

# Border Ownership Representation and Polychronisation in a Spiking Neural Network Model

Clemens Hutter



UNIVERSITY OF  
**OXFORD**



UNIVERSITÄT  
OSNABRÜCK

## Supervisors

Prof. Dr. Simon Stringer    University of Oxford  
Prof. Dr. Gordon Pipa      Universität Osnabrück

July 8, 2018

# Abstract

In this computer simulation study we investigated a hierarchical neural network model of the ventral visual pathway in the primate brain. It consists of four layers of spiking leaky integrate and fire neurons that are connected by bottom-up, top-down and lateral synapses. These synapses are modified through Spike Time Dependent Plasticity while visual stimuli of simple shapes are presented to the network. Similarly to Eguchi and Stringer (2016), we observed the emergence of cells that seem to represent border ownership, that is they seemed to encode the presence of a border in their receptive field and which side of the shape it is located on. However, detailed investigations later indicated that these cells instead encode the presence of some contour in rough regions of the retina and that they can therefore not be considered border ownership cells. This might also be the case for some of the cells reported in Eguchi and Stringer (2016). Further, we investigated the emergence of reliable spatio-temporal patterns of spiking activity (polychronous groups). We initially observed large numbers of informative polychronous groups consisting of two neurons, which is in line with the findings of Eguchi et al. (2018). Additionally, we found that the high values of information in polychronous groups are a side effect of information carried by firing rates of cells. Thus, the precise times of spikes do not seem to carry information in the investigated network. All in all, it might be that the problem with measuring polychronous group information has also occurred in Eguchi et al. (2018), which raises doubts about the reported high information carried by polychronous groups.

# Acknowledgements

I would like to thank Prof. Stringer for making this research project possible as well as for his continued advice, inspiration and enthusiasm. My deep gratitude goes to Nasir Ahmad for his extensive help in all parts of the project. I also want to thank Akihiro Eguchi and Gisbert Teepe for allowing me to continue their project. Further, I would like to thank the other members of the OFTNAI team for their warm welcome and many insight full discussions: James Isbister, Harry Jordan, Daniel Navarro, Karen Verstraelen, Toby Smithe, Elliot Smith and Hannah Dury. Additionally, I want to thank Ann Xavier for her support and her feedback on the report.

Finally, I want to express my appreciation for the funding provided by ERASMUS+ which made my research visit possible.

# Contents

<b>Abstract</b>	<b>i</b>
<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 Hypothesised Mode of Operation of Border Ownership Cells . . . . .	2
1.2.1 Previous Rate Coded Model of Border Ownership by Eguchi and Stringer (2016) . . . . .	2
1.2.2 Hypothesised Solution in a Spiking Network Model . . . . .	4
1.3 Polychronisation . . . . .	6
1.4 <i>Continuous Transform</i> and <i>Trace Learning Mechanisms</i> . . . . .	7
<b>2 Methods</b>	<b>9</b>
2.1 Spiking Neural Network Model . . . . .	9
2.1.1 Network Architecture . . . . .	9
2.1.2 Differential Equations . . . . .	9
2.1.3 Numerical Scheme and Implementation . . . . .	13
2.1.4 Input Representation of Visual Stimuli . . . . .	13
2.2 Stimulus Sets . . . . .	14
2.3 Detecting Oscillations of Neuron Populations . . . . .	18
2.4 Performance Measures . . . . .	19
2.4.1 Single Cell Information . . . . .	19
2.4.2 Single Cell Classification Accuracy . . . . .	21
2.4.3 Spike Pair PG Information . . . . .	23

Contents	iv
<b>3 Results</b>	<b>24</b>
3.1 Firing Rates . . . . .	24
3.2 Oscillations . . . . .	26
3.3 Information in Single Cell Firing Rates . . . . .	27
3.3.1 V4 Boundary Contour Cells . . . . .	28
3.3.2 The Effect of Trace Learning . . . . .	32
3.3.3 V1/V2 Border Ownership Neurons . . . . .	35
3.4 Polychrony . . . . .	41
3.4.1 Spike Pair Information . . . . .	41
3.4.2 Information of Spike Times Relative to the Underlying Oscillation . .	43
<b>4 Discussion</b>	<b>46</b>
4.1 Problems with the Border Ownership Representation . . . . .	46
4.2 Problems with Polychronisation . . . . .	49
4.3 Conclusion and Future Directions . . . . .	52
<b>Abbreviations</b>	<b>54</b>
<b>Bibliography</b>	<b>55</b>

# List of Figures

1.1	Hypothesised mechanism for border ownership cells in a rate coded network	2
1.2	Hypothesised mechanism for border ownership cells in a spiking network	5
1.3	Three neuron binding circuit	5
2.1	Training stimuli	15
2.2	Novel testing stimuli	16
2.3	Stimuli without border at target location	17
2.4	Stimuli with two objects	18
2.5	Example of an informative polychronous group which does not result in high spike pair information	22
3.1	Firing rates in response to example stimuli	25
3.2	Oscillation of neural activity	27
3.3	Information about the border type for training stimuli	29
3.4	Information about the border type for novel stimuli	31
3.5	Gabor filter that have strong connections to three example layer 4 cells	32
3.6	The effect of the trace learning mechanism on different tasks	34
3.7	Performance of border ownership cells for training stimuli	36
3.8	Performance of border ownership cells for novel stimuli	37
3.9	Development of activity and receptive fields of two example border ownership cells	39
3.10	Receptive field of Neuron(31, 23)	39
3.11	Firing rate responses to training and novel stimuli of two example border ownership cells	40
3.12	Spike pair information about border ownership	42
3.13	Polychronous groups given by the first spike times of cells relative to the peak of underlying neural activity	44

4.1	Approximate receptive fields of two border ownership cells reported in Eguchi and Stringer (2016) . . . . .	47
4.2	Schema of stimulus outlines . . . . .	47

# List of Tables

2.1	Model parameters . . . . .	10
-----	----------------------------	----

# Chapter 1

## Introduction

### 1.1 Overview

In cortical areas V1 and V2 a class of so called *border ownership* cells has been found (Zhou et al., 2000). These cells respond to oriented edges in the visual field similar to simple cells in V1. However, the response is additionally modulated by which side of the object the edge is located on. A cell might for example be selective to a vertical edge in it's receptive field, but only if that edge is on the left hand side of the object. However, their receptive field size is too small to capture enough context to determine which side of an object the border is on.

Therefore, this information is likely to come from a brain area with larger receptive fields further up in the hierarchy. Pasupathy and Connor (2001) identified area V4 as a potential site. They found cells in this area that are responsive to a certain contour element occurring at a specified position with respect to the object's centre of mass. These cells are invariant to the object's exact location on the retina. Such a cell would, for example, encode the presence of an object with a straight vertical border on it's left hand side anywhere in a larger region of the retina.

Since a *border ownership* cell is selective to a vertical edge at a tightly specified location (information present in simple cells in V1) but only if the edge is on a certain side of an object (information present in V4), such a cell effectively binds these two aspects together. The feature 'vertical edge at this location' is bound to the object 'something with a vertical edge on it's left border in a broad area of the retina'. Therefore the neuron performs *feature binding*, a crucial open question in visual neuroscience (Von Der Malsburg, 1999).

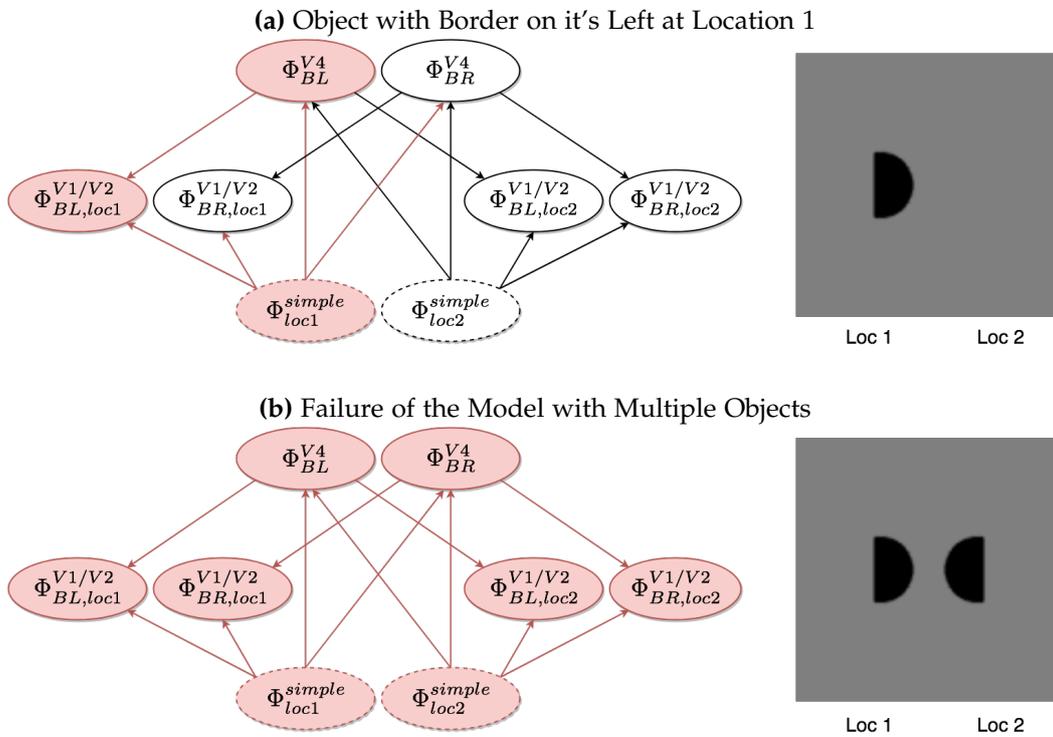
Eguchi and Stringer (2016) proposed a rate coded network, in which border ownership neurons in V1/V2 receive synaptic input from simple cells in V1 and feedback connections from the mentioned contour cells in V4, thus binding these features together. However, their model fails if multiple objects are presented at the same time, since, in a rate coded network, the feed-back from V4 can not be associated with only one of the multiple objects.

In this report we investigate how a *spiking* artificial neural network model might over-

come these limitations by associating the feedback with the corresponding object through patterns in exact spike times. We show how border ownership neurons emerge through visually guided training in a biologically plausible way. We then check in minute detail whether these neurons can be said to constitute binding neurons. Furthermore, we discuss briefly the effect of different learning paradigms. Finally, we will carefully examine the role of exact spike times and polychronisation. (Izhikevich, 2006)

## 1.2 Hypothesised Mode of Operation of Border Ownership Cells

### 1.2.1 Previous Rate Coded Model of Border Ownership by Eguchi and Stringer (2016)



**Figure 1.1:** Hypothesised connectivity of V1/V2 border ownership and V4 contour cells adapted from Eguchi and Stringer (2016). Arrows symbolise connectivity, which might involve intermediate neurons. Active connections or cells are shown in red. The original visualisation in Eguchi and Stringer (2016) additionally shows lateral inhibitory connections and feedforward connections from V1/V2 border ownership cells to V4 boundary contour cells. For clarity, these are omitted in the present figure without changing the core idea. (a) A stimulus with a left border at location 1 is shown. First, the simple cells in V1 encoding a vertical edge at location 1 ( $\Phi_{loc1}^{simple}$ ) are activated. These then drive both  $\Phi_{BL}^{V4}$  and  $\Phi_{BR}^{V4}$  cells. However, only  $\Phi_{BL}^{V4}$  have sufficient contextual input from other cells (not shown here) to become maximally active (Eguchi et al., 2015). Both  $\Phi_{BL}^{V4}$  and  $\Phi_{loc1}^{simple}$  are now active and drive  $\Phi_{BL,loc1}^{V1/V2}$  cells, which respond maximally. (b) The border ownership cells lose specificity when 2 objects are presented. In this case both  $\Phi_{BL}^{V4}$  and  $\Phi_{BR}^{V4}$  cells have sufficient contextual input from the object at location 1 or the object at location 2 respectively and become active. Therefore, all 4 types of border ownership cells receive feedback activation from V4 as well as feedforward activation from the simple cells in V1 and become active, thus losing the desired behaviour. The problem arises since the feedback from V4 cells is not bound to either one of the 2 objects (Eguchi and Stringer, 2016).

This report is largely based on the work by Eguchi and Stringer (2016), who showed how border ownership cells can emerge in a self organising network through visually guided learning. Their model consists of three simulated neural layers modelling the section from V2 to V4 of the ventral visual pathway in the primate brain. These are preceded by a static input layer corresponding to V1, in which the firing rates of simple cells are explicitly given according to the activation of Gabor filters. Each of the simulated cells has a firing rate that evolves over time, governed by a differential equation. However, exact spike times are not explicitly modelled and the firing rate changes on a rather slow temporal scale. Each neuron has afferent feedforward and feedback connections from a topologically corresponding region in the previous and the following layer respectively. The weights of these synapses are modified during the presentation of visual stimuli according to an unsupervised, local learning rule. The network is then presented with a variety of visual stimuli, all of which have a vertical edge on either their left or their right side. This edge is always perfectly aligned with one of two discrete locations on the retina.

During training with these stimuli, a class of neurons that encodes local boundary contour elements, develops in Layer 3 through modified feedforward connections. These neurons correspond to the V4 cells found by Pasupathy and Connor (2001, 2002) and their emergence had already been previously demonstrated in Eguchi et al. (2015). With the simple stimulus set we only consider the class of cells encoding the presence of an object with a vertical border on either its left or right side, denoted by  $\Phi_{BL}^{V4}$  and  $\Phi_{BR}^{V4}$  respectively. For these cells it does not matter whether the border is located at location 1 or 2 since the V4 cells exhibit location invariance in a region of the retina (Pasupathy and Connor, 2001, 2002) and it is simply supposed that only this region of the retina is simulated.

Furthermore, a class of cells resembling the mentioned border ownership cells in V1/V2 develops in earlier layers of the model. With the given stimulus set, 4 types of border ownership cells can be distinguished. They respond to either

- an object with a border on its **left side**, which is situated at **location 1** (denoted by  $\Phi_{BL,loc1}^{V1/V2}$ ).
- an object with a border on its **right side**, which is situated at **location 1** (denoted by  $\Phi_{BR,loc1}^{V1/V2}$ ).
- an object with a border on its **left side**, which is situated at **location 2** (denoted by  $\Phi_{BL,loc2}^{V1/V2}$ ).
- an object with a border on its **right side**, which is situated at **location 2** (denoted by  $\Phi_{BR,loc2}^{V1/V2}$ ).

Eguchi and Stringer (2016) propose, that these neurons work in the following manner (see figure 1.1a): Let us consider for example the presentation of a stimulus with a border on the left of an object that is presented at location 1 and examine the behaviour of a  $\Phi_{BL,loc1}^{V1/V2}$  neuron. It receives feedforward input from simple cells in V1 that respond to a vertical edge at location 1 (denoted by  $\Phi_{loc1}^{simple}$ ). The same simple cells also contribute to the activation of both types of boundary contour cells in V4 ( $\Phi_{BL}^{V4}$  and  $\Phi_{BR}^{V4}$ ). However, the

class  $\Phi_{BL}^{V4}$  is activated the most due to input from other cells in earlier layers that provide the necessary context to determine, that the border is actually on the object's left hand side. As a result, the  $\Phi_{BL,loc1}^{V1/V2}$  cells under examination here, receive feedback input from these higher layer  $\Phi_{BL}^{V4}$  cells, in addition to the already mentioned feedforward input from the class  $\Phi_{loc1}^{simple}$ . Since both of these cell types are very active for the example stimulus, the border ownership cells of type  $\Phi_{BL,loc1}^{V1/V2}$  have strong feedforward and feedback input and are therefore activated the most. Thus, as expected, the most active cells are the ones that encode a border at location 1, which is on the left side of an object ( $\Phi_{BL,loc1}^{V1/V2}$ ).

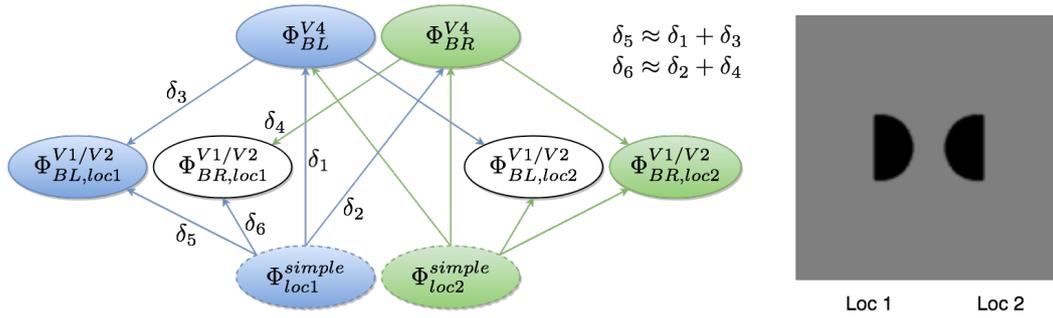
But, as Eguchi et al. (2015) correctly predicted, this model fails when it is presented with two objects, one at location 1 and the other one at location 2. In this case (see figure 1.1b) both  $\Phi_{BL}^{V4}$  and  $\Phi_{BR}^{V4}$  have sufficient contextual input to become active. The information about which of the objects at location 1 or 2 drove the response in  $\Phi_{BL}^{V4}$  or  $\Phi_{BR}^{V4}$  is not maintained in their rate coded activity. Therefore, the feedback from the two types of V4 neurons reaches all four types of border ownership neurons in V1/V2 equally. Since there is an edge present at location 1 and location 2 all of these four types of border ownership cells receive the necessary additional feedforward input from simple cells and all of them become active. Thus, they lose the desired behaviour, because the feedback from V4 is not specific to the object that actually caused that particular V4 cell's activation.

### 1.2.2 Hypothesised Solution in a Spiking Network Model

As described in the previous section, it is impossible in a rate coded model to determine which neurons were activated by the same object or features and which were activated by a different object or feature, a phenomenon known as 'superposition catastrophe' (Von Der Malsburg, 1999). Concretely, as seen in figure 1.1b for an example stimulus, it is not possible to distinguish coactivation of neurons ( $\Phi_{loc1}^{simple}$  and  $\Phi_{BR}^{V4}$  in the example) from causal drive ( $\Phi_{loc1}^{simple}$  drives  $\Phi_{BL}^{V4}$  in the example). But in Eguchi et al. (2018) a possible solution to this problem is developed, that relies on incorporating the exact timing of spikes into the model.

In the real biological brain, neurons emit action potentials that are discrete events localised in time. These spikes then propagate along synapses, which have different delays resulting in the action potentials reaching a postsynaptic cell with some time delay, that varies across synapses. When a spike reaches a postsynaptic cell, it's membrane is slightly depolarised. However, the potential of the depolarised membrane then quickly decays back to it's resting point. But because a cell only spikes if the depolarisation is big enough, *multiple* spikes have to converge on it within a short temporal window to activate it, before the membrane potential can decay again. Such a cell functions as a *coincidence detector*, i.e. detecting that the presynaptic cells emitted action potentials in a temporal succession that reflects the synaptic delays (Eguchi et al., 2018).

With this mechanism Eguchi et al. (2018) postulate the so called 'three neuron binding circuit' (see figure 1.3). It consists of a neuron  $L$  representing a lower level feature, which participates in driving a higher layer neuron  $H$ , that encodes a more abstract feature or object. Both of these neurons then have synapses to a so called 'binding neuron'  $B$  (Eguchi



**Figure 1.2:** In a spiking network with appropriate delays the superposition catastrophe is avoided. Different colours symbolise activation with different temporal signatures. The synaptic delays are such that a feedback spike from a V4 neuron and a feedforward spike from a simple cell only converge on the border ownership cell at roughly the same time, if the simple cell actually drove the V4 cell (see text for mathematical detail).

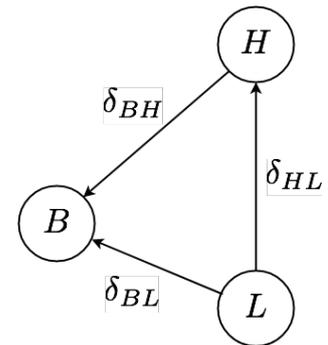
et al., 2018). The synaptic delays satisfy the following equation:

$$\delta_{HL} + \delta_{BH} \approx \delta_{BL} \quad (1.1)$$

where  $\delta_{yx}$  is the synaptic delay for the synapse from neuron  $x$  to neuron  $y$  (see figure 1.3 for illustration). Further, they assume that neuron  $B$  is activated only if two spikes converge on it at the same time. If cell  $L$  is driving cell  $H$ , then a spike emitted by  $L$  at time  $t$  will be followed by a spike emitted by  $H$  at roughly  $t + \delta_{HL}$ , when the spike from  $L$  reaches  $H$ . The spike consequently emitted by  $H$  will then travel along the synapse to neuron  $B$ , reaching it at time  $t + \delta_{HL} + \delta_{BH}$ . Simultaneously, the original spike of neuron  $L$  at time  $t$  also reaches neuron  $B$  along the synapse from  $L$  to  $B$ , since  $t + \delta_{BL} \approx t + \delta_{HL} + \delta_{BH}$  (equation 1.1). Hence, two spikes converge on neuron  $B$  in short succession and elicit a spike. The cell is therefore only active if neuron  $L$  drives neuron  $H$  and not if  $H$  and  $L$  just happen to be co-active.

With spiking neurons, the superposition catastrophe observed in (Eguchi and Stringer, 2016) can thus be overcome, if the synapses have the right kind of delay structure (see figure 1.2). For our example,  $\Phi_{BR,loc1}^{V1/V2}$  cells would not be erroneously active anymore. Since  $\delta_6 \approx \delta_2 + \delta_4$  holds, cells  $\Phi_{loc1}^{simple}$  and  $\Phi_{BL}^{V4}$  therefore have to spike with a time difference equal to  $\delta_2$  to cause activation in  $\Phi_{BR,loc1}^{V1/V2}$ . Only then would their spikes arrive at this neuron at the same time, causing false activation in it. But since  $\Phi_{loc1}^{simple}$  does not participate in driving  $\Phi_{BL}^{V4}$ , it is highly unlikely that they spike with this temporal delay. Therefore,  $\Phi_{BR,loc1}^{V1/V2}$  will not receive the necessary two spikes at roughly the same time and it will not spike.

