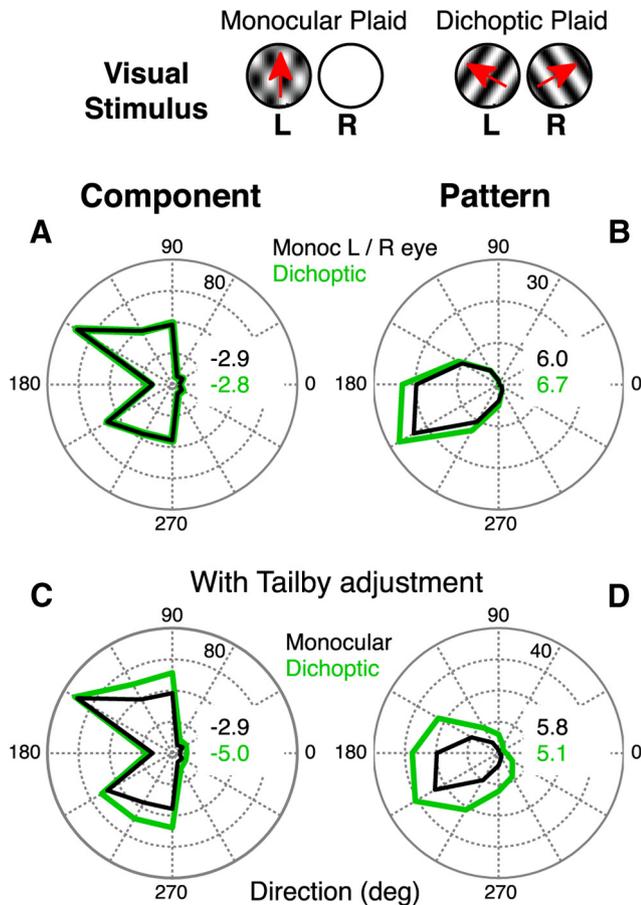
Figure 4. Reduced pattern motion sensitivity with dichoptic plaids in the binocular models. Direction tuning for monocular and dichoptic plaids as in the protocol of Tailby et al. (2010). Pattern index values for the monocular and dichoptic conditions are shown as insets on all plots. **A**, Direction tuning curves for the component cell from Figure 3F, J for monocular 120° plaids (black trace) and dichoptic plaids (green trace). **B**, Monocular and dichoptic plaid direction tuning curves for the pattern cell in Figure 3H, L. **C**, Component cell tuning curves where inhibition is reduced to 40% of monocular strength when the component gratings of the plaid are presented dichoptically, following the method of Tailby et al. (2010). **D**, Tuning curves for the pattern cell model with inhibition reduced to 40% of monocular strength with dichoptic presentation.

$$w^i = \begin{cases} k_{inh} W^i & \text{if } W^i < 0, \\ W^i & \text{otherwise} \end{cases} \qquad (14)$$

The default value of k_{inh} is 1. The weight distributions were either identical in each eye (FP motion tuned, FP models) or shifted 180° in the right eye (3DT models).

MT output nonlinearity
The final processing stage is an output nonlinearity (Fig. 1G). We implemented both the expansive output nonlinearity of Rust et al. (2006) as follows:

$$MT^*(t) = Ae^{(B*MT(t))} \qquad (15)$$

and a simple half-wave rectification nonlinearity. An optional linear scaling and offsetting of this signal are performed, and the result is passed through a Poisson spike generator to produce spiking output. For the model MT cells in Figures 3 and 4, we applied the expansive nonlinearity to fit the spiking data from the example cells of Rust et al. (2006). For the canonical versions of the models with idealized weight distributions, we used a rectification nonlinearity, to better fit results of the dichoptic plaid simulations (see Results).

Visual stimulus
We characterized the models with various combinations of sinusoidal grating stimuli that reproduced key monocular and dichoptic experimental paradigms used in the literature. Many of our stimuli were plaids, meaning that they consisted of the sum of two sinusoidal grating components, where the orientation differed across components. The main protocols are as follows:

Single plaid direction tuning. In this series of stimuli, the overall motion direction of a plaid pattern was varied to generate a direction tuning curve. The plaid had a single, fixed difference in the direction of its two component gratings, namely, 120°. When presented monocularly (both components to the same eye), this is a classical stimulus used to classify pattern and component cell types in MT (Movshon et al., 1985). When presented dichoptically (one component grating to the left eye and the other component grating to the right eye), this is the stimulus used by Tailby et al. (2010) to demonstrate that MT pattern motion sensitivity relies on monocular computations. An important reference for this protocol (see Data analysis) is the single grating direction tuning protocol, which can be achieved by setting the contrast of one of the plaid components to 0.

Plaid matrix protocol. This set of stimuli encompasses a variety of plaid patterns by independently varying the direction of motion of the two component gratings across 12 values between 0° and 360°. This protocol was used to generate a 2D matrix of responses to plaids that differs characteristically between component and pattern MT neurons, as shown in Rust et al. (2006, their Fig. 8). The single plaid protocol (above) corresponds to a particular diagonal within the plaid matrix here.

Interocular velocity difference (IOVD) protocol. Devised by Czuba et al. (2014), this is a dichoptic plaid stimulus where the directions of the component gratings in each eye are separated by either 0° (binocular matched motion, B_{Same}) or 180° (binocular opponent motion, B_{Opp}). The B_{Same} condition corresponds to typical FP motion, whereas the B_{Opp} condition produces the appearance of 3D motion, either directly toward or away from the observer, which arises when motion has equal speed and opposite direction in the two eyes. The temporal frequency of the gratings was also varied independently between eyes at values of 2.4, 4.8, and 18 Hz to generate oblique 3D motion directions (see Results; see Fig. 11A).

The component sinusoidal gratings were presented at 50% contrast for the single 120° plaids and the plaid matrix protocol but were presented at 100% contrast in the IOVD protocol. Stimuli were presented at the optimal SF, 2.4 cyc/deg, for the model units.

Data analysis
To quantitatively assess the degree to which DS units display pattern versus component motion selectivity, we computed pattern index values as described by Tailby et al. (2010). Briefly, we used the single grating direction tuning curve to make predictions of the tuning curve responses to plaids for idealized pattern- and component-selective cells. The component prediction is the sum of two appropriately shifted single-grating tuning curves, whereas the pattern prediction is simply the single-grating tuning curve. For dichoptic plaids, predictions were based on monocular single-grating tuning curves. We computed partial correlations R_c and R_p of the actual responses against the predicted tuning curves, as follows:

$$R_c = \frac{(r_c - r_p r_{pc})}{\sqrt{(1 - r_p^2)(1 - r_{pc}^2)}} \qquad (16)$$

$$R_p = \frac{(r_p - r_c r_{pc})}{\sqrt{(1 - r_c^2)(1 - r_{pc}^2)}} \qquad (17)$$

where r_c is the correlation of the data with the component prediction, r_p is the correlation of the data with the pattern prediction, and r_{pc} is the correlation of the two predictions. These were converted into Z-scores using Fisher's r-to-Z transformation as in Smith et al. (2005) as follows:

$$Z_p = \frac{0.5 \ln\left(\frac{1 + R_p}{1 + R_c}\right)}{\sqrt{\frac{1}{df}}} \qquad (18)$$

where df is the number of degrees of freedom, equal to the number of values in the tuning curve minus 3.

We computed a direction selectivity index (DSI) and monocularity index for our model units as described by Czuba et al. (2014) as follows:

$$DSI = \frac{\left| \sum_{n=1}^{N} R_n e^{i\theta_n} \right|}{\sum_{n=1}^{N} R_n} \qquad (19)$$

$$MI = \frac{|R_{max} - L_{max}|}{R_{max} + L_{max}} \qquad (20)$$

where θ_n and R_n are the direction of motion and response strength for each of the $N = 12$ directions and R_{max} and L_{max} are the maximum responses in the left and right eye, respectively, across all the monocular stimulus conditions. We also computed 3D directions ϕ_n shown in Figure 9A following Czuba et al. (2010, their Eq. 3), as follows:

$$\phi_n = \arctan(V_{REn}, V_{LEn}) - \frac{\pi}{4} \qquad (21)$$

where V_{REn} and V_{LEn} are the stimulus velocities in the right and left eyes, respectively.

Results

A framework for image-computable binocular MT models

We constructed a binocular multistage modeling framework (Fig. 1) to unify concepts from past studies of pattern motion selectivity and to account for the experimental results of recent studies of MT that presented different motion stimuli to each eye. Key elements from past models include the following: front-end DS filters, response normalization, motion opponency, and weighted integration across multiple direction channels. We introduced a parallel left- and right-eye dual stream structure and allowed the flexibility to vary the mixing of left and right eye signals at various points along the streams and the ability to vary the order of operations for opponency and binocular integration.

Our observations and insights from the model are organized below as follows: (1) We first develop image-computable pattern and component cell models that are consistent with past monocular studies, and we extend these in the simplest way to operate on binocular stimuli. (2) We then test and refine the binocular models to account for the loss of pattern motion sensitivity when moving plaid gratings are presented dichoptically. (3) We further test and refine the models to account for 3D motion sensitivity in area MT, making predictions to guide future experimental studies to examine whether there is a relationship between pattern motion and 3D motion sensitivity. (4) Finally, we show our results hold when a canonical binocular disparity computation is included in the models.

Monocular characterization of pattern motion sensitivity in the binocular MT model

Our first goal was to reproduce within our framework the widely studied selectivity of MT cells for pattern motion. MT cells have traditionally been ranked along the continuum from component-DS to pattern-DS based on their responses to plaid stimuli: component cells respond best when either constituent grating of the plaid matches the preferred direction of the cell (measured with single gratings), whereas pattern cells respond best when the overall pattern moves in the cell's preferred direction (Movshon et al., 1985; Rodman and Albright, 1989; Stoner and Albright, 1992).

A thorough characterization of this behavior was performed by Rust et al. (2006) who used a full range of directions for two-component and multicomponent sinusoidal stimuli to monocularly drive MT units that spanned the spectrum from component to pattern cell behavior. They fit the resulting responses with a monocular cascade model (Fig. 3A–D). In the first stage of the cascade model, the responses of 12 DS V1 channels were modeled as direction tuning curves spaced evenly over the full 360° range (Fig. 3A). The responses were then divisively normalized by the sum of two components: a direction-tuned self-normalization term and an untuned term that summed signals across all directions. The model MT unit computed a weighted sum of these normalized V1 outputs (weights for component and pattern example cells are shown in Fig. 3B,C), and the summed signal was passed through an output nonlinearity to produce a spiking response (Fig. 3D). These last three steps correspond to the steps shown in Figure 1B, E, G in our framework, respectively. Rust et al. (2006) found that a primary distinction between fits to component and pattern MT cells was the distribution of weights from V1 to MT: pattern cell weights were broadly direction tuned, with strong, oppositely tuned inhibitory weights (Fig. 3C), whereas component cells had narrowly tuned excitatory weights and sparse, low-strength inhibitory weights (Fig. 3B).

To build representative binocular pattern and component MT units in our framework, we fit our model to example cells for which Rust et al. (2006) presented extensive data. Taking a minimalist approach to building binocular models, we set identical parameter values in both the left and right streams and simply summed the signals from both streams (Fig. 1F) before applying the MT output nonlinearity to generate spikes. We extracted the key parameters (V1 tuning curve bandwidth, normalization strengths, MT weight distributions, and MT nonlinearity) from the data presented in Rust et al. (2006) by inspection of their Figure 6A, E (see Materials and Methods); we chose as representative units of their most component-like and most pattern-like example cells (their Fig. 4A and Fig. 4E, respectively). We tested our binocular models using monocularly presented stimuli to be consistent with experimental conditions. A side-by-side comparison of the recorded MT data (plots reproduced from Rust et al., 2006) and our fits are shown in Figure 3E–L, where we have adopted their plotting conventions to facilitate comparison. The classical single grating and single plaid tuning curves for the model component cell (Fig. 3F) and pattern cell (Fig. 3H) closely match the recorded data (Fig. 3E, G). When tested with a full matrix of plaid combinations (Fig. 3J, L), we also obtained a good qualitative match to the experimental results (Fig. 3I, K). We have also implemented models reproducing data from the other three example cells of Rust et al. (2006, their Fig. 4B–D; see supplemental material).

