Learning Transformation-Invariant Visual Representations in Spiking Neural Networks

Benjamin D. Evans

The Queen's College

University of Oxford

A thesis submitted for the degree of

D.Phil. in Experimental Psychology

Trinity Term 2012

Acknowledgements

Firstly I would like to thank my supervisor, Dr Simon M. Stringer for his support and guidance and many years of friendship. His boundless optimism and enthusiasm swept aside any self-doubts or apprehensions on countless occasions to keep this thesis on track.

A special note of thanks to Amanda for generally looking after me when I barely had the time or energy to take care of myself in the inevitable agony of the deadline.

On a similar note, there are many friends and housemates near and far who have provided moral support, comic relief, an escape for the weekend or simply helped to make this long distance run that little bit less lonely. Tom, Fruity, Graeme, Monkey, Gecko, Andy, Rob, Josh, Dom, Pete, Michelle, Nick, ADK, AKI, Brown, Anna and Claudio — sincerely, thank you.

Thanks also to Jamie for help in preparing the 3D stimuli and along with Dan, Bedeho, Juan and Erica for making the lab an enjoyable place to be, and keeping the 'fraud and abuse' to a minimum!

I would also like to thank the ESRC for funding this research and supporting my privilege to spend yet more formative years living and working in such an intellectually stimulating environment.

Last, but by no means least, I would like to thank my parents for their love and support through the ups and downs of life chasing a doctorate.

Abstract

This thesis aims to understand the learning mechanisms which underpin the process of visual object recognition in the primate ventral visual system. The computational crux of this problem lies in the ability to retain *specificity* to recognize particular objects or faces, while exhibiting *generality* across natural variations and distortions in the view (DiCarlo et al., 2012). In particular, the work presented is focussed on gaining insight into the processes through which transformation-invariant visual representations may develop in the primate ventral visual system.

The primary motivation for this work is the belief that some of the fundamental mechanisms employed in the primate visual system may only be captured through modelling the individual action potentials of neurons and therefore, existing rate-coded models of this process constitute an inadequate level of description to fully understand the learning processes of visual object recognition. To this end, spiking neural network models are formulated and applied to the problem of learning transformation-invariant visual representations, using a spike-time dependent learning rule to adjust the synaptic efficacies between the neurons.

The ways in which the existing rate-coded CT (Stringer et al., 2006) and Trace (Földiák, 1991) learning mechanisms may operate in a simple spiking neural network model are explored, and these findings are then applied to a more accurate model using realistic 3-D stimuli. Three mechanisms are then examined, through which a spiking neural network may solve the problem of learning separate transformation-invariant representations in scenes composed of multiple stimuli by temporally segmenting competing input representations. The spike-time dependent plasticity in the feed-forward connections is then shown to be able to exploit these input layer dynamics to form individual stimulus representations in the output layer. Finally, the work is evaluated and future directions of investigation are proposed.

Long Abstract

This thesis aims to understand the learning mechanisms which underpin the process of visual object recognition in the primate ventral visual system. The computational crux of this problem lies in the ability to retain *specificity* to recognize particular objects or faces, while exhibiting *generality* across natural variations and distortions in the view (DiCarlo et al., 2012). In particular, the work presented is focussed on gaining insight into the processes through which transformation-invariant visual representations may develop in the primate ventral visual system. The primary motivation for this work is the belief that some of the fundamental mechanisms employed in the primate visual system may only be captured through modelling the individual action potentials of neurons and therefore, existing rate-coded models of this process constitute an inadequate level of description to fully understand the learning processes of visual object recognition. To this end, spiking neural network models are formulated and applied to the problem of learning transformation-invariant visual representations, using a spike-time dependent learning rule to adjust the synaptic efficacies between the neurons.

The thesis starts in Chapter 1 by introducing the biological and computational background pertinent to the study of how transformation-invariant visual representations are formed and by providing a review of the associated relevant literature. Next, Chapter 2 details the spiking neural network developed to model the primate ventral visual stream and the methods used to analyse, interpret and visualize the results derived from it.

Chapter 3 investigates the extent to which the rate-coded Hebbian learning mechanisms, of CT (Stringer et al., 2006) and Trace learning (Földiák, 1991), may be implemented with a more biologically accurate spike-time dependent learning rule in a spiking neural network for learning to recognize objects in a model of the ventral visual stream.

Both the Trace and CT learning mechanisms have been previously applied successfully to this problem in rate-coded models to form transformation-invariant representations with both temporal continuity (Wallis and Rolls, 1997; Wyss et al., 2006; Michler et al., 2009; Isik et al., 2012) and spatial continuity (Stringer et al., 2006;

Perry et al., 2010) of the stimuli respectively. The simulations conducted in Chapter 3 explore for the first time how these learning paradigms may be successfully adapted to a spiking neural network featuring STDP in its feed-forward excitatory synapses.

It is shown that with appropriate modification of the stimulus presentation regime and by changing the excitatory synaptic conductance time constant, τ_g^{EfE}, (modelling the dynamics of the postsynaptic ion channels), the model could operate in two dissociable modes to learn translation-invariant representations. In particular, a short synaptic time constant and high degree of spatial overlap between transforms allows the model to operate in a manner akin to CT learning, whereas a long synaptic time constant facilitated a form of Trace learning, whereby the network could associate together completely non-overlapping (orthogonal) views of the stimuli onto the same output neurons.

While both mechanisms are found to be robust to the order of transform presentation within a stimulus block, Trace learning suffers when the transforms of two stimuli were interleaved as expected from a violation of the assumption of temporal continuity of stimuli within natural scenes (Földiák, 1991; Wallis and Rolls, 1997; Sprekeler et al., 2007; Li and DiCarlo, 2008, 2010). The CT learning mechanism is found to require tight synchrony of the input volleys and temporally specific STDP time constants, whereas Trace learning is more robust to spike-timing jitter and performs better with longer STDP time constants.

Following from the simulations of simple stimuli used to investigate the transformation-invariance learning in Chapter 3, the aim of Chapter 4 is to demonstrate that the same principles apply when using more realistic three-dimensional stimuli. Through a combination of the Trace and CT learning mechanisms, it is demonstrated that the network can be trained to recognize both a cone and a cube as they translate across the input layer. Furthermore, it is demonstrated how additional sorts of transformation, such as rotations of view (in depth), may be associated together through the same learning mechanisms thus generalizing the scope of the insights gained from simple abstract stimuli.

Some novel problems arise through using filtered grey-scale images, such as the difficulty in representing particular transforms of low contrast with respect to the grey background, and remedies are discussed along with potentially interesting avenues of investigation which are open to a model using more detailed stimuli.

A central question in understanding the process of biological object recognition is how to form separate representations of objects from natural visual scenes composed of multiple objects. Previous work has largely used stimuli presented in isolation (Wallis

and Rolls, 1997; Stringer et al., 2006; Riesenhuber and Poggio, 1999; Masquelier and Thorpe, 2007), and so is lacking in ecological validity. Rate-coded models in particular suffer in this respect, as a rate-based model of Hebbian learning predicts that stimuli presented together will be associated together leading to a superposition catastrophe (von der Malsburg, 1999). This is a theme addressed in several ways over the course of Chapters 5–7.

In Chapter 5, the standard one-layer feed-forward model is augmented with a lateral connectivity architecture similar to the standard Self-Organizing Map (Kohonen, 1982; Miikkulainen et al., 2005) to address the question of how the extensive lateral connectivity of the visual system may aid learning to recognize objects in natural visual scenes composed of multiple objects. Guided by previous analyses of laterally connected spiking neurons (Nischwitz and Glünder, 1995; Miconi and VanRullen, 2010) it was predicted that the architecture and neuronal properties of the network would enable competing stimulus representations to push each other out of phase. In particular, the hypothesis tested is that with such anti-phase oscillations (perceptual cycles) between input layer representations, a time-sensitive STDP learning rule in the feed-forward synapses will be able to exploit the temporal segmentation of the stimuli as they translate across the input layer to learn separate stimulus representations in the output layer.

The key properties of the network are identified to be a mechanism of self-inhibition, consistent with previous work (Choe and Miikkulainen, 2003; Miconi and VanRullen, 2010), specifically cell firing-rate adaptation, and a gradient in the excitatory synaptic weights (which decrease exponentially with distance). Variations in these properties are investigated with respect to how they enable the network to synchronize features represented by proximal neurons and desynchronize them with respect to features represented by distal neurons, thus facilitating temporal segmentation of the input representations in a way suitable for transformation-invariant learning with STDP.

Following from the conclusions of Chapter 3, specifically that CT learning requires tight synchronization of spikes in an input volley (as slightly later spikes fall into the STDP rule's LTD window) and that Trace learning is less temporally specific (with respect to input synchrony and the STDP time constants), only the CT learning mechanism was explored in these simulations, as a Trace learning paradigm would undermine the dynamics of the temporal segmentation of stimuli.

The work shows how the spatial separation of the stimuli is effectively converted into a temporal separation (through the lateral weight gradient and firing-rate adaptation) in a way that allows separate translation-invariant representations to be formed even

when both stimuli are always presented together (with no statistical decoupling) and moving in lock-step. This is an advancement on previous work with multiple stimulus presentation paradigms in rate-coded networks where, either ecologically implausible training regimes are required to statistically decouple the objects (Stringer et al., 2007; Stringer and Rolls, 2008), or independent motion of the objects is necessary (Tromans et al., 2012).

The extraordinary computational power of neurons lies in their ability to communicate with one another through their synapses. Their connections may be very short or extend across millimetres of cortex (Miikkulainen et al., 2005) and so the delay in conducting their action potentials along their axons (with or without myelination) varies accordingly from the order of one millisecond (Girard et al., 2001) to the order of tens (Swadlow, 1992) or even hundreds of milliseconds (Aston-Jones et al., 1985). Following from these experimental data, Chapter 6 investigates the role of axonal conduction delays, firstly in the feed-forward axons of a variant of the model used in Chapter 3, and secondly as a replacement for the Gaussian profile of lateral excitatory weights used in Chapter 5.

In using a constant axonal conduction delay for the feed-forward connections, both the CT and Trace learning mechanisms are unaffected in their capacity to learn transformation-invariant representations in the output layer when the STDP rule also incorporates the delay term. This is a novel change to the original formulation of Perrinet et al. (2001), which like many other models of STDP, neglects axonal delays (Song et al., 2000). When the unmodified version of STDP is implemented in the model, a pronounced and different effect emerges as the axonal conduction delays are increased between the CT and Trace learning regimes. These are argued to be artefactual however, with the natural conclusion being that when implementing axonal delays in such a model, the STDP rule should also be revised accordingly.

Gaussian distributions of delays in the feed-forward connections are also investigated. Here, another marked difference between the CT and Trace learning mechanisms is discovered, whereby the network performance of CT learning is severely degraded with the increasing standard deviation of the distribution, whereas Trace learning is found to be far more robust, as expected due to their respective differences in sensitivity to the synchrony of the inputs.

The second section of Chapter 6 considers the computational benefits of axonal conduction delays and demonstrates that they can replace the role of the gradient in the excitatory lateral connection strengths to temporally segment multiple stimuli through the dynamics of perceptual cycles. The delays in this section were proportional

to the Euclidean distance between the neurons giving a graduation of delays and thereby extends the simple two neuron case of Nischwitz and Glünder (1995) and the single neural layer of Miconi and VanRullen (2010) which each involved a fixed delay. It is predicted that the optimal delay between the stimulus representations was half the period of oscillation, such that the wave of excitation arrives midway through the cycle of the afferent neurons, encouraging their targets to fire at that point. The results however indicate that a delay $1ms$ lower was very slightly better for network performance, which may be due to the time required for the synaptic conductances to drive up the postsynaptic membrane potential.

In the final body of experimental work, Chapter 7 investigates a third solution to the problem of separating representations of multiple translating stimuli. Rather than imposing a fixed lateral weight structure or attempting to match the lateral delays to the (half) period of oscillation, plasticity is introduced into the lateral connections and a preliminary training phase is introduced to allow them to self-organize through exposure to exemplars of categories. During this preliminary phase, knowledge of the categories is built up in the lateral connections which is then found to enable the network to temporally segment a visual scene composed of two novel exemplars.

While initially the effects of 'early' exemplar exposure are investigated in a single layer of excitatory neurons (and inhibitory interneurons), the second section augments this layer with another and adds feed-forward plastic connections between them, and translating stimuli are presented to the input layer. This model is then used to demonstrate that category learning in the lateral connections allows the output layer to form independent translation-invariant representations for previously unseen, stimulus representations. In addition to the lack of statistical decoupling and independent movement, in contrast to the work of Chapters 5 and 6, the use of distributed stimulus representations meant that spatial separation is not required to temporally segment the stimuli.

The input layer dynamics generated in this model are shown to be robust to wide ranges of several parameters, including the strength of inhibitory and excitatory connections and the adaptation model strength and time constant. The effect upon the subsequent training of the feed-forward connections is also explored, principally through systematically varying the parameters of the STDP model, which is found to be similarly robust to the earlier simulations of this thesis involving multiple stimulus presentation paradigms.

Finally, the thesis concludes with an evaluation of the research presented, a discussion of the themes to emerge throughout the course of the investigations and

some suggestions for future directions in which to progress the field of research.

Contents

List of Figures

List of Tables

Acronym	Meaning	Alternative
AHP	After Hyperpolorization Potential	
aIT	Anterior Inferotemporal Cortex	aITC, AIT, TE
BCM	Bienenstock-Cooper-Munro (rule)	
CT	Continuous Transformation (Learning)	
DFT	Discrete Fourier Transform	
DoG	Difference of Gaussians	
EPSP	Excitatory Postsynaptic Potential	
FFA	Fusiform face area	
FFT	Fast Fourier Transform	
fMRI	Functional Magnetic Resonance Imaging	
gLIF	Conductance-based LIF	
HH	Hodgkin-Huxley	
ICA	Independent Component Analysis	
IF	Integrate-and-Fire	
IPSP	Inhibitory Postsynaptic Potential	
IT	Inferotemporal Cortex	ITC, TE, Brodmann's area 20, 21
LGN	Lateral Geniculate Nucleus	Lateral Geniculate Body, LGB
LIF	Leaky Integrate-and-Fire	
LIP	Lateral Intraparietal Area	
LTD	Long-Term Depression	
LTP	Long-Term Potentiation	
MST	Medial Superior Temporal Area	
MT	Middle Temporal Area	V5, Brodmann's area 19
NMDA	*N*-Methyl-D-aspartate (receptor)	NMDAR
ODE	Ordinary Differential Equation	
OFA	Occipital Face Area	
PDE	Partial Differential Equation	
pIT	Posterior Inferotemporal Cortex	pITC, PIT, TEO
PSP	Post-synaptic Potential	
PSTH	Post-stimulus Time Histogram	Peri-stimulus Time Histogram
RBF	Radial Basis Function	
RC	Resistor-Capacitor	
rOFA	Right Occipital Face Area	
SFA	Slow Feature Analysis	
SNN	Spiking Neural Network	
SOM	Self-Organizing Map	
SRM	Spike-Response Model	
STDP	Spike Time Dependent Plasticity	Spike Timing Dependent Plasticity
STP	Short-Term Potentiation	
TE	See AIT	
TEO	See PIT	
TMS	Transcranial Magnetic Stimulation	
UTL	Unsupervised Temporal Learning	Trace Learning
V1	Primary (striate) visual cortex	Brodmann's area 17
V2	Visual Area 2 (Prestriate Cortex)	Brodmann's area 18
V4	Visual Area 4	V8
V5	See MT	

We are so familiar with seeing that it takes a leap of imagination to realize that there are problems to be solved. But consider it. We are given tiny distorted upside-down images in the eyes, and we see separate solid objects in the surrounding space. From the patterns of stimulation on the retina we perceive the world of objects and this is nothing short of a miracle.

—Richard L. Gregory

1 Introduction

The work presented in this thesis aims to understand how the primate visual system learns to recognize objects and faces irrespective of the details of a particular view. In particular, I seek to show how spiking neural network models provide a more appropriate level of description for this self-organizational process than much of the existing work on transformation-invariance with rate-coded models. This chapter aims to set the context for the work presented in subsequent chapters, by first explaining the challenge of the visual recognition process (that this work addresses) in detail and then explain the rationale for the approach taken in attempting to understand this process. There then follows a broad overview of what is known of the structure and function of the ventral visual system in primates, followed by a review of other modelling work of this process and the learning mechanisms discovered and applied in such work. Having reviewed the pertinent body of work, an overview is then provided of the research undertaken for this thesis and how it contributes to the overall understanding of this field.

1.1 Visual Recognition

Typically without needing a moment's conscious thought, we recognize objects and faces irrespective of variations in their size and position or our perspective of them. How the brain forms such representations capable of serving to identify objects invariantly with respect to many differences from a pair of warped, two-dimensional retinal inputs

is an interesting and non-trivial problem (Gregory, 1966; Pinto et al., 2008). The raw pattern of stimulation on the retinae of the viewer may be very different from one view to the next, (possibly orthogonal) with even the smallest of movements or occlusions, yet the ability of primates to recognize conspecifics and shapes despite these transformations is remarkably automatic, rapid and reliable (Thorpe et al., 1996).

Object recognition may be defined as 'the ability to accurately discriminate each named object ("identification") or set of objects ("categorization") from all other possible objects, materials, textures, other visual stimuli and to do this over a range of identity-preserving transformations of the retinal image of that object' (DiCarlo and Cox, 2007). Despite the apparent ease, lack of conscious thought and speed (Thorpe et al., 1996; Hung et al., 2005; Afraz et al., 2006) with which we accomplish this, object recognition represents a very difficult computational problem due to the infinite differences in retinal images owing, for example, to variations in stimulus position, distance, pose, view, illumination and occlusion (Pinto et al., 2008). While retaining 'specificity' to the identity of particular objects, a visual system (either natural or artificial) must therefore simultaneously exhibit 'generality' across natural variations in the input. This 'invariance problem' is the computational crux of visual object recognition and the major difficulty for artificial vision systems (DiCarlo et al., 2012).

1.1.1 Biological Object Recognition

Vision is clearly a highly important sense for many species of animal, with up to 50% of the neocortex of non-human primates devoted to visual processing (DiCarlo et al., 2012). Many aspects of our evolutionary ancestors' survival, such as the ability to find food, select tools and avoid predators depended upon the rapid and accurate ability to extract object identity from the raw patterns of light falling on their retinae (DiCarlo et al., 2012). As a social species of mammal, it is also particularly important to be able to recognize other individuals over a long period but with the flexibility to adapt the recognition process based upon long-term changes in an individual's appearance or the sets of faces in our particular social environment (Webster et al., 2004). It is therefore an adaptive system and one which learns useful representations in an unsupervised manner (Li and DiCarlo, 2008, 2010), independent of reward (Li and DiCarlo, 2012). As almost a pre-requisite for meaningful social interaction in highly visual species, it is therefore natural that this process should be robust for primates despite the multitude of possible variations in the view of an individual's face but perhaps surprising how difficult it remains for artificial approaches.

Aside from the subjective feeling of recognition, there is a confluence of empirical data demonstrating the ability of the primate and human ventral visual systems (V1→V2→V4→IT, see Figure 1.1) to reliably identify objects and faces across a wide range of conditions. Evidence comes from a growing and disparate group of disciplines including recording and stimulating techniques. Several of the earlier key studies of occipital lobe structures (such as V1) investigated the cat visual cortex (Hubel and Wiesel, 1962), whereas invasive techniques (especially for IT recordings) more commonly use the macaque (*Macaca mullata*), a species of 'Old World' monkey[1] as the model for the primate visual system (Hubel and Wiesel, 1968).

While the macaque's visual system exhibits differences in its anatomy to ours, it shows striking similarities to other primates including the owl monkey, squirrel monkey and bushbaby (Felleman and Van Essen, 1991) accounting for it being the most intensively studied non-human primate (Hegdé and Van Essen, 2007). Furthermore, striking functional similarities have been found between the macaque and human visual systems, for example in the representation of object categories (Tsao et al., 2003) revealed through fMRI, and thus the macaque has become the standard model of the primate visual system. Non-invasive techniques have used human participants where possible, particularly for imaging studies and techniques of indirect stimulation and temporary ablation, and occasionally direct microelectrode recording while patients undergo brain surgery e.g. for pharmacologically intractable epilepsy (Quiroga et al., 2005). Here follows a review of such sources of evidence and explanations of the inferences drawn from them.

One of the primary sources of evidence for the brain's ability to consistently recognize objects is the corpus of single-cell recordings. These neurophysiological studies have revealed that over successive stages of the physiological hierarchy (Felleman and Van Essen, 1991), the primate visual system develops neurons which respond to increasingly complex stimuli (Kobatake and Tanaka, 1994; Tanaka, 1996) from local spots of contrast (Kuffler, 1953), to oriented bars (Hubel and Wiesel, 1959), to oriented angles and curves (Hegdé and Van Essen, 2000), to specific curvatures (Connor et al., 2007) and eventually faces (Desimone, 1991; Tsao et al., 2003). These cells also have also been found to have increasingly large receptive fields (Freeman and Simoncelli, 2011; Kobatake and Tanaka, 1994), with ratios of approximately 1:6:16 for V1:V2:V4 (Pasupathy, 2006). This allows the cells in the temporal lobe (see Figure 1.1, shown

[1] 'Old World' monkeys refers to those native to Africa and Asia (and prehistorically Europe) which differ from 'New World' monkeys (native to Central and South America) most notably by their tails not being prehensile.

in red) to take in information from across the entire visual scene and makes them ideally suited to develop the type of invariant representations observed.

While fMRI is not able to achieve the same spatial resolution as cellular recordings with electrodes, and hence is unable to show identity specific responses, it has been used to demonstrate the functional organization of the macaque visual processing areas. Very similar anatomical differences to the human brain are revealed in the representation of different stimulus categories, for example between human faces, monkey faces, or bodies, from which the category may be reliably decoded (Tsao et al., 2003). Further specialization of face processing areas was subsequently revealed also through fMRI analysis, indicating some areas have view-specific responses, some have a mirror-symmetric tuning (achieving partial view invariance) and other patches have almost full invariance (Freiwald and Tsao, 2010).

Further evidence of cell response properties (and thus confirmation of the changes in the nature of representations throughout the visual processing stream) come from experimental manipulation (stimulation) of cells either directly with microelectrodes (Brindley and Lewin, 1968) or indirectly with non-invasive techniques such as Transcranial Magnetic Stimulation — TMS (Cowey and Walsh, 2000). Here again, there is evidence for a hierarchical processing stream, with gradually more complex representations being constructed from topographically organized features (Felleman and Van Essen, 1991), such as simple cells in V1 (Riesenhuber and Poggio, 1999, 2000, 2002).

Microstimulation of face-cell clusters in Macaque IT biased them to choose 'face' in a 'face/non-face' discrimination task, demonstrating a causal relationship between IT neuron activity and visual object perception and categorization (Afraz et al., 2006). TMS has also been used to temporarily impair face perception (Pitcher et al., 2009) by stimulating the right Occipital Face Area (rOFA) — part of the Fusiform gyrus in the temporal lobe. Interestingly, a triple dissociation was found in impairing processing of faces, bodies and objects for three close areas consistent with a modular organization of object categories in the cortex (Pitcher et al., 2009).

In accordance with the increasingly large receptive field size and growing complexity of cell stimulus preferences along the ventral pathway (Kobatake and Tanaka, 1994), the response latency has been found to increase along the pathway, from approximately $40\,ms$ in the LGN to around $100\,ms$ in anterior Inferior Temporal cortex (see DiCarlo et al., 2012, Figure 3), commensurate with the hypothesized flow of information (VanRullen and Thorpe, 2002). At the end of the ventral visual processing stream the Inferior Temporal visual cortex (IT) of macaques, neurons have been found which

respond with translation (Tovée et al., 1994; Op de Beeck and Vogels, 2000; Hung et al., 2005), size (Ito et al., 1995; Hung et al., 2005) and view invariance (Booth and Rolls, 1998; Logothetis et al., 1995) to objects (Tanaka et al., 1991) and faces (Desimone, 1991). It is the response properties of these cells that the present work seeks to reproduce.

Together, these techniques provide a comprehensive and coherent account of the cell response properties of neuron populations at various stages along the ventral visual stream, and as such, may give some clues as to how invariant representations of objects are constructed by gradually transforming the visual representation (DiCarlo and Cox, 2007). They do however still leave explanatory gaps, offering no real insight into the mechanisms by which this invariance is achieved, or why the visual system is organized in this way. In order to go beyond a mere description, a precise model must be articulated in order to test specific hypotheses about the structure and mechanisms required to accomplish the task of object recognition (as discussed in Section 1.3) but first each layer in the feature hierarchy of the (ventral) visual system is discussed in detail.

1.2 The Primate Visual System

A widely accepted view of the primate visual system known as the 'two-streams hypothesis' was first proposed by Ungerleider and Mishkin from finding a double dissociation between object or location tasks following focal lesions in temporal or parietal lobes respectively (Underleider and Mishkin, 1982) and holds that the processing of information is segregated into two parallel streams known as the 'What' and 'Where' stream (Ungerleider and Haxby, 1994) supporting 'perception' and 'action' respectively (Goodale and Milner, 1992).

While both pathways naturally have their origins in the phototransducers of the retinae and pass through some of the same structures initially, a degree of segregation already appears evident even within the very first structures they have in common. Processing is segregated into magnocellular and parvocellular layers within the LGN, the outputs of which then proceed to the back of the brain where they synapse with neurons in V1 (Felleman and Van Essen, 1991). From here, two separate pathways are easily identifiable — the dorsal stream ('Where') which continues over the top of the brain and the ventral stream ('What') which continues along the underside.

The dorsal stream receives inputs mainly from the magnocellular pathway and has been identified as V1→V2→V3→MT feeding into the (posterior) parietal lobe. It is

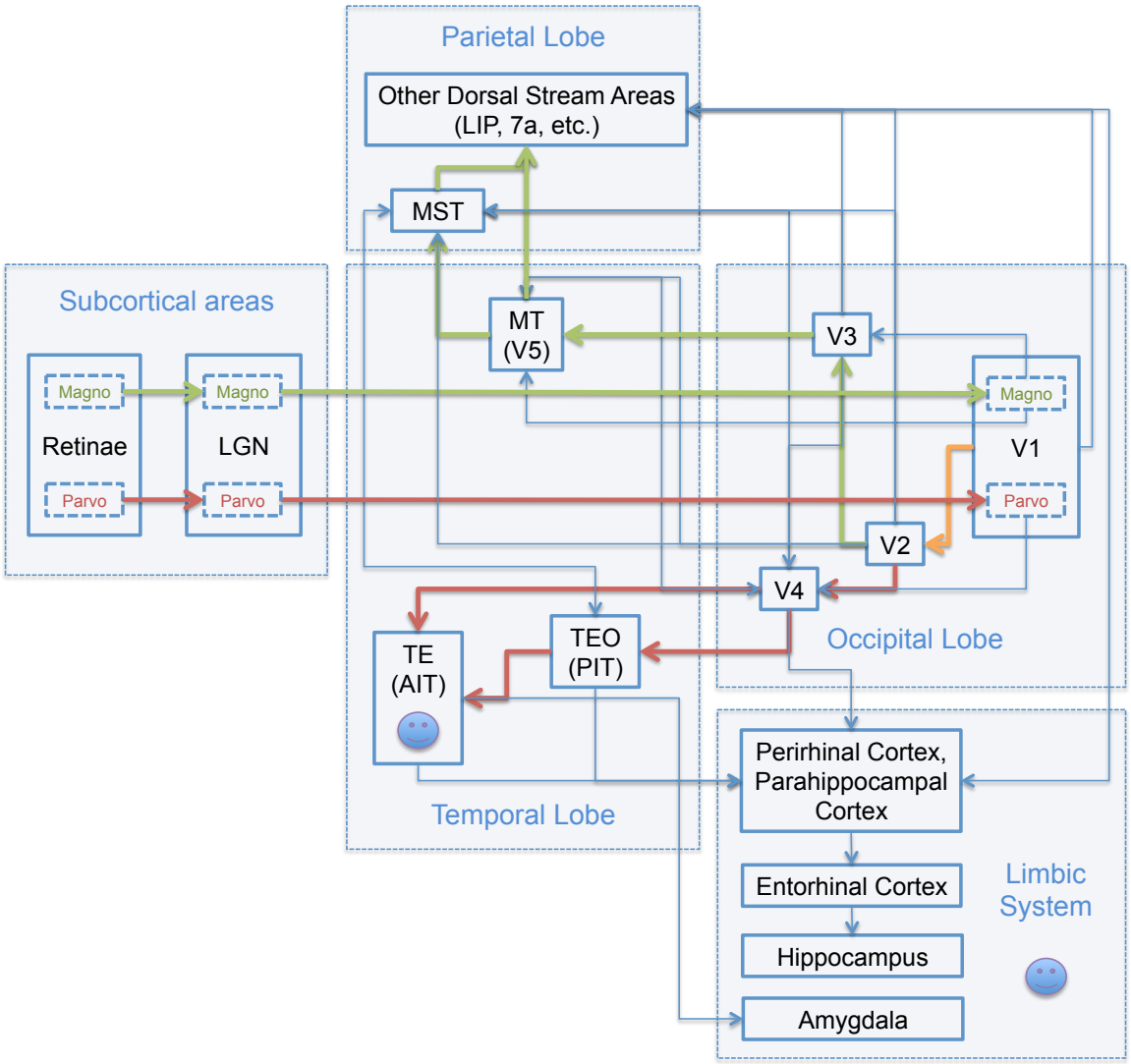


Figure 1.1: Schematic of the Primate Visual System. This figure represents a significant simplification of the true complexity of the visual processing areas and their connectivity (especially back-projections and subdivisions within areas) but it serves to illustrate the main (afferent) projections of the dorsal stream (shown in green) and the ventral stream (shown in red), adapted from Gross et al. (1993) with further detail from Felleman and Van Essen (1991) and Lehky and Sereno (2007). Areas discovered to contain face-selective cells are marked, including areas of the limbic system (Medial Temporal Lobe) including the Hippocampal formation and Amygdala (Quiroga et al., 2005).

believed to be largely concerned with processing spatial and movement information, whereas the ventral stream is concerned with form and colour perception and consists of the cortical pathway V1→V2→V4→IT. Figure 1.1 illustrates the main feed-forward projections and structures involved in these two visual processing streams.

While the 'two-streams' hypothesis has gained general acceptance, the functions of each are possibly not as cleanly separated as first thought. There is some evidence that the ventral stream retains some information pertaining to location which may be decoded from suitable populations of IT cells using linear classifiers (Hung et al., 2005). Similarly, some identity information is retained in the dorsal stream where it was found that some neurons in the lateral intraparietal area (LIP) in posterior parietal cortex respond differently to two-dimensional shapes, irrespective of their size (Sereno and Maunsell, 1998). However, direct comparison shows that the information available for shape discrimination in the dorsal stream is not nearly as powerful as that found in the ventral stream (Lehky and Sereno, 2007).

While there may be some cross wiring at similar levels of each hierarchy (Felleman and Van Essen, 1991) and some resultant shared degree of information between them, the 'two-streams' hypothesis is still a broadly accurate model of visual processing in the primate visual system and has served as a useful framework for much of the empirical and theoretical work relevant to this thesis. Since the functions of interest to this work concern form perception (that is object recognition), further discussion is restricted primarily to the ventral pathway.

There follows an introduction to each of the areas of cortex that form part of the ventral processing stream, and the role that each appears to play in processing the visual information. "From differences in responses at successive stages in the pathway one may hope to gain some understanding of the part each stage plays in visual perception" (Hubel and Wiesel, 1959).

1.2.1 The Eye and Optic Nerve

The function of the eyes is to maintain a clear, focused image of the external world projected onto each retina. They are directed to fixate upon objects of interest by six pairs of orbital muscles (3 antagonistic pairs) which swivel and twist the eyes in their sockets (Carlson, 2000). The eyes channel and focus the light reflecting off the environment and the objects contained within it onto the sensory transducers of the retinas. Here, is the point of contact with the external world where electro-magnetic radiation in the visual spectrum is transduced into electrical energy in the specialized nerve cells lining the back of the eyes, from which point information about the visual scene may be extracted and processed.

The main structures of the eye are illustrated in Figure 1.2. Light first passes through the conjunctiva (a protective skin) and then through the cornea which is responsible for around ⅔ of the bending which helps to steer the course of the light

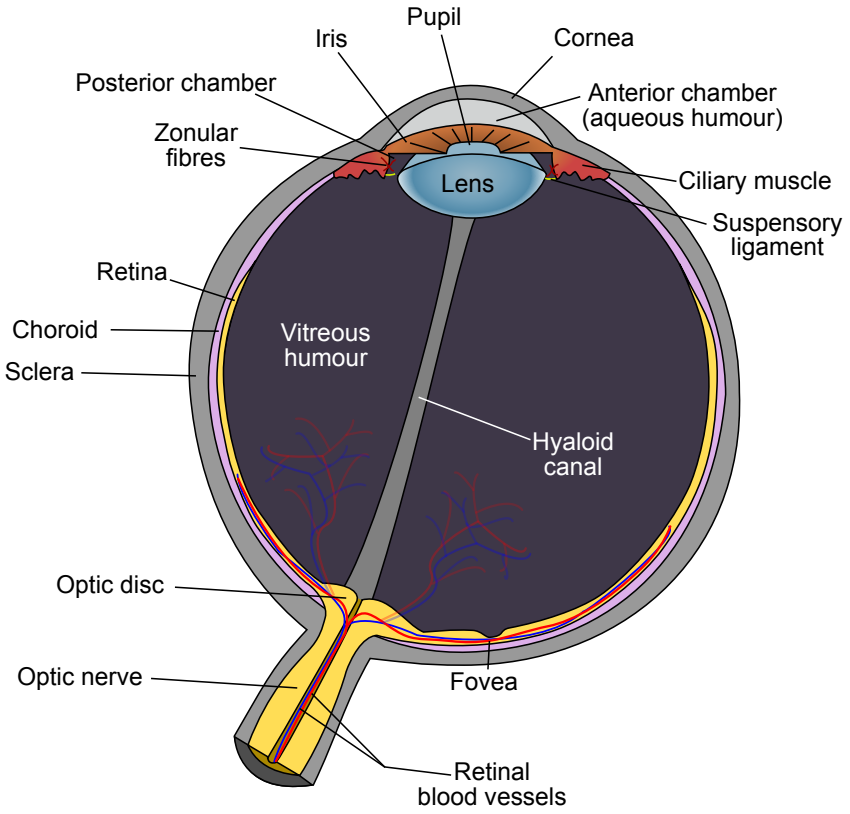


Figure 1.2: Schematic diagram of the human eye (freely reproduced under the Creative Commons license).

through the aqueous humour and onto the lens. The iris controls the size of the pupil, helping to prevent damage to the eye in very bright conditions by shrinking the aperture, and facilitating vision in low light conditions by dilating. Ciliary muscles manipulate the shape of the lens in order to bring objects at a range of distances into focus at the back of the eye on the retina (specifically the *fovea centralis*) becoming more spherical for near objects and flatter for more distant ones (Carlson, 2000).

Each retina is a plate (about $^1/_4mm$ thick) consisting of three layers of nerve cells and two layers of their synapses and dendrites between them with the last layer comprised of approximately 120–125 million photoreceptors (Hubel, 1995). Rod cells are roughly 20 times more numerous than cones (Carlson, 2000) and are specialized for low-light (scotopic) vision as they are sensitive enough to detect a single photon of light, although approximately 5 must be detected for the flash of light to be 'seen' (Hecht et al., 1942). They are distributed more towards the periphery of the retina (Osterberg,

1935) and become saturated (and therefore uninformative) in bright conditions.

Cones are less sensitive, becoming operational in light (phototopic) conditions too bright for rods, and allow for much finer discrimination of detail, essentially giving the benefits of two visual systems. Most species of primates have trichromatic visual systems, meaning that there are three varieties of cone, each containing a different photosensitive pigment with a different spectral sensitivity curve. The cones acquire their names from the wavelengths of the peaks of these curves: Short (β) 420–440 nm, Medium (γ) 534–545 nm and Long (ρ) 564–580 nm (Kandel et al., 2000). The cones allow the visual system to discriminate between different wavelengths of light (independently of intensity) and hence grants the organism colour vision. Cones tend to be distributed towards the centre of the retina and are most densely packed in the fovea (where rods are completely absent) such that visual acuity is at its highest for this small indentation in the retina (roughly one mm across) directly behind the centre of the lens (Osterberg, 1935). Behind the retina is the choroid layer, composed of blood vessels to oxygenate and nourish the retina and cells rich in the dark pigment melanin to absorb any light not captured by the photoreceptors, serving to prevent any back-scattering which would otherwise trigger the extremely sensitive photoreceptors (Hecht et al., 1942) and thus impair vision.

The layer of cells immediately in front of the rods and cones consists of bipolar, horizontal and amacrine cells as illustrated in Figure 1.3. Horizontal cells have relatively long dendrites running parallel to the retina layers which link the transducers with the bipolar cells and similarly amacrine cells link bipolar cells with ganglion cells, accounting for the antagonist surround of retinal ganglion cell receptive fields (Kuffler, 1953). They have also been proposed to adjust the spatio-temporal receptive fields of ganglion cells to adapt to familiar stimuli and enhance sensitivity to novel stimuli through anti-Hebbian learning (Hosoya et al., 2005).

Bipolar cells receive inputs from the transducers and, in most cases, feed directly into the retinal ganglion cells in front of them, which form the third layer of retinal cells (innermost with respect to the curvature). Unusually they have graded (analogue) potentials rather than the digital impulses of axon potentials found ubiquitously throughout the rest of the nervous system (Kandel et al., 2000). This is a curiously backward organization (since it means that light may be blurred by passing through two layers of cell bodies and two layers of dendrites in order to reach the photoreceptors) and illustrates the way evolution tweaks existing phenotypes rather than starting afresh with an optimal design, (as may be found, for example, in some species of

octopus which have the photoreceptors projecting into the eye with the dendrites growing neatly behind them).

The final synapses in the eyes are made at the ganglion cells, whose dendrites are gathered together into a bundle of nerve fibres which plunge back through each retina to form the optic nerves leaving the eyeballs, at which point each eye has a blind spot due to the absence of photoreceptors to make way for the nerve fibres (Carlson, 2000). Although there are only 1 million ganglion cells ($\sim 0.8\%$ the number of photoreceptors), the horizontal and amacrine cells provide an indirect route for the visual information (in addition to the direct route) as Santiago Ramón y Cajal traced in 1900 (Figure 1.3). This indirect path means that the photoreceptors which ultimately stimulate a ganglion cell (through direct and indirect pathways) may be up to 1 mm apart, consistent with the measured receptive field size of a retinal ganglion cell (approximately 3.5°) although this can be much smaller in the fovea.

In and around the fovea (the macular), the mapping from cones to ganglion cells (the direct path) is one-to-one in order to maintain visual acuity with receptive field centres as small as the centre-centre spacing of the cones ($2.5\,\mu m$ or $0.5'$). Towards the periphery however, more photoreceptors converge upon each ganglion cell, accounting for the 125:1 overall ratio of receptors to ganglion cells, and allows light stimulation to be summed over a greater spatial extent making the periphery more sensitive in low light conditions (Hubel, 1995).

The mix of excitatory and inhibitory connections formed through these direct and indirect pathways gives rise to the two types of retinal ganglion cells first identified in the eye of Limulus: on-centre cells are excited by light shone in a central region and inhibited by light shone in the surrounding ring; whereas the equally numerous off-centre cells are functionally opposite in terms of their receptive fields (Kuffler, 1953). The receptive fields of the recorded ganglion cells were found to be usually concentric and covering an area of 1–2 mm or more, shrinking under high background illumination and expanding during dark adaptation (Kuffler, 1953). Diffuse, even illumination evokes little change in response, suggesting that the centre and surround regions are approximately balanced in order to remove the 'DC' component (average value) of the illumination signal. Based upon the original data, the receptive fields of retinal ganglion cells are commonly described as 'Difference of Gaussian' functions with centres and surrounds which sum to 0 (Rodieck, 1965), elegantly capturing much of the structure and function of their prototypical biological counterparts.

These functionally opposing pairs are a common feature of sensory systems, which has the advantage of minimizing firing rates and therefore energy expenditure by

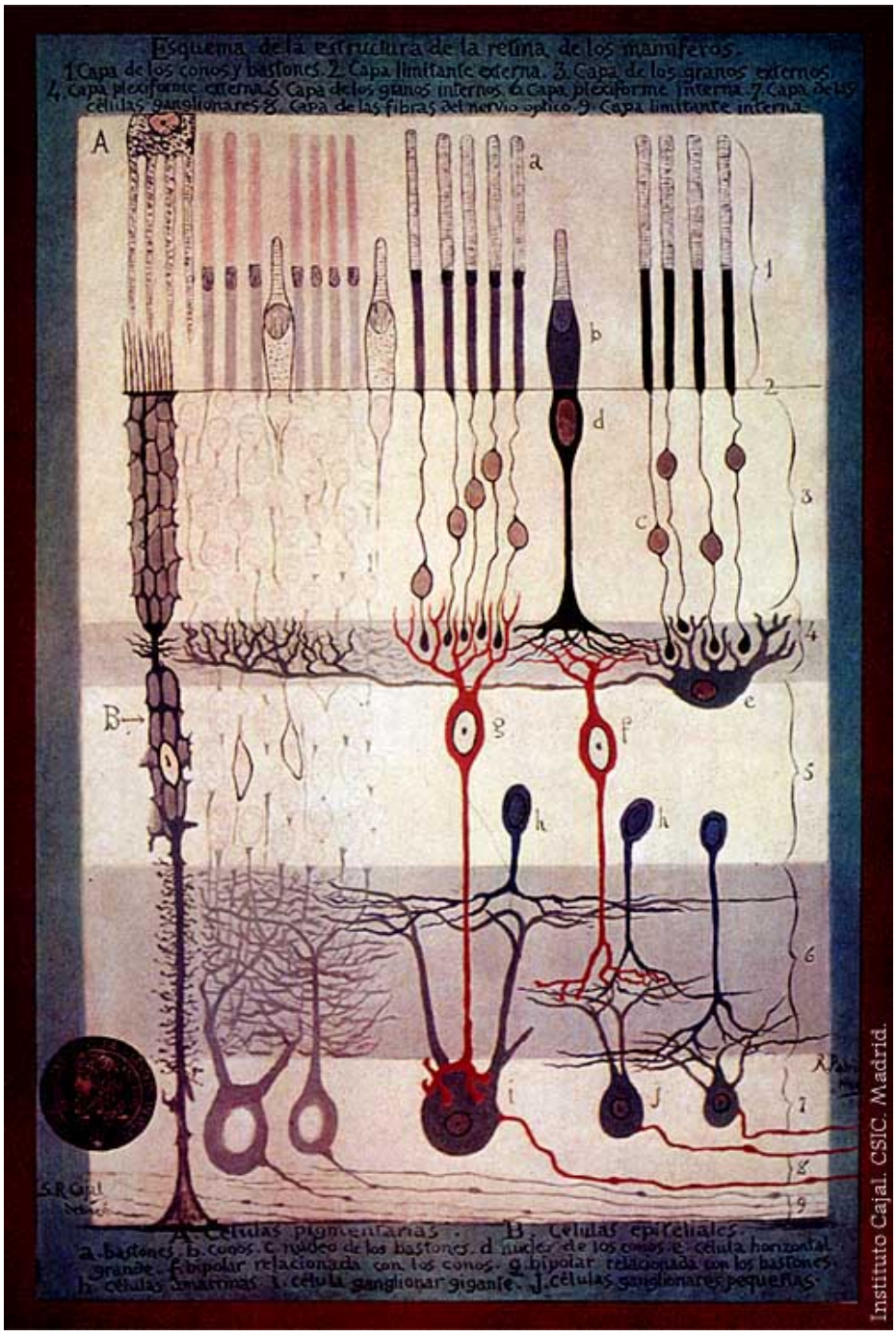


Figure 1.3: The Mammalian retina, from "Structure of the Mammalian Retina" c.1900 by Santiago Ramón y Cajal (freely reproduced under the Creative Commons license). 1: Rod and Cone layer; 2: Outer nuclear layer; 3: Granule layer; 4: External plexiform layer; 5: Inner granular layer; 6: Inner plexiform layer; 7: Ganglion cell layer; 8: Optic nerve fibre layer; 9: Internal limiting membrane. A: Pigmented cells; B: Epithelial cells. a: Rods; b: Cones; c: Rod nucleus; d: Cone nucleus; e: Large horizontal cell; f: Cone-associated bipolar cell; g: Rod-associated bipolar cell; h: Amacrine cells; i: Giant ganglion cell; j: Small ganglion cells.

reducing the range of intensity the neurons need to encode. Rather than having a moderately high firing rate which may move up or down to signify light or dark spots, two opponent systems are used which may both have low background firing rates which rise upon detection of their preferred features. Note however, the background firing rate in dark-adapted cat ganglion cells was found to be typically 20–30 spikes/s (Kuffler, 1953), allowing the cell to signal the inhibitory effect of focused stimulation of its OFF region. Conversely, discharge frequencies of 200–800 spikes/s were found to be within the normal range for the cat's eye under stimulation (Kuffler, 1953).

Similarly, the antagonistic surround of the ganglion cell's receptive fields confers another useful and common property to the visual system. The absolute intensity (or wavelength) of the light at a particular location in the visual field is of little interest — it is the relative intensity (or wavelength) compared with the surrounding locations which is most informative. This way the visual system signals meaningful changes and may therefore economise on firing over uniform areas of light by firing only at background rates for those areas and signalling the changes only at the edges of such areas. A further advantage of the antagonistic surrounds of ganglion cells is that variations in ambient light levels are automatically compensated for, since only relative differences are signalled, thus preserving the appearance of objects.

Having left the back of the eyes, the optic nerve later separates such that the nerve fibres of the right eye coming from the ganglion cells on the left (which are stimulated by the right side of the visual field since the eyes invert the images) cross over (at the Optic Chiasm) to the left where they join the half of the nerve fibres from the left eye's optic nerve which are also stimulated by the same side (right) of the visual field (Hubel, 1995). The optic nerve of the left eye divides in a symmetrical way such that after re-routing each optic nerve, the fibres on the left side of the brain carry only information from the right side of the visual field (aggregated from both eyes) and vice-versa (Miikkulainen et al., 2005). Since each side of the brain controls the movements in the contralateral side, this means that the visual information pertinent to each half of the body is in the right place. Having reorganized the outputs of the eyes, some of the optic nerve fibres continue to structures controlling eye movements and the pupillary light reflex, while for most, the first synapse in the brain is in one of the two lateral geniculate bodies (nuclei).