$\Phi_{BL,loc1}^{V1/V2}$  cells on the other hand, are still correctly active. Because  $\Phi_{loc1}^{simple}$  cells actually drive the  $\Phi_{BL}^{V4}$  neurons, the later cells spike after the previous cells, with a temporal offset equal to the delay of the synapse between the two ( $\delta_1$ ). The freshly emitted spike from  $\Phi_{BL}^{V4}$  then reaches the binding neuron  $\Phi_{BL,loc1}^{V1/V2}$  with a further delay of  $\delta_3$ . At the same time, the original spike from  $\Phi_{loc1}^{simple}$  reaches the binding neuron, since the delay of this synapse is  $\delta_5 \approx \delta_1 + \delta_3$ . As a result the border ownership or binding neuron  $\Phi_{BL,loc1}^{V1/V2}$  will be activated



**Figure 1.3: Three Neuron Binding Circuit** proposed by Eguchi et al. (2018)

by these two spikes arriving simultaneously.

Perhaps most importantly, Eguchi et al. (2018) further argue, that in a network containing multiple connections with varying synaptic delays between each pair of neurons, these kinds of binding circuits can be formed by selectively strengthening synapses with appropriate delays. This relies on spike time dependent plasticity (STDP), a learning mechanism found in the biological brain (Markram et al., 1997; Bi and Poo, 1998), which potentiates a synapse if the postsynaptic neuron spikes shortly after the presynaptic cell and depresses the synapse if the order of spikes is inverted. In this simulation study we investigate whether this mechanism enables a network to develop border ownership neurons, that rely on the type of circuit depicted in 1.2. We can thus formulate the first set of hypotheses:

$H_1$ : In a spiking neural network border ownership cells develop through STDP.

$H_2$ : The border ownership cells are modulated through feedback from V4 contour element cells.

$H_3$ : Exact spike times play a crucial role, as the binding cells act as coincidence detectors (see figure 1.2).

Furthermore, we will present a few points of criticism on the original rate coded border ownership study by Eguchi and Stringer (2016).

### 1.3 Polychronisation

The mechanism just outlined in the previous section 1.2.2 relies on sets of neurons spiking in a well defined temporal sequence that depends on the presented stimulus. Such a spatiotemporal pattern of spiking activity is referred to as a ‘polychronous group’ (Izhikevich, 2006). It has been argued, that these polychronous groups have a vastly higher information capacity than the firing rates of neurons (Izhikevich, 2006; Eguchi et al., 2018).

Previously, polychronous groups have been defined only in terms of trigger events, i.e. the set of neurons with associated spike times that start a spatiotemporal pattern of activity in the network (Martinez and Paugam-Moisy, 2009). Here, however, we include all spikes occurring within this entire chain of activity in the definition of a polychronous group:

**Definition 1.** A **polychronous group**  $PG = \{(N_1, t_1), (N_2, t_2), \dots, (N_k, t_k)\}$  is a set of  $K$  spikes each of which is given as a tuple  $(N_i, t_i)$  of a neuron and a timepoint at which this neuron spikes. Note that adding a constant value to all spike times  $t_i$  does not change the polychronous groups, as only relative spike time differences are relevant. A polychronous group is said to be activated, if the involved neurons spike with these relative time differences.

Eguchi et al. (2018) report the emergence of such polychronous groups in a spiking network very similar to the one presented in this study. Further, they report that these spatiotemporal patterns of activity carry information about the presented stimulus, i.e. a

certain polychronous group is activated repeatedly and reliably if a particular object is present in the stimulus and it is not activated if that object is not present. In this study we aim to replicate these findings in a spiking neural network model of border ownership representation, thus motivating the fourth hypothesis:

$H_4$ : Spike times exhibit reliable temporal patterns that carry information about presented objects.

Additionally, we will provide some points of critique on the analysis techniques employed in Eguchi et al. (2018).

## 1.4 *Continuous Transform and Trace Learning Mechanisms*

During training, we present various objects to the network that all have a vertical straight border on either their left or right hand side that is precisely aligned with one of the two target locations. However, the response of both a V1/V2 border ownership cell and a V4 contour element cell should only be selective to a certain type of contour element (a vertical straight edge) occurring in a certain place relative to the objects centre of mass (left or right of it depending on the cells tuning properties). The cell should be invariant to the shape of the object's remaining contour and to the objects colour (Pasupathy and Connor, 2001, 2002; Zhou et al., 2000).

Further, V4 contour element cells have been reported to have location invariance in a small region of the retina (Pasupathy and Connor, 2001). Thus, the V4 contour element cells in our simulations are additionally required to respond invariantly whether the object is located at location 1 or at location 2, because we suppose that the simulated region of the retina is completely contained within a V4 neuron's receptive field.

Two biologically plausible learning mechanisms have been proposed to achieve these two kinds of invariances: *Continuous Transformation Learning* (Stringer et al., 2006) and *Temporal Trace Learning* (Földiák, 1991).

**Continuous Transformation Learning** (CT) on the one hand, relies on spatial continuity between continuous transforms of a visual stimulus. Visually similar stimuli are likely to show the same object and the CT mechanisms uses this fact by binding them together through a purely associative (Hebbian) learning rule (Stringer et al., 2006). The presentation of an initial transform will activate an arbitrary set of neurons and the synapses involved will be strengthened through the Hebbian learning rule (Hebb, 1949). Presenting a further visually similar transform of the object will then result in the same output neurons to become active, given that there is enough overlap in the set of active input neurons. Consequently, the synapses from the overlapping but slightly different set of active input neurons for the new stimulus will be strengthened. Thus, the output cells will be as active for the newly presented transform as they are for the originally presented one. In this way, all transforms of an object will be mapped onto the same output neurons, as long as there is enough visual similarity between the transforms (Stringer et al., 2006). The CT

mechanism has also been shown to achieve good results in spiking neural network models (Evans and Stringer, 2012) and we will therefore rely on it for the initially mentioned invariance to an object's overall contour shape and its colour.

**Temporal Trace Learning** on the other hand, relies on temporal continuity of objects in the real world. The proposition is that different images, that are projected onto the retina in short temporal succession, are likely to depict the same object. In this learning paradigm, the temporal trace of a neurons previous activity is incorporated into a Hebbian learning rule (Földiák, 1991; Wallis and Rolls, 1997). As with the CT mechanism, an initial stimulus presentation activates an arbitrary set of output neurons. When presenting a further, possibly completely different stimulus shortly after, the neurons still remember their previous activity in terms of this temporal trace and are biased by it to respond more easily. Thus, cells active during the previous stimulus presentation are likely to be activated by the next stimulus presentation, which results in Hebbian strengthening of the synapses involved in the new presentation. Using this mechanism, arbitrarily dissimilar stimuli are mapped onto the same output neurons, as long as they are presented close together in time. Evans and Stringer (2012) have shown this mechanism to also work in spiking neural networks, if synapses bleed current into the postsynaptic neuron for a relatively long period of time, thus biasing the cell to spike more easily upon presentation of the next stimulus. In the following simulations we will therefore rely on a rather large synaptic time constant to achieve a temporal trace like learning effect, which we believe to facilitate the development of the mentioned location invariance in V4 neurons. In summary, the last hypothesis we investigate can be formulated as follows:

*H*<sub>5</sub>: Invariance to colour and overall shape relies on the CT mechanism while location invariance is achieved through the temporal trace mechanism.

## Chapter 2

# Methods

### 2.1 Spiking Neural Network Model

The spiking neural network model explored in this study is, except for the choice of some parameters, identical to the model used in Eguchi et al. (2018).

#### 2.1.1 Network Architecture

The network consists of four neural layers that simulate successive stages along the ventral visual pathway from visual area V2 to area V4. Each layer consists of  $64 \times 64 = 4096$  excitatory pyramidal neurons and  $32 \times 32 = 1024$  inhibitory neurons. The first layer receives input from a set of input neurons that are described later. There are bottom-up (feedforward) and top-down (feedback) synapses between successive layers of excitatory neurons as well as lateral connections between excitatory cells within a layer. All of these synapses are modifiable and may change their weights during training through spike time dependent plasticity (described later). Additionally, there are constant, lateral connections from excitatory cells within a layer to inhibitory cells in the same layer as well as connections from the inhibitory cells back to the excitatory ones. Each cell has a fixed number of afferent synapses of each type, given in table 2.1a. For each postsynaptic cell and each connection type, the appropriate number of presynaptic cells is drawn with replacement from a two dimensional gaussian distribution that is centred at the position corresponding to the postsynaptic neuron in the previous, same or following layer for feedforward, lateral or feedback connections respectively. This may result in multiple connections between a pair of cells as well as different numbers of efferent synapses for different cells. The fan-in-radius (standard deviation) of the gaussians for each synapse type are given in table 2.1a.

#### 2.1.2 Differential Equations

Neuron and synapse constants were chosen to be as biologically realistic as possible based upon the available neurophysiological literature (see table 2.1 for a full list).

**Table 2.1:** Model parameters. Most model parameters are taken from Eguchi et al. (2018), who obtained most integrate and fire parameters from Troyer et al. (1998) (indicated by §) and most plasticity parameters from Perrinet et al. (2001) (denoted by †). Parameters marked with ● were tuned for the simulations reported in this study.

<b>(a) Network parameters</b>				
Layer	1	2	3	4
Number of excitatory neurons within each layer	64 × 64	64 × 64	64 × 64	64 × 64
Number of inhibitory neurons within each layer	32 × 32	32 × 32	32 × 32	32 × 32
Number of feedforward afferent excitatory connections per excitatory neuron ( $E/E$ )	270	180	180	180
Fan-in radius for feedforward afferent excitatory connections to each excitatory neuron ( $E/E$ )	10	50	50	50
Number of feedback afferent excitatory connections per excitatory neuron ( $E/E$ )	5	5	5	-
Fan-in radius for feedback afferent excitatory connections to each excitatory neuron ( $E/E$ )	8	8	8	-
Number of lateral afferent excitatory connections per excitatory neuron ( $E/E$ )	30	30	30	30
Fan-in radius for lateral afferent excitatory connections to each excitatory neuron ( $E/E$ )	14	14	14	14
Number of lateral afferent excitatory connections per inhibitory neuron ( $E/I$ )	60	60	60	60
Fan-in radius for lateral afferent excitatory connections to each inhibitory neuron ( $E/I$ )	8	8	8	8
Number of lateral afferent inhibitory connections per excitatory neuron ( $I/E$ )	60	60	60	60
Fan-in radius for lateral afferent inhibitory connections to each excitatory neuron ( $I/E$ )	4	4	4	4
<b>(b) Parameters for Gabor Filtering of visual images</b>				
Phase shift ( $\phi$ )	0, $\pi$			
Wavelength ( $\lambda$ )	2			
Orientation ( $\theta$ )	0, $\pi/4$ , $\pi/2$ , $3\pi/4$			
Spatial bandwidth ( $b$ )	1.5 octaves			
Aspect ratio ( $c$ )	0.5			
<b>(c) Cellular Parameters</b>				
Excitatory cell somatic capacitance ( $C_m^E$ )	500 pF			
Inhibitory cell somatic capacitance ( $C_m^I$ )	214 pF			
Excitatory cell somatic leakage conductance ( $g_L^E$ )	25 nS			
Inhibitory cell somatic leakage conductance ( $g_L^I$ )	18 nS			
Excitatory cell resting potential ( $V_R^E$ )	-74 mV			
Inhibitory cell resting potential ( $V_R^I$ )	-82 mV			
Excitatory firing threshold potential ( $\Theta^E$ )	-53 mV			
Inhibitory firing threshold potential ( $\Theta^I$ )	-53 mV			
Absolute refractory period ( $\tau_R$ )	2 ms			
<b>(d) Synaptic Parameters</b>				
Range of axonal transmission delays	[1, 10] ms			
Synaptic neurotransmitter concentration ( $a_C$ )	1.0			
Proportion of unblocked NMDA receptors ( $a_D$ )	0.5			
Presynaptic STDP time constant ( $\tau_C$ )	5 ms			
Postsynaptic STDP time constant ( $\tau_D$ )	5 ms			
Synaptic learning rate ( $\rho$ )	increasing up to $4.1 \cdot 10^{-5}$			
Synaptic conductance scaling factor for feedforward excitatory connections from Gabor filters to Layer 1 excitatory cells ( $\Delta g^{E/E}$ )	●			
Synaptic conductance scaling factor for feedforward excitatory connections to excitatory cells in layers 2, 3 or 4 ( $\Delta g^{E/E}$ )	●			
Synaptic conductance scaling factor for feedback excitatory connections to excitatory cells in layers 1, 2 or 3 ( $\Delta g^{E/E}$ )	●			
Synaptic conductance scaling factor for lateral excitatory connections to excitatory cells in layers 1, 2, 3 or 4 ( $\Delta g^{E/E}$ )	●			
Synaptic conductance scaling factor for lateral connections from excitatory cells to inhibitory cells in layers 1, 2, 3 or 4 ( $\Delta g^{E/I}$ )	●			
Synaptic conductance scaling factor for lateral connections from inhibitory cells to excitatory cells in layers 1, 2, 3 or 4 ( $\Delta g^{I/E}$ )	●			
Excitatory reversal potential ( $V^E$ )	0 mV			
Inhibitory reversal potential ( $V^I$ )	-70 mV			
Synaptic time constant for all feedforward, feedback and lateral connections from Gabor filters and excitatory cells to excitatory cells ( $\tau_{E/E}$ )	150 ms			
Synaptic time constant for lateral connections from excitatory cells to inhibitory cells ( $\tau_{E/I}$ )	2 ms			
Synaptic time constant for lateral connections from inhibitory cells to excitatory cells ( $\tau_{I/E}$ )	5 ms			
<b>(e) Parameters for numerical simulation by Forward Euler timestepping scheme</b>				
Numerical step size ( $\Delta t$ )	0.2 ms			

### Cell equations

Each cell is modelled as a standard conductance-based leaky integrate and fire neuron. The membrane potential is updated according to equation (2.1).

$$\tau_m^\gamma \frac{dV_i(t)}{dt} = V_0^\gamma - V_i(t) + R^\gamma I_i(t) \quad (2.1)$$

where  $V_i(t)$  is a given cell's (indexed by  $i$ ) membrane potential at time  $t$ . All parameters in equation (2.1) are specific to the cell class (excitatory or inhibitory), which is denoted by  $\gamma$ . Without any input, the cell's membrane potential decays to its (class dependent) resting potential  $V_0^\gamma$ , the speed of which is determined by the membrane time constant  $\tau_m^\gamma$ . This is defined as  $\tau_m^\gamma = C_m^\gamma / g_0^\gamma$ , where  $C_m^\gamma$  is the membrane capacitance and  $g_0^\gamma$  is the membrane leakage conductance. The cell is further driven by the combined current of excitatory and inhibitory afferent synapses denoted by  $I_i(t)$ .  $R^\gamma$  is the membrane resistance, given by  $R^\gamma = 1/g_0^\gamma$ . Whenever the cell's membrane potential exceeds the firing threshold potential  $\Theta^\gamma$ , a spike is recorded and the membrane potential is explicitly set to the resting potential  $V_0^\gamma$ . After each spike, the cell's membrane potential is kept at the resting potential for the duration of the refractory period  $\tau_R$ .

The current injected into neuron  $i$  by a synapse from neuron  $j$  is equal to the difference between the membrane potential of the postsynaptic cell  $i$  and the synapse's reversal potential multiplied with the conductance of the synapse:  $g_{ij}(t)(\hat{V}^\gamma - V_i(t))$ .  $g_{ij}(t)$  is the conductance of the synapse from neuron  $j$  to neuron  $i$  at time  $t$ , which is driven up by a presynaptic spike and decays to zero otherwise.  $\hat{V}^\gamma$  is the reversal potential of the synapse class. Excitatory synapses, on the one hand, have a reversal potential above the firing threshold of a cell, which results in positive current (since  $\hat{V}^\gamma - V_i(t) > 0$ ) and depolarisation of the cell. Inhibitory synapses, on the other hand, have a reversal potential below or equal to the cell's resting membrane potential, which results in a negative current (since  $\hat{V}^\gamma - V_i(t) < 0$ ) and hyperpolarisation of the cell. The total input current that cell  $i$  receives is the sum of the currents of all synapses of both types and is given by equation (2.2).

$$I_i(t) = \sum_{\gamma \in \{E, I\}} \sum_j g_{ij}(t)(\hat{V}^\gamma - V_i(t)) \quad (2.2)$$

### Synaptic conductance equations

The time course of the conductance of a synapse from neuron  $j$  to neuron  $i$  is governed by the following differential equation:

$$\frac{dg_{ij}(t)}{dt} = -\frac{g_{ij}(t)}{\tau_g} + \lambda \Delta g_{ij}(t) \sum_l \delta(t - \Delta t_{ij} - t_j^l) \quad (2.3)$$

Without presynaptic spikes, the conductance  $g_{ij}(t)$  of the synapse decays to zero, the speed of which is determined by the decay term  $\tau_g$ . Whenever there is a presynaptic spike at time  $t_j^l$  ( $l$  indexes over all presynaptic spikes), the synapses's conductance is increased by the current weight of the synapse  $\Delta g_{ij}(t)$ . This instantaneous increase in conductance is

modelled by a Dirac delta function (equation (2.4)) and it takes place with some temporal delay  $\Delta t_{ij}$  (axonal transmission delay). In this study, each synapse has a delay randomly chosen from the range between 1 ms and 10 ms. Further, a biological scaling constant  $\lambda$  has been introduced to scale the synaptic efficiency  $\Delta g_{ij}$ , which is bounded between zero and one.

$$\delta(x) = \begin{cases} \infty & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{where, } \int_{-\infty}^{\infty} \delta(x) dx = 1 \quad (2.4)$$

### Synaptic learning equations

Except for the lateral synapses to and from inhibitory neurons, the weights ( $\Delta g_{ij}(t)$ ) of all synapses are modifiable. These weights change according to spike time dependent plasticity (STDP), which is governed by differential equations 2.5, 2.6 and 2.7. The core idea of STDP is that a synapse is potentiated (LTP) whenever a postsynaptic spike follows shortly after a presynaptic spike (which implies causality between the two), whereas it is depressed (LTD) whenever a presynaptic spike follows shortly after a postsynaptic one (since this excludes causality). To this end, the recent presynaptic activity ( $C_{ij}(t)$ ) as well as the recent postsynaptic activity ( $D_i(t)$ ) are maintained for a synapse from neuron  $j$  to neuron  $i$ . Both are bounded between zero and one.

The recent presynaptic activity ( $C_{ij}(t)$ ) may be interpreted as the concentration of neurotransmitter (glutamate) released into the synaptic cleft (Perrinet et al., 2001) and it is modelled by equation (2.5).

$$\frac{dC_{ij}(t)}{dt} = -\frac{C_{ij}(t)}{\tau_C} + \alpha_C(1 - C_{ij}(t)) \sum_l \delta(t - \Delta t_{ij} - t_j^l) \quad (2.5)$$

$C_{ij}(t)$  decays to zero governed by the decay term  $\tau_C$  and it is driven up according to the model parameter  $\alpha_C$  whenever a presynaptic spike *arrives* at the postsynaptic cell. A presynaptic spike at time  $t_j^l$  arrives at the postsynaptic cell with a delay equal to the synaptic transmission delay  $\Delta t_{ij}$  and  $C_{ij}(t)$  is increased at that point.

The recent postsynaptic activity ( $D_i(t)$ ) may be interpreted as the proportion of NMDA receptors unblocked by recent depolarisation from back-propagated action potentials (Perrinet et al., 2001). It is modelled by equation (2.6).

$$\frac{dD_i(t)}{dt} = -\frac{D_i(t)}{\tau_D} + \alpha_D(1 - D_i(t)) \sum_k \delta(t - t_i^k) \quad (2.6)$$

$D_i(t)$  also decays to zero governed by the decay term  $\tau_D$  and it is driven up by postsynaptic spikes according to the model parameter  $\alpha_D$ . The increase happens at the exact time  $t_i^k$  when the postsynaptic cell spikes ( $k$  indexes over all postsynaptic spikes), since the receiving synapse is assumed to be arbitrarily close to the postsynaptic cell body. For this reason there is no synaptic transmission delay included in equation (2.6).

The strength of the synaptic weight ( $\Delta g_{ij}(t)$ ) is thus modified according to equation (2.7).

$$\tau_{\Delta_g} \frac{d\Delta g_{ij}(t)}{dt} = (1 - \Delta g_{ij}(t)) C_{ij}(t) \sum_k \delta(t - t_i^k) - \Delta g_{ij}(t) D_i(t) \sum_l \delta(t - \Delta t_{ij} - t_j^l) \quad (2.7)$$

where  $t_i^k$  indexes over all postsynaptic spikes,  $t_j^l$  indexes over all presynaptic spikes and  $\tau_{\Delta_g}$  is a time course variable. Whenever the postsynaptic cell spikes at time  $t_i^k$ , the synaptic weight  $\Delta g_{ij}(t)$  is increased (LTP) the strength of which is determined by the recent presynaptic activity  $C_{ij}(t)$ . That is, the more recent there was a presynaptic spike, the higher is the presynaptic activity  $C_{ij}(t)$  and the stronger is the increase in synaptic weight. On the other hand, whenever a presynaptic spike *arrives* at the synapse, the synaptic weight  $\Delta g_{ij}(t)$  is decreased (LTD), the strength of which is determined by the recent postsynaptic activity  $D_i(t)$ . That is, the more recent there was a postsynaptic spike, the higher is the postsynaptic activity  $D_i(t)$  and the stronger is the decrease in synaptic weight. Note that a weight update takes place whenever there is a pre- or postsynaptic spike. For example, if there is a presynaptic spike immediately followed by a postsynaptic spike then the weight is first decreased slightly (since  $D_i(t)$  is still small) when the presynaptic spike arrives at the synapse and then increased strongly (since  $C_{ij}(t)$  is high from the just arrived presynaptic spike) when the postsynaptic cell spikes shortly after.