Reduction of pattern motion sensitivity with dichoptic plaids

Having created image-computable, binocular models that accounted for the responses of typical pattern-DS and component-DS cells to monocularly presented stimuli, we next examined whether these models could account for the robust loss of pattern-DS behavior when the plaid stimuli were presented dichoptically rather than monocularly (Tailby et al., 2010). This experimental observation provides an important constraint on

the integration of motion information from V1 to MT, indicating that monocular mechanisms are critical for pattern motion sensitivity. To explain their dichoptic plaid results, Tailby et al. (2010) proposed an adjustment to the monocular cascade model of Rust et al. (2006) but did not provide a unified model to account for both monocular and binocular responses. Our goal in this section is to verify that the adjustment proposed by Tailby et al. (2010) works within our binocular image-computable framework. Thus, we first test our simple binocular models with the dichoptic stimuli of Tailby et al. (2010) to observe the effect of dichoptic presentation on pattern sensitivity in our component and pattern model units, and we then implement and test the adjustment by Tailby et al. (2010). In the next section, we apply the intuition gained here to create a unified model to account for pattern selectivity under both monocular and dichoptic conditions.

The first experiment of Tailby et al. (2010) was to measure the pattern indices of MT neurons monocularly in both eyes, from cells spanning the continuum from strongly pattern-DS to strongly component-DS. They found that the measured pattern indices (PIs) were similar across the two eyes. This holds by construction in our model units because the parameters for the left and right streams were identical: the PIs for left and right eye presentation were both −2.9 for the component cell (Fig. 4A, black trace, tuning curves are identical) and both 6.0 for the pattern cell (Fig. 4B, black trace). We consider PI < −1.28 to be indicative of a component cell and PI > 1.28 to be indicative of a pattern cell with values in between unclassified, as in previous studies (Smith et al., 2005; Tailby et al., 2010).

The main finding of Tailby et al. (2010) was that PI values decreased for essentially all MT cells when plaids were presented dichoptically, with one component grating presented to the left eye while the other component grating was simultaneously presented to the right eye. In our model units, the PIs changed very little with dichoptic presentation, with PIs of −2.8 for the component unit (Fig. 4A, green trace) and 6.7 for the pattern unit (Fig. 4B, green trace). The similar PI values for monocular and dichoptic presentation are largely a consequence of the linear combination of the signals in the left and right streams (Fig. 1F), which occurs before the MT output nonlinearity and the identical MT weight distributions across the eyes. Intuitively, splitting the plaid across eyes should produce about the same amount of drive as a monocular plaid because the channels in the left and right eyes have the same weights and respond equally strongly to the component gratings. It is important to note that the results in Figure 4A, B indicate that normalization, raised as a possible key element to the monocular dependence of pattern selectivity by Tailby et al. (2010), do not play a critical role in our models, which have tuned and untuned normalization in accordance with the findings and specific fits of Rust et al. (2006). Specifically, because the normalizations (tuned and untuned) act monocularly, splitting the grating across the two eyes alters the normalization signals within each eye; however, this does not end up having any substantial effect on the pattern index. However, the slight differences in monocular and dichoptic firing rates observed in the pattern model (Fig. 4B, black vs green curves) are a result of the normalization, which is applied to the V1 channels at a monocular stage (Fig. 1B), and this difference in the tuning curves is eliminated if normalization is removed (data not shown). Thus, the simplest binocular extension of the monocular pattern and component models does not explain the experimental observations.

In their study, Tailby et al. (2010) proposed a potential modification to the monocular cascade model by which their dichoptic plaid results could be achieved. They reasoned that the reduction in PI under dichoptic viewing conditions implied that at least some of the mechanisms underlying pattern selectivity must act monocularly. The mechanisms previously identified as essential for pattern selectivity (Rust et al., 2006) were contrast-dependent normalization, integration of V1 inputs over a wide range of preferred directions, and strong opponent motion suppression. Focusing on motion opponency, Tailby et al. (2010) compared plaid direction tuning curves from the monocular model of Rust et al. (2006) with curves from an altered model, used only for dichoptic stimulation, in which the MT inhibitory weights (i.e., the negative weights in Fig. 3B,C) were decreased to 40% of their original strength. This reduction was intended to simulate a decrease in the inhibitory input, relative to that generated by the monocular plaid, when one component of the plaid is shifted to the other eye. They found that they were able to reproduce the reduction in PI from monocular to dichoptic presentation in simulated pattern-DS tuning curves using these two models. We refer to this modified model as the "Tailby adjustment," noting that they compared the monocular stimulus presented to the standard cascade model, against the dichoptic stimulus presented to the modified model.

We verified the validity of the Tailby adjustment by implementing it in both our component and pattern model units. The component unit (Fig. 4C) produced PIs of −2.9 for the standard model and −5.0 for the reduced-inhibition model, whereas the pattern unit (Fig. 4D) generated PIs of 5.8 and 5.1 for the standard and modified models, respectively. Both of these reductions in PI fall within the range of reductions observed by Tailby et al. (2010, their Fig. 2C). These results validate the insight of Tailby et al. (2010), that reducing the inhibition provided by the negative MT weights for the dichoptic plaid, but not for the monocular plaid, could lead to appropriate models for pattern and component cells.

A single circuit model for representing monocular and dichoptic plaid responses

The Tailby adjustment provides a proof of principle that reducing the negative MT weights during dichoptic stimulation can reduce the PI, but it does not provide a unified, plausible circuit model that simultaneously works for both the monocular and dichoptic stimuli. In addition, Tailby et al. (2010) also reported that pattern cells commonly showed larger decreases in PI than we saw in our pattern cell using the Tailby adjustment, with 11 of their 18 pattern cells recorded no longer being classified as pattern selective with dichoptic presentation (Tailby et al. (2010, their Fig. 2C; under their criterion, a cell is classified as pattern if the pattern correlation coefficient significantly exceeds either 0 or the component correlation, whichever is largest). The key mechanism identified by Tailby et al. (2010) was a reduction, but not elimination, in inhibitory strength with dichoptic presentation, which suggests that both monocular and interocular sources of inhibition are at play. We implemented this in our model by: (1) introducing motion-opponent suppression between V1 channels with 180° differences in direction preference, and (2) selectively reducing the strength of the inhibitory weights from V1 to MT. The strength of the motion-opponent suppression between V1 channels (Fig. 1C) was set by a parameter c_{opp} (e.g., $c_{opp} = 0.5$ means the normalized V1 signal from the opponent motion channel is scaled by 0.5 before being subtracted). To vary the strength of MT inhibitory weights, we introduced a
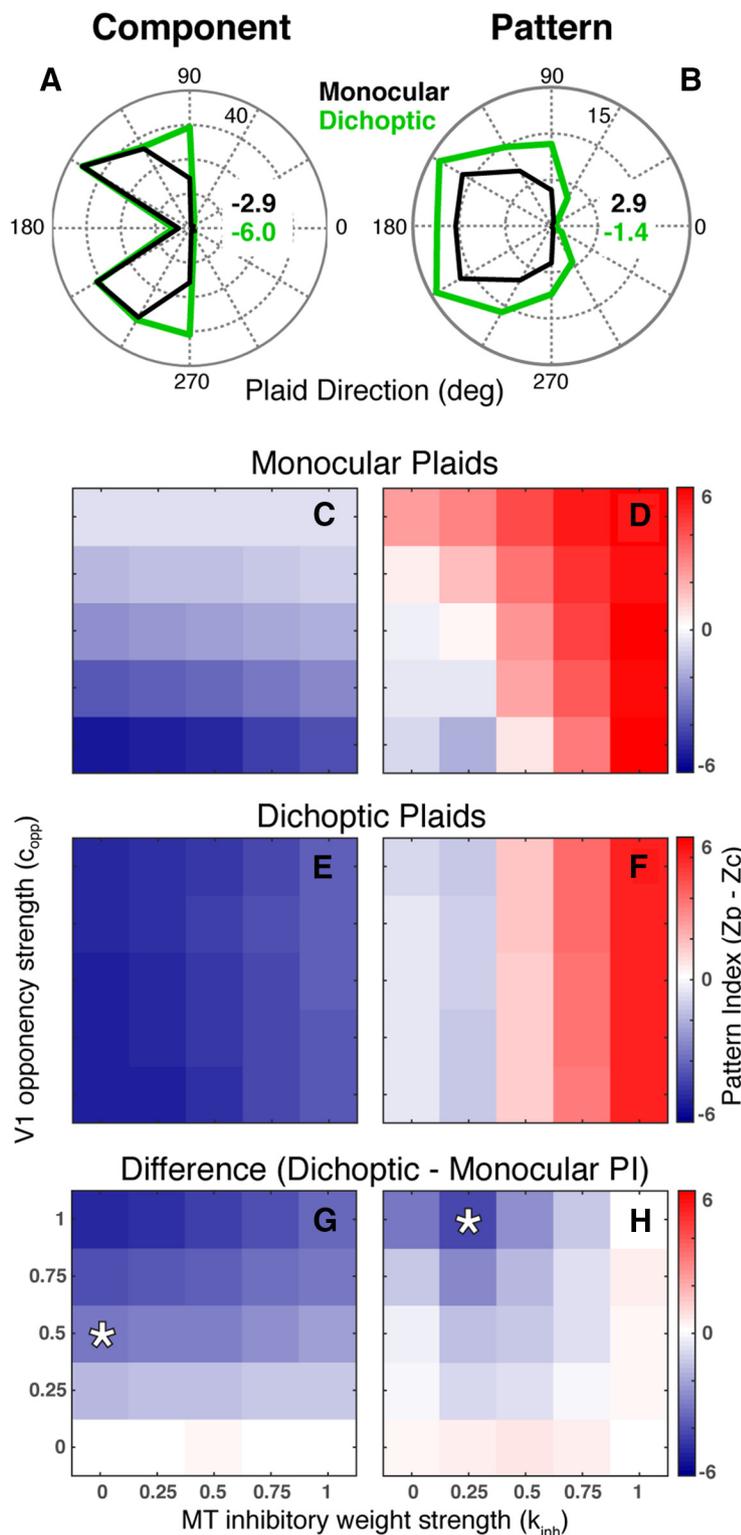
Figure 5. Effect of monocular and binocular opponent motion suppression on loss of pattern sensitivity with dichoptic plaids. ***A***, Plaid direction tuning curves for a model component cell with 50% V1 opponent motion suppression strength and no MT inhibitory weights. Inset, Pattern index values for the two curves. ***B***, Plaid tuning curves for a pattern cell with 100% V1 opponent motion suppression strength and inhibitory weights reduced to 25% of the full-strength weights. Inset, PI values for the two curves. ***C***, Dependence of the monocular PI on amount of V1 opponency and MT inhibitory weight strength in the component cell model. Blue areas of the plot represent negative PI values. ***D***, Monocular PI values in the pattern cell model as V1 opponency and MT inhibitory weight strength are varied. Blue regions represent negative PI values. Red regions of the plot represent positive PI values. ***E***, Dichoptic PI values as opponency and inhibitory strength vary in the component cell model. ***F***, Dichoptic PI values as opponency and inhibitory strength vary in the pattern model. ***G***, The difference between dichoptic and monocular PI as V1 opponency and MT inhibitory strength are varied in the component model. Blue regions represent a drop in PI. White represents no change. Red regions represent an increase in PI from

scaling factor k_{inh} that was applied to only the negative values in the MT weight distribution (Eq. 14). With dichoptic presentation, the V1 motion-opponent suppression between signals driven by the two component gratings is eliminated, reducing the total suppression generated by the plaid stimulus relative to monocular presentation, whereas the suppressive component generated by the MT inhibition is the same in both cases because the MT weight distribution is identical in both left and right eye streams.

