The neurophysiology of stereoscopic vision

Sindre Henriksen

Doctor of Philosophy
Institute of Neuroscience
Newcastle University

September, 2017

Abstract

Many animals are able to perceive stereoscopic depth owing to the *disparity* information that arises from the left and right eyes' horizontal displacement on the head. The initial computation of disparity happens in primary visual cortex (V1) and is largely considered to be a correlation-based computation. In other words, the computational role of V1 as it pertains to stereoscopic vision can be seen to roughly perform a binocular cross-correlation between the images of the left and right eyes. This view is based on the unique success of a correlation-based model of disparity-selective cells – the binocular energy model (BEM). This thesis addresses two unresolved challenges to this narrative. First, recent evidence suggests that a correlation-based view of primary visual cortex is unable to account for human perception of depth in a stimulus where the binocular correlation is on average zero. Chapters 1 and 2 show how a simple extension of the BEM which better captures key properties of V1 neurons allows model cells to signal depth in such stimuli. We also build a psychophysical model which captures human performance closely, and recording from V1 in the macaque, we then show that these predicted properties are indeed observed in real V1 neurons. The second challenge relates to the long-standing inability of the BEM to capture responses to anticorrelated stimuli: stimuli where the contrast is reversed in the two eyes (e.g. black features in the left eye are matched with identical white features in the right eye). Real neurons respond less strongly to these stimuli than model cells. In Chapter 3 and 4, we make use of recent advances in optimisation routines and exhaustively test the ability of a generalised BEM to capture this property. We show that even the best-fitting generalised BEM units only go some way towards describing neuronal responses. This is the first exhaustive empirical test of this influential modelling framework, and we speculate on what is needed to develop a more complete computational account of visual processing in primary visual cortex.

4

Acknowledgements

Contents

List of Figures

Chapter 0

Introduction

0.1 The stereo correspondence problem

Stereoscopic vision is possible because the left and right eyes receive slightly different images of the world. This geometric arrangement gives rise to retinal disparity which can be used to extract depth information from a visual scene. A key challenge in stereo vision is determining which elements in the image seen by the left eye correspond to the elements in the image seen by the right eye. This computationally demanding task is known as the stereo correspondence problem and has been extensively studied [Julesz, 1971, Marr, 1980, Ohzawa et al., 1990, Howard and Rogers, 1995, Backus et al., 2001, Prince et al., 2002a, Banks et al., 2004, Parker, 2007, Henriksen et al., 2016c]. The challenge posed to the brain is particularly evident in a class of stimuli known as random dot stereograms (RDSs, Figure 0.1) [Julesz, 1971]. Random dot stereograms are made up of black and white dots, and the images are identical in the left and right eyes up to a translation of a subset of the pixels. Random dot patterns are interesting because for any given dot in the left eye, there is a very large number of dots in the right eye which it could in principle correspond to. Nevertheless, the brain effortlessly achieves a globally consistent solution to the correspondence problem, and the *disparity* information allows human observers to see complex depth-defined objects in otherwise featureless images. Figure 0.1, for example, contains a silhouette of a stag which is only seen when the left and right images are appropriately fused. Notably, when the contrast of the Julesz-style random dot stereogram is reversed, such that white dots in the left eye correspond to black dots in the right eye, and vice versa, the perception of depth is abolished. These stimuli are known as anticorrelated random dot stereograms [Julesz, 1971, Cogan et al., 1993].

Modern versions of the random dot stereograms often paint black and white dots on a gray background, and use simple objects in depth. In Figure 0.2, the depth object is a disk which either appears in front of or behind the background. These stimuli allow for more complex stimulus manipulations with which to probe the mechanisms of correspondence.
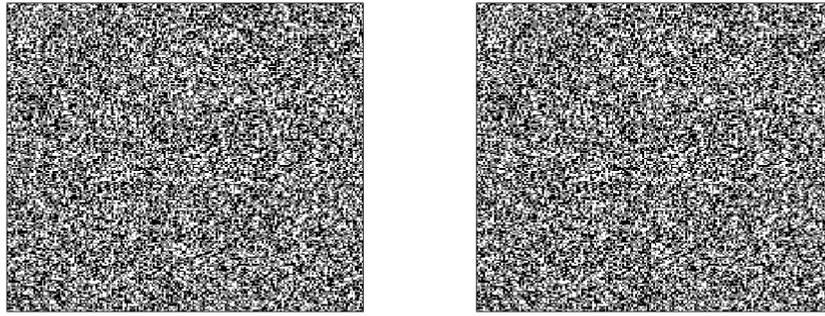
Figure 0.1: A random dot stereogram. When viewed separately, no discernible features are seen. However, when fused, even complex objects can appear in vivid depth. In the above example, a silhouette of a stag can be seen in a separate plane to the background.
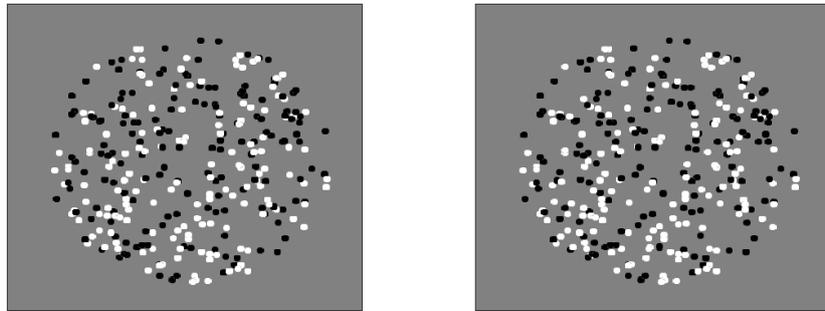
Figure 0.2: Another version of the random dot stereogram, where black and white dots are painted on a gray background. When fused, a stereo-intact observer will perceive a disk either in front of or going into the background. Many modern psychophysical studies use these random dot stereograms to study stereopsis. A stereo intact observer can perceive a central disk in front of the background by crossing their eyes such that the left eye sees the right-hand image and vice versa.

0.2   Mechanisms of correspondence

Julesz's random dot stereograms have been incredibly influential in shaping thinking about stereopsis and correspondence. One of the key contributions of Julesz and his stimuli was that the perception of depth in these stimuli unequivocally showed that stereo vision must be a low-level process. Up until the introduction of the random dot stereogram, researchers generally considered stereopsis to be a high-level cognitive process [Howard, 2002]. The realisation that correspondence occurs at a very early stage of visual processing had the knock-on effect of focusing stereo research on the underlying mechanisms of stereopsis [Julesz, 1971, Marr and Poggio, 1976, Marr and Poggio, 1979], rather than simply investigating the properties of the phenomenon [Von Helmholtz, 1881, Ogle, 1952b, Ogle, 1952a]. Marr & Poggio (1976, 1979) published a series of influential papers, where they proposed algorithms for obtaining stereo correspondence solutions. These early algorithms were generally based on matching image features in the two eyes. For example, in the

algorithm proposed by Marr & Poggio (1979), linear filters obtain zero-crossings of the left and right images. These roughly correspond to the edges of objects, but can be produced by any large local change in luminance. The zero-crossings are then matched across the left and right eyes, and a disparity estimate is obtained through a coarse-to-fine refinement, where the coarse spatial filters constrain the disparity solution space of the fine spatial filters. These models have been greatly influential in both human vision science and computer vision, and while they have been largely discounted as veridical accounts of human stereopsis, certain components of the models, such as pooling across spatial frequencies, have been shown to take place in disparity-selective cells in V1 [Baba et al., 2015, Tanabe et al., 2011a].

The earliest observations of neurons specialised for stereopsis came from Barlow, Blakemore, & Pettigrew (1967) and Pettigrew, Nikara, & Bishop (1968), who reported disparity-selective cells in the primary visual cortex (V1) of cats. These neurons modulate their firing rate as a function of the disparity of the stimulus [Barlow et al., 1967, Pettigrew et al., 1968], and have since been found in the macaque [Hubel and Wiesel, 1970, Poggio and Fischer, 1977], the barn owl [Pettigrew and Konishi, 1976, Nieder and Wagner, 2000], sheep [Clarke et al., 1976], and their existence has been inferred in humans through psychophysical [Read and Eagle, 2000, Kane et al., 2014], electroencephalographic [Norcia et al., 1985], and neuroimaging [Backus et al., 2001] approaches.

0.3   Neuronal models for computing disparity

Ohzawa, DeAngelis, & Freeman (1990) proposed the first computational model of disparity-selective cells in primary visual cortex. This model - the binocular energy model (BEM) - has been incredibly successful in capturing a range of computational properties of disparity-selective V1 neurons. The BEM is the binocular cousin of the more well-known motion energy model [Adelson and Bergen, 1985], and belongs to a large family of models known as linear-nonlinear (LN) models. LN models locally filter the image using either spatial or spatiotemporal linear filters, before applying an output nonlinearity to the response. For the purposes of computer simulations, a useful convention is using vector notation to denote the linear filter and the stimulus, and then define the filtering process as the dot product:

$$v = \boldsymbol{\rho} \cdot \mathbf{s}, \tag{0.1}$$

where $\boldsymbol{\rho}$ and $\mathbf{s}$ refer to the linear filter and stimulus, respectively, represented as vectors, and $\cdot$ denotes the dot product. This form is agnostic about whether the linear filter is spatial or spatiotemporal. If the stimulus vector is spatial, then the vector's entries represent the stimulus at various positions on the retina. If the stimulus vector is spatiotemporal, then each entry represents a luminance value at a particular time and retinal position. The linear filter $\boldsymbol{\rho}$ is roughly analogous

to the receptive field of real neurons. The response of the LN unit is obtained by passing v through a nonlinearity:

$$s = f(v). \tag{0.2}$$

f can take on a range of forms, but common choices are half-wave rectification ($f(v) = \text{Pos}(v)$), half-squaring ($f(v) = [\text{Pos}(v)]^2$), or simply squaring ($f(v) = v^2$). These models are very similar to the units used in modern neural networks, where sigmoids or rectified linear units (RELUs) are generally used [Rumelhart et al., 1986, Nair and Hinton, 2010].

The binocular energy model has filters in the left and right eyes, and selectivity for non-zero disparities is generated by having the left and right filters centred on non-corresponding regions of the left and right visual field (for non-zero disparities). To compute the response of the model, the left and right filter responses are summed before passing the sum through a squaring nonlinearity [Ohzawa et al., 1990, Read et al., 2002].

$$s = (v_L + v_R)^2 = v_L^2 + v_R^2 + 2v_L v_R. \tag{0.3}$$

The subscripts here refer to the left and right eyes. The $v_L^2 + v_R^2$ term means that the model will respond independently to stimulation in either the left or the right eye. This mimics the behaviour observed in real V1 neurons [Ohzawa et al., 1990], many of which can be driven by visual stimulation in either eye. The model's disparity selectivity arises from the cross-term $2v_L v_R$. The BEM has two binocular subunits, each analogous to a simple cell, which together form a model complex cell. The response of complex cells is invariant to stimulus phase in either eye (i.e. they respond equally well to a black bar as they do to a white bar). In order to achieve phase invariance, the BEM uses two binocular simple cells whose linear filters are in quadrature, so that the left linear filter of the first subunit $\boldsymbol{\rho}_{1L}$ is orthogonal to the left linear filter of the second subunit $\boldsymbol{\rho}_{2L}$, and vice versa for the right eye filters. The complete response of the binocular energy model is therefore

$$
\begin{aligned}
C &= (v_{1L} + v_{1R})^2 + (v_{2L} + v_{2R})^2 \\
&= v_{1L}^2 + v_{1R}^2 + v_{2L}^2 + v_{2R}^2 \\
&\quad + 2(v_{1L}v_{1R} + v_{2L}v_{2R}).
\end{aligned} \tag{0.4}
$$

The invariance to stimulus phase arises because the terms $v_{1L}^2 + v_{2L}^2$ and $v_{1R}^2 + v_{2R}^2$ guarantee constant response to equal magnitude stimuli with different phase (in the same way that $\cos^2(x) + \sin^2(x) = 1$).

Figure 0.3 shows a schematic representation of the BEM. Two types of retinal disparity are seen in cortex: position disparity and phase disparity. Position disparity is defined as a position offset
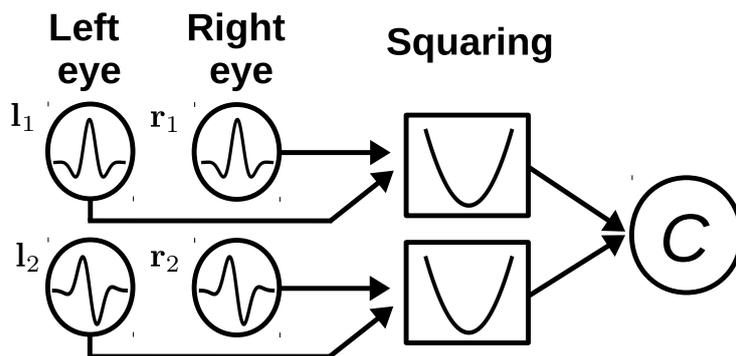
Figure 0.3: Schematic illustration of the binocular energy model. The two left-most columns show the left and right eye linear filters, respectively, and the two rows correspond to the two orthogonal simple cell subunits of the BEM. The left and right image is filtered by the left and right filters, respectively. The filtered response is then summed, passed through a squaring nonlinearity before the responses of the two simple cells ubunits are added to produce the model complex cell response.

between the filters in the two eyes without a change in the profile of the filters. Phase disparity is when the centres are the same across the two eyes, but the profiles (phase) of the left and right filters are different. The binocular energy model has been widely successful because of its ability to capture a range of properties of real cells, including both phase and position disparity. One key success of the BEM is the observation that although V1 complex cells do not modulate their response to stimulus phase in either eye, they do modulate their response to the *binocular phase difference* [Ohzawa and Freeman, 1986]. The reason for this in the BEM can be seen by inspecting Equation 0.4: the cross-terms that produce the disparity modulation are $v_{1L}v_{1R}$ and $v_{2L}v_{2R}$, and these do depend on the phase difference in the left and right image. Another key property of the BEM is that inverting the contrast in one eye – thereby inverting the correlation – also inverts the disparity tuning curve of the model units. This was shown experimentally to be the case, first by Ohzawa et al. (1990) in the cat using individual bars, and later by Cumming & Parker (1997) in the macaque. Cumming & Parker (1997) also showed that while the original BEM predicts a perfect inversion of the disparity tuning curve, real cells also have reduced amplitude to anti-correlated stimuli (Figure 0.4). Because anticorrelated stimuli represent an impossible binocular configuration, this attenuation represents a suppression of false match responses, and suggests that these cells are specialised for the binocular statistics of the natural world [Cumming and Parker, 1997, Tanabe et al., 2011a].

A large corpus of research has attempted to resolve the discrepancy between the binocular energy model and real neurons. Initial attempts used simple extensions, such as appending an output nonlinearity to the BEM [Nieder and Wagner, 2000]. However, an output nonlinearity cannot account for the attenuated response of odd-symmetric cells, so it is at best an incomplete
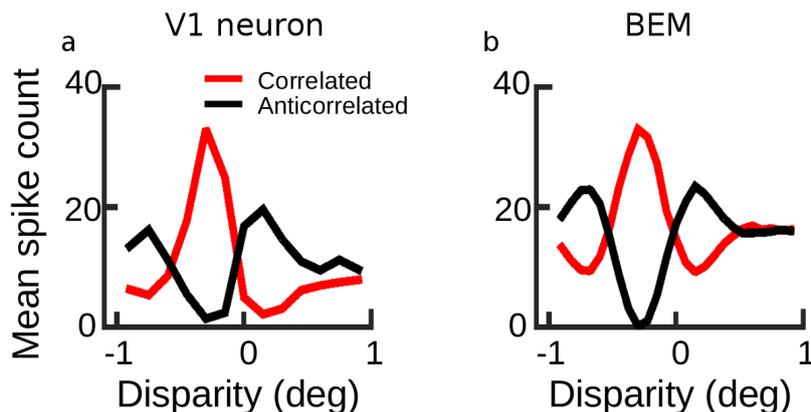
Figure 0.4: Example disparity tuning curve, similar to that reported in Cumming & Parker (1997). The cell (a) shows clear disparity tuning to both correlated (red) and anticorrelated (black) stimuli, but its disparity tuning curve inverts to anticorrelation. The BEM predicts an exact inversion (b), while the cell has an attenuated response to anticorrelated stimuli.

description of V1 disparity selectivity [Read et al., 2002]. Other efforts have focused on modifying the architecture of the BEM. Read, Parker, & Cumming (2002) showed that a model which passes the monocular filter responses through a nonlinearity *prior* to binocular combination can account for a range of disparity tuning curves [Read et al., 2002], including the attenuated response to anticorrelated stimuli. However, a key prediction made by applying such nonlinearities prior to binocular combination is that the temporal modulation frequency in response to drifting gratings should double. This is only observed in some cells [Read et al., 2002], and so most modelling approaches have since maintained the condition of linear binocular combination [Haefner and Cumming, 2008, Tanabe and Cumming, 2008, Tanabe et al., 2011a]. Haefner & Cumming (2008) proposed a two-subunit (2SU) model to account for odd-symmetric cells [Haefner and Cumming, 2008]. It does so by combining two BEM-like cells with different disparity tuning, where neither of the subunits themselves have phase disparity. Tanabe & Cumming (2008) tested the predictions of the 2SU model by varying the spatial frequency and phase components of broadband spatial gratings, and showed that the behaviour of V1 neurons agree better with the predictions of the 2SU model than that of the BEM [Tanabe and Cumming, 2008]. Further expanding on the 2SU framework, Tanabe & Cumming (2011) carried out a spike-triggered analysis of covariance; if the underlying generative process is well-described by a generalised quadratic model, this approach will recover filter elements which span the same subspace as the underlying neuronal subunits [Tanabe et al., 2011a]. Tanabe & Cumming (2011) confirmed that many neurons are indeed composed of many more LN elements than that suggested by the BEM, and also often vary in their position disparity. Interestingly, the authors also found that neurons in V1 are commonly arranged in a push-pull architecture: suppressive units have the opposite response profile to excitatory subunits, meaning excitatory subunits are driven by the same disparities that suppress the response of the suppressive subunits (i.e. disinhibition). The suppressive subunits additionally have the

interesting characteristic of a broader spatial profile than the excitatory subunits. The spike-triggered covariance method demands that the stimulus is independent white noise - independent from pixel to pixel and frame to frame, as well as between the left and right eyes. However, for this stimulus, neither the neurons nor the model units exhibited the characteristic attenuation to anticorrelated frames, and the authors were thus unable to comment on the ability of this framework to capture the response attenuation to anticorrelated stimuli. The origin of the response attenuation to anticorrelated stimuli remains an outstanding question in the field.

0.4 V1 activity and depth perception

Activity in primary visual cortex is a poor correlate of stereoscopic depth perception. Two key observations demonstrate this fact. First, V1 neurons show tuning only to absolute disparity, whereas psychophysically, humans have much greater sensitivity to *relative* disparity [Cumming and Parker, 1999]. Second, while V1 neurons modulate their response to anticorrelated stimuli, depth is only perceived here under very specific circumstances. In the traditional Julesz-style random dot stereograms, depth is never perceived in anticorrelated stimuli. However, Cogan et al. (1993) showed that human observers perceive *veridical* depth in sparse anticorrelated random dot stereograms (although observers have very high thresholds) [Cogan et al., 1993]. More commonly, reversed depth is reported in anticorrelated stimuli. For example, Read & Eagle (2000) showed that reversed depth is perceived by human observers using both 1D and 2D random noise patterns, with a stronger effect being observed for the 1D case [Read and Eagle, 2000]. Tanabe, Ysauoka, & Fujita (2008) noted that reversed depth discriminations can be observed in random dot stereograms where black and white dots are painted on a gray background [Tanabe et al., 2008]. The authors note that the condition for depth perception appears to be that there must be a clear reference plane against which the observers can make the depth judgement. More recently, Hibbard et al. (2014) reported data from a large number of participants suggesting that no reversed depth is seen in anticorrelated random dot stereograms [Hibbard et al., 2014]. However, Hibbard et al. introduced an empty zone between the disparity-defined region and the reference plane (the background), which may have prevented the observers from using the annulus as a point of reference. In general, the literature on depth perception in anticorrelated stimuli is inconsistent, with depth only being reported in specific circumstances. The neuronal mechanisms which lead to reversed depth judgements are not understood. The failure to understand how and when anticorrelation gives rise to reversed depth judgements comes partially from an inadequate understanding of how V1 activity is read out in real networks.

0.5 Disparity processing in higher cortical areas

Cells in higher cortical areas likely inherit their disparity tuning from cells in primary visual cortex. Tanabe & Cumming (2008) showed that neurons in V2 are consistent with the 2SU model discussed

previously. A substantial number of V2 neurons also show tuning to relative disparity, which is absent in V1 [Thomas et al., 2002]. While the anticorrelated modulation in V2 is comparable to that in V1 [Tanabe and Cumming, 2008], there appears to be progressively weaker modulation to anticorrelated stimuli along the ventral pathway. In V4, there is clear evidence that responses to anticorrelated stimuli are greatly weakened compared to the modulation seen in V1 [Tanabe et al., 2004]. By the time the signal reaches IT, there is no significant neuronal modulation to anticorrelated stimuli [Janssen et al., 2003]. The character of disparity tuning appears to be different in the ventral and dorsal streams. Along the dorsal stream, tuning for anticorrelated stimuli is strongly maintained even in MT and MST [Krug et al., 2004, Takemura et al., 2001], prompting the suggestion that the ventral and dorsal pathways might be differentially specialised for different aspects of binocular vision [Parker, 2007].

0.6   Aims and outline

This thesis has two key aims. Recent work by Doi et al. (2011, 2013, 2014) has found that humans can perceive depth in a class of stimuli with mixed correlated and anticorrelated dots (mixed correlation random dot stereograms). When there is an equal number of correlated and anticorrelated dots (half-matched stereograms), the average correlation of the stereogram is zero because the correlated and anticorrelated dots cancel out. In these stimuli, the BEM does not signal disparity, and Doi et al. have suggested that a separate disparity computation extracts disparity from these images. The implication of this work is that a substantial revision in our understanding of the role of V1 binocular neurons in stereoscopic depth perception is needed. The first aim is therefore to examine whether such a major revision is necessary. While Doi et al. show how a binocularly linear model such as the BEM cannot signal depth in these stereograms, this may not imply that V1 neurons cannot signal depth. This is possible precisely because the responses of real neurons to anticorrelated RDSs do not follow the predictions of the BEM. Evaluating whether a revision of the role of V1 in depth perception is necessary is therefore the focus of Chapters 1 and 2. Chapter 1 begins by exploring the recent literature on the subject, and then examines how modified BEM units, which more closely approximate key aspects of the disparity tuning properties of real cells, can account for depth perception in mixed correlation stereograms. We confirm that they can, and that the existing psychophysical literature is consistent with the theoretical properties of modified BEM neurons. The work reported in Chapter 1 is published in Henriksen, Cumming, & Read, *PLoS Comput. Biol.* (2016). Chapter 2 follows on from this and examines whether neurons in primary visual cortex do in fact respond to half-matched stereograms, as our model would predict. We find that neurons in macaque V1 agree well with the properties of our modified BEM units. We conclude that the psychophysical evidence is consistent with known properites of V1 neurons, and that the conceptual revision offered by Doi et al. is unnecessary. The work reported in Chapter 2 is published in Henriksen, Read, & Cumming, *J. Neurosci.* (2016).

The binocular energy model and related neuronal models are central for the field's thinking about stereopsis as a perceptual process. The well-established limitations of the BEM means that we need a better model for understanding the foundational units which facilitate stereopsis (i.e. V1 neurons). However, a better model which can capture properties such as the anticorrelated attenuation in real cells has so far remained elusive. In order to model disparity-selective cells, the field has traditionally used models where the model architecture and parameters are chosen by hand. Recent developments in optimisation routines now allow for the automatic fitting of architectures and model parameters. We are thus in the unique position to test not only whether a particular model gives an adequate description of a particular disparity-selective cell, but rather whether this entire class of models can give an adequate description of a population of disparity-selective cells. The second aim of this thesis is therefore to attempt to develop a full model of V1 disparity selectivity through data-driven (machine learning) modelling efforts. In Chapter 3, we show that even the best-fitting generalisations of the binocular energy model (GBEM) cannot capture key characteristics of real V1 neurons. Specifically, we show that what the field has generally thought of as an attenuated response to anticorrelated stimuli is in fact an enhanced response to correlated stimuli at the preferred disparity of the cell. We call this mechanism correlated boosting. Importantly, real cells exhibit much stronger correlated boosting than model units, suggesting that this arises through a separate mechanism than the one used by the LN framework to create disparity-selectivity. In Chapter 4, we analyse stimulus-driven response variability ("external variance") in the BEM, GBEM, and real cells. Two key results emerge. First, correlated boosting also affects external variance and results in an underestimate of the external variance at the preferred disparity of the cell. Second, real V1 neurons have higher baseline external variance than GBEM units. This means that they modulate their response more strongly to features in the stimulus other than disparity. We discuss the implications of these findings for our understanding of V1 neurons, and chart out potential modifications to existing modelling frameworks to help bridge the gap between data and theory.

Chapter 1

A single mechanism can account for human perception of depth in mixed correlation random dot stereograms

1.1 Introduction

In a recent series of papers, Doi et al. (2011, 2013, 2014) have proposed that two distinct computations contribute to depth perception in cyclopean stimuli [Doi et al., 2011, Doi et al., 2013, Doi and Fujita, 2014]. They postulate a pure correlation mechanism, which depends linearly on interocular correlation, plus an additional "matching" computation, which in their most recent work [Doi and Fujita, 2014] they have suggested may simply be the correlation mechanism plus an additional output nonlinearity. Under some circumstances, this is identical to a single mechanism with a nonlinear response to correlation (Figure 1.1). However, Doi et al. propose that the two mechanisms have distinct spatiotemporal integration properties, and so may be differentially activated by different stimuli. These conclusions are motivated by a series of ingenious psychophysical experiments in which the authors mixed correlated and anticorrelated dots within a single random dot stereogram (RDS). When half the dots in an RDS are correlated and half are anticorrelated (half-matched RDSs; Figure 1.2a), the global binocular correlation of the stimulus is 0. In this case, the authors argue, a pure correlation computation should not be able to detect depth. However, humans can perceive depth in such stimuli [Doi et al., 2011, Doi et al., 2013]. Doi et al. argue that this cannot be explained by a pure correlation mechanism, and propose an additional matching mechanism to account for these data. In addition to depth perception in half-matched RDSs, two pieces of evidence suggest that two separate mechanisms extract disparity in random dot stereograms. Doi et al. (2011) have reported that larger disparities tend to lead to decreased performance for half-matched RDSs and more reversed depth responses to anticorrelated RDSs [Doi et al., 2011]. In a subsequent publication, Doi et al. (2013) reported a similar phenomenon in the temporal domain [Doi et al., 2013]. They investigated dynamic random-dot patterns, in which the dot pattern is periodically replaced with a new random pattern with the same disparity and correlation. They showed that faster dot pattern refresh rates lead to poorer half-matched judgments and more reversed depth responses to anticorrelation. The authors argue that these results again reflect the weighted contribution of two separate mechanisms: one slow matching
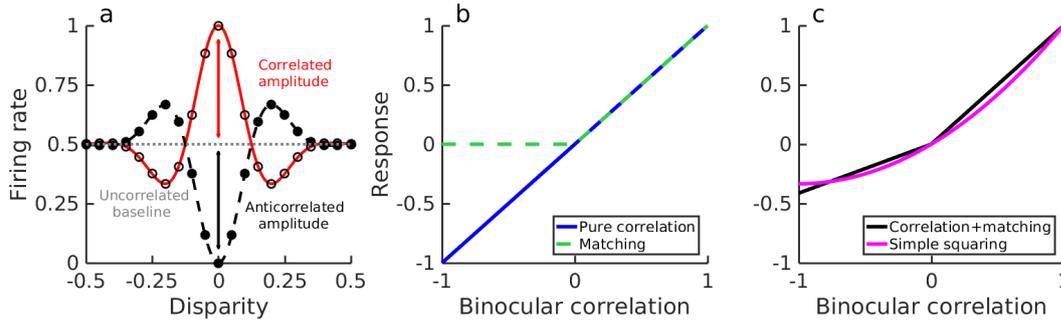
Figure 1.1: Response amplitudes in correlated and anticorrelated stereograms. Binocular energy model response to correlated (open circles, red line) and anticorrelated stimuli (solid circles, black line). The red arrow denotes the correlated amplitude, the black arrow the anticorrelated amplitude and the gray line the uncorrelated baseline response of the model cell. The amplitude ratio is the anticorrelated amplitude (black line) divided by the correlated amplitude (red line). b) The response at the preferred disparity as a function of binocular correlation (blue) which is linear, hence we call this a "pure" correlation computation. The green line shows this response after rectifying in the correlation domain - proposed as a separate computation by Doi & Fujita (2014). c) shows a linear combination of the two lines in b) - a single computation passed through a different nonlinearity (black). As shown, the two computations in b) (i.e. blue and green lines) and the linear combination of the two in (black line in c) are indistinguishable descriptions: b) contains two mechanisms, but their combined response is identical to the black line in c). If the two components in b) differ in some other respects (spatial or temporal properties) it may be possible to demonstrate that a single mechanism cannot reproduce the same behavior. No study has yet attempted to describe behavioral data with a single mechanism like c). We explore a mechanism based on a slightly simpler nonlinearity - squaring the response of the model in a (magenta line in c).

computation, responsible for fine disparity discrimination, and one rapid correlation computation, responsible for coarse disparity discrimination. Figure 1.2b illustrates schematically the performance they expect from these two computations in isolation.

While it is true that half-matched RDSs – stereograms with equal numbers of correlated and anticorrelated dots – have a mean binocular correlation of 0, it is possible that local fluctuations in correlation could be exploited to determine the stimulus disparity [Doi et al., 2013, Doi and Fujita, 2014]. Doi et al. propose that these fluctuations are used by the matching computation, possibly in extrastriate cortex [Doi et al., 2011, Doi et al., 2013]. However, the attenuated responses of V1 neurons to anticorrelated dots [Cumming and Parker, 1997] makes it possible that even V1 neurons could encode disparity in these stimuli. V1 neurons respond more strongly to positive than negative binocular correlation [Cumming and Parker, 1997]. If this attenuated response was generated by a simple output nonlinearity, then their responses to stimuli with high correlation variability may be greater than predicted from the mean correlation alone. In other words, the mean response of the cell may depend on the local correlation variability as well as the mean correlation. Doi & Fujita (2014) explore this with a modified version of a cross-correlation computation which they refer to as "cross-matching". Cross-matching computes the correlation between left
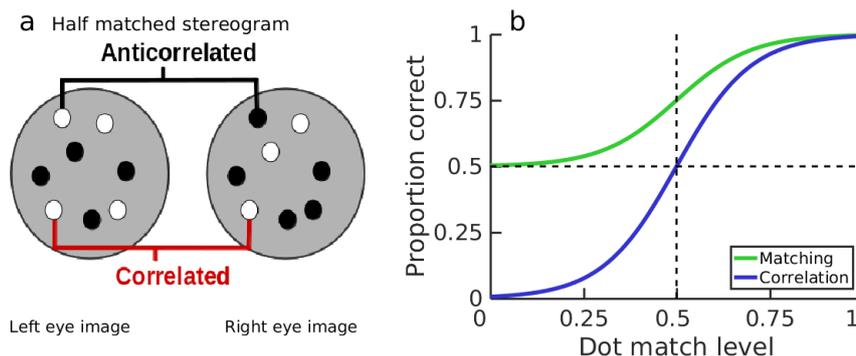
Figure 1.2: Correlation and matching computations in random dot stereograms. a) A schematic representation of a half-matched random dot stereogram. Half the dots are correlated, half the dots are anticorrelated, resulting in a binocular correlation of 0. b) Psychometric functions reflecting the operation of hypothetical matching and pure correlation computations (recreated from Doi et al., 2011) as a function of dot match level. The pure correlation computation (blue line) signals reversed depth at a dot match level of 0 (fully anticorrelated), is at chance at a dot match level of 0.5 (half-matched) and performs perfectly at a dot match level of 1. The matching computation (green line) is at chance for anticorrelated stimuli and gradually increases to perfect performance with increasing percentage of matched dots. The horizontal line shows chance performance, and the vertical line marks a dot match level of 0.5, e.g. the half-matched stereogram shown in (a).

and right images, but then follows this by half-wave rectification. Doi & Fujita (2014) conclude that cross-matching has the necessary properties to serve as the computation underlying their putative match-based computation. Importantly, they still postulate that human stereo vision is subserved by separate pure correlation and match-based computations, with different spatiotemporal properties, and whose contribution to perception vary with properties of the stimulus. The proposition that these two mechanisms have different spatiotemporal properties is crucial if the two-mechanism model is to be distinguished from a single mechanism intermediate between the pure correlation and pure matching models. While differences in spatiotemporal properties are essential in order to separate "pure correlation" and "cross-matching", changes in psychophysical performance with changes in spatiotemporal properties of the stimulus do not necessarily imply that there must be two mechanisms. In principle, changes in performance could be due to changes in the stimulus, even with a single-mechanism model. However, no one has yet explored whether a single-mechanism model can account for the results of Doi et al.

Here we explore the possibility that a single computation can explain depth perception in correlated, anticorrelated, and half-matched random-dot stereograms. Like Doi & Fujita (2014), we use a model that can describe the attenuation observed in V1 neurons – a binocular energy model with an additional output nonlinearity – and show that this can explain responses to half-matched stereograms. Our scheme differs from theirs in that it does not suppose two distinct computations operating in cortex, but rather uses a single mechanism to explain all the psychophysical data. To explore this model, we first investigate the model responses to half-matched RDSs. We then show

that this model can account for a range of previously documented psychophysical phenomena, including effects which Doi et al. have suggested are diagnostic of either a pure correlation or a match-based computation [Doi et al., 2011, Doi et al., 2013, Doi and Fujita, 2014]. Finally, we confirm two new predictions made by the model in human observers: that psychophysical performance should become worse with 1) decreased dot size in half-matched stereograms and 2) in response to rapid temporal modulation in correlation.