### 2.1.3 Numerical Scheme and Implementation

All of the just described differential equations are simulated using the Forward-Euler numerical method with a time step  $\Delta t = 0.2ms$ . The Dirac delta function is discretely approximated as defined in Amit and Brunel (1997). Furthermore, the change in synaptic weight (equation (2.7)) is modelled to be instantaneous with each pre- or postsynaptic spike. Thus,  $\frac{\Delta t}{\tau_{\Delta_g}}$  is replaced by a newly introduced learning rate  $\rho$  in the corresponding finite difference equation.

The model was implemented by James Isbister, Nasir Ahmad and Akihiro Eguchi and it is available at <https://github.com/OFTNAI/Spike>.

### 2.1.4 Input Representation of Visual Stimuli

The first layer of simulated excitatory leaky integrate and fire neurons receives feedforward connections from a set of input cells that resemble the tuning profiles of simple cells in V1. Simple cells respond to local oriented bars and edges within a small receptive field and Gabor filters have been found to provide a good fit for the tuning properties of these cells (Jones and Palmer, 1987; Cumming and Parker, 1999).

The input Gabor filters used in this study are computed by the following equations:

$$g(x, y, \lambda, \theta, \psi, b, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right) \quad (2.8)$$

with the following definitions:

$$\begin{aligned}
x' &= x \cos \theta + y \sin \theta \\
y' &= -x \sin \theta + y \cos \theta \\
\sigma &= \frac{\lambda(2^b+1)}{\pi(2^b-1)} \sqrt{\frac{\ln 2}{2}}
\end{aligned} \tag{2.9}$$

where  $x$  and  $y$  specify the position of a light impulse in the visual field (Petkov and Kruizinga, 1997). The parameter  $\lambda$  is the wavelength ( $1/\lambda$  is the spatial frequency),  $\sigma$  controls number of such periods inside the Gaussian window based on  $\lambda$  and spatial bandwidth  $b$ ,  $\theta$  defines the orientation of the feature,  $\psi$  defines the phase, and  $\gamma$  sets the aspect ratio that determines the shape of the receptive field (Eguchi and Stringer, 2016). An array of eight such Gabor filters is generated at each of  $128 \times 128$  retinal locations with different values for the parameters  $\theta$  and  $\psi$  (see table 2.1b).

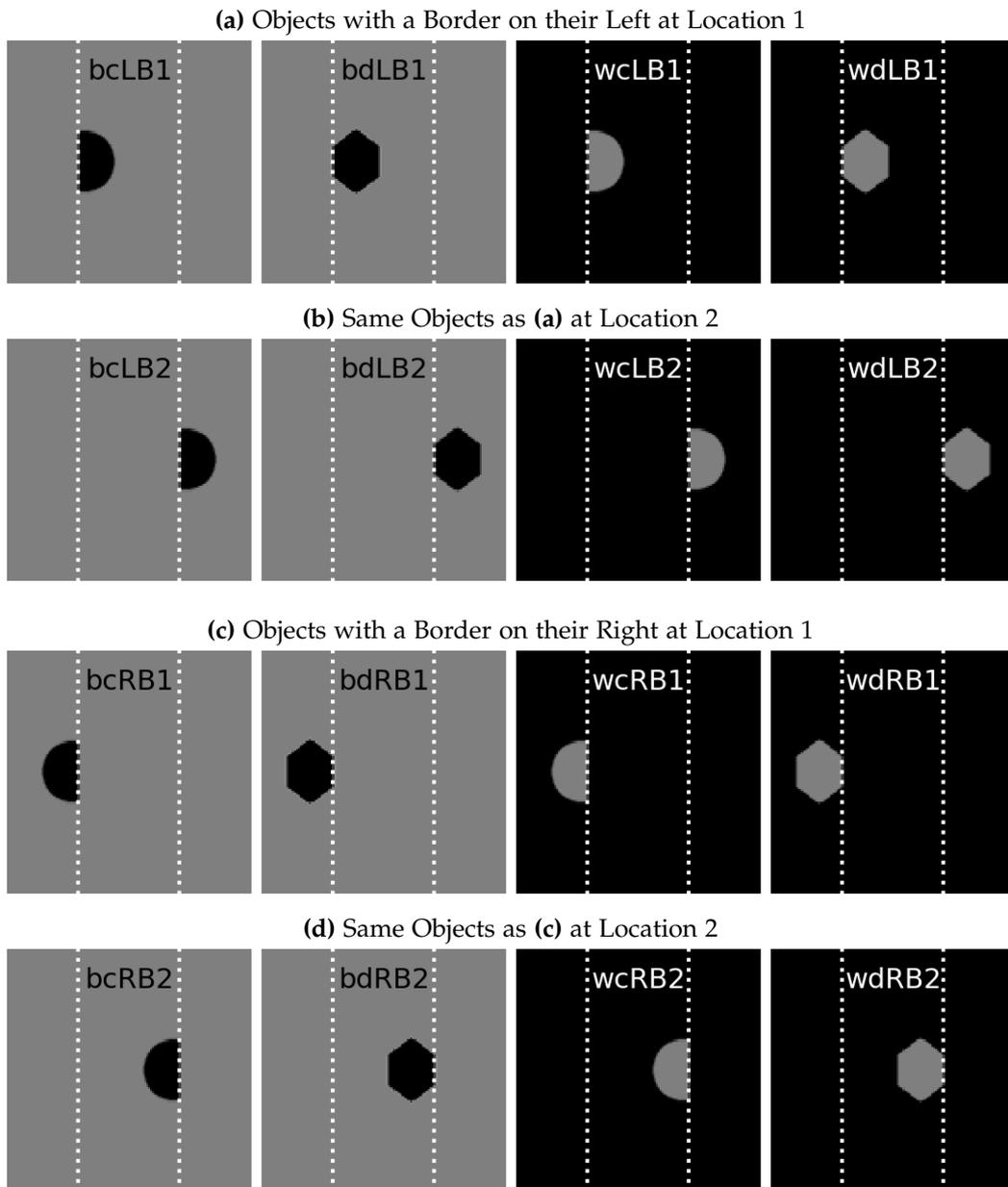
When a stimulus is presented, each of these  $8 \times 128 \times 128$  Gabor filters has a unique activation value and the firing rate of the corresponding input simple cell is set accordingly, while maintaining a constant average firing rate of all input cells at 0.1 Hz. For each input cell, a Poisson spike train is then generated that exhibits this firing rate.

Each excitatory layer 1 cell has 270 afferent synapses from each of the eight layers of input cells (for the eight Gabor filter parameterisations). Similarly to all other synapses, the locations of the presynaptic input cells within the corresponding Gabor filter layer are drawn from a Gaussian distribution centred at the location of the postsynaptic layer 1 neuron.

## 2.2 Stimulus Sets

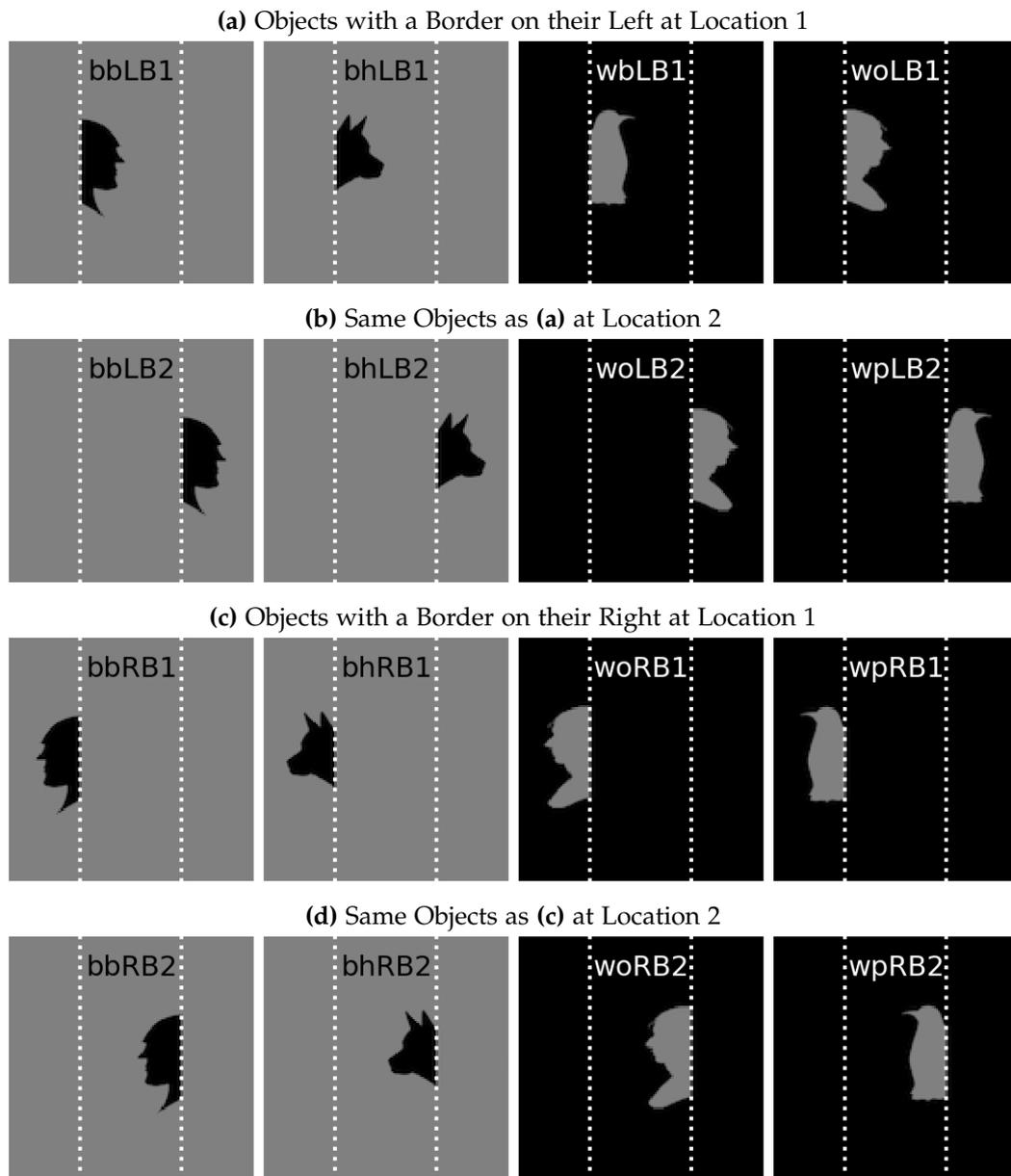
### Training Set

The network is trained and initially tested on the set of training stimuli (figure 2.1). There are eight different objects shown in this stimulus set (figures 2.1a and 2.1c) and each object is displayed in both location 1 (left dotted line in figure 2.1) and location 2 (right dotted line in figure 2.1). In the normal training regime, stimuli with the same object at location 1 and location 2 are switched in short succession, in order to simulate visual saccades. For example, when the object "black half circle with a left border" (bcLB) is presented, then the two stimuli **bcLB1** and **bcLB2** are shown in short succession.



**Figure 2.1:** The set of stimuli that the network is trained and tested on. There are eight objects depicted in this stimulus set and each is presented at two distinct retinal locations (dotted lines). (a & b) the four objects with a vertical straight border on their left hand side at the two retinal locations. (c & d) the four objects with a vertical straight border on their right hand side at the two retinal locations.

## Testing Set



**Figure 2.2:** Additional set of testing stimuli that the network is tested on. The network parameters are always fixed while these stimuli are presented. (a & b) four novel objects with a vertical straight border on their left hand side at the two retinal locations. (c & d) four novel objects with a vertical straight border on their right hand side at the two retinal locations.

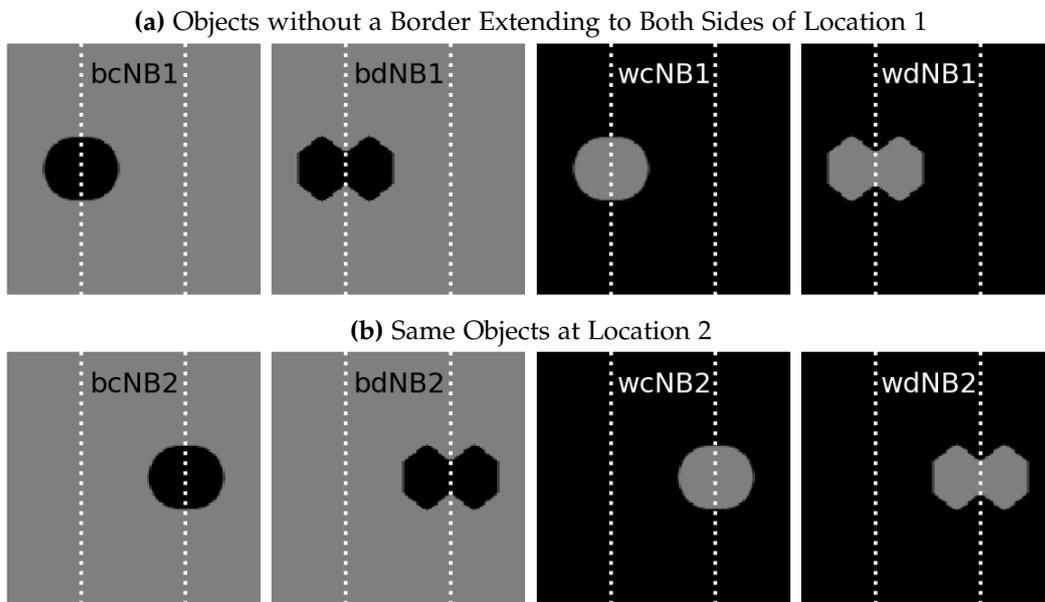
**Set of Stimuli without a Border at a Target Location**

Figure 2.3: Set of stimuli that do not contain a border at any of the target locations.

### Set of Stimuli With Multiple Objects

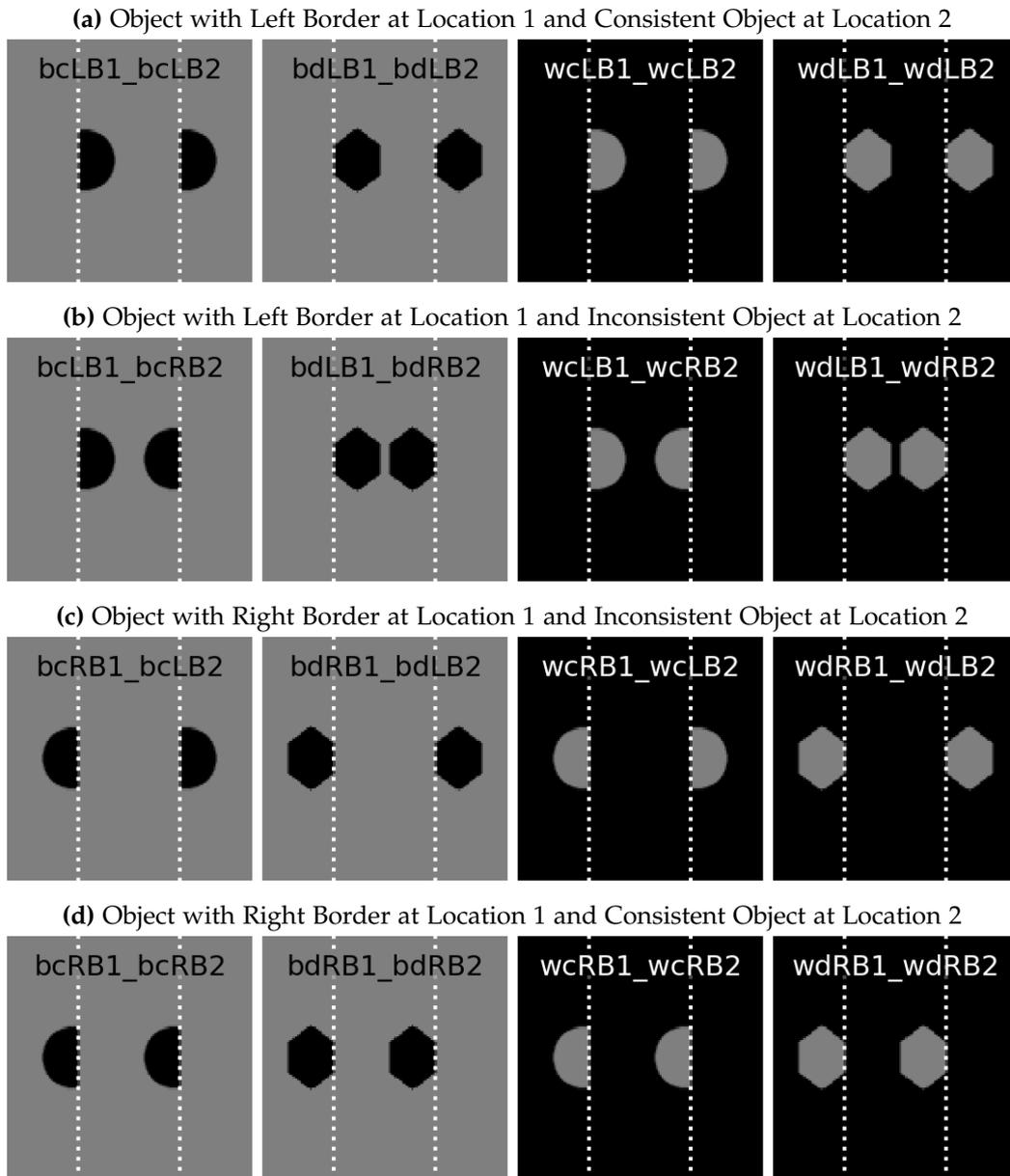


Figure 2.4: Set of stimuli showing two of the objects from the training set (figure 2.1) at the same time.

## 2.3 Detecting Oscillations of Neuron Populations

### Frequency of Oscillation

An analysis of the frequency and reliability of the prevalent oscillations within a population was carried out where a population is defined as all neurons of a certain type (excitatory or inhibitory) in one layer. First, all spikes emitted in the time from 0.75 s to 2.0 s after stimulus onset by any cell in the population are counted in bins of 4 ms

length. Thus, a time course of each population's instantaneous firing rate with a resolution of 4 ms is obtained separately for each stimulus presentation. On this time course a Hanning window is applied to mitigate border discontinuities and a discrete Fourier transform is computed. The result is a frequency spectrum for each stimulus presentation, the maximum of which is considered to be the frequency of the oscillation prevalent in the population for this stimulus presentation. This frequency estimate might differ slightly for different presentations of the same stimulus due to slight variations in the population's activity. The most commonly obtained frequency for 11 presentations of a stimulus is then considered to be the characteristic response frequency for that stimulus.

### Peaks of Activity

Peaks of a population's activity are detected by finding local maxima of the smoothed instantaneous firing rate of the population. The population's instantaneous firing rate is first smoothed by convolving it with a gaussian ( $\sigma = 10ms$ ). Local extrema on this smoothed activity are then considered to be activity peaks if the local extrema's instantaneous firing rate value is at least 30% as high as the maximal instantaneous firing rate for the whole time course. The time points of the population's peak activity obtained in this way are then used to estimate the regularity of the oscillations by computing the variance of the time between two such peaks and as temporal fix points for the polychronous group analysis carried out in section 3.4.2.

## 2.4 Performance Measures

We use concepts from information theory and machine learning to assess the performance of the network. First of all, we outline two measures for investigating how much information a single cell's firing rate carries about the presented stimulus type (sections 2.4.1 and 2.4.2). A cell carries information in its firing rate if it has a certain firing rate response for one particular type of stimulus and a different firing rate response for all other stimulus types. We use the firing rate during the last 0.5 seconds of a 2 seconds stimulus presentation as an estimate of the cell's steady state firing rate in response to the presented stimulus.

Secondly, we describe a metric for the information carried by the occurrence probability of spike pairs in response to the presented stimulus type (section 2.4.3).

### 2.4.1 Single Cell Information

In previous studies, *single cell information* was used to assess how reliable a cell responds with a different firing rate to its preferred stimulus (Rolls and Milward, 2000; Eguchi and Stringer, 2016; Eguchi et al., 2018). Given, that all stimuli are grouped into disjunct *stimulus types (categories)*, denoted by  $s$ , a metric for the informativeness of a cell's firing rate with respect to each of the stimulus types can be obtained. For example, 4 single cell

information values for each of the 4 types of border ownership relations (LB1, LB2, RB1, RB2, see figure 2.1) can be computed.