For simplicity and generality in these tests, we built canonical pattern and component cell models that have idealized weight distributions. The component cell model MT weights consisted of a δ function at 180° (Fig. 1E, solid blue lines) with scalable opponent MT inhibitory weights across three antipreferred channels (−30°, 0°, and 30°; data not shown in Fig. 1E; see Materials and Methods). The pattern cell model had cosine weights (Fig. 1E, solid red lines), also with adjustable scaling for the inhibitory weights. We used simple rectification for the MT output nonlinearity (Fig. 1G, dashed line). We retained the V1 filtering and normalization parameters from our Rust et al. (2006) example component and pattern units. Introducing the canonical models is important because they embody the key factors underlying the differences between pattern and component cells, as shown by Rust et al. (2006): those being (1) a broad weight distribution for pattern versus a narrow distribution for component cells, and (2) strong opponent inhibition (negative weights) for pattern versus negligible inhibition for component cells. The canonical models retain strong pattern and component tuning while ensuring that our results do not depend on specific details of the weight distributions of the two example cells (Fig. 3B,C). Nevertheless, the following results also hold for the particular Rust et al. (2006) example cells we presented in Figures 3 and 4. The canonical models are used in all subsequent sections for testing plausible configurations of the circuitry for dichoptic and 3D motion sensitivity.

Figure 5A shows the plaid direction tuning curves for the canonical component-DS model that has both V1 opp-

←

onent motion suppression and no MT inhibitory weights (c_{opp} = 0.5 and k_{inh} = 0.0). When tested with monocular plaids (black curve), the model component cell has a PI value of −2.9. When tested with dichoptic plaids (green curve), the PI dropped to −6.0. A similar trend holds for the representative model pattern unit with V1 opponency and reduced inhibitory weights (Fig. 5B; c_{opp} = 1.0 and k_{inh} = 0.25), where the PI dropped from 2.9 to −1.4 (compare black and green curves). These decreases in PI, −3.1 and −4.3 for the component and pattern models, respectively, are both within one-half SD of the mean values found in the population data (−2.0, SD 2.2 for component cells, −5.2, SD 3.4 for pattern cells) of Tailby et al. (2010). In contrast to the results above for the Tailby adjustment (Fig. 4C,D), in these simulations the model parameters are the same for both monocular and dichoptic stimulation: for the component cell, 50% V1 opponency and no MT inhibitory weights; and for the pattern cell, 100% V1 opponency and 25% strength MT inhibitory weights. The change in PI with dichoptic stimulation reflects a loss of monocular suppression when one component of the plaid is moved to the other eye, which is not fully recovered by the additional reduced dichoptic MT inhibition.

To explore how the decrease in PI values depended on the amounts of monocular and dichoptic suppression, we varied the strength of V1 opponency (c_{opp}) and the strength of the MT inhibitory weights (k_{inh}) in the models (Fig. 5C–H). Full MT inhibitory weight strength (k_{inh} = 1.0) is defined for both models as shown in Table 1. For both the component (Fig. 5C) and pattern models (Fig. 5D), monocular PI tended to increase with both the strength of V1 opponency and MT inhibition. For the component model (Fig. 5C), the PI values ranged from strongly negative (consistent with a component cell) for no suppression to near 0 (consistent with an unclassified cell) for full monocular and dichoptic suppression. In contrast, the pattern model (Fig. 5D) generated PIs that ranged from near 0 with no suppression to large positive values associated with strong pattern cells for full monocular (V1) and dichoptic (MT-level) suppression. This extends the observation of Rust et al. (2006) that antipreferred direction suppression is key to generating responses to pattern motion, by showing that this suppression can be implemented both monocularly in V1 and binocularly at later stages. In contrast, with dichoptic plaid presentation, we found that in both component (Fig. 5E) and pattern (Fig. 5F) models, PI no longer varies with strength of V1 opponency (along the vertical axis) because this opponency is not engaged by the dichoptic stimulus.

Using these data, we plotted the change in PI from monocular to dichoptic presentation as a function of monocular and dichoptic suppression in both models (Fig. 5G,H). The parameter values used to generate the plots in Figure 5A, B are indicated by asterisks in Figure 5G, H. Both models showed significant regions of the parameter space that produced the decrease in PI observed by Tailby et al. (2010), whereas increases in PI with dichoptic presentation were negligible. Critically, in both models, there was no decrease in PI when V1 opponency strength was 0, indicating that V1 opponent suppression is a key factor in producing the loss of pattern sensitivity. For the component model (Fig. 5G), eliminating dichoptic MT inhibition resulted in a strongly component-like response and the largest drops in PI with the dichoptic stimulus. However, the strongest level of V1 opponency in the component model led to monocular PI values (Fig. 5C, top row) that were too high for component cells. Thus, for the representative component model (Fig. 5A), we chose a V1 opponency strength of 50%, where there is still a large drop in PI between dichoptic and monocular presentation, and the monocular PI =

−2.9 is typical for component cells (Fig. 5G, asterisk). For the pattern model, the largest difference in monocular and dichoptic PI also occurred with 100% V1 opponency (Fig. 5H). PI values also decreased from monocular to dichoptic presentation across a large range of MT inhibitory weight strengths, but unlike the component model, the maximum decrease in PI occurred for non-0 inhibitory weights. Thus, for the representative pattern model (Fig. 5B), we chose a V1 opponency of 100% and MT inhibitory weight strength of 25% (Fig. 5H, asterisk). For both component and pattern cells, the full MT inhibitory weight strength produced the smallest drop in PI with dichoptic presentation.

These results demonstrate the critical role of V1 opponency in our models for explaining responses to dichoptic plaids because both pattern and component units showed no difference between monocular and dichoptic PIs when opponency strength was 0 (Fig. 5G,H). They also predict that there should be a difference in the strength of opponency in the V1 inputs to component and pattern cells: strongly component cells must have V1 inputs that show weaker opponent suppression because stronger suppression results in more pattern sensitivity (Fig. 5C). Furthermore, these results predict that MT pattern-DS cells with higher PIs will tend to have V1 inputs with stronger motion opponency.

In summary, we have applied insights from previous studies (Rust et al., 2006; Tailby et al., 2010) to build binocular component- and pattern-DS cell models that account for MT responses to monocular and dichoptic plaids observed in vivo. Within our model framework, we found that the presence of monocular V1 motion opponent suppression plays the key role in explaining the observations of Tailby et al. (2010), whereas MT inhibitory weights will promote pattern motion sensitivity when stimulated both monocularly and dichoptically. A summary of the values of c_{opp} and k_{inh} that we tested for the plots in Figure 5 can be found in Table 2. A strong prediction of our models is that V1 inputs to MT are moderately to strongly motion-opponent. It is important to note that a large range of the parameter space (V1 opponency strength and strength of MT inhibitory weights) is consistent with the results of Tailby et al. (2010). That their result held true for nearly all of the MT cells that they tested may reflect how widespread motion opponent suppression may be in the V1 inputs to MT. We consider independent evidence for V1 opponency in the Discussion.

Responses to dichoptic plaids in MT models with binocular V1 input

In the models considered above, binocular mixing occurs in MT and the relevant V1 to MT afferents are strictly monocular. This is the simplest circuitry for building binocular MT responses. However, current electrophysiological evidence, although limited, suggests that MT neurons receive input from V1 cells that are themselves primarily binocular, being well driven by each eye. In particular, Movshon and Newsome (1996) found that 11 of 12 neurons projecting from V1 to MT were in OD Groups 3–5 (range 1–7, where 4 has no obvious imbalance) (Hubel and Wiesel, 1962). It is therefore important to determine whether our results are robust across the range of possible ocular dominance strengths of V1 inputs. The models presented in Figures 3–5 show results with monocular V1 inputs, and here we demonstrate these results hold for simple binocular motion integration in V1. In a later section, we also consider models where V1 binocular integration includes a disparity computation, which adds substantially more complexity to the circuitry.
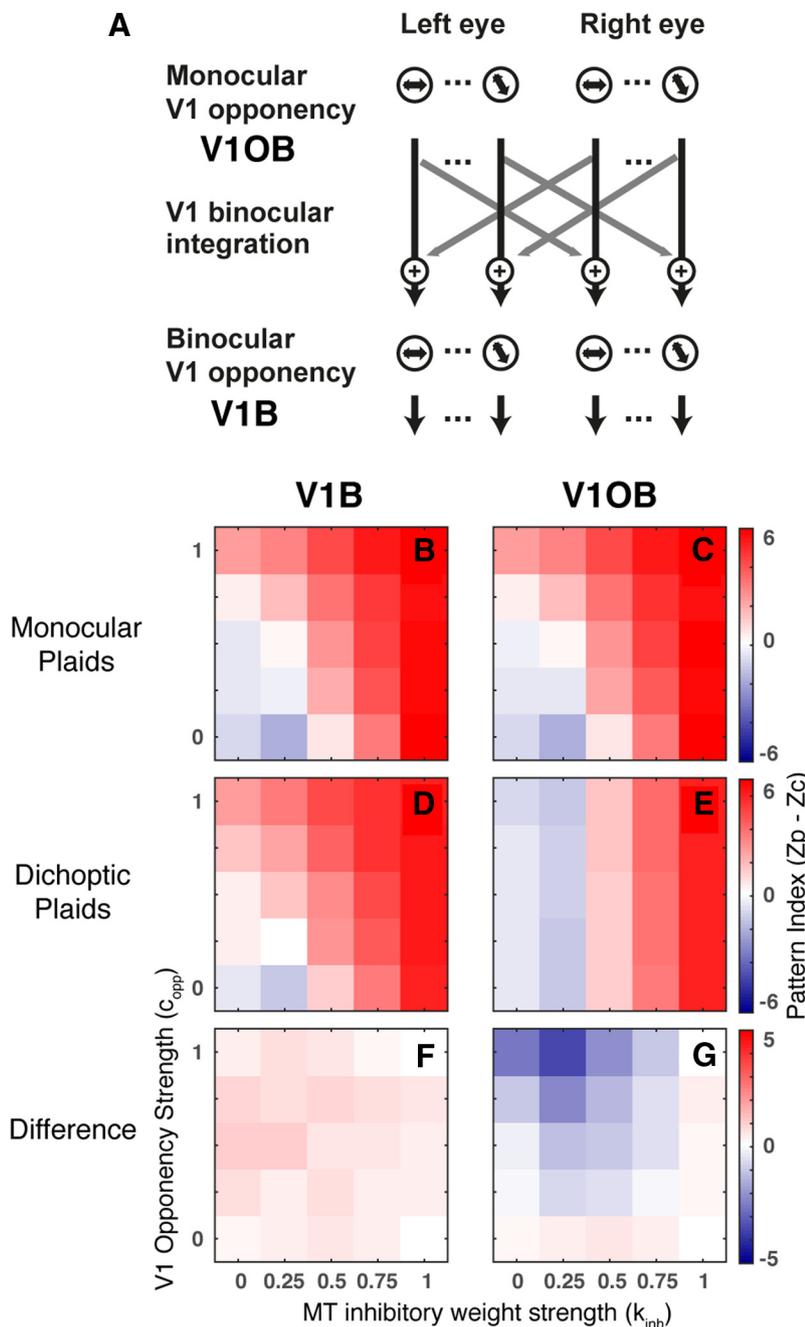
Figure 6. The pattern model with binocular V1 inputs. ***A***, Diagram showing the two alternative models we tested that differ in the location of binocular integration: the V1B model where binocular integration occurs before motion opponency; and the V1OB model, where binocular integration occurs after monocular motion opponent suppression. ***B***, Dependence of the monocular PI on amount of V1 opponency and MT inhibitory weight strength in the V1B model. ***C***, Dependence of monocular PI on V1 opponency strength and MT inhibitory weight strength in the V1OB model. ***D, E***, Dichoptic PI values as opponency and inhibitory strength vary in the V1B (***D***) and V1OB (***E***) models. ***F, G***, The difference between dichoptic and monocular PI as V1 opponency and MT inhibitory strength are varied in the V1B (***F***) and V1OB (***G***) models.