1.2   Methods

All simulations were implemented in BEMtoolbox - a custom Matlab toolbox for simulating binocular neurons. The toolbox is available at `http://github.com/sidh0/BEMtoolbox`. All code used in the current manuscript is available online at `http://github.com/sidh0/hcr16_ploscb` (requires BEMtoolbox).

1.2.1   Random dot stereograms

We created dynamic random-dot stereograms with a varying number of correlated and anticorrelated dots as described by Doi et al. (2011, 2013). Black and white circular anti-aliased dots, $0.09°$ in radius, were painted on a gray background. When the stimulus was half-matched, half the dots were correlated, i.e. had the same luminance in both eyes, while the other half were anticorrelated, i.e. drawn black in one eye and white in the other. There were on average equal numbers of black and white dots for each correlation value. Dots had zero binocular disparity except within the central $2.5°$ of the stimulus. The surrounding annulus had a width of $1°$. The stimulus thus depicted a disparate disk either in front of or behind a zero-disparity background. No subpixel disparity was used. Unless otherwise specified, the dot density was 24%, meaning that if none of the dots overlapped, they would have occupied 24% of the stimulus area. The dots were, however, allowed to occlude one another. The dots were painted in random order so as to prevent any cues arising from occlusion due to either correlated dots systematically occluding anticorrelated dots or the surround dots systematically occluding the center dots. The above applies to both the human psychophysics and the simulations. For the psychophysical investigations, the background was kept 100% correlated for consistency with Doi et al. (2011, 2013). For all simulations, 292×292 pixels were used to simulate the stimulus such that 1 pixel corresponded to $0.03°$. When time was incorporated into the model, the simulations were carried out at a temporal resolution of 1ms.

For the dot size experiment (Figure 1.10), we created half-matched (binocular correlation of 0, match level of 0.5) and correlated RDSs (binocular correlation of 1, match level of 1). We used three different dot sizes: $0.025°$, $0.05°$, and $0.075°$. The dots were circular and anti-aliased. The surrounding annulus was always correlated as per Doi et al. (2011, 2013). The stimuli were presented at a dot pattern refresh rate of 21.25Hz, for 500ms, and at a disparity of $\pm0.075°$. All
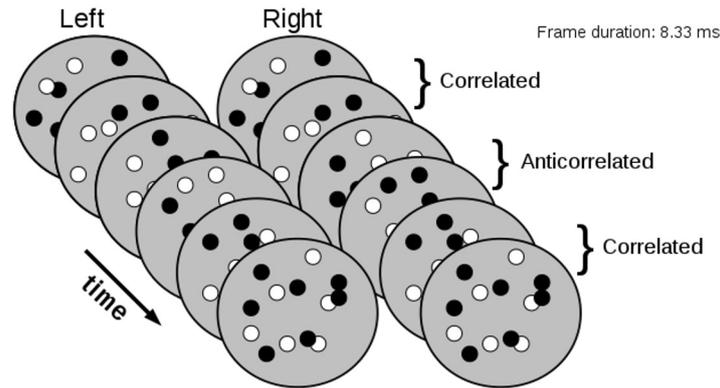
Figure 1.3: Alternating-correlation random dot stereograms. In the illustration, the observers first see two correlated frames, then two anticorrelated frames, then two correlated frames and so on for 500ms. The dot pattern is updated at 120Hz with the correlation alternation rate varying across trials.

other features of the RDSs were as previously described.

1.2.2   Alternating-correlation random dot stereograms

Instead of manipulating binocular correlation over space, as in the half-matched stereogram, we manipulated the binocular correlation over time (Figure 1.3). The stimulus was presented at a constant refresh rate – 120Hz – meaning that a new frame of the dynamic RDS was generated every 8.33ms. The key manipulation was how often the binocular correlation of the dynamic RDS flipped. This could either be at 60, 30, 15, 7.5 or 3.75Hz. At 60Hz, a correlation alternation cycle is completed after two frames, at 30Hz after four frames and so on. Whether a trial started with a correlated or an anticorrelated frame was randomized. For the alternating-correlation RDSs, the surround had the same correlation as the disparity-defined region, meaning that the surround changed correlation with the centre, although its dispariy remained fixed at $0°$. We used a dot density of 200% and six disparities: $\pm 0.2275°$, $\pm 0.1365°$, and $\pm 0.0455°$. The stimulus was presented for 500ms (60 frames) and following the presentation, the observers were asked to report whether the central disk appeared near or far relative to the background using a mouse press. The stimulus is illustrated in Figure 1.3.

1.2.3   Psychophysics experimental procedure

The stimuli were generated in Matlab and displayed using Psychtoolbox [Kleiner et al., 2007]. The stimuli were displayed on a 19" Dell Trinitron CRT monitor. For the dot size experiment, the refresh rate of the monitor was 85Hz and the resolution was 1024×768 pixels. The monitor's luminance output was linearized prior to the experiment. For all experiments, the Michelson contrast was $> 99\%$. For the alternating experiment, the refresh rate of the monitor was 120Hz

and the resolution was 800×600. Dichoptic presentation was ensured through the use of a simple four-mirror haploscope. In both experiments, the observers indicated using a mouse button press whether the central disk appeared near or far relative to the background. For the statistical testing of the effect of dot size (Figure 1.10) we used a Monte-Carlo method equivalent to 1-way ANOVA (which cannot be used here since the data are binomial proportions). For each observer and dot size we generated random draws from a binomial distribution with a fixed probability, equal to the mean across dot sizes for that observer. We then measured the variance in proportion correct across dot size, generating a distribution of values compatible with the null hypothesis.

1.2.4 Observers

Four observers participated in the dot size experiment, three of whom were male. Six observers participated in the alternating experiment, four of whom were male. In the alternation experiment, two observers' data were discarded as they were unable to reliably report depth in 100% correlated stereograms. For both experiments, one of the observers was the author, the rest were naive to the purpose of the experiment. All observers had normal or corrected-to-normal vision using spectacles or contact lenses. Both experiments were approved by the Faculty of Medical Sciences ethics committee at Newcastle University.

1.2.5 Binocular energy model units

The energy model has been described in detail elsewhere [Ohzawa et al., 1990, Qian and Zhu, 1997, Read et al., 2002]. Briefly, we modeled the receptive fields of monocular subunits as two-dimensional Gabors with vertical orientation tuning

$$\rho(x,y) = \exp\left(-\frac{(x - x_0 \pm \Delta x/2)^2}{2\sigma_x^2} - \frac{(y - y_0)^2}{2\sigma_y^2}\right) \cos(2\pi f(x - x_0 \pm \Delta x/2) \pm \varphi/2). \qquad (1.1)$$

x_0 and y_0 denote the horizontal and vertical receptive field centers, respectively, Δx denotes horizontal disparity and φ denotes phase. σ_x and σ_y denote the horizontal and vertical extent of the receptive fields, respectively, and f is the frequency of the Gabor. For all simulations carried out here we used $\sigma_x = \sigma_y$, and the receptive field centers x_0 and y_0 were placed randomly within the disparity-defined region of the stimulus. No phase disparity was used in any of the models. For Figure 1.4, 1.5, and 1.6 there was no temporal component of the receptive field. For Figure 1.5b, 1.8, 1.10, and 1.11 we incorporated time by giving each monocular subunit a biphasic temporal kernel as described by Qian and Freeman (2009).

$$\rho_t(t) = \begin{cases} \frac{1}{\Gamma(\alpha)\tau^\alpha} t^{\alpha-1} \exp(-\frac{t}{\tau}) \cos(\omega t + \phi) & \text{if } t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \qquad (1.2)$$

For all simulations where a temporal component was incorporated, we used $\alpha = 2.5$, $\omega = 4 \times 2\pi$, $\varphi = -\pi$, and $\tau = 0.035$, which gives a temporal kernel with peak response at approximately 4.3Hz. We chose a biphasic kernel because most V1 neurons have temporal kernels that are bandpass [Hawken et al., 1996]. In Figure 1.11, we also used the monophasic temporal kernel from Doi et al. (2013). The spatial receptive field and temporal kernel were separable giving

$$\rho(x, y, t) = \rho(x, y)\rho(t) \tag{1.3}$$

Two binocular simple cells were constructed by squaring the sum of two monocular inputs. This produces a binocular simple cell response

$$S = (V_L + V_R)^2 = V_L^2 + V_R^2 + 2V_L V_R \tag{1.4}$$

where V_L and V_R denote the left and right monocular responses, respectively. The disparity tuning of the model arises from the cross-term $2V_L V_R$. The BEM models a complex cell by combining simple cells whose receptive fields are $\frac{\pi}{2}$ out of phase, generating response invariance to stimulus phase. Combining the simple cell responses, we now have $C = S_1 + S_2$. In order to obtain a cell with an amplitude ratio < 1, we added a static squaring output nonlinearity so that our final model is simply C^2. We computed the disparity tuning curve in Figure 1.4 by calculating the mean response of the model to 20 000 images displayed at 21 disparities, spanning the range of disparities covered by the neurons' responses. For generating Figure 1.5a, we computed the correlated and half-matched response of 30 cells, whose RF size, parameterized by σ, was in the range [0.01°, 0.3°]. We used 11 dot density values, logarithmically spaced from 0.01 to 5.12. The dot size was fixed at 0.09°. We computed the response of each of the 30 cells to 20 000 RDSs per density. For Figure 1.5b, we computed the correlated and half-matched responses for each cell to dynamic RDSs of 11 different frequencies, ranging from 1Hz to 100Hz. The RF sizes and dot size were the same as for Figure 1.5a. We obtained the model responses by averaging across 5000 trials per frequency-relative RF size combination, where each trial had a duration of 10s.

1.2.6 Perceptual decision model

Our model population consisted of 160 neurons tuned to four disparities ($\pm 0.48°$, $\pm 0.03°$). Using only a single neuron at each disparity produced significantly poorer performance than human observers, because performance is limited by the fluctuations in image content from trial to trial.

With 40 neurons in each group, and fitted noise levels, performance was comparable to humans. Disparity selectivity was introduced with a position disparity between left and right eyes, with a phase disparity of 0. We made the assumption that receptive field size, parameterized by the standard deviation of the Gaussian envelope σ, scaled with disparity magnitude. Specifically, we had

$\sigma = 0.023 + 0.41|\Delta x|$, where $|\Delta x|$ is the absolute value of the cell's preferred disparity, measured in degrees. This is similar to previous modeling work that incorporate disparity-size correlations [Allenmark and Read, 2011]. The specific parameters we have chosen here are not critical – many different sizes and disparities will yield very similar results for the half-matched stimuli, though the exact shape of the psychometric function varies. The frequency of the monocular Gabors scaled inversely with the receptive field size: $f = 0.3125/\sigma$, in agreement with physiological estimates [Prince et al., 2002b]. The resulting disparity tuning curves, shown in Figure 1.6, were obtained from computing the average response of each cell to 5000 correlated RDSs per disparity. For the tuning curves, we used 51 disparities spaced from $-1°$ to $1°$, with disparities near the peak of the fine tuning curves being sampled more finely. All other stimulus parameters were as previously described.

Gaussian noise was included in the model, where the variance of the noise at any moment in time was proportional to the response of the cell at that time. This approximates the relationship between mean spike count and variance in real cells. The response of the ith neuron to the kth time point is given as

$$P_{i,k} = C_{i,k}^2/\langle \bar{C}^2{}_i \rangle + \kappa \epsilon_{i,k}, \tag{1.5}$$

where $C_{i,k}^2$ is the squared energy model response of the ith neuron to the kth time point, and $\langle \bar{C}^2{}_i \rangle$ is the mean response of the squared energy model cell to correlated RDSs at the preferred disparity of the cell, presented at 21.25Hz. Dividing by the constant scaling factor $\langle \bar{C}^2{}_i \rangle$ ensures that all cells, have the same maximum response to correlated stimuli at their preferred disparity. In other words, a value of 0.5 in this scheme means that the response was half the mean response to correlated RDSs at the cell's preferred disparity. $\epsilon_{i,k} \sim \mathcal{N}(\mu, \sigma^2)$ is the noise in the model, with $\mu = 0$ and $\sigma^2 = C_{i,k}^2/\langle \bar{C}^2{}_i \rangle$ (i.e. the variance is proportional to the response magnitude of the cell at any given time). κ is a free parameter which governs the magnitude of the noise. This scaling reveals a subtle difference in disparity selectivity with RF size. Because of the final squaring, differences in variability of C lead to differences in mean response. For correlated RDS at the preferred disparity, these fluctuations are correlated in the monocular responses, whereas for uncorrelated RDS they are not. As a result the variability in C is greater for correlated stimuli than for uncorrelated stimuli, and this variability is greater for small RFs than larger ones. Consequently, when scaled by the response to the preferred disparity, smaller RFs show slightly weaker responses to uncorrelated stimuli. Note that the half-matched stimulus introduces additional variation in the binocular correlation, causing responses that are greater than those to uncorrelated dots.

Each neuron also has an antineuron whose response is denoted by $N_{i,k}$. The antineuron response is defined exactly the same as $P_{i,k}$ above, with the same retinal position, except its disparity is the opposite sign. That is to say, if a neuron P has a disparity preference of $0.03°$, then its

antineuron N would have a disparity preference of $-0.03°$. We created an opponent cell by taking the difference of a neuron and its antineuron. To make the decision, each opponent cell's response was summed across time points

$$R_i = \sum_k (P_{i,k} - N_{i,k}).$$ (1.6)

R_i thus reflects the overall opponent (squared) energy for the ith neuron-antineuron pair on a particular trial. If this value is negative, the pair has a mean signal indicative of a negative disparity, and vice versa for positive values.

Within each group of 40 neurons tuned to the same disparity, the models were given non-overlapping RF locations, so that they sampled independent regions of the image. In Figure 1.8, we used 1.5s trials and summed the responses over time in each model neuron. To obtain the decision, we used a straightforward linear readout of the population response: if the summed activity across the neuron-antineuron pairs was negative, then the decision model would report that the stimulus disparity was negative, and vice versa for positive values:

$$\Psi = \begin{cases} 1 & \text{if } \sum_i R_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$ (1.7)

R_i is as defined in Equation 1.6. $\Psi = 0$ and $\Psi = 1$ indicate near and far responses, respectively. The biphasic temporal kernel employed here had a peak response to 4.3Hz, meaning that the on-phase of the kernel has a duration of approximately 125ms. For reference, at 21.25Hz, the dot pattern is updated every 47.06ms and at 120Hz every 8.33ms. We computed the response of the model to a constant number of RDSs consisting of mixed correlated and anticorrelated dots. Correlation varied from -1 (completely anticorrelated) to +1 (completely correlated) in steps of 0.2. A correlation of 0 in this scheme corresponds to half-matched RDSs. We presented 10 000 trials at each disparity-correlation combination; all other stimulus parameters were as previously described.

1.3 Results

1.3.1 Binocular energy model disparity tuning to half-matched stimuli

We created binocular energy model (BEM) units by combining model binocular simple cells whose monocular receptive fields were in quadrature phase. This gives a model complex cell response C which is invariant to stimulus phase. As discussed above, the response of a BEM unit like this is a linear function of binocular correlation. We made the model nonlinear with respect to correlation by adding a static squaring output nonlinearity, giving a final response C^2. The same model was also explored by Read, Parker, and Cumming (2002). We computed the response of both models to correlated, half-matched, and anticorrelated random dot stereograms (RDSs) of

various disparities. Disparity tuning curves for the BEM with and without an output nonlinearity are shown in Figure 1.4. For the binocularly linear BEM (Figure 1.4a), the model's responses to correlated and anticorrelated stimuli have a characteristic symmetry about a horizontal line (the response to zero binocular correlation), i.e. the amplitude ratio between the correlated and anticorrelated responses is 1. For half-matched stimuli, the model has no disparity selectivity. When the model has a static output nonlinearity (Figure 1.4b), the correlated and anticorrelated tuning curves become asymmetric. This asymmetry leads to a modest, but very clear disparity tuning for half-matched RDSs with a peak at the neuron's preferred disparity.
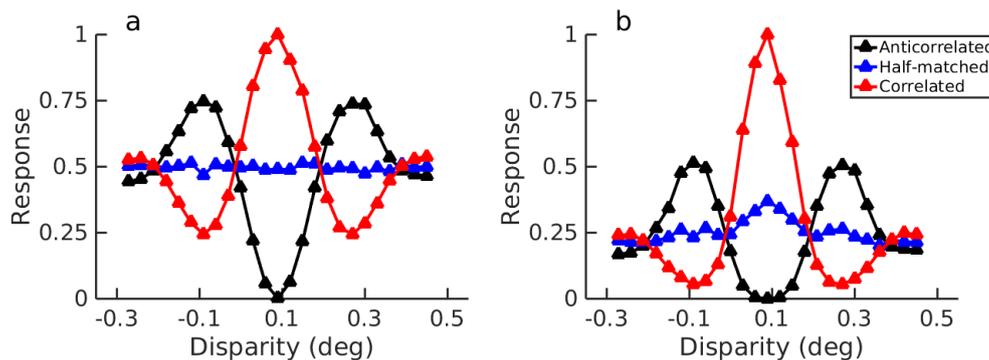
Figure 1.4: Binocular energy model tuning curves. Tuning curves from a binocular energy model in response to correlated, half-matched and anticorrelated RDSs for a binocularly linear model (a) and a model with an output nonlinearity (b).

In the linear BEM (Figure 1.4a), the half-matched nature of the stimulus manifests itself as a variable firing rate, but not as a mean change [Doi et al., 2013]. The disparity tuning in Figure 1.4b arises because the expected value of a squared random variable depends on its variance: $\mathrm{E}[X^2] = \mathrm{E}[X]^2 + \mathrm{Var}[X]$ (other choices for the exponent, as well as other nonlinearities, such as thresholding, will also yield a dependence on variance). For the model with an output nonlinearity, therefore, the high correlation variability in half-matched RDSs gets converted into a mean change in the firing rate. Because the tuning in Figure 1.4b is the consequence of fluctuations in local correlation over the RF, stimulus parameters that affect the variability of these local measures also affect the disparity tuning. Increasing the dot density decreases the correlation variability as there are more dots within a neuron's spatial receptive field. Following earlier studies [Doi et al., 2011, Doi et al., 2013], dot density is defined as the proportion of the stimulus area that would be covered by dots if the dots were not allowed to occlude (although dots were allowed to occlude), hence the units are in proportion coverage. Having fewer, larger dots (while maintaining constant density) generally increases the correlation variability as a single dot fills a larger fraction of the RF with pixels sharing the same correlation. With more, smaller dots, the cell is integrating across a greater number of samples (since the cell is likely to see more independent dots within its receptive field) and so the variability is reduced. Because increasing the dot size while holding density constant is the equivalent of reducing the receptive field size in the model, we use the relative
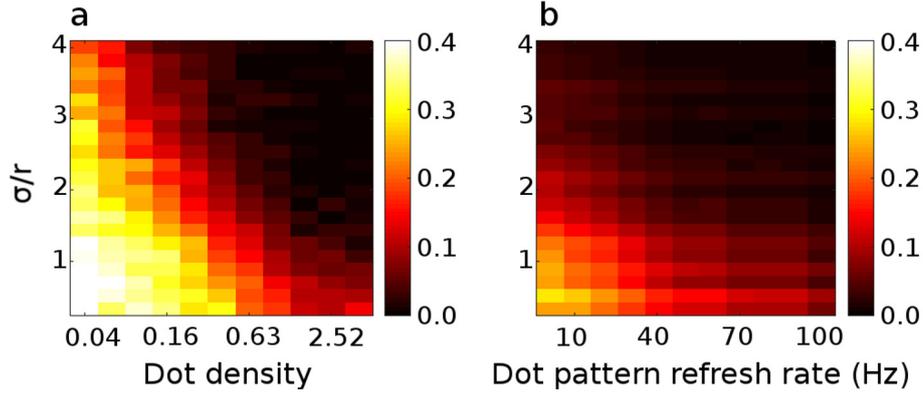
Figure 1.5: Effect of dot density, dot pattern refresh rate and relative receptive field size on disparity tuning in a model cell. a) Normalized half-matched response (Equation 1.1) as a function of density and relative receptive field size (density is here defined as the equivalent proportion of space the dots would occupy if they did not occlude one another). A relative receptive field size of 2 means that $\sigma/r = 2$, i.e. that the standard deviation of the monocular Gabor is twice the dot radius. Colors show the magnitude of the normalized response. A normalized response of 1 would indicate that the response to half-matched stimuli at the preferred disparity is equal to the correlated response; a normalized response of 0 means that the half-matched response is equal to the uncorrelated response. Decreasing the relative receptive field size and reducing the dot density both increase the variability in the correlation level and thus increase the half-matched response. The normalized responses were obtained by computing the average half-matched, correlated, and uncorrelated responses from 20 000 RDS per density-relative RF size combination. b) Normalized half-matched response as a function of dot pattern refresh rate and relative receptive field size. Decreasing the relative receptive field size again increases the half-matched response. Similarly, increasing the dot pattern refresh rate decreases the variability in binocular correlation and thus decreases the half-matched disparity tuning. The normalized responses were obtained by averaging across 5000 trials, 10 seconds in duration, for each frequency-relative RF size combination.

receptive field size, defined as $\frac{\sigma}{r}$, where σ is the standard deviation of the monocular Gabors, and r is the dot radius. The effect of density and relative receptive field size on disparity tuning in our model can be seen in Figure 1.5a, where the strength of disparity tuning for half-matched stimuli is plotted as a proportion of the modulation produced by correlated patterns with the same spatial parameters. We define this normalized mean response from the responses to the preferred disparity:

$$R_{\mathrm{norm}} = \frac{\langle C^2_{\mathrm{hm}} \rangle - \langle C^2_{\mathrm{uncorr}} \rangle}{\langle C^2_{\mathrm{corr}} \rangle - \langle C^2_{\mathrm{uncorr}} \rangle}, \tag{1.8}$$

where C^2_{uncorr}, C^2_{hm}, and C^2_{corr} are respectively the mean response to uncorrelated, half-matched and correlated stimuli at the neuron's preferred disparity. As the relative receptive field size decreases (i.e. smaller receptive field relative to the dot size), modified BEM cells signal disparity more vigorously to half-matched stimuli (relative to equivalent correlated stimuli). It is worth noting that while the amplitude ratio is large for very low densities, this does not necessarily translate to better psychophysical performance on the task. This is because at low densities, the responses to both correlated and anticorrelated patterns are much weaker, so the signal-to-noise ratio is lower. (The variations caused by the monocular image content dominate the model's

response). The effect of this is that there may be poorer performance at very low densities than at higher densities, despite the amplitude ratio being higher at low densities. Consider the case where the density is so low that on some trials no dots are presented in the RF. For these trials performance based solely on this neuron will be at chance, despite the high normalized response ratio averaged across many trials. Even performance based on many neurons would presumably suffer from a lower signal to noise ratio. We will explore the effects of signal and noise in simulations of a psychophysical task below.

Just as increasing the spatial extent of the receptive field (relative to dot size) reduces variability in binocular correlation seen by the model, so increasing the temporal extent of the RF (relative to the pattern update frequency) reduces variability. We explore this using a fixed RF and a changing dot pattern refresh rate. As refresh rates increase, the model cell is integrating more dots within its temporal window, reducing fluctuations in the binocular correlation. Modified BEM cell responses are shown in Figure 1.5b for different receptive field sizes and pattern refresh rates. For low refresh rates and small RF sizes, the cell exhibits substantial disparity tuning to half-matched stimuli. As in Figure 1.5a, when the RF size increases relative to the dot size, the disparity tuning to half-matched stimuli decreases since the cell is integrating across more dots. Similarly, as the dot pattern refresh rate increases, the half-matched disparity tuning decreases since the cell is again integrating across more dots (but now across time rather than across space).

These simulations demonstrate two key properties: the model exhibits less disparity tuning to half-matched stimuli as 1) the receptive field increases in size relative to the dot size, and 2) the refresh rate increases relative to the temporal integration period of the neuron. The first finding is noteworthy because it was observed in Doi et al. (2011) that human observers are better at reporting depth in fine disparity half-matched RDSs than in coarse disparity ones [Doi et al., 2011]. Given the effects of RF size we show, this observation might be accounted for by the well-known size-disparity correlation [Allenmark and Read, 2011, Smallman and MacLeod, 1994, Prince and Eagle, 1999, Tyler, 1974, Prince et al., 2002a], since coarse disparity detectors tend to have larger receptive fields than fine disparity detectors. The second finding is noteworthy because Doi et al. (2013) reported that performance to half-matched stimuli also decreased with increasing pattern refresh rate [Doi et al., 2013]. The authors interpreted this as a shift from a matching computation at low refresh rates to a pure correlation computation at high refresh rates. In the current framework, the decreased disparity tuning with refresh rate reflects temporal integration within a single correlation-based computation, rather than differential activation of two distinct computations with different spatiotemporal properties. Although the match-based computation hypothesized by Doi et al. is similar to our modified correlation-based computation, in our framework, only a single computation is involved. Indeed, our model neurons could be described as the sum of a matching computation and pure correlation computation, just as illustrated in Figure 1.1c, but this is achieved by a single mechanism. Unless the two components differ in some other way (e.g. temporal response), the two descriptions are identical.

1.3.2   Psychophysical decision model

Clearly, the fact that our version of the BEM can signal disparity in half-matched stereograms makes it possible that this explains human psychophysical performance. Additionally, our model neurons lose disparity tuning to half-matched stimuli with increasing receptive field size and dot pattern refresh rate, which is also in agreement with the psychophysical literature [Doi et al., 2011, Doi et al., 2013, Doi and Fujita, 2014]. Finally, the model neurons show weaker responses to anticorrelated dots, which produce weak or absent depth sensations [Doi et al., 2011, Doi et al., 2013, Hibbard et al., 2014, Read and Eagle, 2000, Tanabe et al., 2008]. Thus, at least qualitatively, the signal strength in these model neurons parallel all of the psychophysical phenomena that have been used to suggest two stereo mechanisms. However, these manipulations also influence the ratio of the signal to noise, so the response amplitudes described above cannot simply be compared to psychophysical performance. To test more formally whether our model can explain these psychophysical phenomena, we simulated responses of a small population of neurons and made the perceptual decision based on a straightforward linear readout of the population activity.

The population and decision rule are described in detail in the Methods section. Briefly, our starting-point was the modified binocular energy model units discussed above, whose response is denoted by C^2 in Equation 1.1. For neurons with different RF sizes, we scaled these responses, such that the cells had equal mean responses to correlated stimuli at their preferred disparity. The disparity tuning curves of these cells are shown in Figure 1.6 in response to correlated RDSs. The peak of the disparity tuning curves are identical for all cells, but the uncorrelated baseline is slightly lower for the fine disparity neurons (see Methods). We then applied Gaussian noise, independently for each unit, with variance proportional to the response at any time and the constant of proportionality being a free parameter. When using a fixed stimulus duration, higher pattern refresh rates mean that model responses average over a larger number of RDS images, reducing the stimulus-driven variability between trials. We therefore fit the noise parameter separately for each frequency. If an equal number of frames is used (and hence varying stimulus duration), the same results can be reproduced with fixed noise across frequencies. For simplicity, we restricted ourselves to four preferred disparities $(-0.48°, -0.03°, 0.03°, 0.48°)$, and assumed that receptive field size scaled with disparity magnitude according to $\sigma = 0.023° + 0.41|\Delta x|$. In simulations using only one cell per disparity, random fluctuations in the monocular image content led to performance that was poorer than human subjects. This is further compounded by the fact that local receptive field models, such as the BEM (and its derivatives) measure local instaneous correlation, and not global binocular correlation voer time. We therefore included multiple cells for each disparity, differing only in their locations on the retina. Each cell had non-overlapping receptive fields that were otherwise identical. We found a good match to human performance using 40 cells per disparity, for a total of 160 cells in the model population.
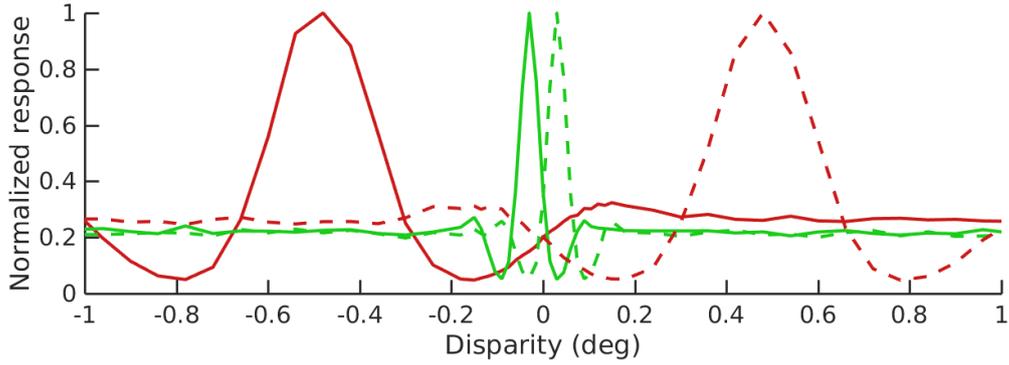
Figure 1.6: Disparity tuning curves for binocular energy model units with a squaring output non-linearity in response to 100% correlated RDSs. Green lines show fine disparity tuning curves, red lines show coarse disparity tuning curves. Dashed lines indicate cells tuned to positive disparities, while solid lines indicate cells tuned to negative disparities. The tuning curves were constructed by computing the responses of four model cells to correlated RDSs with disparities in the range $[-1°, 1°]$. 5000 RDSs were used per disparity.

For the decision rule, we first created opponent cells by taking the difference of a squared energy model neuron and its "antineuron", i.e. the neuron at the same location in the retina but with preferred disparity differing in sign. To make the decision, each opponent cell's response was summed across time points to obtain R_i: the overall opponent (squared) energy for the ith neuron on a given trial (Equation 1.7). A negative value for any given neuron-antineuron pair means that the pair signals a negative (near) disparity, while a positive value means that the pair signals positive (far) disparity. To obtain a decision, we summed the activity across the pool of neuron-antineuron pairs. If this summed value was negative, the model reported that the stimulus had a negative disparity, and vice versa. The decision model is shown schematically in Figure 1.7. The red and green circles show the coarse and fine receptive fields, respectively, corresponding to the red and green tuning curves in Figure 1.6. The tuning curves with negative disparities belong to model units whwich make up the "near" pool and vice versa for cells with positive disparity. The responses of the units in each pool is linearly summed before the responses are compared.

We first consider the effect of disparity magnitude on stereo depth perception in half-matched RDSs. As Figure 1.5 shows, large dots (relative to RF size) produce stronger responses to half-matched stereograms than small dots. For a stimulus with a fixed dot size, this means that neurons with smaller RFs give stronger disparity signals in half-matched stereograms. Because we include a size-disparity correlation in our model, fine disparities will elicit responses predominantly from neurons with smaller RFs. These cells will see larger correlation fluctuations because of their small RF size and thus have a larger response in the half-matched condition. The results from the simulations are shown in Figure 1.8a. For half-matched stimuli (correlation of 0) the model performs
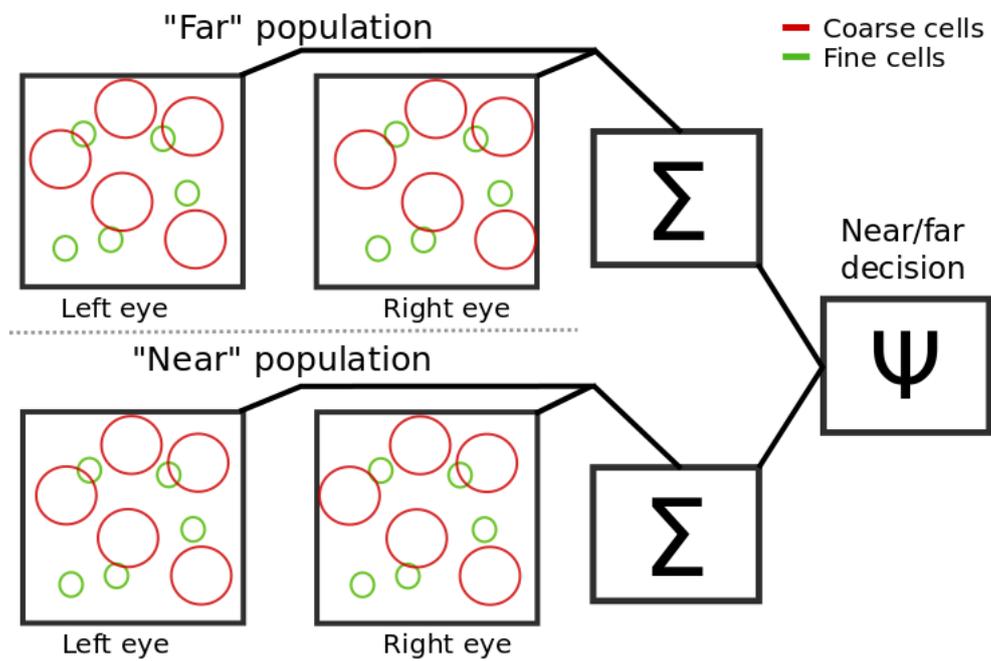
Figure 1.7: Schematic representation of the psychophysical decision model. Two separate pools of cells are used – one tuned to negative or "near" disparities, and one tuned to positive or "far" disparities. Each pool has cells with large receptive fields tuned to coarse disparities, and cells with small receptive fields tuned to fine disparities. The response of each pool is integrated, and the responses are then compared. If the near pool has a larger response, the model guesses that a negative disparity was present, and vice versa for the positive pool.

better in response to fine disparity stimuli than to coarse disparity stimuli. This remains true across a range of correlation values, because even for low correlation values there are local fluctuations in the correlation level. The fine disparity model cells are more sensitive to these fluctuations than the coarse disparity cells, which leads to a leftward shift in the psychometric function. The upward shift at fine disparities arises because here all cells contribute to the disparity judgment (often with opposing signals, see Figure 1.6). However, for coarse disparities, the coarse cells dominate the decision to anticorrelated stereograms. These shifts are similar to that observed by Doi et al. (2011), but our account invokes only a single mechanism.