1.2.2 The Lateral Geniculate Nucleus

At the Lateral Geniculate Nucleus (LGN), the optic nerve fibre fans out in an orderly way such that retinotopic mapping is preserved (likewise the case in V1) as evidenced

by damage to V1 causing scotomas ('blindness' in the corresponding part of the visual field) as though the retina stimulated by that part had been destroyed. Each nucleus is arranged in six layers (with each layer being four to ten cells deep) which are folded slightly and clearly visible through cell-staining techniques. It is from these distinctive folded layers that the LGN gets its name (derived from the Latin 'genu' for knee), as they were likened to a folded knee. Each nucleus receives inputs from one half of the visual field, with the left halves of each retina projecting to the left nucleus etc., (Miikkulainen et al., 2005) and each of the layers receives its inputs exclusively from one eye only with the order (starting from the top of the convexity and working down): right, left, right, left, (constituting the parvocellular layers) left, right (the magnocellular layers) (Kandel et al., 2000).

The LGN also receives inputs from the brainstem reticular formation (which plays a role in attention and arousal) and back-projections from cerebral cortex and there are also lateral connections within the LGN. Cell response properties are, at this stage of processing, still very similar to those of ganglion cells, consisting of a mix of on-centre and off-centre receptive fields with similar responses to colour. Most of the *optic radiations* leaving the LGN project to the primary striate cortex (V1) where cell response properties start to become more complex.

The retinal ganglion cell LGN receptive fields are essentially point-wise spatial sensors. This is an important first stage of visual processing, particularly for forming invariant representations, as the centre-surround opponency allows them to negate illumination and colour changes affecting the whole visual scene and instead save energy by only signalling changes at contrast or colour borders. According to the theory of Hubel and Wiesel (1962), the outputs from such cells may be pooled by aligning their centre regions along an axis to create a degree of orientation selectivity (or tuning) in the form of 'edge detectors' (simple cells) at the next level of processing. Similarly, such simple cells are suggested to be pooled across different locations through convergent feed-forward connections in order to generate location invariant complex cells (Riesenhuber and Poggio, 1999, 2002). Indeed, the selective convergence of simpler features from a previous stage to produce more complex selectivity has been suggested as a general mechanism for all ventral stream shape transformation (Riesenhuber and Poggio, 1999).

1.2.3 Primary striate cortex: V1

Sitting at the back of the brain surrounding the calcarine fissure, the primary striate cortex (V1 or Brodmann's 'Area 17') is common to both the ventral and dorsal visual

processing streams, and is possibly the most studied visual area. It is roughly $2\,mm$ thick in humans and composed of approximately 140 million neurons per hemisphere arranged in 6 layers (Kandel et al., 2000). In each hemisphere, V1 receives its inputs directly from the ipsilateral lateral geniculate nucleus meaning that, like the LGN and superior colliculus, V1 receives information exclusively from the contralateral half of the visual field. There is also a precise retinotopic mapping such that the lower bank of the calcarine fissure responds to the upper half of the visual field and vice-versa (Kandel et al., 2000).

The inputs to V1 synapse principally in Layer 4 (which is itself composed of four sublaminae; $4A, 4B, 4C_\alpha$ and $4C_\beta$) and consist of cells from both the magnocellular and parvocellular pathways (which largely originate from rod and cone cells respectively). Tracing axons in this area in monkeys reveals that segregation of the pathways is maintained at this stage, with magnocellular cells terminating in Layers $4B$ and $4C_\alpha$ while parvocellular cells terminate in Layer $4A$ and $4C_\beta$ (Felleman and Van Essen, 1991). Inputs are also received from the interlaminar region of the LGN (from Koniocellular ganglion cells innervated by 'blue' cones) which terminate in Layers 2 and 3 where they innervate Cytochrome oxidase 'blobs' (Hubel, 1995). Cells in blobs are tuned to low spatial frequencies and colour, making them functionally different to those in the interblob regions, leading to a further subdivision of the parvocellular pathway — the parvo-blob (P-B) and parvo-interblob (P-I) streams (Felleman and Van Essen, 1991).

The 100-fold increase in the number of primate V1 neurons relative to the number of retinal ganglion cells feeding into it (Pinto et al., 2008) and the effect of lesions here resulting in scotomas, suggests that it is a very important area for visual processing. Note that while there is a complete loss of *conscious* awareness of anything in the corresponding part of the visual field (scotoma) from damage to V1, it was discovered that some visual information is still conveyed by subcortical structures enabling patients with brain damage in this area to 'guess' at better than chance level whether a stimulus is present in their blind field — a phenomena discovered by Larry Weiskrantz known as 'blindsight' (Cowey and Stoerig, 1991).

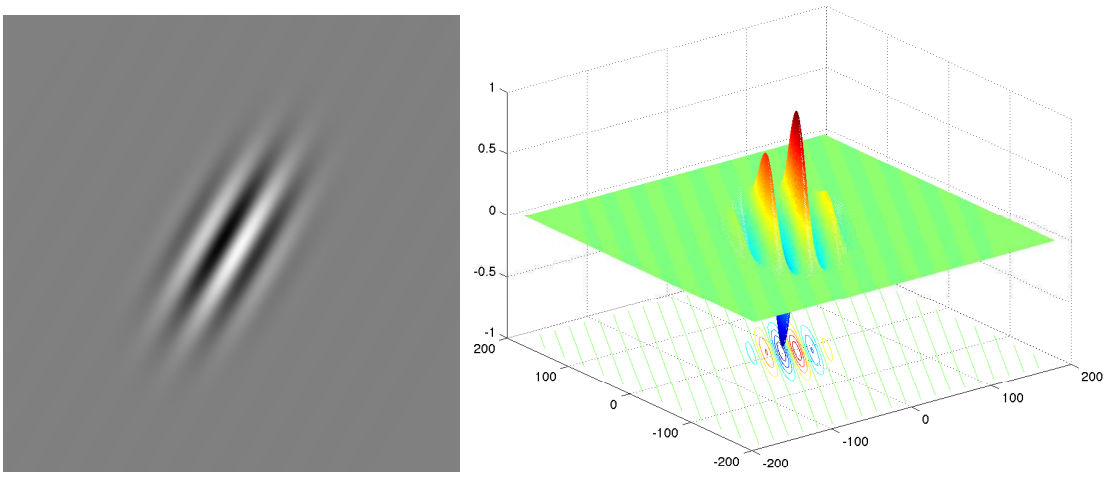
Studies injecting retrograde tracers have also found extensive feedback projections from higher ('downstream') cortical areas in both the ventral and dorsal stream, and even the auditory (A1) and multisensory (STP) cortices (Clavagnier et al., 2004). These feed-back projections are less well understood than the feed-forward connections of the conventional hierarchical hypothesis, but appear to have a modulatory effect upon

the responses of V1 cells and are believed to account for contextual and extra-classical receptive field effects.

One of the most important discoveries about V1 comes from early work with single-cell recordings, where neurons were found to have specific response properties quite different to the circular opponency receptive fields of ganglion and LGN cells. V1 cells (in layers $4C_\alpha$ and $4C_\beta$) in the cat (Hubel and Wiesel, 1962) and monkey (Hubel and Wiesel, 1968) fail to respond to diffuse light in their receptive fields, but prefer bars and edges of specific sizes, orientations and contrasts. Cells in V1 also have a preference for the eye to which the stimulus is presented, changing consistently across the cortex and organized into groups known as ocular dominance columns (Hubel and Wiesel, 1968).

These cells sensitive to edges or bars have come to be known as 'simple cells'. Hubel and Wiesel also discovered two more classes of cells known as 'complex' (which may include motion or short-range translation invariance) and 'hypercomplex' cells (which are sensitive to oriented contour with termination), both of which are believed to be built from the outputs of simple cells. Through building cell tuning properties like these from the receptive field alignment of afferent ganglion cells (Hubel and Wiesel, 1962), it is now well established that the role of V1 is to extract local orientation and frequency information (Hubel and Wiesel, 1968; Connor et al., 2007).

Much of the work conducted on characterizing V1 neurons' response properties (and often those of higher cortical areas) has involved the use of artificial stimuli. Initially these were oriented bars but are more recently oriented sinusoidal gratings within a circular vignette or convolved with two-dimensional Gaussian function. A precise mathematical description has been used to formulate such stimuli known as Gabor functions (Jones and Palmer, 1987) as illustrated in Figure 1.4. While these stimulus sets have provided a useful start in characterizing V1 neurons, some work suggests that the simplification fails to evoke the complex non-linear responses which these neurons have to natural stimuli (David et al., 2004). As such, they found that the model of spatio-temporal receptive fields fit using the natural stimuli were more than twice as accurate in predicting the responses to other natural stimuli than the model fit to the synthetic stimuli. However, in reviewing evidence for the Efficient Coding Hypothesis, Simoncelli (2003) cautions that the notion of what constitutes a 'natural image' and the definition of the neural response are both unconstrained.



(a) Two-dimensional representation (b) Three-dimensional representation

Figure 1.4: Example Gabor function shown in two (a) and three (b) dimensions.

1.2.4 Prestriate cortex: V2

The next destination in the ventral visual system is V2, which receives strong feed-forward connections from V1 (and the thalamic pulvinar nucleus) which in turn casts strong projections back to V1 (Felleman and Van Essen, 1991). Cells here have similar response properties to those in V1, namely preferences for particular orientations, colours and spatial frequencies, but are also modulated by more complex properties (such as illusory contours) and have larger receptive fields. This region also makes strong connections on to area V4 in the ventral stream as well as V3 and MT (V5) in the dorsal stream (Felleman and Van Essen, 1991).

Neurons in V2, like V1, are also primarily responsive to local orientation and spatial frequency, albeit sampling a slightly larger area of the visual field; approximately 1°–2.5° (Smith et al., 2001). Evidence for increasing complexity of stimulus preference comes from finding some cells here that are sensitive to line conjunctions, orientation combinations and illusory contours. As such it appears that processing of curvature begins here (Hegdé and Van Essen, 2007).

Area V2 has a distinct anatomical structure too, consisting of 'thick stripes', 'thin stripes' and 'interstripe' regions as revealed by the technique of Cytochrome Oxidase (CO) staining (Malach et al., 1994). The blob and interblob pathways of V1 project respectively to the 'thin stripes' and 'interstripes' of V2 (Felleman and Van Essen, 1991) continuing the subdivision of the parvocellular pathway. The V2 'thick stripes' receive inputs mainly from the magnocellular pathway, which project on to areas V3 and MT in the dorsal stream (Felleman and Van Essen, 1991).

While the functions of V1 neurons and the nature of their representations is generally accepted to be the encoding of local orientation and spatial frequency, the nature of representations and their basis dimensions for subsequent cortical areas is less clear. Optical imaging and histo-chemistry of V2 in the squirrel monkey (*Saimiri sciureus*) have revealed that highly orientation tuned patches of cells are centred on thick CO stripes and cells with broader orientation tuning are centred on thin stripes (Malach et al., 1994). A systematic analysis of V2 (and higher area) neural response properties has proven difficult due to the high-dimensionality of the 'shape space', the non-linearity of responses beyond V1 and the practical constraints of neurophysiological experiments (Pasupathy and Brincat, 2012). However, it is apparent that many V2 neurons have more complex cell tuning preferences than those in V1, such as the orientation of corners which could not be predicted on the basis of responses to simple bars (Hegdé and Van Essen, 2000).

1.2.5 V4

Of the three main subregions of V2, namely thick strikes, thin stripes and interstripes, V4 only receives projections from the latter two, although V4 does also receive projections from the magnocellular dominated areas V3 and MT (Felleman and Van Essen, 1991). Again, V4 neurons have similar response properties to the cortical areas preceding it (V1 and V2) but with additional complexity. Many cells in V4 are colour selective, while some respond to combinations of colour and form. Rather than simple bars, V4 cells prefer simple geometric shapes as ablations in monkey V4 results in an inability to differentiate shapes, whereas in humans differentiations of hues tends to suffer more (Kandel et al., 2000). The spatial profile of V4 cells' receptive fields are also modulated by attention, which are most likely influenced by the cortical back-projections.

While V1 cells encode *orientation* — a first order derivative of an object's contours, V4 cells are sensitive to the second order derivative — *curvature* (Connor et al., 2007) with some even sensitive to the third order derivative, *spirality* (Gallant et al., 1993). Other studies have found relatively high overlap in the response properties of V4 cells with those in V2 but in general finding grating stimuli were on average more effective than contour stimuli (Hegdé and Van Essen, 2007), in particular for more complex non-Cartesian gratings (Gallant et al., 1993). While there is clear evidence for differences among shape representation between V1, V2 and V4, techniques to analyse the differences do not always exhibit a systematic progression (Hegdé and Van Essen, 2007) as would be expected by the hierarchical model of processing.

A key reason for the differences in conclusions regarding the progression of cell response properties through the ventral stream is the nature of the stimuli used to determine the cell response preferences (Pasupathy and Brincat, 2012). Ascertaining the precise stimulus preferences has proved difficult however, as time constraints of single-cell recordings mean that no stimulus set can ever adequately sample the myriad of features found in the real world (Pasupathy, 2006). Several groups have taken a very principled approach of gradually increasing the complexity of mathematically well-defined stimuli (Gallant et al., 1993; Hegdé and Van Essen, 2007), whereas others have attempted to use a wide ranging selection of natural stimuli to iteratively 'zero-in' on a cell's preferred stimulus (Quiroga et al., 2005) or feature (Tanaka, 2003). Further work suggests that the differences between the response properties of visual areas are more pronounced under natural viewing conditions (David et al., 2004), while another comparative study of object recognition performance with the popular 'Caltech101' image set suggests that tests based upon uncontrolled 'natural' images can be seriously misleading (Pinto et al., 2008).

1.2.6 Inferotemporal Cortex

Neurons in Inferotemporal cortex appear to be arranged in a columnar organization whereby cells located within the same column have similar (but not identical) stimulus preferences, typically responding to the same critical features for other cells in the column but differing by one key dimension (Tanaka, 2003).

Since the majority of V4 outputs project to the posterior Inferotemporal cortex (Pasupathy and Brincat, 2012), Brincat and Connor (2006) applied a similar (but expanded) set of parameterized shape stimuli to analyse the shape selectivity of PIT neurons, finding neurons highly selective for specific configurations of contour segments of more complex structure than in V4. Furthermore, the selectivity of these cells was found to shift from a linear summation of the contour features to a non-linear combination of typically 2–4 distinct contour segments (Pasupathy and Brincat, 2012).

In contrast to arguments for object recognition being a feed-forward process (Thorpe et al., 1996), modelling the transformation of these V4 to PIT responses with a simple recurrent network reproduced the (approximate) $60\,ms$ time course of the shift from linear to non-linear codes for contour-segment summation found in PIT populations (Pasupathy and Brincat, 2012).

Neurons in macaque posterior IT also appear to signal different information at different times. Brincat and Connor (2006) found that the initial response (around $100\,ms$ after stimulus onset) was to ambiguously respond to a stimulus containing

either of several preferred features (simple contour fragments), modelled as a linear summation across two tuning regions in the contour fragment domain. At around 200 ms after stimulus onset however, responses to individual features subside and the neurons are left explicitly signalling feature combinations, producing a sparser and more explicit representation of object shape. This later phase was modelled by a non-linear product term based on two tuning regions.

From the activity of small populations of IT neurons over short time intervals, the identity and category of the presented visual stimulus may be accurately decoded (Hung et al., 2005). These 'what' details of visual stimuli can be generalized over a range of positions and scales and determined using only very simple 'linear classifiers' which may be easily implemented in neurons as a weighted sum of spikes over approximately 100–300 ms (Hung et al., 2005). While the information for identifying visual stimuli is clearly available at the neuronal level in IT, face perception is so important that specialized face-processing regions can even be identified at the gross anatomical level (Connor et al., 2007) with brain imaging techniques such as fMRI (Tsao et al., 2003).

From direct manipulation of IT cells, it is clear that faces are represented here. Microstimulation of face-selective sites here biased macaques to choose 'face' in a face/non-face discrimination task, thus establishing a causal link between object-specific percepts and the neural activity of inferotemporal cortex (Afraz et al., 2006).

Recent evidence from fMRI and multicellular recordings of the macaque temporal lobe reveals six 'face-patches' in the temporal lobe with different degrees of selectivity to head orientation and identity. Cells found in the posterior (Middle Lateral/Middle Fundus) compared to the anterior areas (Anterior Medial) exhibited increasing invariance, response latencies and (at least in some cases) greater selectivity to identity, again suggesting a hierarchical arrangement (Freiwald and Tsao, 2010). Interestingly, one of the intermediate areas in this path (Anterior Lateral) contained cells which demonstrated mirror symmetry for head-orientation selectivity, suggesting that for faces at least, pooling across mirror-symmetric views may be a mechanism through which view-invariant representations are constructed.

In summary, it is well established that the first stage of processing in the ventral visual system is the extraction of local orientation and spatial frequency information (Hubel and Wiesel, 1968; Connor et al., 2007) — a transformation that maximizes sparseness (that is reduces redundant firing) in the representation of natural images (Olshausen and Field, 1996).

From the initial 'decomposition' of an image into its most statistically regular features in V1, over subsequent cortical stages the receptive fields of the neurons

therein become progressively larger and prefer more complex stimuli until neurons respond to particular faces or objects. From the inferotemporal cortex, there are strong projections to the limbic system, where structures are responsible for processing emotion (Amygdala) and forming episodic memories (Hippocampus). Here, the visual representations are combined with information from other sensory modalities to become increasingly abstract. Cells here have been found which not only respond to different views of a particular person (specifically and invariantly), but even to the written image of their name (Quiroga et al., 2005).

It has been demonstrated that visual object identity can be decoded from IT neurons on behaviourally relevant time-scales with very simple circuits (Hung et al., 2005) and that this information then forms the basis of more abstract representations elsewhere in the brain. The challenge remains to understand how such responses are constructed in the first place. In order to do so, models have been implemented to help move from a description of the phenomenology of IT neuron responses to an understanding of the computational mechanisms which produce them.

1.3 Artificial approaches to visual recognition

Object recognition, in the context of artificial visual systems, is an important area of technological development and so has been the subject of much attention in the machine learning community. While biological and artificial object recognition models start with the different goals of reverse engineering the primate visual system and building a working system by any means respectively, each discipline may at times inform the other. Given that the aim of this thesis is to understand the learning processes occurring in the primate visual system, a full review of the machine learning techniques is beyond the scope of this introduction except for where they have provided insight into the primate visual system.

One such example of synergistic progress comes from the machine learning realm of feature extraction. Here, Independent Component Analysis (ICA), a mathematically derived computational method for separating a multivariate signal into additive subcomponents, has provided an information theoretic angle of analysis on the processing of stimuli by the early stages of the visual system (Hyvarinen and Oja, 2000). There are good *a priori* reasons and strong empirical evidence to suggest that the visual system produces an optimal code from its sensory inputs by extracting the independent components from a scene. When applied to images of natural scenes, ICA has consistently yielded components which closely resemble the response profiles

of simple cells in V1 (Bell and Sejnowski, 1997; van Hateren and van der Schaaf, 1998) for both spatial aspects (van Hateren and Ruderman, 1998) and colour (Caywood et al., 2004), and may arise simply as a result of maximizing the sparseness of the representation (Olshausen and Field, 1996). Furthermore, these mathematically derived independent components of a scene can be used as the basis for an artificial system of face recognition (Keck et al., 2005), confirming that these features contain sufficient information to construct the more sophisticated representations required at the end of the ventral visual stream.

This is what might reasonably be expected of an organism evolved to use a compact and efficient neural code. In essence, the early stages of the ventral visual pathway should respond to general features rather than specific objects, whereby visual scenes are deconstructed into a 'visual alphabet'. From this compact set of visual primitives, increasingly complex and particular features or representations may be built up as required through learning, according to the statistics of the individual's environment and the salient stimuli therein. As the application of ICA to natural scenes has demonstrated, these features are the most statistically regular in our visual world and thus natural elements on which to build a useful representation of it.

Solving the problem of object recognition artificially may be seen from two perspectives. A mathematical, machine learning approach may be to find complex, highly non-linear decision functions to operate on retinal image representations, signalling the presence/absence (or likelihood) of each object to be identified. An alternative perspective would be to find simple operations which gradually transform the retinal input representation into another form which may be acted upon by relatively simple decision functions (such as linear classifiers). DiCarlo and Cox (2007) argue that from a computational perspective, this is essentially a difference of terminology, however the latter approach is more productive in understanding biological object recognition since it deconstructs the problem in a way that is consistent with the architecture and response properties of the ventral visual stream. Furthermore, the simple 'decision functions' required at the end of such a processing chain are easily implemented in biologically plausible neural networks as a thresholded sum over weighted synapses (DiCarlo and Cox, 2007; Hung et al., 2005). For these reasons, the work presented in this thesis adopts the latter paradigm insofar as it is more pertinent to reverse engineering the primate ventral visual stream.

In a similar vein, early models of biological object recognition such as three-dimensional structural descriptions are largely omitted, as they were developed in a time when less was known of the detailed anatomical structure of the ventral pathway

and the cell response properties along it. As such, these models were relatively unconstrained by neurophysiology in comparison to their successors such as two-dimensional view-based models which better fit the psychophysical and physiological data (Riesenhuber and Poggio, 2000; Tarr and Bülthoff, 1998). For reviews of these earlier approaches, the reader is referred to Riesenhuber and Poggio (2000) and Palmeri and Gauthier (2004).

In a similar way to David Marr's 'three levels of analysis', it has been argued that progress in understanding a complex brain process such as object recognition is driven by linking phenomena at different levels of abstraction (DiCarlo et al., 2012). Each 'phenomenon' is best explained in terms of the mechanisms operating at the level below. Such a reductionist approach may lead to an infinite regress, so progress relies upon good intuitions about choosing appropriate and useful levels of abstraction and measurement of pertinent phenomena at near-by levels. In particular, for understanding how the primate visual system learns to identify objects, models of the neurons which constitute the ventral pathway appear to be the most appropriate substrate for the explanatory goal. The next section therefore reviews various biologically plausible neural network models consisting of neurally inspired elements structured and processing information in such a way as to mimic the feature hierarchy of the visual cortex in its capacity for visual object recognition.

1.3.1 Biophysically accurate computational models

Brain function may be classed as an example of 'organized complexity' (Johnson, 2001), whereby the system (brain) is comprised of large numbers of individual agents (neurons) acting according to simple local rules to produce an emergent global behaviour. While a neuron's behaviour in isolation is well understood, their collective behaviour is still difficult to predict and understand. In order to understand the phenomena of interest at one level, it must be linked to the mechanisms one level of abstraction below (DiCarlo et al., 2012). Abstracting the behaviour of the individual neuron therefore seems to be an appropriate basis upon which to build a more sophisticated model of the system's behaviour. Such a model may then be used to test the effect of, for example, different learning rules or training environments, and yield some insight into the principles of self-organization employed in this phenomenon, seen in terms of the behaviour of the components. In the words of Bartlett Mel, "The need for modelling in neuroscience is particularly intense because what most neuroscientists ultimately want to know about the brain is the model — that is, the laws governing the brain's information processing functions." (Mel, 2000).

1.3.2 Rate-coded Neurons

The first model of the neuron was essentially a threshold gate with boolean (digital) outputs in response to weighted boolean inputs (McCulloch and Pitts, 1943) as shown in Equation 1.1. If the activation of the neuron reaches some threshold, Θ, then the output, y, is 1, otherwise 0 as shown in Equation 1.2.

$$\begin{aligned} h &= \langle \mathbf{x}, \mathbf{w} \rangle \\ &= \mathbf{x} \cdot \mathbf{w} \\ &= \sum_{i=1}^{N} x_i \cdot w_i \end{aligned} \tag{1.1}$$

$$y = \begin{cases} 1 & \text{if } h \geqslant \Theta \\ 0 & \text{if } h < \Theta \end{cases} \tag{1.2}$$

Later, the model neuron was refined to produced a graded (continuous) output from weighted real valued inputs. This led to the use of more sophisticated transfer functions (rather than a simple threshold) for mapping the cell body activation to its output firing rate, including for example, linear, threshold-linear and sigmoidal (the output is a generalized form of Equation 1.2 where $y = f(h)$). Other variations replaced the simple dot-product operation, $\langle \mathbf{x}, \mathbf{w} \rangle$, between the input activations and the weight vector with a (usually quadratic) distance computation $\|\mathbf{x} - \mathbf{w}\|^2$, as in the case of Radial Basis Function (RBF) networks.

Even with this advance, research with rate-coded neurons was stalled when it was proven by Minsky and Papert that a single layer network could not solve non-linearly separable problems such as the XOR problem (Bishop, 1997). While augmenting the networks with additional layers could, in principle, overcome this limitation, there was no known method for teaching such networks, as it was unclear what the responses of hidden (intermediate) layer neurons should be. This area was revitalized however following the formulation of the back-propagation learning algorithm — so called because the error between the desired and obtained output was propagated backwards through the network with blame apportioned according to the activation of units in each preceding layer. The back-propagation learning algorithm is a supervised but effective means of training a multi-layer perceptron. Aside from needing to be explicitly taught correct answers, the major drawback of back-propagation, is that

[2]N.B. $\|\mathbf{a}\|$ is the 'norm' of vector $\mathbf{a}$ i.e. $\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2 + \cdots + a_N^2}$

it requires a non-local flow of information in the form of the 'blame' or error signal passing back from the output nodes to modify the intermediate weight layers.

While the highly unusual architecture of the cerebellum makes it a plausible candidate as one brain area where an error signal may be utilized (Rolls and Treves, 1998), the hypothesis concerns a *one-layer* error correction network and as such, error-correction is generally considered to be biologically implausible for models of the multi-layer visual system. In accordance with this, most rate-coded models of the visual system feature only local (synapse specific) learning rules.

1.3.3 Rate-coded Neural Networks

One of the earliest 'feature hierarchy' neural network models of the ventral visual stream is the 'Neocognitron' (Fukushima, 1980). Like many of the models which followed, it is based upon the hierarchical model of Hubel and Wiesel (1962) where the outputs of the LGN are arranged to form Simple cells, whose outputs in turn form the inputs for Complex cells and so on for Hypercomplex cells. The feature preferences of the cells increase in complexity and invariance through successive stages of the hierarchy, with a corresponding increase in receptive field size due to the spatial convergence of the feed-forward connections between layers.

The Neocognitron consists of an input layer followed by three (Fukushima, 1980) or four (Fukushima, 1988) 'S-layers' in which feature specificity and complexity is increased, which are interleaved with an identical number of 'C-layers' which build invariance across position (along with some tolerance to distortions of shape or size and some corruption by noise) for the features selected in their preceding S-layers (Fukushima, 1980). Within each layer are a number of 'planes' or sublayers, (originally 24) with cells arranged retinotopically in each plane (topographically according to their receptive fields) for both 'C-cells' and 'S-cells'. Traversing through the planes, there are cells with receptive fields covering approximately the same spatial extent of the visual scene but which are selective for different features, thus representing hypercolumns.

Competition within each layer is implemented through inhibitory interneurons and tuned to effect a 'winner-takes-all' algorithm. The winner within each S-layer then strengthens its afferent synapses along with the corresponding inhibitory 'V-cell' connections being strengthened to the average strength of the excitatory connections. In this way, each S-cell becomes selective for a particular feature and is suppressed by the presence of non-preferred features through the strengthened connections of its V-cell.

Each C-cell has fixed connections from a group of S-cells in the preceding layer which all extract the same feature but from slightly different positions in the visual field (that is, from the same S-plane). The C-cells respond if any of their presynaptic S-cells respond and thus build a degree of translation invariance, through a 'spatial blurring' operation of the S-cells' feature selectivity.

After either an unsupervised (Fukushima, 1980) or a supervised learning regime (Fukushima, 1988), C-cells in the final layer respond selectively to one of the stimuli presented during training, with only one of the final layer C-cells responding to each stimulus category (Fukushima, 1980, 1988). As such, the neocognitron is able to perform pattern recognition on the basis of shape similarity, unaffected by shifts in position. It also features some resilience to corruption through image noise, or small distortions in shape and size, but is largely restricted to achieving invariance in image-plane transformations (Riesenhuber and Poggio, 1999).

The HMAX model of Riesenhuber and Poggio (1999) was originally introduced to model the data of Logothetis et al. (1995) and follows many of the same principles of organization and operation found in the Neocognitron, including a hierarchical build-up of invariances, first to position and scale and then to viewpoint. A core principle of the model is that invariance to *any* transform can be built up by pooling over afferents tuned to various transformed versions of the same stimulus (Riesenhuber and Poggio, 1999).

HMAX features alternating layers of S-cells and C-cells to extract features and build invariances, which have progressively larger receptive fields and prefer progressively more complex stimuli in higher layers — a crucial property for avoiding the trade-off between a combinatorial explosion of feature units or insufficient discriminability (Riesenhuber and Poggio, 2000). Importantly, the complex features in higher layers are, like the Neocognitron, built from simpler features, making the complex features tolerant to local deformations as a result of the invariance properties of their afferent cells tuned to the simpler features.

Like the Neocognitron, HMAX implements a winner-takes-all strategy known as the 'MAX' function rather than using a linear 'SUM' operation over a cell's afferents, as the non-linear MAX function provides better invariance with cluttered visual scenes, the presence of multiple stimuli or changes in size (Riesenhuber and Poggio, 1999). Such non-linear operations are crucial for developing increasingly complex shape selectivity in the visual system, as any number of successive linear transformations can be mathematically reduced to a single linear transform (Pasupathy and Brincat, 2012).

Unlike the Neocognitron, HMAX features fewer layers of cells, with just two S-layers alternating with two C-layers in the original formulation (Riesenhuber and Poggio, 1999) but with a final 'VTU' layer. It also has a slightly more complicated connectivity, whereby S and C levels do not always have to alternate, for example, C1–C2 connections featured in the model, skipping layer S2.

After training, HMAX provided good agreement with the view-tuning range of experimentally observed values recorded by Logothetis et al. (1995) in Monkey IT cells (Riesenhuber and Poggio, 1999). Additionally they found cells in the output layer responding with mirror-view tuning as Logothetis et al. (1995) and other more recent empirical studies (Freiwald and Tsao, 2010) also found.

Both the Neocognitron and HMAX are hand-wired to build location invariance into the architecture insofar as the 'MAX' operation is prescribed for 'C' cells over a particular feature, in turn represented by 'S' cells over many different locations. Other models however, have used a simpler assumption of convergence of connections from all features and leave the network to self-organize with respect to which features to build invariance across.

One such 'bottom-up' approach is VisNet (Wallis and Rolls, 1997), a four-layer hierarchical feed-forward network with convergence to each part of a layer from a small topologically corresponding region of the preceding layer (see Figure 1.5) which is not feature specific. Each layer consists of a competitive network with local graded inhibition providing intra-layer competition between neurons (potentially being winner-takes-all in the extreme case) and providing an important non-linearity to the layer-by-layer processing. Learning occurs in the feed-forward connections with a form of the Trace learning rule (Földiák, 1991), a modified version of Hebbian learning which adjusts synaptic weights according to current firing rates and those of the previous time-step, thus associating together input patterns presented over consecutive time-steps. Since images presented close together in time are likely to be transforms of the same object (Simoncelli, 2003; Simoncelli and Olshausen, 2001), the model learns to associate different views or transforms of an object together until neurons in the final layer have formed a view-invariant representation of the object.

In addition, a new mechanism known as Continuous Transformation (CT) learning was discovered (Perry et al., 2006; Stringer et al., 2006), relying on spatial rather than temporal continuity, which has been found to be robust under a wide range of experimental conditions. Importantly, both the Trace and CT learning mechanisms are local and unsupervised, making VisNet a biologically plausible model of object recognition in this respect.

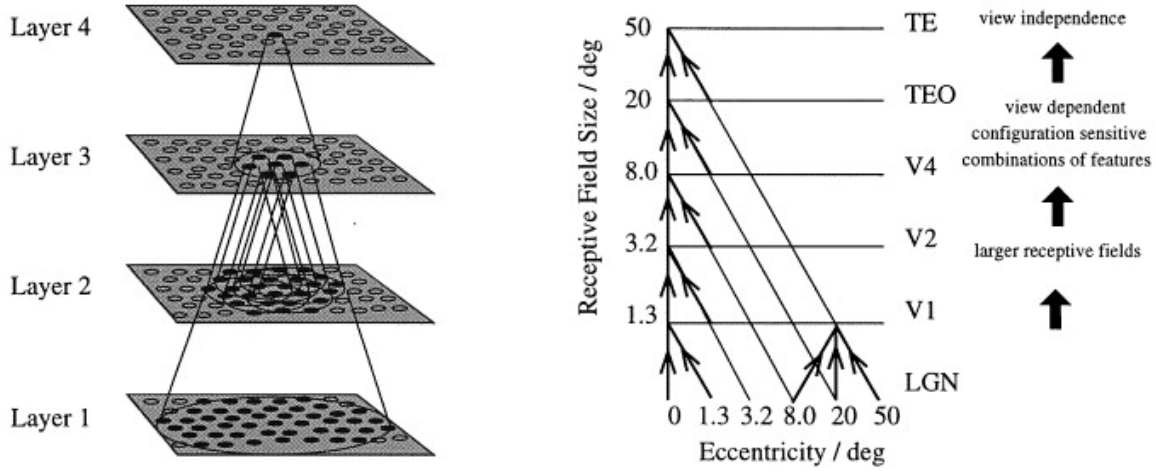


Figure 1.5: The feed-forward architecture of VisNet, a rate-coded model of the ventral visual system (Stringer and Rolls, 2000). Reprinted from Stringer and Rolls (2000), Figure 1, Copyright © (2000), with permission from Elsevier.

The inputs to VisNet consist of images pre-processed with elongated Difference of Gaussian (DoG) filters (Rodieck, 1965) designed to model the cell response properties of V1 cells (localized oriented bar and edge detectors of multiple spatial scales). While DoG filters provide a reasonable approximation to V1 simple cells when considered to be bars and edges, Gabor filters are more commonly used now for V1 filtering, which capture the sinusoidal grating response preferences of these cells (Jones and Palmer, 1987). From the input layer, the connectivity converges such that neurons in the fourth layer sample virtually the whole of the input space, from which the model learns which features from each preceding layer should be associated together.

VisNet has proved to be a successful model of the process of biological recognition across a variety of types of transform, starting primarily with translation invariance (Wallis and Rolls, 1997). At first, simple shapes were used as stimuli for more controlled studies, (in particular 'T', 'L' & '+') but also filtered greyscale images of faces, with each set of stimuli achieving translation invariance across nine locations due to the partial invariance afforded by the converging receptive fields but mostly due to the temporal association of successive transforms through the Trace rule.

In subsequent work, the Trace rule was modified in various ways to examine performance on an expanded translation problem (Rolls and Milward, 2000). Interestingly, the performance was greatly improved by basing the temporal trace on the previous trial without the current trial, highlighting the importance of this learning rule for the model's functioning. However, CT learning was later discovered (Stringer et al., 2006) and was subsequently shown to be sufficient for VisNet to build translation invariant

representations with no trace component at all (Perry et al., 2010) when the stimuli were presented appropriately.

Other simulations have attempted to show that VisNet could cope with learning translation invariant representations on cluttered backgrounds (Stringer and Rolls, 2000). In this work however, the model could only cope when the faces were tested on the cluttered backgrounds and trained on blank backgrounds. Under the reverse matching, VisNet failed to learn about individual faces.

This difficulty led Stringer et al. (2007) to reconsider the notion of a cluttered background as an environment composed of multiple stimuli rather than just noise. With a large enough pool of stimuli to draw from, presenting each possible pairing to the model enabled it to learn individual translation- (Stringer et al., 2007) and rotation-invariant (Stringer and Rolls, 2008) representations of each object, despite the tendency of the Hebbian learning rule to associate them onto the same output neurons and cause a 'superposition catastrophe' (von der Malsburg, 1999).

However, in the original experiments, a pool of 10 objects equated to ${}^{10}C_2 = 45$ pairings, which constitutes such an extensive training regime as to make it biologically implausible as a full account of how we learn separate representations from perhaps just one exposure to a pair of objects. Nonetheless, it was an important first step in recognizing the importance of the statistics of the visual environment as a key to moving away from an ecologically implausible single stimulus presentation paradigm common in most models and will be an important theme for this thesis.

1.3.4 Spiking Neurons

The primary motivation for the work conducted and presented in this thesis comes from the hypothesis that by modelling the precise spike times of neurons (rather than taking an average firing rate as many existing models have done), more complex and powerful dynamics may be captured which provide new insights into the way the primate ventral stream learns to recognize objects invariantly. This section therefore reviews the spiking neuron models which have been formalized and then the most significant networks of them to be applied to the particular problem of object recognition.

Many models of spiking neurons exist offering various compromises between biological detail and computational cost, as summarized in Figure 1.6. The 'mechanisms' used to explain the 'phenomena' of interest (one level of abstraction above) are themselves phenomena requiring their own level of mechanistic explanation (DiCarlo et al., 2012). Since these mechanisms may ultimately prove pertinent to the original phenomena of interest, in choosing a model neuron for this work, it was therefore important to

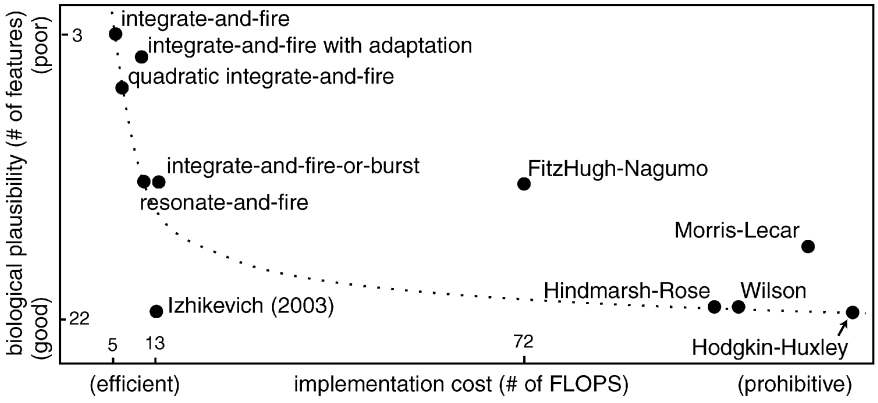


Figure 1.6: Comparison of the biological accuracy (in terms of the number of features captured) for various models of spiking neurons against approximate computational cost, measured in the number of Floating Point operations per millisecond of simulated time. Reproduced with permission from Izhikevich (2004) Copyright © 2004, IEEE.

consider the ease with which a model may be adapted to incorporate new features which may contribute to the emergent behaviour of interest.

Many spiking neuron models have been formalized, offering various degrees of biological detail and computational efficiency. Here, the few most popular and influential are introduced out of the candidates considered for this work with a brief appraisal of their strengths and weaknesses. For mathematical descriptions and evaluations of other notable spiking neuron models including the Quadratic Integrate-and-Fire, FitzHugh–Nagumo, Hindmarsh–Rose and Morris–Lecar neurons, the reader is referred to Izhikevich (2004) and Gerstner and Kistler (2006).

1.3.4.1 Hodgkin–Huxley

One of the oldest and most detailed mathematical models of a spiking neuron is the Hodgkin–Huxley model (Hodgkin and Huxley, 1952) for which they earned a Nobel prize. It derives an expression for the cell membrane potential in terms of the Nernst equations for each of the three ions known to play an integral role in the generation and propagation of action potentials, based upon their experiments with the giant axon of the squid. The passive electrical properties of the resistive bi-lipid cell membrane and the voltage gated behaviours of ion channels within the membrane for sodium and potassium and the non-specified leakage ion (considered to be chlorine) form a set of ordinary differential equations, ultimately describing the evolution of the cell

membrane potential through time (Equation 1.3).

$$C_m \frac{dV}{dt} = -\sum_k I_k(t) \tag{1.3}$$

Here, C_m is the electrical capacitance of the cell membrane, V is the cell membrane potential and $\sum_k I_k(t)$ is the sum of all the ionic currents that pass through the cell membrane. This is further decomposed according to the conductance of each type of ion (g_k), their reversal potentials (E_k) and three gating variables, m, n, and h, which modify the proportion of open ion channels for each type of ion as follows in Equation 1.4.

$$\sum_k I_k = g_{Na} m^3 h (V - E_{Na}) + g_K n^4 (V - E_K) + g_L (V - E_L) \tag{1.4}$$

Additionally, the three gating variables each evolve through time according to their own differential equations as given in Equation 1.5.

$$\begin{aligned} \frac{dm}{dt} &= \alpha_m(V)(1 - m) - \beta_m(V)m \\ \frac{dn}{dt} &= \alpha_n(V)(1 - n) - \beta_n(V)n \\ \frac{dh}{dt} &= \alpha_h(V)(1 - h) - \beta_h(V)h \end{aligned} \tag{1.5}$$

The full model description is given by Equations 1.3–1.5 together with the empirical measurements for each type of ion's reversal potential, membrane conductance and the functions $\alpha_x(V)$ and $\beta_x(V)$ for each gating variable, which were adjusted to fit the data for the giant axon of the squid. For further details please refer to Gerstner and Kistler (2006).

While the Hodgkin–Huxley model is probably the most detailed and accurate mathematical description of a 'point' neuron (especially considering subsequent variants which have extended it to incorporate more species of ions), it can be seen from the sheer number of differential equations to be solved at each time-step (and the analysis shown in Figure 1.6) that this is a prohibitively expensive model for simulating more than a very small number of neurons. Similarly, there are a family of models with spatial extent (that is, not 'point' neurons) which describe the pertinent cell membrane potentials and currents through both time and *space* in the form of partial differential equations. As these are even more computationally expensive and there is no *a priori* reason for believing that this level of detail is necessary, the Hodgkin–Huxley neuron and its variants were ruled out very quickly.

1.3.4.2 Izhikevich

Although the Hodgkin–Huxley (HH) model embodies our most detailed understanding of the mechanism of action potential generation, its prohibitive cost has led other theoreticians to move away from taking the explicit flow of ions as the primitives to explain spike generation and consider 'higher-level' primitives to explain the same emergent properties. One such popular model, which claims to offer as much detail in describing the behaviour of spiking neurons but at a fraction of the computational cost, is the nullcline derived Izhikevich model neuron given in Equations 1.6–1.8 (Izhikevich, 2003, 2004).

$$\frac{dV}{dt} = 0.04V^2 + 5V + 140 - u + I \tag{1.6}$$

$$\frac{du}{dt} = a(bV - u) \tag{1.7}$$

There is also an after-spike resetting condition (where $+30\,mV$ is the peak of the spike, not the potential).

$$\text{If } V \geq +30\text{mV}, \text{then} \begin{cases} V \leftarrow c \\ u \leftarrow u + d \end{cases} \tag{1.8}$$

This model neuron has been used in several large scale simulations of cortical areas (Izhikevich et al., 2004; Izhikevich, 2006) as it is claimed to combine the biological plausibility of HH dynamics with the computational efficiency of Integrate-and-Fire (IF) neurons (Izhikevich, 2003).

While the advantages of this model are clear from Figure 1.6 (Izhikevich, 2004), the constants have no correspondence to measurable physiological properties of real neurons but are calculated from phase-plane analyses. It is also not obvious how to extend the model to incorporate new features which the author has not categorized already, other than by parameter searching which could be a costly pursuit in itself. Izhikevich has inventoried twenty behaviours for this model (Izhikevich, 2004) and with only 13 floating point operations per second (FLOPS) required to simulate $1\,ms$ of biological time, compared to 1200 for HH neurons, this model represents a very appealing plausibility to cost ratio. However, it is still slightly more expensive than the 5–10 FLOPS required for simulating variants of the IF neuron and computational efficiency must be prioritized if there is no particular reason for using the greater range of behaviours.

1.3.4.3 Integrate-and-Fire

The simplest and computationally cheapest spiking neuron model is the leaky Integrate-and-Fire neuron[3] (formally 'LIF', but typically abbreviated 'IF'). The most important simplifying assumption in IF neurons is that the shape of action potentials may be neglected and so spikes are treated as uniform events, defined only by their time of occurrence (meaning also that an absolute refractory period must be explicitly implemented). It is derived from a simple RC circuit (a resistor and capacitor in parallel) driven by an input current, $I(t)$, and formulated as in Equation 1.9.

$$\tau_m \frac{dV}{dt} = -V(t) + RI(t) \tag{1.9}$$

This model had a number of suitable properties for the planned research, in that it is computationally cheap, it neglects details such as the shape of the action potential which were deemed irrelevant for the level of description sought and it may be straight-forwardly extended with additional features and properties, such as firing-rate adaptation, as they are required. Further details and a derivation of the (conductance-based) LIF are given in Chapter 2.

1.3.4.4 Spike-Response Model

The three models so far introduced are all synchronous 'clock-driven' models, whereby time must be broken down into discrete steps and a numerical method used to calculate approximate solutions to the systems of differential equations at each time-step. Another very different approach is an 'event-based' simulation such as Gerstner's Spike-Response Model (SRM). Rather than being defined in terms of differential equations to be solved at each time-step, the SRM expresses the membrane potential at time t as an integral over the past (Gerstner and Kistler, 2006). The consequence for simulation is that the state variables do not need to be updated at every time-step, only upon the arrival of a spike, which may make it a computationally cheaper model under some circumstances where the spike rate is low. However, it is not easy to extend the SRM to include, for example, firing-rate adaptation and especially noise generated by a stochastic process (Brette et al., 2007) nor is it clear that the planned simulations would actually mean that the SRM had a computational advantage over the more straightforward LIF.

[3]With the exception of non-leaky variant which is rarely used due to the lack of decay on the cell membrane potential

1.3.5 Spiking Neural Networks

Many of the large (network) scale simulations of cortical neurons in general use a variant of the LIF or sometimes the Izhikevich model neuron (Izhikevich et al., 2004; Izhikevich, 2006). Here models of the ventral visual stream which have been applied to the problem of object recognition are reviewed.

One recent spiking neural network of the visual system (consisting of conductance-based LIF neurons) has been used to segment a visual scene composed of two objects which consist of oriented texture patches (Wu et al., 2008). By augmenting the network with lateral excitatory connectivity, this ability was found to be enhanced. It was also shown that with another layer of neurons and plastic feed-forward synapses between them, output neurons could be trained to one or the other object selectively, although in a position-dependent way. As such, it addresses an important issue of how the visual system might recognize individual objects in a scene composed of several, but fails to address the crux of real-world visual object recognition — invariance.