We first implemented binocular integration at the V1 level in our model by mixing signals via a weighted sum within a single direction channel across the left and right streams (Fig. 1D). The balance of input from each eye is set by the parameter b; for 50% mixing ($b = 0.5$), both streams of V1 channels in the binocular model are equally well driven by left and right eye channels (i.e., all channels are OD 4, in which case having two sets of channels is redundant), whereas $b = 0.7$ indicates that the left V1 stream is 70% left eye and 30% right eye driven,

and vice versa for the right stream. In Figures 3–5, b was set to 1.0 for purely monocular V1 channels. We tested two possible sequences for binocular combination and opponency in the model (Fig. 6A): either binocular combination came first, the V1B model (data not shown in Fig. 1), or motion opponency came first, which we refer to as the V1OB model (as shown in Fig. 1C,D).

We first tested the FP pattern cell model with 50/50 left/right eye ($b = 0.5$) binocular V1 inputs using the monocular and dichoptic plaid protocol of Tailby et al. (2010) and varied the strength of V1 opponency and dichoptic inhibition as we did for the model with monocular V1 channels (Fig. 5D,F,H). For monocular plaids, the change in PI as V1 opponency and MT inhibitory strength are varied in the V1B (Fig. 6B) and V1OB (Fig. 6C) models is very similar and also matches the results from the model without binocular V1 mixing (Fig. 5D). For dichoptic plaids, however, there is a stark difference between the two versions of the model. The V1B model (Fig. 6D) shows a similar dependence on the suppression parameters as seen for the monocular plaids (Fig. 6B), whereas the V1OB model (Fig. 6E) lacks the dependence of PI on V1 opponency seen in the monocular V1 model (Fig. 5F). Thus, the V1B model produces negligible changes, even slight increases, in PI with dichoptic presentation (Fig. 6F). Because binocular combination precedes opponency in the V1B model, the V1 opponent motion signal is now binocular, allowing the two grating components to suppress each other, even when they are presented to separate eyes (dichoptically), and pushing the PI toward pattern cell values. Thus, the monocular V1 opponency that is important to differentiating monocular and dichoptic PIs is absent, and the decrease in PI with dichoptic presentation cannot be achieved. In contrast, the V1OB model generates the decreases in PI with dichoptic stimulation that is stipulated by the neuronal data of Tailby et al. (2010) (Fig. 6G) and with similar dependencies as in the monocular V1 model (Fig. 5H). The V1OB model retains the same dependency on opponency strength as the original model with monocular V1 inputs because opponency still occurs at a stage with only monocular signals. Similarly, in the component cell model, we found that the PI value decreased for dichoptic plaids when placing binocular integration after opponency but not before (data not shown). Thus, the model strongly predicts that opponent suppression occurs before binocular integration in V1. Table 2 shows a summary of the key parameter values we tested in this section.
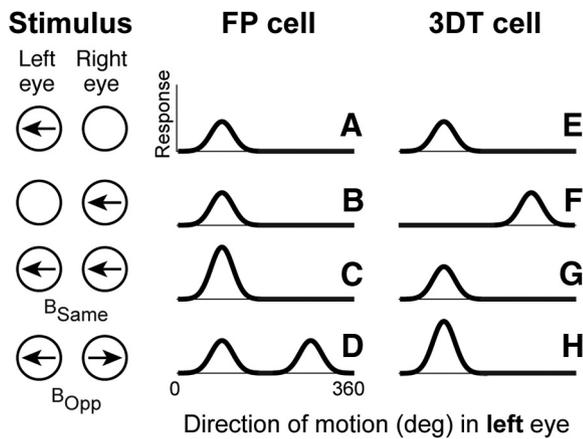
Figure 7. Typical responses for FP and 3DT MT cells from the study of Czuba et al. (2014). Typical plaid direction tuning curves as found by Czuba et al. (2014) using their dichoptic motion protocol with FP cells (left column) and 3DT cells. *A*, Left eye direction tuned response in the FP cell. *B*, Right eye direction tuned response in FP cell. Note the shared tuning with the left eye shown in *A*. *C*, Direction tuned response to B$_{Same}$ in the FP cell. *D*, Orientation tuned response to B$_{Opp}$ found in most FP cells. *E*, Left eye direction tuned response in the 3DT cell. *F*, Right eye direction tuned response in 3DT cell. Right eye tuning curve is shifted 180°, as they found for many 3DT cells. *G*, A direction tuned response to B$_{Same}$ was still typical for 3DT cells despite the disparate monocular tuning. *H*, Strong direction tuned response to B$_{Opp}$ characteristic of 3DT cells.

Tuning for FP and 3D motion in pattern and component units

We now examine whether our binocular pattern and component models, which were developed to explain MT pattern motion selectivity under monocular and dichoptic conditions, can be extended to account for the recently reported sensitivity of MT to 3D motion (Czuba et al., 2014; Sanada and DeAngelis, 2014). Both recent studies found evidence that IOVD cues are the primary drivers of 3D motion sensitivity in MT. We focus here on the results of Czuba et al. (2014), whose stimulus protocol consisted of dichoptic presentation of single sinusoidal gratings to both eyes with variations in drift direction and speed across trials. Specifically, the gratings either drifted in the same direction, which we refer to as B$_{Same}$, or in opposite directions, which we refer to as B$_{Opp}$. Czuba et al. (2014) also varied the speed independently in the two eyes in both of these conditions. The B$_{Same}$ stimulus with equal speed in the two eyes corresponds to classical FP motion. The B$_{Opp}$ stimulus with equal speed and opposite directions in the two eyes corresponds to 3D motion that is purely toward or away from the observer. When the speed varies across eyes, both the B$_{Same}$ and B$_{Opp}$ conditions simulate oblique motion that contains both FP and 3D motion components (Cynader and Regan, 1978).

Using this protocol, Czuba et al. (2014) found that some MT cells, which we will call FP cells, responded best to FP motion, whereas others, which we will call 3DT cells, preferred motion toward or away from the observer. The behavior of FP and 3DT cells is demonstrated with idealized tuning curves for monocular and binocular grating stimuli in Figure 7. In particular, the FP cells (exemplified by Czuba et al., 2014, their Fig. 2*A–D*) showed strong direction-tuning, quantified by a DSI, for the B$_{Same}$ stimulus (Fig. 7*C*), with the same direction preference in each eye when tested monocularly (Fig. 7*A,B*). When tested with the B$_{Opp}$ stimulus, the FP cell tuning curve had a low DSI, typically with two peaks (Fig. 7*D*) corresponding to when the stimulus component in one or the other eye matched the cell's preferred direction. Thus, the FP cell response had no bias for motion toward

versus away from the observer. Our binocular model units characterized above have the same direction tuning in each eye; we thus refer to them as FP models. On the other hand, the 3DT cells of Czuba et al. (2014, exemplified by their Fig. 2*E–H*) preferred opposite directions of motion in each eye (Fig. 7*E,F*). These cells had a direction tuned response with low firing rate for the B$_{Same}$ stimulus (Fig. 7*G*) but had direction tuning with a substantially higher firing rate for the B$_{Opp}$ stimulus (Fig. 7*H*). These 3DT cells were thus associated with 3DT because they were strongly direction tuned when tested with the B$_{Opp}$ protocol, as measured by the DSI being >0.5.

The discovery of 3D versus FP motion tuning in MT raises the question as to whether there is any relationship between 3D tuning and pattern motion sensitivity in MT. Czuba et al. (2014) did not test whether the cells they characterized with binocular motion stimuli were component or pattern-DS for classical monocular stimuli, and so we addressed this from a computational perspective by testing FP and 3D tuned versions of both our component and pattern models. We examined whether there were fundamental differences in the ability of the component and pattern models to generate the typical FP versus 3DT binocular motion tuning shown in Figure 7. As with the dichoptic plaid simulations, we began by testing models where the V1 channels were monocular ($b = 1.0$) as this is the simplest configuration of the models for gaining intuition on how interocular comparisons can give rise to 3D motion tuning. A full summary of the parameter values we varied for the simulations in this section can be found in Table 2.

Our FP component and pattern models (developed above) were tuned for the same preferred direction in the left and right eyes, as can be seen in their monocular single grating tuning curves (Fig. 8*A,B* and Fig. 8*E,F* for component and pattern cells, respectively). When tested with the B$_{Same}$ stimulus, the FP models for both component (Fig. 8*C*) and pattern (Fig. 8*G*) cells responded strongly with a single peak in the direction tuning curve and produced DSIs of 0.8 and 0.7, respectively (see Materials and Methods; DSI defined as in Czuba et al., 2014). When tested with the B$_{Opp}$ stimulus, the component cell produced an orientation-tuned curve with two lower amplitude peaks (Fig. 8*D*), corresponding to when the stimulus in one or the other eye moved in the preferred direction (DSI = 0.0). This matches the responses of many FP cells found experimentally (Czuba et al., 2014, their Fig. 2*C,D*), as idealized in our Figure 7*A–D*. In contrast, the FP pattern cell showed almost no tuning to the B$_{Opp}$ stimulus (Fig. 8*H*, DSI = 0.0). Thus, our representative FP pattern unit, which was fit to the constraints of the dichoptic plaid protocol (Fig. 5*B*), was a poor fit for the typical response pattern of FP neurons in MT.

We then considered what modifications to the canonical FP pattern model would satisfy both the dichoptic plaid and FP motion tuning constraints. The moderate-amplitude, flat tuning curve produced by the FP pattern model with the B$_{Opp}$ stimulus depends on two factors: (1) the broad distribution of excitatory weights, characteristic of pattern cells (Simoncelli and Heeger, 1998; Rust et al., 2006), which drives a spiking response at all directions with the B$_{Opp}$ stimulus; and (2) the weaker MT inhibitory weights required in this model for the dichoptic plaid constraint, which is insufficient to cancel the excitatory drive completely. We tested whether modifying these factors could achieve a pattern model that matched the FP cell data. Varying the second factor, inhibitory strength, alters the amplitude of the firing rate response as direction is varied but does not affect the shape of the tuning curve. Modifying the first factor by narrowing
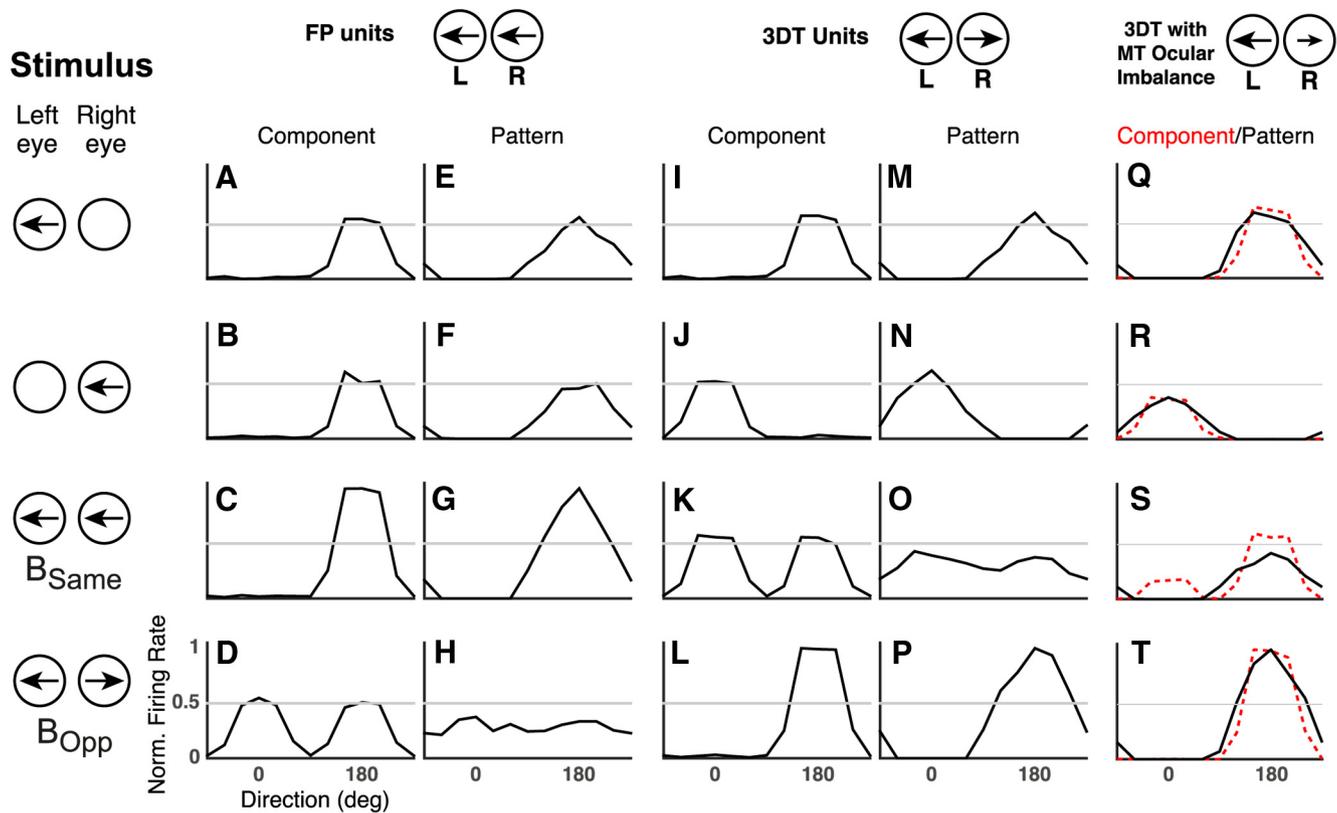
Figure 8. FP and 3DT models. Direction tuning curves obtained with the protocol of Czuba et al. (2014) for monocular gratings, B_{Same} and B_{Opp}. Only one of the speeds we tested is plotted for clarity. Direction varied is that shown to the left eye, with right eye direction either the same or shifted 180°. All tuning curves for a given model are normalized to the maximum firing rate obtained for that model across both the B_{Same} and B_{Opp} protocols. **A–D**, Direction tuning for the FP component cell. Left eye (**A**) and right eye (**B**) tuning is similar. **C, D**, Tuning for B_{Same} and B_{Opp} protocols, respectively. **E–H**, Direction tuning curves for the FP pattern cell. **I–L**, Direction tuning curves for the 3DT component cell. **M–P**, Direction tuning curves for the 3DT pattern cell. **Q–T**, Direction tuning for the 3DT component cell (red) and pattern cell (black) models with right eye MT weights at 50% the strength of the left eye weights ($A_R = 0.5$) and MT inhibitory weights at 100% (component; $k_{inh} = 1.0$) and 75% (pattern; $k_{inh} = 0.75$) strength.