Next we consider the effect of dot pattern refresh rate on model performance. We used the same model as for the disparity magnitude simulations, but used fine disparity stimuli. We presented RDSs at two different frequencies: 5.3Hz and 42.5Hz, with a stimulus duration of 1.5s. Figure 1.8b shows the psychometric functions for the same model in response to low and high refresh rates for fine disparities (0.03°). As the pattern refresh rate increases, the psychometric function moves rightward. This shift occurs because the neuron is integrating across more dot patterns (in this case over time rather than space) which reduces local fluctuations in correlation. A similar result was obtained using an equal number of stimulus frames rather than equal stimulus duration. These results are qualitatively very similar to the data reported by Doi et al. (2011, 2013). The key result from Doi et al. (2011) is reproduced in Figure 1.9a, and the key result from Doi et al. (2013) is reproduced in Figure 1.9b.

Our model produced weak reversed depth in response to anticorrelated stimuli. This is similar to some human studies [Doi et al., 2011, Doi et al., 2013, Read and Eagle, 2000], but the reversed depth reported by our model was also generally stronger than that reported in the literature. This partly reflects the fact that our model cells modulate their activity more than typical V1 neurons. A final output exponent greater than 2 would reduce this, but we present data for the simplest model as it is more tractable. Additionally, responses to anticorrelated stimuli are influenced by factors that are not readily incorporated into simple models. For example, when a zero-disparity annulus is also anticorrelated, depth perception is abolished [Read and Eagle, 2000, Cumming et al., 1998], but when the surround is correlated, depth is sometimes reported [Doi et al., 2011, Doi et al., 2013, Tanabe et al., 2008]. No existing models provide an account of this effect of the surround. While a sufficiently complicated model could undoubtedly explain this important phenomenon, adding additional parameters to the model would make it harder to interpret the success in explaining the results we discuss here. To this extent, the description of reversed depth in our model (and all other extant models) is a simplification. In our model, the extent to which reversed depth is reported is quite sensitive to the shape of the disparity tuning curves used for model neurons. Despite this, the model does make clear that the information available about negative correlations is influenced by these stimulus manipulations. Our configuration illustrates that it is possible to
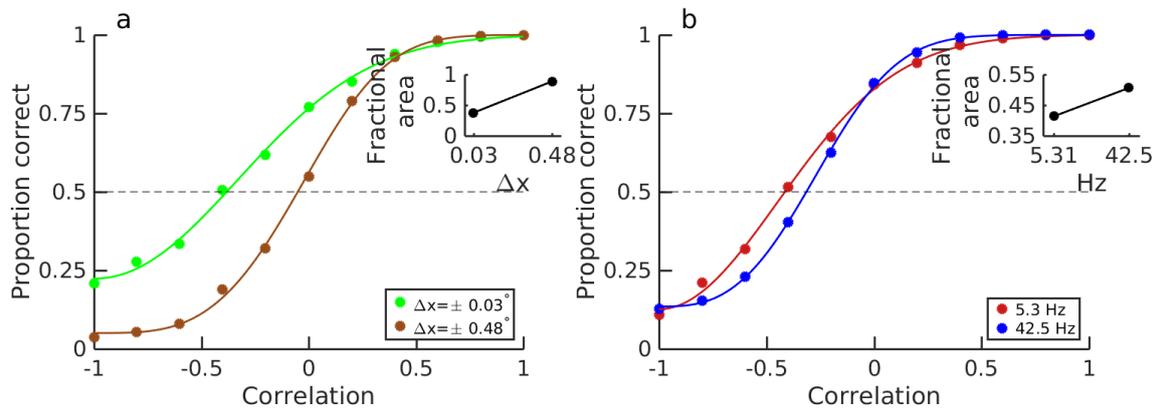
Figure 1.8: Simulated psychophysical performance. Performance of the model (proportion correct responses) as a function of binocular correlation. A binocular correlation of 0 here indicates half-matched stereograms, i.e. stimuli with an equal number of correlated and anticorrelated dots. a) Model performance to fine disparity stimuli ($\pm 0.03°$, green line) is better than in response to coarse disparity stimuli ($\pm 0.48°$, brown line) for many correlation values. b) Model performance to stimuli with low pattern refresh rate (5.3Hz, red line) is better than to stimuli with high refresh rate (42.5Hz, blue line). In a) all stimuli were updated at 21.25Hz, and in b), all stimuli were presented at a disparity of 0.03. Noise was fitted to match the performance of human observers. The insets in both figures show the fractional area, which quantifies the odd-symmetry component of the fitted psychometric function. A fractional area of 1 means that the psychometric function is completely symmetric (deviations from chance performance are exactly proportionate for anticorrelated and correlated stimuli), whereas a fractional area of 0 means that all of different-from-chance performance occurs at positive correlation values (i.e. the psychometric function is completely odd-symmetric).
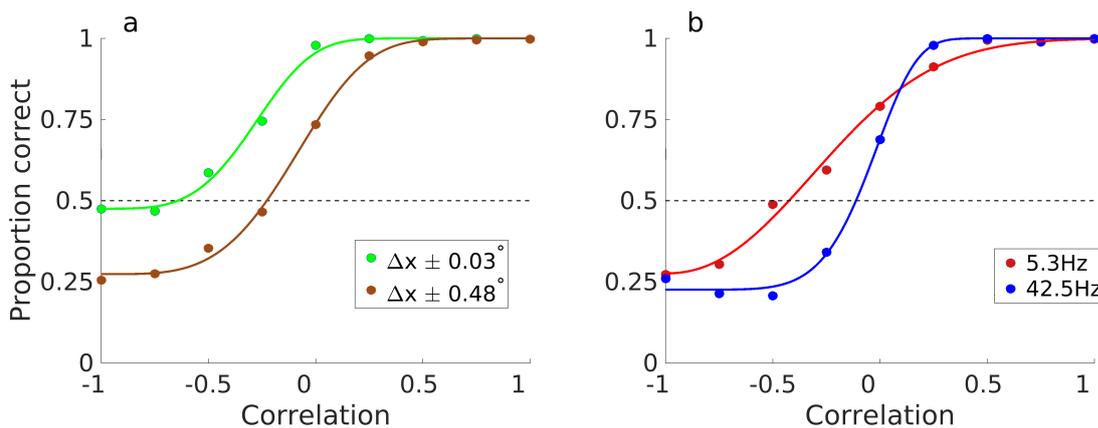
Figure 1.9: Psychophysical performance from Doi et al. (2011; left) and Doi et al. (2013; right). In a), as the stimulus disparity becomes coarser, more reversed depth responses are seen for anticorrelated stimuli, and poorer performance is seen for half-matched stimuli. In b) the same effect is seen with high refresh rates.
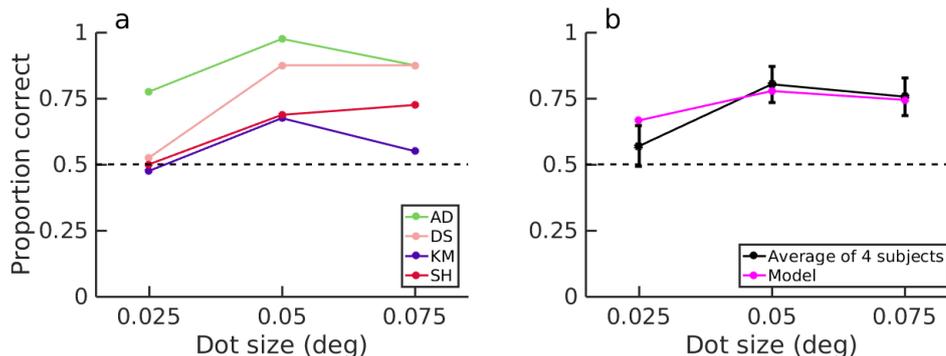
Figure 1.10: The effect of dot size in half-matched RDSs. a) Psychophysical performance for four human observers in response to dynamic half-matched RDSs of three different dot sizes presented at 21.25Hz dot pattern refresh rate. Each data point shows a minimum of 40 trials for each dot size per subject. b) Average performance for the four human observers and for the model shown in Figure 1.6. Error bars show binomial 95% confidence intervals. The noise-level was fitted to match the human observers. Dot density for all stimuli was 24%.

explain the results of Doi et al., including reversed depth, using a single mechanism.

Here we have have used a linear readout of population activity, which is a parsimonious method of neural decoding, and has frequently been used before [Uka and DeAngelis, 2004, Shadlen et al., 1996, Gold and Shadlen, 2007]. It is possible that a more sophisticated decision model could yield performance even more similar to the human observers. Indeed, more complex models, such as those incorporating Bayesian priors [Read, 2002], or using maximum likelihood decoding [Goris et al., 2013], have been shown to capture human psychophysical performance across a range of stimuli and tasks. However, given the simplicity of our model population (only two cell types), and the fact that the model cells do not faithfully reflect the properties of V1 neurons (which show stronger attenuation to anticorrelated stimuli), exploring more complex decision rules seems inappropriate.

1.3.3   The effect of dot size on human and model performance

Our model neurons show reduced responses to half-matched stimuli as dot size (relative to RF size) is reduced. If this correctly captures the nature of signals used to perceive depth, we should expect human performance also to depend on dot size to half-matched RDSs, but not for 100% correlated stereograms. We therefore examined the effect of dot size on depth perception. Figure 1.10a shows the proportion correct as a function of dot size for 4 observers. For the smallest dot size ($0.025°$), performance is not significantly different from chance for three of the observers using 95% binomial confidence intervals. We used a Monte-Carlo simulation to test if dot size had a significant effect on the variance in performance and found that it did ($P = 6.3 \times 10^{-5}$). For correlated stimuli, all observers performed at $\geq 95\%$ correct for all dot sizes. Thus, the decreased performance in response to small dot half-matched RDSs cannot be attributed to changes in the spatial content of

the monocular images. The decreased performance is consistent with predictions made by energy model units with a simple output nonlinearity. Figure 1.10b compares the average response of the human observers with the psychophysical decision model introduced earlier (Figure 1.8). Both the model and the average human performance show a similar decline for the smallest dot size.

1.3.4 Alternating-correlation stereograms: temporal half-matching

We showed above that the effects of pattern refresh rate on performance in half-matched stereograms can be explained by the effects of temporal integration on local fluctuations in correlation. Doi et al. (2013) propose a different explanation, which is that the matching process is slow and so the rapid changes in monocular patterns disrupt the matching computation, leaving the pure correlation computation to dominate at fast refresh rates [Doi et al., 2013]. In order to provide an additional test of these hypotheses we introduce a new stimulus in which the monocular pattern refresh rate is always high, but binocular correlation changes at different rates. Each random dot pattern is either 100% correlated or 100% anticorrelated (with the same disparity), and the correlation value alternates (while monocularly every new monitor frame shows a new pattern, as illustrated in Figure 1.3). We then explored the effect of changes in frequency with which the correlation alternates. By presenting dot patterns at very rapid pattern refresh rates, we should be able to keep the contribution of any putative slow matching computation to a minimum, independent of alternation rate. Doi et al. (2013) showed that the energy seen by their sustained, matched-based mechanism fell by a factor of 2 as the pattern refresh rate increased from 5.3Hz to 43Hz [Doi et al., 2013]. At 43Hz, the highest refresh rate they could present, their sustained and transient channels were seeing equal stimulus energy. In our CRT mirror stereoscope, we use a pattern refresh rate of 120Hz. According to Doi et al's model, this will ensure that the transient channel feeding into the pure correlation mechanism is driven far more strongly than the sustained channel feeding into the match-based mechanism. Perception in this stimulus should therefore be dominated by the pure correlation mechanism.

However, Doi et al's model also predicts that even though the pure correlation mechanism is strongly driven, it must perform at chance in this task. Their definition of a pure correlation mechanism is one that outputs 100% veridical depth for 100% correlated stimuli, 100% reversed depth for anti-correlated stimuli, and is at chance (50%) for half-matched stimuli. Now, at alternation rates which are slow compared to the temporal kernel of this mechanism, the observers are simply seeing a rapidly updating stimulus which periodically flips between being correlated and anticorrelated. Let's say this stimulus has a near disparity. Doi et al's pure correlation mechanism will report alternately "near" and "far" as the correlation flips. Since we randomised the phase of our alternation, over many trials there is no way for their pure correlation mechanism to report "near" more or less often than "far". On average, therefore, performance must be at chance. The

situation is no better for alternation rates which are fast compared to this mechanism. There, both correlated and anticorrelated frames fall within the temporal integration window. The stimulus is effectively half-matched, and by definition, Doi et al's pure correlation mechanism must be at chance. Thus, a pure correlation mechanism, as defined by Doi et al., cannot contribute to above-chance performance with this stimulus. Their match-based mechanism can contribute in principle, but their conclusions about its temporal properties - that it is temporally low-pass - makes depth discrimination in rapidly updating stimuli, such as the alternating-correlation stereograms, a particularly demanding task. According to Doi et al. (2013), depth perception makes use of "a simple, correlation-based representation for more dynamic (faster) and coarser features, and a complex, match-based representation for less dynamic (slower or stationary) and finer features" [Doi et al., 2013]. The implication is that for very rapidly updating stimuli, the relative contribution of the matching computation to depth perception should be very small. Thus, this task presents a particular challenge because the near/far judgment must be based on weak signals from the slow matching computation, and stronger, but alternating and conflicting signals from the fast correlation computation. Yet as Figure 1.11 indicates, our human observers performed well above chance for alternation rates below 30Hz.

We compared the responses of human observers to alternating-correlation RDSs of various alternation frequencies with that of the psychophysical decision model used earlier. Figure 1.11a shows psychophysical performance in response to alternating-correlation RDSs averaged across four subjects for the three disparities employed. Clearly, stimulus disparity had virtually no effect on this task, where task difficulty was manipulated by increasing the alternation rate. Individual psychometric functions, averaged across disparities, are shown for each subject in Figure 1.11b. In both plots, model responses are shown in magenta. At alternation rates below 4Hz, the human observers make accurate judgments, but as the alternation rate increases, performance decreases. At intermediate alternation rates, the human observers can still do the task, but crucially, as the alternation rate increases beyond about 20Hz, the correlation variability decreases and the performance of human observers falls to chance. We conclude that our model accounts excellently for human performance in this stimulus. Using the temporal kernel defined by Doi et al. (2013) for the matching computation, we find a slightly less good fit to the data (see dotted line in Figure 1.11). The main difference is at slow alternation rates, where subjects perform very well as does our mechanism with a bandpass kernel. This model is unable to produce good performance here because of the high monocular pattern refresh rate and its lowpass kernel. Presumably, performance would be even worse in a two-mechanism model where the weak signal from the low-pass matching computation competes with a much stronger, alternating signal from the pure correlation computation.
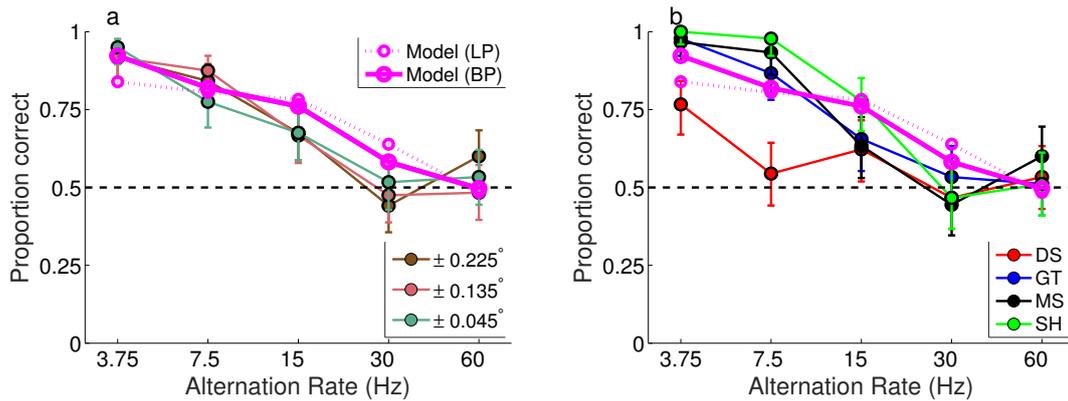
Figure 1.11: Psychophysical performance to alternating-correlation stereograms. Performance of human observers on a front-back discrimination task with alternating-correlation random dot stereograms. a) Each line shows the averaged performance of all subjects for each disparity magnitude. b) Each line shows the performance of an individual observer averaged across the three disparities. The solid magenta line in each plot shows model performance with our band-pass temporal kernel [Model (BP)], and the magenta dotted line shows the model performance with the low-pass temporal kernel used in Doi et al. (2013) [Model (LP)]. In both cases, the model was shown a disparity of 0.03, and noise was fitted to match human observers.

1.4 Discussion

The first step in stereoscopic depth perception is the extraction of disparity from the left and right images. The success of the binocular energy model in describing the responses of disparity-selective cells in primary visual cortex [Ohzawa et al., 1990, Cumming and Parker, 1997, Read et al., 2002] led to correlation-based schemes becoming the dominant explanation for human stereo vision [Filippini and Banks, 2009, Banks et al., 2004, Kane et al., 2014]. Disparity-selective cells in V1 show the hallmark of a correlation computation: inverting the binocular correlation of the stimulus inverts their response, whereas a unit looking for matching features would simply cease to detect any disparity after this manipulation. Recently, Doi et al. have postulated that two distinct computations extract disparity in dense random dot stereograms: one extracts disparity by computing the binocular correlation, while the other extracts disparity by computing the binocular match level of the image. Doi et al. cite three key observations as evidence for the existence of this match-based computation. First is the fact that humans correctly report depth in random dot stereograms made of equal numbers of correlated and anticorrelated dots [Doi et al., 2011, Doi et al., 2013]. Second, performance to these half-matched stimuli degrades with disparity magnitude, in a way that is not seen with correlated RDS [Doi et al., 2011]. Third, in dynamic stimuli where the dot pattern is regularly refreshed, performance declines with increasing refresh rate for half-matched, but not correlated, stereograms [Doi et al., 2013].

1.4.1   A single model accounts for depth judgments in half-matched random-dot patterns

Here we show that all three of these observations can be explained by a model which uses a single nonlinear correlation-based mechanism (the BEM with a static output nonlinearity). The nonlinearity we propose is similar to the nonlinear "cross-matching" model proposed by Doi & Fujita (2014) - our proposed mechanism is well described by the linear sum of a "pure-correlation" mechanism and a "cross-matching" mechanism (Figure 1.1c). Thus the two accounts are closely related. If the pure correlation filter and cross matching filter have the same monocular spatiotemporal RFs, the two accounts are indistinguishable. If the different computations are associated with filters that have different spatiotemporal properties it becomes possible to distinguish one mechanism from two, as two mechanisms will then predict different psychophysical performance. This is exactly what Doi et al. propose to explain psychophysical changes that occur with pattern refresh rate and disparity magnitude. We point out here that many of those changes in psychophysical performance might occur because of the effects of stimulus changes on the activity of a single mechanism. Doi et al. do not report any simulations with single-mechanism models, and so never test the null hypothesis that a single channel suffices. Our results suggest that it may. Importantly we do not claim to falsify a two-computation hypothesis - the isomorphism shown in Figure 1.1c means that any data described by a single computation can also be described with two. All three of these mechanisms (pure correlation, cross-matching, and the BEM with an output nonlinearity) could be described as "correlation-based", since they all start by computing correlation.

Adding an output nonlinearity to the BEM allows the neurons to signal disparity in half-matched stereograms because the correlation fluctuations are converted to a mean firing rate through this nonlinearity. (When there is no nonlinearity, the correlation fluctuations manifest as a variable firing rate, but do not lead to a change in the mean). It follows from this that the larger the correlation fluctuations, the greater the disparity tuning to half-matched stimuli. We showed that the effect of disparity magnitude on half-matched depth perception can be explained if larger receptive fields are used to detect larger disparities (the well-known size-disparity correlation [Allenmark and Read, 2011, Smallman and MacLeod, 1994, Tyler, 1974, Prince et al., 2002a]. A similar point was noted by Doi et al. (2013), who found that larger receptive fields decrease the response variability of standard energy model cells [Doi et al., 2013], and by Doi & Fujita (2014) who extended these findings to a "cross-matching" computation [Doi and Fujita, 2014]. Along the same lines, the observation that depth perception is compromised in rapidly changing half-matched RDSs is compatible with temporal integration within the correlation mechanism, and does not imply a qualitative shift to a different computation.

This model correctly predicts our finding that psychophysical performance decreases with

smaller dot size, and states that this is because smaller dots tend to decrease the local correlation variability. It also correctly predicts that alternating the correlation over time should decrease psychophysical performance because of a reduction in the effective variability in binocular correlation. This is particularly interesting since the alternating stimulus presents particular challenges for both of the mechanisms proposed by Doi et al. Their pure correlation mechanism should be at chance since, by definition, it reports opposite depth sign for opposite correlations, and thus reports either depth sign with equal probability for our alternating stimulus. Our single-mechanism model can straightforwardly account for depth perception in these stimuli. Our model has a single, fixed temporal kernel, and yet it can account simultaneously for the effects both of pattern refresh rate (Figure 1.8b) and of alternation rate (Figure 1.11). Indeed our model describes the data somewhat better than the "cross-matching" mechanism, most tellingly at low alternation rates. Here subjects are close to 100% correct, whereas the sustained temporal properties of Doi et al's cross-matching mechanism means that it only reaches 80% correct.

1.4.2 Aspects of depth perception still unexplained

It has long been recognized that in some situations stereo correspondence that is not based on correlation can be exploited. For example, it is well-known that monocular occlusion (i.e. objects seen by one eye but occluded in the other) can contribute to the perception of depth in humans (so-called da Vinci stereopsis [Nakayama and Shimojo, 1990]). Additionally, patients with binocular vision disorders such as strabismus may show no depth perception with cyclopean stereograms, while having measurable stereoacuity in images with monocularly visible contours [Fricke and Siderov, 1997, Fawcett, 2005, Giaschi et al., 2013]. For isolated monocular targets, humans can correctly report depth for larger disparities than is possible in random dot patterns [Ogle, 1952c, Westheimer and Tanzman, 1956]. This suggests that the human visual system may also contain a separate algorithm, which enables a coarse form of stereopsis even when the correlation-based system is damaged [Read, 2015], at least for sparse images consisting of a small number of monocularly visible objects. There is some evidence suggesting that this system can use head-centric rather than retinotopic coordinates [Read, 2015, Zhang et al., 2010, van Ee and Erkelens, 2010], implying an extrastriate locus. Importantly, however, the match-based computation proposed by Doi et al. is quite different from this mechanism, as it would have to operate on dense random dot patterns.

Additionally, in human observers, depth perception in half-matched and anti-correlated stereograms requires the presence of an adjacent correlated region [Doi et al., 2011, Doi et al., 2013]. This may be related to humans' greater sensitivity to the relative disparity between adjacent surfaces than to the absolute disparity of a surface in isolation. However, the presence of a correlated surround is not sufficient for reversed depth in anticorrelated RDSs [Hibbard et al., 2014, Doi et al., 2014]. As these observations indicate complex interactions between different regions of the visual

field, it is inevitable that they cannot all be explained by models like ours (or that of Doi & Fujita, 2014) that consider only the responses of a population of neurons at one location. That a local model exploiting a single mechanism successfully explains so many phenomena strongly suggests that a single mechanism is responsible.

1.4.3   Properties of V1 neurons

This mechanism closely resembles the known properties of disparity tuning in V1. The critical property is that V1 neurons show weaker tuning to anti-correlated than to correlated stimuli. The reduction becomes more pronounced in later areas [Janssen et al., 2003, Tanabe et al., 2004], but is already present in V1 [Cumming and Parker, 1997]. This asymmetry suggests that V1 neurons should, weakly, encode disparity in half-matched stimuli. This together with the tendency for V1 neurons tuned to large disparities to have larger receptive fields [Prince et al., 2002b] can account for all the psychophysical data regarding half-matched stimuli. As we have shown, a model based on these ideas provides an excellent account of human performance from previous studies [Doi et al., 2011, Doi et al., 2013] (Figure 1.8) and also predicts performance on new stimulus manipulations (Figure 1.10 and 9). Importantly, we use one model while only fitting a noise parameter to explain the psychophysical results in Figures 6, 7 and 9. It is important to note, though, that no one has yet examined the response of V1 neurons to half-matched stereograms. Our model of V1 neurons captures their weaker tuning to anti-correlated stereograms, and predicts that this results in weak tuning to half-matched stimuli. Yet until this prediction has been directly tested in V1 neurons, we cannot be sure it occurs. If V1 neurons do not show disparity tuning for half-matched stimuli, this would give much greater credence to the idea of a separate dot-matching computation in extrastriate cortex. This question is the subject of chapter 2.

Chapter 2

Neurons in striate cortex signal disparity in half-matched random dot stereograms

2.1   Introduction

As seen in the previous chapter, recent work by Doi et al. (2011, 2013, 2014) has suggested that a correlation-based view of early disparity encoding may need to be supplemented by an additional match-based computation. This is based on the observation that humans are able to see depth in random dot stereograms (RDSs) constructed with an equal number of correlated and anticorrelated dots (termed half-matched RDS [Doi et al., 2011, Doi et al., 2013, Doi and Fujita, 2014]). These stimuli have a mean binocular correlation of 0 (because the correlation of the correlated and anticorrelated dots cancel out) and therefore many correlation-based computations, such as the binocular energy model, do not signal disparity here. This led Doi et al. (2013) to propose that an additional "matching computation", possibly performed in extrastriate cortex, accounts for human depth perception in dense half-matched random dot stereograms. If V1 neurons only perform a correlation computation, then this observation implies that humans see depth in half-matched stereograms even though V1 neurons do not signal disparity in their mean firing rate. This would be surprising as V1 activity is generally thought to be a necessary prerequisite for cyclopean depth perception. Indeed, it would provide the first evidence that depth perception can occur without an explicit signal in V1.

Thus, the current literature on half-matched stereograms suggests a radical change in our understanding of the role played by area V1 in depth perception. This argument depends critically on the assumption that V1 neurons perform a correlation computation, as described by the binocular energy model. However, the attenuated responses to anticorrelated RDS already indicate that disparity selective responses in V1 do not simply reflect correlation. Indeed, the computational work from Chapter 1 of this thesis showed that a simple modification to the BEM produces cells which are disparity tuned to half-matched random dot stereograms. This raises the possibility that neurons in V1 do modulate their firing rate with disparity in half-matched stereograms. We therefore investigated the responses of disparity-selective cells in macaque V1 to half-matched random dot

stereograms. We find that these cells do signal disparity (weakly) in the half-matched condition. The model we developed in the previous chapter can explain this finding, and also predicts that the strength of disparity tuning for half-matched stimuli should decrease with increasing dot density. We show that variation in dot density does have this effect on the responses of V1 neurons. The observed responses to half-matched stereograms restore the view that disparity-selective neurons in V1 provide a sufficient substrate for depth judgments in random-dot patterns. The effects of dot density suggest that a simple mechanism can explain these responses.

2.2 Methods

2.2.1 Animal subjects

Two male macaque monkeys (subjects Lem and Jbe) were implanted with head posts, scleral search coils and a recording chamber under general anesthesia. The full experimental procedure is described in detail elsewhere [Cumming and Parker, 1999, Read and Cumming, 2003]. Briefly, subjects viewed separate CRT monitors with each eye through a mirror haploscope. They were required to fixate a bright spot on each CRT, and maintain fixation for 2.1 sec to earn a drop of liquid reward. The window of fixation was typically a box of $0.8° \times 0.8°$ around the fixation spot. One subject was trained to perform a front/back discrimination task with random-dot patterns as described in [Prince et al., 2000]. All experiments were performed at the US National Institutes of Health. All procedures were performed in accordance with the US Public Health Service policy on the care and use of animals. The protocols were approved by the National Eye Institute Animal Care and Use committee.

2.2.2 Model cells

The model cells were constructed exactly as described in [Henriksen et al., 2016a] (Chapter 1 of this thesis) using BEMtoolbox - a Matlab toolbox for simulating binocular energy model cells (available on `https://github.com/sidh0/BEMtoolbox`). In brief, the binocular energy model (BEM) models a complex cell by combining the responses of two binocular simple cell subunits. The simple cell has linear monocular receptive fields described by a Gabor function. For simplicity, here we used identical RFs in the two eyes, so that model cells have a preferred disparity of zero. The responses from left and right RFs are summed and then squared. A binocular complex cell is the sum of two simple cells in quadrature, i.e. with RF phase differing by $\frac{\pi}{2}$. We modeled a cell whose response was a nonlinear function of correlation by including a static squaring output nonlinearity. Thus, the final model is simply

$$C = (S_1 + S_2)^2 \tag{2.1}$$

where S_1 and S_2 are the two simple cell subunits of the BEM model. We computed the mean

response of the model to correlated, half-matched and anticorrelated random dot stereograms at 5% and 24% dot density. 21 disparities were used, evenly spaced between $-0.5°$ and $0.5°$. The model response was averaged across 10 000 presentations for each stimulus condition.

2.2.3 Recording

We recorded extracellular activity from cells in V1 using 24-channel linear multicontact electrodes (V-probes, Plexon), with 50 m spacing between the probes. Behavioral and neuronal data were sampled using Spike2 (Cambridge Electronic Design). The spike waveform data were saved to disk for offline analysis, and spikes were classified offline using custom software. Cells that were well-isolated and exhibited significant disparity tuning to correlated random dot stereograms of both 5% and 24% density as determined by a one-way ANOVA ($P < 0.01$) were included in the analysis. 53/90 cells passed these criteria.

2.2.4 Stimulus

Black and white square dots were painted on a grey background, with disparity applied to the center of the stimulus, keeping a zero-disparity annulus as reference (in order to eliminate monocular cues; without a zero-disparity annulus, the observers might be able to detect a monocular shift in the dot pattern from trial-to-trial). In the region where dots were displaced, monocular dots were painted in order to prevent monocular cues. The stimulus is illustrated in Figure 2.1. For recordings from the operculum (relatively foveal with RF eccentricity $1° - 3.5°$), the disparity-defined region was $3.4°$ in diameter, while the surrounding annulus had a width of $1°$. The annulus had a disparity of $0°$ and a correlation that matched the center. Some recordings were made from neurons in the calcarine sulcus by advancing the probe through the operculum. For these recordings (eccentricities $10° - 13°$), the disparity-defined region was $4.2°$ in diameter while the annulus had a width of $2°$. This was done to ensure that the larger receptive fields in the calcarine were completely covered by the disparity-defined region. For half-matched stimuli, we painted an equal number of correlated and anticorrelated dots. Each dot had an equal probability of being black or white. An illustration of this stimulus is shown in Figure 2.1. Disparity-values were chosen based on disparity tuning curves collected prior to the experiment, ensuring that the range over which cells exhibit disparity tuning was covered in our selection of disparity values. Each cell was tested with at least 9, sometimes as many as 16, disparities. The random dot stereograms were presented dynamically at a pattern refresh rate of 100 Hz. Each dynamic RDS stimulus was presented for 420 ms (i.e. consisting of 42 unique dot patterns), with four stimuli being presented in a given trial with a 100 ms gap (grey screen) between the stimuli. Thus four stimuli were presented in each completed fixation trial (2.1 sec). This allows four stimulus presentations to be completed while only rewarding the monkey once. Because we anticipated weak responses to the half-matched stimuli, they were presented 10 times more frequently than correlated or anticorrelated disparities. On average each
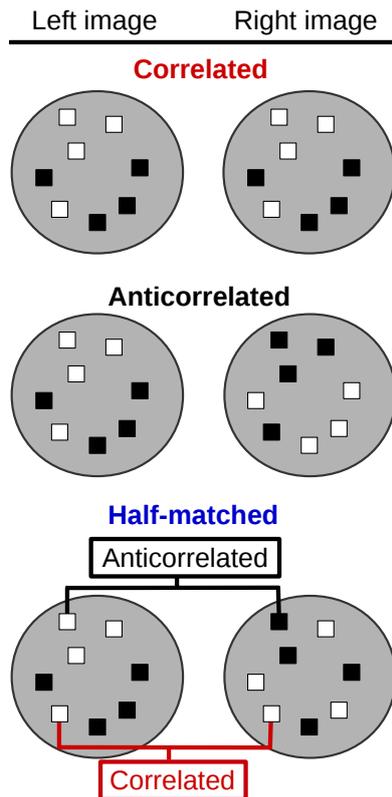
Figure 2.1: Illustration of a random dot stereogram. Black and white dots are painted in the left and right image with disparity applied to the dots. Top row, All the dots have the same contrast in the left and right eyes and so the stereogram is binocularly correlated. Middle row, All the dots have opposite contrast in the two eyes (black is matched with white, and vice versa), and so the stereogram is anticorrelated. Bottom row, The stereogram has an equal number of correlated and anticorrelated dots, and so the stimulus is half-matched.

correlated (or anticorrelated) stimulus was shown 16 times, while each half-matched stimulus was shown on average 161 times. We used two dot density values, 5% and 24%, where dot density is defined as the percentage of the stimulus area that the dots would occupy if they did not occlude one another. The dots were, however, allowed to occlude, but were painted in random order so that correlated dots did not systematically occlude anticorrelated dots or vice versa, and so that the centre did not systematically occlude the surround or vice versa. For the electrophysiological experiments, the monkey simply needed to maintain fixation. The dot size varied depending on the size of the receptive field. Previous modeling work has shown that the ratio between receptive field size and dot size may affect the magnitude of half-matched responses [Henriksen et al., 2016a]. Thus, for eccentric recordings (defined as $> 10°$ eccentricity), the dot size was increased to $0.2°$ or $0.3°$ to compensate for the larger receptive fields (3 sessions, 19 cells). Early recordings were done with $0.2°$ radius dots, but these were subsequently increased to $0.3°$ after recognising that this was a more appropriate size relative to the receptive field size at these eccentric locations. In the remaining recordings the dot size was $0.1°$ (9 sessions, 34 cells). Recordings were either from the operculum (relatively foveal; small dots) or from the calcarine sulcus (relatively eccentric; large

dots). Dots were chosen according to the apparent receptive field size of the neurons.