The single cell information is computed by firstly converting each scalar firing rate of a cell in response to each stimulus presentation to a discrete value. In this study three values are possible: low, medium or high firing rate. In order to obtain this value for a given cell, the range from the minimal to the maximal scalar firing rate in response to all presentations of all stimuli is divided into three equispaced bins. As a result, the cell has one of three discrete response values,  $r \in \{1,2,3\}$ , for each stimulus presentation. Next, stimulus presentations that show stimuli of the same type are combined to obtain the probability distribution  $P(R = r|S = s)$ . It provides the probability for the discrete response value  $r$ , given that the type of the presented stimulus is  $s$ . The single cell information that a particular cell carries about stimulus type  $s$  (e.g. left border at location 1) is defined by equation (2.10).

$$I(s) = \sum_{r \in \{1,2,3\}} P(R = r|S = s) \cdot \log_2 \left( \frac{P(R = r|S = s)}{P(R = r)} \right) \quad (2.10)$$

The following provides an intuitive understanding of this measure. Let  $\mathcal{M}(R) := P(R)$  be the marginal probability distribution over responses without knowing the stimulus category and  $\mathcal{P}_s(R) := P(R|S = s)$  be the posterior probability distribution over responses if we know that the stimulus belongs to the stimulus category  $s$ .

$$\begin{aligned} I(s) &= \sum_{r \in \{1,2,3\}} P(R = r|S = s) \cdot \log_2 \left( \frac{P(R = r|S = s)}{P(R = r)} \right) \\ &= \sum_r P(R = r|S = s) \cdot (\log_2 P(R = r|S = s) - \log_2 P(R = r)) \\ &= \sum_r \mathcal{P}_s(R = r) \cdot (\log_2 \mathcal{P}_s(R = r) - \log_2 \mathcal{M}(R = r)) \\ &= \sum_r \mathcal{P}_s(R = r) \cdot \log_2 \mathcal{P}_s(R = r) - \sum_r \mathcal{P}_s(R = r) \cdot \log_2 (\mathcal{M}(R = r)) \\ &= \left[ - \sum_r \mathcal{P}_s(R = r) \cdot \log_2 \mathcal{M}(R = r) \right] - \left[ - \sum_r \mathcal{P}_s(R = r) \cdot \log_2 \mathcal{P}_s(R = r) \right] \\ &= H_{\mathcal{M}}(\mathcal{P}_s) - H(\mathcal{P}_s) = KL(\mathcal{P}_s \parallel \mathcal{M}) \end{aligned} \quad (2.11)$$

where  $H_{\mathcal{M}}(\mathcal{P}_s)$  is the cross-entropy between  $\mathcal{M}$  and  $\mathcal{P}_s$ ,  $H(\mathcal{P}_s)$  is the entropy of  $\mathcal{P}_s$ , and  $KL(\mathcal{P}_s \parallel \mathcal{M})$  is the Kullback–Leibler divergence from  $\mathcal{M}$  (the marginal distribution over responses) to  $\mathcal{P}_s$  (the posterior distribution of responses when the stimulus category is known).

The Kullback–Leibler divergence "can be understood as a measure of information gain when one revises one's beliefs from the prior probability distribution [ $\mathcal{M}$ ] to the posterior probability distribution [ $\mathcal{P}_s$ ]" (Neuman, 2017, p. 43). In other words, it measures how much information about the cell's discrete firing rate response we gain from knowing the type of the presented stimulus. It is maximal, if a cell has one particular discretised firing rate value (say  $r_1$ ) for one stimulus type (say  $s_1$ ) and different values for all other stimulus

types. In this case, there is high uncertainty (entropy) about the response value before knowing the stimulus type and perfect certainty (zero entropy) about the response value after knowing the stimulus type. How much information was gained by knowing the stimulus type is then inversely proportional to the share that  $r_1$  has in the marginal firing rate distribution. The more unlikely  $r_1$  is in the marginal distribution, the higher is the information gained by knowing that the stimulus type is  $s_1$ , which perfectly determines the response to be  $r_1$ . Since  $r_1$  only occurs for stimuli of type  $s_1$ , the probability of the response  $r_1$  is equal to the probability of a stimulus of type  $s_1$ . The information gain is thus equal to  $-\log_2 P(S = s_1)$ , where  $P(S = s_1)$  is the occurrence probability of that stimulus type. Therefore, the maximum information a cell can have about any stimulus is equal to  $-\log_2 P(S = s_{\min})$ , where  $s_{\min}$  is the least probable stimulus type. If all stimulus types are equally probable, which is the case in this study, the maximum single cell information is given by:

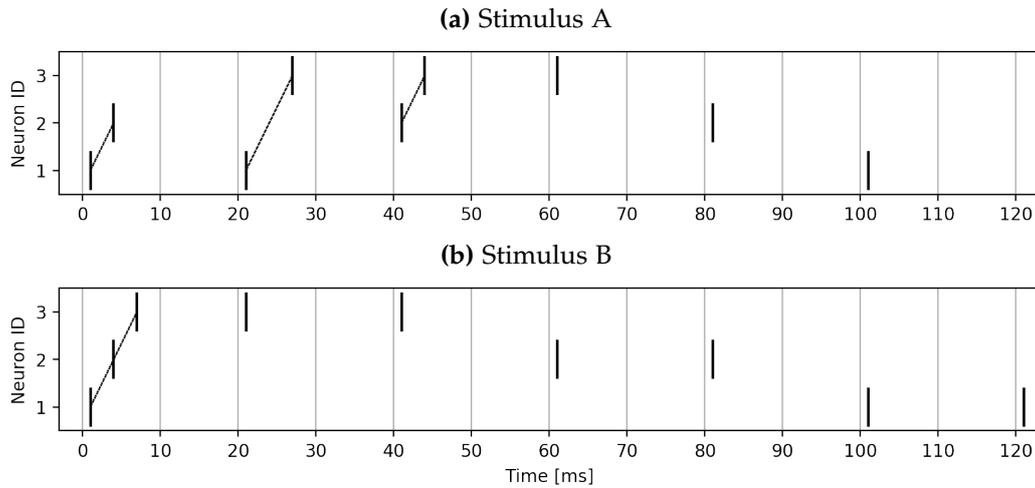
$$\text{Maximum cell information} = \log_2(\text{number of stimulus categories}) \text{ Bit} \quad (2.12)$$

It is important to note that the single cell information is positively biased with fewer stimulus presentations. For example, in the extreme case, where only one stimulus of category A and one of category B are presented, the firing rate responses of a cell to the two presentations are guaranteed to be placed in different firing rate bins, even if they are simply random, thus yielding full information. The more often each stimulus type is presented, the less likely it is that a cell's firing rate reliably ends up in different bins for the different stimulus types by pure chance. Thus, it is important to present each stimulus a fairly large number of times in order to obtain a good estimate of  $P(R|S)$ .

### 2.4.2 Single Cell Classification Accuracy

For stimuli that contain multiple objects (see figure 2.4) the single cell information metric is not applicable. The stimuli depicted in figure 2.4b, for example, contain both an object with a left border at location 1 and an object with a right border at location 2. Thus, these stimuli belong to two stimulus categories at the same time. The random variable  $S$ , however, can not take up multiple values at the same time and the theoretical framework of the single cell information breaks down. Therefore, we introduce in this section the *single cell classification accuracy* as a metric applicable to stimuli with arbitrary numbers of objects. Since these stimuli do not have a unique type, we will consider *objects* of particular types present in the stimulus instead of *stimulus types (categories)*. Each object always has exactly one type, which depends on the task at hand. When investigating contour element cells, for example, there are two types of objects (left border or right border) whereas investigations into border ownership representations are concerned with four types of objects (LB1, LB2, RB1, RB2). Each stimulus can contain an arbitrary number of such objects. Note that the stimuli in figure 2.3 are considered to contain zero objects, as there is no object with a desired characteristic (i.e. border at either location) present in those stimuli.

Inspired by concepts from machine learning, we treat each cell as a classifier that solves as many separate binary classification tasks as there are object types. A cell tries to classify,



**Figure 2.5:** Artificially generated spike trains. Spikes that occur with less than 10 ms delay are connected by dotted lines. In both spike trains, all spike *pairs* have the same probabilities. Spike train (b), however, exhibits a pattern of three spikes,  $PG=\{(1, 1 \text{ ms}), (2, 4 \text{ ms}), (3, 7 \text{ ms})\}$ , while spike train (a) does not.

for each object type, whether an object of that type is present in the current stimulus, based on thresholds of it's firing rate. For each object type, the cell is fitted with one such firing rate threshold and the cell is considered to be 'ON', if it's firing rate is above the threshold. This indicates that the object type is present in the current stimulus. If the cell's firing rate is below the threshold for the object type, it is considered 'OFF', which indicates the absence of an object of that type. The *single cell classification accuracy* of a cell for the object type  $o$  is then the fraction of testing stimulus presentations with or without an object of type  $o$  for which the cell was correctly 'ON' or 'OFF' respectively. Thus, as with single cell information, each cell has as many accuracy values as there are object types. If a cell has perfect accuracy for one object type, that type can be said to be the cell's preferred object as the cell is selectively active for stimuli containing this object type. That is, it's firing rate is always highest (above threshold, 'ON') whenever that object type is present. Consequently, the cell has poor accuracy for all other object types, because it is not possible to find a threshold with which the cell would be considered 'OFF' for it's preferred object type and 'ON' for a different object type, as the preferred object type elicits the highest firing rates which are therefore above any nontrivial threshold.

How are the thresholds for a cell determined? For an independent set of six fitting presentations of each stimulus in the training set (figure 2.1), 100 different threshold values are tried for each object type. For each object type, the threshold that achieves best accuracy (as described above) on these fitting presentations is then selected. The cell's accuracy for all other stimulus sets is computed using these thresholds. Note that the stimulus presentations used for fitting the thresholds are never also used when assessing the cell's accuracy.

### 2.4.3 Spike Pair PG Information

Following Eguchi et al. (2018), we also apply information theory to information conveyed by minimal polychronous groups that consist of two neurons emitting spikes with a fixed temporal delay between the two. As with single cell information (section 2.4.1) this metric is only applicable to stimuli that belong to exactly one *stimulus category*. The core idea is that there are pairs of spikes with a fixed temporal delay that occur frequently for one stimulus category and infrequently for another stimulus category. For each stimulus presentation  $p$ , a probability table is constructed of two spikes of each pair of neurons ( $i$  and  $j$ ) co-occurring with a certain temporal delay  $d$  (equation 2.13). Note that all pairs of cells are considered, regardless of whether there is actually a synaptic connection between the two cells.

$$ProbTable_p(i, j, d) = P(\text{Neuron } j \text{ spikes at time } t - d \mid \text{Neuron } i \text{ spikes at time } t) \quad (2.13)$$

In this study, we only consider cells within one layer and ten values for the temporal delay  $d$  from 0 ms to 10 ms. The probability table therefore has  $4096 \times 4096 \times 10 \approx 1.7 \cdot 10^8$  values and a separate table is generated for each stimulus presentation  $p$ . On this data, the information theoretic measure outlined in section 2.4.1 is computed by treating each position in the probability table (i.e. a pair of two cells with a particular delay) as an entity (analogous to a neuron) and the corresponding value in the table (i.e. probability of the spike pair occurring) as the entity's response to the stimulus presentation (analogous to a neuron's firing rate). Instead of computing single cell information on the firing rates of 4096 neurons, it is thus computed on the occurrence probabilities of roughly  $1.7 \cdot 10^8$  spike pairs.

Eguchi et al. (2018) mention, that a number of two-neuron interactions can chain together and form more complex polychronous groups involving multiple neurons. However, it is important to note that the metric does not necessarily pick up information conveyed by spike patterns involving more than two cells. Figure 2.5 provides a constructed example of this. The spike train for stimulus B clearly contains a polychronous episode of three consecutive spikes emitted by cells 1, 2 and 3 whereas stimulus A does not have such an episode. Thus, the three cells form a polychronous group that is exclusively activated for stimulus B and not for stimulus A. However, this is not reflected in the spike pair information, as all pairs of spikes have the same probability for both stimuli. Therefore, low spike pair information values do not necessarily exclude the possibility of larger polychronous groups conveying information about the stimulus category.

## Chapter 3

# Results

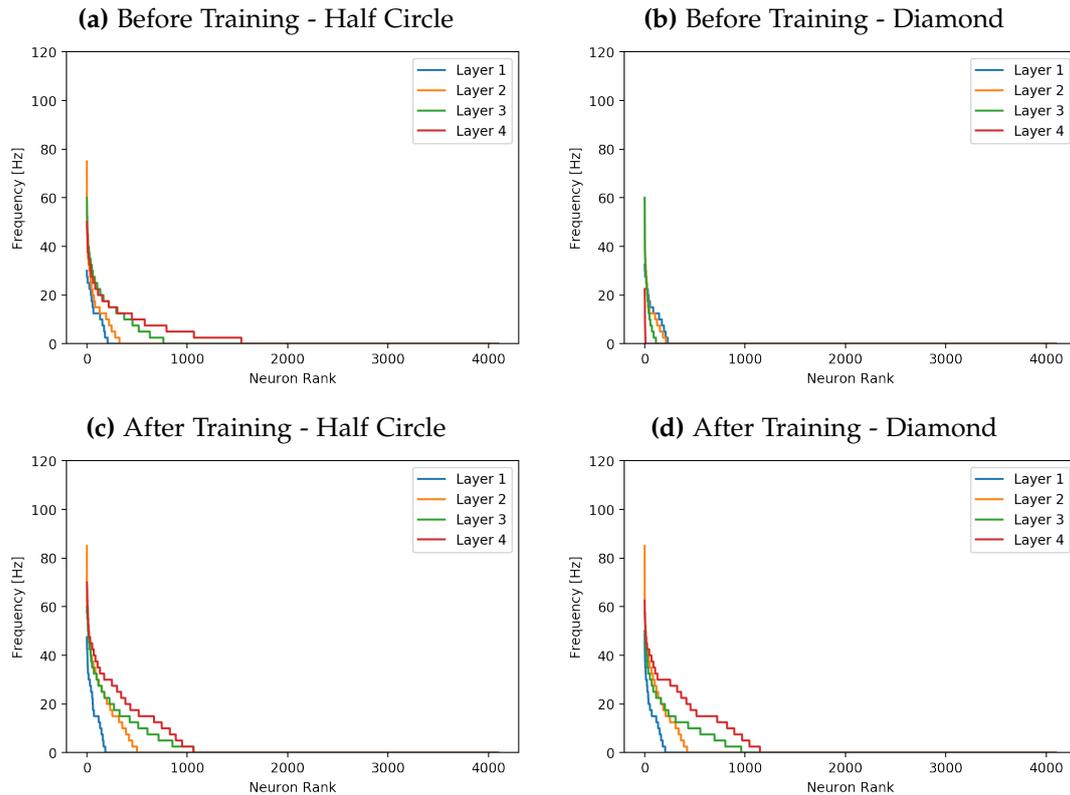
In this simulation study, the network is trained on the training stimulus set displayed in Figure 2.1 for 250 epochs. Each epoch lasts 48 seconds of simulated time and all eight objects are presented in the same order for six seconds each. Within those six seconds, the object is presented at one of the two locations, with the location potentially changing every 0.2 seconds. This is designed to emulate saccades. Consequently, the total time the object remains in each location is 3 seconds. After an object was presented, the network is reset to its resting state before the next object is shown. This emulates the long period of time between occurrences of independent objects in the real world, during which the network would decay to its resting state.

The network is tested before and after training on the training set (Figure 2.1), a testing set (Figure 2.2), a set of stimuli containing objects without a border (Figure 2.3) and a set of stimuli that contain multiple objects (Figure 2.4). In order to obtain a reliable estimate of the probability distribution of stable state firing rates, each stimulus is presented 11 times during testing for two seconds each and the network is reset between the stimulus presentations. On the following pages we explore the network's behaviour based on these testing responses.

### 3.1 Firing Rates

Since this study is aimed at deciphering how vision works in the biological brain, it is important to ensure that the network is in the same realm of firing behaviour as the biological brain. Extremely high firing rates in the simulation might enable types of processing that are different from the processes occurring in the brain. Thus, network parameters were optimised under the constraint of maintaining biologically plausible firing rates throughout training. As can be seen from figure 3.1 for two example stimuli, almost all neurons have firing rates well below 50 Hz, which is comparable to firing rates in the real brain (40 Hz for some border ownership cells in V1/V2 (Zhou et al., 2000)).

The most prominent effect of training is the equalisation of firing rates for different stimuli. Before training, similar but slightly different stimuli can have vastly different



**Figure 3.1:** Firing rates before and after training for a white half circle and a white diamond with a vertical straight border on their left side which is located at location 2. The firing rates are plotted for all excitatory neurons, sorted by their rank. Before and after training the firing rates are within the realm of biological plausibility. Before training - (a) and (b) - different objects (circle or diamond) elicit very different responses. After training - (c) and (d) - the firing rates are very similar for all objects.

network responses. In figure 3.1a and 3.1b for example, the firing rate distribution of the untrained network is plotted in response to a white half circle and a white diamond respectively. Both of these objects had a vertical straight edge on their left hand side that is precisely aligned with retinal location 2. When the half circle is presented, the activity propagates up the layers and a considerable portion of layer 4 neurons is active (Figure 3.1a). On the contrary, when the diamond is presented the activity decreases from layer to layer, quickly dying down. In layer 4 only a handful of neurons are activated for this stimulus. However, when the same diamond is presented in location 1 instead of location 2 the activity does propagate upwards similarly to figure 3.1a (not shown here). Thus it seems, that the combination of network parameters and initial synaptic weight distribution is such that the behaviour is on the brink between these two regimes. For some stimuli, the initial weights of involved synapses are slightly higher (due to the random initialisation), which results in activity increasing from layer to layer. For other stimuli, the initial weights are not as optimal and the activity thus decreases from layer to layer. After training, as exemplified in figures 3.1c and 3.1d, the network's response is very similar for all stimuli, with 1000 to 1300 layer 4 neurons active for any given stimulus. This is likely the result of synaptic weights adapting to the characteristics of the training stimulus set through learning. After training, all synapses activated by any of the training stimuli have increased weights that are more optimal. Thus, the effects of different random

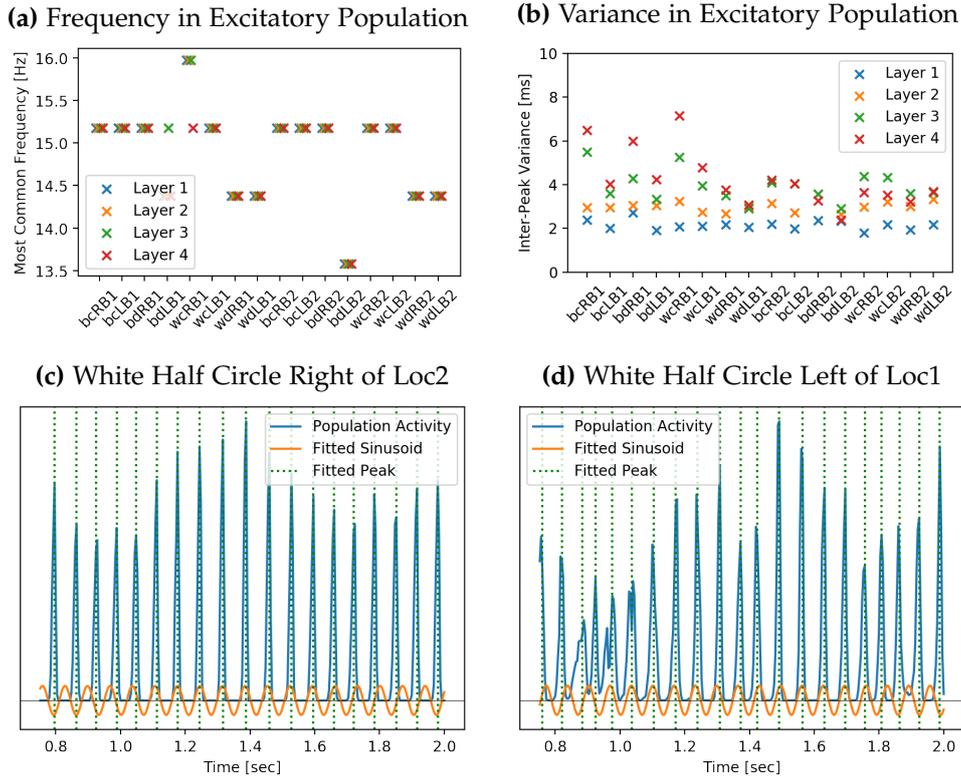
initialisation are mitigated.

Furthermore, training increases the sparseness of activity, especially in higher layers. After training, there are approximately 1000 layer 3 and layer 4 neurons active for all stimuli. In layer 4, this usually is a decrease in the number of active neurons in response to stimuli for which activity reached the fourth layer already before training. However, after training the neurons that are active usually have higher firing rates than before. This effect of fewer but individually more active layer 4 neurons is exemplified by the comparison of figure 3.1a and 3.1c. It is likely the result of competitive learning. As mentioned before, Hebbian learning leads to an increase in synaptic efficiency and therefore higher firing rates of some excitatory neurons for stimuli that contain their preferred object. Through lateral connections these then cause higher activity in the inhibitory neuron population. This, in turn, leads to higher lateral inhibition, which the excitatory cells have to overcome in order to fire. The neurons that were only slightly active before training do not receive enough synaptic feedforward input to overcome this new inhibition and are thus kept from becoming active. This results in *fewer* active neurons with *higher* firing rates.

## 3.2 Oscillations

Like the biological brain, the simulated network exhibits oscillations. In a given layer, the combined activity of all neurons of a certain type (excitatory or inhibitory) follows a sinusoidal pattern (see figure 3.2c for an example). To understand this process, we will consider a layer without any activity at stimulus onset. First, the excitatory neurons are activated by feedforward input. Once some of them begin to spike, this activity in the layer quickly gets further amplified by lateral excitatory connections and the entire population of excitatory neurons reaches its peak activity. Next, this causes lateral activation of the population of inhibitory neurons and it reaches its peak activity. This results in strong lateral inhibition which then quiets down the excitatory neurons, overpowering the still persisting feedforward input. Consequently, the inhibitory neurons stop firing as well, because they do not receive lateral input from the excitatory population any more. In the now completely quiet layer, the inhibitory influence on the excitatory neurons ceases and they can again be activated by the feedforward input. Thus, the process starts again.

These oscillations usually have the same frequency in all layers for a given stimulus (see figure 3.2a) but it varies for different stimuli. The regularity of the oscillations also varies across stimulus presentations (see figure 3.2d for a more noisy example). To estimate how reliable the oscillations are, we found the times of local maxima in the smoothed population activity and computed the variance of the interval in between two such peaks (see dotted lines in figures 3.2c and 3.2d for examples of found activity peaks). As can be seen from figure 3.2b, this variance tends to increase as activity propagates up the layers and it is always lowest in the first layer.