SpikeNet is an event-driven four-layer feed-forward network of leaky integrate-and-fire neurons which seeks to model the speed of processing in the primate visual system by implementing a rank order code, in which neurons are sensitive to the order that spikes arrive in (Delorme et al., 1999; VanRullen and Thorpe, 2002). This is an appealing model as it demonstrates how relatively short time windows (sometimes just one spike per neuron) and small numbers of neurons can quickly and efficiently transmit reliable information about a visual input. While other models may require a long period over which to count spikes (incompatible with the speed of processing) or large numbers of neurons over which to count spikes in shorter intervals (incompatible with the size of the optic nerve), by making a postsynaptic neuron progressively less sensitive to incoming spikes, $n!$ states may be transmitted by just n neurons (VanRullen and Thorpe, 2002).

Since feature or stimulus detectors in later layers will be more excited by their preferred stimulus, they are more likely to fire sooner. As such, SpikeNet offers another perspective on the traditional feature-hierarchy view of the ventral stream, with each level representing stimulus saliency, which is gradually redefined layer-by-layer (VanRullen and Thorpe, 2002). The progressive desensitization (potentially through shunting lateral inhibition) to subsequently arriving spikes in this model effectively implements the usefully non-linear 'MAX' function in the temporal domain. By doing so, this model has been shown to use latency to overcome the 'binding problem' occurring for neurons with large receptive fields viewing scenes composed of multiple objects or clutter (Rousselet et al., 2004).

The input layer of SpikeNet is an array of On-Centre and Off-Centre units, modelling the responses of retinal ganglion cells. Since the only important detail for reconstructing the image based on the neural firing is the order of spikes, these units perform an automatic normalization of the image with respect to contrast and luminance, fitting with the inaccuracy of human vision for estimating the true (absolute) image luminance levels (Thorpe et al., 2000). While luminance and contrast are important invariances for a model of the visual system to build, SpikeNet produces a localized 'face map' and hence is not location invariant.

This model synthesizes empirical data regarding the speed of the human recognition process (100–150 ms) with persuasive theoretical arguments about the inadequacy of Poisson-like rate codes (since each area has only $\sim 10\,ms$ to integrate its inputs), to arrive at a demonstration of rapid and accurate object recognition in a spiking neural network with only feed-forward projections. Notably, simulations have shown that more than 50% of the maximum information potentially available in the model retina is transmitted in the first 1% of spikes (VanRullen and Thorpe, 2002).

It is however, less convincing as a model of the learning process which ultimately must underpin the recognition process. The training scheme for SpikeNet employs a supervised learning algorithm, whereby synaptic weights are fixed as a function of the order in which the inputs fire (Thorpe et al., 2000), yet a full model of the primate visual system must be able to learn in an unsupervised way, most likely by virtue of heuristics forged through evolution (Sprekeler et al., 2007).

In a later extension to the work, a simplified form of STDP (where only the order of pre- and postsynaptic spikes, not the delay between them was used to update synaptic weights) was used to detect statistical regularities in terms of the earliest firing afferent patterns within spike trains (Masquelier and Thorpe, 2007). By presenting the network with natural images from two Caltech data sets (one containing faces, the other, motorbikes), ten class-specific structured visual representations emerged in an unsupervised way which formed a foundation for a radial basis function (RBF) classifier to successfully categorize the presented stimuli (Masquelier and Thorpe, 2007). While these simulation results exhibited specificity and invariance, it was with respect to exemplars from categories, rather than recognizing specific faces or objects over transformations, (for example in size, view or location) and so principally informs our understanding of how features of intermediate complexity may be learned.

As a model of the mature system, it therefore serves a complimentary role to the aims of this work but importantly, does also emphasize the importance of the timing of spikes and not only their number, indicating the necessity of using spiking neural

networks and the pulse codes they transmit over earlier rate-coded models. In the next section, there follows a review of the learning rules employed in neural networks which have attempted to model the learning process for transformation-invariant recognition.

1.4 Learning

The ability to learn means that an organism is able to cope with complex and unfamiliar environments and adapt as they change throughout its lifetime. Accordingly, the brain systems for learning to recognize objects or faces are not fixed at birth, but develop gradually, providing a degree of flexibility necessary to recognize food, predators, group members and other stimuli essential for reproduction and survival. As an individual gathers sensory experience through interaction with their environment, much of the learning is likely to be at least partially unsupervised. As such the brain is likely to utilize heuristics, or general principles, which hold across a broad range of environments to provide a useful mechanism for learning from this data (Sprekeler et al., 2007).

The central dogma of neuroscience is that learning and memory are encoded in the strength of connections between neurons — *synaptic* plasticity underlies *behavioural* plasticity. There are believed to be several mechanisms for how synaptic efficacies may be modified. While the computational goals of learning may change, the algorithm underlying these learning processes is common to many parts of the brain, although variations in implementation may vary throughout an individual's nervous system. This section first reviews biologically plausible, rate-coded learning rules as applied to learning transformation-invariant visual representations, before examining subtleties in modification of synaptic efficacies discovered due to differences in spike timing.

Synaptic learning may occur on short time-scales (in the order of minutes or seconds) which is commonly considered to be short-term plasticity — both short-term potentiation (STP) and short-term depression (STD). However, while object recognition may occur on a similarly short time-scale, the learning *persists* for a significantly longer time-scale (as in normal cases an individual may recognize a face years after first seeing it) and so this section is concerned only with the long-term equivalents: long-term potentiation/depression (LTP/LTD).

1.4.1 Hebbian Learning

As a theory of synaptic plasticity, Hebbian learning has been the dominant paradigm in the biology of learning and memory since it was first articulated over 60 years ago

and is best described in the words of its author (Hebb, 1949):

"When an axon of Cell A is near enough to excite a Cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."

While other more powerful learning rules have been formulated, (for example back-propagation as discussed) they typically lack the biological plausibility of Hebbian learning, for example through relying on non-local information or mechanisms which would be very difficult or impossible to implement in neural circuits. The next section introduces two biologically plausible variants of Hebbian learning rules which have proved very successful in models of transformation-invariant object recognition including VisNet (Wallis and Rolls, 1997; Stringer et al., 2006) and very recently HMAX (Isik et al., 2012). The concept of Hebbian learning is then reconsidered in the light of more recent discoveries about the effect of pre- and postsynaptic spike timings on changes in synaptic weights.

1.4.2 Rate-coded learning rules

While there are many artificial learning rules such as back-propagation which have been successfully applied to a wide variety of supervised learning problems in neural networks, this review focuses on the more biologically plausible unsupervised Hebbian varieties, as these are generally more likely to be closer to the learning occurring in the primate brain[4]. The two most successful Hebbian learning rules in a rate-coded paradigm have been Trace Learning (Földiák, 1991) and Continuous Transformation (CT) learning (Stringer et al., 2006) which exploit the statistics of natural transforms of objects to associate them together based on their temporal and spatial continuity respectively.

1.4.2.1 Trace Learning

Several studies have noted that in the statistics of natural images, those seen contiguously or close together in time are likely to be different views of the same object (Földiák, 1991; Wallis and Rolls, 1997; Sprekeler et al., 2007). This has been proposed in various forms to allow the visual system, or models thereof, to solve the 'invariance' problem by associating together temporally contiguous visual inputs. In a robotics study serving as a physical instantiation of this notion, the neural network processing the video input of the robot as it explored its environment even stabilized on a

[4]One notable exception perhaps is the Cerebellum, which according to Marr's theory has a rare architecture precisely suited to conveying an error signal to correct learning in a supervised way.

hierarchy of processing stages with very similar response properties to those found in the ventral visual stream, through modifying the feed-forward connections with a local learning rule which extracts smoothly varying features (Wyss et al., 2006). The receptive fields of neural units in the lowest level were most sensitive to gratings (parallel oriented bars) similar to the tuning properties of V1 cells, while units in higher levels were found to be tuned to more complex configurations such as curvature and particular objects in the environment, closely matching the way representations are found to build from V2 to IT.

The standard basic form of the Trace rule is a modified version of Hebbian learning, which utilizes a temporal trace of a neuron's recent activity in order to associate together stimuli occurring close together in time (Földiák, 1991; Wallis and Rolls, 1997). The learning rule relies upon the properties of real world objects, whereby stimuli seen close together in time are likely to be different views of the same object which are associated together to achieve transform invariance. In comparison, there will be relatively few associations between different objects, as views of a particular object are stabilized on the fovea for much longer than the saccades from one stimulus to another. In other words, the object identities typically vary more slowly in time than contextual variables or noise (Sprekeler et al., 2007). Furthermore, the huge number of objects in real world vision means that the probability of repeatedly experiencing the same temporal sequence of views between different objects will be low (i.e. after seeing many views of object A, sometimes object B is seen next, other times object C, D or E etc.) such that the associations across different objects, on average, will be very weak relative to associations between different views of the same object.

Learning with the Trace rule is very similar to the standard Hebbian form for rate-coded neurons except that the present *post*-synaptic neuron's activity, y_i is replaced by a trace term $\overline{y}_i$ which is associated with the present activity of the *pre*-synaptic neuron, x_j to modify the synaptic weight between them, Δw_{ij} according to the learning rate constant α as shown in Equations 1.10–1.11. Here η is a parameter which weights the previous trace term with the present activity of the same neuron.

$$\delta w_{ij}^{\tau} = \alpha \overline{y}_i^{\tau} x_j^{\tau} \tag{1.10}$$

$$\overline{y}_i^{\tau} = (1-\eta)\, y_i^{\tau} + \eta \overline{y}_i^{\tau-1} \tag{1.11}$$

This learning rule has been shown to successfully exploit the spatio-temporal statistics of real-world objects to learn transformation-invariant representations of them in models of the ventral visual system (Wallis and Rolls, 1997). Other simulations

have explored variations in the form of Trace learning, including a presynaptic trace $\overline{x}_j$ rather than postsynaptic trace $\overline{y}_i$ for both τ and $\tau - 1$, which found no significant difference in pre- versus postsynaptic forms, but a significant improvement in network performance when the trace did not incorporate activity from the current stimulus presentation but only previous presentations (Rolls and Milward, 2000).

As a test of the hypothesis that visual object invariance learning utilizes the temporal contiguity of object features during natural visual experience, Li and DiCarlo (2008) manipulated images of stimuli presented to a monkey by either changing the presented stimulus for another during the monkey's saccade in the 'swap' condition or continuing to display the same object in the 'natural' condition. In support of the hypothesis in question, they found that after just one hour of the swap condition, there were specific and significant changes in position tolerance for the IT cells recorded. In earlier work, a similar result was obtained with human participants, who came to associate two different objects as the same when they were swapped during the blindness of their saccades after less than one hour of such manipulated conditions (Cox et al., 2005), adding support to the notion that position invariance is constantly adjusting to the spatio-temporal statistics of our visual environment.

While Trace learning has so far been a theoretical construct, there are several plausible biological mechanisms by which it could be implemented by neurons. At the level of neural circuits, lateral excitatory (recurrent) connections could maintain activity in the pyramidal cells after removal of the stimulus. Alternatively, activity could be sustained at the synaptic level by slow time-course of NMDA receptors (in the order of $100\,ms$) for example, due to slow deactivation of the Calcium channels or slow removal of glutamate from the synaptic cleft.

1.4.2.2 Continuous Transformation Learning

Continuous Transformation (CT) learning is based upon the idea that two views of the same object are more similar to one another than two views of different objects (Stringer et al., 2006). Through the standard rate-coded Hebbian learning rule (Equation 1.12), sufficiently similar views are associated together onto the same postsynaptic neurons. While Trace learning relies upon temporal continuity between views (transforms) of an object, CT learning relies upon spatial continuity (Perry et al., 2006).

$$\delta w_{ij} = \alpha y_i x_j \tag{1.12}$$

Under normal viewing conditions, the projection of an object will be stabilized onto the fovea and so for a small transformation (such as a shift or rotation), many of

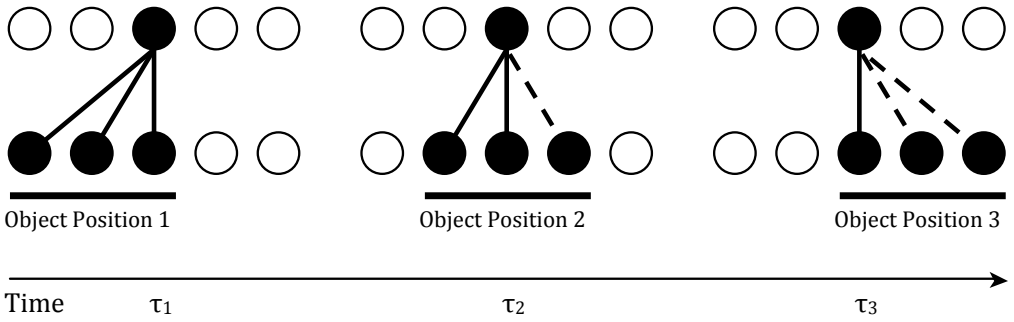


Figure 1.7: Illustration of the CT learning mechanism for translation invariance. In the initial position at the first time-step (τ_1) the input neurons randomly activate a set of postsynaptic neurons (due to the random synaptic weight initialization) and the synaptic connections between them will be strengthened through Hebbian learning. If the second transform at τ_2 is similar enough to the first, the same postsynaptic neurons will be encouraged to fire by some of the same connections potentiated at τ_1 and the input neurons of the second transform will have their synapses potentiated onto the same set of output neurons. This process may continue (τ_3) until there is very little or no resemblance between the current and the initial transforms.

the same features will be present in the two different views, leading to a significant overlap of activity amongst the feature detectors of V1. The neurons representing the first view of the object will be potentiated onto a set of postsynaptic neurons, making them more likely to fire when presented with the same (or similar views) in the future. When the next view is 'seen' by the network, if enough features are shared with the original view then there may be enough of the same input neurons firing to stimulate the same set of postsynaptic neurons again, as the synapses between them were previously strengthened. Since these same postsynaptic neurons are active simultaneously with the new input neurons (which represent the previously unseen features of the same object), they too will have their synapses strengthened onto the same set of postsynaptic neurons by a simple Hebbian learning rule (Equation 1.12). Figure 1.7 shows a simple illustration of the CT mechanism for learning translation invariance over three positions.

Provided that enough similar 'intermediate' views are seen, this mechanism is able to associate many different views of an object together, to be represented by the same set of output neurons, even if some of the views in the set are orthogonal to one another. Since the output neurons representing an object will respond even when only some of the features present in a particular learned view are seen, it can be seen how the postsynaptic neurons can interpolate between previously experienced views

to recognize a completely novel 'intermediate' view.

CT learning has been found to be a robust mechanism for training a neural network to recognize 3-D rotating stimuli under a range of experimental conditions, such as when transforms are presented in a random order, and a condition under which Trace learning fails when: views of different stimuli are interleaved through time (Stringer et al., 2006). While CT does not require temporal continuity in the order of transforms, it is however sensitive to the degree of similarity (for example, degrees of rotation) between transforms (Stringer et al., 2006).

1.4.3 Spike-Time Dependent Plasticity

Spike-Time Dependent Plasticity (STDP) is the phenomenon that the relative timing of two cells' spikes affects the degree and even the direction with which the synaptic efficacy is altered between them. It has received a great deal of attention since the original discoveries in thick tufted Layer 5 neocortical pyramidal cells (Markram et al., 1997) and rat hippocampal cells (Bi and Poo, 1998). While many varieties of STDP have subsequently been discovered (Abbott and Nelson, 2000), the 'classic' form exhibits long-term potentiation (LTP) when the presynaptic spike occurs before the postsynaptic spike (the 'causal' condition) and long-term depression (LTD) when the order of spikes is reversed (the 'acausal' condition).

Importantly, the spiking events must also occur close together in time (in the order of $10\,ms$) for such changes to occur, as the strength of both effects diminish approximately exponentially with increasing separation between spikes, as mapped out in rat hippocampal cell cultures by Bi and Poo (1998) and shown in Figure 1.8. It is also commonly found that the temporal window for depression is slightly larger[5] than for potentiation. This has the useful property of overall weakening the synapses for uncorrelated pre- and postsynaptic spike patterns, eliminating the need for an additional artificial homeostatic regulatory function and helping to stabilize output neuron steady-state firing rates making them independent of variations in input firing rates (Song et al., 2000).

In a recent psycho-physics study of face perception, evidence was obtained suggesting that STDP features throughout the ventral visual stream, including higher temporal cortical areas, making it particularly pertinent to this work. A shift in the perceived midpoint along a continuum of 'morphs' between two faces was obtained by

[5]The size of the LTD (or LTP) function integral may be increased through lengthening the time constant τ_- (τ_+ respectively) or the size of the impulse, A_- due to a postsynaptic spike (or A_+, due to a presynaptic spike respectively).

presenting the two original faces in close succession, (with maximum shift at $\pm 20\,ms$) in a manner consonant with the temporal profile and causal order of STDP, thus reproducing the STDP curves from experimental manipulation at the level of the *system* rather than the *synapse* (McMahon and Leopold, 2012). A significant, but much smaller shift was obtained when the same experimental paradigm was applied to oriented gratings, suggesting not only that STDP is found throughout the visual system, but that it is enhanced in higher visual areas — further supported by the two-fold stimulus size invariance reported, suggesting changes in non-retinotopic areas (McMahon and Leopold, 2012). This data is commensurate with the idea that the low-level visual 'alphabet' of features remain relatively fixed (and are more frequently reinforced) while an individual must retain the ability to rapidly learn and remember new faces throughout life.

Interestingly, the shape of the STDP curves have also been recovered from a purely mathematical analysis of the 'slowness' or 'temporal stability' principle (Sprekeler et al., 2007), thus formally relating STDP to learning mechanisms used for developing transformation-invariant representations, namely Slow Feature Analysis and the Trace rule. In light of the discovered details of this relationship, the authors propose a new functional interpretation of STDP as an implementation of the slowness principle that compensates for neuronal low-pass filters such as the excitatory postsynaptic potential (EPSP).

While much theoretical and experimental work assumed (tacitly or otherwise) that the important neural quantities for communication and learning are the firing rates of the pre- and postsynaptic neurons (averaged over space, time or repeated trials) the all-or-nothing nature of an action potential lead to the idea that the precise timing of the spikes might be important and the subsequent re-evaluation of Hebb's postulate.

In early theoretical work, Gerstner et al. (1993) demonstrated that a learning rule where pre-before-post spikes caused LTP could be used to learn spatio-temporal sequences in a spiking neural network, which was later developed to include an LTD component when the order of spikes was reversed (Gerstner et al., 1996). In these models, the time windows for plasticity were (erroneously) assumed to be in the order of $1\,ms$ but the later model did emphasize the importance of the 'causal' order of spike timing for potentiation (and the reverse for depression) — a hallmark of the STDP phenomenon.

Experimental confirmation was being gathered around the same time, demonstrating the sensitivity to timing through mapping out the time windows, initially for hippocampal neurons (Bi and Poo, 1998). STDP is now a readily accepted and

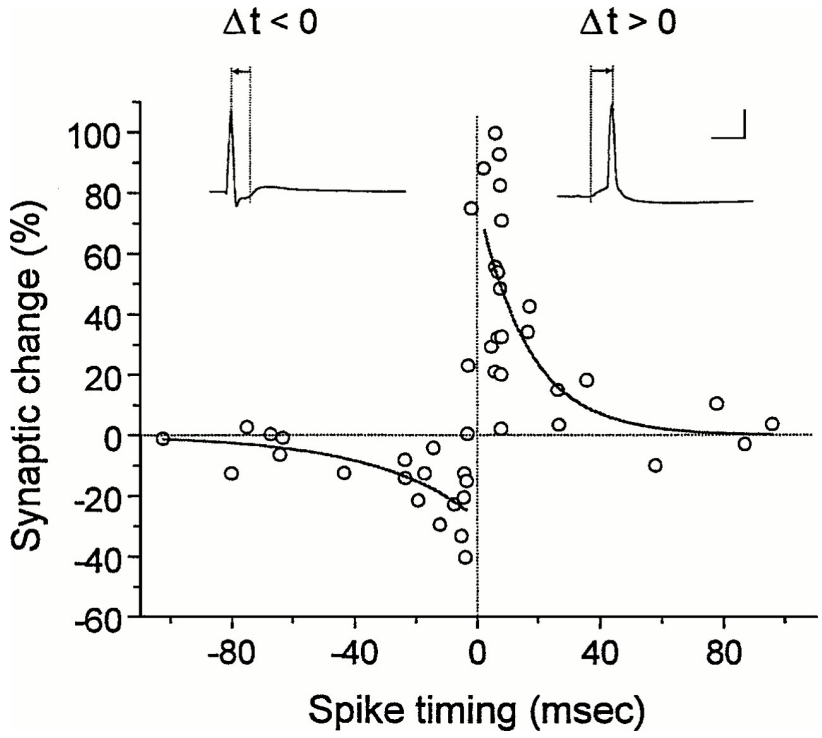


Figure 1.8: Changes in synaptic weight in rat hippocampal cells under electrical stimulation plotted against spike timing (and order) where the data were from Bi and Poo (1998). Reproduced with permission from Bi and Poo (2001), Copyright © 2001, Annual Reviews.

well supported phenomenon in both the neurophysiological and modelling literature, but several open questions still remain concerning the precise details. For example, the earlier work on the frequency dependence of LTP and LTD has yet to be unified with the more recent finding of STDP (Markram et al., 2011). Furthermore, while STDP is believed to be dependent upon postsynaptic Ca^{2+} influx, as evidenced by the elimination of LTP by blocking NMDA receptors (Markram et al., 1997), the exact details of the role played by Calcium (and subsequent or alternative messengers) remain unknown (Bi and Poo, 2001).

While the importance of spike timing in STDP is incommensurate with Shatz's often cited simplification of Hebb's postulate 'cells that fire together wire together' (Markram et al., 2011) and recent evidence from cortical reorganization in cat primary visual cortex (Young et al., 2007), it is consistent with the original statement that synapses from presynaptic neurons are reinforced to the postsynaptic neuron if those cells 'took part in firing it'. Importantly then, because presynaptic spikes must occur slightly before the postsynaptic spikes they contribute to, at a fundamental level, STDP captures the causality of interactions. Accordingly, most models of STDP

are 'phenomenological' in the respect that they are based purely upon the timing of spikes with no other neuronal variables or mechanisms. While they may feature 'plasticity' variables (updated according to the timing of pre- and postsynaptic spikes) they are only abstract place-holders for posited biophysical properties (for example Ca^{2+} concentration) and not directly derived from empirical measurements of such quantities.

The most commonly used formulations consider interactions between pairs of spikes, in either an online (Perrinet et al., 2001) or offline (Delorme et al., 2001; Song et al., 2000) form. Equation 1.13 shows a typical formulation of an 'offline', additive form of STDP (Masquelier et al., 2008) describing the change in synaptic weight between a presynaptic neuron (indexed by j) and a postsynaptic neuron (indexed by i) occurring upon the event of a pre- or postsynaptic spike[6]. The A coefficients adjust the strength of LTP/LTD effects (with $A_{+/-} \geqslant 0$) and the exponents model the decreasing sensitivity of the synaptic weight change with increasing time between events, (parametrized by their respective τ's) although simplifications are sometimes used where only the sign of the time difference (order of spikes) is considered (Masquelier and Thorpe, 2007).

$$\delta w_{ij}(t_j - t_i) = \begin{cases} A_+ \cdot \exp\left\{\frac{t_j - t_i}{\tau_+}\right\} & \text{if} \quad t_j \leqslant t_i \qquad \textbf{(LTP)} \\ -A_- \cdot \exp\left\{-\frac{t_j - t_i}{\tau_-}\right\} & \text{if} \quad t_j > t_i \qquad \textbf{(LTD)} \end{cases} \tag{1.13}$$

The total synaptic weight change induced by a simulation is thus the sum of each of these pair-wise effects across the whole trains of pre- and postsynaptic spikes, indexed by l and k respectively, as shown in Equation 1.14.

$$\Delta w_{ij} = \sum_{l=1}^{N_j} \sum_{k=1}^{N_i} \delta w_{ij} \left(t_j^l - t_i^k\right) \tag{1.14}$$

Equivalently this may be expressed in an 'online' form described by differential equations, as in the example given in Equation 1.15 adapted from (Song et al., 2000). Here P_{ij} and M_i are pre- and postsynaptic 'plasticity variables' respectively, which keep a trace of recent spiking activity and are incremented by presynaptic spikes (indexed by l) and postsynaptic spikes (indexed by k) modelled as Dirac delta functions. One scheme may be derived from the other through the use of Laplace transforms

[6]In alternative notations, the subtraction of spike times is commonly replaced by Δt_{ij} and the order of subtraction may be reversed, in which case $\Delta t < 0$ corresponds to LTP and $\Delta t > 0$ corresponds to LTD (Song et al., 2000).

and both are included here for completeness.

$$\tau_- \frac{dM_i}{dt} = -M_i - A_- \sum_k \delta\left(t - t_i^k\right) \quad \text{and} \quad \tau_+ \frac{dP_{ij}}{dt} = -P_{ij} + A_+ \sum_l \delta\left(t - t_j^l\right) \quad (1.15)$$

Whenever a synapse receives a presynaptic action potential, its weight is updated accordingly: $w_{ij} \leftarrow w_{ij} + M_i(t)$ — note the presynaptic event is paired with the postsynaptic variable. Similarly, upon a postsynaptic spike, the weight is updated according to the presynaptic variable: $w_{ij} \leftarrow w_{ij} + P_{ij}(t)$. Since additive STDP weight updates are unbounded, these updates must then be clipped, typically in the range $[0, 1]$ (before rescaling to a biologically meaningful change in conductance e.g. $\Delta g_{ij} = w_{ij} \cdot g_{max}$, where g_{max} is in the order of nS).

While most models consider only the nearest-neighbour spike pairs, others have considered higher order interactions such as triplets of spikes (Pfister and Gerstner, 2006) to account for the potentiation found in high frequency stimulation (unlike the predicted depression of pair-based models). In further work comparing this triplet model to the original pair-based model, it was found that there was little difference unless the neurons featured cell firing-rate adaptation, in which case the triplet model was much closer to the information rate of their artificially derived 'optimal' model (Hennequin et al., 2010).

Others considered more natural spike trains in their stimulation protocol of rat visual cortical slice Layer 2/3 pyramidal cells, finding that not only is the relative timing between the pre- and postsynaptic neuron important, but also the inter-spike intervals within each neuron (Froemke and Dan, 2002). They reported that including a suppressive interaction between consecutive spikes within each neuron better described the synaptic modifications from complex spike trains recorded *in vivo*, whereby the first spike in each burst is dominant. This model, based upon the spike pair intervals and the spike 'efficacy' (which is set to 0 immediately after a spike and recovers exponentially towards 1), may therefore provide a basis for the development of rank order coding (Delorme et al., 1999), helping to explain the extraordinary speed of processing in the primate visual system (Thorpe et al., 1996).

In addition to phenomenological differences, there are several basic implementational differences too. One such distinction concerns whether the updates are multiplicative (soft bounds), where the update has a weight dependence (Perrinet et al., 2001; Masquelier and Thorpe, 2007) or additive (hard bounds), where the update is independent of the current weight, as in the previous examples (Song et al., 2000; Masquelier et al., 2008).

For multiplicative forms of STDP, *presynaptic* variable P_{ij}, used to increase the weight upon a *postsynaptic* spike, will be multiplied by $(w_{max} - w_{ij}(t))$ while the *postsynaptic* variable M_i, contributing to LTD upon *presynaptic* spikes, is multiplied by $w_{ij}(t)$. The effect is that changes in the weights scale with their present values, becoming smaller as they approach the bounds w^{max} and w^{min} (typically 0), which overcomes the feedback instability of correlation-based Hebbian learning without the need for clipping. As a result, multiplicative and additive formulations give rise to significantly different weight distributions and therefore dynamics, yielding unimodal, more stable synaptic distributions or bimodal distributions with greater competition respectively (van Rossum et al., 2000).

To address the shortcomings of each regime, Gütig et al. (2003) proposed a model where weight dependence has the form of a power law, (as shown in Equation 1.16), which interpolates between these two extremes to provide a balance between stability and sensitivity.

$$\delta w_{ij}(\Delta t) = \begin{cases} \lambda(1-w)^{\mu} \cdot \exp\left\{-\frac{|\Delta t|}{\tau_+}\right\} & \text{if} \quad \Delta t > 0 \qquad \textbf{(LTP)} \\ -\lambda\alpha w^{\mu} \cdot \exp\left\{-\frac{|\Delta t|}{\tau_-}\right\} & \text{if} \quad \Delta t \leqslant 0 \qquad \textbf{(LTD)} \end{cases} \tag{1.16}$$

Here, λ is the learning rate ($0 < \lambda \ll 1$), with $\alpha > 0$ adjusting the asymmetry between LTP and LTD. The non-negative exponent, μ, interpolates between the two schemes, with $\mu = 0$ recovering the weight-independent additive form, $\mu = 1$ recovering the weight-dependent multiplicative form, and intermediate values providing a mixed mode. Note that this model adopts the alternative notation where $\Delta t = t_i - t_j$ (the reverse to that used in the previous formulations).

Experimental data suggest that the ratio of depressing/potentiating synaptic changes increases in a stabilizing manner as synaptic weights grow (Bi and Poo, 1998) but the exact details of the efficacy changes' dependence on the weight remains to be determined by further experimentation, with theoretical results pointing towards a mixture of the two forms (Gütig et al., 2003).

One of the main functional properties of STDP is the reduction of latency (Song et al., 2000). In a volley of input spikes, those occurring soon before the postsynaptic spike will be strengthened, while those occurring soon after will be weakened. If the input pattern (stimulus) is presented again, the previously strengthened synapses of the 'early' spiking neurons will be more likely to cause the same postsynaptic neuron to fire earlier because of the stronger inputs. Thus the mechanism of STDP could be enough to explain the incredible rapidity of information processing in feed-forward

neural networks, such as in the case of human object recognition (Thorpe et al., 1996) as demonstrated in a spiking variant of HMAX (Masquelier and Thorpe, 2007). This property has more recently been applied to a more realistic paradigm of continuous presynaptic population activity, whereby postsynaptic neurons are able to localize a repeating spatio-temporal spike pattern embedded within a larger population of uncoordinated activity (Masquelier et al., 2008).

STDP also conveys a number of other useful properties to neural circuits, such as the ability to convert the timing of individual spikes with millisecond precision into a spatially distributed pattern of persistent synaptic modifications in a 'delay line' architecture (Bi and Poo, 1999), which may be useful for learning temporal sequences. When modelled in conjunction with cell firing-rate adaptation, it has also been found to yield nearly optimal information transmission (Hennequin et al., 2010) making it a suitably efficient learning mechanism for the visual system, especially for the highly compressed neural code exiting the retinae along the optic nerves. Multiplicative forms of STDP also help to stabilize the firing rates of postsynaptic neurons without the need for weight normalization common to many rate-based learning rules (Song et al., 2000) and would be consistent with the greater degree of plasticity found in higher visual areas for novel stimuli than familiar faces (McMahon and Leopold, 2012).

While more experimental work is required in order to determine exactly how many spike interactions should be taken into account for accurate phenomenological models of STDP in different brain areas, they may be superseded by more detailed models which describe the phenomenon in terms of physiologically and pharmacologically measurable quantities such as Calcium ion and signalling protein concentrations or postsynaptic membrane potential. In the meantime (and for the sake of computational efficiency) these models are promising and have been mathematically mapped onto older rate based models such as the BCM learning rule (Pfister and Gerstner, 2006), demonstrating a promising unifying ability.

1.5 Overview of research conducted

Having introduced the subject of this thesis and reviewed the relevant literature, this Chapter closes with an overview of the research conducted and presented hereafter. Chapter 2 then details the spiking neural network developed to model the primate ventral visual stream and the methods used to analyse, interpret and visualize the results derived from it.

Chapter 3 investigates the extent to which the rate-coded Hebbian learning mechanisms, of CT (Stringer et al., 2006) and Trace learning (Földiák, 1991), may be implemented with a more biologically accurate spike-time dependent learning rule in a spiking neural network for learning to recognize objects in a model of the ventral visual stream.

Both the Trace and CT learning mechanisms have been previously applied successfully to this problem in rate-coded models to form transformation-invariant representations with both temporal continuity (Wallis and Rolls, 1997; Wyss et al., 2006; Michler et al., 2009; Isik et al., 2012) and spatial continuity (Stringer et al., 2006; Perry et al., 2010) of the stimuli respectively. The simulations conducted in Chapter 3 explore for the first time how these learning paradigms may be successfully adapted to a spiking neural network featuring STDP in its feed-forward excitatory synapses.

It is shown that with appropriate modification of the stimulus presentation regime and by changing the excitatory synaptic conductance time constant, τ_g^{EfE}, (modelling the dynamics of the postsynaptic ion channels), the model could operate in two dissociable modes to learn translation-invariant representations. In particular, a short synaptic time constant and high degree of spatial overlap between transforms allows the model to operate in a manner akin to CT learning, whereas a long synaptic time constant facilitated a form of Trace learning, whereby the network could associate together completely non-overlapping (orthogonal) views of the stimuli onto the same output neurons.

While both mechanisms are found to be robust to the order of transform presentation within a stimulus block, as expected Trace learning suffers when the transforms of two stimuli were interleaved, as follows from a violation of the assumption of temporal continuity of stimuli within natural scenes (Földiák, 1991; Wallis and Rolls, 1997; Sprekeler et al., 2007). The CT learning mechanism is found to require tight synchrony of the input volleys and temporally specific STDP time constants, whereas Trace learning is more robust to spike-timing jitter and performs better with longer STDP time constants.

Following from the simulations of simple stimuli used to investigate the transformation-invariance learning in Chapter 3, the aim of Chapter 4 is to demonstrate that the same principles apply when using more realistic three-dimensional stimuli. Through a combination of the Trace and CT learning mechanisms, it is demonstrated that the network can be trained to recognize both a cone and a cube as they translate across the input layer. Furthermore, it is demonstrated how additional sorts of transformation, such as rotations of view (in depth), may be associated together through the same

learning mechanisms thus generalizing the scope of the insights gained from simple abstract stimuli.

Some novel problems arise through using filtered greyscale images, such as the difficulty in representing particular transforms of low contrast with respect to the grey background, and remedies are discussed along with potentially interesting avenues of investigation which are open to a model using more detailed stimuli.

A central question in understanding the process of biological object recognition, is how to form separate representations of objects from natural visual scenes composed of multiple objects. Previous work has largely used stimuli presented in isolation (Wallis and Rolls, 1997; Stringer et al., 2006; Riesenhuber and Poggio, 1999; Masquelier and Thorpe, 2007), and so is lacking in ecological validity. Rate-coded models in particular suffer in this respect, as a rate-based model of Hebbian learning predicts that stimuli presented together will be associated together, leading to a superposition catastrophe (von der Malsburg, 1999). This is a theme addressed in several ways over the course of Chapters 5–7.

In Chapter 5, the standard one-layer feed-forward model is augmented with a lateral connectivity architecture similar to the standard 'Self-Organizing Map' (Kohonen, 1982; Miikkulainen et al., 2005) to address the question of how the extensive lateral connectivity of the visual system may aid learning to recognize objects in natural visual scenes composed of multiple objects. Guided by previous analyses of laterally connected spiking neurons (Nischwitz and Glünder, 1995; Miconi and VanRullen, 2010), it was predicted that the architecture and neuronal properties of the network would enable competing stimulus representations to push each other out of phase. In particular, the hypothesis tested is that with such antiphase oscillations ('perceptual cycles') between input layer representations, a time-sensitive STDP learning rule in the feed-forward synapses will be able to exploit the temporal segmentation of the stimuli as they translate across the input layer to learn separate stimulus representations in the output layer.

The key properties of the network are identified to be a mechanism of self-inhibition, consistent with previous work (Choe and Miikkulainen, 2003; Miconi and VanRullen, 2010), specifically cell firing-rate adaptation, and a gradient in the excitatory synaptic weights (which decrease exponentially with distance). Variations in these properties are investigated with respect to how they enable the network to synchronize features represented by proximal neurons and desynchronize them with respect to features represented by distal neurons, thus facilitating temporal segmentation of the input representations in a way suitable for transformation-invariant learning with STDP.

Following from the conclusions of Chapter 3, specifically that CT learning requires tight synchronization of spikes in an input volley (as slightly later spikes fall into the STDP rule's LTD window) and that Trace learning is less temporally specific (with respect to input synchrony and the STDP time constants), only the CT learning mechanism was explored in these simulations, as a Trace learning paradigm would undermine the dynamics of the temporal segmentation of stimuli.

The work shows how the spatial separation of the stimuli is effectively converted into a temporal separation (through the lateral weight gradient and firing-rate adaptation) in a way that allows separate translation-invariant representations to be formed even when both stimuli are always presented together (with no statistical decoupling) and moving in lockstep. This is an advancement on previous work with multiple stimulus presentation paradigms in rate-coded networks where, either ecologically implausible training regimes are required to statistically decouple the objects (Stringer et al., 2007; Stringer and Rolls, 2008), or independent motion of the objects is necessary (Tromans et al., 2012).

The extraordinary computational power of neurons lies in their ability to communicate with one another through their synapses. Their connections may be very short or extend across millimetres of cortex (Miikkulainen et al., 2005) and so the delay in conducting their action potentials along their axons (with or without myelination) varies accordingly from the order of one millisecond (Girard et al., 2001) to the order of tens (Swadlow, 1992) or even hundreds of milliseconds (Aston-Jones et al., 1985). Following from these experimental data, Chapter 6 investigates the role of axonal conduction delays, firstly in the feed-forward axons of a variant of the model used in Chapter 3, and secondly as a replacement for the Gaussian profile of lateral excitatory weights used in Chapter 5.

In using a constant axonal conduction delay for the feed-forward connections, both the CT and Trace learning mechanisms are unaffected in their capacity to learn transformation-invariant representations in the output layer when the STDP rule also incorporates the delay term. This is a novel change to the original formulation of Perrinet et al. (2001), which like many other models of STDP, neglects axonal delays (Song et al., 2000). When the unmodified version of STDP is implemented in the model, a pronounced and different effect emerges as the axonal conduction delays are increased between the CT and Trace learning regimes. These are argued to be artefactual however, with the natural conclusion being that when implementing axonal delays in such a model, the STDP rule should also be revised accordingly.

Gaussian distributions of delays in the feed-forward connections are also investigated. Here, another marked difference between the CT and Trace learning mechanisms is discovered, whereby the network performance of CT learning is severely degraded with the increasing standard deviation of the distribution, whereas Trace learning is found to be far more robust, as expected, due to their respective differences in sensitivity to the synchrony of the inputs.

The second section of Chapter 6 considers the computational benefits of axonal conduction delays and demonstrates that they can replace the role of the gradient in the excitatory lateral connection strengths to temporally segment multiple stimuli through the dynamics of perceptual cycles. The delays in this section were proportional to the Euclidean distance between the neurons giving a graduation of delays and thereby extends the simple two neuron case of Nischwitz and Glünder (1995) and the single neural layer of Miconi and VanRullen (2010) which each involved a fixed delay. It is predicted that the optimal delay between the stimulus representations was half the period of oscillation, such that the wave of excitation arrives midway through the cycle of the afferent neurons, encouraging their targets to fire at that point. The results however indicate that a delay $1\,ms$ lower was very slightly better for network performance, which may be due to the time required for the synaptic conductances to drive up the postsynaptic membrane potential.

In the final body of experimental work, Chapter 7 investigates a third solution to the problem of separating representations of multiple translating stimuli. Rather than imposing a fixed lateral weight structure or attempting to match the lateral delays to the (half) period of oscillation, plasticity is introduced into the lateral connections and a preliminary training phase is introduced to allow them to self-organize through exposure to exemplars of categories. During this preliminary phase, knowledge of the categories is built up in the lateral connections which is then found to enable the network to temporally segment a visual scene composed of two novel exemplars.

While initially the effects of 'early' exemplar exposure are investigated in a single layer of excitatory neurons (and inhibitory interneurons), the second section augments this layer with another and adds feed-forward plastic connections between them, while translating stimuli are presented to the input layer. This model is then used to demonstrate that category learning in the lateral connections allows the output layer to form independent translation-invariant representations for previously unseen stimulus representations. In addition to the lack of statistical decoupling and independent movement, in contrast to the work of Chapters 5 and 6, the use of distributed stimulus

representations meant that spatial separation is not required to temporally segment the stimuli.

The input layer dynamics generated in this model are shown to be robust to wide ranges of several parameters, including the strength of inhibitory and excitatory connections and the adaptation model strength and time constant. The effect upon the subsequent training of the feed-forward connections is also explored, principally through systematically varying the parameters of the STDP model, which is found to be similarly robust to the earlier simulations of this thesis involving multiple stimulus presentation paradigms.

Finally, the thesis concludes with an evaluation of the research presented, a discussion of the themes to emerge throughout the course of the investigations and some suggestions for future directions in which to progress the field of research.

Build it, and you understand it.

—John J. Hopfield

2 Methods

While there are many freely available simulators of neurons covering a large range of scales of detail, from specifying subcellular components (for example *NEURON* and *GENESIS*) to large scale networks of $\mathcal{O}(10^5)$ neurons (for example *NEST*), the model presented was written by myself in order to provide the flexibility and understanding appropriate to this thesis. For a review of other available simulation tools, please see Brette et al. (2007).

2.1 Network Architecture

The ventral visual stream is typically modelled as four or more layers of neurons with excitatory modifiable feed-forward synapses and a mechanism of lateral inhibition, with each qualitatively the same as the last (Fukushima, 1988; Riesenhuber and Poggio, 1999; Wallis and Rolls, 1997; Delorme et al., 1999; Wu et al., 2008) embodying the notion that each layer of the ventral visual system performs fundamentally the same type of computation. However, this work seeks to understand the mechanisms operating at *each* layer that ultimately contribute to achieving transformation-invariant representations, hence a simpler architecture is used in order to facilitate a more detailed inquiry.

Typically, the model consists of two layers of excitatory pyramidal neurons with one layer of modifiable feed-forward synapses between them (as shown in Figure 2.1). Within each layer there are also inhibitory shunting interneurons with non-plastic

lateral synaptic connections to and from the excitatory neurons, to produce a degree of competition between the excitatory neurons. The probability for making each class of synapse (from afferent connections to a given postsynaptic neuron) was specified for each layer and for most presented simulations, the network was fully connected.

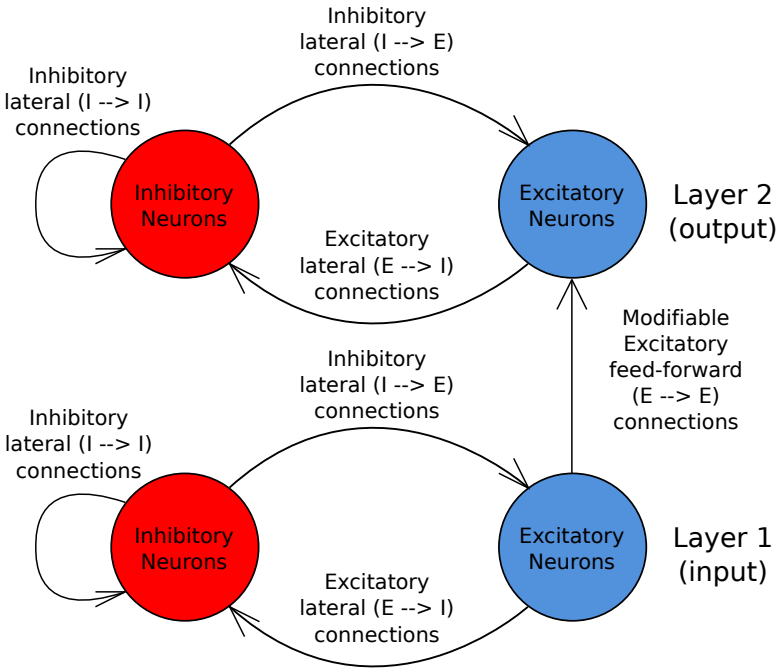


Figure 2.1: Schematic of the two-layer network architecture used. Excitatory neurons within each layer are fully reciprocally connected to the pool of inhibitory interneurons within the same layer by fixed synaptic efficacies. Inhibitory neurons are also fully connected to other inhibitory neurons within the same layer by unmodifiable synapses. The excitatory neurons in the input layer are fully connected to the excitatory neurons in the output layer via plastic feed-forward synapses which are modified through training by an STDP learning rule.

In most simulations, the number of excitatory (principal) cells within each layer was in the order 400–500 with 1/4 as many inhibitory neurons in each layer (Izhikevich et al., 2004; Izhikevich, 2006). Each neuron is based upon the standard conductance-based leaky integrate and fire (LIF) model (see for example Rolls and Treves, 1998; Gerstner and Kistler, 2006) while the equations for STDP at the Excitatory-Excitatory ($E \rightarrow E$) synapses are adapted from Perrinet et al. (2001).

In general, reciprocity is the rule for cortico-cortical connections and in particular, feed-back connections are extensive within the visual cortex (Felleman and Van Essen, 1991). However, given the speed of processing in object recognition or discrimination tasks where categorizing takes only $\sim$100–200 ms (Thorpe et al., 1996; Hung et al.,

2005; Afraz et al., 2006), it appears that feed-forward connections are sufficient for object recognition and so feed-back projections may be reasonably omitted as a starting point. When testing the limits of the speed of processing, feed-back projections do not have enough time to affect the object recognition process, however they may still influence the process in a priming capacity, by attentional shifts (DiCarlo and Cox, 2007; Felleman and Van Essen, 1991), in memory processes or by allowing stimuli outside the 'classical' receptive field to influence stimuli within (Felleman and Van Essen, 1991).

Lateral reciprocal intra-area excitatory (ElE) connections are also found extensively in visual cortex and form a significant theme of investigation in Chapters 5, 6 and 7.

2.2 Neuron model

Understanding any complex system requires careful choice of the level of description, such that important properties are retained whilst removing details unnecessary for understanding the behaviour of interest. For example, the spatial extent of the dendritic tree is neglected (eschewing the use of partial differential equations), along with the conductance-based description of the spiking process (captured by the Hodgkin–Huxley equations). Such simplifications were necessary in order to greatly reduce the computational expense of the simulations and pare down the complexity of the biological system to its core principles in order to explain clearly the phenomena of interest.