To provide quantitative estimates of RF size, we measured responses to thin strips of random dot texture. Vertical strips were placed at a variety of horizontal positions, spanning the RFs of all recorded neurons, and a Gaussian function of position (SD σ_x) was fit to the spike counts. Horizontal strips were used to estimate size in the vertical direction (SD σ_y). RF size was then defined as $4\sqrt{\sigma_x\sigma_y}$.

For neurophysiology experiments, stimuli were presented on two Viewsonic P225f CRT displays, with a resolution of 1280×1024 pixels at 100Hz. At the viewing distance used (89cm) each pixel subtended $0.018°$. The luminance response was measured with a Konica-Minolta LS100 photometer, and linearized with a lookup table. The mean luminance was 40 cd/m^2, and contrast was $> 99\%$.

2.2.5   Quantifying disparity tuning

In order to quantify correlated disparity tuning, we used a standard metric known as the Disparity Discrimination Index (DDI; [Prince et al., 2002b]). The DDI was computed using the square root of the firing rate to ensure equal variances for different disparities/firing rates. The DDI is defined as:

$$\text{DDI} = \frac{R_{\max} - R_{\min}}{R_{\max} - R_{\min} + 2\text{RMS}_{\text{error}}} \qquad (2.2)$$

where $R_{\max}$ and $R_{\min}$ correspond to the maximum and minimum mean square root firing rates on the tuning curve, and $\text{RMS}_{\text{error}}$ is the root mean square error around the mean square root rates in the tuning curves. The DDI gives a measure of how large the peak-to-trough difference in the tuning curve is relative to the intra-stimulus variability. A DDI near 0 thus means that the cell can poorly discriminate the disparities corresponding to the peak and trough of the disparity tuning curve. The DDI approaches 1 as the variability becomes negligible relative to the response range.

In order to quantify disparity tuning to half-matched stimuli, we computed the regression slope between correlated and half-matched responses (type 2 regression, [Draper and Smith, 2014]). The half-matched regression slope estimates the magnitude of disparity tuning to half-matched stimuli as a fraction of that for correlated stimuli. A half-matched slope of 1 would mean that the cell has the same disparity tuning to half-matched stimuli as it has to correlated stimuli; a half-matched slope of 0 would mean either that the cell shows no disparity tuning to half-matched stimuli or that the half-matched tuning is present but has a shape that is uncorrelated with the correlated tuning. We observed no instances of the latter, so we used the slope as an index of response magnitude. We also quantified the anticorrelated disparity tuning equivalently by computing the regression slope between the correlated and anticorrelated responses. If the cells modulated their firing rate strictly as a linear function of correlation, the anticorrelated slope should be -1 (corresponding to an amplitude ratio of 1 and a phase change of π). The anticorrelated slope is closely related to the

anticorrelated amplitude ratio that has been previously used [Cumming and Parker, 1997]. The amplitude ratio uses the amplitude of Gabor functions fitted to each of the tuning curves, which has the advantage that it can capture a broader range of changes in the tuning curve, such as phase shifts other than 0 or π. However, since the ratio must exceed 0, it can overestimate weak modulation, which the slope estimate used here does not. We obtained confidence intervals for the half-matched and anticorrelated slopes by resampling of residuals [Efron and Tibshirani, 1994]. For each cell and stimulus dot match value, we performed a square-root transform on the spike counts, before computing the (square-root transformed) residuals for each disparity. In order to construct a single resampled disparity tuning curve, we drew a sample from the pool of residuals, added this on to the square root of the mean firing rate, and squared the value. This gave us one resampled trial. We repeated this for k_i trials, where k_i is the number of trials (observations) for the ith disparity value. In order to generate half-matched slope confidence intervals, we generated a resampled tuning curve for correlated data, and a resampled tuning curve for half-matched data, and then computed the slope between the two. We repeated this procedure 100 000 times, and obtained the 95% confidence intervals for the slopes. The corresponding procedure was done for anticorrelated data to obtain confidence intervals for anticorrelated slope.

2.2.6   ROC analysis

The ROC curve traces the performance of a binary classifier by plotting the false positive rate versus the true positive rate using a variable threshold; in this case the classifier is a cell's ability to discriminate preferred disparity trials from null disparity trials [Green and Swets, 1966, Tolhurst et al., 1983, Britten et al., 1992]. For each cell, we chose the two disparities with the largest and smallest mean response in response to correlated RDS (i.e. preferred and null disparities). Using the half-matched responses to these disparities, we computed the true and false positive rates by progressively incrementing the classification threshold. This gives us the ROC curve for an individual cell. In order to obtain neurometric performance for the cells to half-matched stimuli, we computed the area under the receiver operating characteristic curve (AUROC). The AUROC varies from 0 to 1. A value of 0 means that the classifier is always incorrect, while a value of 1 means that the classifier is always correct. An AUROC value of 0.5 corresponds to chance performance. Thus, the AUROC as a measure of neurometric performance is equivalent to percent correct as a measure of psychometric performance.

The tuning curves we have collected are available on `https://github.com/sidh0/hm` with an accompanying interactive data browser written in Matlab. Matlab code for generating all figures in the current manuscript is also available on the Github repository.

2.3 Results

2.3.1 Model disparity tuning curves

We have previously shown that a simple modification to the binocular energy model can produce disparity selectivity for half-matched stimuli ([Henriksen et al., 2016a]; chapter 1 in this thesis), by adding a squaring nonlinearity at the output of a traditional binocular complex cell. The result is that positive binocular correlation produces a larger change in activity than negative correlation of the same magnitude. This in turn means that random fluctuations in correlation around a mean of zero produce a larger response than a correlation that is fixed at zero. (This is because the expected value of a squared random variable depends on its variance: $E(X^2) = [E(X)]^2 + Var(X)$, so that the squaring output nonlinearity makes the mean firing rate depend in part on the variance in binocular correlation). The original binocular energy model does not signal depth in half-matched stereograms because its response varies linearly as a function of binocular correlation. Thus, when the mean binocular correlation is zero, the mean response of the model is equal to its uncorrelated response (although the variability of the response is greater in the half-matched case; [Doi et al., 2013, Doi and Fujita, 2014, Henriksen et al., 2016a]). The extent of this variation in binocular correlation will depend on the number of dots contained within the receptive field. More dots within the receptive field reduces the fluctuations in correlation. If dot density (expressed in the fraction of pixels that are covered by dots) is held constant, smaller dots produce more dots in the receptive field. For fixed dot size, higher density also increases the number of dots. Thus, small dots and high dot density both reduce the fluctuation in correlation over the receptive field. Consequently, the $Var(X)$ term is smaller, and the mean response of the cell is lower. Decreasing the dot size and increasing the RF size are functionally equivalent operations; thus both produce the same decrease in the fluctuations in the correlation level seen by the cell. In Figure 2.2 we show the effect of dot density on disparity tuning by plotting the responses of the model neuron described in (Henriksen et al., 2016a) to random dot patterns of two densities. We computed disparity tuning curves in response to correlated, half-matched and anticorrelated random dot stereograms. We used two dot densities - 5% and 24%. Figure 2.2a shows the response of the model to 5% dot density stimuli. The tuning curves to correlated and anticorrelated stimuli are asymmetric due to the output nonlinearity. For the half-matched stimuli, the model cell exhibits clear disparity tuning at the preferred disparity of the cell. At higher dot densities (Figure 2.2b) the half-matched disparity tuning, although still present, is greatly attenuated relative to the 5% density stimuli. Thus, our model predicts that there should be a correlation between the magnitude of half-matched tuning to 5% and 24% density reflecting variation between cells in, for example, the output nonlinearity. It also predicts that the responses to the higher dot density should show weaker modulation. One simple way to appreciate these results is to consider a dot density so low that only one dot ever falls within the RF. Half of the stimuli will be 100% correlated, and half will be 100% anticorrelated. The cell's response will then be the mean of its responses to correlation

and anticorrelation. As density is increased, the fluctuations in correlation are reduced, and the disparity-related response of the cell weakens.
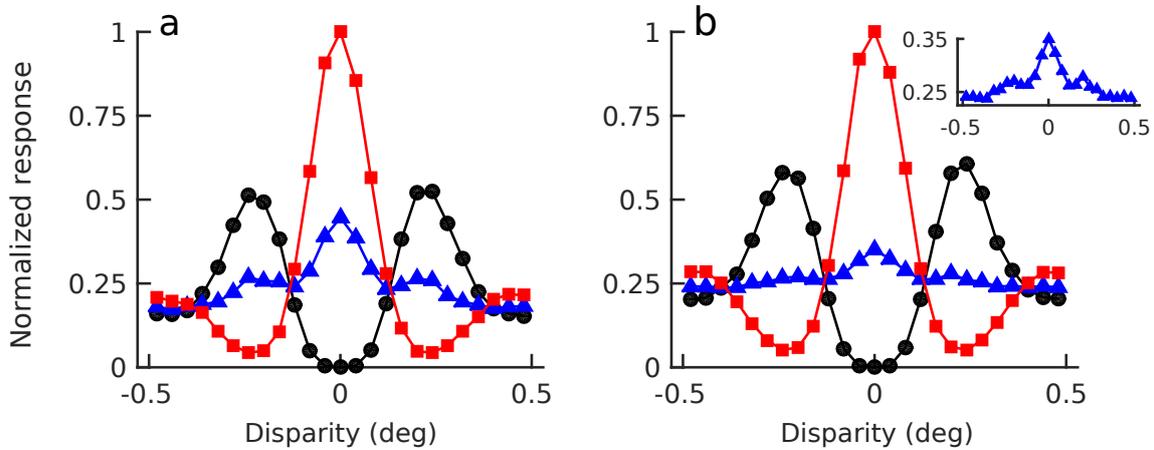
Figure 2.2: Disparity tuning curve of a model cell which signals disparity in half-matched stimuli. In a), the density of the stimulus was 5%, while in b) the density was 24%. The inset shows a magnified view of the response to half-matched stereograms. The model cell is simply a traditional binocular energy model with an additional squaring nonlinearity on the output [Henriksen et al., 2016a]. Increasing the dot density decreases the local correlation variability and thus decreases disparity tuning to half-matched stimuli. The inset in b shows a zoomed-in view of the half-matched tuning curve. Due to the squaring output nonlinearity, the anticorrelated responses around the flanks of the disparity tuning curve are also increased. Each point in a given tuning curve was generated by averaging the responses of the model cell to 10 000 unique dot patterns.

2.3.2   Neuronal responses

We recorded extracellular activity of 53 isolated disparity-selective V1 neurons in response to correlated, anticorrelated and half-matched dynamic random dot stereograms, while two macaque monkeys maintained fixation. We used two dot densities - 5% and 24% - to test the model predictions that the magnitude of disparity tuning to half-matched stimuli should decrease with increasing dot density. Figure 2.3a shows an example disparity tuning curve for a cell in response to 5% dot density stimuli. As in the model, this cell has asymmetric correlated and anticorrelated tuning curves, and a peak in its response to half-matched stereograms at the preferred disparity of the cell. In response to 24% dot density stimuli (Figure 2.3b), the cell's half-matched tuning decreases visibly, while the correlated and anticorrelated responses remain largely unchanged.

In order to quantify the magnitude of disparity tuning to half-matched and anticorrelated stimuli relative to the correlated response, we computed the regression slope between the correlated and half-matched responses (half-matched slope) and between the correlated and anticorrelated responses (anticorrelated slope). Figure 2.3c and d shows this for the 5% and 24% density stimuli, respectively.
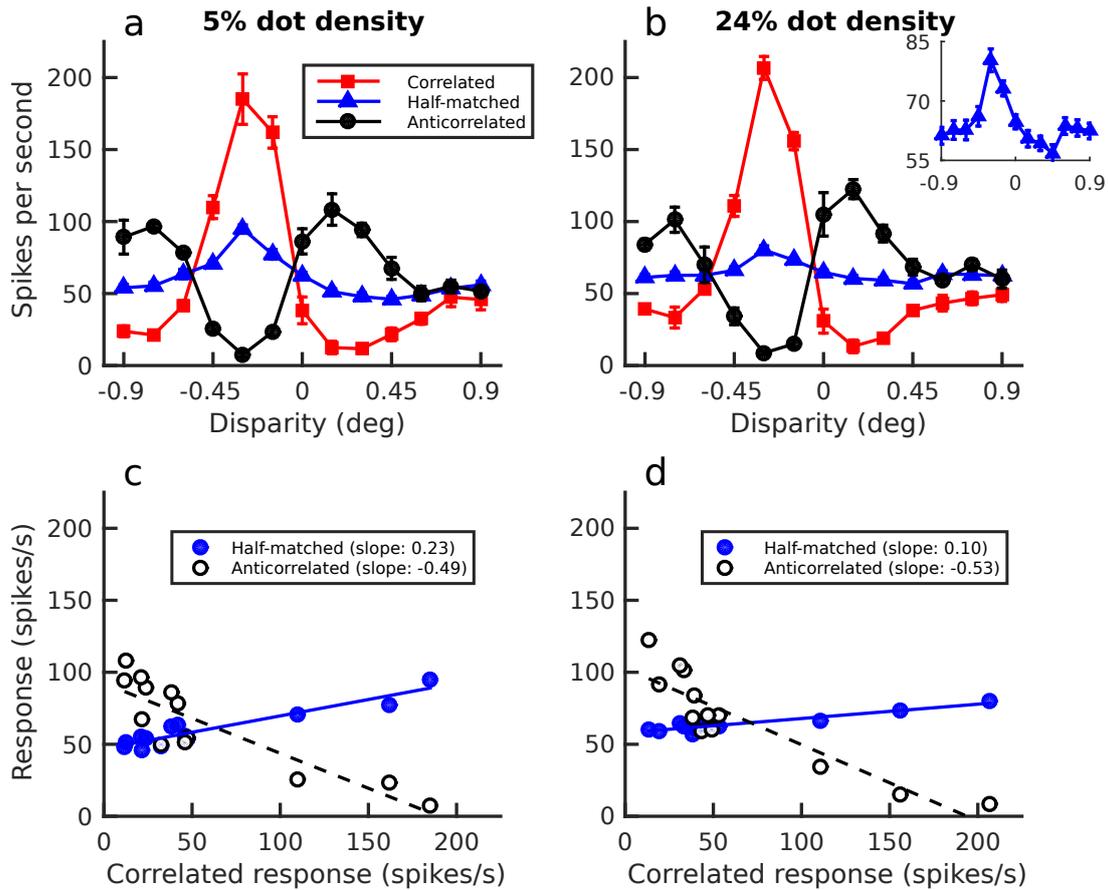
Figure 2.3: Example tuning curves for a cell that is selective for disparity in dynamic half-matched stimuli for both 5% dot density (a) and 24% dot density (b) stimuli. Dot width was 0.3°, and RF width was 1.05°. In b, the inset shows a zoomed in view of the half-matched response. No inset is shown for a as the response is sufficiently large to see unaided. Error bars show ±1 SEM. The lower panel shows the half-matched and anticorrelated responses plotted as function of the correlated response for 5% density (c) and 24% density (d). Lines show type 2 regression fits. 95% bootstrap confidence intervals for half-matched slope was [0.197,0.27] for 5% density, and [0.076,0.125] for 24% density. For anticorrelated slope, the confidence intervals were [-0.591,-0.45] for 5% density and [-0.637,-0.492] for 24% density.

In the example cell shown in Figure 2.3, the anticorrelated slope is ≈ 0.5 for both densities tested. This is typical: across the population, anticorrelated slopes did not differ significantly for 5% vs 24% density ($t(52) = 0.97$, $P = 0.34$, paired t-test).

In contrast, half-matched slopes do depend on dot density. The half-matched slope in the low density case is 0.23 (95% bootstrap CIs: [0.197,0.27]), meaning that the magnitude of half-matched disparity tuning is 23% of that for correlated disparity. In the high density case, the half-matched slope is 0.1 (95% CIs: [0.076,0.125]), or 10% of the correlated tuning. In other words, the strength of disparity tuning has approximately halved in response to increasing the dot density (i.e. decreasing the correlation variability), yet remains significantly greater than 0.

Figure 2.4 summarises this result across the population, showing the half-matched slope as a function of disparity tuning strength, which is quantified with the Disparity Discrimination Index (DDI). The DDI ranges from 0 to 1 and is a measure of a cell's disparity tuning (Prince et al., 2002). Figure 2.4a shows that there is no significant correlation between the DDI and the half-matched slope of a cell for low density (r = -0.02, P = 0.91, Pearson correlation), and only a modest relationship between DDI and half-matched slope in the high density stimuli (Figure 2.4b, r = 0.34, P = 0.01, Pearson correlation). This latter observation might reflect the higher signal-to-noise ratio in neurons with higher DDIs. Under the null hypothesis that V1 cells are, on average, not tuned to disparity in half-matched stereograms, the distribution of half-matched slope values should be centered on 0. In Figure 2.4 the mean half-matched slope is significantly greater than zero for both densities (5%: M=0.14, $t(52) = 11.46$, $P < 10^{-15}$; 24%: M = 0.04, $t(52) = 6.76$, $P < 10^{-7}$). This is also true for both subjects when we consider their data separately (Lem 5%: M=0.15, $t(27) = 8.75$, $P < 10^{-8}$; Lem 24%: M=0.04, $t(27) = 4.55$, $P < 10^{-3}$; Jbe 5%: M=0.14, $t(24) = 7.33, P < 10^{-6}$; Jbe 24%: M=0.03, $t(24) = 5.35, P < 10^{-4}$). Neurons which exhibit significant disparity tuning to half-matched stimuli are shown as red triangles, while those that did not are shown as green circles. For low dot density stimuli (Figure 2.4a), 34/53 cells exhibit significant half-matched disparity tuning, while for high dot density stimuli (Figure 2.4b), 11/53 cells show significant tuning. Thus, on average, V1 neurons transmit a systematic disparity signal even in 24% density half-matched RDSs.
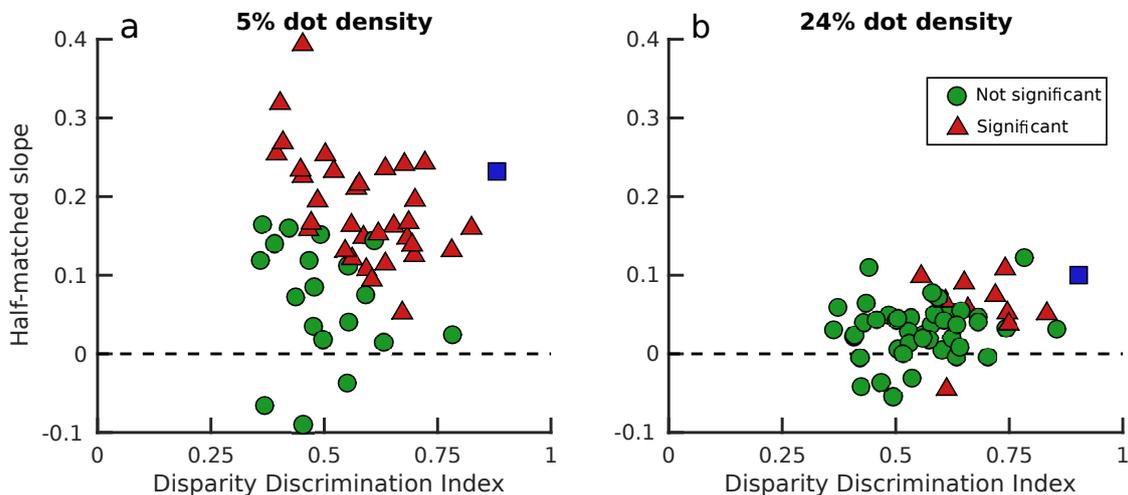
Figure 2.4: Half-matched slope (i.e., the regression gradient of responses to half-matched stimuli as a function of the responses to correlated stimuli) plotted as a function of Disparity Discrimination Index (DDI) for correlated stereograms for 5% density (a) and 24% density (b). The red triangles show cells that exhibit significant disparity tuning to half-matched stimuli at the $P < 0.01$ level (34/53 cells in a, and 11/53 cells in b). The blue square shows the example cell in Figure 2.3. The dashed black line shows the expected half-matched slope under the null hypothesis that V1 cells are not, on average, disparity-tuned to half-matched stimuli. In both plots the points deviate significantly from this prediction.

As noted above, our model predicts that there should be a correlation between the magnitude of half-matched tuning at different dot densities. We do find a moderate correlation between the half-matched slopes at 5% and 24% density ($r = 0.43$, $P = 0.001$, Pearson correlation). Our model also predicts that half-matched tuning should be weaker for stimuli with higher dot density, since these have smaller fluctuations about the mean correlation level of zero. The difference between the 5% and 24% density slopes is indeed highly significant (M=0.14 for 5% vs M=0.04 for 24%, $t(52) = 9.51$, $P < 10^{-12}$, paired t-test). This was also true for both monkeys when considered separately ($P < 10^{-6}$ in both cases).

In our simple model, the magnitude of disparity selective responses depends on the dot size, the dot density, and the receptive field size, since all of these things alter the local variation in correlation (Henriksen et al., 2016a). Despite this, the model predicts a unique relationship between the slope of responses to half-matched vs correlated stimuli observed at 5% density and that at 24% density. Two different combinations of RF size and dot size that produce the same slope at 5% density will also produce the same slope at 24% density. This arises because the only factor that determines the response magnitude for half-matched stimuli relative to correlated stimuli is the variance in local correlation (other factors, such as contrast or spatial frequency content would affect responses to both stimuli equally). Importantly, this means that the model predicts the relationship between slopes (as a function of density) without any fitting of parameters. We show this expected relationship between the half-matched slope for the two dot densities in Figure 2.5 (red line). Although there may be a deviation at large slope values (> 0.3), we have too few neurons with these responses to be clear that this really is a model failure. As a result, over the observed range, the quantitative success of the model is mainly in describing the mean slope magnitudes, rather than the shape of any relationship. Nonetheless, since the model prediction was made without any parameter fitting, this success provides strong evidence that V1 cells signal disparity in these stimuli by exploiting fluctuations in local correlation within the RF. Note that if responses to half matched stimuli represented a contribution from a pure "matching computation" [Doi et al., 2011, Doi et al., 2013, Abdolrahmani et al., 2016, Henriksen et al., 2016a], the data in Figure 2.5 should lie on the identity line, which they do not.

2.3.3 Testing more general models of a single mechanism

The quantitative prediction shown in Figure 2.5 is specific to our particular model: the binocular energy model with a squaring output nonlinearity. But for a wide range of models in which a cell's half-matched response reflects its averaged response to positive and negative fluctuations in binocular correlation, there should be a relationship between a cell's attenuation to anticorrelated stimuli and the magnitude of the half-matched tuning. We assess the attenuation using the anti-correlated slope (i.e. the gradient of the regression line when anticorrelated responses are plotted
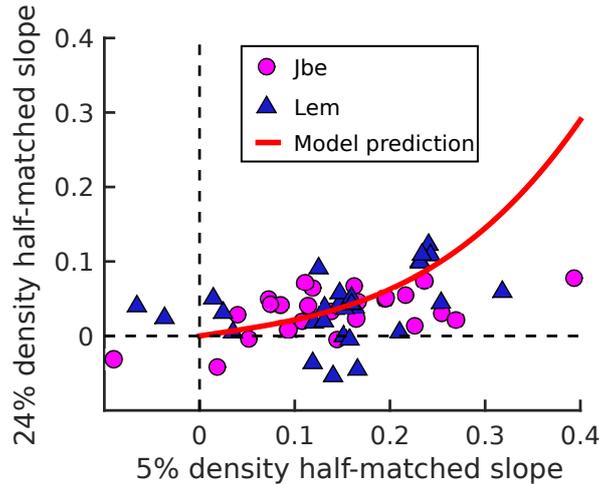
Figure 2.5: Comparing half-matched disparity tuning for two dot densities (5% and 24%). Tuning strength is quantified with the half-matched (HM) slope. Data are shown from 53 cells. Blue triangles show data from monkey Lem, while magenta circles show data from monkey Jbe. The solid red line shows the prediction of the modified binocular energy model introduced earlier (with no free parameters). The model prediction was obtained by computing the 5% and 24% density responses of model cells with different RF sizes.

against correlated). In neurons where responses to anticorrelation shows no attenuation, the mean response to a mixture of correlations with a mean of zero is the same as the response to zero correlation, and so a straightforward prediction is that there should be no tuning for half-matched stimuli: the half-matched slope should be zero when the anticorrelated slope is -1. As the modulation to anticorrelated stimuli gets weaker, this averaging allows fluctuations in correlation to produce stronger responses to half-matched stimuli at the preferred disparity (although responses to half-matched stimuli will always be near-zero when fluctuations are small, e.g. if receptive fields are large compared to dot size [Henriksen et al., 2016a]). Thus, the range of possible half-matched slopes should be maximal when the anticorrelated slope is zero (or positive). In the low density stimuli (Figure 2.6a), there is some support for this. There is a weak positive correlation between the two ($r = 0.25$, $P = 0.07$, Pearson correlation), although this marginally fails to reach significance. For high densities (Figure 2.6b), this trend is not evident or even reversed (r = -0.21, P = 0.13, Pearson correlation). However, there are a number of reasons why this relationship may be obscured. For example, receptive field size affects half-matched slope without affecting anticorrelated slope. Additionally, since the half-matched slopes are all small, it may require considerably more statistical power to reveal any relationship. We have sufficient power to demonstrate that these cells are on average disparity tuned to half-matched RDSs at 24% density, but not for more sophisticated analyses.

In Figure 2.6, the red and blue crosses show the predictions of idealized correlation and matching computations, respectively. A pure correlation computation, such as the BEM, would have perfectly inverted response to anticorrelated, and consequently no response to half-matched (anti-
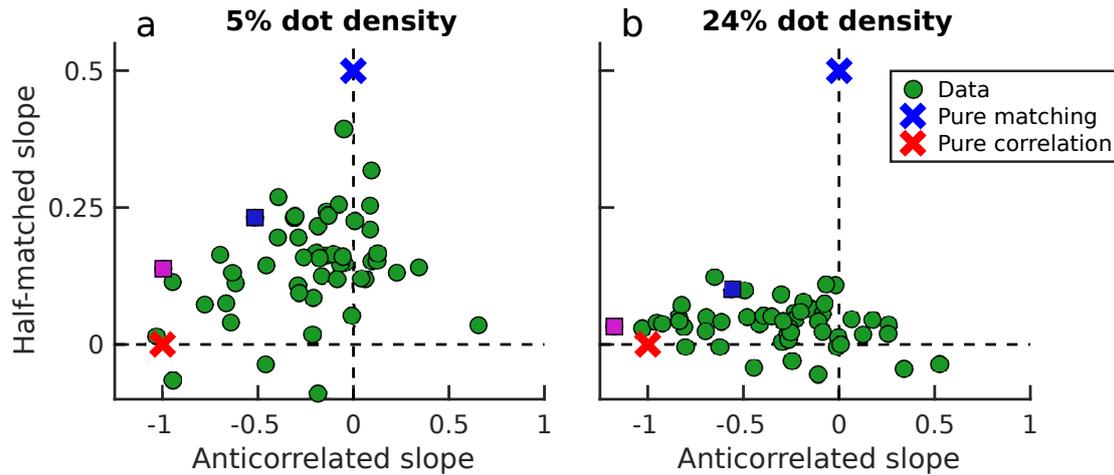
Figure 2.6: Half-matched slope as a function of anti-correlated slope for 5% density (a) and 24% density (b). The blue square shows the example cell in Figure 2.3, and the magenta square shows the example cell in Figure 2.7. If the half-matched response can be inferred from the correlated and anticorrelated responses alone then there should be a positive correlation between half-matched and anticorrelated slope (i.e. less attenuation to anticorrelation would imply smaller half-matched slope). While there is a trend bordering on significance for low density (a), no such relationship is evident in the high density stimuli (b). The red and blue crosses show the expected performance of idealized correlation and matching computations, respectively.

correlated slope of -1, half-matched slope of 0). A pure matching computation would not modulate its response at all to anticorrelated dots, but would have a half-matched amplitude which is half its correlated amplitude (anticorrelated slope of 0, half-matched slope of 0.5). This is clearly not a veridical characterization of the neurophysiological data, which shows instead a cloud centered in between these two extremes and which changes with stimulus parameters such as dot density. This is consistent with the view that disparity tuning in V1 arises from a single nonlinear correlation computation, which can be roughly approximated by appending a squaring on to the BEM.

Many neurons in Figure 2.6 have anticorrelated slopes near 0 or even greater than zero, suggesting there may be a subpopulation of neurons with no disparity-selective response to anticorrelated dots, which seems at odds with the observations in Cumming and Parker (1997). This apparent discrepancy reflects two factors: First, some neurons do show clear modulation to anticorrelated stimuli but without any inversion. Some show tuning of similar shape (these are shown with phase shifts near 0 in Cumming and Parker 1997, and have slopes > 0 here), and some show shapes that differ in other ways (phase shifts neither 0 or π). Second, random fluctuations in a neuron showing no systematic response will produce slope values scattered around zero here, but inevitably produce amplitude ratios > 0 when using fitted Gabor functions.

A less stringent version of the model prediction in Figure 2.5 is that the response to half-matched stimuli should be less than or equal to the average of the correlated and anticorrelated responses.

Only one cell in our dataset deviated significantly from this prediction. This cell, shown in Figure 2.7a for 5% density, shows completely symmetric tuning curves to correlated and anticorrelated stimuli (i.e. an anticorrelated slope not significantly different from -1), yet has a half-matched slope of 0.14 (95% CI: [0.103,0.173]). In other words, this cell's response to half-matched stimuli is greater than that predicted from the average of its correlated and anticorrelated responses. For 24% density stimuli (Figure 2.7b), the cell has an anticorrelated slope that is significantly lower than -1, yet its half-matched slope is again significantly positive (95% CI: [0.005, 0.06]). This means that the cell's half-matched tuning is opposite to that produced by a random mixture of responses to correlated and anticorrelated stimuli. These responses are rare, so it is possible that these cells process disparity in a way that is different from other cells in striate cortex. Alternatively, it may be that our model is too simple to fully describe the behavior of V1 neurons, a point we return to in the discussion. Nonetheless, in 52/53 neurons, the 95% confidence interval for the half-matched slope included the value predicted by the model.