**Figure 3.2:** (a) Most common oscillation frequencies of the neuron populations in 11 presentations of each stimulus. Visual inspection indicates that all layers and both types of neuron populations exhibit the same oscillation frequency for a given stimulus. However, the Fourier transform only computes intensity estimates for discrete frequencies with a spacing of approximately 1 Hz for a 1.2 seconds long signal. If the frequency of oscillation lies in between two of these discrete frequency measurements, then it depends on small variations in the signal which of the two frequencies has the highest intensity estimate. This is likely the cause of different frequency estimates in different layers for the stimuli 'bdLB1' and 'wcrB1'. (b) Variance of the time between two peaks in the population activity, averaged across eleven presentations for each stimulus. (c & d) Activity of the population of excitatory layer 4 neurons for the presentation of a white half circle with an edge on it's left hand side that is aligned with location 2 and a white half circle with an edge on it's right hand side that is aligned with location 1 respectively. The Fourier transform based frequency estimate for each stimulus presentation that was used for subplot (a) is plotted as a sinusoid in orange. The green dotted lines indicate the peaks of population activity as they were found by our algorithm. These are used to compute the variance of the time-interval in between peaks, which is plotted in (b).

### 3.3 Information in Single Cell Firing Rates

Zhou et al. (2000) and Pasupathy and Connor (2001) reported that some cells have reliably higher firing rates, if objects with a certain characteristic are presented whereas they have reliably lower firing rates, if objects without that characteristic are presented. Such a neuron can be said to have a preferred stimulus characteristic (i.e. preferred value of a certain attribute of the presented object, e.g. *relative\_border\_location='left'*) and it responds maximally (in terms of it's firing rate) if and only if the presented stimulus has that characteristic. For example, a neuron's firing rate is always higher for objects that have an edge on their left hand side (*relative\_border\_location='left'*) than for objects with an edge on their right hand side (*relative\_border\_location='right'*). In this section we will investigate this kind

of firing rate based coding of object's attributes by single neurons.

### 3.3.1 V4 Boundary Contour Cells

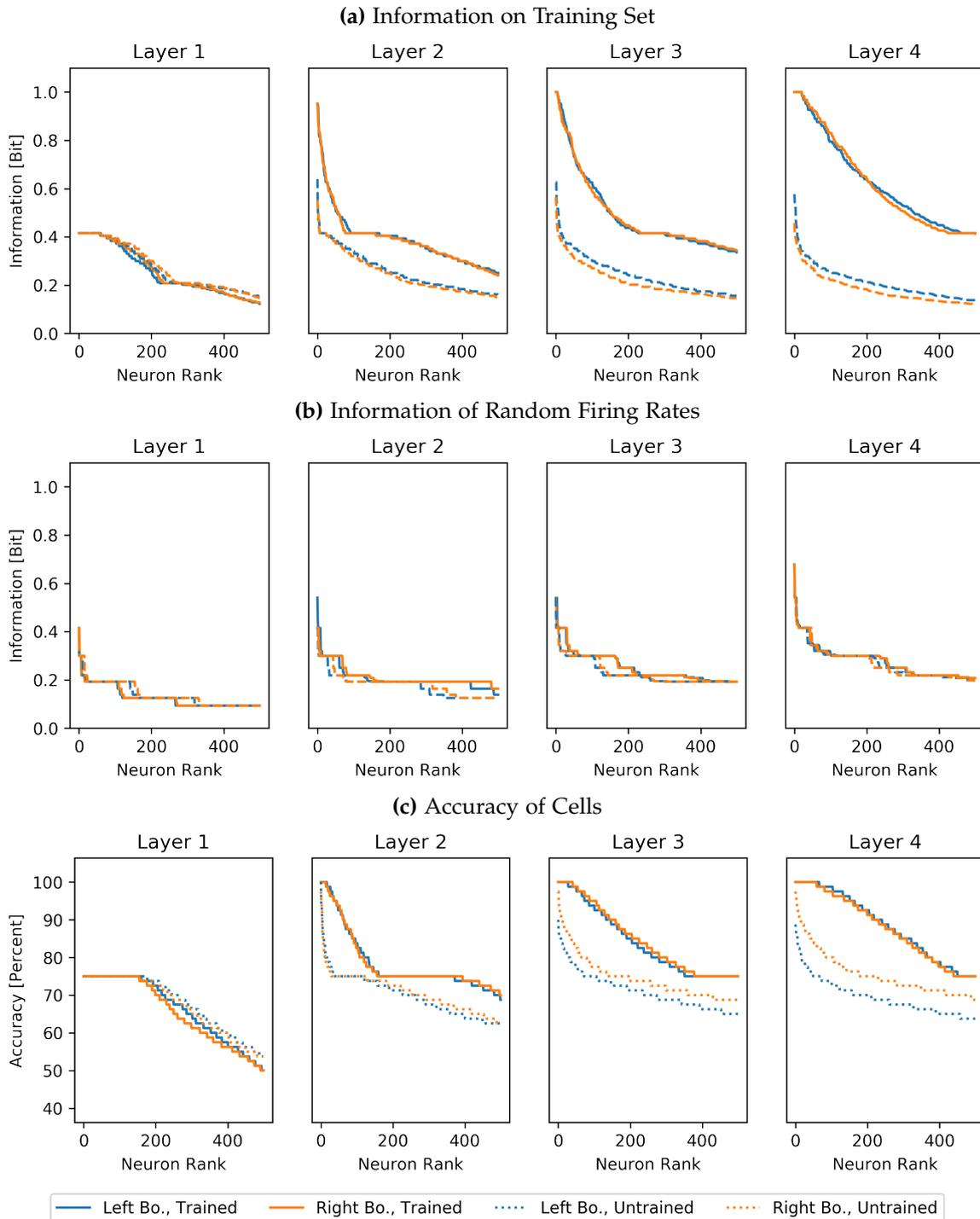
In visual area V4 a class of neurons has been found, that encodes local boundary contour elements of an object with respect to it's centre of mass (Pasupathy and Connor, 2001, 2002). In this simulation we consider two types of such cells: A  $\Phi_{BL}^{V4}$  cell, on the one hand, would have a high firing rate response if there is an object present that has a vertical edge on it's *left* hand side, which is aligned with any of the two retinal locations. If no such object is present then the cell's firing rate is considerably lower. A  $\Phi_{BR}^{V4}$  cell, on the other hand, would have a high firing rate if and only if an object is present that has a vertical edge on it's *right* hand side that is aligned with one of the two retinal locations. As expected, neurons emerge in the stimulation, which appear to exhibit this behaviour. On the following pages we will investigate their responses to the different stimuli sets (see section 2.2) in detail.

#### Information about Training Stimuli

First, we investigate the single cell information (see section 2.4.1) of the V4 contour element cells based on the set of stimuli that were also shown during training. In the fourth layer, 21 cells emerged that carry perfect single cell information of 1 Bit ( $= \log_2(2)$ ) about whether the presented object has it's vertical edge on the left or right hand side (figure 3.3a). Here, each stimulus was presented 11 times (corresponding to 88 presentations of each edge type) to obtain a more accurate estimate of the probability distribution over firing rates for a given edge type, which is used to compute the single cell information.

One concern with the above results is the risk of a multiple comparisons error due to the large number of neurons in each layer. One could argue that each cell has a completely random firing rate that is independent of the type of edge presented. One could further argue, that amongst a large number of these random neurons we would still find a couple of cells that exhibit the desired behaviour, simply because we have so many neurons to select from. The increase in information from before to after training would then be explained by the change in overall firing rates that was reported in section 3.1. To exclude this possible explanation of the results in figure 3.3a, we generated data exactly like just described and computed the information for this random baseline. As can be seen in figure 3.3b this randomisation completely destroys the information and therefore figure 3.3a does indicate, that some cells actually learned to respond differently to their preferred stimulus.

As explained in section 2.4.2 the single cell information metric requires each stimulus to belong to *exactly one* category. In this case, each stimulus has to contain either an object with a right border or one with a left border. This does not hold for some of the testing stimuli, that are investigated in the next section and the *single cell classification accuracy* metric (section 2.4.2) will be used for those stimuli. To achieve comparability of the performance for those novel stimuli and the training stimuli, we also computed the single cell classification accuracy on the set of training stimuli. The threshold, above which a cell



**Figure 3.3:** (a) Single cell information of the 500 most informative cells out of 4096 excitatory cells in each layer. Each stimulus from the training set was presented 11 times. (b) Single cell information baseline, where the firing rate of each neuron is drawn independently from the firing rate probability distribution before and after training. Multiple presentations of the same stimulus are assumed to elicit identical responses, while cells are assigned independently drawn firing rates for different stimuli. (c) Accuracy with which a cell has a higher firing rate (above fitted threshold) for its preferred stimulus and a lower firing rate (below fitted threshold) for a non-preferred stimulus. For each cell, the threshold is fitted to achieve optimal accuracy with 6 presentations of each stimulus. The plot shows the accuracy of this threshold for 5 independent presentations of each stimulus.

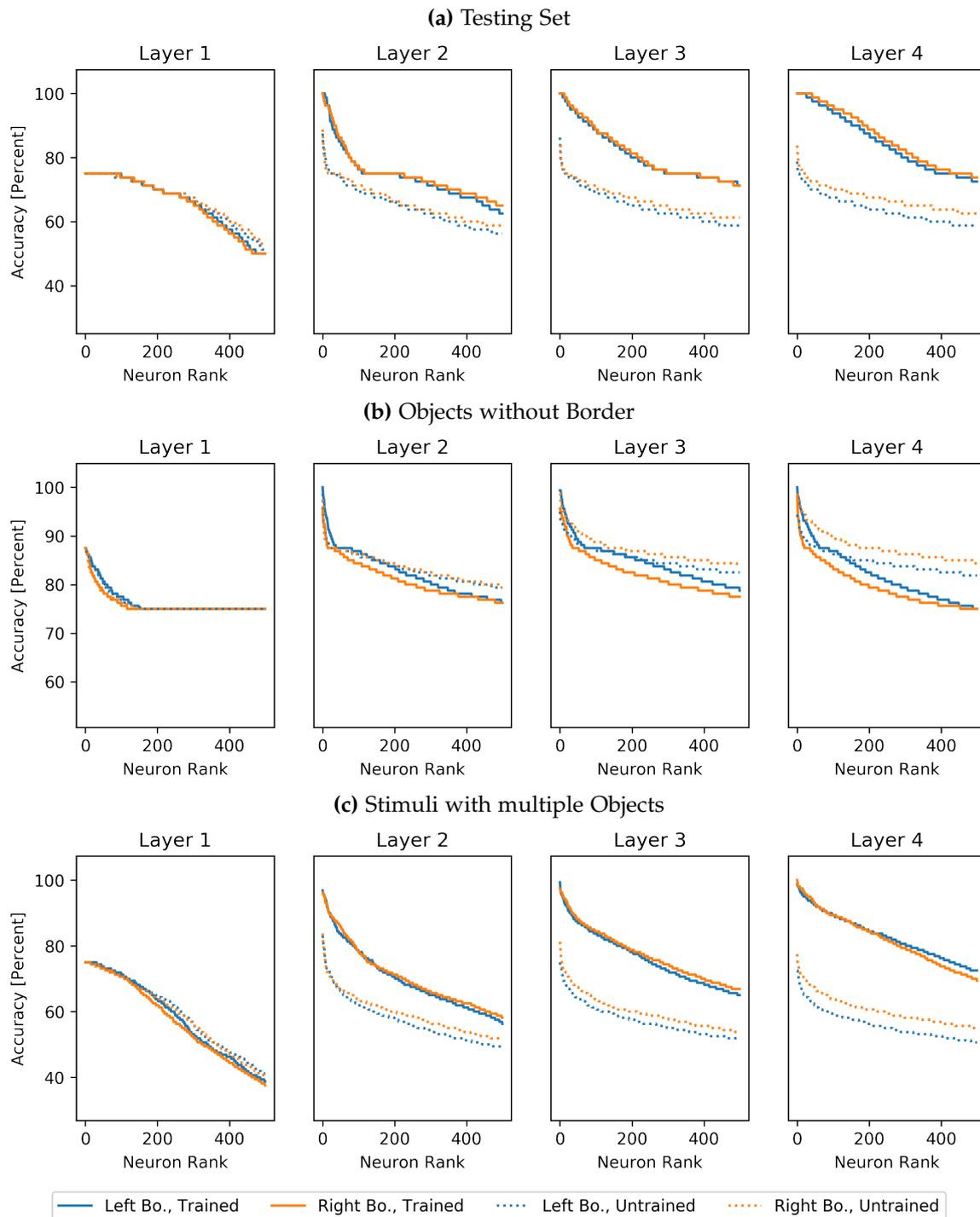
is considered to be 'ON', is fitted with six presentations of each training stimulus and the cell's accuracy (i.e. fraction of correctly classified stimuli) is assessed on five independent presentations of each stimulus. Figure 3.3c shows the classification accuracy with these thresholds for the 500 best neurons in each layer.

### Information about Novel Stimuli

Next, we assessed how robust the information is to different types of control stimuli. First, we present stimuli with novel objects that also always have a vertical edge on their left or right hand side, which is aligned with one of the two retinal locations. These stimuli have the same characteristics as the training set and the V4 contour element cells achieve similar accuracy (figure 3.4a).

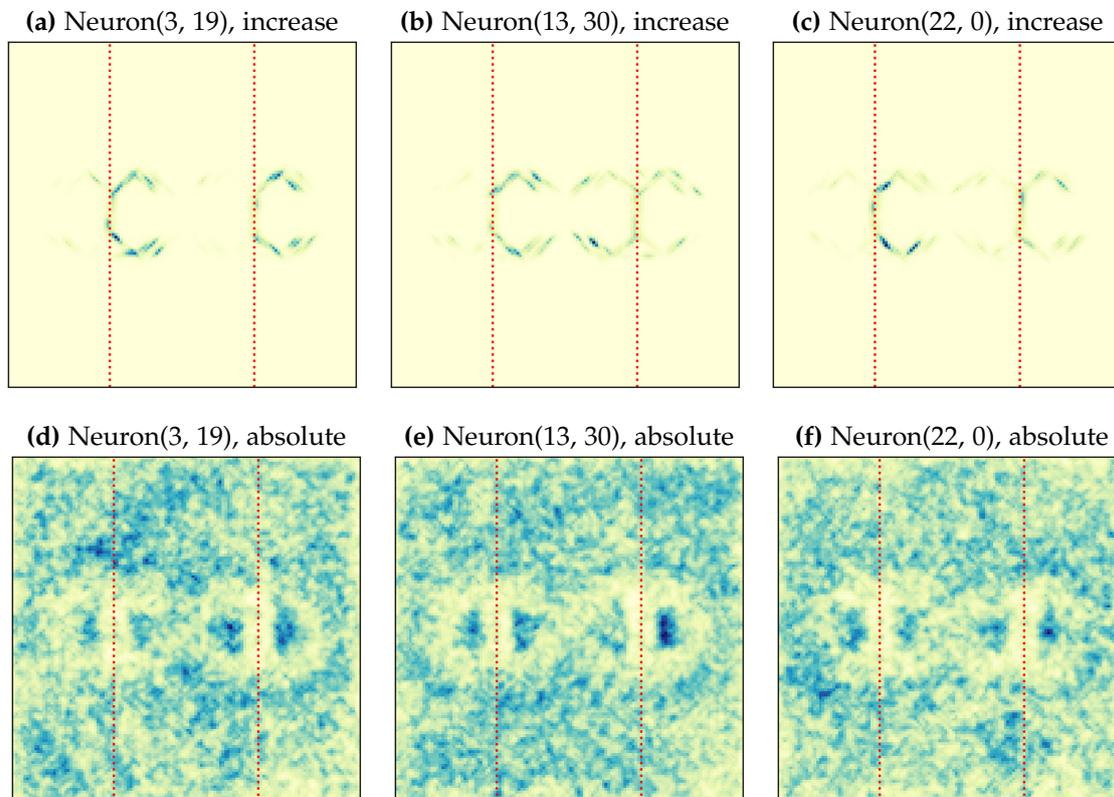
However, the opposite is the case when introducing novel objects to the training set that do not contain a border at either location (figure 2.3). After determining what constitutes a high firing rate indicating the presence of the preferred border type for each cell by fitting a threshold on 6 presentations of each training stimulus, we tested the accuracy of the cells on a set of new presentations showing in equal parts training stimuli and stimuli without a border. To achieve perfect accuracy on this combined stimulus set, a  $\Phi_{BL}^{V4}$  cell, for example, would have to be 'ON' (i.e. firing rate above threshold) for those training stimuli where the object has a border on it's left hand side and 'OFF' (i.e. firing rate below threshold) for the remaining training stimuli and all of the novel stimuli showing objects without a border at the two retinal locations. This is the behaviour we would expect based on the neurophysiological studies (Pasupathy and Connor, 2001, 2002). Figure 3.4b, however, clearly shows that no cell achieves good accuracy when stimuli without a target border are introduced. The supposed contour element cells are falsely active although their preferred contour is not present.

Furthermore, we presented stimuli that show multiple objects. Testing on these stimuli revealed the failure of the rate coded model in the previous simulation study (Eguchi and Stringer, 2016), which motivated this spiking investigation. As just described, we again used the thresholds found for each neuron with 6 presentations of the testing stimuli to decode whether the cell's response to a stimulus is 'ON' (above threshold) or 'OFF' (below threshold). We then computed the accuracy of the cells to a set consisting in equal parts of novel training stimulus presentations and presentations of stimuli with multiple objects. As can be seen in figure 3.4c, the accuracy significantly deteriorates when introducing these novel stimuli to the training set. However, performance increased through training and a few neurons come close to perfect accuracy.



**Figure 3.4:** (a) Accuracy of cells to the set of novel testing stimuli (figure 2.2). The threshold for each cell was fitted on 6 presentations of each stimulus in the training set (figure 2.1) and the accuracy was then computed on 5 presentations of each each stimulus from the testing set (figure 2.2). (b) Accuracy of cells for a stimulus set containing 50% training stimuli (figure 2.1) and 50% stimuli with objects that do not have any vertical border aligned with one of the target locations (figure 2.3). The threshold was fitted on 6 presentations of each training stimulus (figure 2.1) and the accuracy was assessed on the combined set of 5 presentations of each training stimulus and 10 presentations of each stimulus without an edge. (c) Accuracy of cells for a set containing 50% training stimuli (figure 2.1) and 50% stimuli with multiple objects (figure 2.4). The threshold is fitted on 6 presentations of each training stimulus and the accuracy is computed on the combined set of 5 independent presentations of each training stimulus and 5 presentations of each stimulus with multiple objects.

### Gabor Filter Input of V4 Contour Element Cells



**Figure 3.5:** Gabor filter that have strong connections to three exemplary layer 4 cells. All three cells are selectively active for objects with a vertical edge on their left, that is aligned with either retinal location 1 or 2. (a, b, c) Gabor filters are weighted according to the increase through training in synaptic efficiency of the paths that connect them to the layer 4 cells. (d, e, f) Gabor filters are weighted according to the absolute strength of the connections after training.

Finally, we investigated which Gabor filters have strong connections to the V4 boundary contour element cells. Figures 3.5a, 3.5b and 3.5c show Gabor filter, whose connections to three example layer 4 cells have increased through training. All three cells are selectively active for objects with a left edge at either of the two retinal locations. It can be seen that only Gabor filters on the outlines of the objects have increased.

Figures 3.5d, 3.5e and 3.5f show Gabor filters that have the strongest connections in absolute terms to the V4 cells after training. It seems that Gabor filters from the outlines of all objects have weaker connections than the surroundings, even though the Gabor filters with the highest increase in connection strength are also located there.

### 3.3.2 The Effect of Trace Learning

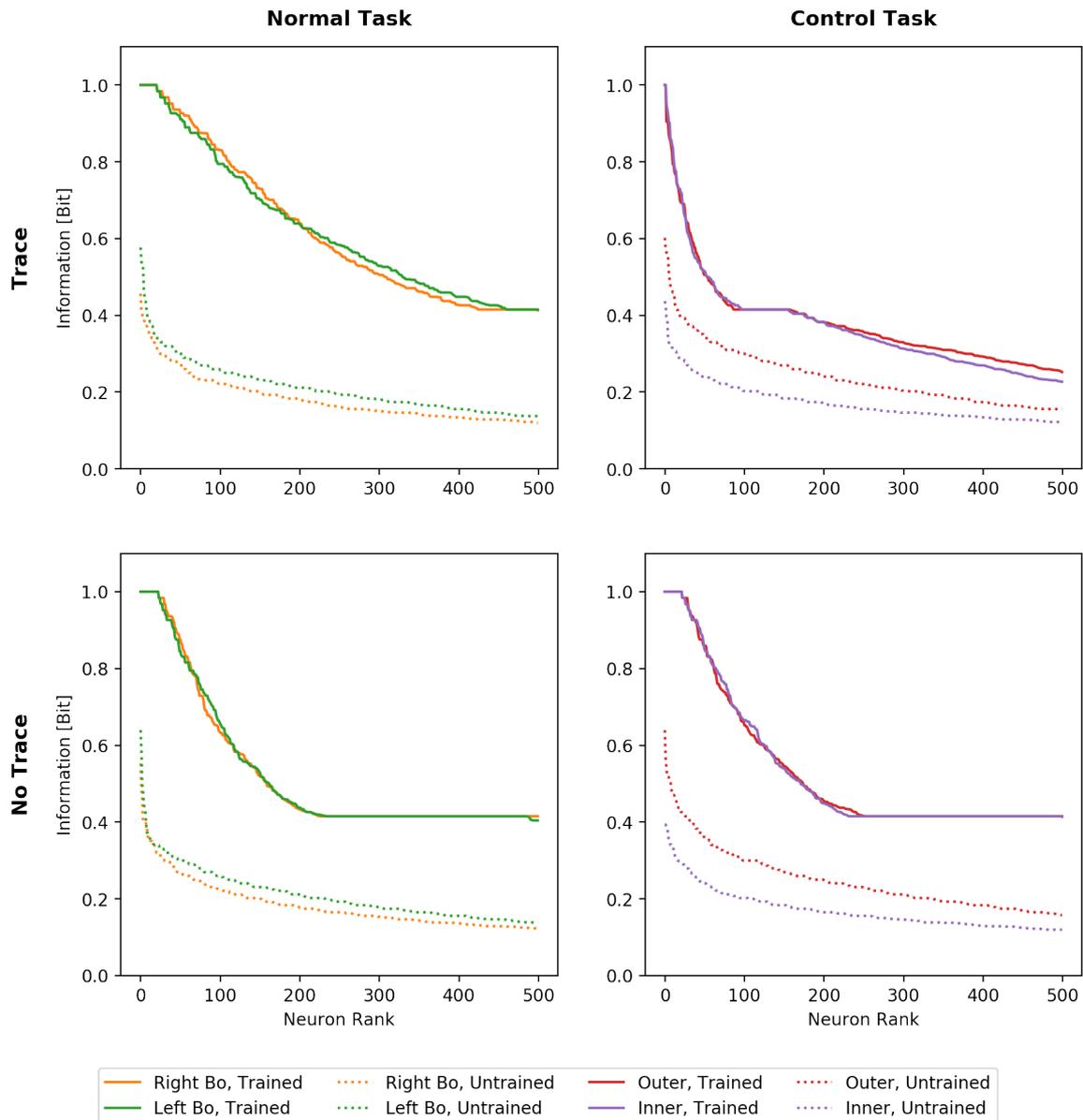
The just described boundary contour element cells found in higher layers display location invariance as a key characteristic. Such a cell would for example respond to an object with a vertical edge on it's left in the same way regardless of which retinal location that edge

is aligned with (1 or 2). In section 1.4, the *trace learning* mechanism was hypothesised to be crucial for this invariance to develop. In this learning paradigm, the network maintains the neural activity of the previous stimulus presentation and this biases the neurons that encode the previous stimulus to also respond to the current one. By alternately presenting each object in short succession at location 1 and 2, cells would, in the *Trace* condition, develop responses that have the above mentioned invariance. While all other results reported in this paper were obtained with a network trained with this interleaved presentation style, we investigate in this section, what effects explicitly disabling the trace mechanism would have. Therefore, in the *No Trace* condition, each object is first presented in one location for 3 seconds, after which the network is reset to its resting state before the same object is presented at the other location. Since the network is reset before each new stimulus, it is impossible for the activity of a previous object presentation to influence the response for the consecutive presentation of the same object at a different location. In this way any trace learning effect is disabled, while each stimulus is still presented for the same amount of time.