the distribution of MT weights could produce an orientation-tuned response to the B_{Opp} protocol (data not shown); however, this had the concurrent effect of lowering the PI produced by the model when tested with monocular plaids, making the resulting pattern unit more component-like.

To build 3DT models, we simply shifted the MT weight distribution in the right stream by 180° relative to that in the left stream (Fig. 1E, dotted lines in right stream). This resulted in a shift between the peaks for monocular left and right eye tuning curves (Fig. 8I,J,M,N). The direction tuning of the 3DT models for the dichoptic stimuli was reversed relative to that of their FP counterparts. Both component and pattern 3DT models were strongly direction tuned to the B_{Opp} stimulus (Fig. 8L,P) with DSI values of 0.8 and 0.7, respectively. The 3DT component unit showed two robust peaks in tuning for the B_{Same} stimulus (Fig. 8K, DSI = 0.0) where the stimulus in one or the other eye matched the unit's preferred direction for that eye. This deviated from the behavior of the recorded 3DT cells, which were typically direction tuned for the B_{Same} stimulus, despite being driven by both eyes about equally well when tested monocularly (Czuba et al., 2014, their Fig. 2E–G), as idealized in our Figure 7E–G. The 3DT pattern model also differs from the typical 3DT cell of Czuba et al. (2014). In particular, the model is not direction tuned for the B_{Same} stimulus (Fig. 8O, DSI = 0.1), whereas the typical 3DT cell shows clear tuning (Fig. 7G), with the preferred direction matching the left eye in this example.

The direction tuning shown by the 3DT cells for both dichoptic stimuli in the protocol of Czuba et al. (2014), despite each eye

differing in preferred direction, suggests an ocular imbalance. Consistent with this, Czuba et al. (2014) found that their 3DT cells were more ocularly imbalanced than their FP population. Thus, we introduced a left-right stream imbalance such that the MT weights from the right stream were weaker than those from the left stream, allowing the left-stream tuning preference to dominate when both eyes were stimulated. We also increased the strength of MT inhibitory weights, to reduce the amplitude of the untuned response seen with the B_{Same} stimulus and thus increase the DSI. Using a fixed scaling factor, A_R, (Fig. 1F), of 0.5 applied to all weights in the right stream and increasing the scaling factor on MT inhibitory weights, k_{inh}, from 0.25 to 0.75, we achieved a DS tuning curve for the B_{Same} stimulus (Fig. 8S, black line, DSI = 0.7) while maintaining DS tuning for the B_{Opp} stimulus (Fig. 8T, DSI = 0.7), thus obtaining a better match to the observed tuning of typical 3DT cells.

We then tested the component 3DT model with an ocular imbalance, also increasing MT inhibitory weights to full strength ($k_{inh} = 1.0$), to determine whether we could achieve a DS tuning curve to the B_{Same} stimulus. This version of the 3DT component model still produced a double-peaked response, although the peak corresponding to the weaker eye had lower amplitude (Fig. 8S, red line, DSI = 0.5). This component model still lacks the necessary strength of antipreferred direction MT inhibition required to completely nullify the excitation generated by the weaker eye with the B_{Same} stimulus, so the resulting tuning curve has two peaks. Increasing the MT inhibitory weights in the model to values that are above those found by Rust et al. (2006) for

component cells could increase the DSI for the B_{Same} stimulus even further. However, we have shown that, even within the range we tested, as inhibitory strength increases, the model produces PI values for plaids that are too high to be classified as a component unit (see Fig. 5C).

Having established the factors that allow 3D motion tuning in the models, we explored what range of these parameters could support DS tuning for the B_{Same} stimulus in the 3DT models. We varied the level of MT ocular imbalance, A_R, and MT inhibitory weight strength, k_{inh}, in the component (Fig. 9A) and pattern (Fig. 9B) 3DT models and plotted the resulting DSI values. Figure 8Q–T (asterisks) indicates the parameter values used for the tuning curves. The highest level of MT imbalance we tested, $A_R = 0.5$, results in monocularity indices of 0.26 and 0.23 in the component and pattern models, respectively, both close to the median reported for 3DT cells in Czuba et al. (2014) (0.24). In the component model (Fig. 9A), only an MT weight imbalance with the right stream at 50% of left stream strength and full-strength MT inhibitory weights produced a DSI of 0.5. With the pattern model (Fig. 9B), a much larger range of the parameter space produced DS responses, and with DSI values >0.7. The highest DSI values occur with the strongest MT inhibitory weights. With strong dichoptic inhibitory inputs, the inhibition generated by one stream, in response to an antipreferred stimulus, can suppress the excitation when the other stream is presented as a preferred stimulus. With the ocular imbalance, there will be responses to the dominant eye stream's preferred stimulus but not the weaker stream, resulting in a DS tuning curve and a high DSI. The component model, lacking strong dichoptic inhibition, will always generate tuning curves with two peaks, corresponding to each eye's preferred direction, and produce lower DSI values. Although 75% of the cells in Czuba et al. (2014) had a DSI >0.5 for the B_{Same} stimulus, not all 3D tuned cells did (their Fig. 3B). Thus, some subset of 3D tuned units may be consistent with our component model. Alternatively, if there are many 3DT component cells with a DSI > 0.5 for the B_{Same} protocol, this implies that there is an additional source of dichoptic, opponent inhibition in MT component cells that is not captured in the model fits to the monocular stimulus protocols.

Having achieved 3D motion tuning in the model with monocular V1 inputs, we then tested whether this tuning was robust to varying V1 ocular dominance. We built a 3DT pattern model by shifting the MT weight distributions between left and right eye streams (Fig. 1E), with V1 motion opponency preceding binocular integration as in the FP version of the V1OB model (Fig. 6A). We use the V1OB model circuitry here, with opponency preceding V1 binocular integration, as we have shown that this ordering
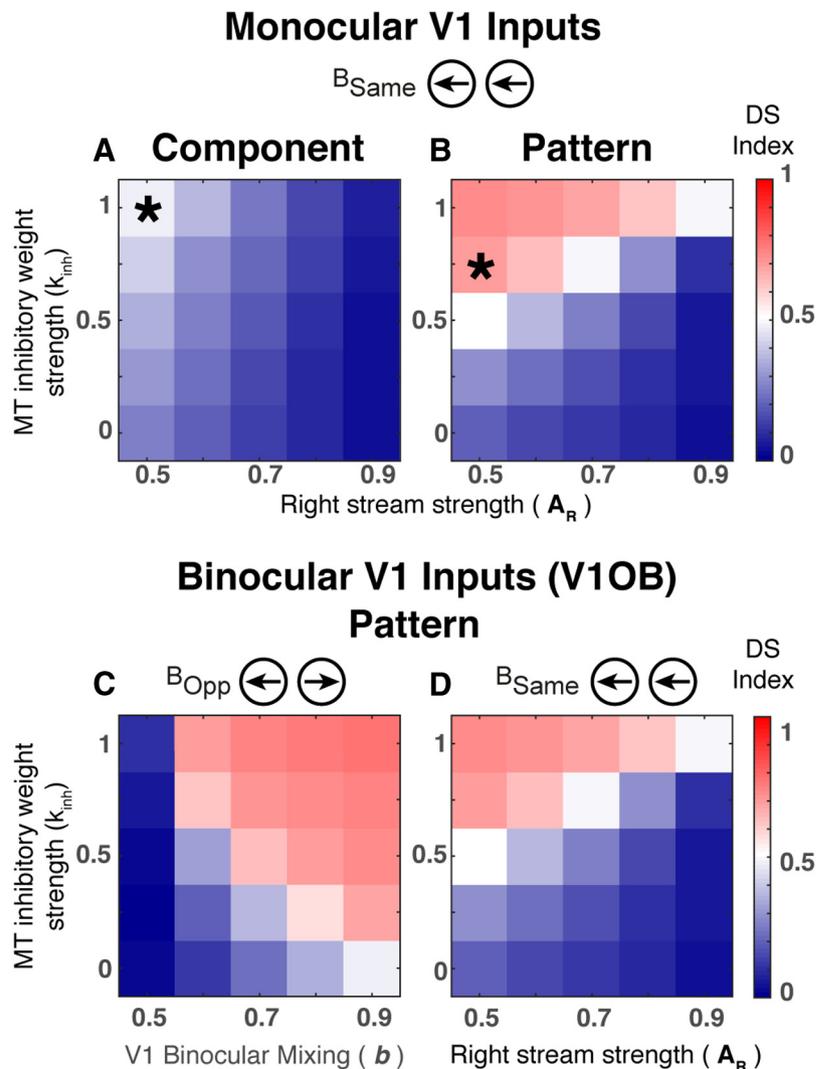
Figure 9. Direction selectivity indices for binocular motion in the 3DT models. **A**, Plot of DSI for B_{Same} as MT inhibitory weight strength and MT ocular imbalance level are varied in the 3DT component model. Blue regions represent DSI values <0.5. White areas represent DSI = 0.5. *Parameter values used for the component model tuning curves plotted in Figure 8Q–T. **B**, DSI for binocular matched motion in the pattern model as inhibitory strength and MT ocular imbalance are varied. Red regions represent DSI values >0.5. *Parameters chosen to generate the pattern model tuning curves in Figure 8Q–T. **C**, Plot showing how DSI for the B_{Opp} stimulus varies as MT inhibitory strength and V1 ocular dominance are varied in the 3DT V1OB model. **D**, Plot of DSI obtained with the B_{Same} stimulus as MT inhibition strength and MT ocular imbalance level are varied in the 3DT V1OB model.

of the V1 stages is needed to explain dichoptic plaid results (Fig. 6). The strength and relative ordering of V1 opponent suppression do not affect 3DT model output using the protocol of Czuba et al. (2014; data not shown) because only one motion component is shown to each eye for the dichoptic Czuba et al. (2014) stimuli; thus, the V1OB and V1B configurations of the model (Fig. 6A) give similar results.