2.3.4   Neurometric performance

The analysis above demonstrates that neurons in V1 do carry a weak but systematic signal about disparity in half-matched stereograms. This analysis does not demonstrate whether the disparity tuning is sufficient to account for psychophysical behavior. We chose our high density (24%) because that value has been used in previous psychophysical studies [Doi et al., 2011, Doi et al., 2013, Henriksen et al., 2016a]. If the weak tuning to half-matched stimuli we find with this density is not sufficient to account for psychophysical performance, it might be necessary to postulate a separate matching computation, as hypothesized in the literature [Doi et al., 2011, Doi et al., 2013, Doi and Fujita, 2014]. To evaluate neuronal performance, we computed the neurometric performance of the cells using the area under the receiver operator characteristic curve (AUROC). The ROC curve was computed for each cell by comparing responses to its preferred disparity (i.e. the disparity where the cell had the highest mean firing rate) and responses to its null disparity (i.e. disparity with lowest mean firing rate). Preferred and null disparities were defined on the basis of responses to correlated stereograms. The AUROC values are shown in Figure 2.8a for 5% dot density stimuli and in 2.8b for 24% stimuli. This then estimates how reliably an ideal observer could discriminate a half-matched stimulus at the preferred disparity from one at the null disparity, given only the spike counts of the neuron. These can then be compared with psychophysical performance, also expressed as % correct. The neurometric performance is lower than the published performance of human observers. Human performance is often above 80% correct on half-matched stereograms, although there is substantial variability between individuals [Doi et al., 2011, Doi et al., 2013]. However, there are a number of important differences between the stimulus conditions used in the psychophysics and that used here. Most importantly, the published human studies used foveal viewing of stimuli that were much larger than typical foveal receptive fields, giving them much more
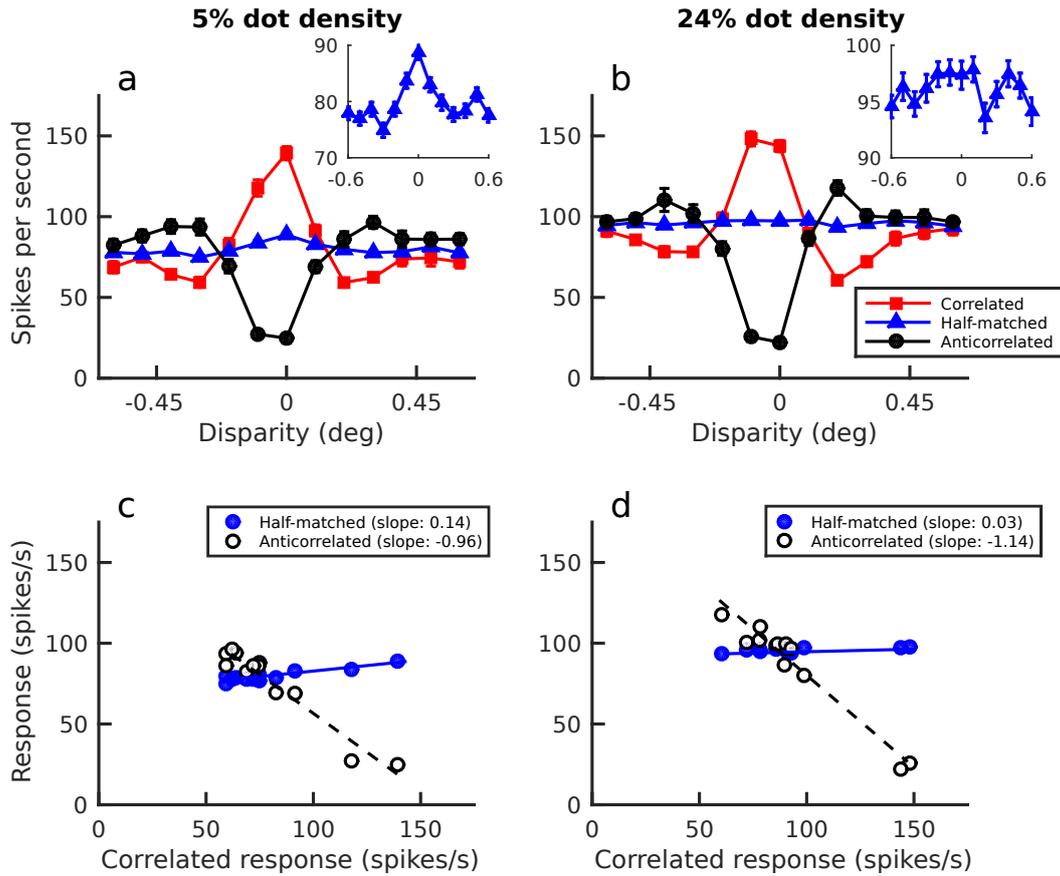
Figure 2.7: An unusual example cell that exhibits significant disparity tuning to half-matched stimuli despite having symmetric tuning curves to correlated and anticorrelated dot-patterns. a) and b) shows the tuning curves for 5% and 24% density dot-patterns, respectively. Error bars show ±1 SEM. 95% bootstrap confidence intervals for the half-matched slope was [0.103,0.173] for 5% density and [0.005, 0.06] for 24% density. The corresponding confidence intervals for anticorrelated slope was [-1.147,-0.876] and [-1.319,-1.066] for 5% and 24%, respectively. To 24% density stimuli (b), the cell has an anticorrelated slope that is significantly less than -1 yet still exhibits significant disparity-tuning to half-matched stimuli. The insets in a and b show a zoomed in view of the half-matched response. The lower panel shows the half-matched and anti- correlated responses plotted as a function of the correlated response for 5% and 24% density (c and d, respectively).

information than any single V1 neuron. We trained one of our animals to perform a discrimination task, and then measured performance using stimuli matched to those used during recording. For the recording sessions, the stimuli used at a given eccentricity were identical except for small changes in position (necessary to center the stimulus on recorded RFs). The psychophysics used the same stimulus configuration, with the location set to the mean of those used in the recording sessions. The animal performed at 70% correct at the eccentric location, and 65% correct at the more foveal location. This stimulus was larger than typical receptive fields (chosen to ensure that the RFs of all cells recorded in a session were covered by the stimulus, even when considering fixational eye movements). We therefore repeated the psychophysical measures changing only the size of the region with disparate dots to match measured RF sizes. (RF size was estimated by the standard

deviation of a Gaussian fit to the measures of minimum response field. The stimulus diameter was set to be 8 times the mean of these standard deviations, still more than adequate to cover the receptive field). Here the animal achieved only 51% correct, poorer than the mean AUROC (and not significantly greater than 50%). Thus, when care is taken to match the information available to individual neurons and the psychophysical observer, the ability of single neurons to detect disparity in half-matched stereograms is sufficient to account for psychophysical performance.
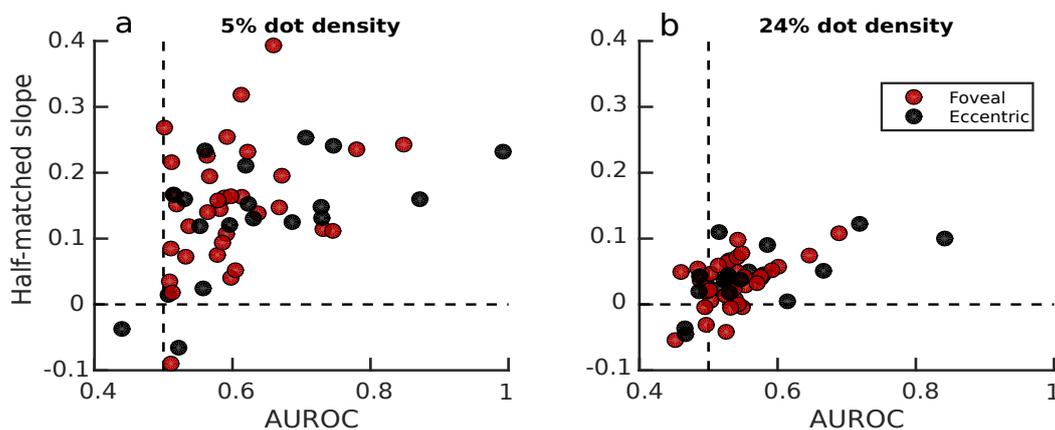
Figure 2.8: Area under the receiver operating characteristic curve (AUROC) for all cells for 5% dot density (a) and 24% dot density (b). Red dots show data from more foveal recordings ($< 5°$ from fixation), and black dots show data from eccentric recordings ($> 10°$ from fixation). Only monkey Lem had eccentric recordings. The AUROC estimates the percent correct on near/far task that can be achieved using the spike counts from a single neuron (see Methods).

2.4   Discussion

Disparity-selective V1 cells probably provide the initial substrate for binocular depth perception, at least in dense cyclopean stimuli such as random dot stereograms (RDSs). Disparity-selective cells in V1 appear to carry out a local correlation-based computation, similar to that described by the binocular energy model (BEM). Depth perception in half-matched random dot stereograms - stimuli with an equal number of correlated and anticorrelated dots - has been proposed as evidence that a separate stereo matching computation operates in cortex [Doi et al., 2011, Doi et al., 2013, Doi and Fujita, 2014]. This is based on the observation that a computation that modulates its response strictly as a linear function of correlation, such as the BEM, cannot report depth in these stimuli. However, it is well-known that disparity-selective cells in V1 often have attenuated responses to anticorrelated stimuli, which is also unlike the BEM. We have previously shown that a simple model that reproduces attenuated response to anticorrelated RDS can also produce disparity selectivity for half-matched RDS [Henriksen et al., 2016a]. This raises the possibility that V1 neurons might signal disparity in half-matched stereograms. Here, we show that disparity-selective neurons in primate V1 do show systematic disparity selectivity to half-matched RDSs. These properties suggest that V1 neurons carry out a nonlinear correlation computation, inter-

mediate between a "pure correlation" and "pure matching" computation. We propose that these cells are the initial neuronal substrate for depth perception in half-matched RDS. This nonlinear response to binocular correlation may represent the effect of mechanisms that reduce responses of V1 neurons to "false" matches [Henriksen et al., 2016c].

In the model which prompted this work, this tuning arises from fluctuations in the local binocular correlation within the receptive field. Any stimulus manipulation which decreases the local correlation fluctuations should decrease the magnitude of the model's disparity tuning to half-matched stimuli. In our experiments, we decreased correlation fluctuations by increasing dot density. We found that this reduces half-matched disparity tuning in real neurons, as predicted by the model. It is noteworthy that a number of psychophysical observations suggest that local correlation fluctuations are also required for depth perception [Doi et al., 2013, Henriksen et al., 2016a], providing further evidence that V1 neurons are indeed the neural substrate for the psychophysics.

Doi et al. (2014) have proposed a particular instantiation of a matching computation, known as "cross-matching". This is closely related to the BEM, but only contains a half-wave rectified binocular term. If one incorporates monocular terms into this model, then this is very similar to the squared model we have used here. Our choice for the squaring is simply that it is a variant of the BEM that has been explored multiple times [Read et al., 2002, Tanabe and Cumming, 2008, Henriksen et al., 2016a], and that the squaring gives a clear algebraic dependence on variance. The choice of nonlinearity is therefore not a significant difference between these studies [Henriksen et al., 2016a]. The distinguishing claim by Doi et al. is not that there are cells whose response is a nonlinear function of correlation (this was shown in Cumming & Parker, 1997), but rather that "Two distinct computations feed the disparity signals for stereoscopic depth perception. One computes disparity based on binocularly matched patterns, while the other computes the cross-correlation of binocular images." ([Doi et al., 2011], p. 11). The fact that neurons at the very first stage of disparity processing respond to both types of signal suggests that the two computations may not be distinct.

A recent study found that V4 neurons also respond selectively to disparity in half-matched stereograms [Abdolrahmani et al., 2016]. Given the results we present here, it is possible that the responses they report are simply inherited from V1 neurons. In principle, the effects of dot density that we demonstrate in V1 might be used to determine if responses in extrastriate cortex simply reflect a summation over V1 inputs. Responses in extrastriate cortex should show a similar dependence on dot density. However, quantitative predictions are difficult without precise information about the properties (especially RF size) of the set of V1 inputs to a given neuron.

Fluctuations in binocular correlation result in disparity tuning to half-matched stimuli in any

system which shows attenuated responses for anticorrelated patterns (such as real V1 neurons, [Cumming and Parker, 1997]). Therefore, increasing the variability of the correlation will increase the mean response. This is true regardless of the mechanism that produces the attenuation. For our quantitative modelling, we used a very simple modification to the BEM (a squaring output nonlinearity). This nonlinearity was chosen simply because the expected value of a squared random variable has a clear algebraic dependence on variance. Many other nonlinearities would also produce such a dependnece. However, there are several reasons to believe this simple model is not an accurate description of the mechanism producing attenuation in V1 neurons [Cumming and Parker, 1997, Read et al., 2002, Haefner and Cumming, 2008, Tanabe et al., 2011a]. Possibly as a result, some quantitative aspects of the data were not captured well by this model (e.g. the lack of a clear relationship between anticorrelated slope and the range of half-matched slopes in Figure 2.6). It is particularly worth noting that most V1 neurons behave as if they sum multiple subunits each of which resembles a BEM [Tanabe et al., 2011a, Tanabe and Cumming, 2014], and that many of these subunits have suppressive effects. If the asymmetrical response to correlation/anticorrelation is different within each subunit, our simplified model is unlikely to reproduce the neuronal behavior.

Although we show that there is a weak signal in V1 neurons in response to half-matched RDSs, this on its own does not prove that the signal is sufficiently strong to account for psychophysical performance. Comparisons of neuronal and psychophysical behavior typically compare neurometric and psychometric thresholds [Britten et al., 1992, Parker and Newsome, 1998, Prince et al., 2000, Uka and DeAngelis, 2003, Nienborg and Cumming, 2006, Nienborg and Cumming, 2014, Gu et al., 2008]. For half-matched stimuli, this is harder to do because the sensation of depth is very weak. In many subjects, no disparity, however large, produces 100% correct performance. As a result, there are no published psychometric thresholds for disparity in half-matched stimuli. We therefore compared neurometric and psychometric performance for a single disparity value (many times threshold in correlated stimuli), using the area under the receiver operating characteristic curve (AUROC) as a measure of neurometric performance. We found that that the most selective neurons match psychophysical performance, but the majority are substantially poorer. However, these psychophysical measures were made with stimuli much larger than typical V1 RFs. In one animal we measured performance with a stimulus only double the measured size of the RFs, and found that performance was then poorer than most neurons. Stimulus size may play a particularly important part in half-matched stimuli, where random fluctuations in the stimulus are the only source of a useable signal. As these are independent at different locations, the useful signal increases with size. It therefore seems likely that the neurometric performance of the V1 cells reported here is more than enough to account for the psychometric performance of human and monkey observers.

Although disparity-selective cells in V1 seem to explain depth perception in half-matched RDSs, they may not explain all aspects of stereoscopic depth perception. One case is binocular stimuli

in which the left and right images contain isolated monocular targets. Here, subjects can report the depth sign for disparities that are larger than any V1 neuron has been shown to signal [Ogle, 1952c, Westheimer and Tanzman, 1956]. This may depend on signals in V1 that are separate from those carried in disparity-selective neurons (such as monocular responses). Nonetheless, in dense stimuli, such as RDS, it seems that disparity-selective signals in V1 provide a substrate that is sufficient to support psychophysical performance in most disparity-based tasks that have been studied.

In summary, the responses of disparity-selective V1 neurons resemble the energy model in that their response depends on the correlation between the left and right images. They differ in showing weaker modulation to anticorrelated stimuli than correlated stimuli. In principle, this asymmetry could lead to discernible responses to half-matched RDS, despite the fact that the mean binocular correlation is 0, and indeed V1 neurons seem to behave in this way. Depth perception to half-matched RDSs is therefore compatible with the view that disparity-selective neurons in striate cortex provide the substrate for stereo depth perception in dense cyclopean stimuli.

Chapter 3

Correlated boosting: a specialised computation for disparity extraction in primary visual cortex

3.1   Introduction

A major unsolved question in the stereo literature is how exactly disparity-selective cells obtain their disparity tuning. This was first raised by Cumming & Parker (1997) who computed disparity tuning curves for correlated and anticorrelated random dot stereograms, and showed that real cells have weaker modulation to anticorrelated stimuli than to correlated stimuli [Cumming and Parker, 1997]. This property is not captured by the binocular energy model (BEM) and a large number of subsequent papers have attempted to explain the reason for this discrepancy [Nieder and Wagner, 2000, Read et al., 2002, Read and Cumming, 2003, Read, 2005, Tanabe and Cumming, 2008, Tanabe et al., 2011a, Tanabe et al., 2011b, Samonds et al., 2013, Henriksen et al., 2016a, Henriksen et al., 2016c, Henriksen et al., 2016b]. With a few notable exceptions [Read et al., 2002, Samonds et al., 2013], these models generally take the form of a linear-nonlinear (LN) cascade. LN cascades begin with LN subunits, which consist of an initial linear filtering step followed by a subsequent nonlinearity, such as half-wave rectification or half-squaring. The models are known as "cascades" since the LN elements are then combined linearly before being passed through a final output nonlinearity. This model architecture is shown schematically in Figure 3.1. These models are very similar to a shallow (three-layer) neural network, and, when used in the binocular domain, implement a generalised version of the binocular energy model. We therefore refer to this broad class of models as the generalised binocular energy model (GBEM).

It is currently unknown whether generalised binocular energy models can capture the attenuated anticorrelated response seen in real cells. Goncalves & Welchman (2017) recently trained a three-layer binocular convolutional neural network to detect disparities in natural images. The authors found that when trained on these stimuli, the model units had disparity tuning curves with attenuated responses to anticorrelated stimuli, much like real V1 neurons [Goncalves and Welchman, 2017]. The earliest attempt to do something similar in real cells was carried out by Anzai, Ohzawa, & Freeman (1999), who carried out singular value decomposition (SVD) on the binocular interaction receptive fields of real neurons. While the authors successfully recovered binocular
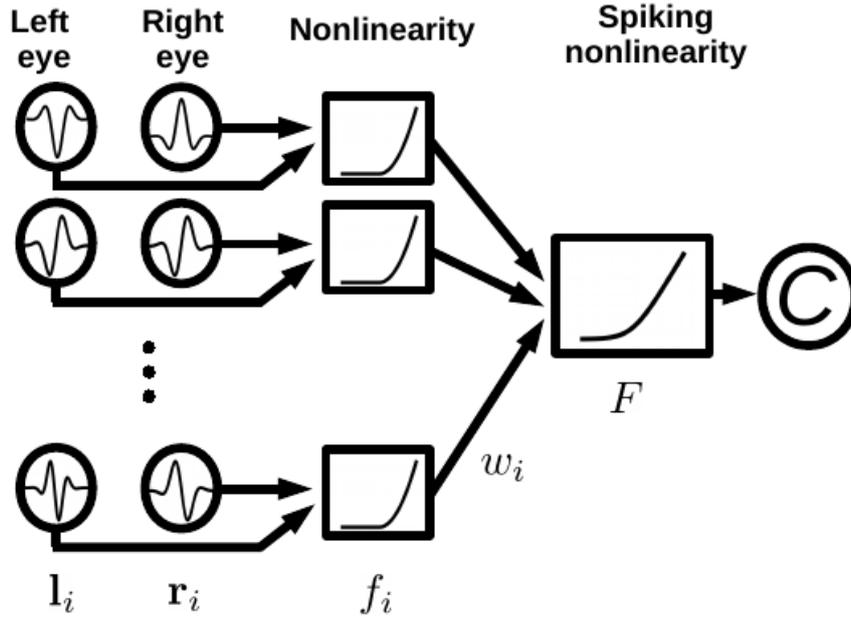
Figure 3.1: The generalised binocular energy model (GBEM). Starting from the left, the image is linearly filtered in the left and right eyes; the filtered responses $L_i = \rho_L^{(i)} \cdot s_L$ and $L_i = \rho_R^{(i)} \cdot s_R$ are then combined and passed through a nonlinearity $f_i(L_i + R_i)$. Finally, a weight w_i is given to each subunit, before the responses of all subunits are summed together and passed through a final spiking nonlinearity F.

subunits using this technique, they did not include anticorrelated stimuli [Anzai et al., 1999], and so are unable to comment on this aspect of neuronal computation. The most comprehensive effort to directly model the responses of V1 neurons to both correlated and anticorrelated stimuli based on spiking data was carried out by Tanabe & Cumming (2011). The authors performed a spike-triggered analysis of covariance using independent binocular Gaussian noise as their stimulus. Tanabe & Cumming found that their GBEM units did not exhibit the attenuated anticorrelated response to random line stereograms which was exhibited by their neuronal counterparts. However, when the V1 cells were tested on the independent binocular Gaussian noise (with which the models were fit), the cells did not show attenuated responses to negatively correlated stimuli either [Tanabe et al., 2011b]. It is perhaps less surprising that the model units did not show the attenuated anticorrelated response if this was not present in the data to begin with.

It is at present unclear whether real neurons achieve their attenuation to anticorrelated stimuli through an LN-cascade. The purpose of the present work is to address the discrepancy between the works of Tanabe & Cumming (2011) and Goncalves & Welchman (2017). Using recently developed optimisation routines [McFarland et al., 2013], we fit GBEM units to neuronal data recorded from V1 in the macaque. The key advance from Tanabe & Cumming (2011) is that this method allow us the fit GBEM units even when the image sequences are not pure noise. For the current purposes, this means that sequences of different disparities, and with correlations of $+1$ or -1 can be used.

The consequence of this is that unlike in Tanabe & Cumming (2011), the attenuated anticorrelated response is present in the neuronal data. Importantly, the attenuated anticorrelated response is recovered by our model subunits. We show that the attenuated anticorrelated response seen in real neurons is in fact a boosting of true matches, as opposed to a suppression of false matches. The extent of the correlated boosting is underestimated by even the best fitting GBEM, suggesting that the mechanism for correlated boosting cannot be captured by a simple LN cascade.

3.2 Methods

3.2.1 Animal subjects

Two male macaques (Macaca mulatta) were implanted with scleral search coils, head posts, and a recording chamber under general anaesthesia. The full procedure is described elsewhere [Cumming and Parker, 1999, Read and Cumming, 2003]. For the experiment, subjects viewed two gamma-corrected CRT monitors through a custom mirror haploscope. The subjects were required to maintain fixation on a central box in order to receive a reward. All experiments were performed at the National Institutes of Health in the US, and complied with the US Public Health Service policy on the use and care of animals. The protocols received approval by the National Eye Institute Animal Care and Use committee at the National Institutes of Health.

3.2.2 Recording

We recorded extracellular activity from neurons in primary visual cortex (V1) using laminar multi-contact electrodes (U-probes, Plexon for monkey Jbe; V-probes, Plexon for monkey Lem). Each electrode had 24 linearly arranged probes spaced 50 μm apart. The data was sampled with Spike2 (Cambridge Electronic Design), and the full waveform data was saved to disk for offline analysis. Spikes were subsequently analysed offline using custom spike-sorting software. Neurons were included in the analysis if they were well-isolated and disparity-tuned (as determined by a permutation test, $P < 0.01$). 95/197 cells met these criteria, giving 65 cells from monkey Lem, and 30 from monkey Jbe.

3.2.3 Stimulus

The stimulus apparatus was the same as in Chapter 2. The animal was shown a 1D random noise pattern consisting of either black, gray, or white bars. The orientation of the stimulus was chosen so as to match the orientation preference of the cell (measured using circular patches of 1D noise at zero disparity) as closely as possible. When the stimulus orientation was sufficiently different from the preferred orientation (e.g. because there were multiple cells with different orientation preferences), there was typically no disparity selectivity to the 1D noise stimulus. The bars were 0.0946° in width, and the pattern consisted of 42 bars in each eye. The stimulus could be either

binocularly correlated, anticorrelated, or uncorrelated. Stimulus disparities were selected based on the disparity tuning observed in the first measurements after fixing the orientation. Disparity was applied to the stimulus by wrap-around (i.e. bars displaced off the right end of the stimulus would be appended to the beginning of the stimulus; this has the effect of keeping the frequency power spectrum the same for the left and right images). We only used disparities which were integer multiples of the bar width, and always applied disparity orthogonally to the orientation of the bar pattern. A new stimulus pattern with a new disparity and correlation was shown every 30ms. A single trial lasted for 3s, corresponding to 100 independent noise patterns. We also implemented a two-pass procedure by duplicating trials, such that the same exact sequence of noise patterns occurred twice for most trials. This was done to facilitate the analysis discussed in Chapter 4.

3.2.4 Generalised binocular energy model

We fit a generalised form of the binocular energy model using the framework developed by McFarland & Butts (2013). The model takes the general form

$$C = F\left[\sum_i^N w_i f_i(L_i + R_i)\right],$$ (3.1)

where f_i is the subunit nonlinearity for the ith subunit, w_i is the weight given to the ith subunit (constrained to be either -1 or +1, corresponding to a suppressive or excitatory subunit, respectively), and F is the final spiking nonlinearity of the model unit. The weights of each subunit, dictating the relative contributions of each subunit to the final model response are fit using the filter coefficients. L_i and R_i are the response of the ith left and right filters, and are further defined as

$$
\begin{aligned}
L_i &= \rho_L^{(i)} \cdot s_L, \\
R_i &= \rho_R^{(i)} \cdot s_R.
\end{aligned}
$$ (3.2)

Here s_L and s_R refer to a vector representation of the stimulus presented to the left and right eyes, respectively. $\rho_L^{(i)}$ and $\rho_R^{(i)}$ refer to vector representations of the linear spatiotemporal filters for the left and right eye of the ith subunit. The number of spatial elements in the filter was simply the number of independent pixels in the stimulus, which was 42 for the left and right eyes (84 total). The number of temporal elements in the filter was 15, sampled at 10ms, corresponding to a maximum temporal kernel of 150ms. Thus, the total number of elements for the binocular spatiotemporal filter was 1260. The subunit nonlinearity f_i can in principle take on a range of forms, but for the current purposes we have constrained it to always be a thresholded square. In symbols,

$$f(x; \theta) = \text{Pos}(x - \theta)^2,$$ (3.3)

Where Pos refers to half-wave rectification, and θ is the threshold parameter. The spiking non-linearity F is a softplus rectifier, which is a smoothly varying rectifier function (with well-defined derivatives everywhere). It takes the form

$$F(z; \alpha, \beta, \gamma) = \alpha \ln(1 + \exp[\beta(z - \gamma)]), \tag{3.4}$$

where α, β, and γ approximately correspond to magnitude, slope, and threshold parameters, respectively. The term in equation 3.1 is a continuous firing rate, and the spike count is obtained by passing the continuous rate through some discretisation procedure. For all analyses, we passed the continuous rate through a Poisson random number generator (`poissrnd` in Matlab) to obtain discrete spike counts.

The optimisation routine developed by McFarland & Butts (2013) allows for fitting various components of this framework to empirical data by optimising the (log) likelihood of the model parameters given the data. Specifically, we can directly fit the coefficients of the linear spatiotemporal filters L_i and R_i, the thresholds θ_i of the subunit nonlinearities, and the parameters α, β, and γ of the spiking nonlinearity. While the log likelihood surface is not guaranteed to be convex, McFarland & Butts note that appropriate steps, such as L1 and smoothness regularisation, can in general prevent the routine from converging to local minima [McFarland et al., 2013]. L1 regularisation penalises the L1 norm (the taxicab distance from the origin) of the linear filter coefficients, ensuring that the filter coefficients take on sensible values and also encourages sparsity of the filter coefficients (Tibshirani, 1996). Smoothness regularisation penalises the Laplacian of the filter coefficients, ensuring that the second derivatives of the filters are small everywhere. This prevents abrupt, physiologically implausible changes in the filters (since such changes would correspond to large second derivatives). More generally, both forms of regularisation help prevent overfitting to the training data. The numbers of excitatory and suppressive subunits have to be optimised through cross-validation. In order to do this, we first split the data into a training set and a validation set. In order to prevent leakage of the training data into the validation data, the sets were split by trial instead of frame. For each cell, 75% of trials belonged to the training data, and the remaining 25% of trials belonged to the validation data. All identical trials (i.e. two-pass trials) were kept in the same set (either validation or training) to ensure independence of the two sets. Since there is no straightforward way of determining what was on a cell's receptive field prior to the onset of a trial, we discarded the first 200ms of each trial. We then performed a grid search on the number of excitatory and suppressive subunits, computing the log likelihood of the model on the validation data (the cross-validated log likelihood) for each combination of excitatory and suppressive subunits. Our hyperparameter search space was from 1 to 12 excitatory subunits, and from 0 to 5 suppressive subunits, yielding a total of 72 hyperparameter combinations for each cell. We capped the number of excitatory and suppressive subunits at 12 and 5, respectively, since we observed no cases of cells which were best modelled by 12 excitatory or 5 suppressive subunits. We

repeated this procedure at least 3 times for each excitatory-suppressive combination, and used the mean cross-validated log likelihood for model selection. The hyperparameter combination with the highest mean cross-validated log likelihood was used in the subsequent analysis. In general, the cross-validated log likelihood is fairly stable across iterations and so the best parameter combination is not greatly affected by either number of repeats or aggregation rule (i.e. mean, min, max, and median all give very similar results). For each cell, we obtained cross-validated spike count predictions by running five-fold cross-validation. We first split the data into five equal subsamples (by trial, as noted above), and then fitted the GBEM on four of the subsamples. This allows us to obtain spike count predictions for hitherto unseen data (i.e. cross-validated predictions). We did this for each of the subsamples, resulting in cross-validated predictions for all trials in our dataset. Note that this results in predictions from slightly different models, but allows us to test this set of model parameters as opposed to a particular instantiation of the GBEM.

The disparity tuning curve captures how well a GBEM unit can capture responses across disparities. In order to describe how the GBEM units captures the cell's responses *within* disparities (e.g. as in Figure 3.4b, d, and f), we first identify frames with the appropriate disparity/correlation, and then run the forward correlation procedure as specified in the Methods. This gives us a set of spike counts associated with each frame for both the cell and the GBEM. We then bin the model spike counts with an equal number of samples in each bin, and then compute the average spike count in the cell for each bin. For a perfect model, these points should lie on the identity line. More generally, if a model unit captures the cell's response within a disparity, then the predicted firing rate and the observed firing rate should be closely related.

In order to compare our model responses to the population responses of cells in Tanabe & Cumming (2011), we also computed the responses of the model units to binocular Gaussian 1D noise. Each pixel in these stimuli had a value which was drawn from a normal distribution with unit variance, and was independent in the two eyes and also independent from frame to frame. A new pattern was generated every 10ms.

3.2.5   Disparity tuning curves

In order to compute disparity tuning curves, we performed a forward correlation analysis. For a given disparity and correlation, we first identified all the patterns which were presented with the given stimulus parameters. For each pattern, we computed the number of spikes observed in bins $t_{\max}-1$, $t_{\max}$, and $t_{\max}+1$, where $t_{\max}$ is the time bin where we observed the largest variance across disparities. In other words, we computed the response in a 30ms window around the peak response for each neuron. The mean spike count was then computed for every disparity/correlation, giving

a mean spike count for each disparity-correlation combination. The same exact procedure was performed for both the real cells and their model counterparts, with the only difference being whether the sequence of spike counts was predicted or observed. In order to construct "tuning curves" for the independent binocular Gaussian noise, we used a procedure like that used by Tanabe et al (2011). We first computed the normalised binocular cross-correlation function of each image frame, extracting a correlation value for each disparity. For each disparity, we then identified the frames with the top and bottom 20% of correlation values, which correspond to our correlated and anticorrelated frames, respectively. We then used these frames to trigger the forward correlation procedure, computing disparity tuning curves as previously specified. With this procedure, a single frame can be used in multiple disparities if the magnitude of binocular correlation exceeded our threshold for more than one disparity.

3.3 Results

3.3.1 Example model subunits

From the fitting procedure, we obtain a GBEM unit with an optimal number of excitatory and suppressive subunits for each cell. The linear filters for the GBEM fit to cell lemM322c1 are shown in Figure 3.2a. For this cell, all subunits have a phase disparity between the two eyes (i.e. the profile of the filters differ in the left and right eyes). Figure 3.2b shows the subunit nonlinearities for the excitatory and suppressive subunits. In this case, the fitted thresholds are similar for the different subunits, except for one of the excitatory subunits (bronze line in top panel of Figure 3.2b). While this subunit's filter coefficients are much lower (excitatory subunit with bronze outline in Figure 3.2a), its threshold is much more negative, meaning that this subunit gives a large constant output to a blank screen. At the same time, the subunit's dynamic range remains substantial despite the small variation in the filter coefficients. This is because the output nonlinearity has a much higher slope when the threshold takes on a large negative value compared to when it is around zero. Figure 3.2c shows the "tuning curve" to correlated stimuli for the excitatory pool and the suppressive pool (i.e. summed over excitatory and suppressive subunits, respectively). The pooled responses are normalised such that the baseline (median) response is zero. The normalisation is necessary for visualisation purposes since the excitatory subunit in Figure 3.2b has a high constant output irrespective of the stimulus, which introduces a much higher baseline response for the excitatory pool. The tuning curves of the excitatory and suppressive pools have opposite phases: this is the familiar push-pull organisation for binocular neurons introduced by Read & Cumming (2007) and found in V1 neurons by Tanabe, Haefner, & Cumming (2011). A similar push-pull organisation was previously uncovered by Ferster (1988), who showed that orientation selective simple cells in the cat receive both excitatory OFF and suppressive ON inputs to the ON region of their receptive fields (and vice versa for the OFF region) [Ferster, 1988].
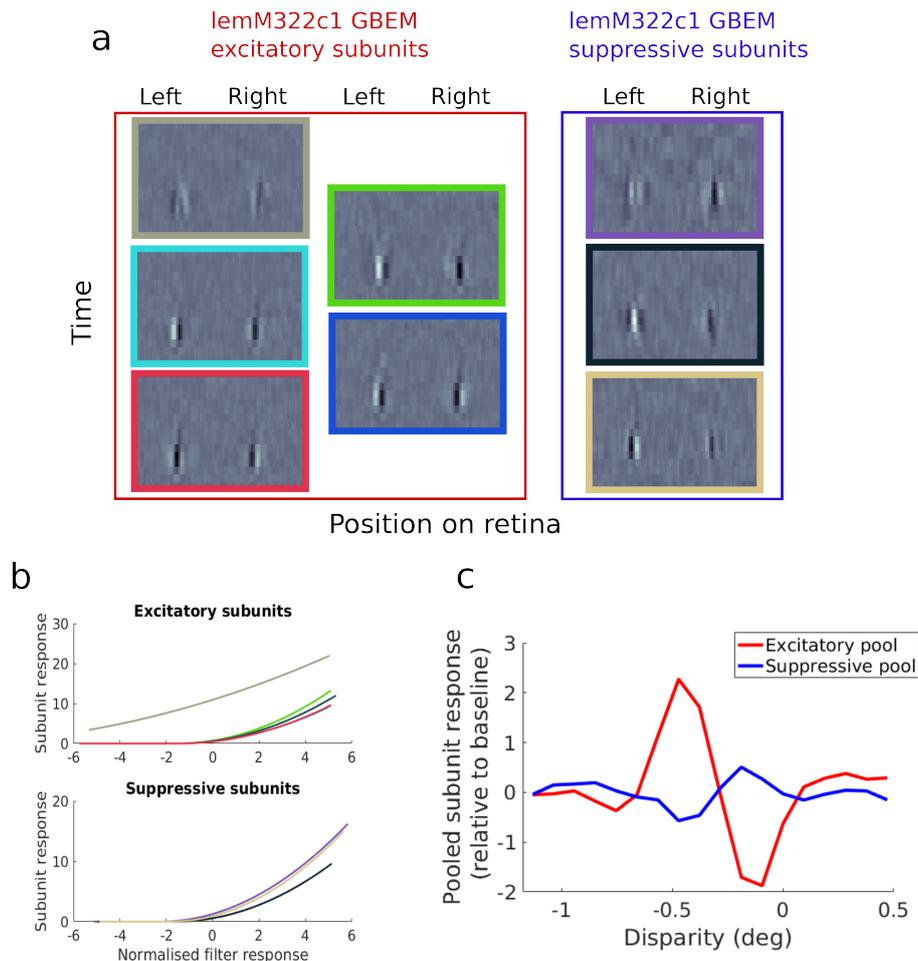
Figure 3.2: a) Filters of excitatory and suppressive subunits recovered for cell lemM322c1. The resulting GBEM unit has 5 excitatory subunits (red outline), and 3 suppressive subunits (blue outline). The filters are shown here as spatiotemporal filters, with the vertical axis denoting time and horizontal axis denoting space. Note that half of the pixels correspond to the left eye and the other half to the right eye (separated by a vertical bisection). b) Subunit responses as a function of the normalised filter response for both excitatory (top) and suppressive (bottom) subunits. The filter response will necessarily be centred on zero. In order to normalise the filter responses, we divide by the standard deviation. The normalised filter response is therefore a z-score where a value of ± 1 corresponds to one standard deviation from the mean (which as noted is necessarily zero). A blank screen corresponds to a filter response of 0 in this scheme (though the converse is not true: a filter response of 0 does not necessarily imply a blank screen). Each line shows a different subunit, with the colours mapping onto the outline of the spatiotemporal filters in a). c) Disparity tuning curves for the excitatory (red) and suppressive (blue) pools. The baseline (median) response of the pools has been normalised to zero so that a meaningful comparison can be made. The excitatory pool has a much higher baseline response than the suppressive pool.