The neuron model used in these simulations is the Conductance-based Leaky Integrate-and-Fire (gLIF) model, which is best described in terms of its commonly used variants. Like the standard Integrate-and-Fire (IF) neuron model (Burkitt, 2006a,b), which is derived only from the law of capacitance ($Q = CV$), the gLIF cell's membrane potential is driven up by currents entering the cell body, until it crosses a threshold, (ϑ) at which point it emits a spike (modelled by a Dirac delta function) and is reset before continuing to run. However, the ordinary differential equation (ODE) describing the membrane potential is of the same form as for the Leaky Integrate-and-Fire (LIF) neuron (derived from a Resistor and Capacitor in parallel in Section 2.2.1), which unlike the standard IF neuron, has a leakage term to model the natural permeability of the membrane. This leakage term allows the cell potential to decay back to rest in the absence of stimulating currents, reflecting the non-equilibrium diffusion of ions across the membrane. This means that the 'memory' of previous stimulation (or any perturbation from rest) gradually fades with time,

unlike the simpler IF model neurons which, after a sub-threshold degree of stimulation, would remain at an excited (or inhibited) state until the threshold is crossed and the cell potential is reset.

The standard LIF neuron typically models the incoming current as the sum of presynaptic spikes with an exponential decay. The gLIF however differs by modelling incoming spikes as opening ion channels in the postsynaptic cell membrane and allowing the flow of these ions across the membrane (through the channels) according to the difference in the synaptic class's (presynaptic cell's) reversal potential and the postsynaptic cell's membrane potential. In this respect, the gLIF shares some of the properties of the seminal Hodgkin–Huxley model (HH), yet is simplified in the respect that it lacks the additional complexity of ordinary differential equations describing the evolution through time of the gating variables for activation and inactivations of the ion channels.

2.2.1 Derivation from a Resistor-Capacitor circuit

The Leaky Integrate-and-Fire (LIF) Neuron displays the minimal ingredients of membrane dynamics. The basic circuit from which it is derived (shown in Figure 2.2 and adapted mainly from Gerstner and Kistler 2006) consists of a capacitor C in parallel with a resistor R, driven by a current $I(t)$. In it's simplest form, incoming spikes drive up the incoming current $I(t)$ directly, whereas in the closely related conductance-based LIF, incoming spikes open ion channels allowing a flow of charge governed by the dynamics of the membrane (see later).

In an RC circuit, the current may be divided into two components:

$$I(t) = I_R + I_C \tag{2.1}$$

The first component, I_R is the resistive current which flows through the resistor R. From Ohms Law, this may be expressed as $I_R = {}^{V}/_{R}$ (where V is the voltage across the resistor). The second component, I_C is the current charging the capacitor, C. From the definition of an ideal capacitor, $C = {}^{Q}/_{V}$ and the definition of current as the rate of flow of charge, we obtain the integral form of the capacitor equation:

$$V(t) = \frac{Q(t)}{C} \tag{2.2}$$

$$= \frac{1}{C}\int_{t_0}^{t} I_C(\tau)d\tau + V(t_0) \tag{2.3}$$

Where $V(t_0)$ is the constant of integration added to represent the initial voltage. Taking the derivative and multiplying by C thus yields:

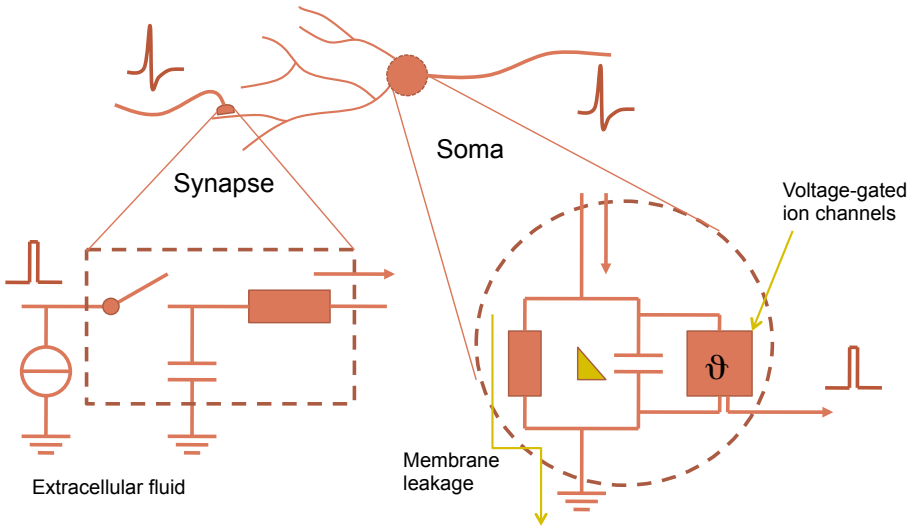


Figure 2.2: Schematic of the synapse and soma for the leaky integrate-and-fire neuron.

$$I_C(t) = \frac{dQ(t)}{dt} \tag{2.4}$$

$$= C\frac{dV(t)}{dt} \tag{2.5}$$

The total current may thus be expressed as:

$$I(t) = \frac{V(t)}{R} + C\frac{dV(t)}{dt} \tag{2.6}$$

Finally, multiplying by R and defining the membrane time constant, $\tau_m = RC$ yields the standard form:

$$\tau_m \frac{dV}{dt} = -V(t) + RI(t) \tag{2.7}$$

In the LIF neuron, the shape of action potentials are not explicitly modelled (in contrast to the more detailed ion channel dynamics of e.g. the Hodgkin–Huxley model). Instead spikes are discrete binary events described entirely by their firing time t^k. A spike occurs when the voltage across the capacitor reaches a threshold ϑ.

$$t^k : V(t^k) = \vartheta \tag{2.8}$$

Immediately after the spike occurs, the membrane potential is reset to the after-spike hyperpolarizing potential V_H where it is held for a refractory period τ_R.

$$\lim_{t \to t^k; t > t^k} V(t) = V_H \tag{2.9}$$

After the absolute refractory period, the integration is restarted at time $t^k + \tau_R$ (with initial condition $V(t^k + \tau_R) = V_H$) as described in Equation 2.7.

The time course of the membrane potential when stimulated with a constant current ($I(t) = I_0$) may be calculated analytically by integrating Equation 2.7 with the initial condition $V(t^k) = V_H$ as follows:

$$V(t) = I_0 R + (V_H - I_0 R) e^{-\frac{t - t^k}{\tau_m}} \tag{2.10}$$

In order to find the time between two successive spikes (t^k and t^{k+1}), we set $V(t) = \vartheta$ (thus giving us the time between two threshold crossings).

$$\vartheta = I_0 R + (V_H - I_0 R) e^{-\frac{t^{k+1} - t^k}{\tau_m}} \tag{2.11}$$

By rearranging to give the time between two successive spikes, the firing rate under constant current is simply $f = {}^1/_{\Delta t}$, where $\Delta t := t^{k+1} - t^k$.

$$\frac{1}{f} = \Delta t = \tau_m \cdot \ln \left[\frac{V_H - I_0 R}{\vartheta - I_0 R} \right] \tag{2.12}$$

If an absolute refractory period is incorporated into the model, then the time between spikes is simply extended by the refractory period and the firing rate is given by:

$$f = \frac{1}{\tau_R + \tau_m \cdot \ln \left[\frac{V_H - I_0 R}{\vartheta - I_0 R} \right]} \tag{2.13}$$

Figure 2.3 shows the frequency of firing obtained in a single model neuron plotted against the injected current using the standard parameters from Troyer et al. (1998), given in Table 2.2. The 'f-I curves' are plotted from a neuron simulated with cell membrane noise (see Section 2.6) and are the mean values obtained averaged over 20 different random seeds (with the standard deviation for each value indicated by the error bars). A curve is plotted for both the standard LIF neuron model and the modified version featuring cell firing-rate adaptation (see Section 5.2.2) which has the effect of linearizing the response curve (Koch, 1999, p337).

The simulation data of Figure 2.3 are in good agreement with the analytically calculated current injection threshold for spiking of 52.5nA. This is calculated as the point where the currents injected into the cell exactly balance those leaking out, resulting in no net change, that is, ${}^{dV(t)}/_{dt} = 0$ in Equation 2.16, and by setting $V(t) = \Theta$ (the firing threshold) before substituting in the other numerical constants of the model (Table 2.2) and solving for I. In the simulations presented, there is

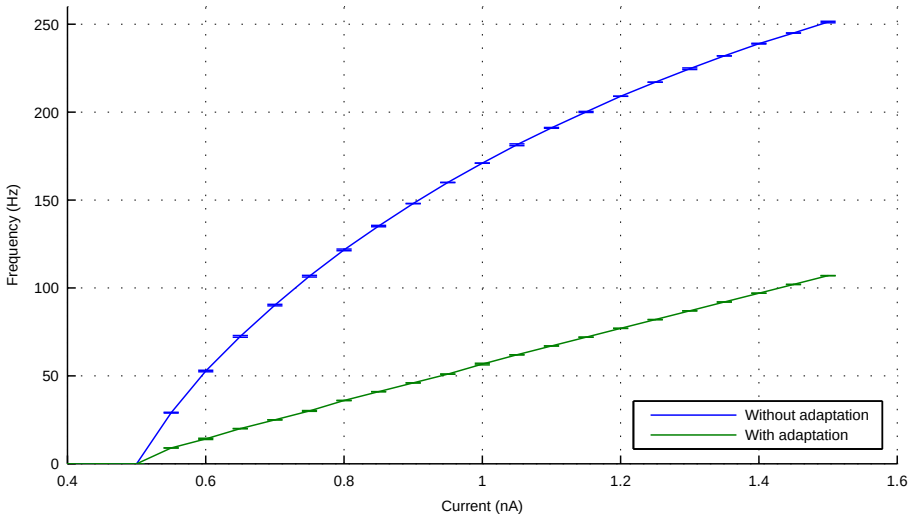


Figure 2.3: Firing frequency plotted against direct stimulating current for a single noisy neuron averaged over 20 random seeds, with and without firing-rate adaptation.

effectively some variation around this threshold for when spikes occur due to the effects of noise in the cell membrane potential.

Presynaptic spikes are low-pass filtered at the synapse and generate an input current pulse. In the simplest form of the LIF, the total incoming presynaptic current is composed of the sum of presynaptic spikes (here indexed by k), multiplied by their synaptic efficacy or 'weight'.

$$I_i(t) = \sum_j w_{ij} \sum_k \delta(t - t_j^k) \tag{2.14}$$

In a more detailed 'conductance-based' model, the incoming spikes open ion channels which then flow (in either direction according to the type of synapse) across the membrane and so the synaptic weight is defined as the change caused by a single presynaptic spike in the membrane conductance of the postsynaptic cell, Δg. Since spikes affect the conductance, the resulting change in the incoming presynaptic current depends upon the state of the postsynaptic cell's membrane potential ($V_i(t)$) at the time of the spike's arrival.

$$I_i(t) = \sum_j g(t - t_j^k) \left(\hat{V} - V_i(t)\right) \tag{2.15}$$

The conductance for each synapse is governed by its own differential equation, ensuring the conductance decays to 0 in the absence of any spikes as described in

Equation 2.19.

2.2.2 Differential Equations

Depolarization of the neuron's membrane potential is described by Equation 2.16 and the cell (and synapse) constants were chosen to be as biologically accurate as possible based upon the available neurophysiological literature (see Tables 2.2 and 2.3 for a full list).

The cell membrane potential for a given neuron (indexed by i) is driven up by presynaptic excitatory conductances (or direct current injection) and down[1] by presynaptic inhibitory conductances, decaying back to its resting state over a time course determined by the properties of its membrane.

$$C_m \frac{dV_i(t)}{dt} = g_0 \left(V_0^{\gamma} - V_i(t)\right) + I_i(t) + I_i^{ext}(t) \tag{2.16}$$

Here C_m represents the membrane capacitance and g_0 represents the membrane leakage conductance (the inverse of the membrane resistance: $1/R_m$) and the membrane time constant, τ_m, is defined as $\tau_m = C_m/g_0$. V_0 denotes the resting potential of the cell of a particular class (indexed by γ), $I_i(t)$ represents the total synaptic current (described in Equation 2.17) and $I_i^{ext}(t)$ models the injected current.

The total synaptic current is the sum of currents of all presynaptic neurons of each type (excitatory and inhibitory) with inhibitory currents being negative.

$$I_i(t) = \sum_{\gamma} \sum_{j} g_{ij}(t) \left(\hat{V}^{\gamma} - V_i(t)\right) \tag{2.17}$$

Here $\hat{V}$ represents the reversal potential of a particular class of synapse (denoted again by γ) which consists of Excitatory and Inhibitory neurons $\{E, I\}$ and j indexes the presynaptic neurons of each class.

2.2.3 Synaptic Conductance Equations

The synaptic conductance of a particular synapse, $g(t)$, (indexed by ij) is governed by a decay term τ_g and a Dirac delta function for when spikes occur, which correspond

[1] More precisely, inhibitory synaptic conductances drive the postsynaptic cell membrane potential towards the inhibitory reversal potential, typically below the membrane resting potential for hyperpolarizing inhibition or near to resting potential for shunting inhibition

to the first and second terms of Equation 2.19 respectively. The Dirac delta function is defined as the derivative of the Heaviside function or as follows:

$$\delta(x) = \begin{cases} \infty & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{where, } \int_{-\infty}^{+\infty} \delta(x)dx = 1. \tag{2.18}$$

The Dirac delta function is an infinitely large and infinitely short impulse in the differential equation, meaning that it instantaneously adds a finite amount (equal to its coefficient) to the conductance of that particular synapse when its argument is 0 (in this case, when the action potential reaches the synapse). As such, this model simplifies the shape of its biological counterparts from being a rapid but finite rise and fall in the synaptic conductance to a discrete jump (or a jump directly in the cell membrane potential in the case of an action potential arriving at the non-conductance-based model).

The conduction delay for a particular synapse is denoted by Δt_{ij} and each spike is indexed by l as a separate train for each presynaptic neuron. A biological scaling constant, λ (set in all simulations to be $5\,ns$), with further scaling for each class of synapse, have been introduced to scale the synaptic efficacy Δg_{ij} which lies between unity and zero.

$$\frac{dg_{ij}(t)}{dt} = -\frac{g_{ij}(t)}{\tau_g} + \lambda \Delta g_{ij}(t) \sum_l \delta\left(t - \Delta t_{ij} - t_j^l\right) \tag{2.19}$$

For further details of these parameters, please refer to Table 2.3.

2.2.4 Numerical Scheme: Cell Equations

In order to numerically approximate the differential equations at particular points in time (time-steps), the Taylor series may be used to represent the solution function as the infinite sum of terms calculated from the function's derivatives at each point.

$$y(t + \Delta t) = y(t) + \frac{y'(t)}{1!}\Delta t + \frac{y''(t)}{2!}\Delta t^2 + \frac{y'''(t)}{3!}\Delta t^3 + \cdots + \frac{y^{(n)}(\tau)}{n!}\Delta t^n \tag{2.20}$$

$$= \sum_{n=0}^{\infty} \frac{y^{(n)}(t)}{n!}\Delta t^n \tag{2.21}$$

where τ is a value between t and $t + \Delta t$ and $y^{(n)}(t)$ is the n^{th} derivative of y at point t.

Typically numerical schemes are based upon truncating this series after a certain number of terms. In the case of the Forward Euler scheme (Atkinson and Han, 2004), the true solution is approximated by truncating the infinite sum after two terms (see

Equation 2.23, which gives an error term which shrinks linearly with the step size — $\mathcal{O}(\Delta t)$.

$$y(t+\Delta t) = y(t) + \frac{y'(t)}{1!}\Delta t + \frac{y''(\tau)}{2!}\Delta t^2 \tag{2.22}$$

$$= y(t) + \Delta t y'(t) + \mathcal{O}(\Delta t^2) \tag{2.23}$$

Equation 2.23 shows the local truncation error (LTE) for the Forward Euler scheme used, which is the error induced at every time-step due to the truncation of the Taylor series. In evaluating the overall accuracy of the scheme, we must consider the global error which is the absolute difference between true solution and the computed solution at a given time-step. While the LTE may be $\mathcal{O}(\Delta t^2)$, it is reasonable to neglect rounding-off errors and assume that the global error at the n^{th} time-step ($y(t+n\Delta t)$) is n times the LTE. Since $n \propto {}^{1}/_{\Delta t}$ the global error should be proportional to ${}^{\text{LTE}}/_{\Delta t}$. This means that for the first-order Forward-Euler scheme, the global error scales according to $\mathcal{O}(\Delta t)$.

While other schemes converge much faster such as Runge–Kutta (Atkinson and Han, 2004), Bulirsch–Stoer and Parker–Sochacki (Stewart and Bair, 2009) and therefore offer a much higher degree of accuracy for the same step-size, the Forward Euler scheme is chosen for its simplicity in order to facilitate more rapid changes to the model during its development.

The differential equations described above are converted to finite difference equations and simulated using the Forward Euler numerical scheme with a time-step $\Delta t = 0.02\,ms$. In the finite difference equations, the Dirac delta function has been replaced by the discrete approximation, $S(x)$ as defined in Amit and Brunel (1997), which is defined below.

$$S(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.24}$$

Finally, in the original description, the change in synaptic weight (Equation 2.29) was instantaneous and so ${}^{\Delta t}/_{\tau_{\Delta g}}$ is defined to be a learning rate constant, ρ, in the corresponding finite difference equation.

Equation 2.25 shows the finite difference equation for updating a neuron's cell potential (V) in full.

$$V_i^{t+1} = V_i^t + \frac{\Delta t}{C_m}\left(g_0(V_0^\gamma - V_i^t) + \sum_\gamma \sum_j g_{ij}^t \left(\hat{V}^\gamma - V_i^t\right) + I_i^{ext}\right) \tag{2.25}$$

2.2.5 Numerical Scheme: Synaptic Conductance Equations

Below, Equation 2.26 describes the update of the presynaptic conductance. In these finite difference equations, the Dirac delta function (which is continuous) has again been replaced by the discrete approximation, $S(x)$ (Amit and Brunel, 1997).

$$g_{ij}^{t+1} = g_{ij}^{t} - \frac{\Delta t g_{ij}^{t}}{\tau_g} + \lambda \Delta g_{ij}^{t} \sum_{l} S\left(t - \Delta t_{ij} - t_j^l\right) \tag{2.26}$$

2.3 Spike-Time Dependent Plasticity

As discussed in Chapter 1, the model of STDP used throughout this thesis is phenomenological and incorporates a weight dependence in the updates making it multiplicative (van Rossum et al., 2000; Gütig et al., 2003), being based upon the work of Perrinet et al. (2001). It is implemented in an 'online' form, meaning that a set of ordinary differential equations describing the LTP and LTD components are updated each time-step and another equation describing their interaction is updated upon a pre- or postsynaptic spike occurring at that particular synapse. In addition to the previously introduced equations describing the cell membrane potential and synaptic conductances, this model of STDP requires another two state variables (and corresponding ODEs) per synapse and one more per postsynaptic neuron.

2.3.1 Synaptic learning (Plasticity) Equations

The following differential equations describe the STDP occurring at each modifiable Excitatory-Excitatory ($E \rightarrow E$) synapse. Here i labels the *post*-synaptic neuron. The recent *pre*-synaptic activity, $C_{ij}(t)$, is modelled by Equation 2.27 which may be interpreted as the concentration of neurotransmitter (glutamate) released into the synaptic cleft (Perrinet et al., 2001) and is bounded by $[0, 1]$ for $0 \leqslant \alpha_C < 1$. Here, t_j^l is the time of the l^{th} spike emitted by the j^{th} presynaptic cell.

$$\frac{dC_{ij}(t)}{dt} = -\frac{C_{ij}(t)}{\tau_C} + \alpha_C \left(1 - C_{ij}(t)\right) \sum_{l} \delta\left(t - \Delta t_{ij} - t_j^l\right) \tag{2.27}$$

The presynaptic spikes drive $C_{ij}(t)$ up at a synapse according to the model parameter α_C and the current value of $C_{ij}(t)$ which then decays back to 0 over a time course governed by τ_C.

The recent postsynaptic activity, $D_i(t)$, is modelled by Equation 2.28 which may be interpreted as the proportion of unblocked NMDA receptors as a result of recent

depolarization through back-propagated action potentials (Perrinet et al., 2001). Here t_i^k is the time of the k^{th} spike emitted by the i^{th} postsynaptic cell.

$$\frac{dD_i(t)}{dt} = -\frac{D_i(t)}{\tau_D} + \alpha_D \left(1 - D_i(t)\right) \sum_k \delta\left(t - t_i^k\right) \tag{2.28}$$

Unlike with the conduction of presynaptic action potentials to postsynaptic neurons, there is no conduction delay associated with D_i since the cell body is assumed to be arbitrarily close to the receiving synapses, and it is the same for a given (postsynaptic) neuron rather than each of its synapses, since the effects of a postsynaptic spike are assumed to have an equal impact on all receiving synapses.

The strength of the synaptic weight, $\Delta g_{ij}(t)$, is then modified according to Equation 2.29, which is governed by the time course variable $\tau_{\Delta g}$.

$$\tau_{\Delta g} \frac{d\Delta g_{ij}(t)}{dt} = \left(1 - \Delta g_{ij}(t)\right) C_{ij}(t) \sum_k \delta\left(t - t_i^k\right) - \Delta g_{ij}(t) D_i(t) \sum_l \delta\left(t - \Delta t_{ij} - t_j^l\right) \tag{2.29}$$

Note that the *post*synaptic spike train (indexed by k) is now associated with the *pre*synaptic state variable (C) and vice versa. If C is high (due to recent presynaptic spikes) at the time of a postsynaptic spike, then the synaptic weight is increased (LTP) whereas if D is high (from recent postsynaptic spikes) at the time of a presynaptic spike then the weight is decreased (LTD).

The weight updates are also multiplicative, meaning that the amount of potentiation decreases as the synapse strengthens, as has been found experimentally (Bi and Poo, 1998). This weight-dependent potentiation yields a normal distribution of synaptic efficacies rather than pushing each weight to one extreme or the other (van Rossum et al., 2000) as the subsequent simulations show.

2.3.2 Numerical Scheme: Plasticity Equations

$$C_{ij}^{t+1} = C_{ij}^t - \frac{\Delta t C_{ij}^t}{\tau_C} + \alpha_C \left(1 - C_{ij}^t\right) \sum_l S\left(t - \Delta t_{ij} - t_j^l\right) \tag{2.30}$$

$$D_i^{t+1} = D_i^t - \frac{\Delta t D_i^t}{\tau_D} + \alpha_D \left(1 - D_i^t\right) \sum_k S\left(t - t_i^k\right) \tag{2.31}$$

In Perrinet's original description, the change in synaptic weight was instantaneous and so $\dfrac{\Delta t}{\tau_{\Delta g}}$ is defined to be a learning rate constant, ρ, in the equation for synaptic

learning below (Equation 2.32).

$$\Delta g_{ij}^{t+1} = \Delta g_{ij}^{t} + \rho \left(\left(1 - \Delta g_{ij}^{t}\right) C_{ij}^{t} \sum_{k} S\left(t - t_i^k\right) - \Delta g_{ij}^{t} D_i^t \sum_{l} S\left(t - \Delta t_{ij} - t_j^l\right) \right) \tag{2.32}$$

2.4 Computational Complexity

The conductance-based leaky integrate-and-fire neuron is a detailed model of a biological neuron and consists of one ordinary differential equation for each cell body and one more for each synapse. (With the online model of plasticity used, there is an additional ODE for each cell body and two more for each synapse.) Each time-step in a clock-driven implementation consists of two parts: (1) state updates and (2) the propagation of spikes.

The cost of the update phase therefore scales according to $\mathcal{O}(N + N \times S)$, where N is the number of neurons and S is the average number of (incoming) synapses per neuron[2]. Using the approximation $S \approx p \times N$ where p is the average probability of forming a presynaptic connection with another neuron, (either in the previous layer or laterally within the same layer) then the cost of the update is $\mathcal{O}(N(1 + p \times N)/\Delta t)$ per second of biological time. Thus the time-cost for the update phase depends strongly on the size of the network model and the precision of the simulation.

If we assume an average firing rate F for each neuron, then $F \times N$ spikes will be produced per second of biological time. Furthermore, given there are as many postsynaptic connections as there are presynaptic connections, then each spike will be propagated to $p \times N$ neurons, meaning a total of $F \times pN^2$ propagations (and therefore synaptic weight updates) per second of biological time[3]. These operations will consist of adding 'weights' or conductance differences to state variables and are relatively cheap operations. In this analysis, the cost of adjusting the synaptic weight upon a spike event for the STDP model is also subsumed into the cost of 'propagation'.

The total computational cost (per second of biological time) is therefore the sum of these two components weighted by respective costs — C_U, the cost of an update

[2]N.B. As a simplification to speed up computations, the input layer neurons have no incoming synapses; instead current is driven directly into their cell bodies corresponding to abstract patterns or filter outputs.

[3]N.B. In a simpler model without learning, the synaptic conductances can be agglomerated on each postsynaptic neuron such that there is only one update to $\tau_G dG/dt = -G$ per neuron per time-step (where $G = \sum_s g_s$) rather than for each synapse, $\tau_g dg_s/dt = -g_s$, where $s = 1, ..., S$. This would mean that the cost of an update scales only linearly with the size of the network as $\mathcal{O}(N/\Delta t)$.

and C_P, the cost of propagating a spike (Brette et al., 2007):

$$\mathcal{O}\left(C_U \times \frac{N + p \times N^2}{\Delta t} + C_P \times F \times p \times N^2\right) \tag{2.33}$$

Although left out of the above analysis for simplicity, synaptic spike transmission delays are an important part of the model and further add to the cost of the computations (albeit only slightly). During the update phase at each time-step, if a threshold crossing is detected in the neuron's cell potential, a spike is inserted into a circular array of scheduled synaptic events with each element corresponding to a time bin (which is preallocated to be large enough to accommodate the maximum possible number of spikes which could occur in the delay period). Using modular arithmetic in circular arrays in this way means that each synaptic event is stored and retrieved exactly once, with the following computational cost per second of biological time (Brette et al., 2007):

$$C_D \times F \times N \times p \tag{2.34}$$

Where C_D is the (low) cost of storing and retrieving synaptic events in the circular arrays. As such, delays increase the cost of propagation by a small multiplicative factor. Similarly, adding additional features to the model, such as cell firing-rate adaptation will add another differential equation to the cell body and hence add to the constant cost of the update, C_U, rather than changing the underlying computational complexity analysis.

Given that the computational cost of the model scales with N^2, it is important to keep the size of the model to a minimum possible in order to keep the simulations tractable. The cost is also heavily dependent upon the size of the time-step and so the largest possible value should be used without qualitatively affecting the results.

2.5 Mesh Refinement

When using a numerical scheme to solve differential equations such as the Forward Euler scheme, it is necessary to ensure not only that the schemes are stable (i.e. the solution converges) which holds when the condition in Equation 2.36 is satisfied; the Courant–Friedrichs–Lewy condition (1928), but also that the numerical approximations calculated are done so with sufficient accuracy to reliably represent the true solutions. While the threshold for stability is known analytically, the largest usable time-step is

found through investigation.

$$\left|1 - \frac{\Delta t}{\tau}\right| < 1 \tag{2.35}$$

$$\Rightarrow \quad 0 < \Delta t < 2\tau \tag{2.36}$$

In order to test that the Forward Euler scheme had an appropriate time-step, a mesh refinement study was carried out by decreasing the numerical time-step, Δt, until no further significant change was found in the numerical results. By inspection, while $\Delta t = 0.1\,ms$ was sufficient for a qualitative understanding of the dynamics, the solution appeared to converge at $\Delta t = 0.02\,ms$, and so this was used throughout the thesis.

2.6 Gaussian white noise in the LIF neuron

The brain is an inherently noisy system and as such, it is important to investigate the effects of this noise upon its ability to process information from its senses. Considering the brain as a macroscopic system, there are two common sources from which this noise is introduced (*i*) the randomly arriving external Poisson spike trains and (*ii*) statistical fluctuations due to the spiking of individual neurons (discrete binary events) in a finite sized network (Rolls and Deco, 2010, p267).

On a microscopic scale at the level of the individual neuron, inputs to the neuron may still be modelled as a Poissonian spike train (except for at the very beginning of neural pathways in sensory transducers). Additionally, there is a source of stochastic fluctuations in the cell potential. This models effects such as noise in the diffusion of neurotransmitters at the synaptic cleft and in the synaptic currents due to the probabilistic transitions of the voltage-gated ion channels in the cell membrane. While this is most naturally implemented by adding Gaussian white noise to the synaptic currents (or conductances), it may also be approximated by adding noise into the sum of synaptic currents (Brunel and Hakim, 1999; Masquelier et al., 2009).

Qualitatively, the potential effects on the spike trains coming out of the cells of introducing such noise may be of two kinds: (*i*) jitter, whereby the spikes occur earlier or later than they otherwise would have (in a deterministic set of equations) and (*ii*) insertion or deletion, whereby noise may force the cell potential over the firing threshold, thus creating a spike which otherwise would not have occurred, and conversely, the cell potential may be randomly brought away from the firing threshold, deleting a spike which otherwise would have occurred. These two measures may be used to characterize the similarity between two biological spike trains and have

been successfully applied to several measures of synchrony in spike trains such as the Victor-Purpura metric (for a comparison of different metrics see Kreuz et al., 2007).

To achieve these noisy variations in spike times, the driving force which makes the differential equations stochastic is known as Brownian motion, which is ultimately integrated to find the value of the cell potential (originally the position of a pollen grain) at a given time-step.

2.6.1 Brownian Motion and the Wiener Process

Brownian motion was first recorded by the botanist Robert Brown and later described mathematically by Albert Einstein. It is the physical process which characterizes the erratic, seemingly random motion of (originally) a pollen grain on the surface of water, due to its continual bombardment by water molecules. The resulting motion can be viewed as a scaled random walk over a finite time interval and is *almost surely*[4] continuous.

The Wiener process is the mathematical description of Brownian motion. The Wiener process over the period $[0, T]$ is a random variable $W(t)$ that depends continuously on $t \in [0, T]$ and has the following definition.

1. $W(0) = 0$ with probability 1
2. $W(t) - W(s) \sim \mathcal{N}(0, t - s)$ For $0 \leq s < t \leq T$
3. $\langle (W(t) - W(s))\,(W(v) - W(u)) \rangle = 0$ For $0 \leq s < t < u < v \leq T$

Here $\sim \mathcal{N}(0, t - s)$ means Normally distributed with 0 mean and variance $(t - s)$. Equivalently, this may be expressed as $\sim \sqrt{t - s}\mathcal{N}(0, 1)$ where $\mathcal{N}(0, 1)$ is the standard Gaussian distribution with zero mean and unit variance:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{2.37}$$

To incorporate this model of Gaussian white noise into the ordinary differential equation describing the cell membrane potential, we first make the common simplification of taking the basic LIF neuron model (Masquelier et al., 2009; Goodman and Brette, 2008).

$$\tau_m \frac{dV_i(t)}{dt} = E - V_i(t) + RI_i(t) \tag{2.38}$$

[4]In probability theory, 'almost surely' (a.s.) describes an event which happens with probability one. In some complex cases relating to different kinds of infinity there is a subtle distinction between 'almost surely' and 'surely', however in this context they may be regarded as synonymous.

This is equivalent to the form given in Equation 2.7 with a non-zero reversal potential E and also corresponds to Equation 2.16 after dividing by the cell membrane conductance g_0.

Adding the noise process described above we have:

$$\tau_m \frac{dV_i(t)}{dt} = E - V_i(t) + RI_i(t) + \sigma \cdot \xi(t) \cdot \sqrt{\tau_m} \tag{2.39}$$

Here, $\xi(t)$ is a Wiener (Gaussian) variable (where $\xi(t)$ represents $\frac{dW}{dt}$) satisfying the definition of the Wiener process such that $\langle \xi \rangle = 0$ and $\langle \xi(t)\xi(s) \rangle = \delta(t-s)$, where $\delta(\cdot)$ is the Dirac delta function and σ tunes the amplitude of the noise (the standard deviation of the noise in units of Volts) since ξ has unit variance. The noise term, ξ, is importantly scaled by (the square root) of the time constant, τ_m, which means that the amplitude of the noise is scaled up or down as the system speeds up (short τ_m) or slows down (long τ_m) respectively. The dimension of the ξ term is $time^{-1}$ and so ξ is scaled by $\sqrt{\tau_m}$ to make the equation dimensionally consistent.

2.6.2 Discretization of Stochastic differential equations

For solving stochastic differential equations computationally, it is necessary to consider discretized Brownian motion where $W(t)$ is specified at discrete t values. The time-step is set to $\Delta t = {}^{T}/_{N}$ for some positive integer, N and let W_i denote $W(t_i)$ where $t_i = i\Delta t$. From the definition, $W_0 = 0$ and:

$$W_i = W_{i-1} + dW_i, \qquad i = 1, 2, ..., N \tag{2.40}$$

where each dW_i is an independent random variable of the form $\mathcal{N}(0, \Delta t)$ since our time interval is $t - s = \Delta t$.

From the original definition of Brownian motion as the random bombardment from molecules in a liquid, the Wiener process, $W(t)$, is the sum of infinite impulses, with the difference from one moment to the next, dW_i being normally distributed with standard deviation $\sqrt{\Delta t}$. When we come to write the finite difference equation for the stochastic differential equation, this difference is directly added on to the process as a random variable η_i drawn from the standard normal distribution ($\sim \mathcal{N}(0, 1)$). This standard Gaussian variable therefore needs to be scaled by $\sqrt{\Delta t}$ since $dW_i = \sqrt{\Delta t} \cdot \eta_i$. Applying the Euler–Maruyama method, a generalization of the Euler method for stochastic differential equations, (Higham, 2001) we obtain the numerical solution:

$$V_i^{t+1} = V_i^t + \frac{\Delta t}{\tau_m}\left(E - V_i^t + RI_i^t\right) + \sigma\sqrt{\frac{\Delta t}{\tau_m}}\eta_i \tag{2.41}$$

To understand why the noise term is scaled by $\sqrt{\Delta t}$, let us consider the variance of noise with amplitude σ and mean, $\mu_W = 0$ over a fixed time $T = N\Delta t$.

$$\begin{aligned} Var(W) &= E\left[(\sigma - \mu_W)^2\right] \\ &= \frac{\sum_{i=1}^{N}(\sigma_i - \mu_W)^2}{N} \\ &= \frac{N(\sigma - 0)^2}{N} \\ &= \sigma^2 \end{aligned} \tag{2.42}$$

This is the variance of a single driving noise term over the N time-steps. For the full model of Brownian motion, we are interested in the variance of the sum of such noise terms.

The Bieneymé formula states that for independent variables, $Var\left(\sum_i X_i\right) = \sum_i Var(X_i)$. The variance of the sum of noise terms over N time-steps is therefore:

$$\begin{aligned} Var\left(\sum_{j=1}^{N}\sigma_j\right) &= \sum_{j=1}^{N} Var(\sigma_j) \\ &= \sum_{j=1}^{N}\sigma_j^2 \\ &= N\sigma^2 \\ &= \frac{T\sigma^2}{\Delta t} \end{aligned} \tag{2.43}$$

From this we note that the variance increases with the length of the time period, T, as it should. However, the Brownian motion is also undesirably dependent upon N and so will change as we refine the time-step of numerical integration. In order to avoid this artefact of discretization, we need to scale the variance of the noise, $\sigma^2 \propto \Delta t$. Substituting $\sqrt{\Delta t} \cdot \sigma$ for σ,

$$\begin{aligned} Var(W) &= E\left[(\sqrt{\Delta t}\cdot\sigma - \mu_W)^2\right] \\ &= \frac{\sum_{i=1}^{N}(\sqrt{\Delta t}\cdot\sigma_i - \mu_W)^2}{N} \\ &= \frac{N(\sqrt{\Delta t}\cdot\sigma - 0)^2}{N} \\ &= \Delta t\sigma^2 \end{aligned} \tag{2.44}$$

Now calculating the variance of the sum of noise terms:

$$\begin{aligned} Var\left(\sum_{j=1}^{N}\sqrt{\Delta t}\cdot\sigma_j\right) &= \sum_{j=1}^{N} Var(\sqrt{\Delta t}\cdot\sigma_j) \\ &= \sum_{j=1}^{N}\Delta t\cdot\sigma_j^2 \\ &= N\cdot\Delta t\cdot\sigma^2 \\ &= T\sigma^2 \end{aligned} \tag{2.45}$$

The discrete approximation now fulfils the second condition of the Wiener process (with $T = t - s$) meaning that the difference in the Wiener process between one non-overlapping period, $W(t)$ and another $W(s)$ will be normally distributed with variance T (standard deviation $\sqrt{T}$) when the σ scaling is unity.

2.6.3 Pseudo-random Number Generator

To provide a source of (pseudo)-randomness for the noise in the cell membrane potential (in addition to, for example, shuffling procedures for randomizing stimuli and transform presentations), a set of pseudo-random number generator routines were included in the model. The GNU Scientific Library (GSL) routines (Galassi et al., 2009) were selected for this purpose as they are thread-safe (meaning they can be used throughout the parallel regions of code with consistent results each time) and are freely distributed under the GNU General Public License. The particular generator used by these routines was set to be the default 'Mersenne Twister' algorithm `mt19937` with the seed being set at run-time automatically according to the system clock or passed as an argument for repeated simulations.

2.7 Model Parameters

In each chapter of research, parameters may vary according to the simulation paradigm and model features being explored. In such cases, the differences are presented in the appropriate methods sections and the remaining parameters are the default parameters used throughout this thesis, presented here in Tables 2.1–2.3. The cellular parameters presented in Table 2.2 are taken from Troyer et al. (1998) and were originally derived from neurophysiological recording in the Guinea pig (McCormick et al., 1985). The synaptic parameters are presented in Table 2.3 and are also taken from Troyer et al.

(1998) where appropriate, or Perrinet et al. (2001); Perrinet (2003) where they pertain to the STDP model used. Otherwise, some parameters had to be tuned to the particular simulations to compensate for artefacts such as the effect of a small network.

General Network Parameters	**Symbol**	**Value**
Cue current	I^{ext}	1.0 nA
Cue period {training, testing}	t_{cue}	{100, 250} ms
Numerical Integration Time-step	Δt	0.02 ms
Network Parameters		
No. layers	N_L	2
No. excitatory cells per layer	N_E	512
No. inhibitory cells per layer	N_I	128
No. afferent excit. connections per excit. neuron	S_{EE}	512
No. afferent excit. connections per inhib. neuron	S_{EI}	512
No. afferent inhib. connections per excit. neuron	S_{IE}	128
No. afferent inhib. connections per inhib. neuron	S_{II}	0

Table 2.1: General parameters used throughout the simulations.

2.8 Network Inputs

The stimuli used for much of this work are represented by injecting a small amount of current directly into the cell bodies of a particular set of excitatory input neurons continuously throughout the cue period. This pattern of stimulated neurons is gradually shifted across the input layer representing successive transforms of the stimulus (see Figure 3.1).

The size of a stimulus and the amount of neurons each of its transforms is shifted by allows us to precisely control the degree of overlap between transforms of each stimulus. Spatial continuity is crucial to the functioning of the CT mechanism, whereas the trace mechanism requires temporal continuity to associate successive transforms together, (which can be controlled independently through model time constants). In this way, we may eliminate the operation of one mechanism to study the other in isolation and hence disentangle their contributions to the network's capacity for invariance learning.

During training, the set of stimuli are presented in a random order with all transforms for a given stimulus being presented in succession (typically in a random but continuous direction) before presenting the next stimulus's transforms. Presentation of all stimuli in this manner constitutes one training epoch, with a typical training phase consisting of ten such epochs.

Cellular Parameters	Symbol	Value
Excitatory cell somatic capacitance	C_m^E	500 pF
Inhibitory cell somatic capacitance	C_m^I	214 pF
Excitatory cell somatic leakage conductance	g_0^E	25 nS
Inhibitory cell somatic leakage conductance	g_0^I	18 nS
Excitatory cell membrane time constant	τ_m^E	20 ms
Inhibitory cell membrane time constant	τ_m^I	12 ms
Excitatory cell resting potential	V_0^E	-74 mV
Inhibitory cell resting potential	V_0^I	-82 mV
Excitatory firing threshold potential	Θ^E	-53 mV
Inhibitory firing threshold potential	Θ^I	-53 mV
Excitatory after-spike hyperpolarization potential	V_H^E	-57 mV
Inhibitory after-spike hyperpolarization potential	V_H^I	-58 mV
Excitatory reversal potential	$\hat{V}^E$	0 mV
Inhibitory reversal potential	$\hat{V}^I$	-70 mV
Absolute refractory period	τ_R	2 ms

Table 2.2: Cellular parameters used in the simulations. The leaky integrate-and-fire parameters used by default throughout this thesis were taken from Troyer et al. (1998) (derived originally from McCormick et al. 1985).

Synaptic Parameters	Symbol	Value	
Synaptic neurotransmitter concentration	α_C	0.5	†
Proportion of unblocked NMDA receptors	α_D	0.5	†
Presynaptic STDP time constant	τ_C	[3, 75] ms	†
Postsynaptic STDP time constant	τ_D	[5, 125] ms	†
Synaptic learning rate	ρ	0.1	†
Plastic ($E \to E$) synaptic conductance range, CT	$\lambda \cdot \Delta g^{EE}$	[0, 4] nS	*
Plastic ($E \to E$) synaptic conductance range, Trace	$\lambda \cdot \Delta g^{EE}$	[0, 1.25] nS	*
Change in synaptic conductance ($I \to E$)	$\lambda \cdot \Delta g^{IE}$	[0.5, 2.5] nS	*
Change in synaptic conductance ($E \to I$)	$\lambda \cdot \Delta g^{EI}$	5.0 nS	*
Change in synaptic conductance ($I \to I$)	$\lambda \cdot \Delta g^{II}$	5.0 nS	*
Excitatory–Excitatory synaptic time constant	τ_{EE}	{2, 150} ms	{§, *}
Inhibitory–Excitatory synaptic time constant	τ_{IE}	5 ms	§
Excitatory–Inhibitory synaptic time constant	τ_{EI}	2 ms	§
Inhibitory–Inhibitory synaptic time constant	τ_{II}	5 ms	§

Table 2.3: Synaptic parameters used in the simulations. The synaptic conductance time constants were taken from Troyer et al. (1998) (derived originally from McCormick et al. 1985) as indicated by § except where τ_g^{EfE} was lengthened to model the Trace learning rule. Plasticity parameters (denoted by †) are taken from Perrinet et al. (2001). Parameters marked with * were tuned for the reported simulations.

After completion of training, learning is switched off (prohibiting further synaptic modification) and the network is presented with all transforms of all stimuli in order (resetting the neurons to their resting state between transforms) and the resultant firing in both input and output layers is saved for analysis.

2.9 Analysis

Having abstracted away from some of the biological detail and built a model of the visual system, the next issue lies in understanding the model's behaviour. A significant part of this process is finding an accurate and reliable way to evaluate the extent to which the model has learnt to solve a given problem (namely transformation-invariant object recognition). In order to evaluate the model we must have some understanding of the patterns of activity or 'neural code' with which it communicates (Quian Quiroga and Panzeri, 2009).

Here I will introduce the various techniques of analysis commonly employed throughout this work to understand and quantify the behaviour of the neural networks. Other types of analysis for particular types of simulations are introduced as required in the pertinent chapters.

2.9.1 Spike Rasters and Histograms

One of the most basic ways to visualize spike data is with a raster plot. Spike rasters show the activity of a population of neurons in a raw form so the data may be inspected before applying any kind of analysis.

In spike rasters, time is plotted on the x-axis typically on a scale in the order of milliseconds, while the neurons (often arbitrarily numbered) are stacked on the y-axis. As spikes are binary events, they are simply marked with a bar or dot at the times they occur in a row for each neuron. At a glance, raster plots give a feel for the coordinated activity of the whole population of neurons and are used extensively throughout this thesis.

Another very useful summary of a population's activity is a spike histogram where time is binned on the x-axis and the spikes within each bin are counted across all neurons, (often normalized) and then plotted as a frequency distribution. Smoothing kernels are also sometimes applied to lessen some of the artefacts arising from the discretization of a continuous process.

While spikes are discrete all-or-nothing impulses (with a neuron's *spike train* denoted by the sum of Dirac delta functions; $\sum_i \delta(t - t_i)$), much of the information

available in the outputs of a brain area is contained in the *firing rate* (Tovée et al., 1993). This is computed by counting the number of spikes in a sampling window, that is in the time interval $[t, t + \Delta t]$. As Δt becomes vanishingly small (in the limit of very small sampling windows and infinitely many trials), the probability of a spike occurring is given by $f(t)\delta(t)$ where $f(t)$ (often denoted $r(t)$) is the instantaneous firing rate of the neuron (spikes per unit time). Plotting $f(t)$ as a function of time is a standard way of visualizing network activity known as a post-stimulus time histogram or *PSTH* (Koch, 1999; Dayan and Abbott, 2005) and is frequently presented along with raster plots to provide a visual summary of the network's activity.

Typically however, the nervous system has to make behavioural decisions based upon the spike train it observes (rather than the average of many spike trains from repeated trials). In some cases though, the temporal average (over repeated trials) for a single postsynaptic neuron may be replaced by an ensemble average (over a population of neurons) to approximate $f(t)$.

$$\langle f(t)\rangle_T = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{\Delta T}\int_t^{t+\Delta T}\sum_{i=1}^{n_j}\delta(t' - t_{ij})dt' \tag{2.46}$$

Here, the first sum indexed by j is over the n identical trials (or neurons for ensemble coding) and the second sum indexed by i is over all n_j spikes at time t_{ij} of the j^{th} trial (or neuron). Note that $\int_0^T f(t) = n$ where n is the number of spikes in the interval $[0, T]$ averaged over trials.

Equivalently, the following notation may also be used as in (Dayan and Abbott, 2005):

$$r(t) = \frac{1}{\Delta t}\int_t^{t+\Delta t} d\tau\langle\rho(\tau)\rangle \tag{2.47}$$

Where $\langle\rho(\tau)\rangle$ is the trial-averaged neural response function ($\rho(t) = \sum_{i=1}^{n}\delta(t - t_i)$) used to re-express sums over spikes as integrals over time.

The instantaneous firing rate, $f(t)$ cannot be determined from finite data sets however and there is no unique way to approximate it. One simple approach is to divide time into discrete bins of duration ΔT and divide the spike count within these bins by the time duration such that the instantaneous firing rate is approximated by the spike-count firing rate over the duration of the bin. Mathematically, the limit $\Delta T \to 0$ should be taken but practically the bin size must be large enough such that there are enough spikes within the interval to obtain a reliable estimate of the instantaneous firing rate. The choice of bin size is therefore an important decision, since decreasing ΔT will increase the temporal resolution by yielding estimates at

narrower intervals of time but at the cost of lessening the ability to distinguish between different firing rates. Objective methods have been devised to choose an optimal bin size from the spike count statistics alone based upon minimizing the mean integrated squared error between the estimated rate and the unknown underlying rate (Shimazaki and Shinomoto, 2007).