We compared a network trained in the *Trace* condition to a network trained in the *No Trace* condition for two different tasks. The *normal* task, on the one hand, is the same as described in the previous section. Each cell is assessed based on its ability to distinguish an object with a vertical edge on its left hand side from one with a vertical edge on its right hand side. Which retinal location the edge is aligned with, should not change the cell's response in this case and the interleaved presentation style in the *Trace* condition is designed to facilitate exactly this invariance. The *control* task, on the other hand, is semantically meaningless and cells are judged on their ability to distinguish stimuli with an object close to the edge of the retina (corresponds to a right edge at location 1 or a left edge at location 2) from stimuli with an object in the centre of the retina (corresponds to a left edge at location 1 or a right edge at location 2).

As can be seen from the first column in figure 3.6, the number of maximally informative neurons for the *normal* task does not decrease when disabling a temporal trace of previous neural activity. This is surprising as trace learning was hypothesised to be crucial for the development of the location invariance needed in the *normal* task (section 1.4). It seems that without the trace effect (i.e. in the *no trace* condition) different cells learn to respond reliably to different arbitrary subsets of stimuli. If a cell's subset coincides with what we defined as an object category (e.g. object with edge on its left in the *normal* task) the cell has maximal information. This is the case for a fair number (23) of cells in the *no trace* condition.

Since the subset of stimuli that a cell responds to is basically random and we simply picked the cells for which this subset coincides with our semantic definition of an object, it is unsurprising that we also find informative neurons for the semantically meaningless *control* task in the *No Trace* condition (figure 3.6 second column). When trained in the *No Trace* condition, the network achieves very similar information scores for both groupings of stimuli into objects (*normal* or *control* task). Thus, without trace learning, some maximally informative neurons seem to emerge that encode any arbitrary grouping of sufficiently dissimilar stimuli.



**Figure 3.6:** Performance of layer 4 neurons for different tasks with or without allowing a temporal trace of previous neural activity. In the *normal* task, on the one hand, a cell’s information is measured about whether it can distinguish a right edge at location 1 or 2 from a left edge at location 1 or 2. In the semantically meaningless *control* task, on the other hand, cells were judged according to their ability to distinguish an outer object (left of location 1 or right of location 2) from an inner object (right of location 1 and left of location 2). In the *Trace* condition, each object is shown at location 1 and 2 in short succession while allowing a temporal trace of neural activity. In the *No Trace* condition, the network was reset after each object presentation and no trace learning can therefore take place.

However, trace learning does have a beneficial effect, especially on mediocre cells, as is shown by higher information values for the *normal* task of neurons ranked between 100 and 400 (figure 3.6). Yet, an even more prominent effect of trace learning is the decrease of performance for the *control* task (right column in figure 3.6). For this task, an informative cell has to respond differently to stimuli, that were bound together by the interleaved presentation style in the *Trace* condition. That is, a cell would have to respond differently

for an object with its right edge aligned with location 1 and an object with its right edge aligned with location 2, since the former is close to the edge of the retina (first category in the control task) whereas the latter is in the centre of the retina (second category in the control task). In the *No Trace* condition some cells encode this arbitrary grouping of stimuli, while barely any do in the *trace* condition. This is likely due to cells losing their specificity to this arbitrary subset of stimuli, as other stimuli are also bound to the cells' preferred set by the trace mechanism.

From these experiments, it appears that the trace learning mechanism is not as crucial for the development of location invariance as originally hypothesised. Hypothesis  $H_5$  in section 1.4 can therefore not be confirmed.

### 3.3.3 V1/V2 Border Ownership Neurons

The second class of cells investigated in this study are border ownership neurons which were found in visual areas V1 and V2 (Zhou et al., 2000). These cells encode the conjunction of the two attributes border type (border on the object's left or right hand side) and object location (retinal location 1 or 2). Since both of these attributes can take two values, there are four types of border ownership cells (see section 1.2.1).

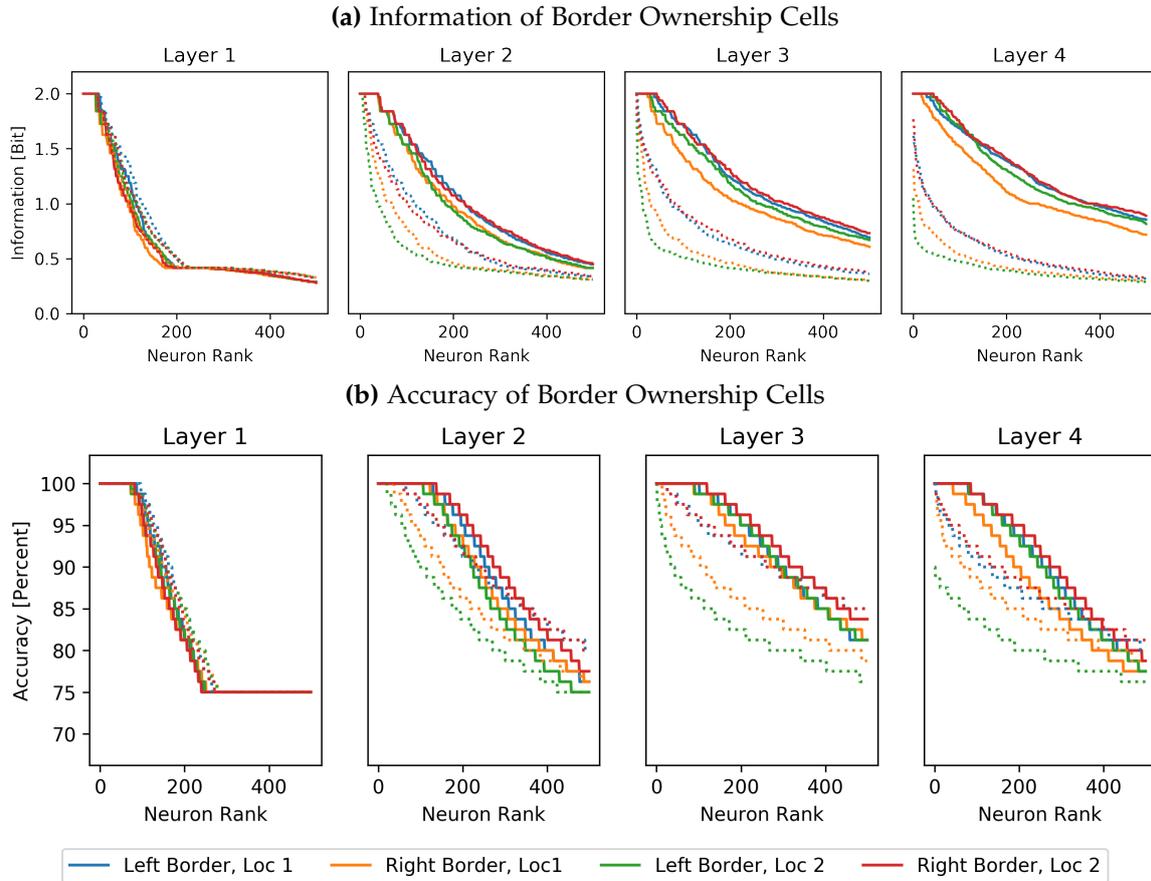
#### Border Ownership Information for Different Stimulus Sets

As hypothesised ( $H_1$  in section 1.2.2), a large number of border ownership cells can be found in the trained network (figure 3.7), if it is tested with the training stimuli (figure 2.1). Interestingly, higher layers exhibit at least as many border ownership cells as the first layer. Further it should be noted that, in the first layer, the border ownership cells did not develop through learning but were present already in the randomly initialised network. In higher layers, on the other hand, the information does increase during training.

Next, we tested the border ownership cells on the novel set of testing stimuli (figure 2.2). As previously described, each stimulus in this set also has exactly one vertical edge on the object's left or right hand side aligned with one of the two locations. The response of border ownership cells is largely as expected for these novel stimuli but performance does decrease a little bit (figure 3.8a).

More interestingly, introducing objects without a border at either of the two retinal locations (figure 2.3) massively decreases the accuracy of border ownership cells, especially in the first layer (figure 3.8b). This indicates that many border ownership cells have firing rates above the threshold, which determines that the cell is "ON", even if there is no border present in their supposed target location (retinal location 1 or 2). It calls into question whether many of the alleged border ownership cells in this simulation actually do encode border ownership as reported in Zhou et al. (2000).

Most importantly, we tested the border ownership response on the set of stimuli that contain two objects at the same time. In Eguchi and Stringer (2016), this stimulus set caused catastrophic breakdown of border ownership representation in the network, as the

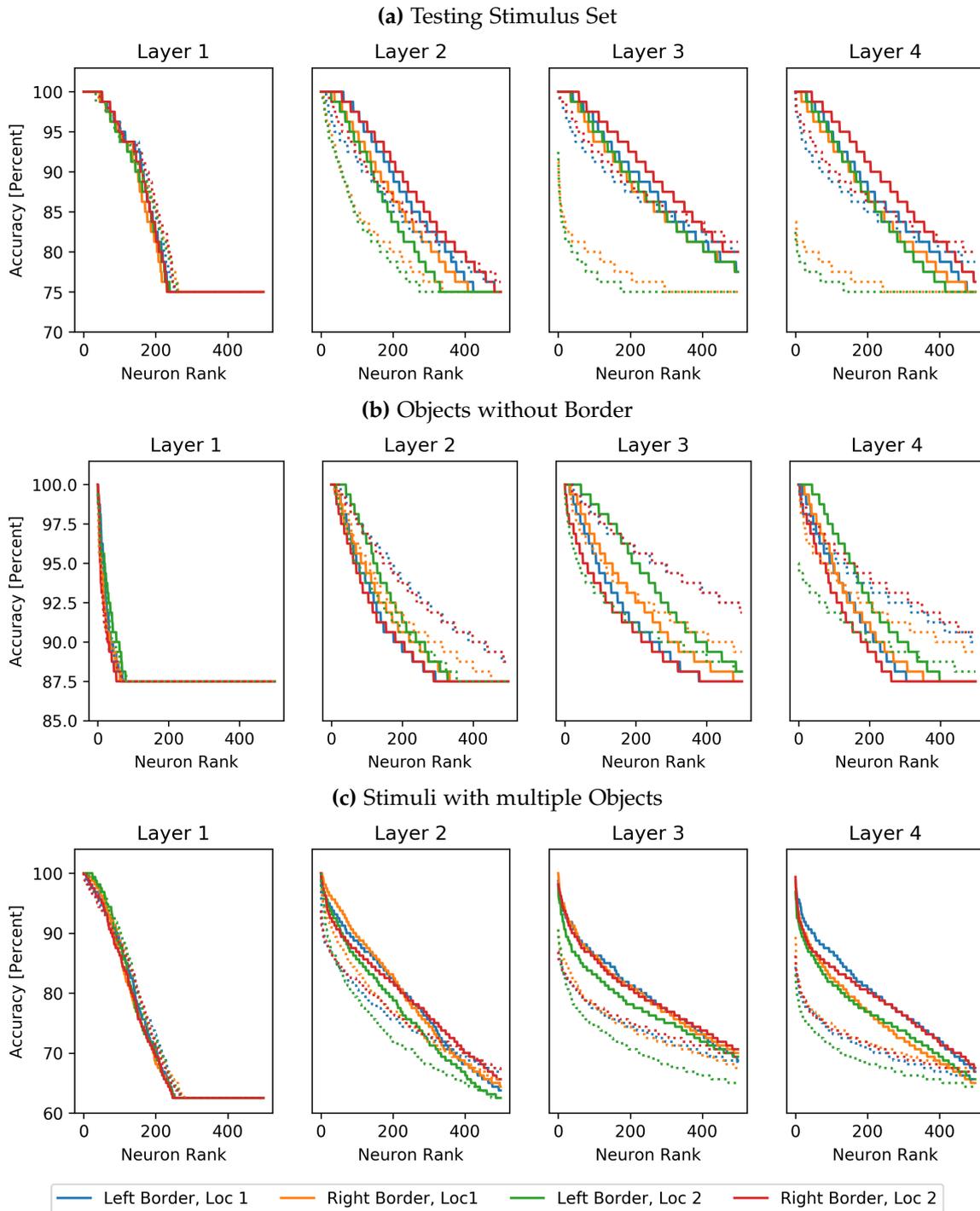


**Figure 3.7:** Performance of border ownership cells on the training stimulus set (figure 2.1). Information after training is indicated by thick lines, while information before training is represented by dotted lines. **(a)** Single cell information of the 500 most informative cells out of 4096 cells in each layer. Each stimulus from the training set was presented 11 times. Since there are 4 types of border ownership categories, the maximum information is 2 Bit ( $= \log_2(4)$ ) **(b)** Accuracy with which a cell has a higher firing rate (above fitted threshold) for its preferred stimulus and a lower firing rate (below fitted threshold) for a non-preferred stimulus. For each cell, the threshold is fitted to achieve optimal accuracy with 6 presentations of each training stimulus. The plot shows the accuracy of this threshold for 5 independent presentations of each training stimulus.

rate coded feedback from V4 neurons is not specific to the object that caused the activation. As outlined in section 1.2.2, we assumed that incorporating the dynamics of spike times into the simulation would solve this problem. That is, temporal patterns in relative spike times would make the V4 feedback specific to the object that caused the activity. However, figure 3.8c surprisingly shows poor performance of the border ownership representation for the stimulus set containing 50% presentations of training stimuli and 50% presentations of stimuli with two objects. Thus, a consistent border ownership response is not maintained when a distractor object is additionally presented on the retina, even though explicit spike times were incorporated into the model.

In summary, the quantitative analysis of border ownership representation brought up the following unexpected findings:

- Many border ownership cells respond incorrectly if there is no edge present at their target location.



**Figure 3.8:** Performance of border ownership cells on novel stimuli. Information after training is indicated by thick lines, while information before training is represented by dotted lines. **(a)** Accuracy of cells to the set of novel testing stimuli (figure 2.2). The threshold for each cell was fitted on 6 presentations of each stimulus in the training set (figure 2.1) and the accuracy was then computed on 5 presentations of each each stimulus from the testing set (figure 2.2). **(b)** Accuracy of cells for a stimulus set containing 50% training stimuli (figure 2.1) and 50% stimuli with objects that do not have any vertical border aligned with one of the target locations (figure 2.3). The threshold was fitted on 6 presentations of each training stimulus (figure 2.1) and the accuracy was assessed on the combined set of 5 presentations of each training stimulus and 10 presentations of each stimulus without an edge. **(c)** Accuracy of cells for a set containing 50% training stimuli (figure 2.1) and 50% stimuli with multiple objects (figure 2.4). The threshold is fitted on 6 presentations of each training stimulus and the accuracy is computed on the combined set of 5 independent presentations of each training stimulus and 5 presentations of each stimulus with multiple objects.

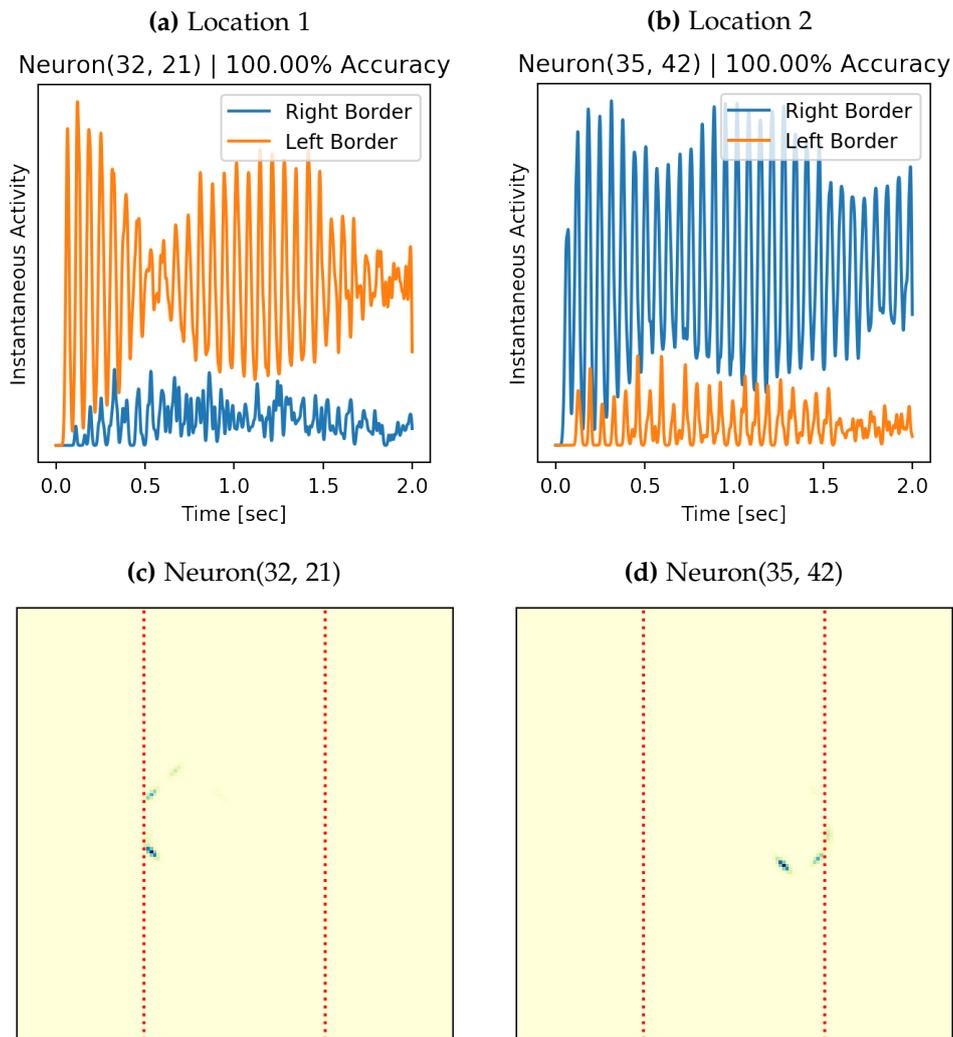
- In layer 1, all border ownership cells already exist before training.
- The border ownership response still breaks down when a distractor object is present.

### Detailed Investigation of Single Border Ownership Cells

To understand these unexpected findings we now proceed to look at some border ownership cells in the trained network in more detail. First, we investigate the hypothesis that the border ownership response relies on feedback from V4 contour element cells ( $H_2$  in section 1.2.2). Eguchi and Stringer (2016) reported, in accordance with experimental findings (Zhou et al., 2000), that the border ownership preference in V1/V2 cells only emerges some time after stimulus onset, after feedback from V4 is integrated. The cells are first equally active to both a left or a right edge in their target location and this initially invariant response is only subsequently modulated depending on which side of the object the border is located on (Eguchi and Stringer, 2016, Fig 9). However, in this simulation there are no cells that exhibit this behaviour. All border ownership cells that exhibit nonzero firing rates to any type of border at their target location, show a much higher firing rate to their preferred border type already from stimulus onset. In figures 3.9a and 3.9b the time course of activity averaged over 44 presentations of stimuli with each edge type is displayed for two exemplary neurons. It can be clearly seen that these cells exhibit their stimulus preference already from the beginning and that they do not rely on feedback which is only available later.

Then how do the cells know which side of the object the border is located on? To understand this, we have to look at the Gabor filters that have increased their connection strength to the border ownership cells through training. This gives some indication for a cell's preferred stimulus. It becomes apparent that the border ownership cells have input connections not just from the location of the target border but also have increased afferent synapses from Gabor filters located on the outline of the training objects. As can be seen from figures 3.9c and 3.9d, these strengthened connections from other parts of the object's contour already provide sufficient context information to determine whether the object extends to the left or to the right of the retinal target location. Thus, feedback from V4 contour element cells can not add new information because the border ownership cells are already supplied with the necessary context directly through their afferent feedforward connections from the input layer.