3.3.2   Example disparity tuning curves

Once we have the GBEM fits for each cell, we can obtain a disparity tuning curve for both the cell and the model. The cells are fitted using five-fold cross-validation (see Methods for details); this enables us to get cross-validated model spike counts for all of our stimuli. In other words, the spike

counts we use for the subsequent analyses are predicted spike counts of unseen data. We compute the disparity tuning curves using a forward correlation procedure, counting the number of spikes in a 30ms time window centred around the time of the maximal disparity-related response of the cell (again, further details are found in the Methods). Figure 3.3 shows the disparity tuning curve for two cells where the GBEM model has done a very good job at capturing the disparity tuning of the cell. Figure 3.3a shows the tuning curve for cell jbeM012c7 which is known as a tuned excitatory cell by Poggio & Fischer's (1977) nomenclature. Figure 3.3b shows the corresponding GBEM fit. It is important to highlight that the model was not fit to the disparity tuning curve of the cell; rather, the optimisation routine was given the luminance values on the screen and the spike times. Thus, the ability to capture disparity tuning means that the model has learned a nonlinear binocular interaction which was not explicit in the inputs. The GBEM does a good job at capturing both the correlated and anticorrelated responses of the cell, and also captures the weak response attenuation of the cell to anticorrelated stimuli. Figure 3.3c shows the tuning curve of another example cell known as a tuned inhibitory cell [Poggio and Fischer, 1977]. These cells are relatively uncommon in cortex and respond most vigorously to a stimulus which is binocularly anticorrelated, i.e. to stimuli which are impossible in naturalistic viewing. A success of the original BEM is that it can capture this type of disparity tuning by incorporating phase disparity between the left and right subunits. Indeed, the GBEM readily accounts for the disparity tuning of this cell (Figure 3.3d), and does so by recovering linear filters which have phase disparities of approximately π.

Most GBEM fits capture the overall shape of disparity tuning, and for some cells the magnitude of disparity tuning is also well-captured (e.g. Figure 3.3). However, most GBEM units underestimate the magnitude of disparity tuning. This becomes particularly evident when the GBEM and cell responses are superimposed. Figure 3.4a shows an example odd-symmetric cell where the cell's response is very well-captured. As previously noted, the response here is cross-validated, meaning that the model responses are shown for stimuli with which the model has not been fit. Thus, Figure 3.4a represents a substantial success of the GBEM architecture to account for average responses across disparities. In order to assess the ability of the GBEM to account for responses within disparities, we plot the mean rate in the real cell as a function of the predicted rate of the model, binned according to the predicted rate within each disparity/correlation (see Methods). If the model correctly predicts the firing rate of the cell within a disparity, then these points should be monotonically increasing within each disparity. The solid dots show the responses to the preferred disparity of the cell – shown by the dashed blue line in Figure 3.4a – for correlated (red) and anticorrelated (black) stimuli. The crosses show the anti-preferred disparity - shown by the dashed green line in Figure 3.4a. If the model captured the responses to stimuli within a disparity, then the predicted and observed rates should be appropriately rank ordered (i.e. higher predicted rate should correspond to a higher observed rate). Indeed, that is generally the case for
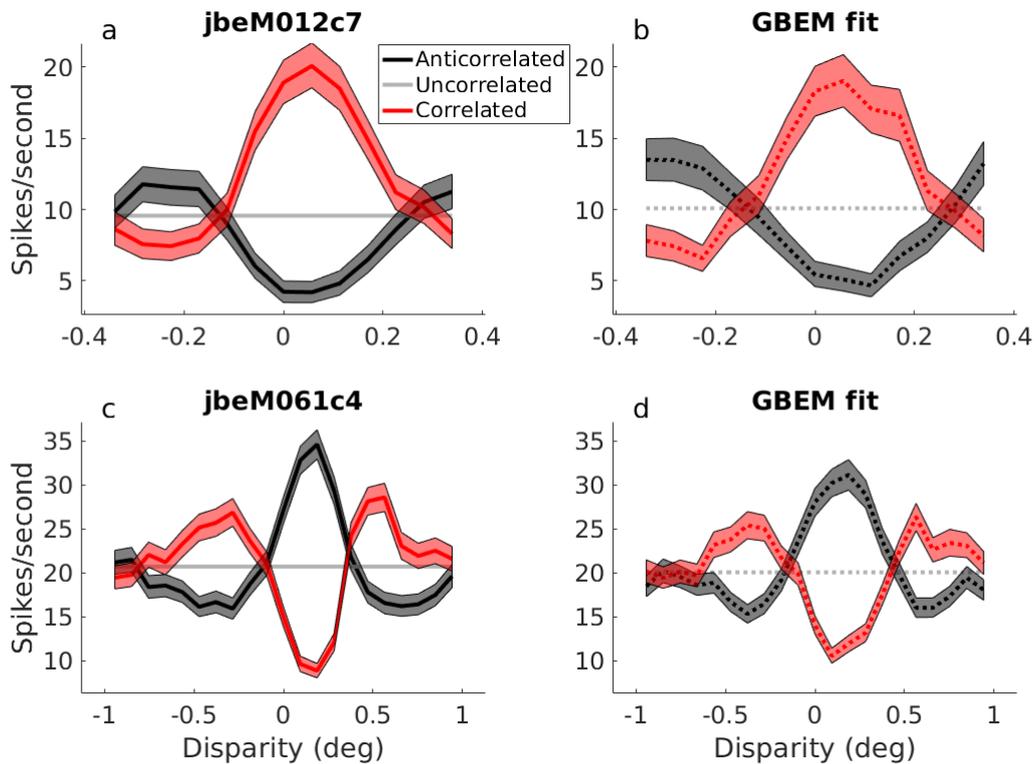
Figure 3.3: Disparity tuning curves for two example cells (a,c) and their corresponding GBEM fit (b,d). Correlated responses are shown in red, anticorrelated responses are shown in black, and uncorrelated responses are shown in gray. The number of spikes per second were calculated in a 30ms window centred around the peak temporal response of the cell. The shaded regions show 95% bootstrap confidence intervals for the responses.

both the preferred and anti-preferred disparity. However, not all cells produce perfect fits. Figure 3.4c shows a cell which is moderately successful: it captures the overall shape of the tuning curve, but underpredicts the magnitude of disparity tuning. In particular, it underpredicts the response of the cell to correlated stimuli. The effect of this can be seen in Figure 3.4d, where the responses to the anti-preferred disparity (red and black squares), as well as the anticorrelated responses to the preferred disparity (black circles) all lie on roughly the same curve. The correlated responses are notably shifted up and to the left. This means that with the recovered model structure, there is no single output nonlinearity that can simultaneously account for the responses to correlated and anticorrelated stimuli. The implication is that the GBEM structure itself is not appropriate for simultaneously describing the response to correlated and anticorrelated stimuli. It also suggests that the correlated responses at the preferred disparity are in some way "special" in the way they are processed by this V1 neuron.

For other cells, we observe a much more spectacular failure to capture the disparity tuning. Figure 3.4e shows a cell where the model exhibits very little disparity selectivity. Despite this fail-
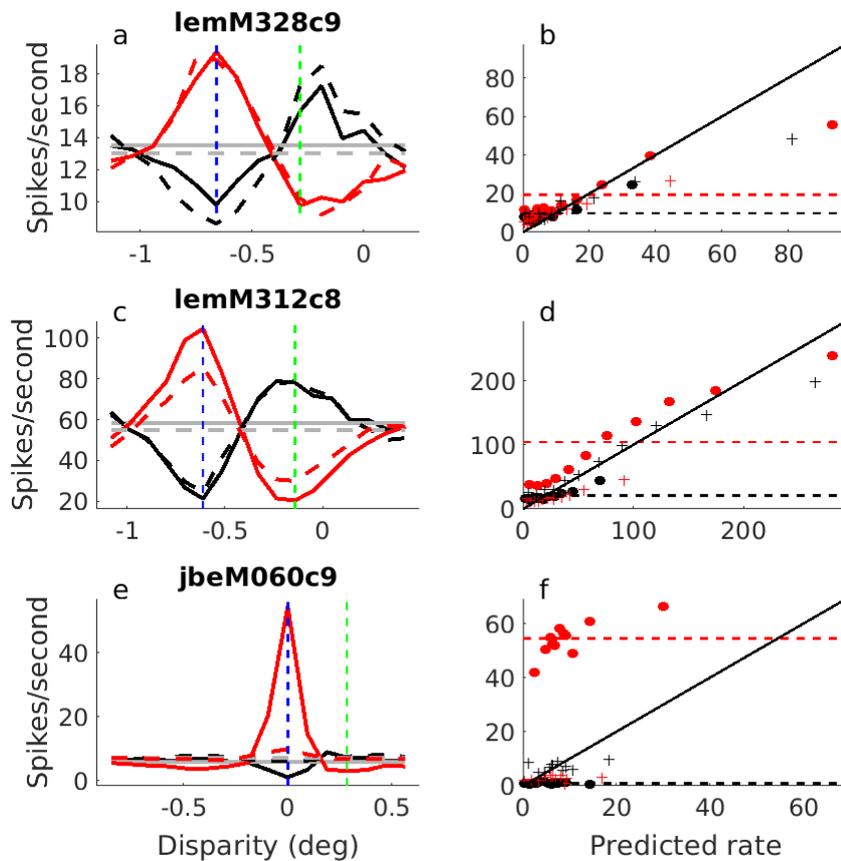
Figure 3.4: Disparity tuning curves for three example cells which capture the diversity of model fits. Here, the response of the model (dashed lines) is overlaid on the response of the cell (solid lines). Panels a, c, and e show the disparity tuning curves, whereas panels b, d, and f show the observed spike rated plotted as a function of the predicted spike rate (binned according to the predicted spike rate). Each point in b, d, and f shows a separate disparity/correlation combination (black dots are anticorrelated, red dots are correlated). Filled circles show the preferred disparity of the cell (dashed blue line in a, c, and e), and pluses show the anti-preferred disparity (dashed green line in a, c, and e).

ure to capture disparity tuning, Figure 3.4f shows that the rank ordering of the model responses within disparities are nevertheless broadly preserved. In other words, the model can account for a large proportion of the cell's responses to variations in the stimulus, but simply fails to account for disparity-selectivity. This means that for a given disparity, a large response in the cell to a particular luminance pattern does generally correspond to a larger response in the model as well. However, the GBEM selectively fails to capture the binocular interaction between the left and right eyes. Indeed, the model fits can be broadly grouped into the three categories highlighted in Figure 3.4. Of particular interest are the model units that fail to capture any disparity tuning (e.g. Figure 3.4e); this does not represent a general fitting failure, as evidenced by Figure 3.4f, but

is, as noted, a selective failure to capture disparity tuning. If a GBEM unit failed to capture less than 30% of the variance in disparity tuning of the corresponding cell, we categorised this model fit as failing to capture disparity tuning. Out of the 95 disparity-selective cells in our dataset, 28 fits failed to adequately capture disparity tuning (30% of disparity-selective cells).

Cells such as that in Figure 3.4e illustrate an important point: although it is trivial to hand-tune a GBEM unit which is able to better capture the disparity tuning curve of these cells, the GBEM is not fit to the tuning curves themselves. Instead, the input to the model is simply the luminance patterns on the screen, and so the GBEM attempts to capture the full range of monocular, binocular, and temporal dynamics of the cell. Thus, while a hand-tuned model might be able to capture the tuning curve, it would do much worse than the fitted GBEM in predicting the response of the cell to actual (unseen) random line stereograms.

A key feature in Figure 3.4 is the underestimation by the GBEM of the magnitude of disparity tuning in real cells (Figure 3.4c). Indeed, the largely absent disparity tuning of the GBEM unit for jbeM060c9 (Figure 3.4e) can be seen as a more extreme failure of the sort seen in lemM312c8 (Figure 3.4c). The effect of underestimating the response to correlated stimuli (at the preferred disparity) is that real cells have stronger anticorrelated attentuation than the GBEM units. This is because the anticorrelated attenuation depends on the magnitude of the correlated response, which is boosted in real cells. In order to quantify this, we compute the regression slope between the correlated and anticorrelated response. We will refer to this metric as the relative anticorrelated response of the cell (note that this is different from the metric used in Cumming & Parker, 1997, but the same as that used by Henriksen et al., 2016b). Figure 3.5b shows this graphically. If the relative anticorrelated response is -1, then the neuron responds as strongly to anticorrelated stimuli as to correlated, but with a sign inversion (as in the standard binocular energy model) and thus exhibits no response attenuation to anticorrelated stimuli. A relative anticorrelated response of 0 corresponds to a neuron which either does not modulate its response to anticorrelated stimuli or whose anticorrelated response is modulated orthogonally to the correlated response. In practice, the latter is rare and we observed no such cases in the current study. We therefore take this metric to be an index of relative anticorrelated response. The relative anticorrelated response for jbeM012c7 (Figure 3.3a) is -0.69 (95% CI: [-0.76, -0.62]), and the relative anticorrelated response for its GBEM fit (Figure 3.3b) is -0.74 (95% CI: [-0.85, -0.63]). Thus, the GBEM readily captures much of the anticorrelated response of this cell. For jbeM061c4 (Figure 3.3c), the GBEM also captures the relative anticorrelated response of the neuron well (Figure 3.3d), and this is again reflected in the relative anticorrelated response (jbeM061c4: -1.01, 95% CI: [-1.09,-0.923]; GBEM: -1.11, 95% CI: [-1.21, -1.00]). For the example cells in Figure 3.4, the relative anticorrelated responses are -0.62 [-0.80,-0.45] for lemM328c9 and -0.95 [-1.04,-0.87] for its GBEM fit; -0.67 [-0.74,-0.60] for lemM312c8 and -0.95 [-1.03,-0.87] for its GBEM fit; and -0.12 [-0.13,-0.11] for jbeM060c9 and -0.59

[-0.67,-0.51] for its GBEM fit.
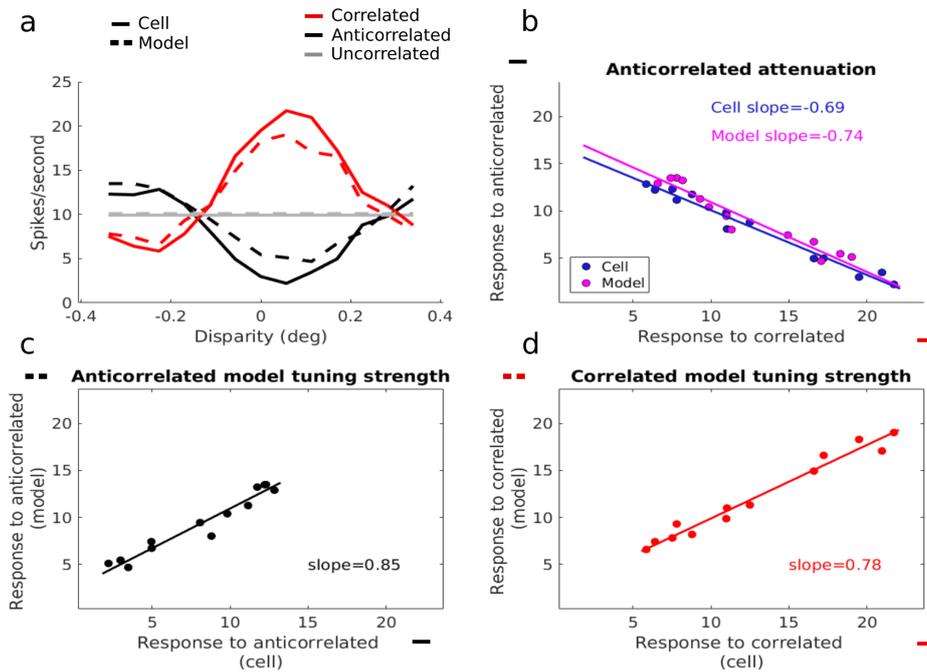
Figure 3.5: a) A disparity tuning curve for jbeM012c7 (solid lines), and its model fit (dashed line). b) Anticorrelated response as a function of the correlated response for a cell (blue dots) and its corresponding GBEM fit (magenta dots). The relative anticorrelated response is defined as the slope of the regression line, shown for both the model and cell. A relative anticorrelated response of 0 means that the cell does not modulate its response to anticorrelated stimuli, whereas an relative anticorrelated response of -1 means that the cell inverts its response perfectly just like in the binocular energy model. c) Anticorrelated response of the model as a function of the anticorrelated response of the cell. The regression slope defines the anticorrelated model tuning strength, which quantifies how strongly the model unit is tuned to anticorrelated stimuli relative to the cell. If the anticorrelated model tuning strength is 0.5, then the model unit modulates its response half as strongly to disparity for anticorrelated stimuli compared to that of the cell. d) Correlated response of the model as a function of the correlated response of the cell. The regression slope defines the correlated model tuning strength, which quantifies how well the model's shape and magnitude of disparity tuning matches that of the cell.

3.3.3 Population summary

In order to summarise the inability of the model to capture the cell's tuning curve across the population, we can plot the relative anticorrelated response for the neurons against the relative anticorrelated response for their corresponding GBEM fits (Figure 3.6a). Two key points are worth observing. First, there is a strong positive correlation between the relative anticorrelated response in the cells and that seen in the model units ($r = 0.61, p < 0.001$, Pearson correlation). This is notable since Tanabe & Cumming (2011) found no relative anticorrelated response in the model units recovered with their spike-triggered analysis of covariance. Thus, although it has long been recognised that LN cascades can in principle explain relative anticorrelated response, this is the first direct evidence that this explanation is at least partially correct. However, the vast

majority of points in Figure 3.6a lie beneath the identity line, meaning that the GBEM systematically predicts a more negative (i.e. stronger) relative anticorrelated response than is observed in the cells. This suggests that although the LN cascade goes some way towards accounting for the relative anticorrelated response seen in V1 neurons, it is not on its own a sufficient explanation.

The GBEM fails because it underpredicts the magnitude of disparity tuning. However, there are three potential reasons for this failure. The first possibility is that the GBEM fails because it is unable to account for the anticorrelated response. This could be the case, for example, if these cells have developed a specialised mechanism for suppressing false matches [Goncalves and Welchman, 2017]. The second possibility is that the GBEM is unable to capture the correlated response. Similar to the anticorrelated case, a failure to capture the correlated response could be because V1 neurons have enhanced responses to true matches which cannot be easily captured by the GBEM framework. The third possibility is that there is failure to capture both the anticorrelated and the correlated responses. To investigate this, we first define a model tuning strength metric. The model tuning strength is simply the regression slope between the cell and the model's tuning curves (illustrated in Figure 3.5c and d for the anticorrelated and correlated case, respectively). If the model tuning strength is 0.5, then the model's disparity tuning is 50% of the real cell's tuning (provided that the shape of the disparity tuning curve has been appropriately captured, which is generally the case in our data). We can compute the model tuning strength separately for correlated and anticorrelated responses, yielding a metric for how well a GBEM unit is able to capture the shape and magnitude of the cell's disparity tuning for correlated and anticorrelated stimuli. Figure 3.6b shows the anticorrelated model tuning strength on the vertical axis, and the correlated model tuning strength on the horizontal axis. The vast majority of cells in this plot lie above the diagonal, suggesting that the GBEM is better able to capture the anticorrelated responses than the correlated ones. This was confirmed by a paired samples t-test comparing the anticorrelated model tuning strength ($M = 0.52$) against the correlated model tuning strength ($M = 0.32$): $t(94) = 5.90$, $P < 10^{-7}$. The fact that points generally lie below 1 for both correlated and anticorrelated stimuli means that the models underpredict the strength of disparity tuning for both correlated and anticorrelated stimuli. However, the magnitude of the model failure is notably greater for correlated stimuli compared to anticorrelated stimuli.

The previous analysis confirms that the correlated responses of the cells are more problematic for the GBEM than anticorrelated responses. This is noteworthy since it suggests that the established way of thinking about the shortcomings of the BEM – that it is not able to capture the anticorrelated response – is incorrect. Instead, real neurons tend to have a larger response to candidate true matches at one specific preferred disparity, which neither the BEM nor its generalisation (the GBEM) is able to capture.

While the previous analysis shows that the GBEM units are failing to capture the correlated
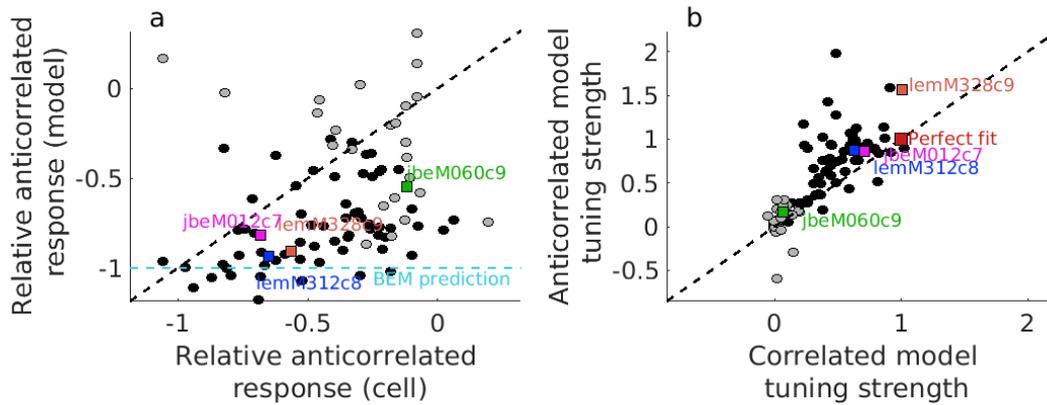
Figure 3.6: The relative anticorrelated response for the model against the relative anticorrelated response for the cell. The dashed black line shows the identity line, while the dashed cyan line shows the prediction of the BEM (perfect inverted tuning curve to anticorrelated). b) The model tuning strength for the anticorrelated tuning curve, against the same for the correlated tuning curve. The dashed black line again shows the identity line, and the red square marker shows the point of perfect model fit. Cells which produce poor GBEM fits ($< 30\%$ accounted for variance in the disparity tuning curve) are shown in gray in both plots.

responses, it is unclear from this whether the anticorrelated responses also contribute significantly to the model failure. The results in Figure 3.6 are consistent with the idea that both the anticorrelated and correlated responses systematically contribute to the model failure, but to different extents. In order to explore this question, we investigated whether the failure of the GBEM to capture the relative anticorrelated response in real cells was related to the model tuning strength for correlated and anticorrelated responses. Figure 3.7a shows the model tuning strength for correlated responses against the relative anticorrelated response in the cell. There is a strong negative correlation between the two ($r = -0.58$, $p < 10^{-9}$, Pearson's r). That is, the disparity tuning curves for correlated stimuli are systematically captured less well for cells which show stronger attenuation to anti-correlated stimuli. Figure 3.7b shows the equivalent plot for the model tuning strength to anticorrelated responses. This relationship is weaker, but significant ($r = -0.34$, $p = 0.001$). Interestingly, the effect in Figure 3.7b is largely due to cells such as jbeM060c9 which fail to account for the disparity tuning curve of the real cells. These are shown highlighted in gray and largely cluster in the bottom right of the plot. Excluding cells with bad fits (e.g. those that can account for less than 30% of the variance in the cell's disparity tuning curve, gray circles in Figure 3.7b), makes the relationship non-significant ($r = -0.19$, $p = 0.13$). The results in Figure 3.7a remain highly significant ($p < 10^{-7}$) when excluding poor fits. Taking the results of Figure 3.6 and 3.7 together suggests that "attenuated anticorrelated response" is the wrong way to think about the response properties of disparity-selective cells. Instead of "attenuated anticorrelated response", real neurons show "correlated boosting": that is, they have an amplification of their response to candidate true matches (correlated stimuli at the preferred disparity of cell) rather than an attenuation or suppression of false matches per se. This amplification is largely responsible for the

failure of the GBEM to capture the disparity tuning curve of real cells.
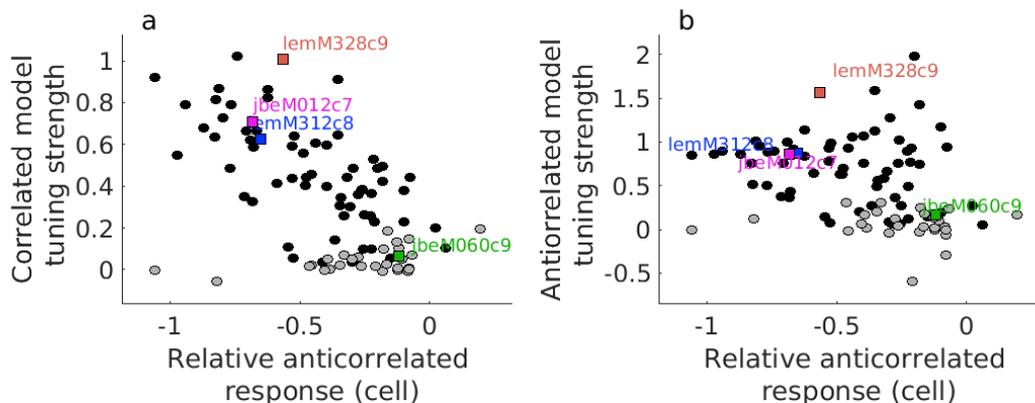
Figure 3.7: Correlated (a) and anticorrelated (b) tuning curve slope plotted as a function of the cell's relative anticorrelated response. Tuning curve slope quantifies how well the magnitude of disparity tuning in the cell is captured by its GBEM fit, and this can be computed separately for the correlated and anticorrelated tuning curves. While the model's correlated tuning strength is significantly related to the cell's anticorrelated slope (a), the model's anticorrelated tuning strength (b) is not. In other words, a cell's magnitude of anticorrelated attenuation (i.e. its relative anticorrelated response) is predictive of whether the correlated response will be well-described by a GBEM unit. Cells which produce poor GBEM units (< 30% variance accounted for in the disparity tuning) are shown in gray.

3.3.4   Binocular Gaussian noise

The most comprehensive attempt at directly modelling disparity-selective cells in V1 by learning model parameters from data was performed by Tanabe & Cumming (2011). In order to retrieve the linear filters for each subunit, the authors first performed a spike-triggered analysis of covariance, and then optimised the nonlinearities for each subunit. In principle, this approach should give very similar results to the analysis we show here. It is noteworthy then that Tanabe & Cumming (2011) found no relative anticorrelated response in their model cells when tested on 1D line patterns, where the real neurons exhibited strong relative anticorrelated response. The key differences between that paper and the current work are 1) the method of model estimation, 2) different stimuli due to the requirements of their spike-triggered covariance method, and 3) the duration of each individual noise pattern (10ms in Tanabe & Cumming, 2011; 30ms in the present work). Specifically, the spike-triggered analysis of covariance approach used by Tanabe & Cumming requires a white noise stimulus which is independent in the two eyes. The consequence of using independent noise in the two eyes is that there are very few frames with extreme binocular correlation values (e.g. close to -1 or 1). Interestingly, the authors showed that when the cells were tested on the same independent noise patterns with which the models were fit, the cells did not exhibit systematic attenuation to anticorrelation either. It is then perhaps not surprising that the model units did not reproduce the anticorrelated attenuation in the cell, since these nonlinearities were not evident

in the data on which the models were fit [Tanabe et al., 2011b].

An independent test of our GBEM units is then to explore whether they can reproduce this characteristic feature of real cells highlighted by Tanabe & Cumming (2011). The stimulus used to fit the GBEM is very similar to the RLS stimuli used by Tanabe & Cumming to construct their attenuated tuning curves, so we know that the model units express relative anticorrelated response (or more appropriately, correlated boosting) in this case. Thus, the critical test is whether our model units also express an absence of that attenuation to independent binocular Gaussian noise. The procedure for calculating disparity tuning curves from independent noise data was the same as in Tanabe & Cumming (2011) and is documented in the Methods section. Figure 3.8a shows a tuning curve computed using Tanabe & Cumming's independent noise stimulus for the example GBEM unit shown in Figure 3.3a (jbeM012c7). While the model unit previously showed clear relative anticorrelated response (M=-0.74, 95% CI: [-0.85, -0.63]), this is now not significantly different from -1 (M=-0.95, 95% CI: [-1.08,-0.82]. We can summarise this across the population by plotting the relative anticorrelated response to the 1D RLS stimulus against the relative anticorrelated response to the independent Gaussian noise stimulus (Figure 3.8). Just as in Tanabe & Cumming (2011), our model units cluster around -1 (no attenuation), and at the population level is not significantly different from -1 ($t(94) = 1.53$, $p = 0.13$). Thus, our model units readily capture both the attenuated anticorrelated response/correlated boosting exhibited by real cells to 1D RLS stimuli, and also the absence of that attenuation in independent Gaussian noise.
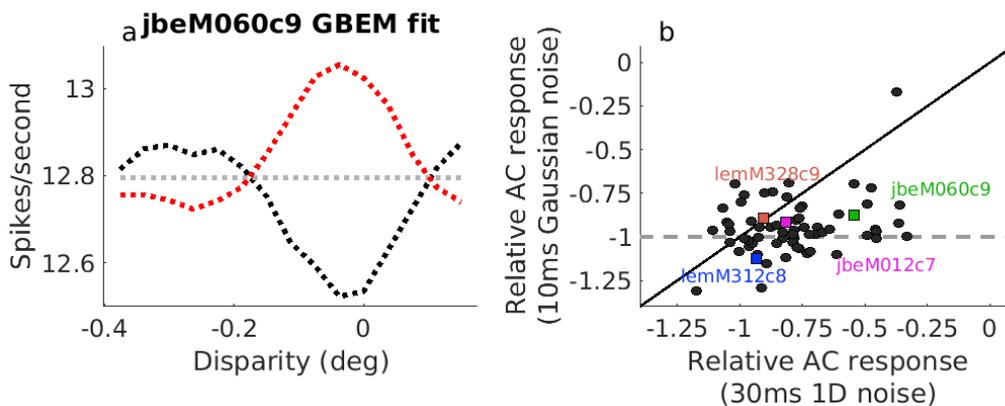
Figure 3.8: The predicted disparity tuning curve of cell jbeM012c7 in response to binocularly uncorrelated Gaussian noise (using forward correlation as in Tanabe et al), based on the GBEM fit. The tuning curve of the cell and its GBEM fit to 30ms 1D noise stereograms are shown in Figure 3.4c and d, respectively, where strong attenuation is present. Note that this attenuation is not present in model responses to binocularly uncorrelated Gaussian noise. b) The population summary for all GBEM units. The relative anticorrelated response to 1D noise is shown on the horizontal axis, and the relative anticorrelated response to Gaussian noise is shown on the vertical axis.

3.4   Discussion

The binocular energy model has been enormously successful in capturing a range of properties of disparity-selective cells. One prominent success is its ability to account for the inverted tuning curves for anticorrelated stimuli [Cumming and Parker, 1997]. However, a complete account of the relatively weaker response to anticorrelated stimuli reported by Cumming & Parker (1997), and thus a complete model of disparity encoding by V1 neurons, has remained elusive. In response to this, a range of modified BEM units have been proposed [Nieder and Wagner, 2000, Read et al., 2002, Tanabe and Cumming, 2008, Tanabe et al., 2011a, Tanabe et al., 2011b, Samonds et al., 2013, Henriksen et al., 2016a, Goncalves and Welchman, 2017], but it is unclear which, if any, of these model frameworks can account for neuronal activity in V1.

In this paper, we made use of developments in optimisation routines which allowed us to fit the components of generalised binocular energy model units to spiking data from V1. The data was generally well-described by the model units, confirming that this class of models can serve as a model of first approximation. However, even the best-fitting models underestimated the magnitude of the correlated response at the preferred disparity of the cell. It is important to note that this is not simply a failure of the fitting procedure, such as a failure to find the global maximum likelihood. Although it is possible to create disparity tuning curves by hand that better capture the cells' tuning curves, the GBEM was not optimised for fitting the actual tuning curves themselves. Rather, the GBEM finds a model which can best map image sequences to firing rates, and in doing so captures temporal and monocular dynamics, as well as binocular interactions.

We observed that the tendency for the GBEM to underestimate the correlated response was related to how much response attenuation the cell exhibited: cells that show relatively weaker modulation to anticorrelated stimuli were fit more poorly by our model framework. However, this was specific to the correlated responses. In other words, when cells had very large correlated responses relative to anticorrelated responses, the model failed to capture the correlated response to a larger degree than the anticorrelated response. This strongly suggests that disparity-selective V1 cells have a specialised mechanism for strengthening the response of candidate true matches, which likely does not originate from a standard linear-nonlinear cascade. We call this mechanism correlated boosting.