In the 3DT model with monocular V1 inputs ($b = 1.0$), 3D motion tuning is achieved by computing sums and differences of V1 signals across the two eye streams. If the V1 channels have perfect binocular balance, this computation would no longer be possible because it requires the comparison of monocular motion signals, and in the case with exactly balanced left and right eye V1 inputs, the "left" and "right" eye streams are identically driven. We examined this by varying the V1 binocular balance and MT inhibitory weight strength in the 3DT V1OB pattern model and measuring the resulting DSI values for the B_{Opp} stimulus (Fig. 9C). We found that the 3DT model with 50/50 V1
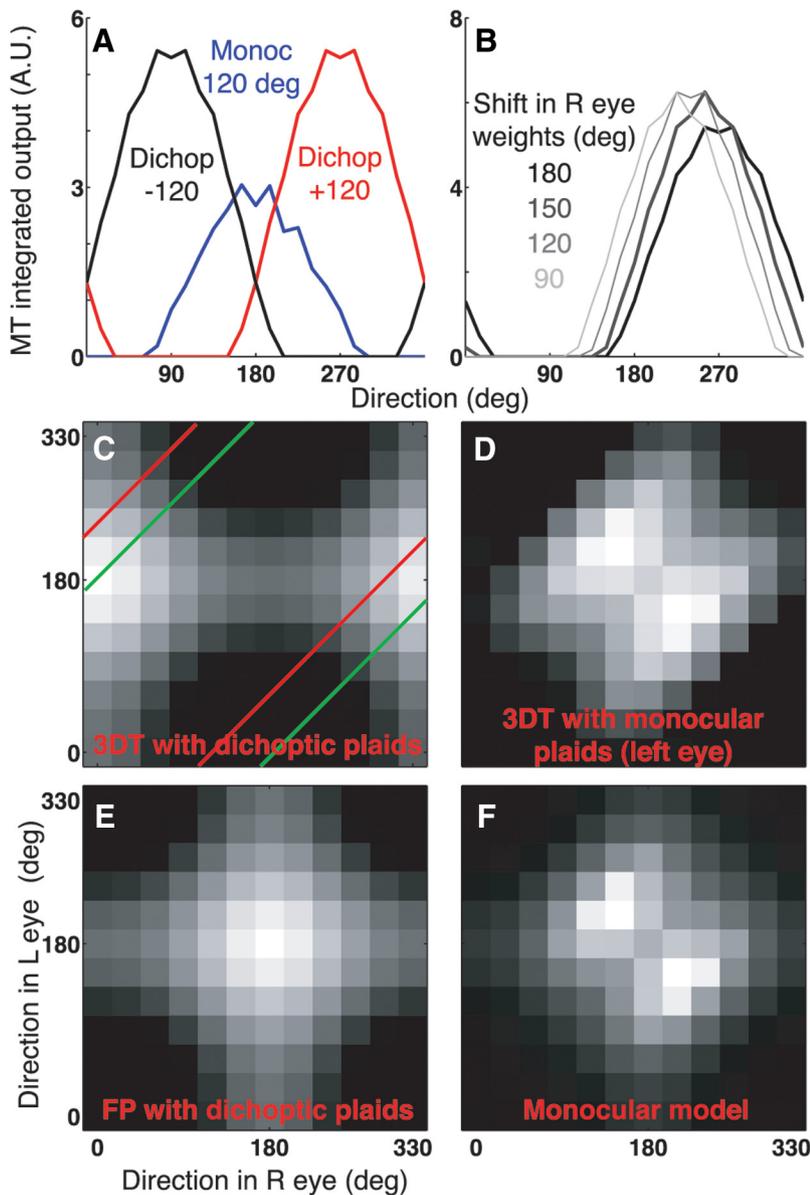
Figure 10. Dichoptic plaids in the 3DT pattern model. For all panels, response plotted is the MT output signal after the rectification nonlinearity is applied. **A**, Direction tuning for monocular (blue trace) and −120° (black trace) and 120° (red trace) dichoptic plaids in the 3DT pattern model. **B**, Dichoptic plaid direction tuning in the 3DT pattern cell as the shift in the weight distribution to the right eye is varied from 180° (black trace) to 90° (lightest gray). **C**, Matrix of responses to dichoptic plaids in all direction combinations (as in Fig. 3L) for the 3DT pattern model. Red lines indicate the stimulus space sampled by the protocol of Tailby et al. (2010). Green lines indicate the B_{Opp} protocol of Czuba et al. (2014). **D**, Response matrix to all monocularly (left eye) presented plaids in the 3DT pattern cell model. **E**, Response matrix to all dichoptic plaids from the FP pattern cell model for comparison. **F**, Response matrix for a monocular version of the pattern cell model with no V1 opponency or reduced inhibitory weights (as in Rust et al., 2006).

similar dependence of direction tuning strength on these parameters in the model with binocular V1 inputs. Thus, the results we described for our model with monocular V1 channels can be reproduced with binocular V1 inputs, as long as a V1-level ocular imbalance is included.

In summary, our models predict that tuning for FP and 3D motion in MT is coincident with the spectrum of component: pattern motion sensitivity. The binocular FP component model that we built according to the constraints of Tailby et al. (2010) fits the response pattern of the FP neurons described by Czuba et al. (2014) with no modifications. The best fit to the data for the FP pattern model involves a modification that decreases the pattern sensitivity of the model. Conversely, the responses of 3DT neurons are best fit by a pattern model with an ocular imbalance, whereas the component model is unable to fit the data without the addition of dichoptic inhibition that increases the pattern sensitivity of the component model. This is a striking and untested prediction of our binocular models: that 3DT cells in MT may be associated with pattern motion sensitivity, whereas FP motion tuned cells will be associated with component selectivity.

Dichoptic plaids with 3DT cells
We have presented models of MT cells that are able to reproduce the results of two key studies on dichoptic motion processing in MT. Another test we performed was to examine the tuning of the 3DT model to the dichoptic plaid protocol of Tailby et al. (2010), which we used to characterize the FP models in Figures 5 and 6. This is interesting to consider because Tailby et al. (2010) may have encountered such cells during their study, and exploring the stimulus space more thoroughly may reveal predictions that can be used to validate or dismiss the models.

Figure 10A shows tuning curves for monocular (blue trace) and dichoptic plaids with 120° (red) and −120° (black) differ-

ences in component grating direction. With the dichoptic plaids, the peak direction in the 3DT pattern unit shifts by 90°. The direction of this peak shift reverses when the component gratings are swapped between the two eyes. We found that the magnitude of the shift was determined by the interocular difference in direction preference. Figure 10B shows the effect of shifting the weight distribution in the right eye, with successively lighter traces for each 30° shift in the preferred direction of the right eye. The tuning curve shifts 15° for each 30° shift in interocular direction preference. The strength of the MT inhibitory weights had no effect on the magnitude of the shift: even in the absence of inhibitory weights, the peak was still shifted 90° (data not shown). We found identical results for the 3DT com-

binocular mixing produced flat direction tuning curves, resulting in uniformly low DSIs, with the B_{Opp} protocol of Czuba et al. (2014) for all MT inhibitory strengths (Fig. 9C, $b = 0.5$). As we increased the ocular imbalance in the V1 channels (increasing b), strong direction tuning could be produced in the V1OB model with appropriate MT inhibitory strengths, indicating that a sufficient IOVD signal remained to generate 3DT. Setting the V1 binocular mixing to 70% dominant eye/30% nondominant eye ($b = 0.7$), we then tested how MT inhibitory weight strength and the MT level ocular imbalance, A_R, affected DSI values for the B_{Same} protocol (Fig. 9D), as we had done for the 3DT pattern model with monocular V1 inputs (Fig. 9B). We found a very

ponent unit as well (data not shown). Thus, our model predicts that 3DT cells with 180° differences in preferred direction between the two eyes will show a shift in preferred direction of 90° when tested with dichoptic plaids.

To fully explore the response patterns of the binocular models to all plaids, we characterized the representative FP and 3DT pattern units (Figs. 5B, 8Q–T) using the protocol of Rust et al. (2006) with plaids of all direction combinations in 30° steps. We tested both dichoptic (Fig. 10C,E) and monocular stimulation (Fig. 10D). Figure 10C (red line) shows the slice through stimulus space of the plaids that form the dichoptic 120° plaid protocol of Tailby et al. (2010), whereas the green line shows the dichoptic opponent motion "plaids" that make up the B_Opp protocol of Czuba et al. (2014). The activated region of the 3DT model's plot (Fig. 10C) is shifted by 180° in the full stimulus space along the axis of grating direction in the right eye compared with the FP model's plot (Fig. 10E). This reflects the 180° shift in preferred direction of the right eye in the 3DT model. The preferred stimulus of the 3DT pattern cell was a "counterphase" dichoptic plaid (components at 0° and 180°), matching the B_Opp stimulus of Czuba et al. (2014) (Fig. 10C). The binocular FP pattern unit showed a preference for stimuli centered around 180° motion in each eye when tested dichoptically (Fig. 9E), as expected by construction of its tuning curves, but without the trough in the plaid matrix seen when stimulating the binocular model monocularly (Fig. 10D). This is because of the loss of monocularly driven tuned normalization when the plaid is presented dichoptically. When tested monocularly (Fig. 10D), the 3DT unit showed a similar preference as a purely monocular version of the model (Fig. 10F), but the region of activation was elongated in the stimulus space because of the reduced inhibition in the binocular model.

It is unclear whether the experiments of Tailby et al. (2010) would have uncovered this response pattern for 3DT units. They did report that most of their cells had very similar direction tuning in both eyes, which is consistent with the report by Czuba et al. (2014) that a small proportion of MT cells had monocularly opposed direction tuning (their Fig. 5). Tailby et al. (2010) did not report the preferred directions measured with monocular and dichoptic plaids, so shifts in the tuning curve peaks may have been overlooked. Given that the PI is the difference between the pattern and component partial correlation coefficients, a shift in preferred direction with dichoptic presentation should result in a reduction in PI. This follows because the resulting tuning curve for the dichoptic plaid should be poorly correlated with the tuning curve prediction from the monocular gratings, which would have a different preferred direction. This is what we found for our pattern 3DT unit: the PI values were 3.1 and −1.3 for monocular and dichoptic plaids, respectively, with the pattern correlation dropping from 4.3 to −0.7 and component correlation also decreasing from 1.2 to 0.6.
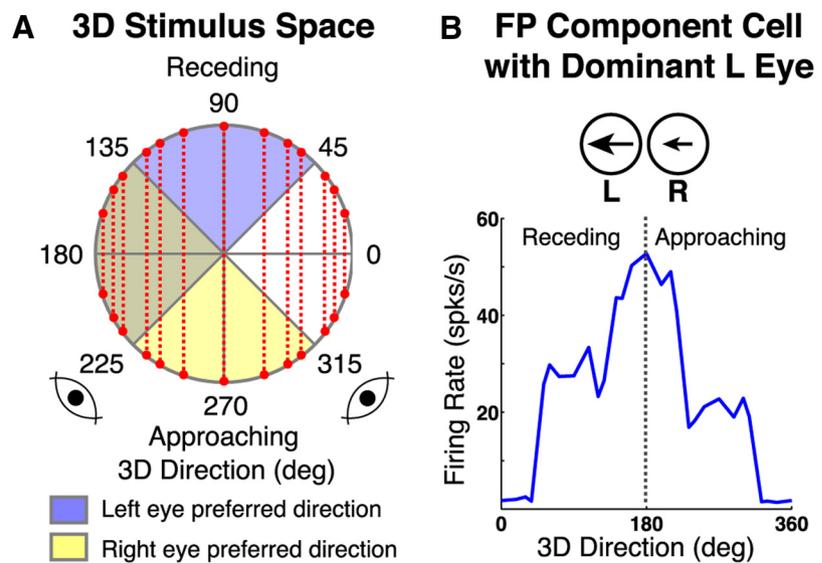
Figure 11. A model for a 3D motion biased cell using an MT ocular imbalance. *A*, Diagram showing the 3D motion direction space generated by varying interocular stimulus direction and speed (Cynader and Regan, 1978; Czuba et al., 2014). Position of the eyes is indicated by icons at 225° and 315°. The speeds we used to generate the 3D motion directions were 1, 2, and 7.5 deg/s. All speeds drove the MT neuron well monocularly. Red dots around the circumference of the circle represent the 3D directions sampled by our protocol. Dashed red lines connect the stimuli, which were matched in the statistical test for 3D bias as used by Czuba et al. (2014). Blue shaded half-circle represents the stimulus space where the left eye in our model FP component cell sees its preferred direction of motion. Yellow shaded half-circle represents the same for the right eye. *B*, 3D direction tuning for the FP component cell with an ocular imbalance such that the right eye weights are 70% the strength of the left eye weights. From 0° to 180° are all receding motions, whereas from 180° to 360° are approaching motions.