Estimates of the firing rate produced in this way will always be an integer multiple of $\frac{1}{\Delta T}$ (since spike counts will be integer values). This can be avoided by varying the bin size such that there are a fixed number of spikes within each bin. The spike counts are similarly divided by the variable bin width to approximate the firing rate.

The calculated firing rate estimates may also depend upon the placement of the discrete time bins. To remedy this artefact, a single bin (time window of ΔT) may be slid along the spike train providing a spike count within the window at each location from which to estimate the firing rate. It is often desirable to smooth the spike trains with such a convolution kernel in this way as it avoids the arbitrariness of bin placement and appears to provide a better temporal resolution, however, rates calculated from overlapping window locations will be correlated as they will be based upon some of the same spikes. The simplest kernel is the square window which relates to the approximate firing rate as follows:

$$\tilde{f}(t) = \sum_{i=1}^{n} w(t - t_i) \tag{2.48}$$

where the window function is defined as

$$w(t) = \begin{cases} \frac{1}{\Delta T} & \text{if} -\Delta T/2 \leqslant t < \Delta T/2 \\ 0 & \text{otherwise} \end{cases} \tag{2.49}$$

The rectangular window produces a jagged curve for the firing rate estimate because of its discontinuous shape. However, virtually any kernel filter may be used provided that it goes to 0 outside a region near $\tau = 0$ and that its time integral is equal to 1 (Dayan and Abbott, 2005). To this end, a commonly used alternative to the rectangular time window of Equation 2.46 is the Gaussian-window smoothed function with standard deviation σ:

$$w(\tau) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\tau^2}{2\sigma^2}} \tag{2.50}$$

Thus the firing rate estimate becomes:

$$\langle f(t) \rangle_T = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(t-t')^2}{2\sigma^2}} \sum_{i=1}^{n_j} \delta(t' - t_{ij}) dt' \tag{2.51}$$

The Gaussian kernel introduces another hyper-parameter into the analysis, namely the standard deviation, σ, (i.e. the size of the region either side of τ used to calculate the firing rate estimate), which will also affect the resultant estimates. Based upon a similar procedure to the Mean Integrated Squared Error (MISE) method for choosing an optimal bin size (Shimazaki and Shinomoto, 2007), there is an algorithm for choosing the optimal kernel bandwidth (σ) which is then extended to allow for a variable kernel bandwidth in conformity with the data (Shimazaki and Shinomoto, 2010).

However, applying kernels to smooth the data in this way is not adopted as a practise for the probability distributions estimated for the information analysis, as it would introduce correlations between neighbouring bins, thus violating the 'independence' condition (Treves and Panzeri, 1995).

2.9.2 Fourier and Power Spectra

The data visualization techniques so far discussed are straightforward ways of displaying time series data. A Fourier spectrum instead takes the same information and represents it in the *frequency* domain rather than the *time* domain by decomposing the signal into its constituent frequencies. In so doing, a Fourier spectrum makes explicit the energy carried by different frequencies in a time-varying signal. This gives them a wide variety of applications in science and engineering, for example in filtering out noise and interference at particular frequencies, but also for characterizing the main frequencies of oscillation in a population of neurons.

The Fourier spectrum is plotted from taking a Fourier transform of a signal and calculating the amplitude (magnitude) of its frequency components (note that plotting instead the square of the magnitude is known as the power spectrum). The principle underlying a Fourier transform is to express a function (data set) in terms of the sum of its projections onto a set of basis functions. In particular, a Fourier transform decomposes a function (signal) into an infinite sum of sines and cosines. The Fourier transform $F(u)$ of $f(x)$ may be taken as in Equation 2.52 where u is the frequency (provided $f(x)$ is continuous and integrable).

$$F(u) = \int_{\infty}^{-\infty} f(x) e^{-i2\pi ux} dx \tag{2.52}$$

This can be rewritten using Euler's formula (Equation 2.53) to make explicit the relationship with sine and cosine functions as in Equation 2.54.

$$e^{i\theta} = \cos\theta + i\sin\theta \tag{2.53}$$

$$F(u) = \int_{\infty}^{-\infty} f(x) \left(\cos(2\pi ux) - i \sin(2\pi ux)\right) dx \tag{2.54}$$

Note, even if $f(x)$ is a real function, $F(u)$ is complex since the exponent (sine term) of the Fourier transform in Equation 2.52 (Equation 2.54) contains the imaginary number, $i := \sqrt{-1}$.

Since spikes are binary events, the groups of them must be binned as described above to provide an appropriate signal for Fourier analysis. As time is discretized in the binning process, the function (signal) may not be integrated analytically, which necessitates the use of the Discrete Time Fourier Transform (DFT).

The DFT approximates the area under the curve (integral) with, for example, the trapezoid rule. In so doing, the computational cost of calculating the DFT on N data points scales by $\mathcal{O}(N^2)$ making it generally an expensive operation. For this reason, another approximation of the full integral, the Fast Fourier Transform (FFT), is used instead as its computational cost scales as $\mathcal{O}\left(N \log_2(N)\right)$, provided that the size of the set of data is a power of two.

The FFT algorithm used in this work is from the Matlab library function `fft(x)` which is based upon the Cooley–Tukey algorithm. It implements the transform for a vector x of length N as follows.

$$X(k) = \sum_{n=1}^{N} x_n e^{-2\pi i (k-1)\frac{n-1}{N}} \qquad k = 1, ..., N \tag{2.55}$$

2.9.3 Weight matrices

Given that a central theme of this thesis concerns how spiking neural networks self-organize through learning, synaptic weight matrices are often plotted before and after learning to show the change in structure. Input cells are plotted on the x-axis (conventionally indexed by j) and output (postsynaptic) neurons are plotted on the y-axis (conventionally indexed by i) with the synaptic weights between them plotted as temperature maps. Hot colours (that is to say red hues) represent large weight values and thus indicate how excited the output neurons will be for a given set of input neurons.

2.9.4 Information Analysis

In order to determine how well a neuron has learnt to respond to a particular object, we wish to know how much information is conveyed through the cell's response r about

the particular stimulus s which is being observed. To be informative in the context of transformation-invariance, the responses of a given neuron should be both specific to a particular stimulus and yet generalize across some or all of its transforms. With an ensemble of output neurons, all of the transforms for each stimulus should evoke a response such that the identity of the stimulus may be reliably decoded from the responses of the populations of neurons. In this respect, different transforms of the same stimulus are considered as repeated presentations of that stimulus, perturbed by noise (or other uninformative differences) which should be ignored. Using the techniques of Shannon's Information Theory, we may quantify how much information the output neurons of the network convey before and after learning. For a general introduction to Information Theory see MacKay (2003), Bishop (1997) or Rolls and Treves (1998).

Neurons are inherently noisy and as such (to a particular degree of precision, or 'binning') the same response may be evoked by two different stimuli. This noise will also lead to trial-by-trial variability, which (for informative neurons) is likely to yield different distributions of responses to different stimuli (even though there may be considerable overlap in these response distributions). It is therefore necessary to express the stimulus-response relationship probabilistically and build a table of probabilities $P(s, r)$ from which the information content may be calculated.

2.9.4.1 Stimulus-specific single-cell Information measure

From Shannon's definition, the total amount of information (entropy) in a set of responses (spike trains) is given by the following formula.

$$H(R) = -\sum_{r \in R} P(r) \log_2 P(r) \tag{2.56}$$

The entropy of the responses in itself is not particularly useful — instead we wish to calculate the information about *particular* stimuli conveyed through the neuron's responses. By applying the Sum and Product rules of probability, we may obtain an expression for the quantity of information conveyed about each stimulus.

$$P(x) = \sum_{y} P(x, y) \tag{2.57}$$

$$= \sum_{y} P(x|y) \cdot P(y) \tag{2.58}$$

Thus, the information is separated into the components contributed by each stimulus by first substituting in the expression $P(r) = \sum_s P(r|s)P(s)$ into the first term of Equation 2.56 and rearranging (since $P(s)$ is not dependent upon the sum over r).

$$\begin{aligned} H(R) &= -\sum_{r\in R}\sum_{s\in S} P(r|s)P(s)\log_2 P(r) \\ &= -\sum_{s\in S} P(s)\sum_{r\in R} P(r|s)\log_2 P(r) \end{aligned} \tag{2.59}$$

This reformulation allows us to obtain an expression for the total entropy (information) associated with each particular stimulus.

$$H(s,R) = -\sum_{r\in R} P(r|s)\log_2 P(r) \tag{2.60}$$

Not all of this information will be about the stimulus, for example, natural variation in the spike rates due to noise or other unrelated processes in the brain will contribute to the total entropy. When the stimulus is held fixed, any variability will be due to these unrelated factors which should therefore be subtracted from the total. The entropy arising from variations in the spike train for a given stimulus is as follows.

$$H(R|s) = -\sum_{r\in R} P(r|s)\log_2 P(r|s) \tag{2.61}$$

By subtracting this quantity from the total entropy for a given stimulus (Equation 2.60) we have an expression for the mutual information between a stimulus and the set of responses (the net stimulus information).

$$I(s,R) = -\sum_{r\in R} P(r|s)\left[\log_2 P(r) - \log_2 P(r|s)\right] \tag{2.62}$$

This is the first information measure used for performance analysis and represents the difference between the *a priori* uncertainty (entropy) and the residual uncertainty remaining once r is known. This stimulus specific information measure attains its maximum value $I(s) = -\log_2 P(s)$ (i.e. $I(s) = \log_2 |S|$, assuming a uniform probability distribution across all stimuli is chosen[5]) only if the relationship between the stimulus and response (conditional probabilities) becomes completely deterministic; $P(s|r) = 1$ for $s = s(r)$, and $P(s|r) = 0$ otherwise (Rolls et al., 1997b). The stimulus specific

[5]In general (for uniform or non-uniform stimulus probability distributions), the maximum information which may be transmitted is the entropy of the stimuli: $H(S) = -\sum_{s\in S} P(s)\log_2 P(s)$

information measure (Equation 2.63) is here rearranged into a more conventional form (Rolls and Milward, 2000; Elliffe et al., 2002).

$$I(s, R) = \sum_{r \in R} P(r|s) \log_2 \frac{P(r|s)}{P(r)} \tag{2.63}$$

2.9.4.2 Stimulus-specific single-cell Information measure algorithm

Having defined the desired performance measure, the procedure to compute it follows closely from previous work (Rolls and Milward, 2000; Elliffe et al., 2002). During testing, each transform of each stimulus is presented to the input layer of the network. Each neuron is reset (allowed to settle) after presentation of each transform such that the activity due to one transform does not affect the responses to later transforms.

To build the necessary probability distributions (or estimates thereof), it is first necessary to regularize the neural firing rate data in some way (typically by binning) since sampling a finite set of data from a continuous variable (firing rate) will likely result in a set of unique responses meaning that each response uniquely identifies its stimulus ($P_N(s|r) \in \{0, 1\}$, where P_N denotes an estimate of the true probability from a finite set of N data). Without regularizing the data first, we therefore obtain a measure of the entropy of the stimulus set only, $H(S)$, rather than the transmitted information (i.e. the stimulus-specific information of Equation 2.63).

Accordingly, after testing, the spike responses of each output neuron to each transform of each stimulus are placed into equispaced bins and the corresponding firing rate for each cell is calculated by dividing by the bin width. From such finite samples of firing rates, frequency distributions for each cell's firing rates (to each transform of each stimulus) are constructed, from which unconditional probability distributions are calculated through normalization, $P(r_i) = {}^{f_i}/_{(\sum_{s \in S} |T_s|)}$ (where $|T_s|$ is the size of the set of transforms for stimulus s and i indexes the bin). This procedure yields an estimate of $P(r)$ for each cell. The firing rates for each cell over all transforms of a *given stimulus* are also binned to construct conditional probability distributions for each cell to each stimulus, $P(r|s_i)$.

Based upon these firing rate probability distributions, the stimulus-specific single-cell information $I(s, R)$ was calculated according to Equation 2.63 for each cell over each stimulus in turn. This yields the quantity of information in a set of responses R of a single cell about a specific stimulus s.

Good performance for a cell would entail specificity, meaning a large response to one or a few stimuli regardless of their position (transform) and small responses

to other stimuli, as the most informative cells respond to one or very few stimuli preferentially. Rather than the average amount of information a cell conveys about the whole set S of stimuli (known as the mutual information), we therefore compute instead the maximum amount of information a neuron conveys about *any* of the stimuli, $\max_{s \in S} [I(s, R)]$ and plot this for each output cell.

2.9.4.3 Multiple-cell Information measure

If all the output cells learnt to respond to the same stimulus then there would be no discriminability and the information about the set of stimuli S would be poor. Conversely, a highly distributed neural representation where the same cells responded differently to each transform would require a very large ensemble of neurons in order for downstream neurons to generalize across transforms. Recordings from such ensembles in the macaque temporal lobe visual cortex reveal a sparsely distributed (but not local) encoding scheme with analysis revealing a very large representational capacity over a reasonably small cell population (Rolls et al., 1997a). Furthermore, the brain typically has to make decisions on single events by evaluating the activity of large populations of neurons in a limited time (Quian Quiroga and Panzeri, 2009) which individually are noisy and thus variable in their discharge patterns from trial to trial, emphasizing the need to extract information from an array of such neurons.

As such, a multiple cell information measure (Equation 2.65) is used to calculate the information about the set of stimuli from a population of up to $5 \cdot |S|$ output neurons. This population consisted of the subset of up to five cells per stimulus which conveyed, as single cells (according to the stimulus-specific single-cell information measure of Equation 2.63), the most information about their respective stimuli. The cells were randomly drawn from these most informative individual neurons as the ensemble size was systematically varied from 1 to $5N_S$, (where $N_S = |S|$, the cardinality of the set of stimuli). If the cells have become tuned to different stimuli (or subsets of stimuli) then this measure should increase with the number of different cells used up to the theoretical maximum of the entropy of the set of stimuli, $H(S)$. If the stimulus probability distribution is uniform, the theoretical maximum is simply $\log_2 N_S \, bits$, attained only if those cells have become tuned to different stimuli such that the stimulus identities of all transforms may be correctly decoded based upon their firing rates.

Ideally, we would like to calculate the mutual information, $I(\mathbf{S}; \mathbf{R}) = H(\mathbf{R}) - H(\mathbf{R}|\mathbf{S})$, (or equivalently $I(\mathbf{S}; \mathbf{R}) = H(\mathbf{S}) - H(\mathbf{S}|\mathbf{R})$; see Equation 2.64), which is the average amount of information about the observed stimulus, calculated from the

responses of all cells after a single presentation of a stimulus. Here, $\mathbf{R}$ is the response matrix, consisting of response vectors composed of the responses from each cell.

$$I(\mathbf{S};\mathbf{R}) = \sum_{s \in S} P(s) I(s, \mathbf{R}) \tag{2.64}$$

That is to say, the mutual information between the whole set of stimuli S and of responses $\mathbf{R}$ is the average across stimuli of this stimulus-specific information. However, the response vector is multidimensional and the minimum number of trials required to adequately sample the response space grows exponentially with the dimensionality of the response space (Quian Quiroga and Panzeri, 2009). As such, the information cannot be measured directly from the sampled probability distributions $P(\mathbf{r}, s)$ (the relationship between a stimulus, s, and the response vector of the firing of a set of neurons to it) because the dimensionality of the response vectors is too large to be adequately sampled. Instead, the data must be regularized (e.g. binned) and a decoding procedure must be used e.g. maximum likelihood, Bayesian or dot-product decoding, to estimate the stimulus (s') that gave rise to each trial's response vector, $\mathbf{r}$ and hence estimate the probability of it being each stimulus. This auxiliary probability table of stimuli, s decoded stimuli, s' (rather than high dimensional firing rate responses) spans a limited set of only $|S|$ values, (rather than the large number of values of the multi-dimensional rate responses) thus ameliorating the curse of high-dimensionality. From these probabilities, the mutual information between the set of actual stimuli S and the decoded estimates S' is calculated according to Equation 2.65.

$$I(s, s') = \sum_{s,s' \in S} P(s, s') \log_2 \frac{P(s, s')}{P(s)P(s')} \tag{2.65}$$

2.9.4.4 Multiple-cell Information measure algorithm

As was necessary for computing the stimulus-specific single-cell information, the data (neural firing rates) are binned to yield finite-sample estimates of the pertinent underlying probability distribution ($P_N(r_i|s)$ and $P_N(r_i)$, where the $_N$ signifies an approximation to the true probability distributions based upon a sample of N data). Aside from pure discretization, other forms of regularization have been investigated to prevent over-fitting, including convolution with a continuous distribution (kernel) with and without discretization of the response space and neural network fitting of the conditional probabilities (Panzeri and Treves, 1996) but do not present a good reason to adopt them over the simplicity of binning.

2.9.4.5 Decoding and Cross-validation algorithm

To prevent over-fitting, a cross-validation technique is used in the decoding algorithm. A jack-knife procedure is used, where the data from one trial (transform) are reserved for 'testing' while all others are used for 'training' (as the basis for comparison in the decoding procedure). In earlier work, this was found to be the more effective than a conventional cross-validation procedure (with which different proportional splits between test and training data were investigated), as the decoding algorithm's predictions are better, the more data it can base them on (Rolls et al., 1997a).

For each transform of each stimulus in turn, the firing rate of the C cells forming the ensemble to the stimulus is designated the 'test' data and the means and standard deviations of responses are calculated for each stimulus separately (without the 'test' data) to form the distribution of responses used as the 'training' data. From these regularized neural firing rates, a pseudo-simultaneous population response vector, $\mathbf{r}$, is constructed of length C for each stimulus. Each element, r_c^s is the firing rate to the 'test' transform of a particular stimulus for one of the C cells considered in an ensemble. This vector is then compared to the 'training' data vectors (the mean response vector to each stimulus calculated without the 'test' responses) for those same randomly selected C cells. A decoding procedure (detailed below) is used to calculate the probabilities that the response vector of 'test' data under consideration was elicited by each stimulus in turn, $P(s'|\mathbf{r})$.

A decoding procedure is used to classify the inputs by estimating the stimulus s' that gave rise to the particular firing rate response vector on each trial based upon the similarity to the training data response distribution for each stimulus (Rolls et al., 1997a). From this estimate, a probability table (confusion matrix) of $P(s, s')$ is then constructed of the real stimuli s and the decoded stimuli s' (whereby a perfectly trained network would have probabilities of $1/N_S$ on the diagonal and 0 elsewhere). For each neuron, a probability distribution of $P(s')$ is also constructed from the frequencies with which each stimulus is classified. Since $P(s)$ is chosen *a priori*, typically with each stimulus being presented with equal probability $1/N_S$, these procedures together provide estimates for all relevant probability distributions, from which the mutual information is calculated.

The decoding procedure used for calculating the multiple cell information measure in this work is known as 'Bayesian decoding' which is designed to extract as much information as is possible for any decoding procedure (thereby giving an upper limit on the information available to subsequent neurons) by reconstructing the Bayesian *posterior* conditional probabilities, $P(s'|\mathbf{r})$ from the data (Földiák, 1993).

The probability distribution for the stimuli, $P(s)$ is set by the experimenter (known *a priori*, hence this is the *prior*) and the training data of mean and standard deviation of responses to each stimulus allow us to estimate the likelihood, $P(r|s)$. By Bayes' theorem, we obtain an expression for posterior probability as shown in Equation 2.66. Since $P(r)$ is independent of the stimuli, it plays the role of a normalization factor, ensuring that the posterior probabilities sum to unity (Bishop, 1997) which may be implemented by normalizing such that $\sum_{s'} P(s'|\mathbf{r}) = 1$.

$$\begin{aligned} P(s|r) &= \frac{P(r|s)P(s)}{P(r)} \\ &= \frac{P(r|s)P(s)}{\sum_s P(r|s)P(s)} \end{aligned} \tag{2.66}$$

First the responses of each cell in the ensemble are fitted to a Gaussian distribution (parameterized by the mean and standard deviation of the training data for each stimulus), whose amplitude at r_c yields the probability $P(r_c|s')$ of each cell's firing rate given that it belongs to the firing rate distribution of that stimulus. These Gaussian distributions are truncated at zero, meaning that when $r_c = 0$, the best estimate of $P(r_c|s')$ is the fraction of that cell's training trials yielding zero firing (Rolls et al., 1997a).

Since there is typically more than one cell in the ensemble, the posterior probability distribution must be calculated based upon multiple responses, $P(s|r_1, r_2, ..., r_C) = \frac{P(r_1,r_2,...,r_C|s)P(s)}{P(r_1,r_2,...,r_C)}$. Assuming that the variability of neural responses are statistically independent (independent and identically distributed), $P(r_1, r_2, ..., r_C|s) = P_1(r_1|s)P_2(r_2|s)...P_C(r_C|s)$, the joint probability $P(\mathbf{r}, s')$ may then be estimated as shown in Equation 2.67 (where r_c is the firing rate of cell c in the ensemble) from these conditional probabilities.

$$P(\mathbf{r}, s') = P(s') \prod_{c=1}^{C} P(r_c|s') \tag{2.67}$$

Normalizing this joint probability distribution as described then yields an estimate of the desired posterior probability, $P(s'|\mathbf{r})$ as given in Equation 2.68. This is repeated for each stimulus in the set to form a posterior probability distribution.

$$P(s'|\mathbf{r}) = \frac{P(s') \prod_c P(r_c|s')}{\sum_{s'} P(s') \prod_c P(r_c|s')} \tag{2.68}$$

Having obtained a posterior probability distribution over the likely (decoded) stimuli given the ensemble's response to a particular transform, the confusion matrix

of $P(s, s')$ may be updated in one of several ways (Quian Quiroga and Panzeri, 2009). One such decoding method is the maximum *a posteriori* (MAP) approach, whereby the responses are decoded as the stimulus which gives the maximum probability across the distribution of decoded stimulus probabilities i.e. $s^{MAP} = \arg\max_{s'} (P(s'|\mathbf{r}))$ or drawing an inference (Bishop, 1997). If a quantifier of the information in this response is still desired, 'one' is added to the frequency tally in the corresponding element of the confusion matrix and normalized after all decoding procedures to yield probabilities, $P(s, s')$. Alternatively, the Bayesian decoding method preserves more information by accumulating the probabilities of each decoded stimulus against each actual stimulus (and normalizing to give the correct marginal distributions as before). From these probabilities obtained with either method, the mutual information may then be calculated as described in Equation 2.65.

These procedures may be thought of as complimentary since Information theory quantifies the total information available in the neural responses by considering the whole distribution of posterior probabilities, both likely and *unlikely* stimuli (not all of which may be available to the organism) while the decoding procedure simply predicts the most likely stimulus (effects a decision) and may therefore be more usefully and directly related to the final behavioural outcome (Quian Quiroga and Panzeri, 2009). Given that this thesis models the ventral visual stream in isolation (without any behavioural output), the information analysis is most useful and it provides a single number summary of the population responses and provides an upper bound on the amount of knowledge which could be available to the organism.

2.9.4.6 Correction procedure

In estimating the transmitted information based upon probability distributions with a small number of trials, a systematic error is introduced which tends to bias the information upwards, that is, on average lead to overestimates of the information content of the responses (Treves and Panzeri, 1995). In essence, this is the *limited sampling problem* whereby the true probability distributions, for example $P(r)$ must be estimated from distributions constructed with a limited amount of data, $P_N(r)$ (where $\lim_{N\to\infty} P_N(r) = P(r)$). While this is not a major shortcoming, since the use of optimal decoders means that we are calculating an upper limit on the information content anyway, it behoves us to find a tight upper bound by correcting for such systematic biases where possible.

Intuitively, estimates of both the 'response entropy', $H(\mathbf{R})$ and the 'noise entropy', $H(\mathbf{R}|\mathbf{S})$ are increased with the number of trials, since entropy is a measure of variability

Algorithm 1 Multiple cell information algorithm

Require: $N_O = \sum_s N_T(s)$ ▷ N_O: Total number of observations (transforms) across all stimuli
function MULTINFO(Spike trains)
for $C \leftarrow 1 : C_{max}$ **do** ▷ Loop over ensemble size
$N_C \leftarrow 100 \cdot (C_{max} - C + 1)$ ▷ N_C: No. of iterations adjusted for ensemble size
for $i \leftarrow 1 : N_C$ **do** ▷ Loop over iterations to smooth information
$\mathbf{e} \leftarrow permute(C_{max})(1 : C)$ ▷ Shuffle the C_{max} cells and select the first C
for $s \leftarrow 1 : N_S$ **do** ▷ Loop over the N_S stimuli
for $tt \leftarrow 1 : T_s$ **do** ▷ Loop over the T_s 'test' transforms of stimulus s
$\mathbf{r} \leftarrow \mathbf{R}(\mathbf{e}, s, tt)$
$\overline{\mathbf{R}}^{train} \leftarrow mean_t[\mathbf{R}(\mathbf{e}, s, t \neq tt)]$
$\mathbf{D}^{train} \leftarrow std_t[\mathbf{R}(\mathbf{e}, s, t \neq tt)]$
$P(s') \leftarrow decode[P(s), \mathbf{r}, \overline{\mathbf{R}}^{train}, \mathbf{D}^{train}]$
$d \leftarrow \arg\max_{s'} P(s')$
$P(s, d) \leftarrow \frac{P(s')}{N_O}$
$P_{MAP}(d) \leftarrow P_{MAP}(d) + 1$
end for
end for
$I \leftarrow \sum_{s,s'} P(s, s') \log_2 \frac{P(s,s')}{P(s)P(s')}$
$\mathbf{I}(C) \leftarrow {}^{I}/_{N_i} - C_1$ ▷ Smooth by N_C and subtract correction term
$I_{MAP} \leftarrow \sum_{s,s'} P_{MAP}(s, s') \log_2 \frac{P_{MAP}(s,s')}{P(s)P(s')}$
$\mathbf{I}_{MAP}(C) \leftarrow {}^{I_{MAP}}/_{N_C} - C_1$ ▷ Smooth by N_C and subtract correction term
end for
end for
return $\mathbf{I}, \mathbf{I}_{MAP}$
end function

and more limited sampling is less likely to capture the full extent of the variability. Since $H(\mathbf{R})$ depends upon all the data, it will therefore be less biased (that is, grow closer to its true value) than $H(\mathbf{R}|\mathbf{S})$, which is calculated from a more restricted set of data. Given that the mutual information (the quantity we would ideally wish to calculate) is $I(\mathbf{S}; \mathbf{R}) = H(\mathbf{R}) - H(\mathbf{R}|\mathbf{S})$ and that the *bias* of the response entropy is far less than the *bias* of the noise entropy (i.e. the second term, *subtrahend* is a smaller proportion of its true value than the first term, *minuend* is of its true value), the net result is typically strong upwards bias in the mutual information (Panzeri et al., 2007).

Part of the problem comes from the need to regularize the data, as a finite number

of responses from a continuous output variable will almost certainly all be different from one another and so will uniquely (but artificially) identify the stimulus (Panzeri and Treves, 1996). The subsequent choice of how to discretize the response space by binning involves a compromise. Choosing too few bins may not adequately differentiate the distributions of responses to each stimulus and result in a large loss of information. However, choosing too large a number would not control finite size sampling distortion adequately and may leave many bins unoccupied. If kernel methods are then used to smooth the data in order to fill these bins, the condition of independence is violated (Treves and Panzeri, 1995).

As a simple heuristic, the choice of the number of bins used in the information analysis routines in this work was set to the number of transforms per stimulus and bins are equispaced over the range of responses.

Having binned (regularized) the data, a correction procedure is then used as originally formulated by Treves and Panzeri (1995). Remarkably, the first correction term depends only upon the sizes N_S and B of the stimulus and response sets, but is invariant with respect to the probability distributions of stimuli and responses (Treves and Panzeri, 1995). A full expansion may be found in Panzeri and Treves (1996) but all other terms may be neglected since the m^{th} correction term, C_m is proportional to N^{-m}, so only the first term is given here in Equation 2.69 which is the only term commonly used (Sugase et al., 1999).

$$C_1 = \frac{1}{2N \ln 2} \left\{ \sum_s \tilde{B}_s - \tilde{B} - (N_S - 1) \right\} \tag{2.69}$$

Here, N is the total number of trials performed, $\tilde{B}_s$ is the number of 'relevant' response bins for the trials with stimulus s (i.e. bins with non-zero occupancy probability: $P(i|s) > 0$, where i denotes a response 'bin') and similarly $\tilde{B}$ denotes the number of response bins where $P(i) > 0$ across all stimuli, although a Bayesian technique to estimate $\tilde{B}_s$ was also tried (Panzeri et al., 2007). In the case where all bins have a non-zero probability of being occupied for every stimulus, the original result is recovered from Treves and Panzeri (1995):

$$C_1 = \frac{1}{2N \ln 2} \{(N_S - 1)(B - 1)\} \tag{2.70}$$

Since subtracting the correction term from the calculated information removes the upward bias *on average*, there are times when it will subtract too much or too little — even adding to the total and increasing it beyond the theoretical limit ($H(s)$, which given an even stimulus presentation probability is simply $\log_2(N_S)$) or making the

final value negative. In such cases the, final information values were clipped at 0 and $\log_2(N_S)$ to remove such artefacts.

2.9.4.7 Averaging and smoothing

In selecting a particular ensemble of neurons to be decoded, C neurons were drawn randomly (without replacement) from the total pool of $C_{max} = 5N_S$. To average out the statistical effects of this random draw procedure and generate smoother values of information, the procedures described are repeated for a large number of iterations for each ensemble size (response vector dimensionality). As the number of cells in the ensemble was varied from 1 to C_{max}, the procedure was repeated over, N_C iterations, for each ensemble size, (C), where $N_C = 100 \cdot (C_{max} - C + 1)$ and the resulting information values obtained were weighted by the appropriate number of iterations. Pseudocode of the full multiple stimulus information algorithm is given in Algorithm 1.

2.10 Simulation Strategy

Considering how the rapidly the computational expense of LIF simulations escalates as the network grows in size, the simulation strategy requires careful consideration. According to Moore's law, which states that the number of transistors which may be built into an integrated circuit doubles every two years[6] for a fixed cost, the speed of computers should grow exponentially. Until fairly recently (approximately the past 5–10 years) this was evident in the increasing clock speeds of successive generations of CPUs. The rate of increase in clock speeds has greatly slowed or even plateaued lately though (typically in the region of 3–4 GHz), since engineers are approaching the thermodynamic limits of the current semiconductor processes and technologies known as 'The Power Wall'. Essentially, small increases in clock speeds (achieved by cramming more transistors into the same space) lead to increasingly large heat outputs which may no longer be adequately cooled through inexpensive traditional techniques (e.g. copper or aluminium air-cooled heat-sinks). Vendors of supercomputers may continue to overcome this hurdle through increasingly sophisticated (and costly) cooling techniques but commodity processor manufacturers have opted to freeze clock speeds and instead increase the number of processing cores (additional processors) in the same chip. In order to harness the additional computational power of modern CPUs it is therefore necessary to embrace parallel programming. In the next section I

[6] In his original formulation the transistor count doubled every year.

survey the various models of parallel computing and assess their applicability to the simulation of clock-driven spiking neural networks.

2.10.1 Shared memory architecture

The first distinction to draw in parallel computing systems is whether or not all processors have direct access to the same memory or not. Systems with shared memory have the advantage that data may be shared among different processors much more quickly than their counterparts with a distributed architecture since the internal circuitry of a computer has latency orders of magnitude lower than the networking fabric used to connect distributed systems. This lowers the thread-synchronization overheads which are necessary at the end of each time-step to ensure data-dependency requirements are not violated. This remains problematic for parallelizing any numerical ODE solver, but there is still significant opportunity for a performance increase from parallelizing within each time-step, particularly if the workload is large compared to the thread (data) synchronization overhead (as becomes the case for large neural networks).

2.10.1.1 Specialized Hardware

There are several specialized types of processor which may be tailored to a specific problems such as SpiNNaker (Jin et al., 2010) which is capable of simulating up to 10^9 Izhikevich neurons in real time and Application-specific integrated circuits (ASIC). However, specialized solutions tend to have a high initial cost and are soon overtaken by the performance gains of general purpose processors according to Moore's law. While other specialized solutions are more general in their use, such as Digital Signal Processing (DSP) boards, field-programmable gate arrays (FPGA), or IBM's Cell microprocessor, their esoteric architectures can make programming them difficult and therefore slow the development and optimization processes.

2.10.1.2 General Purpose GPU computing

One particularly notable approach using a specialized form of hardware is that of graphics processing unit (GPU) programming. Graphics card programming has recently attracted a lot of interest in the scientific community and has produced impressive performance increases with both rate-coded (Pinto et al., 2009) and (Izhikevich) spiking neural networks (Nageswaran et al., 2009a,b). However, given that these techniques are still in their infancy (with the initial specifications of the programming languages

only being released in February 2007 (CUDA) and December 2008 (OpenCL)), a more mature standard and tool-set was sought to avoid spending too much time on implementing the model (as opposed to exploring its parameter space and running biologically relevant simulations). At the time of writing, GPU programming appears to be a highly promising direction for improving simulation efficiency but developing such a model currently requires a significant degree of practical 'research and experimentation' in its own right.

2.10.1.3 Multicore computing and Symmetric multiprocessing

Virtually all commodity CPUs are now multi-core (two or more identical processors on the same chip), making them symmetric multiprocessing (SMP) systems (where each core shares the system memory and accesses it via a bus (or equivalent). Bus contention (i.e. competition between cores for memory access bandwidth) limits the number of cores from scaling efficiently, however, a recent shift away from the traditional front-side bus (FSB) model to using an integrated memory controller with a point-to-point connection architecture (such as Intel's QPI technology) has mitigated this to some extent and workstations with 12 cores (spread over two CPUs) are available at the time of writing. Various pipe-lining techniques such as Intel's Hyper-threading further raise the number of (pseudo-)cores in a system to 24. Since these systems are readily available and use a standard programming model with mature parallelization instruction sets such as 'OpenMP' (Chandra et al., 2001), this option was attractive for its less steep learning curve, substantial support base and open-source standards-based compilers.

2.10.2 Distributed memory architecture

Distributed computing systems are composed of multiple distinct nodes, each with their own processor(s) and memory. Communication between nodes is by means of a network and is significantly slower than communication within a node. They are therefore best suited to 'embarrassingly parallel' problems where each part of the computation is largely (or completely) independent of each other part requiring little (or no) communication between nodes e.g. ray-tracing and genetic algorithms.

2.10.2.1 Cluster computing

A cluster is a set of (typically homogeneous) nodes connected by a high speed local area networking fabric for inter-node communication. This may use standard

networking protocols e.g. TCP/IP for Beowulf clusters or more specialized, lower latency networking fabrics. Beowulf clusters communicating via MPI have been successfully used to simulate $\mathcal{O}(10^{11})$ neurons for one second of biological time[7] — see Izhikevich and Edelman (2008) for a scaled-down version. At the time of writing, clusters constitute the majority of the top 500 fastest supercomputers[8], however require a highly significant capital investment.

2.10.2.2 Grid computing

While similar in concept to cluster computing, grid computers are not necessarily all on the same local area network, but are distributed over the internet and communicate with the standard TCP/IP protocol. They are often individuals' personal computers which run jobs when the screen saver is active and hence use only spare CPU cycles. While communication between the nodes is slower and grids are typically highly heterogeneous, the sheer number of nodes may make them valuable computing resources (in the order of PFLOPS[9]) for example SETI@home, Folding@Home and OpenMacGrid.

2.10.3 Conclusion: Simulation strategy

Neural networks are not generally suited to distributed memory architectures owing to the high degree of communication between different parts of the network. This is particularly the case for clock-driven algorithms since all processing must be synchronized at least at the end of each time-step to satisfy data-dependency restrictions, meaning that the transfer of data for this stage would be the limiting factor in the performance of such a simulation.

The most suitable architecture is therefore a shared memory system, with the standard multicore CPU model being preferred for its availability, cost-effectiveness and mature tool sets (however for future research, GPU programming would be the preferred model as a compliment to other threads running on a multicore CPU). This also allows for convenient speed-improvements as faster workstations are acquired over the course of the research, since the same code will run faster without any alteration. For this purpose, the 'C' programming language with OpenMP for multi-threading were chosen for the very efficient binaries they produce

[7] http://www.izhikevich.org/human_brain_simulation/Blue_Brain.htm: Simulation of Large-Scale Brain Models

[8] http://www.top500.org/lists/2012/06

[9] 1 P(eta)FLOP := 10^{15} Floating point operations per second.

However, with even a moderately detailed neural network, there may be $\mathcal{O}[10^1, 10^2]$ parameters. This parameter space is often non-linear and so to explore even small portions of it in a systematic way requires a vast amount of simulation time. Since a simulation with one parameter set is entirely independent of a simulation with a different parameter set — an 'embarrassingly parallel' problem — this process may be greatly facilitated by utilizing a distributed memory system as a complimentary level of parallelism. While each simulation is parallelized across the cores of a given node to reduce individual simulation runtime, different parameter sets may be run asynchronously across different nodes (and the results collected and analysed independently) to also increase the speed of parameter-space exploration.

To this end, I developed a grid using Apple's Xgrid technology, which uses the spare CPU cycles of all of the group's workstations and anyone else's across the Department or internet in general who wishes to aid the simulations.

Let us assume that the persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability... When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

—Donald O. Hebb

3 Learning Mechanisms

The ventral visual pathway achieves object and face recognition by building transformation-invariant representations from elementary visual features. In previous computer simulation studies with rate-coded neural networks, the development of transformation-invariant representations has been demonstrated using either of two biologically plausible learning mechanisms, Trace learning and Continuous Transformation (CT) learning. However, it has not previously been investigated how transformation-invariant representations may be learned using these mechanisms in a more biologically accurate spiking neural network. A key issue is how the synaptic connection strengths in such a spiking network might self-organize through Spike-Time Dependent Plasticity (STDP) where the change in synaptic strength is dependent on the relative times of the spikes emitted by the pre- and postsynaptic neurons rather than simply correlated activity driving changes in synaptic efficacy. Here we present simulations with conductance-based integrate-and-fire (IF) neurons using a STDP learning rule to address these gaps in our understanding. It is demonstrated that with the appropriate selection of model parameters and training regime, the spiking network model can utilize either Trace-like or CT-like learning mechanisms to achieve transformation-invariant representations.

3.1 Introduction

The increasingly complex cell response properties of the primate ventral visual stream strongly suggest the functional organization of this pathway is that of a feature

hierarchy. Cells in the early stages (V1) are found to be sensitive to oriented bars and edges appearing in particular locations on the retina (Hubel and Wiesel, 1968). Information analysis of natural scenes reveals these features to be the most statistically independent components of such images (Bell and Sejnowski, 1997; van Hateren and van der Schaaf, 1998) and hence the most natural 'building-blocks' for such a system. Through successive layers, there follows a convergence of receptive fields allowing neurons at the end of the pathway in anterior Inferotemporal cortex (aIT) to view the entire retina and respond to increasingly complex stimuli (Tanaka, 1996). Here, and more recently in the medial temporal lobe (Quiroga et al., 2005), neurons have been found which respond with translation (Op de Beeck and Vogels, 2000), size (Ito et al., 1995) and view invariance (Booth and Rolls, 1998) to objects (Tanaka et al., 1991) and faces (Desimone, 1991).

Several groups have attempted to understand how elementary features may be combined into more complex view-invariant representations of whole objects with hierarchical feed-forward neural network models such as the *Neocognitron* (Fukushima, 1988), the *SEEMORE* system (Mel, 1997), the *HMAX* model (Riesenhuber and Poggio, 1999) and *VisNet* (Wallis and Rolls, 1997). These models are all composed of 'rate-coded' neurons (McCulloch and Pitts, 1943) which consist of applying a non-linear function (e.g. threshold or sigmoid) to a weighted sum of inputs (Boolean, or real values) which they receive at each computational step[1].

Within this paradigm, two main biologically plausible learning mechanisms have been discovered which explain how different views of the same object may be bound together and recognized as the same entity. The first of these, *Trace learning* (Földiák, 1991), relies upon temporal continuity, while the second, *Continuous Transformation (CT) learning* (Stringer et al., 2006), relies upon spatial continuity to associate together successive transforms and build view-invariant representations in later layers. While the properties of these mechanisms have been explored extensively in rate-coded models, it remains an open question as to how they might map onto a more biologically realistic spiking-neuron paradigm.

Spiking Neural Networks (SNN) can solve problems at least as complex as those that rate-coded models can solve (Šíma and Orponen, 2003), which in turn have greater computational power than Turing machines, and as such have been applied to a wide variety of problems, including modelling object recognition (Michler et al.,

[1]These early neuron models were designed to show that the elementary components of the brain could compute elementary logic functions. The belief commonly held at the time being that intelligence is based upon symbolic reasoning, which in turn rests upon the foundations of logic.

2009). By more faithfully modelling the electrical properties of neurons, spiking neural network model parameters may be more meaningfully mapped onto the biophysical properties of their real counterparts. This motivates the use of the conductance-based 'leaky' integrate-and-fire (LIF) model (described in Section 3.2) over models which are computationally cheaper or have a less apparent correspondence to measurable biological parameters such as the Spike Response Model (Gerstner and Kistler, 2006) or Izhikevich's nullcline derived model (2003).

Since time is explicitly and accurately modelled in SNNs, they allow quantitative investigation of the time-course of processing on such tasks (Thorpe et al., 2000; VanRullen and Thorpe, 2002) providing further arguments against rate-coding on the basis that Poisson rate-codes are too inefficient to account for the rapidity of information processing in the human visual system[2] (VanRullen and Thorpe, 2001; Thorpe et al., 1996). Furthermore, SNNs allow the investigation of qualitative effects such as the selective representation of one stimulus over another by the synchronization of its population of feature-neurons as found in neurophysiological studies (Fries et al., 2002; Kreiter and Singer, 1996). Similarly, the phenomenon of Spike-Time Dependent Plasticity and its effect upon learning transformation-invariant representations may only be investigated by modelling individual spikes which is of great importance to the present research.

Hebb originally conjectured that synapses effective at evoking a response should grow stronger (Hebb, 1949), capturing a causal relationship between the two neurons. This was eventually simplified (partly for the purposes of rate-coded models) to become interpreted as any long-lasting synapse-specific form of modification dependent upon correlations between presynaptic and postsynaptic firing. This is usually expressed in the form $\delta w_{ij} = k y_i x_j$, where δw_{ij} is the change in synaptic strength, k is a learning rate constant, and x_j and y_i are the firing rates of the presynaptic and postsynaptic neurons (see e.g. Rolls and Treves, 1998).

Progress in neurophysiology has shown however, that the all-or-nothing nature of an action potential means that the information may be conveyed by the number *and* the timing of action potentials (Ferster and Spruston, 1995; Maass and Bishop, 1999), typically neglecting their size and shape in modelling. In other words neurons communicate by a *pulse* code (a time series of discrete binary events) rather than simply a *rate* code (a moving average level of activity) which has been convincingly demonstrated in the sensory systems of several organisms, such as echolocating bats (Kuwabara and Suga, 1993) and the visual systems of flies (Bialek et al., 1991).

[2]Typically, only 100–150 *ms* is required to respond to complex stimuli

It is also now well established that *synaptic plasticity* is sensitive to the relative timing of the presynaptic and postsynaptic spikes (Dan and Poo, 2006; Markram et al., 1997), typically becoming approximately exponentially less sensitive as the time difference increases (Bi and Poo, 1998). This has been found to take several forms in different brain regions (Abbott and Nelson, 2000) but here we focus on the form observed in retinotectal connections and neocortical and hippocampal pyramidal cells where $pre \rightarrow post$ spike pairs lead to synaptic potentiation (with greater effect over shorter intervals) and the opposite ordering of spikes leads to synaptic depression.

The challenge now is to investigate how the timing of spikes affects the self-organization of the system applied to the problem of developing transformation-invariant representations and understanding how the CT and Trace learning mechanisms, which have been developed in the context of rate-coded models, might fit into a model of Spike-Time Dependent Plasticity (STDP).

3.2 Methods

3.2.1 Network Architecture

While the ventral visual stream is typically modelled as four or more layers of neurons with excitatory modifiable feed-forward synapses and a mechanism of lateral inhibition, here we seek to understand the mechanisms operating at each layer which ultimately may lead to transformation-invariant representations, hence a simpler architecture is used.

The model consists of two layers of excitatory pyramidal neurons with one layer of modifiable feed-forward synapses between them (as shown in Figure 2.1). Within each layer there are also inhibitory interneurons with non-plastic lateral synaptic connections to and from the excitatory neurons to produce a degree of competition between the excitatory neurons.

For all presented simulations in this chapter, 400 excitatory neurons and 100 inhibitory neurons were used in each layer, with full connectivity. Each neuron is based upon the standard conductance-based leaky integrate and fire (LIF) model (see for example Rolls and Treves, 1998) while the equations for STDP at the Excitatory–Excitatory ($E \rightarrow E$) synapses are adapted from Perrinet et al. (2001).

3.2.2 Neuron model

Depolarization of the neuron's membrane potential is described by Equation 3.1 and the cell (and synapse) constants were chosen to be as biologically accurate as possible based upon the available neurophysiological literature (see Table 3.1 for a full list).

The cell membrane potential for a given neuron (indexed by i) is driven up by presynaptic excitatory conductances (or direct current injection) and towards $\hat{V}^I$, the inhibitory reversal potential (typically down) by presynaptic inhibitory conductances, decaying back to its resting state over a time course determined by the properties of its membrane.

$$\tau_m^\gamma \frac{dV_i(t)}{dt} = V_0^\gamma - V_i(t) + R^\gamma I_i(t) + R^\gamma I_i^{ext}(t) + \sigma \cdot \xi(t) \cdot \sqrt{\tau_m^\gamma} \tag{3.1}$$

Here τ_m represents the membrane time constant, defined as $\tau_m = {}^{C_m}/_{g_0}$, where C_m is the membrane capacitance, g_0 is the membrane leakage conductance and R is the membrane resistance, $(R = {}^{1}/_{g_0})$. V_0 denotes the resting potential of the cell (indexed by γ along with these other class-specific parameters), $I_i(t)$ represents the total synaptic current (described in Equation 3.2) and $I_i^{ext}(t)$ models the injected current.