Previous work has expanded on the BEM to account for the relatively weaker anticorrelated responses in three separate ways. The first and simplest is by simply appending an output nonlinearity to the BEM. This is by far the most common method for generating anticorrelated attenuation (e.g. [Nieder and Wagner, 2000, Read et al., 2002, Henriksen et al., 2016a]), but there has been little evidence showing that this is the actual explanation. Read et al. (2002) pointed

out that this explanation does not work for odd-symmetric cells since the correlated and anticorrelated responses are equally affected by the output nonlinearity in this case (thus there is no way to selectively attenuate the anticorrelated responses, or more appropriately, to selectively boost the correlated responses). Notably, a simple output exponent can be easily approximated by the thresholded square nonlinearity on the subunits in the GBEM (Figure 3.9). Thus, if an output nonlinearity was the correct explanation, our modelling approach would have revealed it.
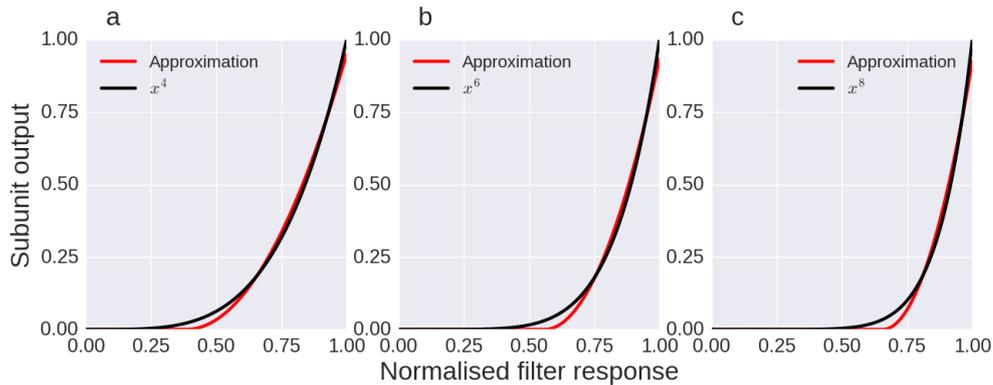

Figure 3.9: Illustration of how a thresholded square can approximate an arbitrary output exponent. The black line shows $g(x) = x^\gamma$ for $\gamma = 4$, 6, and 8 in plots a, b, and c, respectively. The red line shows the nonlinearity $f(x) = \alpha \mathrm{Pos}(x - \theta)^2$, where values of α and θ were fit to minimise the sum squared error for each plot. α is a scalar which simulates the effect of the filter coefficients (though it is outside the nonlinearity in this case for simplicity), and θ corresponds to the threshold. In each case, the thresholded square is readily able to capture the steepness of the high exponent nonlinearity. Not shown is $\gamma = 2$ since this is trivially captured by the rectified square by setting $\alpha = 1$ and $\theta = 0$.

The second approach for generating anticorrelated attenuation is to use multiple binocular subunits with different disparity tuning. This method was introduced by Haefner & Cumming (2008). To produce odd-symmetric model cells that have attenuated responses to anticorrelated stimuli, this model uses two subunits which are both even-symmetric with an output nonlinearity, tuned to different disparities. If two such subunits feed into a complex cell, one exciting the cell and one suppressing it, the complex cell will have odd-symmetric disparity tuning with anticorrelated attenuation. Again, if this was the correct explanation, then the GBEM fitting procedure should have come up with such model cells. Interestingly, for cells with odd-symmetric tuning, the resulting GBEM consistently had binocular subunits which themselves had odd-symmetric tuning. If the two-subunit (2SU) model by Haefner & Cumming (2008) was a correct description of odd-symmetric disparity tuning, then the model should have recovered subunits with offset, even-symmetric disparity tuning. This is therefore the first evidence that odd-symmetric disparity tuning of V1 neurons does not generally arise through this type of 2SU-type architecture.

The third approach was pioneered by Samonds et al. (2013), where they used a recurrent net-

work made up binocular energy model-like units. The authors found that anticorrelated response attenuation can be accounted for by such models. Recurrent models are interesting because they can easily model a range of complex phenomena known to operate in cortex. For example, feed-forward models such as the GBEM are not ideal for modelling complex temporal dynamics. In contrast, Samonds et al. (2013) used their recurrent binocular network to successfully model the sharpening of disparity tuning over time [Samonds et al., 2009, Samonds et al., 2013]. Recurrent models are particularly promising given the recent success of recurrent neural networks in matching or exceeding human performance on a range of complex visual tasks, such as image captioning [Vinyals et al., 2014] and object recognition [Liang and Hu, 2015]. Cells with recurrent connections remain a strong potential candidate for accounting for disparity tuning in real cells. However, this explanation has not yet been tested in real V1 neurons.

A model framework which has been much less explored is incorporating nonlinearities on the monocular filter responses. Read et al. (2002) successfully described the tuning curves of a number of V1 neurons by creating models which incorporated a rectified squaring prior to binocular combination. While the GBEM can readily generate such monocular subunits, it is greatly limited in its ability to easily capture complex binocular interaction between the subunits. This is because monocular subunits in the GBEM can only interact through the final (softplus) spiking nonlinearity, which is much less flexible than the thresholded square used on the individual subunits. Interestingly, we observed several cells which had very dominant monocular responses (i.e. the values of the filter coefficients of one eye were much greater than in the other eye), and whose disparity tuning was poorly accounted for by the GBEM. One possibility is that another architecture, such as that proposed by Read et al. (2002) is better able to account for the responses of such cells.

A fifth possibility is that incorporating models with pairwise nonlinearities might be able to capture a range of complex dynamics. For example, a cell whose final spiking nonlinearity is a thresholded AND gate, i.e.

$$F(g_1, g_2) = \begin{cases} g_1 + g_2 & \text{if } g_1 > \theta_1 \text{ and } g_2 > \theta_2, \\ 0 & \text{otherwise,} \end{cases} \quad (3.5)$$

could potentially display the correlated boosting effect given appropriate subunits. Similar nonlinearities have been explored in relation to contrast normalisation [Simoncelli and Heeger, 1998, Carandini and Heeger, 2012], but little attention has been given to these types of computations in the disparity domain.

A final possibility is that adding layers to the GBEM framework will allow the model to learn more complex models which still generalise well to unseen data. Indeed, one of the remarkable

features of deep convolutional neural nets is their ability to learn complex, nonlinear features from unstructured data [LeCun et al., 2015]. It is well-known that given enough subunits, even models such as the GBEM can in principle approximate any arbitrary function [Hornik, 1991]. However, if the model is not a good approximation of the underlying generative process, then the learned model parameters will fail to generalise to new, unseen data. For a deep generalised binocular energy model framework the question therefore becomes whether such a model can 1) provide a compact account of disparity-selectivity in cortex, and 2) help illuminate the mechanisms which gives rise to interesting properties of real cells. Deeper binocular models (i.e. with 2 or more "hidden layers") are a promising avenue for future research.

Chapter 4

Response variability in disparity-selective cells

4.1 Introduction

Chapters 1 and 2 [Henriksen et al., 2016a, Henriksen et al., 2016c] examined how response variability in the binocular energy model can get converted into a mean rate change through an output nonlinearity. The choice of output exponent was a squaring because squaring gives a clear algebraic dependence on variance (i.e. $\mathrm{E}[X^2] = (\mathrm{E}[X])^2 + \mathrm{Var}[X]$). Interestingly, Chapter 3 showed how an output exponent cannot be the right explanation for the characteristic boosting of correlated responses at the preferred disparity of the cell, since the subunits can easily approximate any arbitrary output exponent (see Figure 3.9). Thus, if this was the explanation, the GBEM would have recovered such subunits.

Response variability is an aspect of neuronal computation in disparity-selective cells that has been largely overlooked. The response variability of the cell can be separated into two components: internal (or intrinsic) variance and external (or stimulus-driven) variance. The internal variance is simply the variability in firing which cannot be accounted for by changes in the stimulus. Neurons appear to exhibit strong stochastic properties, and so will not generally repeat the same identical set of spikes even though a sequence of stimuli is repeated exactly. Commonly, the spiking process is modelled as a Poisson process, which has a Fano Factor (variance divided by mean) of 1. The external variance of the cell is the variance in firing which is associated with changes in the stimulus. Equivalently, it is that variance which is *not* stochastic variability. The total variance of the cell can be written as a sum of the two

$$\mathrm{Var}_{\mathrm{total}}[R] = \mathrm{Var}_{\mathrm{int}}[R] + \mathrm{Var}_{\mathrm{ext}}[R]. \tag{4.1}$$

Here, $\mathrm{Var}_{\mathrm{total}}[R]$, $\mathrm{Var}_{\mathrm{int}}[R]$, and $\mathrm{Var}_{\mathrm{ext}}[R]$ refer to the total, internal, and external variance, respectively. External or stimulus-related response variability is a particular challenge for a system trying to decode information about the disparity of an image, independent of the scene content. In the case of the random-dot stereograms used throughout this thesis, the observer has to correctly extract the disparity of the stereogram while ignoring the particular pattern of dots. Even if a V1

cell is highly disparity-tuned, it may carry relatively little information about disparity, independent of dot pattern, if its external variance is also high In Chapter 1 and 2, it was the external variance of the binocular energy model that carried the disparity signal which ultimately allowed a BEM unit with an output nonlinearity to signal disparity in its mean rate.

External variance has received very little attention in the disparity domain. The notable exception is the work by Burge & Geisler (2014) who created so-called "log likelihood" model neurons. These model units are a special case of the GBEM, and modulate their firing rate in proportion to the log likelihood of the disparity given the stimulus. Consequently, these model units have very little external variance and can encode disparity in a statistically optimal manner [Burge and Geisler, 2014]. However, perfect disparity encoding comes at a cost: neurons that encode disparity perfectly will have to ignore variability in the stimulus other than disparity. Real neurons are unlikely to be log likelihood neurons since disparity-selective cells also modulate their firing in response to other features of visual stimuli, such as contrast, orientation, spatial frequency content, and so on. Despite the fact that understanding external variance has clear implications for understanding disparity encoding and the stereo correspondence problem, so far, no one has explored the character of external variance in disparity-selective cells. In this chapter, we first derive theoretical properties of the BEM's external variance, and show that in the BEM, the external variance is proportional to the square of the mean spike count. We then show data from a procedure which is able to decompose the total response variance of real cells into external and internal variance (as in Equation 4.1). These results show that the predictions of the BEM are in fact qualitatively met: real cells do modulate their external variance with mean spike count. However, in real cells, this relationship is linear, not quadratic. Lastly, we perform simulations using the model GBEM units fitted in Chapter 3 and show that although the added flexibility of the GBEM allows the model units to better capture the linear dependence of the external variance, the GBEM units underpredict the overall external variance in the real cells.

4.2 Methods

4.2.1 GBEM units and cells

The data used in the present analysis was a subset of the data reported in Chapter 3. Out of the 95 disparity-selective cells in Chapter 3, 75 cells were recorded using the two-pass procedure described below and were included in the analysis. The GBEM units were the same as those recovered in Chapter 3 and all model responses were cross-validated as in Chapter 3.

4.2.2 External variance and the two-pass procedure

In order to compute the external variance, we record the responses of neurons in response to two identical trials. We define the internal variability for a given disparity and correlation as

$$\mathrm{Var}_{\mathrm{int}}(R) = \frac{1}{2N} \sum_i^N (R_i^{(1)} - R_i^{(2)})^2. \tag{4.2}$$

Here, $R_i^{(1)}$ and $R_i^{(2)}$ refer to the spike count on the first and second pass of the ith stimulus sequence. N is the total number of unique stimulus sequences (and $2N$ is therefore the total number of trials). Thus, $R_i^{(1)}$ and $R_i^{(2)}$ denote the responses to two *identical* stimulus sequences. If there was no internal variance in the cell, then $R_i^{(1)} = R_i^{(2)} \ \forall i$, and so the internal variance (Equation 4.2) would be zero. In such a case, all the response variance of the cell would be external variance. When neuronal spiking is not entirely determined by the stimulus sequence (as is generally the case in mammalian cells), we can obtain the external variance by subtracting out the internal variance from the total variance.

$$\mathrm{Var}_{\mathrm{ext}}(R) = \mathrm{Var}_{\mathrm{total}}(R) - \mathrm{Var}_{\mathrm{int}}(R). \tag{4.3}$$

For completeness, the total variance is

$$\mathrm{Var}_{\mathrm{total}}(R) = \frac{1}{2N-1} \sum_i^N \left[(R_i^{(1)} - \bar{R})^2 + (R_i^{(2)} - \bar{R})^2 \right] \tag{4.4}$$

. $\bar{R}$ refers to the mean rate for a particular disparity and correlation.

As in Chapter 3, we computed the spike count associated with a particular frame by computing the number of spikes in a 30ms window centred around t_{max}. t_{max} is the time at which the where the variance across disparities is maximal.

4.3 Results

4.3.1 The binocular energy model

The response of the binocular energy model to any given stimulus can be compactly summarised as

$$C = (L_1 + R_1)^2 + (L_2 + R_2)^2, \tag{4.5}$$

where L and R refer to the left and right eye responses and the subscripts refer to the two binocular subunits of the BEM. The two subunits are constructed such that the receptive fields which give rise to the responses L_1 and L_2 are 90° out of phase (i.e. they form a quadrature pair), giving statistically independent responses (and equivalently for R_1 and R_2). Unless we introduce either

a stochastic process (e.g. Poisson spiking) or a hidden variable (e.g. an internal gain parameter [Goris et al., 2013]), all variance in the model response will be external variance.

In general, for spatiotemporal receptive fields, and in non-natural stimuli (e.g. random dot stereograms), the distribution of the monocular filter responses (L_1, R_1, L_2, R_2) will be Gaussian. The mean response of the filter depends on the characteristics of the filter and the stimulus, but for stimuli which have on average an equal number of black and white pixels, the mean is 0 by convention. In studies which look at natural images, luminance and contrast normalisation is generally performed prior to feeding the images into the model [Burge and Geisler, 2014]. The sum of two normal random variables - L and R - with zero mean is simply another normally distributed variable with zero mean and variance depending on the covariance between L and R. We can write $B = L + R$, and rewrite Eq 4.5 as

$$C = B_1^2 + B_2^2. \tag{4.6}$$

We can now see that the distribution of C is simply the sum of two independent squared random variables. If the binocular responses B_1 and B_2 both had unit variance, this would simply be a χ^2-distribution with 2 degrees of freedom:

$$C \sim \chi^2(k=2) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} \exp(-x/2) = \frac{1}{2}\exp(-x/2). \tag{4.7}$$

As shown, when $k = 2$, this reduces to an exponential distribution. In general, $\mathrm{Var}(B) = \mathrm{Var}(L + R) \neq 1$. The effect of disparity and correlation is to change the variance of B, which in turn gives rise to the mean change in C. Importantly, the variance of B is affected by the *covariance* of L and R. This is because

$$\mathrm{Var}(L + R) = \mathrm{Var}(L) + \mathrm{Var}(R) + 2\mathrm{Cov}(L, R), \tag{4.8}$$

and so the BEM is actually computing the covariance between the left and right eye responses. Note that for half-matched stereograms, such as those used in Chapter 1 and 2, the covariance term is larger for half-matched stimuli at the preferred disparity. Indeed, in general, the disparity tuning arises by changes in the covariance between L and R. However, changing the variance of B does not change the nature of the distribution: as we change the correlation and disparity, we effectively change the length constant of the exponential distribution, but the BEM distribution will remain an exponential so long as $\mathrm{Var}(B_1) = \mathrm{Var}(B_2)$. In fact, this will often not be true. In the general case, the response distribution of the BEM is the sum of two squared normal random variables with different variances, whose analytical form can be derived [Moschopoulos, 1985], but is much less straightforward. The variances of the filter responses will generally not be widely different, and so the exponential approximation in Equation 4.7 is a useful simplification. As a

purely empirical observation, the BEM distribution is very nearly exponential for all disparities. One of the properties of the exponential distribution is that the variance is equal to the square of the mean. Thus, for all disparities,

$$\mathrm{E}[C]^2 \approx \mathrm{Var}[C]. \tag{4.9}$$

The BEM therefore makes the rather remarkable prediction that the spike count variance should increase with the mean spike count (and specifically with the mean spike count squared).

Our stimulus was a 1D dynamic random line stereogram (see Methods for details), and the tuning curves were computed by forward correlation using a 30ms window around the peak response of the model unit. Figure 4.1 shows the disparity tuning curve (a) with the corresponding external and internal variance (b and c) for an example BEM unit. The model cell was shown a dynamic RLS, where a new stereogram with a different disparity and correlation was presented every 30 ms. In order to summarise the dependence of the variance on the spike count, we fit a scalar slope term to the variances. For the internal variance, there is a clear linear dependence, as dictated by the Poisson spiking in the model (the slope is guaranteed to be exactly 1.0). For external variance, there is a quadratic relationship between spike count and external variance, as our derivation above shows. Naturally, this quadratic relationship is not well described by a simple scaling of the mean spike count.

4.3.2 Example neurons

In order to explore the above prediction in real cells, we recorded neuronal spiking data from 75 V1 neurons in the macaque visual cortex. The full procedure is documented in Chapter 3. Briefly, the stimulus was a sequence of 1D noise patterns which could either be binocularly correlated or anticorrelated at a range of disparities, or uncorrelated. Each frame was on the screen for 30 ms, and were presented in 3s trials. Disparity tuning curves were computed by summing the number of spikes observed in a 30 ms window, centred on the maximum disparity-related response of the cell. We compute the external variance by subtracting the internal variance from the total variance within a disparity (see Equations 4.2 and 4.3) Figure 4.2a shows the tuning curve of an example cell (lemM326c7). Figure 4.2b and c plot the external and internal variance, respectively, in the same format. Two points are worth noting. First and foremost, the BEM prediction that external variance should increase with firing rate clearly holds as the external variance follows the mean spike count very closely (Figure 4.2b). Second, whereas the BEM predicts that the relationship between external variance and mean spike count should be quadratic, Figure 4.2d shows that this relationship can be adequately described by a linear relationship 4.2. In other words, in this particular cell, variance increases linearly with firing rate.

The linear relationship between mean spike count and external variance is very consistent in
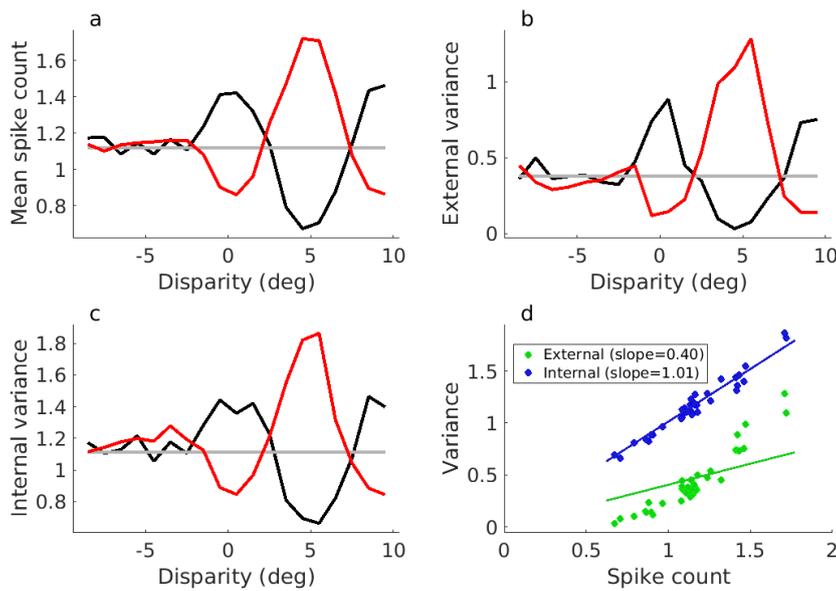
Figure 4.1: A binocular energy model unit's response to a random dot stereograms. a) The disparity tuning curve computed through forward correlation in a random line stereogram. b) The external variance as a function of disparity; as per the derivation above the external variance is proportional to the square of the mean spike count. c) The internal variance for the BEM unit, which simply reflects Poisson spiking (i.e. so that internal variance is equal to the mean spike count). d) External and internal variance plotted as a function of the mean spike count. Regression lines are computed without allowing an offset, such that when the mean spike count is 0, both variances must also be 0. While the external variance is quadratically related to the mean spike count, internal variance is linearly related to it due to Poisson spiking.
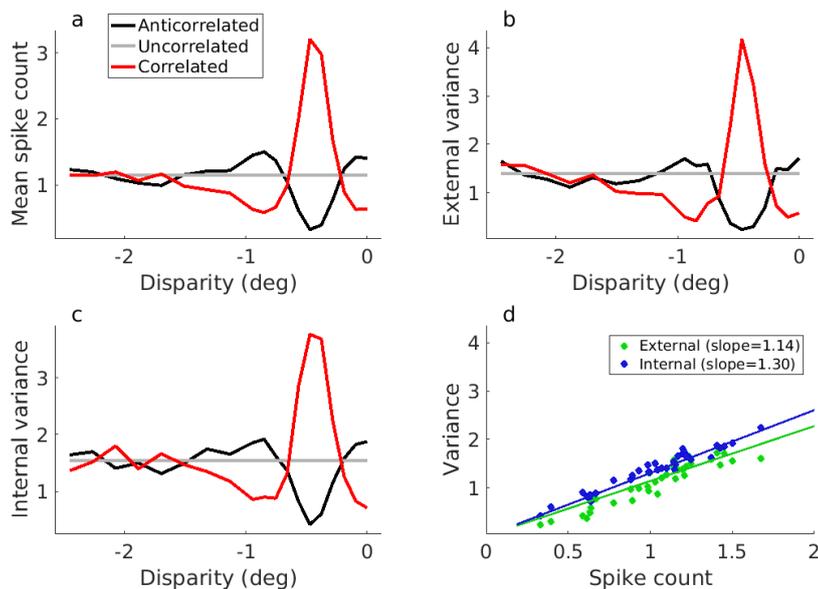

Figure 4.2: a) Disparity tuning curve for cell lemM326c7. External variance (b) and internal variance (c) are also shown for the same cell. d) External and internal variance plotted as a function of mean spike count. In contrast to the BEM (Figure 4.1), lemM326c7's external variance can be well described by as a linear function of the mean spike count.
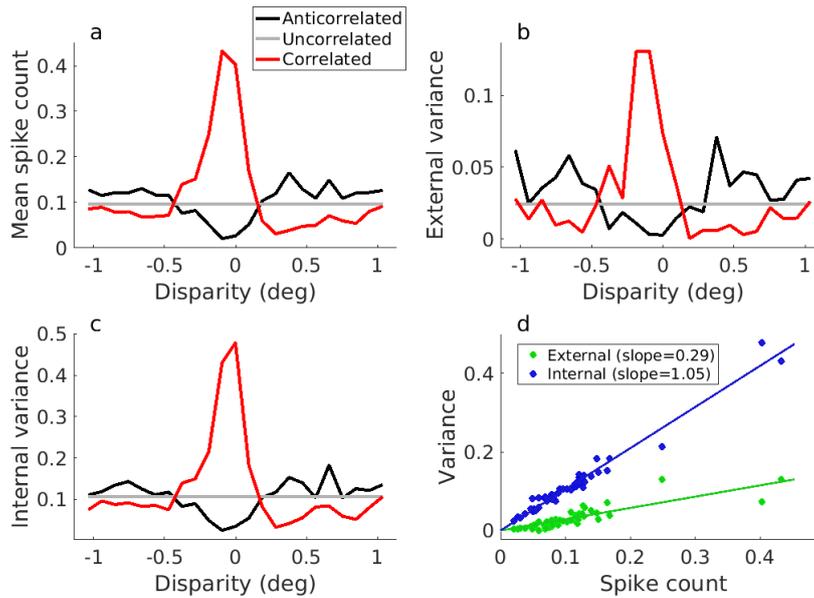
Figure 4.3: a) Disparity tuning curve for jbeM056c5. External variance (b) and internal variance (c) are also shown for the same cell. d) External and internal variance plotted as a function of mean spike count. As with lemM326c7 (Figure 4.2), this cell's external variance can also be well described by as a linear function of the mean spike count.

real cells. Figure 4.3 shows another example cell which demonstrates this relationship.

4.3.3   Generalised binocular energy model units

The binocular energy model imposes tight constraints on the form that a given model neuron can take. These constraints give rise to the prediction that the external variance should be quadratically related to the mean spike count. However, in the GBEM, the final spike rate reflects the sum of multiple subunits, some of which may be suppressive. Thus, the GBEM is not constrained to have a quadratic relationship between mean spike count and external variance. Testing whether external variance can be well described by the GBEM represents a novel and challenging test for the GBEM framework. It will also allow us to understand whether such model neurons capture the range of stimulus selectivity seen in real cells.

To explore this, we used the models fitted in Chapter 3. Note that these models were fitted such that when a trial was repeated for the two-pass procedure, both passes were put either in the validation set or in the training set. This was done to maintain independence between the training and validation sets, ensuring that the external variance is calculated completely on unseen data. Figure 4.4 shows a GBEM unit for cell lemM326c7 (Figure 4.2). In this model unit, the external variance follows the mean spike count (Figure 4.4), as in the real cells and the BEM. However, visually, the external variance appears to be neither as linear as the real cells, nor as quadratic as the BEM (Figure 4.4b and d).
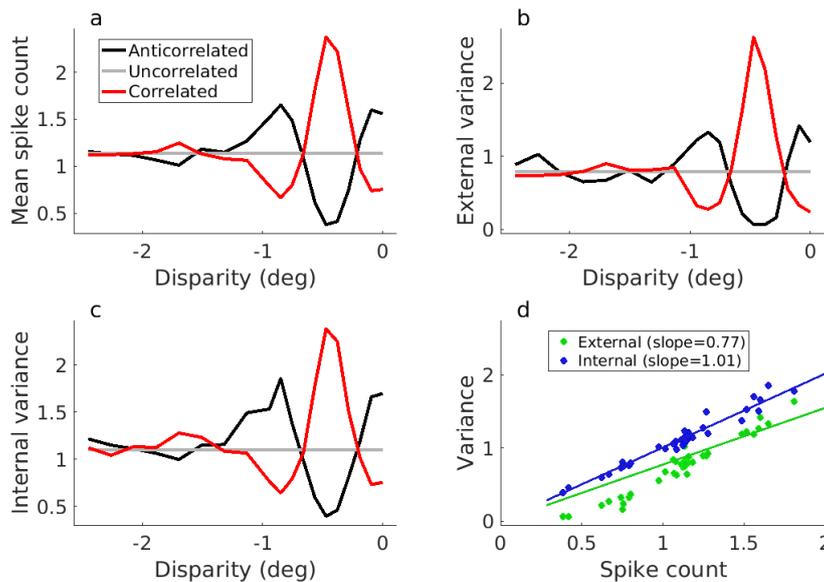
Figure 4.4: A GBEM unit fit to cell lemM326c7. a), b), and c) The mean spike count, external variance, and internal variance, respectively, all as a function of disparity. d) External and internal variance as a function of mean spike count.

Across the population, this is a weak, but systematic effect. In fitting the external variance as either a linear or quadratic function of the mean spike count, we find that $47/75$ (63%) of the cells are best described as a linear function of the mean spike count (defined as where R^2 for the linear model is greater than R^2 for the quadratic model). This contrasts with the GBEM units where $32/75$ (43%) are best described as a linear function of the mean spike count. The effect is significant for the cells, comparing the median R^2 for the linear model (median=0.82) to that of the quadratic model (median=0.78), $Z = 2.98$, $P = 0.003$ (Wilcon signed sum rank test). For the GBEM, there was no significant difference between the linear (median = 0.81) and the quadratic (median = 0.82) models, $Z = -1.12$, $P = 0.23$. Although the effect in Figure 4.4 is visually compelling, the relationship between external variance and mean spike count is often noisy for a given cell, even though it appears systematic across cells. Taken together, these results suggest that the GBEM goes some way towards capturing the linear dependence of external variance on mean spike count that is observed in real neurons.

We next turn our attention to comparing the external variances between the cell and the model. By superimposing the external variance of a GBEM unit onto that of the real cell, we can get a better sense of the GBEM's performance in capturing the underlying response statistics of the cell. Figure 4.5a shows the cell and the GBEM superimposed for the mean spike count, and Figure 4.5b shows the same for the external variance. For the external variance, we can see two key features: 1) the amplitude of the external variance modulation with disparity is larger in the real cell than in the GBEM, and 2) there is a vertical shift in the external variance for the GBEM.
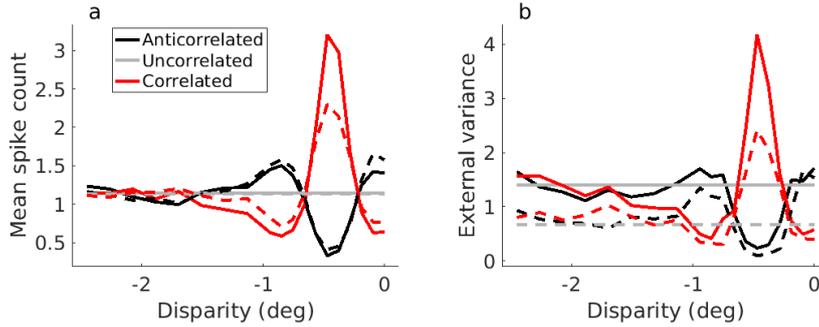
Figure 4.5: The mean spike count and external variance of cell lemM326c7 with the corresponding curves for its GBEM fit superimposed. The mean spike count (a) shows the effect of failing to capture the mean rate reported in Chapter 3 (correlated boosting). The external variance in the GBEM qualitatively mirrors the external variance in the cell, but there is a quantitative disagreement between the two.

4.3.4   Population summary

Given the known failure of the GBEM to capture the correlated response, it is likely that the smaller amplitude of the external variance reflects the same underlying failure. In order to quantify this effect, we plot the model tuning strength from Chapter 3 – the regression slope between model response and cell response – and the corresponding metric for external variance (i.e. the regression slope between the model external variance and cell external variance). We call this latter metric simply the external variance (EV) slope. As with the model tuning strength, if the EV slope is 1, it means that the magnitude of change in the external variance is perfectly captured by the model. On the other hand, if the EV slope is 0, it means that the model does not capture any of the disparity-related external variance modulation seen in the real cell. Importantly, if the failure to capture the amplitude of the external variance is due to the same underlying failure that is seen in the mean spike count (i.e. correlated boosting), then the model tuning strength and the EV slope should be positively correlated. Figure 4.6a shows these metrics plotted against one another for the population of cells. There is a clear and very significant relationship between the two ($r = 0.6, P < 10^{-5}$, Spearman correlation). This provides compelling evidence that the underestimate of the disparity-related modulation in external variance is due to the same mechanism which gives rise to the correlated boosting of the mean spike count. However, there is a second feature of the external variance in the GBEM which cannot be accounted for by correlated boosting. Indeed, this is evident in Figure 4.5 since the GBEM captures the mean spike count at all disparities except the preferred disparity of the cell (i.e. correlated boosting), but the external variance is underestimated for all disparities. Thus it is unlikely that the correlated boosting mechanism is responsible for the failure to capture the average external variance. To quantify the prevalence of external variance underestimation, we computed the average external variance for each cell and for its GBEM fit only on the uncorrelated responses. This allows us to examine the average external variance for data where the correlated boosting mechanism does not operate.
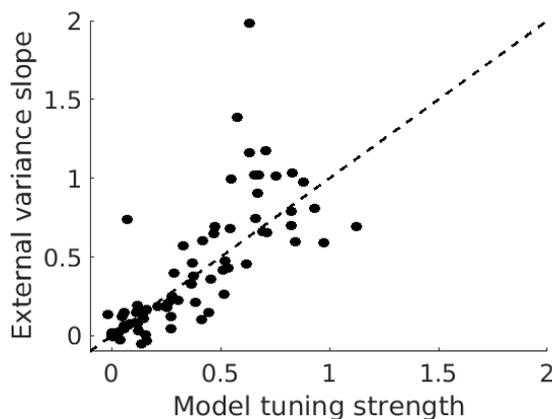
Figure 4.6: External variance slope plotted as a function of the model tuning strength. External variance slope is defined as the regression slope between external variance in the model and external variance in the cell. An external variance (EV) slope of 1 therefore means that the model is perfectly able to describe the magnitude of changes in external variance in the cell. Model tuning strength is as defined in Chapter 3 and is the regression slope between the mean spike count in the model and that in the cell. As with the EV slope, a model tuning strength of 1 means that the model is perfectly able to describe the disparity tuning curve of the cell. If the failure of the GBEM to capture changes in external variance with disparity can be accounted for by the GBEM's inability to capture changes in mean spike count with disparity, then there should be a positive correlation between EV slope and model tuning strength.

Comparing the average external variance of the cell to that of the GBEM, we see that in 70/75 cases, the cells had a higher external variance to uncorrelated stimuli. The average ratio of model external variance to cell external variance was 0.65, meaning that the model external variance was, on average, 65% that of the cells. We confirmed that this difference in the external variance was significant with a Wilcoxon ranked sum test, $Z = 6.57$, $P < 10^{-10}$.

4.4   Discussion

Compared to the mean rate of disparity-selective cells, the variability in response owing to stimulus variation has received very little attention. In this chapter, we have examined the theoretical properties of the binocular energy model (BEM) and showed that the BEM predicts a quadratic dependence of the external variance on the mean spike count. Recording from neurons in V1, we showed that BEM prediction is qualitatively borne out in real cells - V1 neurons do modulate their external variance with mean spike count. However, they do so linearly as opposed to quadratically as in the BEM. Using the GBEM units from Chapter 3, we showed that many of the best-fitting GBEM fits share the properties of the BEM: their external variance can often be better described as a function of the mean spike count squared rather than the mean spike count. Lastly, we showed that GBEM units quantitatively fail to capture two elements of the external variance: the magnitude of disparity-related modulation and the baseline external variance seen in real cells.

The latter two failures represent challenges to the GBEM framework. The result that failure to capture the external variance is related to the correlated boosting is intuitive: if the external variance increases with mean spike count, then we would expect that a failure to capture the mean spike count should result in a failure to capture the external variance. Thus, it seems likely that solving the question of correlated boosting should also resolve the issue of external variance near the preferred disparity of the cell. The second result that the GBEM has a lower baseline external variance than the cell (even to uncorrelated stimuli) is more puzzling. One possibility is that the noise model used in our GBEM units is inappropriate for the cells. However, internal variance is generally higher in real cells than assumed by Poisson spiking (i.e. Fano Factor is generally greater than 1), and so it is unclear how this would in turn cause *lower* external variance in the GBEM units. Lower external variance means that the cells have a much stronger selectivity for certain aspects of the stimulus compared to the GBEM units. The fact that the cells have higher external variance than the GBEM units means that they are responding more to some stimuli, and less to others, than predicted by the model. Since all stimuli are binocularly uncorrelated, this must be due to an aspect of the stimulus other than disparity. One interpretation is that, just as the cells show "correlated boosting" leading to greater than predicted responses to their preferred disparity, they also have other mechanisms which boost their responses to other stimulus features, perhaps some particular pattern of contrast for example. Understanding what these stimulus features are would represent an initial step for developing better models of stimulus encoding in cortex.