In summary, it is possible to reconcile the results of Tailby et al. (2010) and Czuba et al. (2014) for 3DT neurons consistent with our model. A full stimulus protocol that characterizes tuning for plaids both monocularly and dichoptically would expose the shift in tuning predicted by our models and reveal any connection between pattern motion selectivity and 3D motion sensitivity, as predicted above.

3D motion biased tuning in MT cells

The FP and 3DT cells described by Czuba et al. (2014) are two ends of a spectrum of sensitivity to 3D motion observed in their MT population. They described most of their cells as 3D biased, indicating that these cells did not show strong direction selectivity (as defined by a DSI > 0.5) for the B_Opp protocol but did have a statistically significant preference for receding or approaching directions of motion for oblique 3D directions. Figure 11A (red points on the circle) shows the oblique 3D directions generated when speed (1, 2, and 7.5 deg/s) and direction (0° or 180°) are varied between the two eyes, as in Czuba et al. (2014) (see Materials and Methods). Points in the top half of the plot correspond to motion away from the observer, whereas those in the bottom half correspond to motion toward the observer. Cells that are 3D biased responded more strongly overall to the set of stimuli with 3D directions that fall in one-half of the stimulus space (receding or approaching) than to those that fall in the other half, as measured by a matched Wilcoxon signed rank test (matched stimuli are shown connected by dashed red lines in Fig. 11A). Czuba et al. (2014) also found that 3D biased cells tended to have similar direction tuning in both eyes, similar to the FP cells, and >50% of 3D biased cells had a monocularity index < 0.125, indicating they tended to be ocularly balanced.

Czuba et al. (2014) and others have noted that 3D biased responses may arise from differences in speed tuning between the two eyes, but here we explore an alternative circuitry for

generating a 3D biased tuning curve. We exploited a difference in the monocular distribution of motions generating the 3D directions to build a model unit that would be classified as 3D motion biased. Consider an FP unit that prefers leftward motion. The left eye for this unit would see its preferred direction motion in Figure 11A (semicircle shaded in blue), whereas the right eye would see its preferred direction in the yellow semicircle. Noting that the top quadrant (blue region) corresponds to receding motion where only the left eye sees its preferred direction while the bottom quadrant (yellow region) corresponds to approaching motion where only the right eye sees its preferred direction, we reasoned that an FP unit with both eyes tuned to leftward motion and an ocular imbalance favoring input from the left eye should show a tuning bias for motion directed away from the observer.

Figure 11B shows a direction tuning curve for such an FP component unit stimulated with the oblique directions represented by Figure 11A (red points). The 3D direction tuning curve is skewed toward receding motions (3D directions between 0° and 180°). This result is highly statistically significant ($p < 0.002$; Wilcoxon signed rank test). The unit is tuned to 170°, slightly away from pure FP leftward motion. In addition, even though an ocular imbalance in the component unit was necessary to achieve these results ($A_R = 0.7$), the monocularity index value for this cell is 0.12, falling within the range of ocular imbalances that Czuba et al. (2014) tk;2found for their 3D biased cells. The DSI for this unit tested with the B_{Opp} protocol is 0.1; thus, this unit is 3D biased, not 3D tuned, by the criteria of Czuba et al. (2014).

Importantly, 3D biased cells constructed in this way must conform to the following pattern. If an MT unit is left eye-dominant, it will be biased for receding motion if both eyes prefer leftward FP motion and biased for approaching motion if both eyes are tuned for rightward motion, and vice versa if the right eye is the dominant eye. This is a specific and readily testable prediction of this model

←

PI as V1 opponency and MT inhibitory strength are varied in the component (**G**) and pattern (**H**) models. **I**, Plot showing how DSI for the B_{Opp} stimulus varies as MT inhibitory strength and V1 ocular dominance are varied in the 3DT pattern model with disparity selectivity. **J**, Plot of DSI obtained with the B_{Same} stimulus as MT inhibition strength and MT ocular imbalance level are varied in the 3DT pattern model.

Figure 12. MT models with binocular disparity tuning. **A**, Plots of disparity tuning in the FP component model with binocular disparity for correlated (blue trace), anticorrelated (red trace), and uncorrelated (black trace) dynamic random dot stimuli. **B**, Disparity tuning in the FP pattern model with binocular disparity with dynamic random dots as in **A**. **C, D**, Dependence of monocular PI on V1 opponency strength and MT inhibitory weight strength in the FP component (**C**) and pattern (**D**) models with disparity. V1 opponency precedes disparity computation in these models (see Materials and Methods). **E, F**, Dichoptic PI values as opponency and inhibitory strength vary in the component (**E**) and pattern (**F**) models. **G, H**, The difference between dichoptic and monocular

for producing MT units with 3D motion biased responses.

MT models with motion and binocular disparity selectivity

We have presented the simplest models that could capture observations from recent binocular motion studies, but a complete binocular model of MT requires disparity tuning, which is expressed by the majority of MT cells (Maunsell and Van Essen, 1983; DeAngelis and Uka, 2003). Because V1 inputs to MT are thought to be primarily binocular (Movshon and Newsome, 1996), and disparity is computed when binocular integration occurs, it is reasonable and parsimonious to build a model where the V1 channels are jointly motion and disparity tuned, conferring both motion and disparity selectivity to the model MT unit. This leaves the key question of whether our results from the motion-only models still hold when we incorporate the constraints on the circuitry necessary for generating jointly tuned responses.

The jointly tuned model we built incorporates cascaded computations for both motion and binocular disparity energy to create complex (i.e., phase-invariant) DS and disparity-tuned V1 channels. The two main changes to the framework for the motion models shown in Figure 1 were as follows: (1) the initial motion energy stage (Fig. 1A) was opened up to allow the even and odd linear filter channels to remain separate through the normalization and opponency stages; and (2) the opponency and V1 integration stages (Fig. 1C,D) were replaced by the circuitry shown in Figure 2, which implements the binocular disparity energy model (Ohzawa et al., 1990). Because the models require monocular opponency with subsequent rectification preceding binocular integration, the disparity circuitry has a form that is very similar to that proposed by Read et al. (2002) for tuned-excitatory cells. All the V1 channels share the same disparity preference, in this case for 0 disparity (tuned excitatory). These jointly tuned V1 inputs are then subject to the same MT weights and output nonlinearity as used in the motion-only models to produce spiking MT responses. We tested our jointly tuned MT units, built with the same parameters as our canonical FP models (Fig. 5A,B), with correlated and anticorrelated dynamic random dot stimuli used to characterize binocular disparity tuning. Both the component (Fig. 12A) and pattern (Fig. 12B) models showed strong tuning for disparity, with the tuning curve for anticorrelated dots inverted and reduced in amplitude as is typical of V1 and MT disparity-tuned cells (Cumming and Parker, 1997; Krug et al., 2004). This attenuation is not seen in the classical disparity energy model but has been shown to arise when the subunits of the disparity energy computation are rectified (Read et al., 2002), as occurs in our model at the opponency stage (Fig. 2B), which includes rectification (see Materials and Methods, Eq. 10).

In the jointly tuned models, we repeated our simulations testing the dependence of pattern motion sensitivity on monocular and dichoptic opponent suppression, as presented above for the motion-only models (Figs. 5C–H, 6C,E,G). Both the component (Fig. 12C,E,G) and pattern (Fig. 12D,F,H) models with disparity tuning share similar trends with the nondisparity models, most critically the requirement for monocular opponency to see the drop in PI observed with dichoptic presentation (Fig. 12G,H). The main difference is that the monocular V1 opponency has an even stronger effect on PI values than in the motion models, leading to large increases in PI in both the component (Fig. 12C) and pattern (Fig. 12D) models for any value of opponency >0 compared with monocular plaids in the motion-only models. This reflects the fact that the rectification, which occurs after opponency, precedes the squaring in the disparity energy computation, thus resulting in a stronger suppressive effect.

We then tested a 3DT version of the disparity-tuned pattern model with the dichoptic protocols of Czuba et al. (2014), comparing the dependence of DSI on MT inhibitory weight strength and V1 (Fig. 12I) or MT (Fig. 12J) level ocular imbalances, as we did for the motion-only model with binocular V1 inputs (Fig. 9C and Fig. 9D, respectively). The results are very similar between the motion-only and joint motion-disparity tuned models, showing that our results are robust and apply to circuits having a wide variety of mechanisms of binocular integration at the V1 level, including no integration, simple additive combination, or plausible disparity energy computations.

Discussion

We have presented the first image-computable binocular model of pattern and component motion selectivity in MT and found evidence for a previously unpredicted and untested relationship between pattern motion sensitivity and tuning for 3D motion. We built component and pattern units that reproduce key binocular response features of MT cells that have not been previously explained. Our unified models can account for the decreased response to pattern motion when plaids are presented dichoptically (Tailby et al., 2010) and the recently reported tuning for 3D motion (Czuba et al., 2014). We found that motion opponent suppression is key to explaining decreased sensitivity to global motion when plaids are presented dichoptically. This mechanism explains the observed drop in pattern index for cells across the full range of pattern motion selectivity. Our models reproduced MT responses to dichoptic plaids using V1 channels ranging from strictly monocular to balanced binocular V1 responses, and the same mechanisms were relevant for the decrease in pattern sensitivity across the range of ocular dominances. Using this same circuitry, we also constructed FP, 3D-biased, and 3D-tuned MT model units that fit the experimentally observed responses to the dichoptic motion protocol of Czuba et al. (2014). We showed that FP and 3DT units qualitatively fit with the expected responses for component and pattern cells, respectively. We found that model units that are sensitive to IOVD signals can be built from binocular V1 inputs, provided the V1 channels are ocularly imbalanced. Last, we showed that our results all hold in the models when binocular disparity computations are incorporated in the V1 stages.

Model predictions

Our models generated the following major predictions: (1) V1 inputs to MT will exhibit motion-opponent suppression, and this opponency will be stronger in the inputs to pattern cells than in the inputs to component cells. Opponent suppression can be measured by comparing responses to drifting and counterphase gratings (Qian and Andersen, 1995; Thiele et al., 2000) presented monocularly to V1 neurons identified as projecting to MT. (2) Opponent motion suppression occurs before binocular integration in V1. There is experimental evidence for a significant monocular source of motion opponent suppression in MT neurons (Majaj et al., 2007). (3) The 3D-tuned cells will be pattern-selective when tested monocularly, whereas FP-tuned will be component-selective. This can be tested experimentally by adding monocular 120° plaid tuning to the protocol of Czuba et al. (2014). (4) The 3D-biased cells without interocular differences in speed tuning can show 3DT depending on their ocular dominance and direction preference, with left eye-dominant MT units being biased for receding motion if tuned for leftward FP motion, and approaching motion if tuned for rightward motion, and vice versa for right eye-dominant units. This correspondence can be

measured within 3D motion protocols that test oblique 3D directions (Maunsell and Van Essen, 1983; Czuba et al., 2014). (5) MT units with opposite direction preferences in each eye will show a shift in preferred direction with dichoptic plaids. This can be tested electrophysiologically, by recording the responses of 3D-tuned MT neurons to dichoptic and monocular plaids. Although Tailby et al. (2010) did not report tuning shifts in their cells, it is unclear whether 3D-tuned cells were included in their dataset. They reported that most, but not all, cells had similar tuning across eyes, and their methods for isolating units may have overlooked cells with large interocular differences in direction preference.