In addition, Gaussian white noise was added to the cell membrane potential with zero mean and standard deviation $\sigma = 0.015 \cdot (\Theta - V_H)$ as used by Masquelier et al. (2009). Here, $\xi(t)$ is a Wiener (Gaussian) variable as described in Section 2.6.

The total synaptic conductance is the sum of conductances of all presynaptic neurons of each type (excitatory and inhibitory) with inhibitory conductances being negative.

$$I_i(t) = \sum_\gamma \sum_j g_{ij}(t) \left(\hat{V}^\gamma - V_i(t) \right) \tag{3.2}$$

Here $\hat{V}^\gamma$ represents the reversal potential of a particular class of synapse (denoted again by γ) which consists of Excitatory and Inhibitory neurons $\{E, I\}$ and j indexes the presynaptic neurons of each class.

The synaptic conductance of a particular synapse, $g(t)$, (indexed by ij) is governed by a decay term τ_g and a Dirac delta function for when spikes occur, which correspond to the first and second terms of Equation 2.19.

The same STDP model is used to modify the feed-forward excitatory synapses between the principal cells of each layer as described in Section 2.3.1 (Perrinet et al., 2001). This is a multiplicative form of STDP, meaning that the amount of potentiation

decreases as the synapse strengthens, as has been found experimentally (Bi and Poo, 1998). Multiplicative weight-dependent potentiation yields a normal distribution of synaptic efficacies rather than pushing each weight to one extreme or the other (van Rossum et al., 2000) as would be the case with an additive form of STDP, which can be seen in results section.

3.2.3 Numerical Scheme

The differential equations described above are converted to finite difference equations and simulated using the Forward Euler numerical scheme with a time-step $\Delta t = 0.02\,ms$. In the finite difference equations, the Dirac delta function has been replaced by the discrete approximation, S(x) as defined in Amit and Brunel (1997). Finally, in the original description, the change in synaptic weight (Equation 2.29) was instantaneous and so $^{\Delta t}/_{\tau_{\Delta g}}$ is defined to be a learning rate constant, ρ, in the corresponding finite difference equation.

3.2.4 Training and Stimuli

Stimuli are represented by injecting a small amount of current directly into the cell bodies of a particular set of excitatory input neurons continuously throughout the cue period. This pattern of stimulated neurons is gradually shifted across the input layer representing successive transforms of the stimulus (see Figure 3.1).

The size of a stimulus and the amount of neurons each of its transforms is shifted by allows us to precisely control the degree of overlap between transforms of each stimulus. Spatial continuity is crucial to the functioning of the CT learning mechanism, whereas the Trace learning mechanism requires temporal continuity to associate successive transforms together. These stimulus properties may be controlled independently of the model time constants. In this way, we may eliminate the operation of one mechanism to study the other in isolation and hence disentangle their contributions to the network's capacity for invariance learning.

During training, the set of stimuli are presented in a random order with all transforms for a given stimulus being presented in succession before presenting the next stimulus's transforms. Presentation of all stimuli in this manner constitutes one training epoch, and the total training period comprised of ten such epochs.

After completion of training, learning is switched off (prohibiting further synaptic modification) and the network is presented with all transforms of all stimuli in order

Parameter	**Symbol**	**Value**	**Reference**
Cue current	I^{ext}	1.0 nA	*
Cue period {training, testing}	t_{cue}	{100, 250} ms	
Time-step	Δt	0.02 ms	
Network Parameters			
No. layers	N_L	2	
No. excitatory cells per layer	N_E	400	
No. inhibitory cells per layer	N_I	100	
No. afferent excit. connections per excit. neuron	S_{EE}	400	
No. afferent excit. connections per inhib. neuron	S_{EI}	400	
No. afferent inhib. connections per excit. neuron	S_{IE}	100	
No. afferent inhib. connections per inhib. neuron	S_{II}	100	
Cellular parameters			
Excitatory cell somatic capacitance	C_m^E	500 pF	§
Inhibitory cell somatic capacitance	C_m^I	214 pF	§
Excitatory cell somatic leakage conductance	g_0^E	25 nS	§
Inhibitory cell somatic leakage conductance	g_0^I	18 nS	§
Excitatory cell membrane time constant	τ_m^E	20 ms	§
Inhibitory cell membrane time constant	τ_m^I	12 ms	§
Excitatory cell resting potential	V_0^E	-74 mV	§
Inhibitory cell resting potential	V_0^I	-82 mV	§
Excitatory firing threshold potential	Θ^E	-53 mV	§
Inhibitory firing threshold potential	Θ^I	-53 mV	§
Excitatory after-spike hyperpolarization potential	V_H^E	-57 mV	§
Inhibitory after-spike hyperpolarization potential	V_H^I	-58 mV	§
Excitatory reversal potential	$\hat{V}^E$	0 mV	§
Inhibitory reversal potential	$\hat{V}^I$	-70 mV	§
Absolute refractory period	τ_R	2 ms	§
Synaptic parameters			
Synaptic neurotransmitter concentration	α_C	0.5	†
Proportion of unblocked NMDA receptors	α_D	0.5	†
Presynaptic STDP time constant	τ_C	[3, 75] ms	†
Postsynaptic STDP time constant	τ_D	[5, 125] ms	†
Synaptic learning rate	ρ	0.1	†
Plastic ($E \to E$) synaptic conductance range, CT	$\lambda \cdot \Delta g^{EE}$	[0, 4] nS	*
Plastic ($E \to E$) synaptic conductance range, Trace	$\lambda \cdot \Delta g^{EE}$	[0, 1.25] nS	*
Change in synaptic conductance ($I \to E$)	$\lambda \cdot \Delta g^{IE}$	[0.5, 2.5] nS	*
Change in synaptic conductance ($E \to I$)	$\lambda \cdot \Delta g^{EI}$	5.0 nS	*
Change in synaptic conductance ($I \to I$)	$\lambda \cdot \Delta g^{II}$	5.0 nS	*
Excitatory–Excitatory synaptic time constant	τ_{EE}	{2, 150} ms	*
Inhibitory–Excitatory synaptic time constant	τ_{IE}	5 ms	§
Excitatory–Inhibitory synaptic time constant	τ_{EI}	2 ms	§
Inhibitory–Inhibitory synaptic time constant	τ_{II}	5 ms	§

Table 3.1: Parameters used in the learning mechanisms simulations. Most integrate and fire parameters were taken from Troyer et al. (1998) (derived originally from McCormick et al. 1985) as indicated by §. Plasticity parameters (denoted by †) are taken from Perrinet et al. (2001). Parameters marked with * were tuned for the reported simulations.

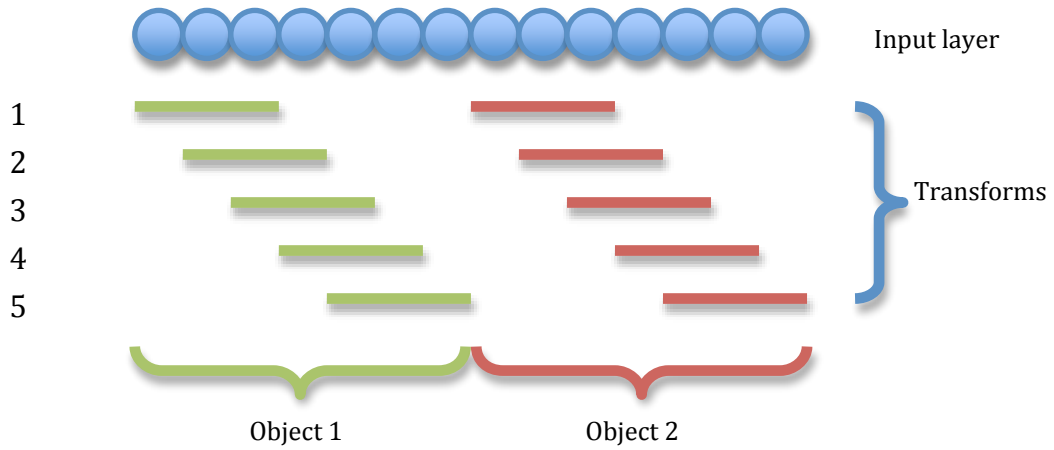


Figure 3.1: Illustration of the transforms of two stimuli. The input layer is divided into as many equal portions as there are stimuli and all transforms of a particular stimulus are confined to that stimulus's portion of the input neurons. In this illustration, there are five transforms per object shifting by one neuron.

(resetting the neurons to their resting state between transforms) and the resultant firing in both input and output layers is saved for analysis.

3.2.5 Performance Measures

Two information-theoretic[3] measures are used to assess the network's performance which reflect the extent to which cells respond invariantly to a particular stimulus over several transforms but differently to other stimuli (for more details see Rolls and Milward (2000); Elliffe et al. (2002)). The work presented has used spiking neural networks because we believe that the richer dynamics they model are critical for learning to solve the problem of object recognition (transformation-invariant cell responses). However, analysis of macaque visual cortical neuron responses has found that after learning, the majority of the information about stimulus identity is contained within the *firing rates* rather than the detailed timing of spikes (Tovée et al., 1994). As such, we adopt a dual approach whereby the network self-organizes through spiking dynamics but the information content with respect to stimulus identity is assessed through the output cell's firing rates.

During testing each transform of each stimulus was presented to the input layer of the network. Each neuron was reset (allowed to settle) after presentation of each transform such that the activity due to one transform did not affect the responses to

[3]For a general introduction to Information Theory see MacKay, 2003.

later transforms. After testing, the spikes of each output neuron were placed into a different bin for each transform of each stimulus and the corresponding firing rate for each cell was calculated. Based upon these firing rates, the stimulus-specific single-cell information $I(s, R)$ was calculated according to Equation 3.3, which gives the amount of information in a set of responses R of a single cell about a specific stimulus s. The set of responses, R consisted of the firing rate of a cell to every stimulus presented in every location.

$$I(s, R) = \sum_{r \in R} P(r|s) \log_2 \frac{P(r|s)}{P(r)} \tag{3.3}$$

Good performance for a cell would entail stimulus specificity (with generality across most or all transforms of that stimulus), meaning a large response to one or a few stimuli regardless of their position (transform) and small responses to other stimuli. We therefore compute the maximum amount of information a neuron conveys about *any* of the stimuli rather than the average amount it conveys about the whole set S of stimuli (which would be the mutual information).

If all the output cells learnt to respond to the same stimulus then there would be no discriminability and the information about the set of stimuli S would be poor. To test this, the multiple cell information measure is used which calculates the information about the set of stimuli from a population of up to ten output neurons. This population consisted of the subset of up to five cells which had, according to the single cell measure, the most information about each of the two stimuli. Ideally, we would calculate the mutual information (the average amount of information about which stimulus was shown from the responses of all cells after a single presentation of a stimulus, averaged across all stimuli), however the high dimensionality of the neural response space and the limited sampling of these distributions is prohibitive.

Instead, a decoding procedure is used to estimate the stimulus s' that gave rise to the particular firing rate response vector on each trial. From this a probability table is then constructed of the real stimuli s and the decoded stimuli s', from which the mutual information is calculated (Equation 3.4).

$$I(s, s') = \sum_{s, s' \in S} P(s, s') \log_2 \frac{P(s, s')}{P(s)P(s')} \tag{3.4}$$

A Bayesian decoding procedure is used for this purpose, whereby the firing rates of each cell in the ensemble vector to each transform of each stimulus in turn is fitted to a Gaussian distribution parameterized by these means and standard deviations of each

cell's responses to all other transforms of each stimulus separately to yield an estimate of $P(r_c|s')$. Taking the product of these probabilities over all cells in the response vector with $P(s')$ and then normalizing the resultant joint probability distribution gives an estimate of $P(s'|\mathbf{r})$, (Földiák, 1993). These probability distributions are factored into a confusion matrix of $P(s, s')$ over many iterations to smooth the effects of randomly sampling the output cells. From this decoding and cross-validation procedure, the probability tables are constructed for calculating the multiple cell information measure, further details of which may be found in Rolls et al. (1997a). This measure should increase up to the theoretical maximum $\log_2 N_S \, bits$, (where N_S is the number of stimuli), as a larger population of cells is used, only if those cells have become tuned to different stimuli.

3.3 Results

In the simulations described below we investigated invariance learning in a spiking neural network with Spike-Time Dependent Plasticity utilizing two different learning mechanisms. For details of the methods and parameters used for the following simulations, please refer to Section **??** and Table 3.1 respectively.

3.3.1 Continuous Transformation Learning

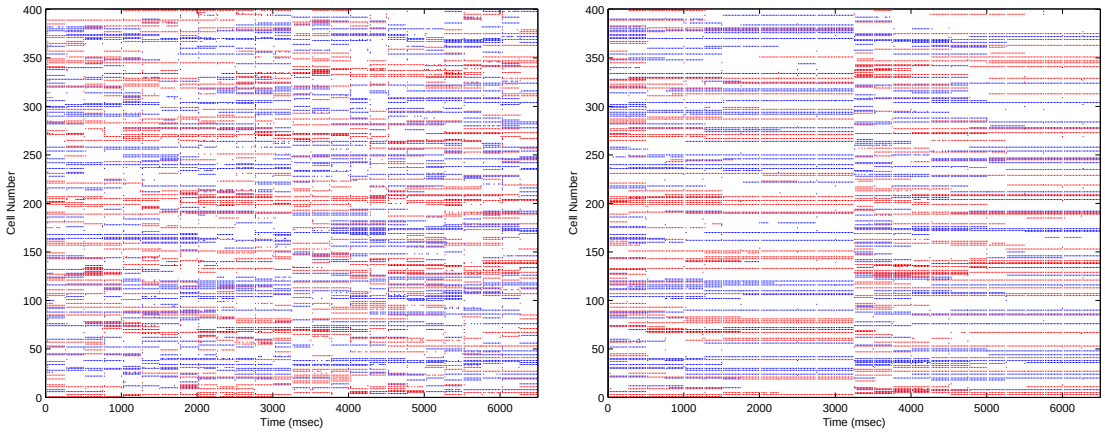
Continuous Transformation (CT) learning relies upon the spatial continuity of continuously transforming stimuli, a purely associative (Hebbian) learning rule in the feed-forward connections between layers, and lateral competition within each layer to associate together successive transforms of a stimulus (Stringer et al., 2006). Presentation of an initial transform to the presynaptic input cells will excite one or more postsynaptic neurons and through the Hebbian learning rule, will strengthen the feed-forward synapses between those cells. If there is enough overlap (similarity) between the initial transform and a subsequent transform, then the subsequent transform will excite the same postsynaptic neurons. These postsynaptic neurons will then also strengthen their synaptic connections from the input cells representing the subsequent transform. This process can continue across a series of overlapping transforms until they are all mapped onto the same output cells. Since similar images are more likely to be transforms of the same object than different stimuli, the CT mechanism provides an explanation for how transformation-invariant representations may develop in the ventral visual system.

In this set of simulations, the parameters were chosen to encourage the operation of the CT learning mechanism (Stringer et al., 2006) while excluding any trace-like effects (Földiák, 1991). To this end, spatial overlap between successive transforms was generally kept high (13 transforms per stimulus each covering 56 neurons and shifting by 12 neurons per transform by default). Also a short time constant of $2\,ms$ was used for the Excitatory–Excitatory (feed-forward) synaptic conductances, τ_{EE}. These conditions were hypothesized to support a CT-like learning mechanism in a spiking neural network.

3.3.1.1 Invariance Learning with CT

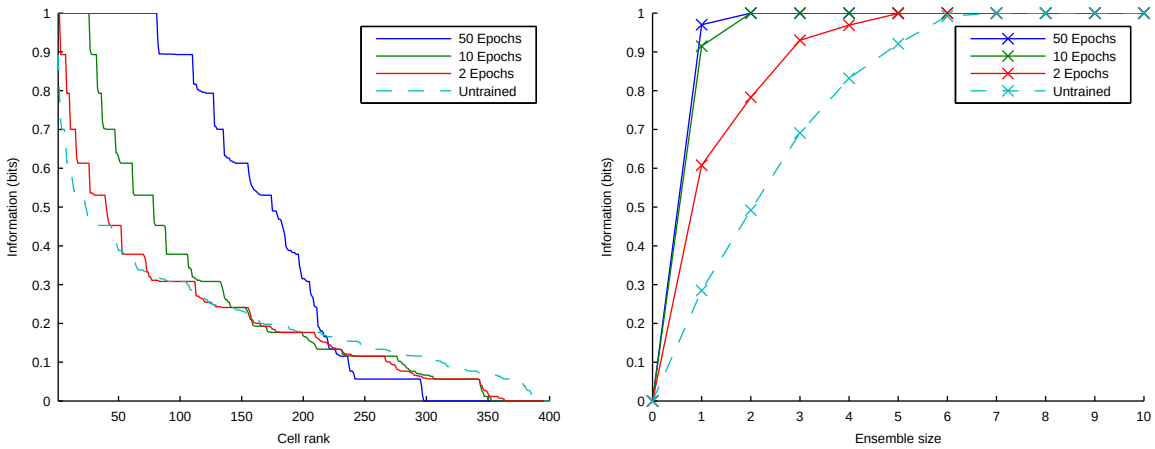
This simulation demonstrates the formation of transformation-invariant representations of two stimuli in the output cells through STDP as illustrated by the raster plots in Figure 3.2 (which contrast the untrained with the trained network) and the information plots of Figure 3.3a and 3.3b. The level of inhibition in the input layer had to be tuned so that the spikes from additional neurons from successive transforms (in the input layer) could be brought into phase with those already firing from the previous transforms. While the feed-forward excitatory weights were plastic and hence modified through learning, their maximum level was set to $4\,nS$ to achieve a reasonable level of output layer activity for the network size and connectivity. It can be seen from the pre-training raster plot (Figure 3.2a) that before learning, output neurons respond to a random set of transforms of each stimulus. However, post-training (Figure 3.2b), there are several cells which are responsive across the whole set of transforms for the first stimulus which are presented contiguously over the first $3250\,ms$, and several other cells which respond to all transforms of the second stimulus presented contiguously over the second period of $3250\,ms$.

In accordance with the raster plot, the $I(s, R)$ (single-cell information measure) plots show many more cells in the network have attained the maximum information content ($1\,bit$) than in the untrained case, demonstrating both transformation-invariance and stimulus specificity. Examining the $I(s, s')$ (multiple cell information measure) plots shows that the maximum information about the stimulus set S is reached with fewer than the 10 available cells of the output ensemble of the highest scoring cells (in terms of their $I(s, R)$ values), thus confirming that both stimuli are represented invariantly.



(a) Before Training (b) After Training

Figure 3.2: Raster plots of output layer cells before and after training from the CT baseline simulation. Transforms 1–13 of Stimulus 1 were presented during 0–3250 ms, followed by transforms 1–13 of Stimulus 2 presented during 3250–6500 ms. Before training (a) the output cells respond randomly to transforms of each stimulus. After training (b), the raster plot shows cells sensitive to all transforms of stimulus one (0–3250 ms) and other cells sensitive to all transforms of stimulus two (3250–6500 ms).



(a) Single Cell Information (b) Multiple Cell Information

Figure 3.3: Information plots for the baseline CT learning demonstration of translation-invariant representations for varying degrees of training: untrained, 2 epochs, 10 epochs and 50 epochs. The single cell information analysis (a) shows that for ten epochs of training, approximately 40 output cells have achieved the maximal information content of 1 bit. With 50 training epochs, the number of cells reaching 1 bit of information rises to approximately 100. The multiple cell information plot (b) confirms that both of the stimuli are represented by cells which are exclusively tuned to one stimulus or the other.

3.3.1.2 Temporal Specificity of STDP

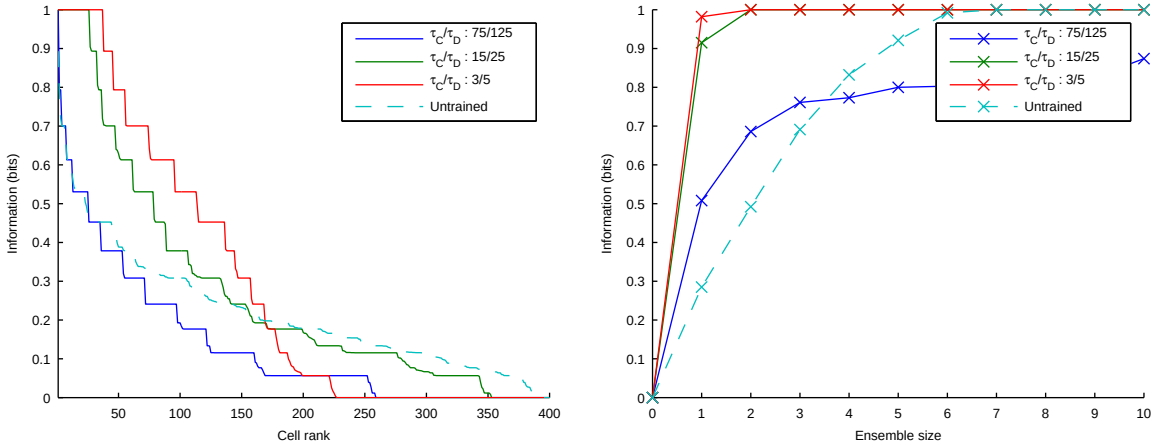
By default, the learning time constants, τ_C and τ_D, used in these simulations are 15 and 25 ms in accordance with Perrinet (2003). Here we reran the same simulations but shortened or lengthened these time constants by a factor of five (maintaining the same $^3/_5$ ratio) to give $3/5\,ms$ and $75/125\,ms$ for τ_C/τ_D respectively. Figure 3.4a shows a trend of a much greater information content in the network with the shorter (more temporally specific) time constants ($3/5\,ms$) with the accompanying $I(s, s')$ plot confirming that both stimuli are being represented (see Figure 3.4b). Network performance drops however with the longer (less temporally specific) STDP time constants 75/125 as the learning rule is less capable of capturing the temporally specific causal relationship of the input/output spike volleys.

The effect of shortening the STDP time constants is that after a pre–post spike pairing results in LTP, the following presynaptic spike from the next wave comes a relatively long time after the initial pair, such that the effect of its post–pre LTD is significantly lessened. The synaptic weight distributions in Figure 3.5 support this, exhibiting a peaked distribution of synaptic efficacies arising from the initially flat uniform distribution (as expected from a multiplicative model of STDP in the standard case, $\tau_C = 15\,ms$, $\tau_D = 25\,ms$, Figure 3.5b) and more peaked distributions with shorter STDP time constants (Figure 3.5c) indicating more specific learning. The higher proportions of large weights with the shorter learning time constants are what might be expected from an unbalanced learning rule dominated by LTP when waves of input spikes are widely spaced relative to the time delay until the postsynaptic spikes which they cause. In contrast, the weight distribution with the longer STDP time constants is smoother, indicating a less trained layer of synaptic weights (Figure 3.5a).

3.3.1.3 Lateral Inhibition and Synchrony

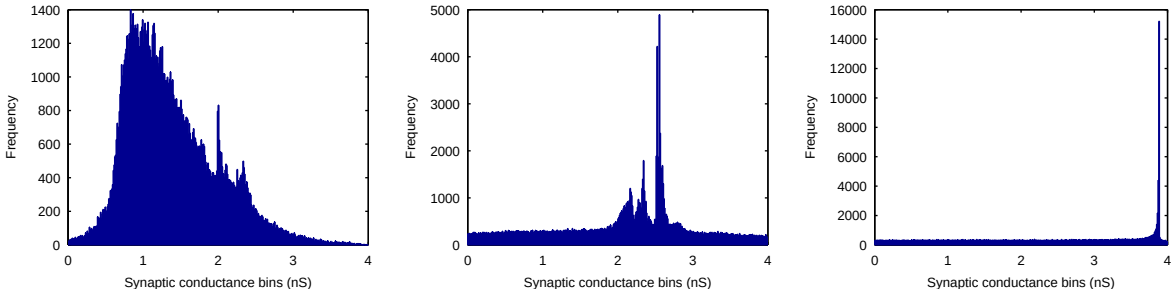
From earlier simulations, it is apparent that this training paradigm and the STDP model are very sensitive to the effects of the strength of inhibition on the synchronization of input spikes. We therefore systematically varied the strength of Δg_{IE}, the $Inhibitory \rightarrow Excitatory$ conductances (which were non-plastic) to understand these effects in more detail.

Figure 3.6 shows that as the level of inhibition is reduced and the cell membrane potential noise begins to cause jitter in the spike timings, the new input layer neurons from successive transforms no longer fire in phase with those neurons from previous



(a) Single Cell Information (b) Multiple Cell Information

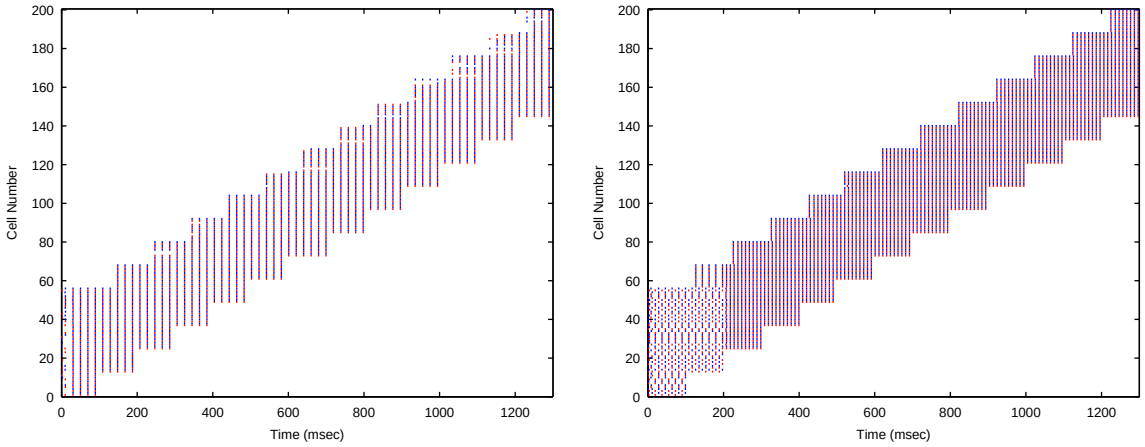
Figure 3.4: Information plots of varying STDP time constants, τ_C and τ_D (CT learning). With shorter time constants the single cell information content (a) is seen to increase as learning becomes more temporally specific. The multiple cell information (b) demonstrates that for short plasticity time constants, both stimuli are represented by the ensemble of output cells. STDP (with sufficiently temporally specific time constants) together with the synchronization of neuronal firing is here able to facilitate the learning of transformation-invariant representations.



(a) $\tau_C = 75\,ms, \tau_D = 125\,ms$ (b) $\tau_C = 15\,ms, \tau_D = 25\,ms$ (c) $\tau_C = 3\,ms, \tau_D = 5\,ms$

Figure 3.5: Synaptic Weight Distributions of varying STDP time constants (CT learning). Compared to the standard case $\{\tau_C = 15\,ms, \tau_D = 25\,ms\}$ (b), longer plasticity time constants $\{\tau_C = 75\,ms, \tau_D = 125\,ms\}$ (a) result in a smoother, more distributed profile of synaptic weights. With shorter time constants $\{\tau_C = 3\,ms, \tau_D = 5\,ms\}$ (c), the distribution is seen to become more peaked with larger synaptic weights as learning becomes more temporally specific and the weight updates experience more LTP.

transforms. This reduces invariance learning in the output layer, where the information content can also be seen to be reduced (Figure 3.7).



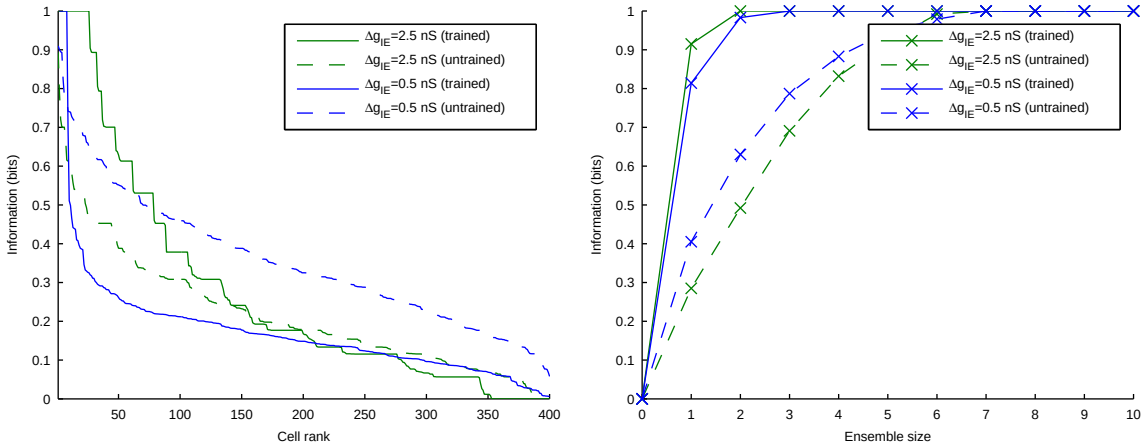
(a) $\Delta g_{IE} = 2.5\,nS$ (b) $\Delta g_{IE} = 0.5\,nS$

Figure 3.6: Raster plots of the training inputs (for all transforms of Stimulus 1 only) for two levels of inhibitory conductance. The raster plot produced with the standard inhibitory conductance strength of $\Delta g_{IE} = 2.5\,nS$ shows synchronized input volleys across all transforms of the stimulus (a). It can be seen at the lower inhibitory strength ($\Delta g_{IE} = 0.5\,nS$) that the neurons within some of the transforms become desynchronized with respect to one another (b).

3.3.1.4 Degree of overlap

Previous rate-coded simulations have shown that CT learning requires a high degree of resemblance among adjacent members of a set of transforms in order to associate them together (Stringer et al., 2006). If this mechanism is being employed in the present spiking model, its performance should suffer by reducing this transform similarity. This was tested by removing intermediate transforms leaving only every 2^{nd} or 3^{rd} transform from the original sets of 13 transforms per stimulus (with a consecutive transform overlap of 44 neurons) such that there were only 7 or 5 transforms per stimulus respectively. Since they still occupied the same proportion of the input layer, the degree of overlap between any two consecutive transforms was correspondingly lower, being 32 or 20 neurons respectively.

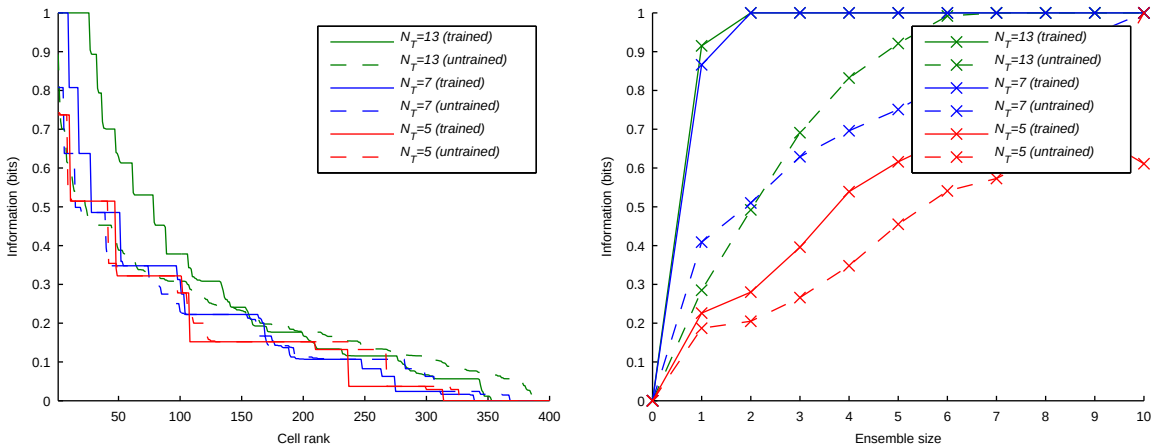
It can be seen from Figure 3.8 that by reducing the spatial overlap between successive transforms of each object, the information content of the output layer declines (despite there being fewer transforms to associate together) since there are fewer cells that respond invariantly across all transforms of a given stimulus. This



(a) Single Cell Information (b) Multiple Cell Information

Figure 3.7: Information plots for different degrees of inhibitory strength (CT learning). As the inhibitory strength decreases, the single cell information in the network declines markedly (a) with a milder decrease in the multiple cell information (b). This is due to the increased difficulty for the whole set of input neurons representing a particular transform of a stimulus to fire in synchrony.

confirms that the network is learning invariance by a spiking equivalent of the CT learning mechanism.

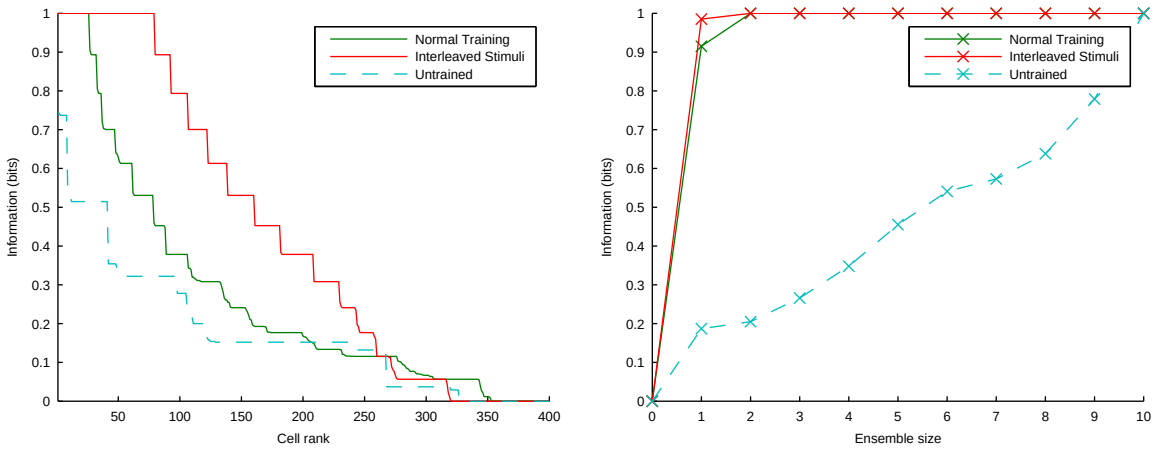


(a) Single Cell Information (b) Multiple Cell Information

Figure 3.8: Information plots for three different degrees of overlap between successive transforms. The information content of the output cells can be seen to fall, in the single cell measure (a) and the multiple cell measure (b), as the degree of overlap is reduced. This is a property of CT learning which requires a sufficient degree of spatial overlap between transforms to build invariance.

3.3.1.5 Interleaved Transforms

Since the degree of similarity between any two transforms of a stimulus is the same regardless of when they are presented to the network, under a CT learning regime it should not matter whether the transforms are seen close together in time or not. One of the key properties of CT learning is therefore its ability to enable a network to learn about stimuli, even when their transforms are interleaved with those of another stimulus (analogous to learning to recognize two faces or objects as the viewer saccades back and forth between them). To test this hypothesis we presented transforms of each stimulus alternately i.e. $S_1^{t_1}, S_2^{t_1}, S_1^{t_2}, S_2^{t_2}, \ldots, S_1^{t_n}, S_2^{t_n}$. If neurons are able to develop transformation-invariant responses with this training paradigm, it proves the learning mechanism is not utilizing a temporal trace.



(a) Single Cell Information

(b) Multiple Cell Information

Figure 3.9: Information plots showing the difference in training the network with the transforms of each stimulus presented consecutively and by interleaving transforms of both stimuli (CT learning). By interleaving the collections of transforms, the single cell information content has not declined (a) while the multiple cell information (b) confirms that both stimuli are still represented. This is a property of CT learning, which does not require different transforms of a stimulus to be seen consecutively in time.

Here it is evident from the information analysis (Figure 3.9) that the network has managed to learn about the individual stimuli with this additional constraint. Both the single and multiple information measures show not just comparable results to the consecutive presentation of each stimulus's transforms during training but surprisingly, an improvement over the standard case. Examining the input layer rasters, this enhancement to learning from interleaving the stimuli seems to be due to the fact that

under normal consecutive training, the first one or two spike volleys of a new transform have not yet recruited the additional neurons (which were not part of the previous transform) due to the lateral inhibition suppressing them. In contrast, neurons in the overlapping region are under constant stimulation from the injected current and so continue to fire, whereas those neurons exclusive to the previous transform stop firing when no longer stimulated with direct current.

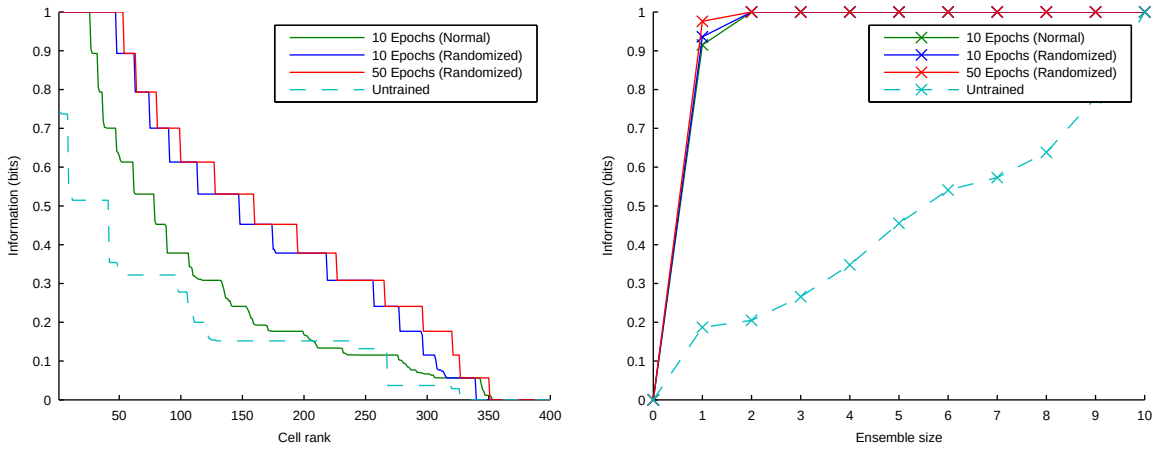
When the stimuli are interleaved however, all of the input neurons representing the new transform are stimulated by current injection at the same time (rather than their cell membrane potentials starting at different points in the stimulation cycle) and so fire simultaneously from the very first volley. Using a stimulating direct current of $1\,nA$ with the cell body parameters and network connectivity given in Table 3.1, the neurons will fire approximately five complete volleys of spikes in the $100\,ms$ presentation period ($50\,Hz$). The ultimate effect of this training difference is that in the interleaved case, each transform will be represented by five complete spike volleys (as opposed to only three of four in the standard case) and hence will be trained more fully (with more useful weight updates) over the same training duration.

3.3.1.6 Randomized Transform Order

CT learning is also able to form transformation-invariant representations when the individual transforms of an object are presented in a random order during training. This is analogous to learning to recognize a face or object from a number of random 'snapshot' views rather than seeing it move smoothly. The consequence of such a training regime is that there is not necessarily any overlap between two consecutive transforms in time. At the beginning of training when the feed-forward weights are randomly initialized, this training regime may mean that different output neurons learn to respond to different subsets of each stimulus' transforms, thus making it harder for the similarity-based CT mechanism to associate all related (overlapping) transforms together onto the same output neurons. If however, there is a sufficient number of such training epochs and degree of competition in the output layer, eventually each transform will be randomly followed by a similar enough transform such that the same postsynaptic cell is fired which eventually learns invariance across the whole set of transforms.

Initially randomizing the order of transforms degraded the network performance as expected. However, building upon the learning enhancement found in the previous

simulations with interleaving the stimuli, simulations were repeated with simultaneously randomized transform order and interleaved stimuli. Figure 3.10 demonstrates that the network is able to cope with randomizing the order of the transforms.



(a) Single Cell Information (b) Multiple Cell Information

Figure 3.10: Information plots showing the effect of training the network by randomizing the order of transforms (CT learning). Despite this paradigm representing more difficult learning conditions for a CT mechanism, the single cell (a) and multiple cell (b) information measures demonstrate good performance. CT learning is thus able to build invariance if enough overlapping transforms are presented to the network over time, irrespective of the temporal sequence they are presented in.

3.3.2 Trace Learning

Trace learning utilizes the temporal continuity of objects in the world to learn transformation-invariant representations (Földiák, 1991). The mechanism relies upon the proposal that over short time scales, successive images are more likely to be transforms of the same object rather than different objects. The trace learning rule (Földiák, 1991; Wallis and Rolls, 1997) uses these temporal statistics of visual input by incorporating a temporal trace of the previous (typically postsynaptic) neural activity into a simple Hebbian learning rule, which helps to associate recent activity in the postsynaptic neuron (due to a previous transform) with present activity in the presynaptic neurons (due to the current transform). Through further Hebbian synaptic modifications, successive transforms may become associated together onto the same output cells leading to transformation-invariant neurons.

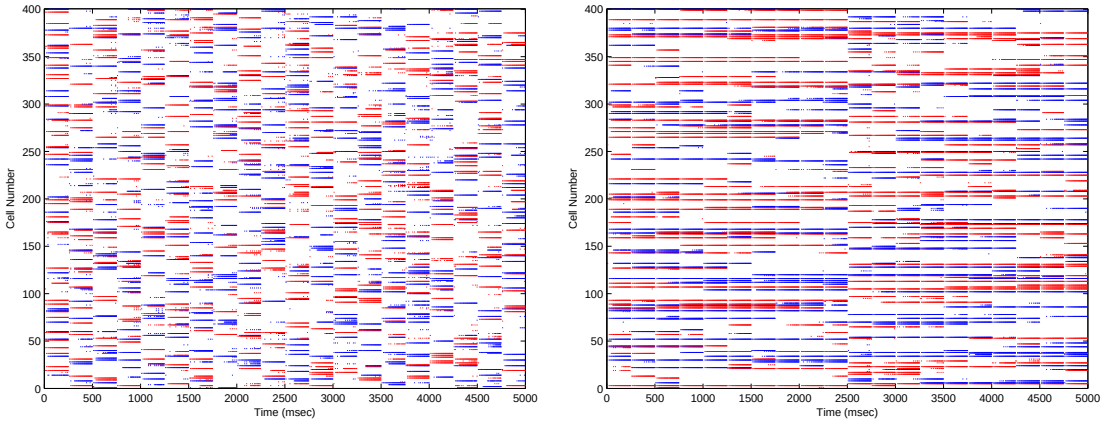
In contrast to the CT simulations (Section 3.3.1), here we lengthen the synaptic time constant τ_{EE} to $150\,ms$ to explore the hypothesis that by continuing to bleed

current into a postsynaptic neuron, the activity generated by one transform in the postsynaptic neuron may be associated with the next. In this way, a temporal trace effect may be achieved, allowing a spiking neural network to learn through temporal rather than spatial continuity.

3.3.2.1 Invariance Learning with a temporal trace

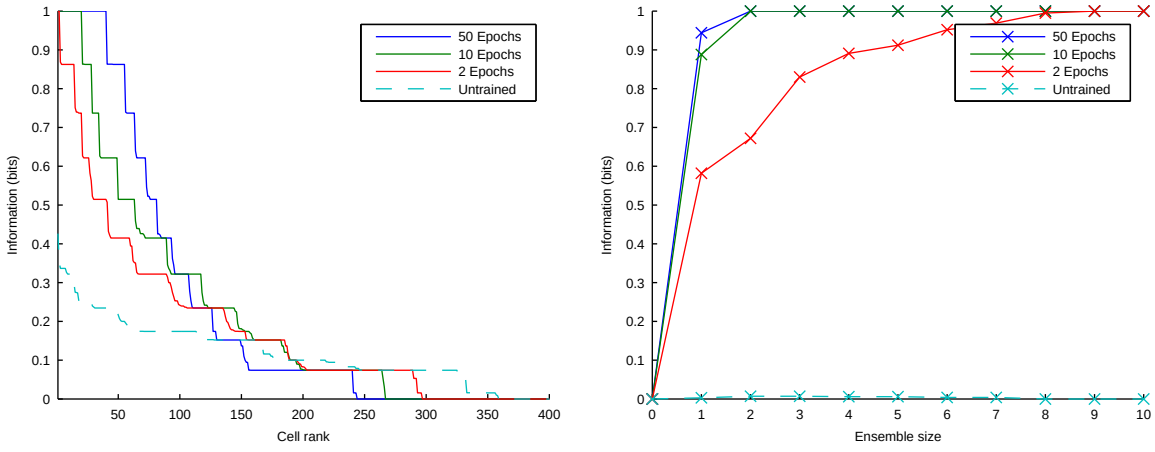
This simulation demonstrates the formation of transformation-invariant representations in the output cells through STDP and a trace-like effect from longer $E \rightarrow E$ synaptic time constants. The other parameters remained the same as in the CT simulations except that the maximum strength of the plastic feed-forward excitatory synapses was reduced to $1.25\,nS$ (from $4\,nS$) to compensate for the greater degree of excitation arising from the longer feed-forward synaptic time constant. Also the stimuli were changed such that in the following trace simulations, there are 10 transforms for each of the two stimuli (consisting of 20 neurons each) which are shifted by 20 neurons for each transform such that there is no spatial overlap between transforms. Since these transforms are orthogonal, any CT effects from spatial overlap are eliminated. Additionally, since the spatio-temporal statistics of natural stimuli tend to have different transforms of the same stimulus closer together in time more frequently than transforms of different stimuli, the neurons were allowed to settle between presentation of the two sets of transforms (stimuli) to effectively reduce the temporal continuity between different stimuli so as to avoid introducing an artificial trace effect between them. These changes allow for a controlled investigation of whether orthogonal transforms may be linked together by a trace-like learning mechanism by lengthening the excitatory synaptic conductance.

Due to the random initialization of the feed-forward weights, output neurons before training respond to a random set of transforms of each stimulus (Figure 3.11a), whereas after training Figure 3.11b shows both stimuli are represented by cells which are invariant to most transforms of their respective stimuli, while the information plots (Figure 3.12) confirm that both stimuli may be identified with a small ensemble of output neurons. In earlier simulations without allowing the neurons to settle between each set of transforms (not shown here), the multiple cell information measure was found to drop with further training. This was caused by the association of the two stimuli together since they are presented consecutively in time during training with long synaptic time constants, so the last transform of the first stimulus was still active as the first transform of the second stimulus was presented, thereby leading to their association.