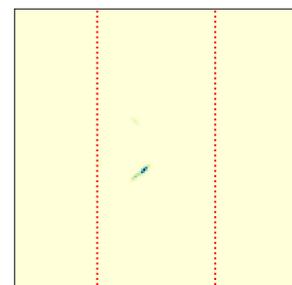
Further, many putative border ownership cells are not actually located on that particular border at all. Instead, they are situated on the outer contour of the objects and thus do not require any border to be present at their target location to become active (see figure 3.10 for an example). All they, therefore, encode is that the object present in the stimulus extends to the left/right of their retinal target location (1 or 2). The cells are not at all influenced by whether there is *also* a border at that retinal location. Such a cell would look like a border ownership cell when assessed on the training set, since, for example, "border on the object's left hand side at retinal location 1" is equivalent to "object extending to the right of retinal location 1" for the training stimulus set. However, for the set of stimuli without a border at one of the two retinal locations (figure 2.3) this equivalence does not



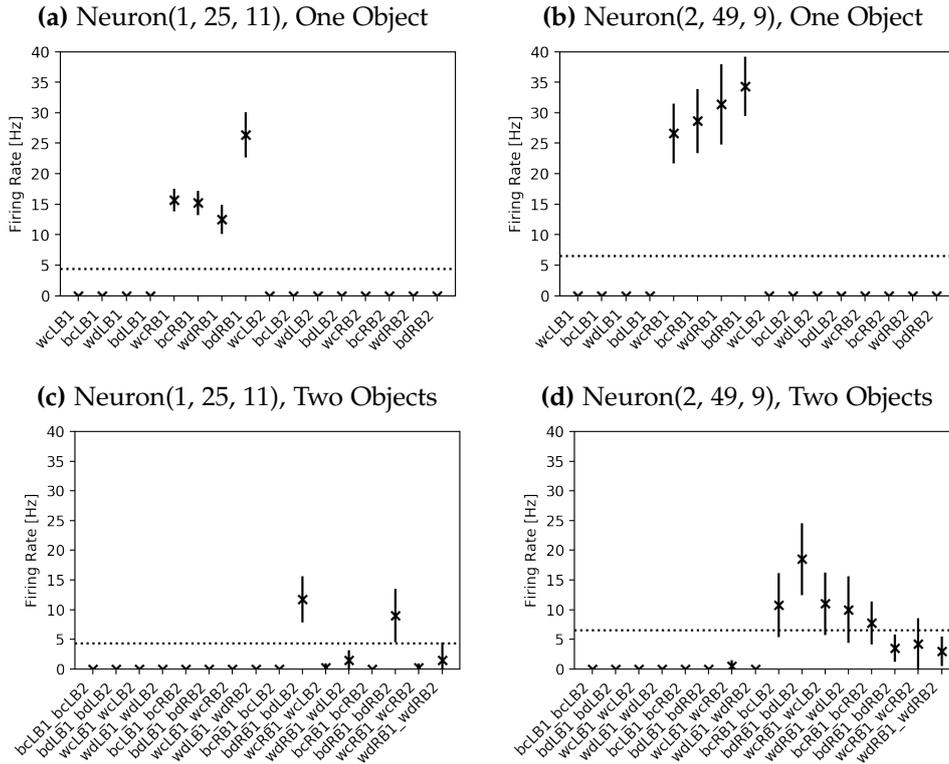
**Figure 3.9:** (a & b) Instantaneous activity of two sample neurons averaged over 11 presentations of each stimulus. This corresponds to 44 presentations of each border type. (c & d) Input Gabor filter for these two cells weighted by how much their synaptic strength increased through training.

hold, since, for this set, objects extending to the right of location 1 do not also have a border at location one. This explains the incorrect response of most border ownership cells when tested with the set of stimuli without a border at either retinal location (figure 3.8b).

Additionally, these cells located on the remaining outline of the object might also explain the big population of alleged border ownership cells (when assessed with the training set) already in the initial network. Such a cell has afferent connections exclusively from Gabor filters that are activated by the outer outline of objects. The cell from figure 3.10, for example, has connections only from Gabor filters at the outer contour of objects, which - on the training set - always have a left border at location 1. Since that cell has afferent synapses exclusively from Gabor filters that can only be activated by these objects, the exact weights of the synapses do not matter



**Figure 3.10:** Input Gabor filters to the first layer border ownership cell. (Neuron(31, 23))



**Figure 3.11:** For two border ownership cells of type  $\Phi_{BR,loc1}^{V1/V2}$  (border on object's right at location 1), the mean firing rates across 11 presentations of each stimulus are plotted. The bars indicate the standard deviation of the responses across the 11 presentations. The dashed line indicates the firing rate threshold above which a cell is considered to be 'ON'. **(a & b)** Responses of the two sample cells in Layer 1 and Layer 2 to the training stimuli. It can be seen that they are strongly active for objects with a vertical edge on their right hand side at location 1. **(c & d)** Responses of the two cells to stimuli with multiple objects. The cells should be active for all eight stimuli in the right half of the plots as these stimuli contain an object with a vertical edge on it's right at location 1 in conjunction with different distractor objects at location 2. However, the cells are incorrectly silent for many of them.

that much. Therefore, the cell is already from the beginning, selectively active for stimuli that show an object with a left border at retinal location 1, for the simple reason that such an object also extends to the right of retinal location 1 where all it's input Gabor filter are located.

The only remaining unexplained observation is the decrease in performance when multiple objects are presented at the same time (figure 3.8c). In the beginning of this section we already presented evidence indicating that the border ownership cells in the lower layers do not to rely on feedback from higher layers. Thus, the failure for the stimulus set containing distractor objects (figure 2.4) is unlikely to be the result of unspecific feedback, as it was the case in Eguchi and Stringer (2016) (see section 1.2.1). To confirm this, we examined the firing rate responses of all border ownership cells for the stimulus set with two objects. If the failure on this set would be the result of the mechanism explained in section 1.2.1 (Eguchi and Stringer, 2016), then we would expect the border ownership cells at location 1, for example, to be incorrectly active if there is an inconsistent object (i.e. with the other border type) present at location 2. All incorrect responses should, in this case, be **false positives** (firing rate above threshold). However, the performance decrease in this

spiking study is the result of a large number of **false negative** responses (firing rate below threshold) of the border ownership cells.

This behaviour, that is found in most border ownership cells, is exemplified by the two cells investigated in figure 3.11. They are selective to objects with a right border at location 1 (see figures 3.11a and 3.11b). To be proper border ownership cells, they should therefore be active to all those novel stimuli that have an object with a right border at location 1 (right half of the stimuli in figures 3.11c and 3.11d). If their failure would be due to unspecific feedback from V4 (Eguchi and Stringer, 2016), as explained previously, then they would be active for all stimuli that have an object with a right border at *any* location (all but the leftmost four stimuli in figures 3.11c and 3.11d). However, in reality neither is the case and the cells are falsely inactive for some stimuli.

The reason for the incorrect inactivity of the border ownership cells when two objects are presented seems to lie in the lateral inhibition. Clearly, two objects on the retina activate more excitatory cells than just one. As a result, there is higher activity in the inhibitory population which, in turn, will exert higher inhibitory pressure on the excitatory border ownership cells. This will keep some of the border ownership cells from becoming active (above threshold). In accordance with this is also the behaviour of Neuron(2, 49, 9) in figures 3.11b and 3.11d. It's activity is actually higher if there is an inconsistent (left border) distractor object at location 2 than if there is a consistent (right border) distractor object at location 2. This could be explained by the fact that a distractor object with a left border at location 2 (remaining contour right of location 2) is further away from the preferred object at location 1 than a distractor object with a right border at location 2 (remaining contour left of location 2). Thus, the only locally acting lateral inhibition caused by the more distant distractor has less effect on the border ownership cells at location 1. According to the unspecific feedback hypothesis (Eguchi and Stringer, 2016), on the contrary, we would have expected highest activity for a consistent distractor, which is clearly not the case here.

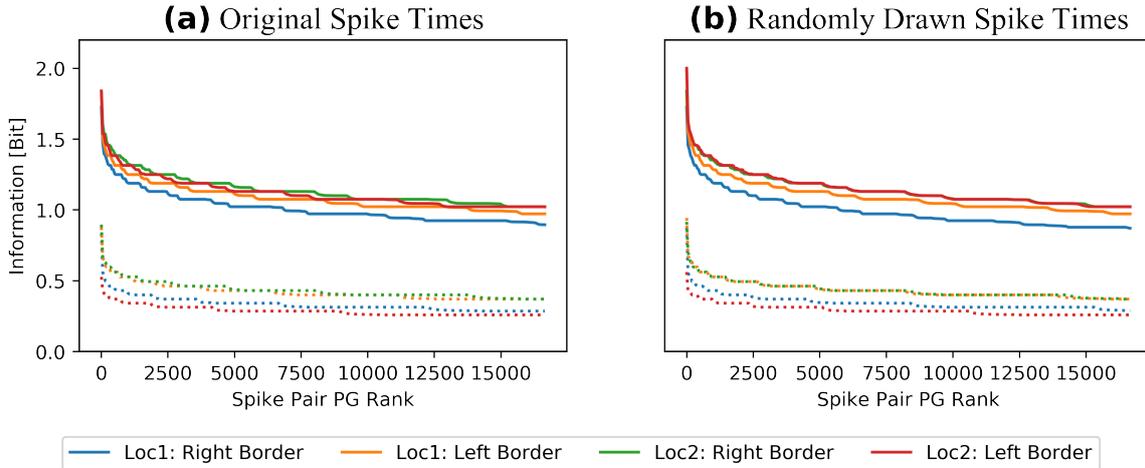
## 3.4 Polychrony

We hypothesised (section 1.3) that polychronous groups would emerge in the network through training and that binding neurons (section 1.2.2) would form as part of these polychronous groups, to represent border ownership. In this section, we investigate whether cells spike in reliable temporal patterns that carry information about the presented stimulus, which is a precondition for the binding circuit outlined in section 1.2.2.

### 3.4.1 Spike Pair Information

To achieve comparability with Eguchi et al. (2018), we initially consider polychronous groups consisting of only two neurons, emitting one spike each with a reliable time delay between the two. See section 2.4.3 for a full description of the spike pair information metric.

We measured the information about the four types of border ownership relations (left



**Figure 3.12:** Information of the most informative spike time pairs. There are  $n\text{-time-delay-bins} \cdot n\text{-cells-in-layer} \cdot n\text{-cells-in-layer} = 10 \cdot 4096 \cdot 4096 \approx 1.7 \cdot 10^8$  spike pairs and only the 0.1% most informative ones are plotted. Solid lines indicate information in the trained network, while information in the initial network is plotted as dashed lines. **(a)** Spike pair information as measured in the network. **(b)** Spike pair information after each spike was *randomly* assigned the spike time of some other spike within the corresponding stimulus presentation. This maintains the spike count of each neuron (i.e. firing rate) as well the overall spike time distribution of the population (i.e. oscillations as described in section 3.2).

or right at location 1 or 2) carried by spike pair PGs in Layer 4. The hypothesised mechanism, according to which the V1/V2 border ownership cells work (section 1.2.2), requires V4 contour element cells to spike with different temporal patterns depending on which retinal location the object's border is aligned with. For this reason, we investigate the information about border ownership (conjunction of border type and location) carried by spike pairs in layer 4. As mentioned previously, the information is positively biased with a decreasing number of stimulus presentations. Therefore, we present each stimulus 11 times to obtain reasonable estimates of the probabilities for two spikes co-occurring with a fixed time difference. As seen in figure 3.12a, there are no spike pair PGs carrying the maximum information of 2 Bit with this testing regime. However, when each stimulus is presented only once, one would find around 300,000 fully informative spike pair PGs for each border type (not shown here). While there are no spike pair PGs with full information, there still appears to be a very clear increase in information through training (figure 3.12a).

Further, we verified whether this increase in spike pair information is actually indicating the emergence of individual pairs of cells spiking in a reliable pattern which would allow for coincidence detecting neurons (like the hypothesised binding neuron). To this end, we computed the spike pair information after artificially 'breaking' these patterns in spike times of particular cells, expecting a large decrease in information on this artificial data. Specifically, a new spike time is assigned to each spike of each neuron. This new spike time is randomly drawn without replacement from all spike times of all layer 4 cells within the same stimulus presentation. The randomisation procedure maintains the firing rates of the cells (since each cell still has the same number of spikes, just with new times) and the overall distribution of spike times (i.e. oscillations, since each spike time still oc-

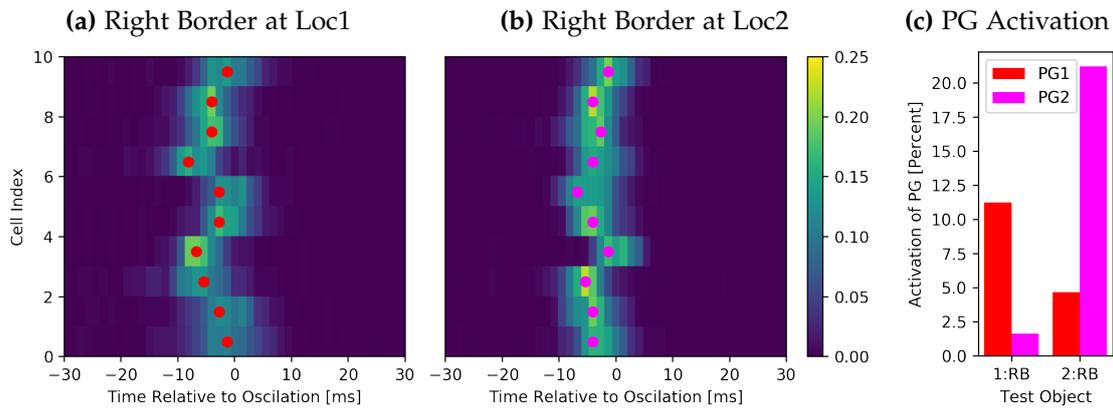
curs, just associated with a new neuron). However, it does break temporal patterns of spike times that are tied to the **identity** of a cell, which a coincidence detector cell would require. For example, if there was a repeating pair of two spikes (with the same delay) of two presynaptic cells of a coincidence detecting neuron, then the pair of spike times would still be in the data after the randomisation, but associated with two different cells each time. Therefore, if there is information carried by certain cells spiking with fixed relative times, then we would expect this information to disappear after the randomisation. Figure 3.12b shows the spike pair information after this procedure. In conflict with our hypothesis, the spike pair information after applying the randomisation is almost exactly the same as the information in the real data. Thus, it can be concluded that there is no information carried in the exact times at which particular cells spike.

But why then, does the spike pair information increase so notably through training? It turns out that the increase in spike pair information is a side effect of the increase in information carried by the firing rates of single cells (figure 3.7). To understand this, let us consider the example of a cell that does not spike for any of the stimuli before training, which corresponds to a firing rate based single cell information of 0 Bit. After training, this imaginary cell has a high firing rate for all stimuli containing an object with an edge on it's left hand side at location 1 and a firing rate of 0 Hz for all other stimuli. This corresponds to the maximal single cell information of 2 Bit. Let us further assume that the exact spike times of this cell are completely random (i.e. drawn from the overall spike time distribution). Before training, all spike pairs involving this cell trivially have a probability of 0% for all stimulus types. This clearly results in a spike pair information of 0 Bit. However, after training the cell fires, if and only if the stimulus shows it's preferred object. This, results in a small but **nonzero** probability for **all** spike pairs involving this cell (as long as the other cell in the pair also spikes), even if the times of the spikes are in fact random. For non preferred stimuli, the probabilities of all spike pairs involving our example cell are still zero, as the cell does not spike for these stimuli. Thus, the spike pair probabilities differ measurably between the preferred and the non preferred stimuli and the information carried by each spike pair involving the example cell is thus considerably larger than 0 Bit. This example easily illustrates how an increase in firing rate based single cell information also leads to an increase in spike pair information, even if the spike times are random. If the exact times of spikes were truly regular, then we would expect to see spike pair information in the real data (figure 3.12a) that is well above this random baseline shown in figure 3.12b.

In conclusion, we can therefore reject hypothesis  $H_3$  (section 1.2.2), since there are no pairs of cells with reliable relative spike times, which would be required by a coincidence detecting binding neuron.

### 3.4.2 Information of Spike Times Relative to the Underlying Oscillation

As outlined in section 2.4.3, the spike pair analysis conducted in the previous section might miss informative patterns that involve more than two spikes. To estimate the information contained in repeating patterns of more than two spikes, one ideally has to consider all



**Figure 3.13:** (a & b) Histograms of the relative first spike times of ten neurons within the corresponding peak of population activity. All of the ten cells are in the fourth layer and are only active for stimuli with a right border anywhere on the retina. For each cell and each peak in the population of excitatory layer 4 activity, the cell's first spike time is determined relative to this peak (see figure 3.2c for an example of how these peaks are found). The fraction of oscillation peaks for which a cell had its first spike at a particular relative time is indicated by the colour in the image. Dots indicate the maximum of the histogram for each cell, i.e., how much time before the population of all excitatory layer 4 neurons reaches its maximal activity each cell is most likely to spike. The histograms are computed for stimuli with a right border at location 1 (subfigure a) and for stimuli with a right border at location 2 (subfigure b). The temporal pattern for the two is clearly different. (c) Fraction of oscillation peaks for which the two polychronous groups are activated. PG1 and PG2 are given by the most probable relative spike times of the ten neurons for stimuli with a right border at location 1 or 2 respectively. The polychronous groups are marked with dots in subfigures a and b. A PG is considered activated in an oscillation if at least seven of the ten cells emit their first spike within 2 ms around the time given by the PG.

subsets of all spikes emitted by all cells within one stimulus presentation and check if some of these subsets only differ by a constant time factor added to all involved spike times. This, however, is computationally infeasible. To make the analysis tractable, we use the strong oscillations of neural activity prevalent in layer 4 (see section 3.2) as temporal fixed points. For each cell, we can then compute its first spike time relative to this fixed point for all oscillations of a stimulus presentation and we can analyse whether there is information in these times.

We only consider ten layer 4 cells, that are exclusively active for stimuli with a right border anywhere on the retina. As hypothesised in section 1.2.2, these cells should exhibit different temporal spike patterns indicating whether the border was aligned with location 1 or location 2. Figures 3.13a and 3.13b show the probability of the ten cells' spikes to occur at certain times relative to the population's activity peaks for these two cases. It can be clearly seen that the times of the first spikes of the cells differ, depending on where the border is located. For each of the two possible locations, the most probable relative spike time for each cell constitutes a polychronous group. PG1, for example, assigns each cell its most likely relative spike time when a right border is present at location 1 (PG1 is depicted by dots in figure 3.13a). PG2 is defined analogously as each cell's most probable spike time for stimuli with a right border at location 2 (PG2 is depicted by dots in figure 3.13b). We then measured the fraction of oscillations for which PG1 and PG2 are activated depending on the stimulus type (border at location 1 or 2). Here, a PG is considered to be activated, if at least seven of the ten cells emit their first spike (relative to the population's

peak activity) with maximally 2 ms difference from the spike time given by the PG. Figure 3.13c shows the probability of the two PGs being active for the two stimulus types. As expected, PG1 is mostly active for stimuli with the border at location 1 and PG2 is mostly active for stimuli with a border at location 2.

To get a more precise estimate of the information encoded in these times, we use a random forest (Breiman, 2001) to classify the border location based on the relative spike times of the mentioned ten cells. The input to the classifier is a ten-dimensional vector with the ten relative spike times of the previously introduced cells. For all layer 4 population peaks during a stimulus presentation, such a vector is generated. The desired output of the classifier for all of these vectors is **1**, if the stimulus had its right edge at location 1 and **2**, if it was aligned with location 2. In this way, training data is constructed from all oscillations occurring during six presentations of each training stimulus (figure 2.1). After the classifier is trained on this data set, it is tested on data obtained from all oscillations occurring during five *independent* presentations of each stimulus. The random forest classifier was able to assign the correct location of the right border in the presented stimuli for 91.2% of the oscillations in the testing data. This shows that there is considerable information in the relative spike times of the ten cells. It goes beyond the information carried by the activation of the cells, since all ten cells are active for both presented stimulus types.

However, this does not yet prove that the relative spike times convey more information than the relative scalar firing rates (not just active or inactive) of the ten cells. To test this, we carried out the same procedure as described in the previous paragraph while simply replacing the relative spike time of a cell with its scalar firing rate. Instead of one 10-D vector for each oscillation within a stimulus presentation, we now only have one such vector for each stimulus presentation. When using absolute firing rates instead of spike times, the classifier achieves a significantly higher accuracy of 99.6%. This suggests, that the information in first spike times is simply a mirror of the information in the firing rates, as cells with higher firing rates are likely to spike earlier in time. In this way, stimulus dependent variations in the firing rate of a cell also result in (more noisy) variations of first spike times, which are picked up by the classifier. But the exact times of spikes do not, in fact, carry more information than what is already present in the cells' relative firing rates.

In conjunction with the results from the previous section, this is strong evidence against hypothesis  $H_4$  from section 1.3, stating that there is information in reliably repeating temporal spike patterns. Or at least it confirms that information in spike times does not go beyond the information already carried by the cells' firing rates.