Accounting for disparity selectivity in MT

While focusing on explaining responses to binocular motion stimuli, we have also presented preliminary models incorporating disparity selectivity to show that our results hold in such models. There are many other properties of disparity tuning in MT that must be considered to build a complete jointly tuned MT model. Changing disparity (CD) signals may provide another cue for perception of motion-in-depth. Sanada and DeAngelis (2014) used stimuli that isolated IOVD and CD cues for MID and found evidence that both cues contributed to MID responses in MT neurons, although IOVD cues were far more effective. Applying their protocol will require incorporating sensitivity to changing disparity into our model framework. Disparity studies may provide additional constraints on sources of opponent inhibition, as disparity-tuned motion-opponent suppression has been described in studies of transparent motion in MT (Qian and Andersen, 1994; Bradley et al., 1995). Opponent suppression may be mediated by reciprocal interactions between disparity channels in MT or from interactions between disparity-tuned V1 subunit inputs to MT, as proposed in models of IOVD processing (Sabatini and Solari, 2004).

Incorporating disparity tuning at the V1 level into our MT framework opens up many complex questions. Discrepancies between disparity tuning in V1 and MT suggest that MT does not simply inherit disparity tuning from V1 (for review, see Cumming and DeAngelis, 2001). The strength and type of disparity tuning of V1 inputs to MT have not been characterized. Moreover, a recent study (Ponce et al., 2008) showed that disparity tuning in macaque MT was decreased when V2 was cooled, whereas motion selectivity was unaffected, suggesting that V2 inputs contribute significantly to disparity in MT. Combining data from V1 and V2 studies is therefore necessary for building and testing joint motion-disparity MT models.

Sources of motion opponent suppression

The mechanism we identified as critical to explaining responses to dichoptic plaids is monocular V1 motion opponency. Qian and Andersen (1995) tested for opponency in V1 DS neurons by comparing responses to single drifting gratings with those for counterphase gratings, the latter which consist of concurrent and colocalized preferred and antipreferred motion, presented binocularly. They found that the vast majority of neurons showed opponent suppression, although at moderate strengths. Qian and Andersen (1995) did not identify whether the source of this suppression was monocular or binocular. Furthermore, the strength of suppression shown by the specific V1 cells that project to MT remains unknown. Additional evidence for monocular opponency contributing to suppression has been presented by Majaj et al. (2007) in a study comparing contrast response functions in MT neurons when opponent motion gratings were presented monocularly or dichoptically. They found larger shifts, indicating greater suppression, for monocular presentation. A similar protocol was used by Thiele et al. (2000), showing suppression in MT neurons by comparing responses to drifting and counterphase gratings, although dichoptic presentation was not tested. Thus existing evidence supports both motion opponency in V1 DS neurons and a monocular source of MT opponent suppression. To test explicitly for monocular opponent suppression in V1, the protocol of Majaj et al. (2007) could be used in V1 DS cells, comparing responses to drifting and counterphase gratings presented monocularly (testing for monocular opponent suppression) and optionally also dichoptically (dichoptic suppression). This would complete the link, showing that monocular opponency in V1 could contribute to the opponency seen in MT.

Previous models of MT neurons

There are many existing monocular models of motion processing in MT (Heeger, 1987; Grzywacz and Yuille, 1990; Nowlan and Sejnowski, 1995; Simoncelli and Heeger, 1998; Bowns, 2002; Perrone and Thiele, 2002; Pack et al., 2004; Perrone, 2006; Rust et al., 2006; Tsui et al., 2010). Each has limitations, such as not representing circuitry with realistic neural elements, or being fit for a particular stimulus protocol and not easily generalizable to arbitrary visual stimuli; however, they also include aspects of motion sensitivity in MT that should be included in future iterations of our binocular models. Such features include multiple SF and TF channels (Simoncelli and Heeger, 1998), and modeling the spatial surround of V1 neurons (Tsui et al., 2010), which may provide a physiological mechanism for tuned normalization (Rust et al., 2006). Future models should also include spiking circuits to build V1 DS inputs (Baker and Bair, 2012), which opens up the possibility for cross-correlation studies between V1 and MT.

Modeling of 3D motion perception and psychophysical data

Psychophysical and fMRI studies implicate MT as an important stage for 3D motion processing (Rokers et al., 2011), with the role of IOVD and CD cues being debated (Harris et al., 2008; Czuba et al., 2010; Rokers et al., 2011). Previous work has shown that MID stimuli can confound CD and IOVD cues in energy models that compute CD and IOVD signals (Peng and Shi, 2014). A key benefit of our image-computable framework is that the same stimuli used in psychophysical protocols can be tested and refined directly in circuit models of MT processing. This is valuable because stimuli that are most relevant to generate perceptual effects can differ substantially from those used to characterize cortical neurons, the latter often being tailored to maximize single neuron responses. Further bridging the gap between perceptual relevance and physiological optimality will require extending our model to large-scale populations at the MT level to facilitate studies of perceptual read-out.

Notes

References

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2:284–299. CrossRef Medline

Adelson EH, Movshon JA (1982) Phenomenal coherence of moving visual patterns. Nature 300:523–525. CrossRef Medline

Albright TD, Desimone R, Gross CG (1984) Columnar organization of di-

rectionally selective cells in visual area MT of the macaque. J Neurophysiol 51:16–31. Medline

Baker PM, Bair W (2012) Inter-neuronal correlation distinguishes mechanisms of direction selectivity in cortical circuit models. J Neurosci 32:8800–8816. CrossRef Medline

Bowns L (2002) Can spatio-temporal energy models of motion predict feature motion? Vision Res 42:1671–1681. CrossRef Medline

Bradley DC, Qian N, Andersen RA (1995) Integration of motion and stereopsis in middle temporal cortical area of macaques. Nature 373:609–611. CrossRef Medline

Cumming BG, DeAngelis GC (2001) The physiology of stereopsis. Annu Rev Neurosci 24:203–238. CrossRef Medline

Cumming BG, Parker AJ (1997) Responses of primary visual cortical neurons to binocular disparity without the perception of depth. Nature 389:280–283. CrossRef Medline

Cynader M, Regan D (1978) Neurones in cat parastriate cortex sensitive to the direction of motion in three-dimensional space. J Physiol 274:549–569. CrossRef Medline

Czuba TB, Rokers B, Huk AC, Cormack LK (2010) Speed and eccentricity tuning reveal a central role for the velocity-based cue to 3D visual motion. J Neurophysiol 104:2886–2899. CrossRef Medline

Czuba TB, Huk AC, Cormack LK, Kohn A (2014) Area MT encodes three-dimensional motion. J Neurosci 34:15522–15533. CrossRef Medline

DeAngelis GC, Uka T (2003) Coding of horizontal disparity and velocity by MT neurons in the alert macaque. J Neurophysiol 89:1094–1111. Medline

Felleman DJ, Kaas JH (1984) Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. J Neurophysiol 52:488–513. Medline

Grzywacz NM, Yuille AL (1990) A model for the estimate of local image velocity by cells in the visual cortex. Proc R Soc Lond B Biol Sci 239:129–161. CrossRef Medline

Harris JM, Nefs HT, Grafton CE (2008) Binocular vision and motion-in-depth. Spat Vis 21:531–547. CrossRef Medline

Heeger DJ (1987) Model for the extraction of image flow. J Opt Soc Am A 4:1455–1471. CrossRef Medline

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154. CrossRef Medline

Jazayeri M, Wallisch P, Movshon JA (2012) Dynamics of macaque MT cell responses to grating triplets. J Neurosci 32:8242–8253. CrossRef Medline

Krug K, Cumming BG, Parker AJ (2004) Comparing perceptual signals of single V5/MT neurons in two binocular depth tasks. J Neurophysiol 92:1586–1596. CrossRef Medline

Majaj NJ, Tailby C, Movshon JA (2007) Motion opponency in area MT of the macaque is mostly monocular. J Vis 7:96.

Maunsell JH, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey: II. Binocular interactions and sensitivity to binocular disparity. J Neurophysiol 49:1148–1167. Medline

Movshon JA, Newsome WT (1996) Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J Neurosci 16:7733–7741. Medline

Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1985) The analysis of moving visual patterns. In: Pattern recognition mechanisms (Chagas C, Gattass R, Gross C, eds), pp 117–151. New York: Springer.

Nowlan SJ, Sejnowski TJ (1995) A selection model for motion processing in area MT of primates. J Neurosci 15:1195–1214. Medline

Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science 249:1037–1041. CrossRef Medline

Pack CC, Gartland AJ, Born RT (2004) Integration of contour and terminator signals in visual area MT of alert macaque. J Neurosci 24:3268–3280. CrossRef Medline

Patterson CA, Wissig SC, Kohn A (2014) Adaptation disrupts motion integration in the primate dorsal stream. Neuron 81:674–686. CrossRef Medline

Peng Q, Shi BE (2014) Neural population models for perception of motion in depth. Vision Res 101:11–31. CrossRef Medline

Perrone JA (2006) A single mechanism can explain the speed tuning properties of MT and V1 complex neurons. J Neurosci 26:11987–11991. CrossRef Medline

Perrone JA, Thiele A (2002) A model of speed tuning in MT neurons. Vision Res 42:1035–1051. CrossRef Medline

Ponce CR, Lomber SG, Born RT (2008) Integrating motion and depth via parallel pathways. Nat Neurosci 11:216–223. CrossRef Medline

Qian N, Andersen RA (1994) Transparent motion perception as detection of unbalanced motion signals: II. Physiology. J Neurosci 14:7367–7380. Medline

Qian N, Andersen RA (1995) V1 responses to transparent and nontransparent motions. Exp Brain Res 103:41–50. Medline

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

Rodman HR, Albright TD (1989) Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). Exp Brain Res 75:53–64. Medline

Rokers B, Czuba TB, Cormack LK, Huk AC (2011) Motion processing with two eyes in three dimensions. J Vis 11:1–19. CrossRef Medline

Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. Nat Neurosci 9:1421–1431. CrossRef Medline

Sabatini SP, Solari F (2004) Emergence of motion-in-depth selectivity in the visual cortex through linear combination of binocular energy complex cells with different ocular dominance. Neurocomputing 58:865–872.

Sanada TM, DeAngelis GC (2014) Neural representation of motion-in-depth in area MT. J Neurosci 34:15508–15521. CrossRef Medline

Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. Vision Res 38:743–761. CrossRef Medline

Smith MA, Majaj NJ, Movshon JA (2005) Dynamics of motion signaling by neurons in macaque area MT. Nat Neurosci 8:220–228. CrossRef Medline

Stoner GR, Albright TD (1992) Neural correlates of perceptual motion coherence. Nature 358:412–414. CrossRef Medline

Tailby C, Majaj NJ, Movshon JA (2010) Binocular integration of pattern motion signals by MT neurons and by human observers. J Neurosci 30:7344–7349. CrossRef Medline

Thiele A, Dobkins KR, Albright TD (2000) Neural correlates of contrast detection at threshold. Neuron 26:715–724. CrossRef Medline

Tsui JM, Hunter JN, Born RT, Pack CC (2010) The role of V1 surround suppression in MT motion integration. J Neurophysiol 103:3123–3138. CrossRef Medline

Zeki SM (1974) Cells responding to changing image size and disparity in the cortex of the rhesus monkey. J Physiol 242:827–841. CrossRef Medline