(a) Before Training (b) After Training

Figure 3.11: Raster plots of output layer cells before and after training from the trace baseline simulation. Transforms 1–10 of Stimulus 1 were presented during 0–2500 ms, followed by transforms 1–10 of Stimulus 2 presented during 2500–5000 ms. Before training (a), the output cells respond to random subsets of transforms of each stimulus. After training (b), some cells are sensitive to most or all transforms of Stimulus 1 (0–2500 ms), while other output cells are sensitive to most or all transforms of Stimulus 2 (2500–5000 ms).



(a) Single Cell Information (b) Multiple Cell Information

Figure 3.12: Information plots for the baseline trace learning demonstration of translation-invariant representations for varying levels of training: untrained, 2 epochs, 10 epochs and 50 epochs. The single cell information analysis (a) shows fewer maximally informative cells than in the CT simulations but the multiple cell information plot (b) confirms that both stimuli are represented by cells which are exclusively tuned to one stimulus or the other.

3.3.2.2 Temporal Specificity of STDP

Lengthening the synaptic conductance time constant, τ_{EE}, may affect the dynamics of synaptic plasticity in unforeseen ways, so it was important to explore a range of values for the plasticity time constants as for the first set of CT simulations. As before, $\tau_C = 15\,ms$ and $\tau_D = 25\,ms$ were used as standard for the learning time constants but here they are shortened and lengthened by a factor of five (keeping the same $3/5$ ratio) for comparison (while τ_{EE} remains fixed at $150\,ms$).

The results are shown in the information plots of Figure 3.13 and the synaptic weight distributions of Figure 3.14. In contrast to the previous CT simulations, the network performance degrades with more temporally specific (shorter) STDP time constants (Figure 3.13) but improves with longer, less specific STDP time constants (the reverse trend). Similarly, this opposite trend is borne out by the synaptic weight distributions (Figure 3.14) exhibiting a smoother profile (indicating less useful training) for short STDP time constants and a more peaked profile for longer, less temporally specific STDP time constants.

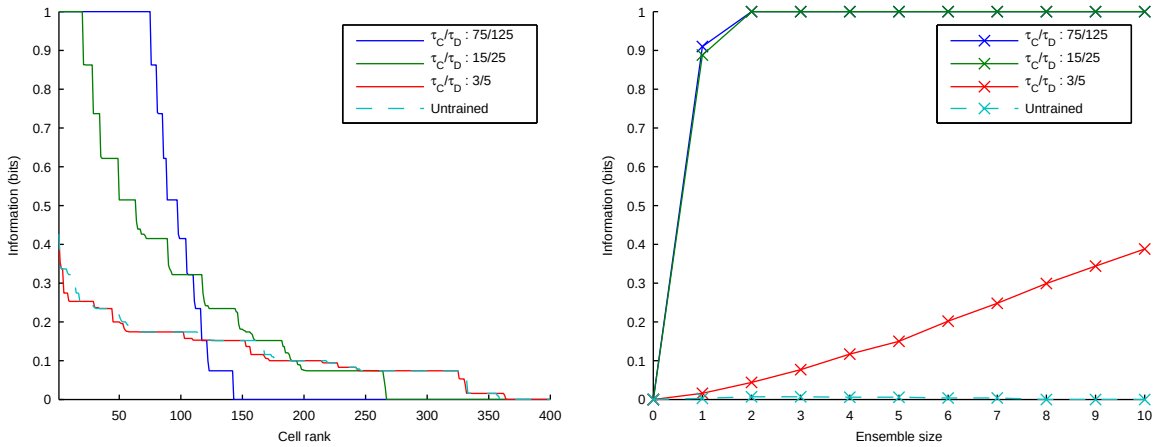
This reverse effect may be understood in the context of the two learning mechanisms whereby CT learning performs best with tightly synchronized, temporally-specific causal spike volleys, hence a temporally specific form of STDP is most appropriate. In contrast, trace learning requires activity to continue over an extended period of time between different transforms in order to associate them together, and as such the relationship it needs to capture is less temporally specific and thus a less specific form of STDP is better suited for this purpose.

3.3.2.3 Lateral Inhibition and Synchrony

As the level of inhibition is reduced, and the effects of timing jitter from the cellular membrane potential noise become more prominent, the new input layer neurons from successive transforms no longer fire in phase with those neurons from previous transforms and the information content of the output layer (Figure 3.15) can be seen to be reduced.

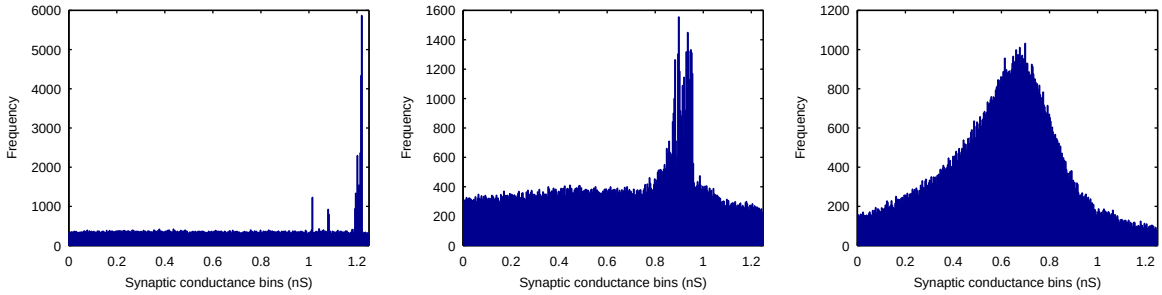
3.3.2.4 Interleaved Transforms

By interleaving transforms of the two stimuli alternately through time, transforms from different stimuli should be associated together by their temporal continuity with a trace-like mechanism. Unlike in the previous CT simulations (where the association is not time-dependent, only similarity dependent), this inter-stimulus association should



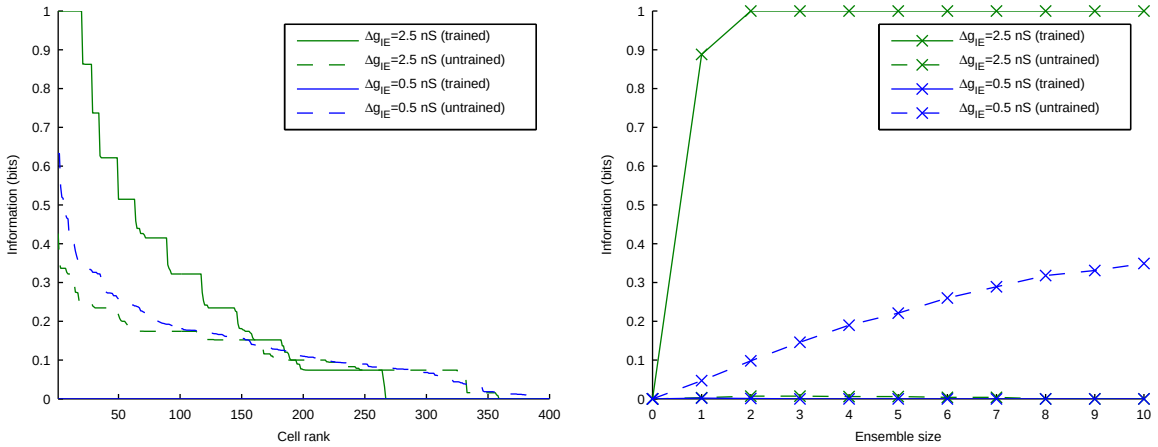
(a) Single Cell Information (b) Multiple Cell Information

Figure 3.13: Information plots of varying STDP time constants, τ_C and τ_D (trace learning). Both single cell (a) and multiple cell (b) information measures show an increase in network performance with longer, less temporally specific STDP time constants and a decrease in performance with shorter STDP time constants. This is the reverse of the trend found in the equivalent CT simulations.



(a) $\tau_C = 75\,ms, \tau_D = 125\,ms$ (b) $\tau_C = 15\,ms, \tau_D = 25\,ms$ (c) $\tau_C = 3\,ms, \tau_D = 5\,ms$

Figure 3.14: Synaptic Weight Distributions of varying STDP time constants (trace learning). In contrast to the standard case $\{\tau_C = 15\,ms, \tau_D = 25\,ms\}$ (b), the more peaked distribution for longer STDP time constants $\{\tau_C = 75\,ms, \tau_D = 125\,ms\}$ indicates more useful learning under this regime (a). Contrary to the equivalent CT simulations, the use of shorter STDP time constants $\{\tau_C = 3\,ms, \tau_D = 5\,ms\}$ (c) yields a smoother, less trained profile.



(a) Single Cell Information (b) Multiple Cell Information

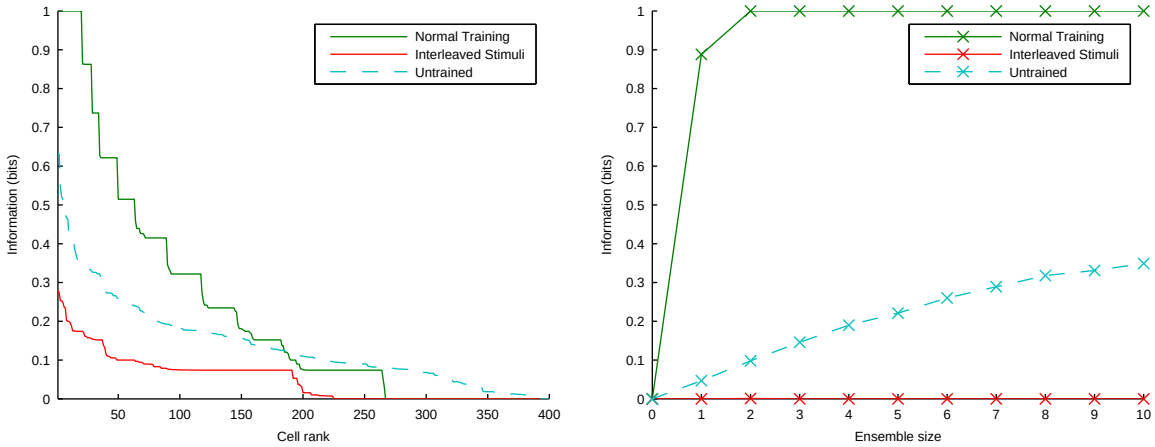
Figure 3.15: Information plots for different degrees of inhibitory strength (trace learning). As the inhibitory strength decreases the information in the network declines, exemplified in both single (a) and multiple cell (b) information measures.

lead to a large drop in information since the network will be unable to distinguish between the two stimuli. The neurons were not allowed to settle between presentations of different stimuli (as with previous trace simulations) as this would negate the effect of interleaving the stimuli and undermine the purpose of this section of simulations.

From Figure 3.16 it is evident that interleaving the transforms of the two stimuli has significantly reduced the information content of the network as expected. In the interleaved case, the single-cell information content (Figure 3.16a) has dropped to a poorer level than the untrained case (tested with a random uniform distribution of synaptic weights) as transforms from each stimuli have been associated together, meaning the output cells are less able to discriminate between stimuli than in their initial untrained, random state. From the multiple-cell information plot (Figure 3.16b) it is clear that virtually all transforms of all stimuli have become associated together since even using the ten best single-cell information neurons barely raises the multiple cell information measure above 0 *bits* since the cells are unable to discriminate one stimulus from the other.

3.3.2.5 Randomized Transform Order

If the network is using a temporal trace to associate orthogonal transforms together, randomizing the order of those transforms within a stimulus block, but still presenting all transforms of one stimulus followed by all transforms of the other, should not significantly degrade its performance. Moreover, there is good reason to expect that

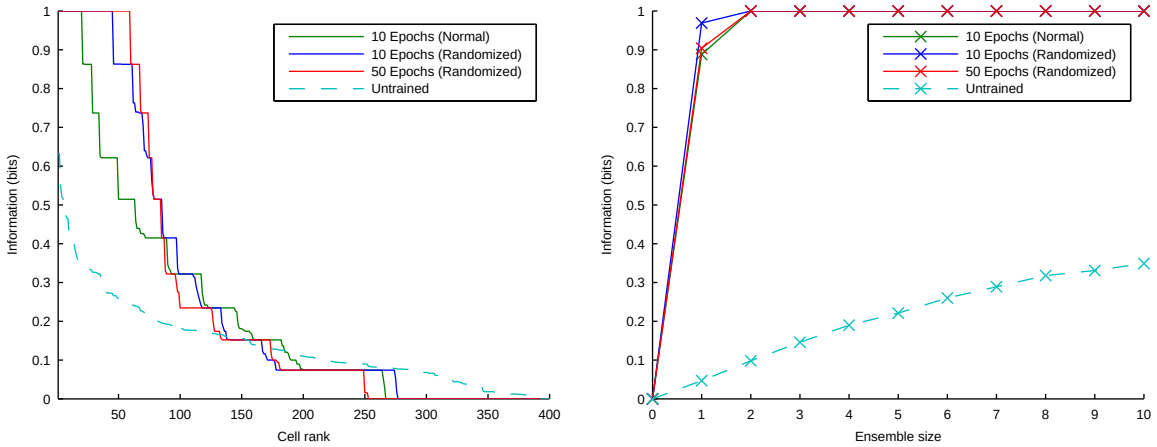


(a) Single Cell Information (b) Multiple Cell Information

Figure 3.16: Information plots showing the difference in training the network with transforms of each stimulus presented consecutively and by interleaving transforms of both stimuli (trace learning). By interleaving the collections of transforms, the single cell (a) and multiple cell (b) information measures indicate that performance has dropped to lower levels than obtained with a random (untrained) network. This is a typical property of Trace learning and is in marked contrast to the equivalent CT learning results with short synaptic time constants.

the performance should be improved slightly, as this training paradigm will help to associate each transform, $S_1^{t_n}$, with each other from the same stimulus rather than just its neighbouring transforms, $S_1^{t_{n-1}}$ and $S_1^{t_{n+1}}$.

It is clear from Figure 3.17 that with the longer synaptic conductance time course ($\tau_{EE} = 150\,ms$) and the same degree of training, the randomized transform case has performed better than the standard non-randomized paradigm as expected, improving further with more epochs of training. In previous simulations this training paradigm proved difficult for the CT mechanism with an initially random set of feed-forward weights, since different pools of output neurons were stimulated by randomly ordered transforms (due to having less spatial overlap between consecutive transforms on average). In the case of randomly ordered transforms with the trace learning mechanism however, the lack of spatial overlap between transforms is irrelevant as the same pool of output neurons is kept active for all the transforms of a particular stimulus, by virtue of the longer synaptic time constants which allows current to continue bleeding into the postsynaptic cell, and the consecutive presentation of all transforms of a particular stimulus (albeit not necessarily in order).



(a) Single Cell Information

(b) Multiple Cell Information

Figure 3.17: Information plots showing the difference in training the network by presenting the transforms sequentially and by randomizing the order of transforms (trace learning). It is evident from the single cell (a) and multiple cell (b) information measures that trace learning exhibits better performance with randomized transformation order. This is because under the randomized training regime, there is greater scope for associations to be made between more pairs of transforms (rather than each with just its sequential neighbours) over the course of many training epochs.

3.4 Discussion

In the above simulations we have shown that a biologically realistic spiking neural network with spike-time dependent plasticity can operate in two very different ways to achieve transformation-invariant representations. These simulations lend more biological plausibility to the Trace and CT learning mechanisms, which may be utilized by the same model with slight differences in the training environment or the physical parameters of the neurons.

With short synaptic conductance time constants in the feed-forward connections between the pyramidal neurons (τ_{EE}) in successive layers, the model works similarly to the CT learning mechanism. In this case, the network requires the transforms to be spatially overlapping (as a direct consequence of the learning mechanism) but can cope with interleaving the transforms of different stimuli and thus bears the characteristics of the equivalent rate-coded mechanism (Stringer et al., 2006). Importantly, this mechanism is sensitive to the strength of lateral inhibition, which under optimal conditions serves to maintain the synchronous firing of neurons representing the novel part of an unseen transform with those already potentiated from previous learning of another transform. Without this effect of lateral inhibition, these novel neurons will

most likely fire outside the time window for significant LTP, and may possibly come after the postsynaptic neuron has fired leading to LTD.

Lengthening the very same synaptic conductance time constants (τ_{EE}), enables the model to work with a Trace learning mechanism. In this case the network uses temporal continuity to associate together orthogonal (completely non-overlapping) transforms and consequently fails to develop invariance and stimulus specificity if the transforms of different stimuli are interleaved. While these properties are the same as for the classic McCulloch–Pitts neuron, it is interesting to note that in such a rate-coded model, the trace term is associated with the presynaptic or (more commonly) postsynaptic *neuron* (Rolls and Milward, 2000). In contrast, in a conductance-based spiking neural network, the trace can instead be associated with the individual *synapses* between two connected neurons. This is a measurable property of biological neurons and suggests where to focus neurophysiological investigation aiming to understand invariance learning mechanisms.

While the Trace and CT learning mechanisms have been studied here in isolation, it seems likely that a combination of both would be employed to varying degrees depending upon the statistics of the inputs to each layer of the brain. In early layers (e.g. V1), the patterns of stimulation are likely to change more from one transform to another since the neurons here are highly specific in their sensitivity to a location and orientation. In later layers however, such as Inferotemporal cortex (IT), the invariance built in the earlier layers will mean that inputs to these cells are less changeable from one transform to another. Having passed through several layers of pyramidal cells with lateral inhibition acting at each stage, the spike volleys representing a stimulus may also become more synchronized (Diesmann et al., 1999). Under these conditions, we therefore expect that as the similarity between transforms increases through the layers, the CT mechanism will become more prominent and trace effects will become less important, which would be evidenced by progressively quicker synaptic conductance decays (shorter time constants).

If it is the case that the ventral visual system uses an effective synaptic time constant between the two extremes presented in the simulations here, we would therefore predict that the type of learning occurring for any given stimuli would be highly dependent on how those stimuli are presented, for example with rapidly transforming (and hence spatially dissimilar successive views) leading to more of a Trace learning regime, whereas temporally separate exposures would require a high degree of similarity between the views for the CT mechanism to work.

The work presented here is a first step towards understanding how the Trace and CT learning rules may be utilized in a spiking neural network, and as such will naturally have limitations. So far, the model has been presented with orthogonal, non-overlapping 'toy' stimuli rather than the more distributed, spatially overlapping stimuli found in the natural visual world. Whilst we acknowledge that these highly idealized representations are somewhat lacking in ecological validity, they were employed in order to isolate each learning mechanism in a precise and identifiable way. Further work would benefit however from exploring these learning mechanisms with more natural, spatially overlapping stimuli.

A further limitation concerns the Trace learning mechanism. By lengthening the time constant of the feed-forward synaptic conductances, τ_{EE}, the excitatory activity reaching the output neurons decays more slowly and results in much higher firing rates in the output neurons (approximately 200 spikes/s) than in the CT simulations (approximately 50 spikes/s). While these rates are still within the realms of biological plausibility, they are towards the edge of it and so the conclusions would be on firmer ground through exploring additional mechanisms to reduce these high firing rates.

3.4.1 Future Directions

In understanding the dynamics of learning transformation-invariant representations in spiking neural networks, we have only demonstrated *translation* invariance so far. A natural extension to this body of work would therefore be to investigate this learning process with other kinds of transforms commonly found in natural visual scenes and investigated in rate-coded models including, for example, rotations (Stringer et al., 2006), occlusions (Stringer and Rolls, 2000) and changes in scale (Wallis and Rolls, 1997). This would provide a more general understanding of the variations in the problems of visual object recognition that the visual system must overcome.

Furthermore, the use of realistic 3-D shapes and faces will also allow the model to be more directly compared to psychophysical data, both in terms of the effects on representations formed from exposure to realistic images (David et al., 2004; Felsen and Dan, 2005; Felsen et al., 2005; Simoncelli, 2003) and testing if invariance learning may be achieved at natural speeds of transformation (e.g. rotation). Neuronal parameters such as the synaptic time constants (e.g. τ_{EE}) and the learning time constants (τ_C and τ_D) may be crucial to invariance learning with realistic stimuli. Exploring the interaction between the speed of transformation of objects and the parameters of the model should lead to concrete predictions which may be tested against neurophysiological data. For example this may reveal an upper-threshold of

stimulus movement speed which still allows transformation-invariant representations to form, or even that our visual systems typically use a number of static views to learn invariance.

Natural stimuli will also test the model's ability to learn transformation-invariant representations with effectively distributed, overlapping representations rather than the orthogonal non-overlapping representations employed so far. This would mean that the network could no longer appear to solve the problem through learning about retinal location.

In addition to enhancing the ecological validity of the stimuli and their presentation paradigm, the model itself could be modified to incorporate additional features found in its biological counterpart including lateral excitatory connectivity, cell firing-rate adaptation and multiple layers of feed-forward weights, some or all of which may prove to be necessary for solving the more complex invariance learning problems, for instance, with natural scenes composed of multiple objects.

3.4.2 Conclusion

In the work presented here, we have demonstrated how a spiking neural network may exhibit two very different modes of invariance learning, which share the characteristic properties of their rate-coded counterparts. This was achieved in a single model by changing, (most notably), the time constant of the feed-forward synaptic conductances and the properties of the stimulus sets. Through developing more biologically accurate spiking models in this way, we may build upon insights from previous work to more fully understand the detailed mechanisms of visual invariance learning in the brain.

The critical act in formulating computational theories turns out to be the discovery of valid constraints on the way the world is structured – constraints that provide sufficient information to allow the processing to succeed.

—David C. Marr

4 Realistic Stimuli

4.1 Introduction

Idealized abstract stimuli are a useful and precise tool for understanding the principles of operation in a model of the ventral visual stream in detail. However, the same principles should be shown to be applicable in a scaled-up model using images of realistic three-dimensional objects as inputs. Following directly on from Chapter 3, this chapter aims to demonstrate that the conclusions drawn from using abstract stimuli are equally valid when using more realistic images of a cone and a cube as stimuli translating across the input layer.

In addition, the training paradigm utilizing abstract stimuli used throughout this thesis is limited to investigations of translation. One principled reason for using realistic stimuli therefore, is the ability to investigate other forms of transformation, and demonstrate that the principles also hold for those. Accordingly, the work presented in this chapter also features a demonstration of the same stimuli being rotated in depth as in many of the studies with VisNet (Wallis and Rolls, 1997).

In order to present realistic stimuli to the network, the input layer is converted to model the cell responses found in V1. As a first approximation, V1 cell response properties were characterized as 'edge' and 'bar' detectors through experimentation with the cat (Hubel and Wiesel, 1959) and macaque (Hubel and Wiesel, 1968) primary visual cortex. These cells were further categorized according to their preference for oriented lines within a particular region of the visual field as 'Simple' (location

specific), 'Complex' (invariant to location forming the majority of cells in V1) and 'Hypercomplex' cells (which are sensitive to end-stopped oriented bars, i.e. length, although these are now regarded as subclasses of simple and complex cells). Further examination of the receptive fields of V1 cells has revealed that their preferred stimuli are actually oriented sinusoidal gratings and as such, their receptive fields are now commonly modelled as Gabor functions (Jones and Palmer, 1987).

These receptive field models are described in detail in the Methods section before the presentation of the results from simulations of filtered images translating and rotating in depth are plotted and analysed. The results are then discussed before returning to the idealized and computationally much cheaper stimuli in subsequent chapters.

4.2 Methods

The model parameters and architecture are similar to that described for the simulations in Chapter 3. The principle difference however, is in the input layer, which has been scaled up to represent a 32×32 two-dimensional image with a column of 16 feature detectors modelled by Gabor filters at each location. The input layer thus consists of over $20,000$ neurons (including excitatory and inhibitory neurons in a ratio of $4:1$) making it prohibitively expensive except for the demonstrations of invariance learning with realistic stimuli provided in this chapter.

4.2.1 Oriented Gabor Filters

A Gabor function is an oriented sinusoidal grating within (convolved with) a Gaussian envelope. An example Gabor function (along with its component functions) is illustrated by Figure 4.1 in one dimension.

Mathematically, a Gabor function is described as follows:

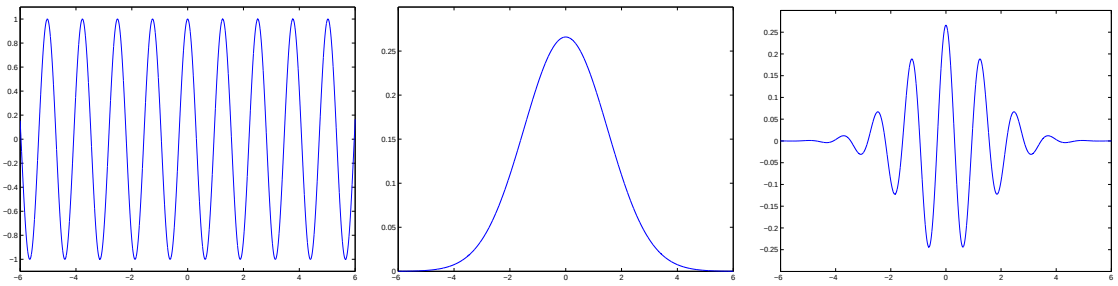
$$g_{\lambda,\theta,\phi,\sigma,\gamma}(x,y) = \exp\left(-\frac{x_\theta^2 + \gamma^2 y_\theta^2}{2\sigma^2}\right)\cos\left(\frac{2\pi x_\theta}{\lambda} + \phi\right) \tag{4.1}$$

with the following definitions:

$$x_\theta = x\cos\theta + y\sin\theta \tag{4.2}$$

$$y_\theta = -x\sin\theta + y\cos\theta \tag{4.3}$$

where x and y specify the position of a light impulse in the visual field (Petkov and Kruizinga, 1997).



(a) Sinusoidal (b) Gaussian (c) Gabor function

Figure 4.1: The sinusoidal, $y = \sin\left(5 \times \left(x + \frac{\pi}{2}\right)\right)$, and Gaussian, $\mathcal{N}(0, \frac{9}{4})$, components of a one-dimensional Gabor function.

The standard deviation of the Gaussian factor, σ, is controlled indirectly through λ, and b as follows:

$$\frac{\sigma}{\lambda} = \frac{1}{\pi}\sqrt{\frac{\ln 2}{2}} \cdot \frac{2^b + 1}{2^b - 1} \tag{4.4}$$

The ellipticity of the Gaussian factor is determined by γ, the spatial aspect ratio and is found in the range [0.23,0.92] (Petkov and Kruizinga, 1997). In an alternative formulation, the standard deviation may be split into independent dimensions where $\sigma_x \equiv \sigma$ and γ is absorbed into σ_y such that $\sigma_y = {}^{\sigma_x}/_{\gamma}$. The exponent is therefore expressed as $-\frac{1}{2}\left(\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2}\right)$ in this alternative description.

The parameter λ is the wavelength (meaning that ${}^1/_{\lambda}$ is the spatial frequency) of the (co)sine factor. The ratio ${}^{\sigma}/_{\lambda}$ determines the half-response spatial frequency bandwidth of simple cells and thus the number of parallel excitatory and inhibitory stripes in their receptive fields. Typically, three to five such parallel excitatory and inhibitory zones are found in the receptive field of a simple cell (Petkov and Kruizinga, 1997). The bandwidth is measured in octaves, meaning that a unit increase means that there are double the number of excitatory and inhibitory regions within the receptive field and is equivalently related to the ratio ${}^{\sigma}/_{\lambda}$ as follows:

$$b = \log_2 \frac{\frac{\sigma}{\lambda} + \frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}}{\frac{\sigma}{\lambda} - \frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}} \tag{4.5}$$

Through neurophysiological studies of the cat, b has been found to be in the range [0.5,2.5] octaves (weighted mean 1.32 octaves) and [0.4,2.6] octaves in the macaque (median 1.4 octaves). Despite the spread, most cells are in the range [1.0,1.8] octaves (Petkov and Kruizinga, 1997). Some propose a sharpening of the spatial frequency bandwidth through the visual system down to one octave.

The parameter θ specifies the orientation of the normal to the parallel zones, in other words, the orientation of the bar or edge to which the Gabor filter is tuned.

Lastly, ϕ is the phase offset for the (co)sine factor (relative to the centre of the Gaussian) and concerns the symmetry of the Gabor function, being symmetric at $\phi = 0$ (on-centre) and $\phi = \pi$ (off-centre). At $\pm\pi/2$ the function is antisymmetric or odd and all other cases are asymmetric mixtures of the two. Symmetric phases therefore yield the response properties of 'bar-detector' simple cells, whereas asymmetric phases model 'edge-detectors'.

4.2.2 Stimulus Preparation

The first step in image-processing is to resize the image to be 32×32 pixels and embed it in the centre of a 64×64 grey background. The significance of this step is that when the image is filtered and cropped back to 32×32 in accordance with the size of the input layer, there will not be any artefacts from the filters detecting the edge of the stimulus borders.

If the image is colour, it is next converted to monochrome using the MATLAB[1] function `rgb2gray` which forms a weighted sum of the Red, Green, and Blue components in accordance with sensitivity to the human eye as follows:

$$I = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B \tag{4.6}$$

It is common practice to next subtract the mean intensity from the image (Pinto et al., 2008), however this is unnecessary since the Gabor filters are generated in such a way as to sum to zero. Stimuli are instead next divided by their standard deviation.

An array of Gabor filters are then generated with the parameters given in Table 4.1. These filters then have their mean set to 0 and their Euclidean norm set to 1. These Gabor filters are plotted in Figure 4.2 for a wavelength of $8\,pixels/cycle$. The other set of filters with a wavelength of $4\,pixels/cycle$ are not shown as they are identical in every respect but their size.

The MATLAB function `conv2` is then used to convolve the image with each filter in two dimensions (convolution in the spatial frequency domain is equivalent to multiplication in the spatial domain), after which the filtered stimulus is cropped back to its original size. Each filter output is then rescaled by dividing by the convolution of the filter with itself in order to bound the filter output values in the range $[-1, 1]$. Finally, each filter output is divided by the standard deviation of all the filter outputs so that the vector of filter outputs has unit variance.

[1]The Mathworks Inc., Natick, Massachusetts.

Parameter	**Symbol**	**Value**
Phase shift	ψ	$\{-\frac{\pi}{2}, \frac{\pi}{2}\}$ *radians*
Wavelength	λ	$\{4, 8\}$ *pixels/cycle*
Orientation	θ	$\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ *radians*
Spatial bandwidth	b	1 *octaves*
Aspect ratio	γ	0.5

Table 4.1: Typical parameters used for constructing sets of Gabor filters.

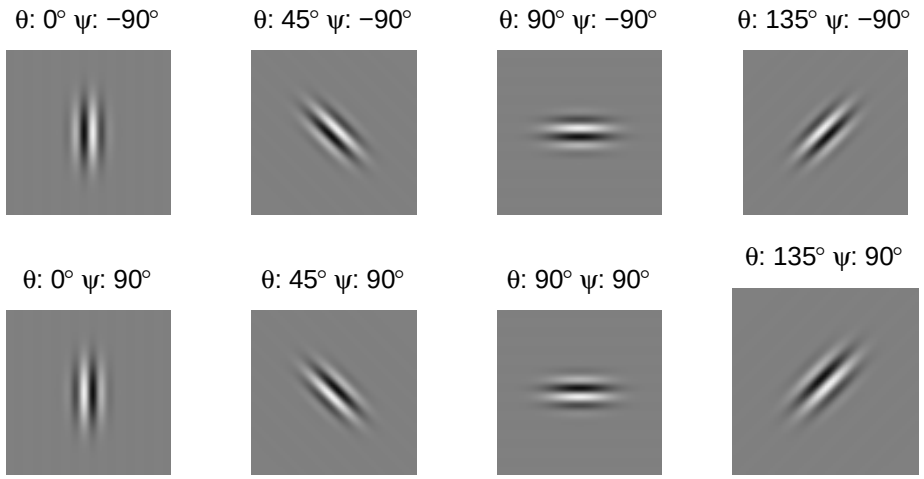


Figure 4.2: The Gabor filters used in stimulus preparation plotted for one scale ($\lambda = 8\,pixels/cycle$).

4.2.3 Network Architecture

The network consisted of two layers of excitatory LIF neurons with a layer of feed-forward plastic synapses between them trained with the STDP learning rule described in Section 2.3.1 (Perrinet et al., 2001). Each layer's excitatory cells are reciprocally connected to their own pool of inhibitory interneurons in a ratio of 4 : 1 through non-plastic synapses. To reduce the simulation run-time further, the probability of each inhibitory neuron making a presynaptic connection with each excitatory neuron in the input layer was reduced to 0.1. Given that there were so many excitatory neurons in the input layer to drive the inhibitory interneurons, this did not appear to affect the dynamics and the network was otherwise fully connected.

To accommodate the 16 Gabor filters at each of the 32×32 input layer locations, there were $16,384$ excitatory neurons in the input layer and the output layer size was also increased to 256 excitatory neurons. Correspondingly, there were 4096 and 64 inhibitory neurons in the input and output layers respectively.

The scaling factor for the strength of the feed-forward excitatory conductances was tuned to $0.875\,nS$ and the general model parameters are otherwise as described in Table 3.1.

Initially an excitatory synaptic time constant τ_g^{EfE} of $2\,ms$ was used to eliminate the Trace learning mechanism and study CT learning in isolation but after unsuccessful preliminary simulations, the longer time constant of $150\,ms$ was used to switch to a Trace learning mechanism. However, the stimuli were also transformed (i.e. shifted or rotated) continuously in order to allow the CT learning mechanism to contribute to invariance learning.

4.2.4 Stimulus Presentation

Through many preliminary simulations, the filtered outputs were scaled with a mean and standard deviation of the current both equal to $0.1nA$, which was then injected into the cell bodies of the input layer excitatory neurons.

Each transform of each stimulus was presented for $100\,ms$ in sequence during training in a random direction for a total of ten epochs. During testing, plasticity in the feed-forward connections was deactivated and the transforms of each stimulus were presented sequentially for $250\,ms$.

For these simulations, a relatively large time-step of $\Delta t = 0.2\,ms$ had to used in order to make the simulation run-times tractable. This was not considered to be too problematic though, since the time constant of the excitatory synaptic conductance was set to encourage the model to operate using the Trace learning mechanism which was found to make the behaviour less sensitive to the precise timing of action potentials.

4.3 Results

4.3.1 Translating Stimuli

For the translation simulations, a set of two stimuli are used, each of which is presented in nine locations, translating across the input layer in an equivalent way to the simple stimuli of Chapter 3. The stimulus set for these simulations is shown in Figure 4.3.

To illustrate the effect of the filtering process, the central transform (T_5) of the cone is filtered and plotted in Figure 4.4.

The network was trained in the way described in Chapter 3 and the output layer (excitatory) raster plots before and after training are given in Figure 4.5. It can be seen that similarly to the abstract stimuli of Chapter 3 (for example, Figure 3.11), before

Figure 4.3: The complete translating stimulus set of the cone and cube comprising of nine locations in a horizontal plane for each stimulus.

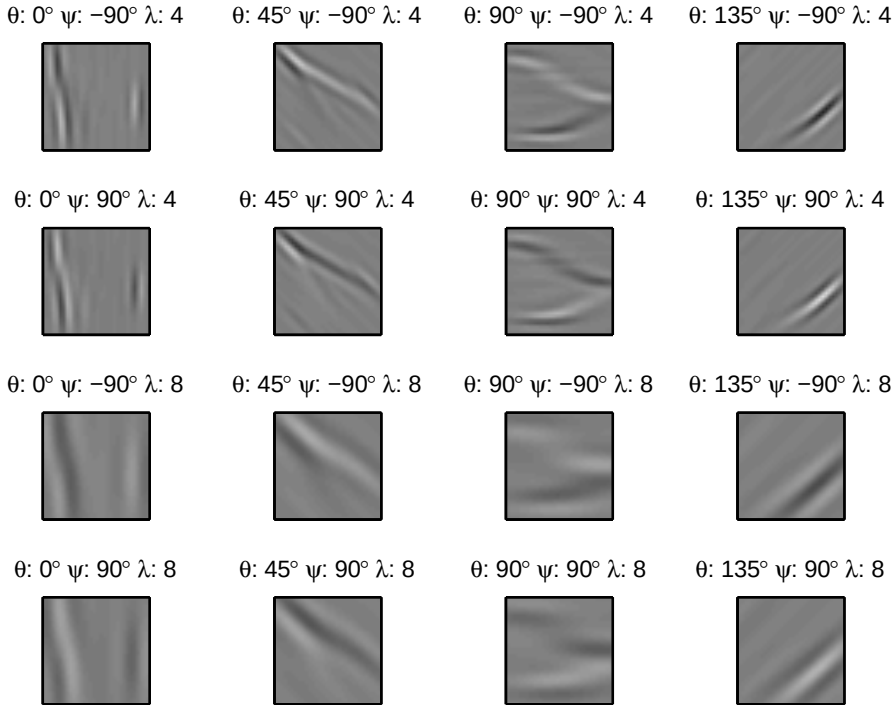
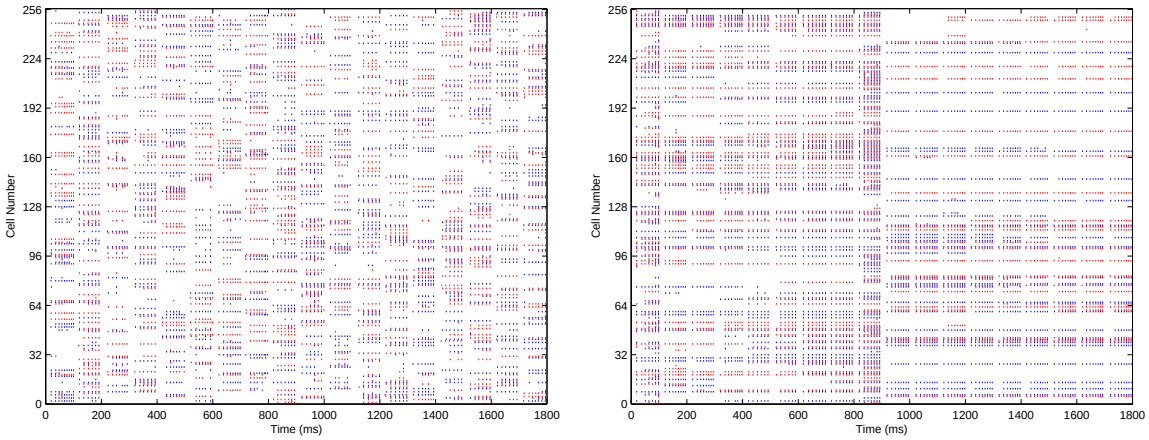


Figure 4.4: The complete filter outputs for the central transform (T_5) of the cone.

training the output neurons respond randomly to either stimulus in various locations (Figure 4.5a). After training however (Figure 4.5b), many output neurons respond selectively to either the cone or the cube across all of their locations, indicating that the model has succeeded in developing transformation-invariant representations from realistic stimuli.

This is confirmed by the information analyses of the output layer plotted in Figure 4.6. Before training, both measures indicate poor information content in the output layer. After training however, the single cell analysis (Figure 4.6a) shows that approximately 60 neurons have achieved the full 1 *bit* of information. The multiple



(a) Before Training (b) After Training

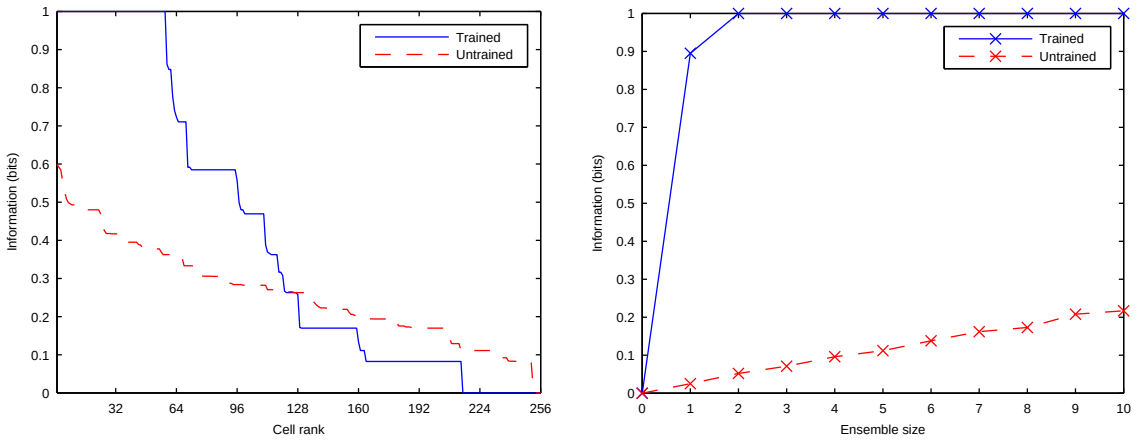
Figure 4.5: Raster plots of the output layer principal cells before (a) and after training (b) with the translating realistic stimuli. The cone (Stimulus 1) is presented in each of the 9 locations (transforms) consecutively from 0–900 ms, followed by the 9 locations of the cube (Stimulus 2) during 900–1800 ms. The neurons respond randomly to both stimuli in various locations before training (a) but after training (b), many can be seen to respond specifically to either the cone or the cube, invariantly across each of their 9 locations.

cell information plot (Figure 4.6b) confirms that both stimuli are represented in the output layer, attaining the maximum amount of information with an ensemble size of 2 neurons or more.

To assess the relative contributions of Trace and CT learning in achieving these invariance results, the simulation was repeated with the same random seed but interleaving the transforms of the two stimuli as in Chapter 3. If the effect is purely due to the Trace learning mechanism, this training paradigm should greatly reduce the network performance.

The output layer raster plot after learning is shown in Figure 4.7, however the pre-training raster is not plotted since it is identical to the raster shown in Figure 4.5a. In contrast to the earlier post-training raster plot where the transforms of each stimulus were presented consecutively (Figure 4.5b) it appears that many output neurons respond to transforms from both stimuli suggesting a decrease in network performance.

The information analyses for the interleaved stimulus paradigm are presented in Figure 4.8, along with those from the standard training regime for ease of comparison. The number of stimuli attaining the maximal single cell information content has significantly decreased (Figure 4.8a) corroborating with the inferences drawn from



(a) Single Cell Information (b) Multiple Cell Information

Figure 4.6: Information plots for Single and Multiple cell analysis before and after training with the translating realistic stimuli. The single cell analysis (a) shows that approximately 60 output neurons have attained the maximal information content of 1 *bit*. The multiple cell analysis (b) confirms that both stimuli are represented among the most informative neurons across all of their locations (transforms), as the information limit is reached with an ensemble size of just 2 neurons.

the raster plot. Correspondingly, the multiple cell information measure (Figure 4.8b) also indicates a decrease in performance compared to the standard training paradigm of presenting stimuli consecutively. Interestingly, both the single and multiple cell information measures for the interleaved case still outperform the random (untrained) network, implying that there may be sufficient overlap between the transforms of each stimulus for the CT learning mechanism to operate and help to build some translation invariance in the output neurons. Therefore, both Trace and CT learning appear to be contributing to invariance learning in the first simulation reported in this chapter.

4.3.2 Rotating Stimuli

In this section, the same stimuli are used but rotating in depth through 8 views, each 45° apart to extend the work to encompass another form of transformation (in addition to translations) in order to demonstrate that the same principles apply. The more challenging *class-specific* transformations such as rotations (as opposed to *generic* transformations such as translations) can only be done with the realistic three-dimensional stimuli used in this chapter. The set of views for each stimulus are shown in Figure 4.9. The only change between this simulation and the previous section (where the cone and cube translated) is with respect to the stimuli and a corresponding

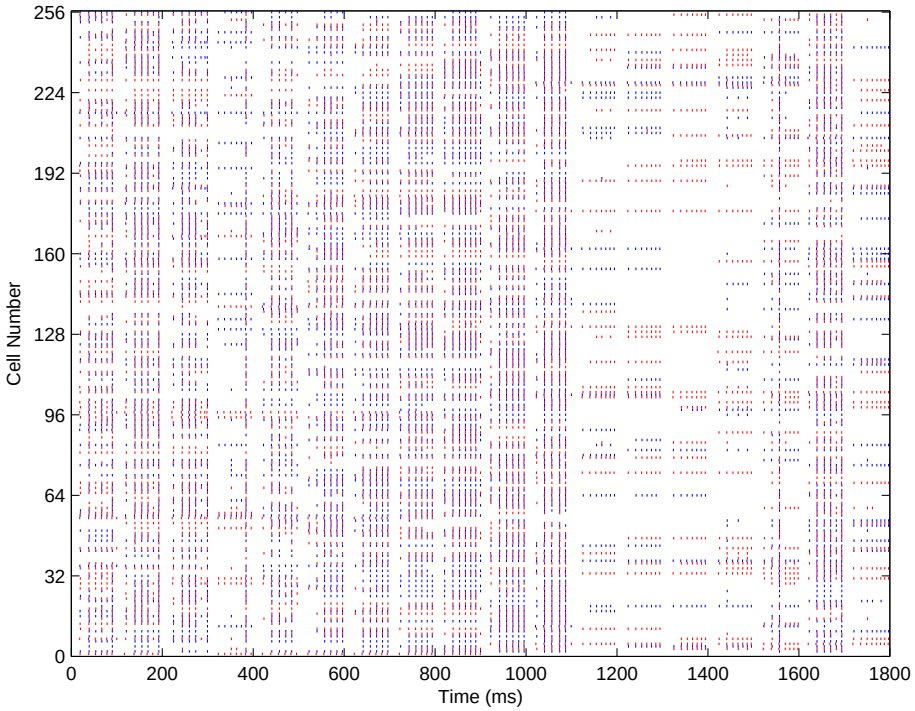
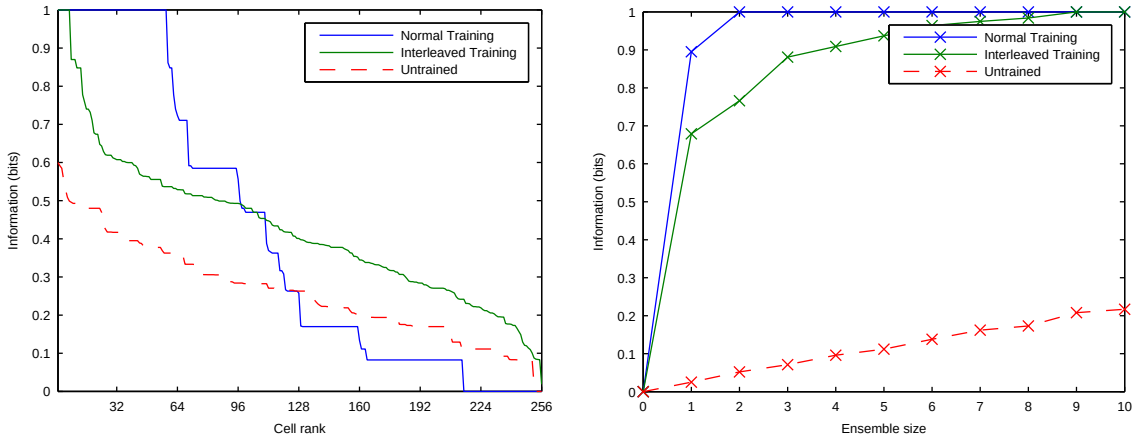


Figure 4.7: Raster plots of the output layer principal cells after training with interleaving transforms of the translating realistic stimuli. The cone (Stimulus 1) is presented in each of the 9 locations (transforms) consecutively from 0–900 ms, followed by the 9 locations of the cube (Stimulus 2) during 900–1800 ms. It can be seen that after training by interleaving the views of each stimulus, many neurons now respond to transforms of both stimuli.

decrease in the total training and testing time due to there being one fewer transform of each stimulus. All other details of the model's architecture, parameters and training regime remain unchanged.