One assumption made by our modelling framework is that neuronal spiking is a stochastic Poisson process. A Poisson process constrains the internal variance to be equal to the mean spike count. It is, however, well-known that real neurons are generally super-Poisson, i.e. variance commonly exceeds the mean spike count, but rarely goes below it. It is unclear what the exact implication of this violated assumption is on the external variance. That being said, it seems unlikely that the systematically lower internal variance from Poisson spiking would cause a systematically lower external variance in the cells.

In the context of disparity-selective cells, external variance represents selectivity for stimulus features other than disparity. Neurons multiplex signals in this way in order to encode a rich representation of the stimulus. However, this multiplexing puts a fundamental limit on the amount of information which any one V1 neuron can transmit about disparity. Interestingly, in this scheme, the BEM is a poor disparity detector. This is because for a given mean rate, the exponential distribution is the maximum entropy distribution in the range $[0, \infty)$ (Park & Bera, 2009). This is noteworthy since it means that within each disparity, the BEM is encoding the maximum amount of information about the stimulus. However, this necessarily means that encoding the disparity of the stimulus is limited as information about disparity is prone to be confused with information about other aspects of the stimulus. At the level of individual cells, external variance therefore rep-

resents a substantial problem for decoding disparity. However, it is possible to construct schemes that can effectively cancel out part of this variability. For example, a standard BEM which also receives suppressive inputs from its monocular subunits can end up with a response which depends purely on the binocular interaction term. Such a model's response is simply $C' = 2L_1R_1 + 2L_2R_2$, and will be better able to decode disparity compared to the BEM. A similar point was noted by Burge & Geisler (2014), who showed that it is possible to create statistically optimal disparity detectors using a relatively modest number of LN subunits. These model units are termed log likelihood (LL) neurons as they modulate their response with the (log) likelihood of the neuron's preferred disparity given the stimulus. Such neurons respond with very little variability to a particular disparity, allowing the disparity to be decoded with a high degree of certainty. In this chapter, we have shown that V1 neurons exhibit very high external variance – in fact much higher external variance than the best-fitting GBEM units. This means that V1 cells are not LL neurons. However, although there is no evidence that such neurons exist in V1, it is possible that neurons very similar to LL neurons exist in higher cortical areas (e.g. IT).

Chapter 5

General discussion and future directions

The current view of the early stages of stereopsis is that disparity is initially encoded by locally filtering the images in the two eyes, and then computing something very similar to a binocular cross-correlation [Qian and Zhu, 1997, Fleet et al., 1996]. This binocular cross-correlation computation was originally represented by the BEM and has been hugely successful in accounting for a range of properties of both disparity-selective cells in V1 and psychophysical performance in humans. The success of the BEM has meant that it is often the first step in modelling perception/behaviour. For example, stereo resolution (i.e. the smallest disparity change that can be detected) is much worse than luminance resolution (i.e. the smallest luminance change than can be detected). Nienborg et al. (2004) showed that V1 neurons with larger receptive fields have lower stereo acuity. These results agreed well with the predictions of the BEM, which proposes that the envelopes of the linear filters constrain the smallest disparity change that can be detected. Importantly, when studying complex stimuli such as dynamic random dot stereograms, V1 neurons are modelled using early-stage spatiotemporal linear filters, and so the BEM makes predictions about the detectability of disparity changes in space as well as in time [Nienborg et al., 2004]. Subsequent studies have found compelling evidence that both the low spatial and temporal resolution of stereo vision (compared to resolution in the luminance domain) is due to the constraints imposed by the linear filtering stage, as predicted by the BEM [Banks et al., 2004, Kane et al., 2014, Allenmark and Read, 2012]. Thus, in the case of stereo resolution, the BEM provides a useful simplification which can provide meaningful insights into neuronal and perceptual processes.

However, despite these successes, the correlation-based view of early stereo vision can also be misleading. Recent experiments by Doi et al. (2011,2013) have brought into question the ability of a correlation-based view to account for disparity computations in cortex. Specifically, the authors argued that an additional match-based computation is required to account for the psychophysical performance seen in human observers in response to mixed correlation random dot stereograms. The first aim of this thesis was to examine whether such a revision is necessary. One key observation is that humans can see depth in half-matched random dot stereograms, even though the binocular correlation is zero in such stimuli. Since the BEM modulates its response proportionately

to binocular correlation, it does not (on average) signal disparity in stimuli with an equal number of correlated and anticorrelated dots. However, real neurons are not linear in this way [Cumming and Parker, 1997]. Chapter 1 showed how these psychophysical findings can be accounted for by a decision model which performs a straightforward readout of a model population where the individual cells better approximate activity in primary visual cortex. A key component of this model was the use of BEM units with a squaring output nonlinearity to encode disparity. These model units signal disparity even in half-matched stereograms, where the correlation is (on average) zero, allowing a linear decision model to correctly report depth in this case. Chapter 2 confirmed that real V1 neurons exhibit tuning to random dot stereograms with an equal number of correlated and anticorrelated dots, and so provides strong support for the view that a single nonlinear correlation computation carried out by V1 neurons is responsible for extracting disparity in this case.

The work in Chapter 1 and 2 highlights the need for better models of disparity-selective cells. The second aim of this thesis was to improve on existing models by building generalised BEM units through a data-driven machine learning approach. While the BEM and its variants have been able to describe a range of properties of V1 neurons, no one has yet been able to explicitly test this framework in real cells. The reason for this is two-fold. First, up until recently, there were no special-purpose optimisation algorithms that could test this hypothesis exhaustively [McFarland et al., 2013]. Second, these methods require large volumes of neuronal data (several hours), which are typically hard to obtain with single electrode recordings. Many modern array techniques (including the V- and U-probes used here) permit for much longer recordings from each neuron. These developments allow us to test the GBEM class of models by fitting the model architecture and parameters to the spiking data recorded in cortex. The GBEM is a version of an artificial neural network, which is now widely used in both academia and industry [LeCun et al., 2015]. Developing and testing a population of GBEM units fit to neuronal data was the key aim of Chapter 3. Our results show that the GBEM can account for some, but not all of the computational properties of V1 neurons. It is notable that the GBEM recovered disparity-selectivity in most cases even though our fitting procedure did not specify disparity or correlation; instead the input was simply the luminance pattern on the retinas, and the output was the spiking activity of the cell. Surprisingly, we found that real cells have an amplification of their responses to correlated stimuli at their preferred disparity, which model cells do not exhibit. This property had previously been thought to be an attenuated response to anticorrelated stimuli. Instead, we show that this is likely a mechanism for V1 neurons to boost their responses to candidate true matches. The view that neurons are boosting their response to candidate true matches provide an important constraint on the development of better models of V1 neurons.

Work here offers the hope that we will soon have a good enough account of V1 that we can begin to understand how these are subsequently used in extrastriate cortex, and what additional

processing is required to support perception. One area where GBEM units may provide insight into human depth perception is the effect of dot contrast polarity on depth discrimination performance. This effect was first reported by Harris & Parker (1995), who showed that human observers are better able to discriminate a depth step-edge in random dot stereograms with mixed black and white dots compared to stereograms that have either all white or all black dots. The performance increase for mixed polarity stimuli was consistent with the idea that the observers effectively used twice as many dots in the mixed polarity case (as determined by an ideal observer analysis). Harris & Parker argued that this suggests that there are two separate channels for disparity extraction of black and white stimuli. Read et al. (2011) reproduced the polarity effect reported by Harris & Parker, but showed that the BEM does not exhibit stronger tuning to mixed polarity stimuli. Additionally, since the work by Harris & Parker, there has been little further evidence for distinct ON and OFF channels involved in the computation of disparity. Thus, the results from Harris & Parker (1995) seems at odds with our understanding of disparity encoding in primary visual cortex. Recently, Goncalves & Welchman (2017) built a model which may shed light on this. The authors created a three-layer binocular convolutional neural network which was trained to identify disparity in natural images. When tasked with detecting a depth step-edge in either mixed polarity or unipolar RDSs, the model performed better on the mixed polarity stimuli, with the performance advantage again being consistent with the magnitude of the effect seen by Harris & Parker. However, according to Goncalves & Welchman, the performance increase reflects a change in the balance of excitation and inhibition depending on whether the model is seeing unipolar or mixed polarity stimuli. It is presently unclear whether this is what is happening in real neurons. The GBEM units reported here are fitted to neuronal data and so provide a direct way of testing this explanation.

An additional perceptual observation which better models of V1 neurons may be able to account for is reversed depth in anticorrelated stimuli. Read & Eagle (2000) reported reversed depth in anticorrelated noise stereograms, and Doi et al. (2011, 2013) have reported reversed depth in anticorrelated random dot stereograms. Reversed depth in anticorrelated stereograms has also been observed in our laboratory [Henriksen, 2014]. One key feature which is necessary to facilitate reversed depth perception is a correlated surround. Interestingly, Hibbard et al. (2014) have reported results from a large experiment where no reversed depth was perceived with anticorrelated stereograms with a zero-disparity correlated surround. However, there were a number of differences between the stimulus used by Hibbard et al. and that used by Doi et al (2011, 2013). One such difference is that Hibbard et al. included a zone with no dots which surrounded the disparity-defined region. Intuitively, the correlated surround provides an unambiguous zero-disparity reference against which to to make the depth judgement. However, it is at present unclear what the neuronal basis for this is. One possibility is that the GBEM units presented in this thesis provide a different signal at the boundary depending on the presence or absence of a gap between

the disparity-defined centre and the correlated surround. A related possibility is that there is an interaction between anticorrelated dots at one disparity (e.g. $< 0°$), and correlated dots at another, neighbouring disparity (e.g. $0°$). A very similar effect was originally reported by Neri et al. (1999), and may account for the apparent importance of a zero-disparity, correlated surround in inducing reversed depth judgements. Future studies which incorporate the GBEM units developed here may be useful for addressing this question.

Indeed, current models of V1 neurons seem to be converging on a design which is now very common in machine vision. Deep neural networks have been enormously successful in matching or exceeding human performance on a range of tasks, including image recognition [Liang and Hu, 2015], tumour detection and classification [Havaei et al., 2017], and optical character recognition [Ciregan et al., 2012]. The success of deep learning is commonly attributed to its ability to learn complex features from the data [LeCun et al., 2015]. This makes it uniquely powerful for doing tasks for which the data is unstructured, but from which meaningful features can be recovered. It is likely no accident that the low-level features that convolutional neural nets learn are very similar to the linear filters used in linear-nonlinear models of visual neurons [Krizhevsky et al., 2012]. However, these are greatly simplified models of neuronal computation. In this thesis, we have showed that a three-layer network cannot veridically describe disparity computations in primary visual cortex. This is because real neurons exhibit domain-specific specialisations, such as correlated boosting (boosting of candidate true matches in the disparity domain). It is possible that deeper architectures will result in models which generalise well to unseen stimuli. Deep nets achieve their success by learning complex nonlinear features which are useful for a given task. Thus, deep nets may be successful in predicting the behaviour of real neurons even if the model structure does not actually match the underlying physiology. Importantly, the mechanism which gives rise to correlated boosting in the deep idealised network may not be related in any straightforward manner to how this happens in real neurons. For example, suppose a nonlinear dendritic computation gives rise to the thresholded AND gate noted in Equation 3.5. Suppose also that this mechanism can be approximated by a network with several hidden layers, each of modest width. In this case, it is far from clear whether the deep network will on its own give any sort of insight into the underlying computation carried out by V1 neurons. Thus, although deeper networks present a viable research target for studying neuronal computation, they may on their own be insufficient for generating an actual understanding of what these neurons do. In principle, deep learning techniques can be applied to our existing dataset, although it is generally the case that more complex models require more data, which may be a limiting factor. Some cells in the dataset reported in Chapters 3 and 4 have relatively large amounts of data relative to that needed by the GBEM (this can be seen with a learning curve, plotting cross-validated model performance as a function of data amount). These cells are therefore good candidates for fitting more complex models.

An interesting possibility is that if deep neural nets can be fitted to neuronal data, then they may be used as proxies for real cells despite being black boxes. In such a scheme, a very complex black box model could be used to test hypotheses and even fit other models which are designed to extract insights and computational principles. Electrophysiological experiments are presently invasive and very expensive, and so having a large number of model cells which generalise well over a large set of stimuli would constitute a major breakthrough in both research efficiency and animal welfare. A series of recent papers have begun exploring deep neural networks as articles of research in their own right [Cadieu et al., 2014, Cichy et al., 2016].

The methods and techniques for modelling disparity-selective cells explored and discussed in this thesis are summarised in Table 1.

Method	Summary
BEM + output nonlinearity	Used in Chapter 1 and 2, and tested as general explanations in Chapter 3 and 4. Not viable as a solution to either odd-symmetric [Read et al., 2002] or even-symmetric disparity-tuned cells (shown in this thesis).
Two subunit (2SU) model	Tested in Chapter 3 and 4. Only clear candidate for odd-symmetric tuning curves with anticorrelated attenuation. GBEM units fitted to odd-symmetric cells consist of subunits which themselves have odd-symmetric tuning curves. Suggests odd-symmetry does not arise through a 2SU architecture with units of different disparity preference.
Recurrent networks	Examined by Samonds et al. (2013); not examined in this thesis. One of the key remaining hypotheses for how disparity tuning (especially correlated boosting) arises in cortex.
Monocular nonlinearities	The GBEM cannot easily capture both monocular nonlinearities and complex binocular interactions, so not exhaustively tested in the current thesis. This approach, introduced by Read et al. (2002), remains a candidate for accounting for the disparity tuning of at least some cells in V1.
Pairwise nonlinearities	Some work done in the domain of contrast, but not explored in stereo vision and not addressed in this thesis One candidate mechanism for how disparity tuning arises in cortex.
Deep neural networks	Currently not explored as a possible mechanism for V1 computations, but is a viable candidate mechanism for disparity selectivity.

Other approaches may be useful in elucidating some of these questions. For example, studying the delay in information transmission may enable us to deduce whether a particular aspect of pro-

cessing (e.g. correlated boosting) is due to recurrent connections. As discussed above, recurrent networks are a very plausible candidate for generating veridical models of V1 neurons. Similarly, studying the connectivity between regions may help constrain realistic models of neurons, especially for understanding how areas downstream from V1 generate their disparity tuning.

Stereo vision has a long history of being at the forefront of linking neuronal activity to perception [Julesz, 1971, Marr, 1980, Cumming and Parker, 1997]. This is in large part because stereo vision is a model system where the challenges posed to the brain are very well understood. Thus, although we still do not have a full understanding of the initial stages of disparity computation, we are making rapid progress in understanding the computations in V1, and how these computations relate to visual perception.

References

[Abdolrahmani et al., 2016] Abdolrahmani, M., Doi, T., Shiozaki, H. M., and Fujita, I. (2016). Pooled, but not single-neuron, responses in macaque v4 represent a solution to the stereo correspondence problem. *Journal of neurophysiology*, 115(4):1917–1931.

[Adelson and Bergen, 1985] Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, 2(2):284 – 299.

[Allenmark and Read, 2012] Allenmark, F. and Read, J. C. (2012). Conjunctions between motion and disparity are encoded with the same spatial resolution as disparity alone. *Journal of Neuroscience*, 32(41):14331–14343.

[Allenmark and Read, 2011] Allenmark, F. and Read, J. C. A. (2011). Spatial stereoresolution for depth corrugations may be set in primary visual cortex. *PLoS Computational Biology*, 7:e1002142.

[Anzai et al., 1999] Anzai, A., Ohzawa, I., and Freeman, R. D. (1999). Neural mechanisms for processing binocular information ii. complex cells. *Journal of Neurophysiology*, 82(2):909–924.

[Baba et al., 2015] Baba, M., Sasaki, K. S., and Ohzawa, I. (2015). Integration of multiple spatial frequency channels in disparity-sensitive neurons in the primary visual cortex. *Journal of Neuroscience*, 35(27):10025–10038.

[Backus et al., 2001] Backus, B. T., Fleet, D. J., Parker, A. J., and Heeger, D. J. (2001). Human cortical activity correlates with sterescopic depth perception. *Journal of Neurophysiology*, 86(4):2054 – 2068.

[Banks et al., 2004] Banks, M. S., Gepshtein, S., and Landy, M. S. (2004). Why is spatial stereoresolution so low. *Journal of Neuroscience*, 24:2077 – 2089.

[Barlow et al., 1967] Barlow, H. B., Blakemore, C., and Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *Journal of Physiology*, 193:327 – 342.

[Britten et al., 1992] Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, 12(12):4745–4765.

[Burge and Geisler, 2014] Burge, J. and Geisler, W. S. (2014). Optimal disparity estimation in natural stereo images. *Journal of vision*, 14(2):1–1.

[Cadieu et al., 2014] Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., and DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS computational biology*, 10(12):e1003963.

[Carandini and Heeger, 2012] Carandini, M. and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62.

[Cichy et al., 2016] Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., and Oliva, A. (2016). Deep neural networks predict hierarchical spatio-temporal cortical dynamics of human visual object recognition. *arXiv preprint arXiv:1601.02970*.

[Ciregan et al., 2012] Ciregan, D., Meier, U., and Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642–3649. IEEE.

[Clarke et al., 1976] Clarke, P., Donaldson, I., and Whitteridge, D. (1976). Binocular visual mechanisms in cortical areas i and ii of the sheep. *The Journal of physiology*, 256(3):509–526.

[Cogan et al., 1993] Cogan, A. I., Lomakin, A. J., and Rossi, A. F. (1993). Depth in anticorrelated stereograms: Effects of spatial density and interocular delay. *Vision Research*, 33:1959 – 1975.

[Cumming and Parker, 1997] Cumming, B. G. and Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, 389(1):280 – 283.

[Cumming and Parker, 1999] Cumming, B. G. and Parker, A. J. (1999). Binocular neurons in v1 of awake monkeys are selective for absolute, not relative, disparity. *Journal of Neuroscience*, 19(13):5602–5618.

[Cumming et al., 1998] Cumming, B. G., Shapiro, S. E., and Parker, A. J. (1998). Disparity detection in anticorrelated stereograms. *Perception*, 27:1367 – 1377.

[Doi and Fujita, 2014] Doi, T. and Fujita, I. (2014). Cross-matching: a modified cross-correlation underlying threshold energy model and match-based depth perception. *Frontiers in Computational Neuroscience*, 8(127).

[Doi et al., 2013] Doi, T., Takano, M., and Fujita, I. (2013). Temporal channels and disparity representations in stereoscopic depth perception. *Journal of Vision*, 13(13):1 – 25.

[Doi et al., 2011] Doi, T., Tanabe, S., and Fujita, I. (2011). Matching and correlation computations in stereoscopic depth perception. *Journal of Vision*, 11(3):1 – 16.

[Doi et al., 2014] Doi, T., Tanabe, S., and Fujita, I. (2014). Reversed depth perception in anticorrelated random-dot stereograms: when it is perceived and when it is not. http://www.plosone.org/annotation/listThread.action?root=78048.

[Draper and Smith, 2014] Draper, N. R. and Smith, H. (2014). *Applied regression analysis*. John Wiley & Sons.

[Efron and Tibshirani, 1994] Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

[Fawcett, 2005] Fawcett, S. L. (2005). An evaluation of the agreement between contour-based circles and random dot-based near stereoacuity tests. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 9(6):572–578.

[Ferster, 1988] Ferster, D. (1988). Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *Journal of Neuroscience*, 8(4):1172–1180.

[Filippini and Banks, 2009] Filippini, H. R. and Banks, M. S. (2009). Limits of stereopsis explained by local cross-correlation. *Journal of Vision*, 9:8.1–8.18.

[Fleet et al., 1996] Fleet, D. J., Wagner, H., and Heeger, D. J. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision research*, 36(12):1839–1857.

[Fricke and Siderov, 1997] Fricke, T. R. and Siderov, J. (1997). Stereopsis, stereotests, and their relation to vision screening and clinical practice. *Clinical and Experimental Optometry*, 80(5):165–172.

[Giaschi et al., 2013] Giaschi, D., Narasimhan, S., Solski, A., Harrison, E., and Wilcox, L. M. (2013). On the typical development of stereopsis: Fine and coarse processing. *Vision Research*, 89:65 – 71.

[Gold and Shadlen, 2007] Gold, J. I. and Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, 30:535–574.

[Goncalves and Welchman, 2017] Goncalves, N. R. and Welchman, A. E. (2017). what not detectors help the brain see in depth. *Current Biology*, 27(10):1403–1412.

[Goris et al., 2013] Goris, R. L., Putzeys, T., Wagemans, J., and Wichmann, F. A. (2013). A neural population model for visual pattern detection. *Psychological review*, 120(3):472.

[Green and Swets, 1966] Green, D. and Swets, J. (1966). *Signal detection theory and psychophysics*. Wiley.

[Gu et al., 2008] Gu, Y., Angelaki, D. E., and DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque mstd. *Nature neuroscience*, 11(10):1201–1210.

[Haefner and Cumming, 2008] Haefner, R. M. and Cumming, B. G. (2008). Adaptation to natural binocular disparities in primate v1 explained by a generalized energy model. *Neuron*, 57(1):147–158.

[Havaei et al., 2017] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., and Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31.

[Hawken et al., 1996] Hawken, M., Shapley, R., and Grosof, D. (1996). Temporal-frequency selectivity in monkey visual cortex. *Visual neuroscience*, 13(3):477–492.

[Henriksen, 2014] Henriksen, S. (2014). The temporal dynamics of disparity extraction is consistent with a single mechanism underlying stereopsis. Master's thesis, Institute of Neuroscience, Newcastle University.

[Henriksen et al., 2016a] Henriksen, S., Cumming, B. G., and Read, J. C. (2016a). A single mechanism can account for human perception of depth in mixed correlation random dot stereograms. *PLoS Comput Biol*, 12(5):e1004906.

[Henriksen et al., 2016b] Henriksen, S., Read, J. C., and Cumming, B. G. (2016b). Neurons in striate cortex signal disparity in half-matched random-dot stereograms. *Journal of Neuroscience*, 36(34):8967–8976.

[Henriksen et al., 2016c] Henriksen, S., Tanabe, S., and Cumming, B. (2016c). Disparity processing in primary visual cortex. *Phil. Trans. R. Soc. B*, 371(1697):20150255.

[Hibbard et al., 2014] Hibbard, P. B., Scott-Brown, K. C., Haigh, E. C., and Adrain, M. (2014). Depth perception not found in human observers for static or dynamic anti-correlated random dot stereograms. *PLoS One*, 9(1):1 – 9.

[Hornik, 1991] Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257.

[Howard, 2002] Howard, I. P. (2002). *Seeing in depth, Vol. 1: Basic mechanisms*. University of Toronto Press.

[Howard and Rogers, 1995] Howard, I. P. and Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford University Press, USA.

[Hubel and Wiesel, 1970] Hubel, D. H. and Wiesel, T. N. (1970). Stereoscopic vision in macaque monkey: cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature*, 225:41–42.

[Janssen et al., 2003] Janssen, P., Vogels, R., Liu, Y., and Orban, G. A. (2003). At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. *Neuron*, 37(4):693–701.

[Julesz, 1971] Julesz, B. (1971). *Foundations of Cyclopean Perception*. MIT Press.

[Kane et al., 2014] Kane, D., Guan, P., and Banks, M. S. (2014). The limits of human stereopsis in space and time. *Journal of Neuroscience*, 34:1397 – 1408.

[Kleiner et al., 2007] Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). Whats new in psychtoolbox-3. *Perception*, 36(14):1.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

[Krug et al., 2004] Krug, K., Cumming, B. G., and Parker, A. J. (2004). Comparing perceptual signals of single v5/mt neurons in two binocular depth tasks. *Journal of neurophysiology*, 92(3):1586–1596.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

[Liang and Hu, 2015] Liang, M. and Hu, X. (2015). Recurrent convolutional neural network for object recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3367–3375.

[Marr, 1980] Marr, D. (1980). *Vision*. MIT Press.

[Marr and Poggio, 1976] Marr, D. and Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, 194:283 – 287.

[Marr and Poggio, 1979] Marr, D. and Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London B*, 204:301 – 328.

[McFarland et al., 2013] McFarland, J. M., Cui, Y., and Butts, D. A. (2013). Inferring nonlinear neuronal computation based on physiologically plausible inputs. *PLoS computational biology*, 9(7):e1003143.

[Moschopoulos, 1985] Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics*, 37(1):541–544.

[Nair and Hinton, 2010] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

[Nakayama and Shimojo, 1990] Nakayama, K. and Shimojo, S. (1990). Da vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision research*, 30(11):1811–1825.

[Nieder and Wagner, 2000] Nieder, A. and Wagner, H. (2000). Horizontal-disparity tuning of neurons in the visual forebrain of the behaving barn owl. *Journal of Neurophysiology*, 83(5):2967–2979.

[Nienborg et al., 2004] Nienborg, H., Bridge, H., Parker, A. J., and Cumming, B. G. (2004). Receptive field size in v1 neurons limits acuity for perceiving disparity modulation. *Journal of Neuroscience*, 24(9):2065–2076.

[Nienborg and Cumming, 2006] Nienborg, H. and Cumming, B. G. (2006). Macaque v2 neurons, but not v1 neurons, show choice-related activity. *Journal of Neuroscience*, 26(37):9567–9578.

[Nienborg and Cumming, 2014] Nienborg, H. and Cumming, B. G. (2014). Decision-related activity in sensory neurons may depend on the columnar architecture of cerebral cortex. *Journal of Neuroscience*, 34(10):3579–3585.

[Norcia et al., 1985] Norcia, A. M., Suiter, E. E., and Tyler, C. W. (1985). Electrophysiological evidence for the existence of coarse and fine disparity mechanisms in human. *Vision research*, 25(11):1603–1611.

[Ogle, 1952a] Ogle, K. N. (1952a). Disparity limits of stereopsis. *AMA archives of ophthalmology*, 48(1):50–60.

[Ogle, 1952b] Ogle, K. N. (1952b). On the limits of stereoscopic vision. *Journal of Experimental Psychology*, 44(4):253 – 259.

[Ogle, 1952c] Ogle, K. N. (1952c). On the limits of stereoscopic vision. *Journal of experimental psychology*, 44(4):253.

[Ohzawa et al., 1990] Ohzawa, I., DeAngelis, G. C., and Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, 249(1):1037 – 1041.

[Ohzawa and Freeman, 1986] Ohzawa, I. and Freeman, R. D. (1986). The binocular organization of simple cells in the cat9s visual cortex. *Journal of Neurophysiology*, 56(1):221–242.

[Parker, 2007] Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature*, 8:379 – 391.

[Parker and Newsome, 1998] Parker, A. J. and Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annual review of neuroscience*, 21(1):227–277.

[Pettigrew et al., 1968] Pettigrew, J., Nikara, T., and Bishop, P. (1968). Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. *Experimental Brain Research*, 6(4):391–410.

[Pettigrew and Konishi, 1976] Pettigrew, J. D. and Konishi, M. (1976). Neurons selective for orientation and binocular disparity in the visual wulst of the barn owl (tyto alba). *Science*, 193(4254):675–678.

[Poggio and Fischer, 1977] Poggio, G. and Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology*, 40(6):1392–1405.

[Prince et al., 2000] Prince, S. J., Pointon, A. D., Cumming, B. G., and Parker, A. J. (2000). The precision of single neuron responses in cortical area v1 during stereoscopic depth judgments. *Journal of Neuroscience*, 20(9):3387–3400.

[Prince et al., 2002a] Prince, S. J. D., Cumming, B. G., and Parker, A. J. (2002a). Range and mechanism of horizontal disparity encoding in macaque v1. *Journal of Neurophysiology*, 98:209 – 221.

[Prince and Eagle, 1999] Prince, S. J. D. and Eagle, R. A. (1999). Size-disparity correlation in human binocular depth perception. *Proceedings of the Rocyal Society of London B*, 266:1361 – 1365.

[Prince et al., 2002b] Prince, S. J. D., Pointon, A. D., Cumming, B. G., and Parker, A. J. (2002b). Quantitative analysis of responses of v1 neurons to horizontal disparity in dynamic random dot stereograms. *Journal of Neurophysiology*, 87:191 – 208.

[Qian and Zhu, 1997] Qian, N. and Zhu, Y. (1997). Physiological computation of binocular disparity. *Vision Research*, 37:1811 – 1827.

[Read, 2002] Read, J. C. (2002). A bayesian model of stereopsis depth and motion direction discrimination. *Biological cybernetics*, 86(2):117–136.

[Read, 2015] Read, J. C. (2015). Stereo vision and strabismus. *Eye*, 29:214–224.

[Read and Cumming, 2003] Read, J. C. and Cumming, B. G. (2003). Measuring v1 receptive fields despite eye movements in awake monkeys. *Journal of neurophysiology*, 90(2):946–960.

[Read, 2005] Read, J. C. A. (2005). Early computational processing in binocular vision and depth perception. *Progress in Biophysics and Molecular Biology*, 87(1):77 – 108.

[Read and Eagle, 2000] Read, J. C. A. and Eagle, R. A. (2000). Reversed stereo depth and motion direction with anti-correlated stimuli. *Vision Research*, 40:3345 – 3358.

[Read et al., 2002] Read, J. C. A., Parker, A. J., and Cumming, B. G. (2002). A simple account for the response of disparity tuned V1 neurons to anticorrelated images. *Visual Neuroscience*, 19(1):735 – 753.

[Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323:533–536.

[Samonds et al., 2009] Samonds, J. M., Potetz, B. R., and Lee, T. S. (2009). Cooperative and competitive interactions facilitate stereo computations in macaque primary visual cortex. *Journal of Neuroscience*, 29(50):15780–15795.

[Samonds et al., 2013] Samonds, J. M., Potetz, B. R., Tyler, C. W., and Lee, T. S. (2013). Recurrent connectivity can account for the dynamics of disparity processing in v1. *The Journal of Neuroscience*, 33(7):2934 – 2946.

[Shadlen et al., 1996] Shadlen, M. N., Britten, K. H., Newsome, W. T., and Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience*, 16(4):1486–1510.

[Simoncelli and Heeger, 1998] Simoncelli, E. P. and Heeger, D. J. (1998). A model of neuronal responses in visual area mt. *Vision research*, 38(5):743–761.

[Smallman and MacLeod, 1994] Smallman, H. S. and MacLeod, D. I. A. (1994). Size-disparity correlation in stereopsis at contrast threshold. *Journal of the Optical Society of America A*, 11:2169 – 2183.

[Takemura et al., 2001] Takemura, A., Inoue, Y., Kawano, K., Quaia, C., and Miles, F. (2001). Single-unit activity in cortical area mst associated with disparity-vergence eye movements: evidence for population coding. *Journal of Neurophysiology*, 85(5):2245–2266.

[Tanabe and Cumming, 2008] Tanabe, S. and Cumming, B. G. (2008). Mechanisms underlying the transformation of disparity signals from v1 to v2 in the macaque. *Journal of Neuroscience*, 28(44):11304–11314.

[Tanabe and Cumming, 2014] Tanabe, S. and Cumming, B. G. (2014). Delayed suppression shapes disparity selective responses in monkey v1. *Journal of neurophysiology*, 111(9):1759–1769.

[Tanabe et al., 2011a] Tanabe, S., Haefner, R. M., and Cumming, B. G. (2011a). Suppressive mechanisms in monkey v1 help to solve the stereo correspondence problem. *Journal of Neuroscience*, 31(22):8295–8305.

[Tanabe et al., 2011b] Tanabe, S., Haefner, R. M., and Cumming, B. G. (2011b). Suppressive mechanisms in monkey v1 help to solve the stereo correspondence problem. *Journal of Neuroscience*, 31(22):8295–8305.

[Tanabe et al., 2004] Tanabe, S., Umeda, K., and Fujita, I. (2004). Rejection of false matches for binocular correspondence in macaque visual cortical area v4. *The Journal of neuroscience*, 24(37):8170–8180.

[Tanabe et al., 2008] Tanabe, S., Yasuoka, S., and Fujita, I. (2008). Disparity-energy signals in perceived stereoscopic depth. *Journal of vision*, 8(3):22–22.

[Thomas et al., 2002] Thomas, O. M., Cumming, B. G., and Parker, A. J. (2002). A specialization for relative disparity in v2. *Nature Neuroscience*, 5(5):472–478.

[Tolhurst et al., 1983] Tolhurst, D. J., Movshon, J. A., and Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research*, 23(8):775–785.

[Tyler, 1974] Tyler, C. W. (1974). Depth perception in disparity gratings. *Nature*, 251:140 – 142.

[Uka and DeAngelis, 2003] Uka, T. and DeAngelis, G. C. (2003). Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *Journal of Neuroscience*, 23(8):3515–3530.

[Uka and DeAngelis, 2004] Uka, T. and DeAngelis, G. C. (2004). Contribution of area mt to stereoscopic depth perception: choice-related response modulations reflect task strategy. *Neuron*, 42(2):297–310.

[van Ee and Erkelens, 2010] van Ee, R. and Erkelens, C. J. (2010). Stereo-vision: head-centric coding of retinal signals. *Current Biology*, 20:R567 – R568.

[Vinyals et al., 2014] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. (2014). Show and tell: A neural image caption generator. *CoRR*, abs/1411.4555.

[Von Helmholtz, 1881] Von Helmholtz, H. (1881). *Popular lectures on scientific subjects*. D. Appleton.

[Westheimer and Tanzman, 1956] Westheimer, G. and Tanzman, I. J. (1956). Qualitative depth localization with diplopic images. *Journal of the Optical Society of America*, 46(2):116 – 117.

[Zhang et al., 2010] Zhang, Z.-L., Cantor, C. R., and Schor, C. M. (2010). Perisaccadic stereo depth with zero retinal disparity. *Current Biology*, 20:1176 – 1181.