## Chapter 4

# Discussion

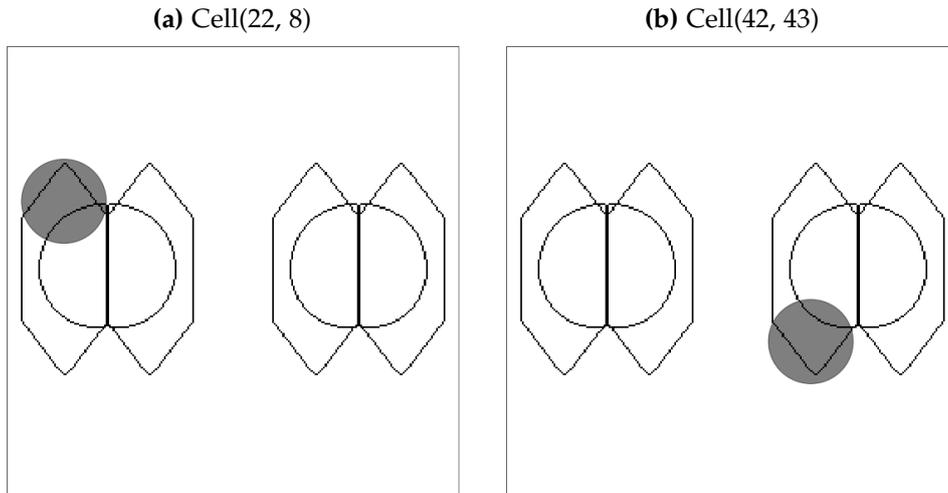
In light of the just presented results, the hypotheses outlined in the first chapter can not be confirmed. Specifically, no unambiguous border ownership cells developed in this particular simulation study and there is no indication that the mechanism from section 1.2.2 is utilised by the network. Further, there is no evidence of informative polychronous groups developing in the network. However, the shortcomings of the network's behaviour were in some cases rather opaque and not directly obvious from the first analysis. These flaws were only discovered after further in-depth analyses, which was not carried out in previous studies about border ownership and polychronisation conducted on similar networks (Eguchi and Stringer, 2016; Eguchi et al., 2018). On the following pages, we will summarise the problems found in this study and outline whether previous work might lack the analysis required to exclude the same problems. Finally, we will make some suggestions of how to possibly mitigate these problems in further studies.

### 4.1 Problems with the Border Ownership Representation

Initially, border ownership cells seemed to have developed in the network when only assessing their performance on the training set. However, these putative border ownership cells turned out to have major shortcomings.

#### Border Ownership Cells on the Outer Outline of Objects

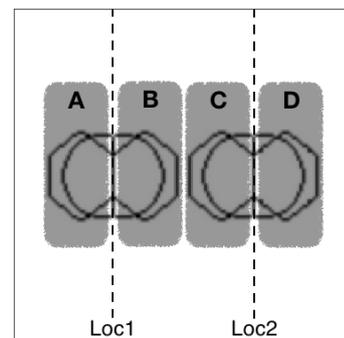
Most of the alleged border ownership cells are found on the outer outline of the training objects or receive strong afferent synapses from those regions of the retina (see the cell depicted in figure 3.10 for example). Such cells can not be considered border ownership cells, since no border is actually necessary for them to be active. The cells simply encode, that there is *something* in their receptive field. Specifically, cells that seemed to be border ownership cells of types  $\Phi_{BR,loc1}^{V1/V2}$ ,  $\Phi_{BL,loc1}^{V1/V2}$ ,  $\Phi_{BL,loc2}^{V1/V2}$  and  $\Phi_{BL,loc2}^{V1/V2}$  do, in fact, only encode the presence of a contour in the retinal regions A, B, C or D respectively (see figure 4.2). This is a much simpler explanation for the high performance of supposed border ownership



**Figure 4.1:** Areas from which 95% of all afferent synapses of two layer 1 cells are drawn. The cells were reported to encode border ownership in the previous rate coded study (Eguchi and Stringer, 2016, figure 7). The input area of a cell is reconstructed from the parameters given in table 1 in Eguchi and Stringer (2016), i.e. the radius of the grey circle is  $2 \times \sigma$ , where  $\sigma$  is the standard deviation of the gaussian that feedforward synapses from the input to the first layer are drawn from. **(a)** A cell reported to encode a left border at location 1. **(b)** A cell reported to encode a right border at location 2.

cells for the set of training stimuli. When only assessed on a stimulus set like the one used for training this network (figure 2.1), it is thus more likely that high performance of "border ownership representation" is a result of cells responding to a contour in a certain part of the retina (area A, B, C or D in figure 4.2). This possibility can not be excluded, if a *particular border type at a particular location* is equivalent to *some contour in a particular area* (as shown in figure 4.2) for all testing stimuli. Specifically, it has to be ensured, that a border at the cell's target location is necessary to elicit a response in it, in order to confirm that the cell actually encodes border ownership. Presumably the easiest way to do this, is including testing stimuli that do not have such a border, as was done in this study (see figure 2.3 for those stimuli). This breaks the equivalence of *border type at particular location* and *some contour in rough area of retina*, leading to a breakdown of the putative border ownership representation.

In the previous study on border ownership representation in a rate coded network (Eguchi and Stringer, 2016), however, this was not explicitly tested. The border ownership cells found by Eguchi and Stringer (2016) might therefore also just respond to contours in one of four rough areas of the retina, as illustrated in figure 4.2. In Eguchi and Stringer (2016), the actual locations of four layer 1 border ownership cells are given in figure 7. From the cells' location and the fan-in-radius for feedforward connections to the first layer (Eguchi and Stringer, 2016, table 1) we can reconstruct from which



**Figure 4.2:** Schema of the outlines of all training stimuli in this study.

area of the retina those cells have afferent synapses. Two of these four cells are not located on the retinal location that one would expect (figure 4.1). They are instead located on

the outline of the objects. Thus, it is very likely that the two cells do not actually encode border ownership but rather the presence of a contour in that rough area of the retina, as described previously. Therefore it stands to reason that at least some of the alleged border ownership cells reported by Eguchi and Stringer (2016) behave in this way and thus do not really encode border ownership. This would have been noticed when also testing on a set of stimuli that do not have borders at either location (figure 2.3).

#### Lack of Feedback from V4

A further problem in the present study is the lack of feedback from V4 contour element cells. The integration of feedback from area V4 is a crucial part of the hypothesised mechanism for border ownership representation in this study (see section 1.2.2) as well as in the previous rate coded work (Eguchi and Stringer, 2016). In the simulated network discussed in this report, however, no evidence for such feedback was found. Specifically, the time course of activity in border ownership cells indicates that they do not rely on feedback input, which is only available some time after stimulus onset (figure 3.9). This is in line with the above outlined finding that border ownership cells simply encode rough areas of the retina.

Conversely, Eguchi and Stringer (2016) reported some cells that do show an only later emerging specificity to the relative location of the border. This is evidence that feedback input influences at least those particular border ownership cells in the rate coded study. However, it can still not be completely excluded that the later emergence of specificity in those cells is the result of lateral excitatory or inhibitory effects instead of the result of excitatory feedback. Furthermore, finding a few examples of later emerging specificity does not allow the conclusion that all cells with high border ownership information behave in this way.

#### Failure of Border Ownership Representation with Multiple Objects

A final issue with the border ownership cells in the current study is their failure to respond correctly to multiple objects being presented in one stimulus (figure 3.8c). However, this failure is not the result of unspecific feedback from V4 (see section 1.2.1), as was hypothesised in Eguchi and Stringer (2016). According to this hypothesis, one would expect a border ownership cell of type  $\Phi_{BR,loc1}^{V1/V2}$ , for example, to be incorrectly active, even if the object with a right border is actually located at location 2. The right border at location 2 would activate V4 contour element cells that are selective for a right border anywhere on the retina. Their unspecific feedback would then activate the border ownership cells of type  $\Phi_{BR,loc1}^{V1/V2}$ . According to this unspecific feedback hypothesis, border ownership cells should be active whenever there is an arbitrary border at their preferred location and an object with the preferred border type anywhere on the retina, since the feedback information about the border type of the later object is falsely associated with the edge location of the former object. Consequently, all incorrect responses of border ownership cells should be a result of false positive activity.

However in this study, we found that border ownership cells are incorrectly *inactive* for some of the stimuli with multiple objects (false negative responses). As detailed in section 3.3.3, this false inactivity is a result of stronger lateral inhibition when multiple objects are presented, which keeps cells from becoming active. Thus, the finding that border ownership cells have poor performance for stimuli with multiple objects does not prove that the reason for their failure is unspecific feedback as outlined before. The true nature of the border ownership cells' failure only becomes clear when looking at their actual responses to the novel stimuli with multiple objects.

In Eguchi and Stringer (2016), on the other hand, the failure of the border ownership representation on the novel stimuli is not further investigated. Thus it is possible that the breakdown of information in that study is also the result of lateral inhibition being at play. To prove conclusively, that the reason for the decrease in information is actually unspecific feedback, as hypothesised in Eguchi and Stringer (2016), one has to confirm that the firing rate responses of border ownership cells actually do behave in the way previously outlined. That is, cells should be incorrectly *active* (false positive response) whenever the correct border type is *anywhere* on the retina.

### Summary

It is clear that none of the cells investigated in the present study behave like border ownership cells. Furthermore, many of the problems with the border ownership representation were discovered only after in-depth analysis, which was, however, not carried out in Eguchi and Stringer (2016). It is likely that at least some of the border ownership cells reported in Eguchi et al. (2018) are subject to the same issues. Thus, further analysis as detailed above would be helpful to confirm and better understand the behaviour of the rate coded network.

## 4.2 Problems with Polychronisation

The second major hypothesis investigated in this report regards the emergence of informative polychronous groups, which are selectively active for particular stimulus categories. Further, the *binding neuron* was hypothesised as a coincidence detecting cell that utilises the fact that presynaptic spikes arrive at different times depending on the synapse that they are transmitted by (see section 1.2.2). Since the transmission delays vary only by a few milliseconds between synapses, those cells are required to detect the coincidence of spike arrivals with millisecond precision. It was anticipated that these binding neurons would form as part of informative polychronous groups, as explained in the first chapter of this report. Previous work by Eguchi et al. (2018) on a similar spiking neural network model presented evidence for both hypotheses.

In the study by Eguchi et al. (2018) as well as in the current study, a rather long synaptic time constant  $\tau_g$  of 150 ms is used in order to facilitate trace learning (see section 1.4). However, this also means that a presynaptic spike influences the postsynaptic cell

for a rather long period of time, since the time course of the synaptic conductance is governed by that parameter (see equation 2.3). For example, 5 ms after a spike reached the synapse, the synaptic conductance is still 96.7% of what it was right at the time of the spike arrival. Thus it seems unlikely that cells could be able to detect coincidences in presynaptic spike arrivals with millisecond precision. Specifically, suppose two spikes arrive at a coincidence detecting cell at exactly the same time. Both of these synapses then cause 100% of their maximum current inflow to the postsynaptic cell, which, according to the hypothesis, is sufficient for a postsynaptic spike. If one of the spikes arrives 5 ms before the other spike, the conductance of that synapse would still be at 96.7% of its maximal value when the second spike arrives. It seems unlikely that this minor difference is enough to have the cell fall short of reaching the spiking threshold, which would be expected for a temporally precise coincidence detector. Therefore it seems rather implausible that spike times are utilised by the network with millisecond precision.

Nonetheless, Eguchi et al. (2018) found high levels of information in spike pair PGs. This was interpreted as evidence for the emergence of informative polychronous groups in the network. Similar results were also obtained in the study at hand, although no fully informative spike pair PGs were found here. Eguchi et al. (2018), on the other hand, report a large number of fully informative spike pairs. The reason for this discrepancy is that each stimulus was only presented twice in that study while there were 11 presentations of each stimulus in the study at hand. This explains the overall higher spike pair PG information values in the previous study, due to the positive bias of the information analysis with fewer stimulus presentations as outlined in section 2.4.1. Furthermore, similarly to the previous study, we also found a high increase in spike pair information about the border ownership through training (figure 3.12a). However, a deeper analysis revealed that this spike pair PG information is in fact only a side effect of information carried by the firing rates of cells, which became clear when we obtained almost identical information levels after randomising the times at which cells spike (see section 3.4.1). Consequently, a strong increase in spike pair PG information is not sufficient to prove that informative polychronous groups emerge in the network through training, as was hypothesised by Eguchi et al. (2018). In order to prove this, one has to show that the information in the real spike train far exceeds the information in a random spike train that exhibits the same firing rates of each individual cell. However, this comparison was not done in Eguchi et al. (2018) and, given the similar network architecture and parameters, it is thus likely that the apparent spike pair PG information in that study is also simply a side effect of informative firing rates as it is the case in the study at hand.

Eguchi et al. (2018) present further evidence for the informativeness of polychronous groups in form of the specificity of trigger events. A trigger event is a set of 3 cells with a common postsynaptic cell and a spike time associated with each cell such that the postsynaptic cell fires when the 3 presynaptic cells fire with their specified times. The authors then count how often each trigger event occurs in the spike train of the network for different stimuli (i.e. how often the three cells fire with the specified relative times). For each stimulus category, they found a substantial number of trigger events that occur exclusively in spike trains elicited by stimuli of that category (Eguchi et al., 2018, figure

11). This is presented as further evidence for the emergence of informative polychronous groups.

The result, however, can just as well be explained by stimulus specific firing rates of cells which manifest in completely random spike trains. How often a trigger event is activated then depends only on the firing rates of the cells in the trigger set. On the one hand, if all three cells have a high firing rate for a stimulus, then it is likely that a fair number of triplets out of the large number of spikes emitted by them will have the correct relative timings to "activate" the trigger event. If one of the cells is inactive, on the other hand, then the trigger event is not activated at all. This illustrates how stimulus specific activity or inactivity of cells can also cause the stimulus specific occurrence of trigger events. Thus, it is clear that the specificity of trigger events is also not sufficient proof for the existence of informative polychronous groups that rely on precise spike times.

But, Eguchi et al. (2018) also present evidence for polychronisation that can not easily be explained by informative firing rates. First of all, they report that the amount of spike pair information increases as the STDP time constants  $\tau_C$  and  $\tau_D$  decrease, which makes the synaptic plasticity more temporally specific (Eguchi et al., 2018, figure 8). If the exact times of spikes are truly not utilised by the network and the spike pair information is in fact just a mirror of the information in firing rates, then one would not expect it to benefit from more temporally precise learning. This finding is therefore in conflict with the previous argument that the high levels of spike pair information in Eguchi et al. (2018) are simply a side effect of informative firing rates. However, one has to add that decreasing the STDP time constants also decreases the total amount of change in synaptic efficiency that occurs with each spike. The presumed effect of temporally more specific STDP due to shorter STDP time constants could therefore also be a side effect of generally decreased synaptic plasticity. Thus, the interpretation of these findings is not completely clear without further investigations.

Secondly, Eguchi et al. (2018) report that STDP is able to selectively strengthen synapses that have a more optimal transmission delay. Specifically, they report that for most pairs of cells connected by two synapses with different transmission delays, one of the two synapses is potentiated while the other one is depressed (Eguchi et al., 2018, figure 9). This indicates that the synaptic plasticity does in fact treat synapses differently even if they only vary by a few milliseconds. However, it remains unclear to what extent the network is able to utilise these slight variations in synaptic transmission delays when information is propagated.

In summary, it is clear that the network reported in this paper does not exhibit polychronous groups, that depend on millisecond precise spike times of particular cells. Furthermore, we provided criticism of the analysis methods used in Eguchi et al. (2018). Most importantly we showed that it is indispensable to compare the spike pair PG information obtained from the actual spike train to a baseline obtained from a spike train with random spike times. This was not done in Eguchi et al. (2018), which calls into question whether the authors' hypotheses about polychronisation can be confirmed fully.

### 4.3 Conclusion and Future Directions

While no border ownership cells were found that work as hypothesised in section 1.2, this does not disprove the hypothesis altogether. The hypothesised mechanism is in principle able to explain how border ownership cells might form in the real biological brain and it is still possible that future simulation work will find such a mechanism employed. In the following section we will address how some of the problems discovered in this study might be overcome in future work.

#### Border Ownership Cells

A critical problem is that cells in the first layer receive input from the outer contour of the objects (see figure 3.9c for example), which enables them to determine what kind of border is present at a particular location directly from the feedforward input. To deal with this, first, it has to be ensured that the border ownership cells are actually located at the sight of the retinal location in question rather than being on the outline of the objects, as is the case for the cells in figure 4.1.

Second, one has to make sure that the area of the retina from which a layer 1 cell receives feedforward input is small enough so that it does not also include parts of the outer contour of the objects. To this end, one has to reduce the fan-in-radius for connections from the Gabor input layer to the layer 1 excitatory neurons. However, reducing the fan-in-radius for these connections comes at the cost of losing information about the input stimulus. Note that there are six input cells with different orientation selectivity at each location of the retina. Thus, the representation of the stimulus is very dense in the input layer containing the intensities of multiple orientations at each point on the retina. However, the outlines of the visual objects only cover a small proportion of the retina and the edge sensitive Gabor input cells are therefore active only in this small region of the retina. Consequently, Layer 1 cells receive input only from this small but information dense area of the input layer. To maintain the orientation information that is present in the input layer, multiple layer 1 excitatory cells have to receive input from the same spot in the retina, but from different input cells with different orientation selectivities. Thus, it is important that the receptive fields of neighbouring layer 1 cells overlap to a certain degree, such that they can cover the same region of the retina and might obtain different orientation selectivity for that region through different random connections. However, the smaller the fan-in-radius is for connections from the Gabor input layer to the layer 1 excitatory cells, the smaller is the overlap of receptive fields between two neighbouring cells. Thus, the information about the exact orientation of edges is more likely to be lost. Furthermore, it also means that fewer layer 1 cells receive any input at all since the outlines of the presented objects are only covered by few of the narrow receptive fields of layer 1 cells.

There is, however, a solution to this problem. One can reduce the fan-in-radius to the first layer while also making the first layer more dense (i.e. more cells in layer 1). As a result, more neurons would cover the same area of the retina with high overlap between the receptive fields of neighbouring cells. At the same time, they would still have narrow

receptive fields, which solves both problems. Thinking further, one could also envision a network architecture in which each consecutive layer has fewer neurons than the previous one. This would allow denser early layers while keeping the computational effort constant. Furthermore, it would resolve border issues that lead to an inhomogeneous number of efferent synapses within a layer.

### **Polychronisation**

As explained in the previous section, the main obstacle for the development of polychronous groups is the rather large synaptic time constant  $\tau_g$ . It results in spikes having a long lived effect on the receiving cells which makes it harder for cells to detect coincidences with temporal precision. As outlined in section 1.4, a rather large value was chosen for  $\tau_g$  in order to facilitate temporal trace learning. However, the findings of section 3.3.2 suggest that this mechanism is not crucial for the development of location invariance after all - at least for the stimulus set at hand. Further, one can rely more on the continuous transform learning paradigm for the development of location invariance. Thus it becomes possible to reduce the value of the synaptic time constant, which is likely to greatly increase the networks ability to polychronise.

# Abbreviations

<b>CT</b>	Continuous Transformation (Learning Mechanism)
<b>LIF</b>	Leaky Integrate and Fire (Neuron Model)
<b>LTD</b>	Long Term Depression
<b>LTP</b>	Long Term Potentiation
<b>PG</b>	Polychronous Group
<b>STDP</b>	Spike Time Dependent Plasticity

# Bibliography

- Amit, D. J. and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex *Cerebral Cortex* 7 237-52. *Cerebral cortex (New York, N.Y. : 1991)*, 7:237–252.
- Bi, G. Q. and Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 18(24):10464–10472.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.
- 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.
- Eguchi, A., Isbister, J., Ahmad, N., and Stringer, S. M. (2018). The Emergence of Polychronization and Feature Binding in a Spiking Neural Network Model of the Primate Ventral Visual System. *Psychological Review (under review)*.
- Eguchi, A., Mender, B. M. W., Evans, B. D., Humphreys, G. W., and Stringer, S. M. (2015). Computational modeling of the neural representation of object shape in the primate ventral visual system. *Frontiers in Computational Neuroscience*, 9(August).
- Eguchi, A. and Stringer, S. M. (2016). Neural network model develops border ownership representation through visually guided learning. *Neurobiology of Learning and Memory*, 136:147–165.
- Evans, B. D. and Stringer, S. M. (2012). Transformation-invariant visual representations in self-organizing spiking neural networks. *Frontiers in Computational Neuroscience*, 6(July):1–19.
- Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3(2):194–200.
- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley, New York.
- Izhikevich, E. M. (2006). Polychronization: Computation with Spikes. *Neural Computation*, 18(2):245–282.

- Jones, J. P. and Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1187–1211.
- Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275(5297):213–215.
- Martinez, R. and Paugam-Moisy, H. (2009). Algorithms for structural and dynamical polychronous groups detection. *ICANN'2009, International Conference on Artificial Neural Networks*, 5769:75–84.
- Neuman, Y. (2017). *Mathematical Structures of Natural Intelligence*. Springer.
- Pasupathy, a. and Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *Journal of neurophysiology*, 86(5):2505–2519.
- Pasupathy, A. and Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, 5(12):1332–1338.
- Perrinet, L. U., Delorme, A., Thorpe, S. J., and Samuelides, M. (2001). Network of integrate-and-fire neurons using Rank Order Coding A: how to implement spike timing dependant plasticity. *Neurocomputing*, 38–40(1–4):817–822.
- Petkov, N. and Kruijinga, P. (1997). Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells. *Biological Cybernetics*, 76(2):83–96.
- Rolls, E. T. and Milward, T. (2000). A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural computation*, 12(11):2547–2572.
- Stringer, S. M., Perry, G., Rolls, E. T., and Proske, J. H. (2006). Learning invariant object recognition in the visual system with continuous transformations. *Biological Cybernetics*, 94(2):128–142.
- Troyer, T. W., Krukowski, A. E., Priebe, N. J., and Miller, K. D. (1998). Contrast-Invariant Orientation Tuning in Cat Visual Cortex: Thalamocortical Input Tuning and Correlation-Based Intracortical Connectivity. *Journal of Neuroscience*, 18(15):5908–5927.
- Von Der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron*, 24(1):95–104.
- Wallis, G. and Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, 51(2):167 – 194.
- Zhou, H., Friedman, H. S., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience*, 20(17):6594–6611.

# Declaration of Authorship

I hereby certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other university.

---

signature

---

city, date