The network was trained by presenting each view of each stimulus sequentially before saving the results for analysis. The raster plots of the output layer before and after training are shown in Figure 4.10. It appears that many cells learnt to respond invariantly to the eight views of the cone, but the cube was not so well represented.

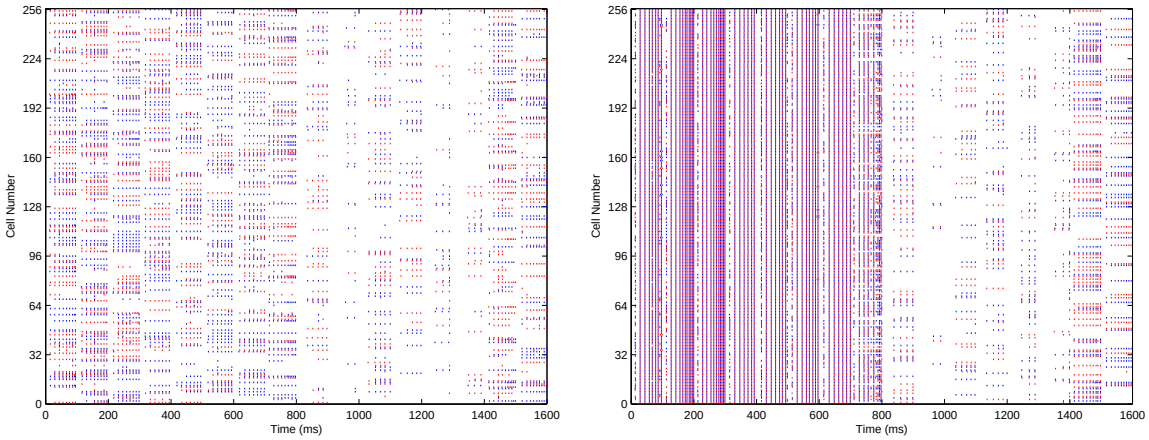
From inspecting the information analysis plot of Figure 4.11, it appears that almost 150 neurons attained the maximum single cell analysis information level of 1 bit after training (Figure 4.11a) and thus have achieved transformation-invariant recognition of one of the three-dimensional rotating stimuli. However, looking at the multiple cell information plot (Figure 4.11b), an ensemble size of 4 neurons was required to reach the full information limit of 1 bit, confirming the inference drawn from the raster plot that cube is under-represented and the output neurons have only achieved partial



(a) Single Cell Information (b) Multiple Cell Information

Figure 4.8: Information plots for Single and Multiple cell analysis before and after training with interleaving transforms of the translating realistic stimuli. The single cell analysis (a) shows only a handful of output neurons have attained the maximal information content of 1 *bit* (compared to approximately 60 with the normal training paradigm), representing a significant degradation of performance. The multiple cell analysis (b) also shows a significant degradation of performance for the interleaved presentation paradigm compared to the standard case, how the full 1 *bit* of information is still achieved with an ensemble of 9 neurons.

Figure 4.9: The complete rotating stimulus set of the cone and cube, comprising of eight views (45° apart) of each stimulus.



(a) Before Training (b) After Training

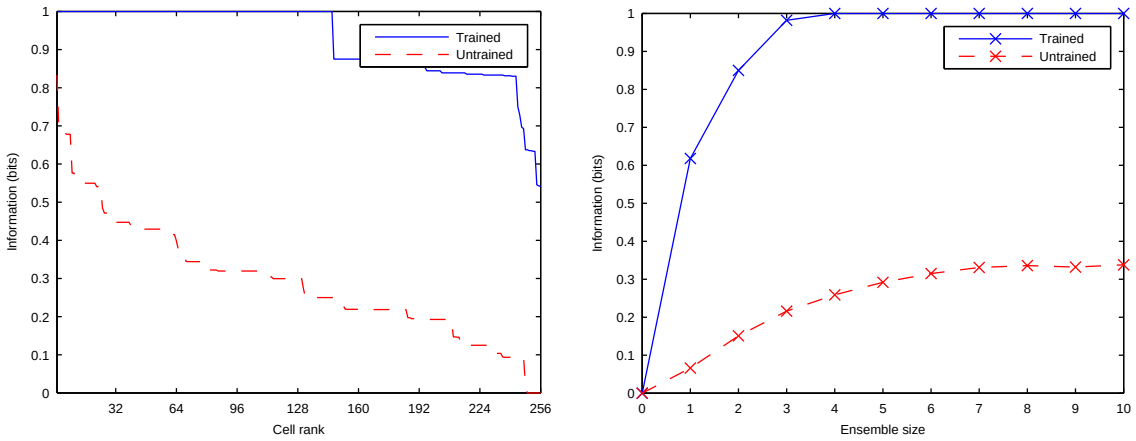
Figure 4.10: Raster plots of the output layer principal cells before (a) and after training (b) with the rotating realistic stimuli. The cone (Stimulus 1) is presented in each of the 8 views (transforms) consecutively from 0–800 *ms*, followed by the 8 views of the cube (Stimulus 2) during 800–1600 *ms*. The neurons respond randomly to both stimuli in various locations before training (a). After training (b), many neurons can be seen to respond specifically to the cone (Stimulus 1) invariantly across each of its 8 views although the cube (Stimulus 2) appears to under-represented.

invariance over some but not all views.

The most likely reason for this is attributable to some views of the cube 'vanishing' against the mid-grey background when the view is such that the lighting renders most of the visible surfaces at a similar luminance level to the background (see for example Figure 4.9, Stimulus 2, Transform 7). The consequence of this is that the Gabor filters do not successfully detect the borders (and other features) of the shapes and so do not generate much activity in the input layer to represent that particular view.

4.4 Discussion

Initially the translation simulations were attempted with short synaptic conductance time constants in accordance with a CT paradigm. The expectation was that the width of filters would provide a degree of overlap for the operation of the CT learning mechanism. In particular, horizontal edges were expected to provide the spatial overlap required by the CT learning mechanism since these filters would be elongated and aligned in the direction of the translations. From inspection of the input layer rasters however, it was confirmed that there was very little overlap between transforms. After considerable time spent adjusting the filters and parameter searching, attempts



(a) Single Cell Information (b) Multiple Cell Information

Figure 4.11: Information plots for Single and Multiple cell analysis before and after training with the rotating realistic stimuli. The single cell analysis (a) shows that approximately 150 output neurons have attained the maximal information content of 1 *bit*. The multiple cell analysis (b) confirms that both stimuli are represented among the most informative neurons across all of their locations (transforms), as the information limit is reached with an ensemble size of just 4 neurons.

to make the model work in a purely CT learning paradigm were abandoned and the synaptic conductance time constants were lengthened to induce the Trace learning mechanism. As the results show, this proved to be more successful.

The difficulties for the CT learning mechanism may be ameliorated however in a more detailed model of the input layer. Connecting features with similar orientation preferences through lateral connections, as appears to be the case in the visual cortex (see Miikkulainen et al., 2005, p23–24), may help to produce contour integration and allow CT to work with translations.

Another issue to arise was the disappearance of some transforms against the background, when the surface of the shapes matched the luminance of the grey behind them. This is a problem for real world object recognition too, as in a sense, it is the basis upon which camouflage may work. However, the primate visual system is likely to be less prone to this sort of problem than the neural network model explored above. One reason for which, may be that primate vision is stereoscopic, whereby the two slightly different views from each eye help to ensure there is more chance that one view at least will not suffer from this problem. Binocular vision of course, also provides useful depth information further helping to separate an object from the background. This problem could also potentially be ameliorated by implementing a trichromatic model of the ventral stream where even if the luminance borders disappeared, edges

due to changes in colour would still be visible. Either of these features, namely binocular or colour vision represents a significant challenge to implement, but would provide a much more realistic model of the visual system to investigate.

In conclusion, using realistic filtered objects as stimuli has shown that the same principles for learning transformation-invariant representations may operate as when using the simpler abstract stimuli of the previous chapter. Importantly, it was also shown how other forms of transformations such as rotations may be investigated by using more realistic three-dimensional stimuli and that the insights gained from simple translating stimuli are still applicable. This model came with its own set of challenges however, including tuning the parameters of the filter set to be general purpose and robust. Perhaps the most significant challenge was the simulation runtime in the order of days, making more thorough investigations currently impractical. Nonetheless, these investigations have pointed the way to a richer model for future work, and for present purposes, have left the simulations with more abstract stimuli reported in this thesis on firmer ground.

The brain is a tissue. It is a complicated, intricately woven tissue, like nothing else we know of in the universe, but it is composed of cells, as any tissue is. They are, to be sure, highly specialized cells, but they function according to the laws that govern any other cells. Their electrical and chemical signals can be detected, recorded and interpreted and their chemicals can be identified; the connections that constitute the brains woven feltwork can be mapped. In short, the brain can be studied, just as the kidney can.

—David H. Hubel

5 Lateral Connectivity

5.1 Introduction

In the primate visual system, increasingly complex representations are developed at successively higher layers in the ventral stream hierarchy (Kobatake and Tanaka, 1994; Tanaka, 1996) until individual neurons respond to selectively to particular faces (Desimone, 1991) or objects (Tanaka et al., 1991). In a way that still eludes many artificial systems, these neurons also respond invariantly to a range of transformations of their preferred stimuli including translations (Tovée et al., 1994; Op de Beeck and Vogels, 2000; Hung et al., 2005), changes in size (Ito et al., 1995; Hung et al., 2005) and view (Booth and Rolls, 1998; Logothetis et al., 1995). Models which have sought to understand the formation of such transformation-invariant representations in Inferotemporal cortex have largely used training paradigms where stimuli are presented individually. An important question concerning this learning process therefore remains — how can the visual system become selective for *individual* objects (or faces) when it only experiences natural scenes composed of *multiple* objects?

Rate-coded models of this process suffer from the 'superposition catastrophe' whereby the coactivity of inputs representing several stimuli seen together leads to their joint representation, further exacerbated by rate-based Hebbian learning whereby the stimuli are associated onto the same (simultaneously coactive) output neurons (von der Malsburg, 1999). In order to avoid this problem, rate-coded neural network models of the visual system are commonly trained by presenting only one stimulus

at a time (Wallis and Rolls, 1997; Perry et al., 2006; Stringer et al., 2006), leaving the training paradigm lacking in ecological validity. However, recent research has uncovered a number of mechanisms for overcoming this problem.

VisNet, a model of the ventral stream consisting of a hierarchical, feed-forward series of rate-coded competitive networks (Wallis and Rolls, 1997) was presented with multiple stimuli transforming (shifting or rotating) in different combinations. It was found that if the pool of stimuli was large enough and a sufficient number combinations was presented during the learning phase, the statistical decoupling between the objects forced the competitive networks to form individual representations of the stimuli in the output layer (Stringer and Rolls, 2008; Stringer et al., 2007). However, achieving this required an extensive training regime where objects were repeatedly seen in different combinations, leaving the problem of how objects may be disentangled even when they are *always* (or very often) seen together.

Another mechanism discovered in VisNet solving the 'superposition catastrophe' of multiple object presentations was found to depend upon independent movement of the stimuli (Tromans et al., 2012). Although there were only two stimuli in each experiment (negating the possibility of statistical decoupling), presenting the objects rotating at different speeds allowed the competitive networks to similarly form transformation-invariant *separate* representations in the network's output layer. While independence of movement is typically a reasonable assumption to make of objects in a natural scene, it may not always be valid, (for example when neither the objects nor the viewer moves) leaving the possibility that simple spatial separation of objects may need to be sufficient to learn independent representations of them.

Traditionally, visual perceptions have been assumed to be represented by the activation level or firing-rate of neurons, known as the *spike-count hypothesis*. Indeed, previous work has suggested that the average firing rate over a short window from the onset of the stimulus, $\langle f(t) \rangle_T$, is the relevant code for transmitting information (Tovée et al., 1993) with estimates for T of 5–10 ms (Koch, 1999, p.333), 20, or even up to 50 ms (Gerstner and Kistler, 2006, p.21). While the majority of the information of output neurons' responses may be contained within their firing rates (Tovée et al., 1993), the timing of their action potentials may still be important for how networks self-organize, potentially allowing them to overcome the limitations inherent in a simpler model.

In one spiking neural network, the 'binding problem' of separating stimuli within large receptive fields was overcome through an attentional mechanism implemented by selectively reducing the firing threshold of particular neurons throughout the layers,

whose receptive fields fell in the attended region VanRullen and Thorpe (2002). While attention may play a role in some circumstances, there must still be an automatic mechanism for unattended scene segmentation. Previous work with a network of spiking neural networks has found that, under the right conditions, competing populations of neurons will tend to push one another out of phase and thereby alternate their respective perceptual representations through time in a phenomena dubbed 'perceptual cycles' (Miconi and VanRullen, 2010). This mechanism has been demonstrated to allow for both segmentation and binding (feature linking) of a visual scene (Choe and Miikkulainen, 1998) and may be used by spiking neurons to overcome the difficulties presented by multi-object training paradigms.

Once such an antiphase dynamic is established in the inputs, it is hypothesized that postsynaptic excitatory cells in subsequent layers will be able to learn (through spike-time dependent plasticity in the feed-forward synapses) about each object independently of the others as they translate across the input layer, without having to rely upon statistical decoupling through presenting many stimulus combinations and the extensive training which that process requires. Hence, the binding and segmentation occurring naturally through the temporal code of the network's inputs should allow separate pools of transformation-invariant cells for each separate object to rapidly and naturally form in the output layer.

In the earlier work of Chapter 3, it was demonstrated how a more biophysically accurate spiking neural network, where individual action potentials were explicitly modelled (rather than a time-window average of activity) could form transformation-invariant representations of objects presented individually during training (Evans and Stringer, 2012). In this chapter it is demonstrated how such a spiking neural network can utilize the richer spiking dynamics to learn separate transformation-invariant representations of visual stimuli which are always seen together and always moving together, but which are located in separate positions in the visual field. This is achieved by combining a SOM network architecture with adapting spiking neurons learning through a spike-time dependent learning rule. There follows an introduction to the key components required for the operation of this model in more detail and a summary of how they interact to achieve separate transformation-invariant representations.

5.1.1 Lateral connections and self-organizing maps

Lateral connections are commonly found throughout the visual cortex and are believed to explain the formation of the similarly ubiquitous feature maps, for example, of orientation preference of V1 simple cells (Miikkulainen et al., 2005). While the

clustering of cells with similar response properties into such maps may be regarded as an epiphenomenon, it has also been proposed that the extensive lateral connectivity fulfils a functional role in the organization of visual perceptions.

Since features spatially close to one another in a visual projection are likely to belong to the same stimulus, one such function may be in helping to synchronize common features of a stimulus to produce a coherent percept (and desynchronize them with respect to another simultaneously presented stimulus) in what is known as the binding hypothesis (Engel et al., 1991). In essence, this enables an individual to view a scene with, for example, a red square and a blue triangle, and know that the square, not the triangle, has the property of redness, while the property of blue is associated with the triangle alone, without the need to have a 'blue-triangle' cell and a 'red-square' cell (and many more cells to represent other specific combinations of properties too). This is typically believed to be mediated by the natural γ-oscillations found in the cortex and elegantly avoids the combinatorial explosion of combination cells which would otherwise be required in a system of representation where conjunctions of perceptual primitives have to be represented explicitly.

So far, there is a strong accord for this idea from convergent sources of evidence. Direct physiological evidence for this hypothesis comes from observing the synchronized oscillations of visual neurons with similar response properties when presented with a common input, for example orientation preference (Gray et al., 1989). Similarly, psycho-physical experiments which investigate the perception of stimuli resulting from the manipulation of their respective phases of presentation show that stimuli presented synchronously are harder to discriminate between. Usher and Donnelly (1998) found that when flashing a target and background asynchronously, subjects' performance in correctly determining where the target stimulus was presented was consistently higher than when they were flashed synchronously, and performance further increased with a larger phase difference. Consistent with these results are those from real-time optical imaging studies which demonstrate that the spontaneous firing of single neurons is tightly linked to the cortical networks in which they are embedded (Tsodyks et al., 1999).

Such features of lateral connectivity therefore allow the temporal binding of the proximal features of one stimulus, in antiphase or perhaps alternate γ-cycles to features of another (more distant) object. We now turn our attention to neural network modelling studies which have further investigated the conditions which are necessary for such dynamics.

	Excitatory connection	Inhibitory connection	
Axonal delay	Desynchrony	Synchrony	Slow PSP decay
No axonal delay	Synchrony	Desynchrony	Fast PSP decay

Table 5.1: General conditions under which synchrony may be achieved in populations of spiking neurons. Principal cells may be reciprocally connected with excitatory or inhibitory synapses (or equivalently via inhibitory interneurons), either with or without delays.

5.1.2 Conditions for synchronous cell assemblies

Previous work has revealed several key features required of a model to form synchronous assemblies of neurons representing a particular stimulus ('feature linking') and to generate an antiphase relationship between competing (input) representations. One of which is the requirement for (short-range) lateral excitatory connections between principal excitatory cells which form a mutually supportive basis for synchronizing the spike volleys of spatially proximal features of a particular object (while the inhibitory interneurons help to desynchronize between objects). The second requirement is either conduction delays (Nischwitz and Glünder, 1995; Miconi and VanRullen, 2010), varying postsynaptic potential decay rate (Choe and Miikkulainen, 2003) or cell firing-rate adaptation (Reitboeck et al., 1993; Choe and Miikkulainen, 1998; Stoecker et al., 1996). Together, these two requirements act to generate periodic firing in each population of principal cells.

The conditions for synchronization were studied in detail for pairs of neurons with conduction delays (Nischwitz and Glünder, 1995). In general, four regimes were identified in their analysis of a simple two neuron system with excitatory or inhibitory coupling, and then confirmed with larger scale simulations. These regimes are as follows: (1) mutual excitatory connections without delays cause synchrony (quickly); (2) mutual excitatory connections with delays cause desynchrony; (3) excitatory to inhibitory connections without delays cause desynchrony; and (4) excitatory to inhibitory connections with delays cause synchrony (slowly). Similarly, $E \rightarrow E$ synapses with fast PSP decay and $E \rightarrow I$ synapses with slow PSP decay lead to synchrony, whilst the opposite combinations lead to desynchrony (Choe and Miikkulainen, 2003). This is summarized in Table 5.1.

In the work presented here, we chose cell firing-rate adaptation as the mechanism by which periodic firing is generated through a 'time-delayed neuronal self-inhibition mechanism' (Liu and Wang, 2001) as this is a common feature of many spiking neuron models and found throughout the brain. Interestingly, when operating in

conjunction with Spike-time dependent plasticity (STDP) it has also been found to yield almost optimal information transmission (Hennequin et al., 2010). It was found in the simulations described below that this mechanism facilitated a homogeneous population of principal cells to display the emergent behaviour of interest.

5.1.3 Cell firing-rate adaptation

Cell firing-rate adaptation is a prominent feature of cortical neurons, in particular of the excitatory pyramidal cells. One model of this neuronal property is derived from Calcium-gated (Ca^{2+}) Potassium (K^{+}) channels which, when opened by recent cellular action-potentials, allow the flow of Potassium across the cell membrane. This is known as the after-hyperpolarization current (I_{AHP}), which is activated by calcium ions that enter through voltage-gated L-type channels during an action potential (Tanabe et al., 1998), and leaks out of the cells through the membrane with a resultant shunting effect upon the cell potential. This effectively makes the neuron more 'leaky' and hence means that it is harder to build up the voltage difference required to fire another action potential. The potassium channels then shut with a particular time course (governed, for instance, by the dispersal rate of the Calcium) such that the adaptation current (I_{AHP}) exponentially decays to 0 (Rolls and Treves, 1998; Liu and Wang, 2001).

Alternatively, firing-rate adaptation may be modelled phenomenologically as a 'dynamic-threshold' model whereby there is a temporary rise in cell firing threshold upon the occurrence of action potentials, making it similarly harder for the cell to fire after recent activity. This dynamic threshold also decays back to its resting state with a characteristic time-course governed by the time constant of the model (Liu and Wang, 2001).

Both models are qualitatively similar in their behaviour (Liu and Wang, 2001) but the potassium-current model was chosen since it has a more direct correspondence with physiologically measurable quantities and is closer to the observed behaviour of cortical neurons in that the time constant of adaptation, τ_{Ca}, increases as a function of the input current intensity (Ahmed et al., 1998), whereas the reverse is predicted for the voltage-threshold model.

5.1.4 Overview of model dynamics

Here it is demonstrated how the Gaussian profile of excitatory lateral connections (effectively modelling the short-range excitation of a SOM architecture) helps to syn-

chronize the discharges of local clusters of neurons in the input layer which represent an individual visual object while the long-range (global) inhibitory connections desynchronize the action potentials between spatially separate clusters of input neurons (which represent different visual objects). The two visual input representations are thus pushed out of phase with respect to each other in this manner, with the cell firing-rate adaptation ensuring that one representation does not continually suppress the other but that the volleys of spikes oscillate between the different stimuli on a time-scale of roughly $100\,ms$.

With the dynamics of the input layer representations settling into the described anti-phase oscillations, the strength of the plastic feed-forward excitatory connections projecting to the next layer are selectively modified through the spike-time dependent learning rule. Specifically, there is long-term potentiation (LTP) if the presynaptic spikes occur in the order of $10\,ms$ before the postsynaptic spikes and long-term depression (LTD) if this order is reversed (Bi and Poo, 1998). If separate output neurons fire between the oscillations of the two input representations, they will experience LTP for only one stimulus (and LTD for the other if the frequency of oscillation is sufficiently high). The effect of this is that separate pools of output neurons (determined by the initial random feed-forward connectivity) learn to respond selectively to only one of the synchronized input clusters representing a particular stimulus.

This mechanism may be combined with Continuous Transformation (CT) learning to achieve transformation-invariant output representations. This is achieved by continuously transforming (shifting) the stimuli on the input layer. Due to the similarity between transforms within a set belonging to a particular stimulus, the Continuous Transformation (CT) learning mechanism (Stringer et al., 2006; Evans and Stringer, 2012) is able to build the desired output representations using the same spike-time dependent learning rule in the feed-forward connections.

It is shown how these neural mechanisms may operate together during learning in order to produce output cells that respond selectively to a specific stimulus in a transformation-invariant manner when the visual objects have always been presented moving together in lockstep during training but physically separated in space. This network behaviour relies upon the explicit modelling of the times of spikes, together with STDP in order to obtain the necessary dynamics and would not be possible in earlier *rate-coded* models such as VisNet (Stringer et al., 2007; Stringer and Rolls, 2008; Tromans et al., 2012).

5.2 Methods

5.2.1 Model Architecture

The model consists of two layers of adapting excitatory integrate-and-fire neurons (each with its own separate pool of inhibitory interneurons), which were connected in a simple feed-forward architecture (with one layer of spike-time dependent plastic conductance based synapses between them). Excitatory neurons were fully connected to the inhibitory neurons within their own layer and also fully connected to all other excitatory cells in the same layer with a weight decreasing exponentially with distance (as detailed in Section 5.2.3). This gave the effect of short-range lateral excitation coupled with diffuse (long-range) global inhibition as a variant of the common 'Mexican Hat' weight profile. All synapses were of fixed efficacy except the feed-forward excitatory weights which were updated during learning with a form of spike-time dependent plasticity described in Section 2.3.1 (Perrinet et al., 2001).

The input layer contained 512 excitatory cells (arranged in one dimension) to provide enough room for translating stimuli, while the output layer contained 256 excitatory cells (arranged in a two-dimensional evenly spaced grid of 16×16). Each layer had one quarter as many inhibitory interneurons with full connectivity to the excitatory cells within their respective layers.

5.2.2 Neuron model description

The neurons used in this work are conductance-based leaky integrate-and-fire neurons with Gaussian noise added to the cell membrane potential with zero mean and standard deviation $\sigma = 0.015 \cdot \left(\Theta - V^H\right)$ Masquelier et al. (2009), as described previously in Chapter 2, Equations 2.16–2.19.

For a complete model description, please refer to Table 5.2 for the parameters used in the presented simulations.

In order to model the property of cell firing-rate adaptation, a small change is made to the standard leaky-integrate-and-fire model which describes the cell membrane potential, $V(t)$, to incorporate the calcium-gated potassium currents, as shown in Equation 5.1 and another differential equation describing the Potassium channel dynamics, shown in Equation 5.2, with further details upon the cell membrane potential reaching firing threshold given in Equation 5.3.

$$C_m \frac{dV(t)}{dt} = g_0 \left(V_0^\gamma - V(t)\right) - g_{AHP}[Ca^{2+}](V(t) - V_K) + I(t) + \sigma \cdot \xi(t) \cdot \sqrt{\tau_m} \quad (5.1)$$

Here the time constant of the cell membrane is rewritten as its component parts, the Capacitance, C_m and leakage conductance, g_0 (inverse of the membrane resistance, R_m), such that $\tau_m = RC_m = {}^{C_m}/_{g_0}$. As in the standard model, V_0^γ is the membrane reversal potential (with γ denoting the class of neuron, either Excitatory or Inhibitory) and $I(t)$ represents the sum of currents flowing into the cell from inhibitory synapses, excitatory synapses and direct 'electrode' stimulation. Modelling the Gaussian noise process, $\xi(t)$ is a Wiener variable as described earlier in Section 2.6.

$$\frac{d[Ca^{2+}]}{dt} = -\frac{[Ca^{2+}]}{\tau_{Ca}} \tag{5.2}$$

$$\text{If } V(t) = \Theta, \text{then} \begin{cases} V(t) \to V_H \\ [Ca^{2+}] \to [Ca^{2+}] + \alpha \end{cases} \tag{5.3}$$

While g_0 is the leakage conductance of the standard leaky integrate-and-fire model neuron, $g_{AHP} \cdot [Ca^{2+}]$ is the effective K^+ conductance of the adaptation mechanism where g_{AHP} is typically $15\,\mu S/cm^2$ (tuned to $30\,mS$ in total) and a single spike increases the cytoplasmic concentration of calcium $[Ca^{2+}]$ by $\alpha = 200\,nM$ (Liu and Wang, 2001). The time constant for the decay rate of $[Ca^{2+}]$ back to its initial value of 0 is governed by the time constant $\tau_{Ca} = 50\,ms$ (Equation 5.2). Finally, V_K, the reversal potential of potassium is set to $-80mV$. The adaptation current from the open potassium channels is referred to as $I_{AHP} = g_{AHP}[Ca^{2+}]\,(V(t) - V_K)$.

This system of differential equations describing the dynamics of the cell bodies, synaptic conductances and synaptic plasticity are discretized with a Forward Euler numerical scheme and simulated with a numerical time-step Δt of $0.02\,ms$.

5.2.3 Lateral connectivity

To create a self-organizing map-like spatial structure, the Euclidean distances between all principal (excitatory) neurons within a layer are calculated according to the following equation:

$$d_{i,j} = \sqrt{min(|x_i - x_j|, ||x_i - x_j| - S_x|)^2 + min(|y_i - y_j|, ||y_i - y_j| - S_y|)^2} \tag{5.4}$$

Here, S_x and S_y are the sizes of the x and y dimensions respectively, which together with the 'min' functions implement periodic boundary conditions, such that one-dimensional layers become circular and two-dimensional layers become toroidal. All excitatory neurons within the input layer are then connected to every other excitatory neuron up to 5σ (assuming the probability of connection, $p(ElE) = 1$)

with their synaptic weight becoming exponentially weaker with increasing Euclidean distance according to Equation 5.5. For a simple illustration of the profile of lateral excitatory connectivity see Figure 5.1.

$$\Delta g_{i,j} = \frac{\phi_{ElE}}{\sigma_E \cdot \sqrt{2\pi}} \cdot exp\left\{ \frac{-d_{i,j}^2}{2\sigma_E^2} \right\} \tag{5.5}$$

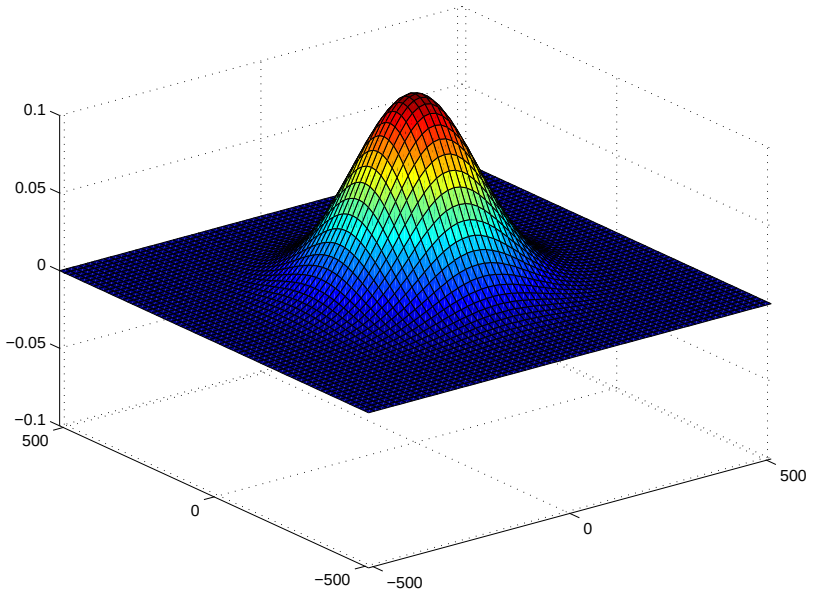


Figure 5.1: Plot to illustrate the strength of the lateral excitatory connections (z-axis) over a two-dimensional network layer.

The default cell body and synaptic parameters (Troyer et al., 1998) and noise parameters (Masquelier et al., 2009) were used throughout these simulations unless otherwise indicated which may be found in Table 2.2 and the time constant of the excitatory feed-forward synaptic conductance, τ_{EfE} was set to $2\,ms$ in line with a CT learning mechanism (Evans and Stringer, 2012) as explored in Chapter 3. The strengths of excitatory feed-forward synapses ($\lambda \cdot \Delta g^{EfE}$) were plastic, modified by the STDP learning rule in the range $[0, 3.75]\,nS$ (lower than the default due to the simultaneous presentation of multiple stimuli) and the strengths of the lateral excitatory connections ($\lambda \cdot \Delta g^{ElE}$) were set to a maximum of ${}^{100}/_{\sqrt{2\pi} \cdot \sigma_E}\,nS$, (and $\sigma_E = 32$ by default) with an exponential decrease with increasing Euclidean distance as described in Section 5.2.3. Similarly, default values were used for the STDP model (Perrinet et al., 2001), as shown in Table 2.3, except where they were systematically varied to assess their effect upon network performance.

5.2.4 Stimuli and Training

Stimuli were represented by injecting tonic current into spatially separate pools of input-layer neurons. Contiguous blocks of neurons within these pools were gradually shifted across the input layer to represent successive overlapping transforms of each stimulus which which may be associated by Continuous Transformation learning (Stringer et al., 2006). In the initial simulation with two stimuli, a stimulus consisted of 64 neurons which was presented in 13 locations (transforms), with a shift of 16 neurons between each adjacent transform giving an overlap of 75% for the facilitation of the CT learning mechanism.

During testing phases, all stimuli were presented individually to measure how the network responded specifically to each stimulus. However during training, the stimuli were presented together, such that the network never learnt about them in isolation. The untrained network was first tested in a 'pre-training' phase by presenting all transforms of all stimuli sequentially, each for a cue period of $500\,ms$ to provide a baseline behaviour to contrast with the effects of learning in the feed-forward synapses.

After saving the network outputs and resetting the dynamic variables for each cell and synapse, the training phase began where all transforms of all stimuli were presented for $500\,ms$ per transform. For combinations of stimuli, the direction of shift between transforms was randomly chosen but with each stimulus in the presented pair shifting in the same direction at the same rate (lockstep). The presentation of all transforms of the pairs of stimuli constituted one epoch of training, and the training phase consisted of ten epochs in total. Figure 5.2 illustrates the multi-stimulus presentation paradigm with two stimuli and five transforms each over one epoch of training.

Once the network had been trained and the dynamic variables (except synaptic weights) reset, the 'post-training' testing phase was simulated in an identical way to the pre-training testing phase. The final outputs were then saved and analysed as detailed in Section 2.9.

5.3 Results

The results section first demonstrates the input layer dynamics and the ability of feed-forward plastic connections to take advantage of them in order to form transformation-invariant representations of each stimulus separately before exploring how robust this effect is to variations in key parameters. The stimuli are then expanded to a larger set

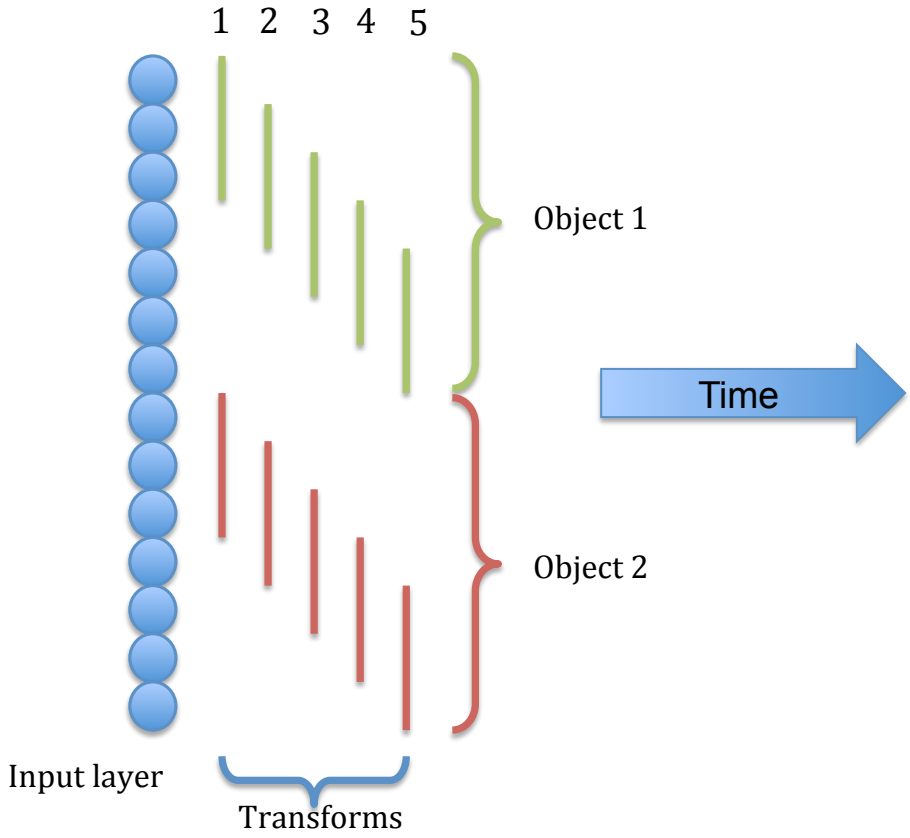


Figure 5.2: Diagram to illustrate the simultaneous presentation of the pair of stimuli translating across the input layer during training.

of four (simultaneously presented during training) to show how the same mechanism may be applied in more ecologically valid scenes composed of many stimuli.

5.3.1 Synchronization of object representations in the input layer

In the first simulation presented, a network was built with two layers of excitatory neurons (each with a separate pool of inhibitory neurons) as described in Equations 5.1–5.3 with parameters specified in Table 5.2. The differential equations were simulated with a Forward Euler scheme with a time-step of $0.02\,ms$ unless otherwise stated. The lateral connectivity was specified (between excitatory neurons within each layer only) according to the methods described in Section 5.2.3. Two stimuli were presented simultaneously to the network during training (but individually during testing) by injecting a current of $0.75\,nA$ into the cell bodies of the sets of input

neurons representing the particular transform of a particular stimulus for $500\,ms$ during training (or $500\,ms$ during testing).

While both stimuli are represented simultaneously and with equal strength, the input layer neurons rapidly adjust the timing of spikes such that each stimulus is represented separately to the other through time. That is, the constellation of features representing a stimulus are synchronized with respect to one another and desynchronized with respect to the features of the other stimulus (see Figure 5.3, bottom). Throughout the course of stimulation, these two competing representations alternate (as shown in the PSTH, Figure 5.3, top) with a regular frequency of approximately $12.5\,Hz$, meaning that combined, the input layer exhibits γ-frequency oscillations at slightly less than $25\,Hz$.

Looking at the combined PSTH (Figure 5.3) it is clear that the two competing populations of input neurons representing each stimulus have pushed one another out of phase, as the volleys of spikes (and frequency bars) for each stimulus are interleaved through time. This is confirmed in the correlations of Figure 5.4. The cross-correlation (Figure 5.4b) shows that the volleys representing the two stimuli are positively correlated with lags of approximately $\pm\{45, 135, 225\}\,ms$ and anti-correlated elsewhere meaning that they are separated by a period, p, of approximately $45\,ms$ (with the positive correlations corresponding to $\{p, 3p, 5p\}$). Furthermore, the autocorrelations for each stimulus' populations of input cells (Figures 5.4a and 5.4c, either side of the cross-correlation) show that the volleys are repeating through time approximately every $90\,ms$ $(2p)$.

To explain this phenomenon, consider the features (excitatory input neurons) representing a particular stimulus. For both populations (representing each stimulus), the external stimulation in terms of time course and amplitude is identical such that initially the neurons representing both stimuli fire together (as can be seen in the first $\sim 50\,ms$ of the raster plot in Figure 5.3). However, the noise in the neurons' cell potentials means that one population (or subpopulation) will by chance, quickly come to dominate the initial competition. These cells transmit action potentials to their neighbouring cells, thus raising the cell potentials of those nearby neurons and encouraging neurons which represent features of the same object to also fire. Compared to excitatory neurons representing other features of the same object, those representing the second object are spatially much further away, and so do not receive as much excitation through the lateral connections, which exponentially decrease in strength with distance (see Equation 5.4). Instead, they receive a wave of inhibition from the mutually connected global population of inhibitory interneurons which suppresses

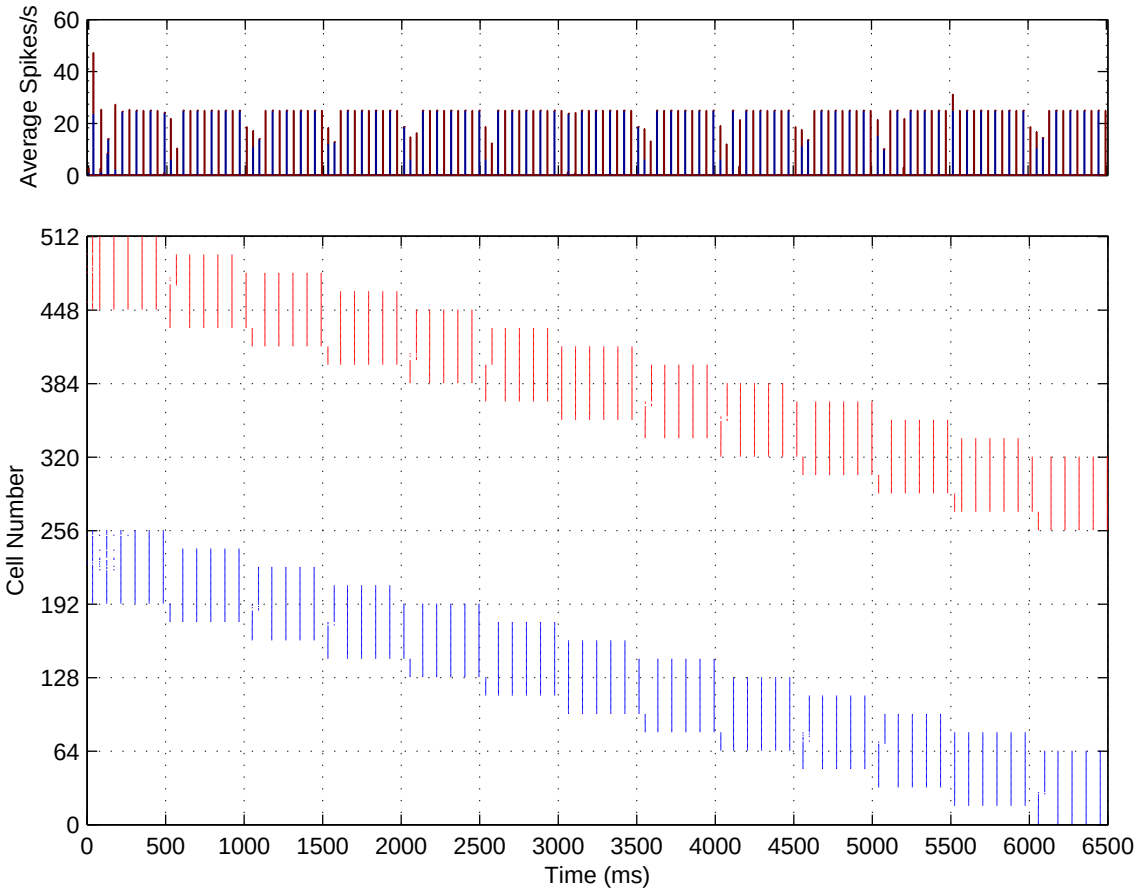
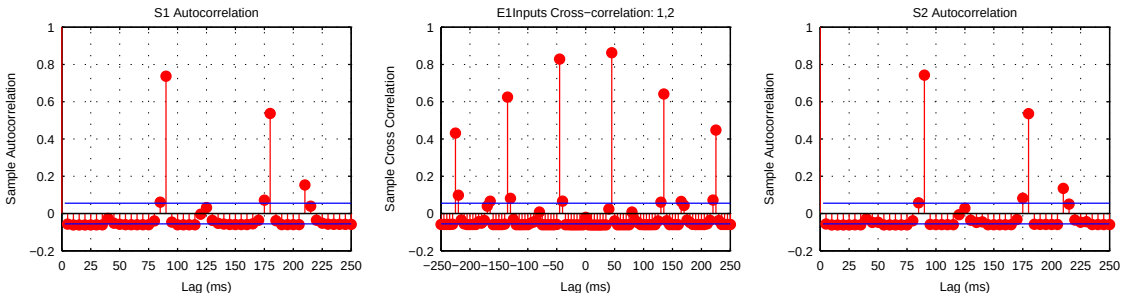


Figure 5.3: Post-stimulus time histogram of global activity in the input layer separated according to stimuli (Top) with the spike raster of the input layer for the simultaneous presentation of each stimulus over all transforms (Bottom). The first stimulus consists of 64 neurons and its transforms are represented by the first (contiguous) half of the input layer (neurons 1–256) over which it gradually moves, while transforms of the second stimulus occupy the second half of the input layer (neurons 257–512).



(a) Autocorrelation: S_1 (b) Cross-correlation (c) Autocorrelation: S_2

Figure 5.4: Autocorrelations and Cross-correlation of the two stimuli in the input layer representations.

their firing. Since the principal cells representing the first object have now fired, they will have adapted through the Calcium-mediated Potassium currents and so it will be relatively harder for them to fire soon after. Neurons in the population representing the first stimulus are therefore less able to stimulate the inhibitory interneurons and thereby suppress the second population of input neurons, which are therefore able to fire their own synchronized spike volleys. The second population of cells then temporarily suppresses the first population by the same interneuron-mediated interaction, until they too self-inhibit through adaptation of their firing-rates and the cycle repeats.

5.3.2 Learning transformation-invariant representations

The process described for the formation of antiphase input representations, when coupled with feed-forward plastic weights is shown here to lead to the formation of transform invariant representations in the output layer. If Spike-Time Dependent Plasticity (STDP) is used, the output neurons will be selective in terms of which population of input neurons they associate with as each population is separated through time. This process persists as the objects translate across the input layer until all transforms of a particular stimulus are potentiated onto a distinct set of output neurons.

In contrast to the precise timings of the action potentials shown in Figure 5.3, it would be very difficult for the output layer to distinguish between the two input stimuli on the basis of firing rates alone (given the full excitatory feed-forward connectivity between the layers). The 13 transforms of the 'compound' training stimulus are plotted in Figure 5.5 for comparison.

In Figure 5.6 we present results demonstrating the formation of cells in the output layer which are selective to one of the two stimuli presented to the network during training, yet invariant to most or all transforms of that particular stimulus. For the testing phase shown in these raster plots, each transform of Stimulus 1 is presented in sequence for $500\,ms$, followed by each transform in sequence of Stimulus 2 also for $500\,ms$ per transform. In the case of the trained network, this leads to a shift after $6500\,ms$ where a separate population of output neurons become active to represent the transforms of the second stimulus.

The effect of training is also clear from the structure in the weight matrix of feed-forward excitatory synaptic conductance strengths as shown in Figure 5.7. These synaptic weights are initialized to random values drawn from a uniform distribution (top) but attain a clear structure through the process of training (bottom). After

Parameter	**Symbol**	**Value**
Cue current	I^{ext}	0.75 nA
Cue period {training, testing}	t_{cue}	{500, 500} ms
No. of training epochs	N_{epochs}	10
Time-step	Δt	0.02 ms
Network Parameters		
No. layers	N_L	2
No. excitatory cells per layer	N_E	{512, 256}
No. inhibitory cells per layer	N_I	{128, 64}
Prob. of E cell synapsing with afferent feed-forward E cell	$p(EfE)$	$\{\sim, 1.0\}$
Prob. of E cell synapsing with afferent lateral E cell	$p(ElE)$	$\{1.0, 0.0\}$
Prob. of E cell synapsing with afferent I cell	$p(IE)$	$\{1.0, 1.0\}$
Prob. of I cell synapsing with afferent E cell	$p(EI)$	$\{1.0, 1.0\}$
Prob. of I cell synapsing with afferent I cell	$p(II)$	$\{0.0, 0.0\}$
Standard deviation of lateral connection strength	$\sigma_E^{lat.}$	$[0, 256]$

Table 5.2: Network parameters used in the SOM simulations.

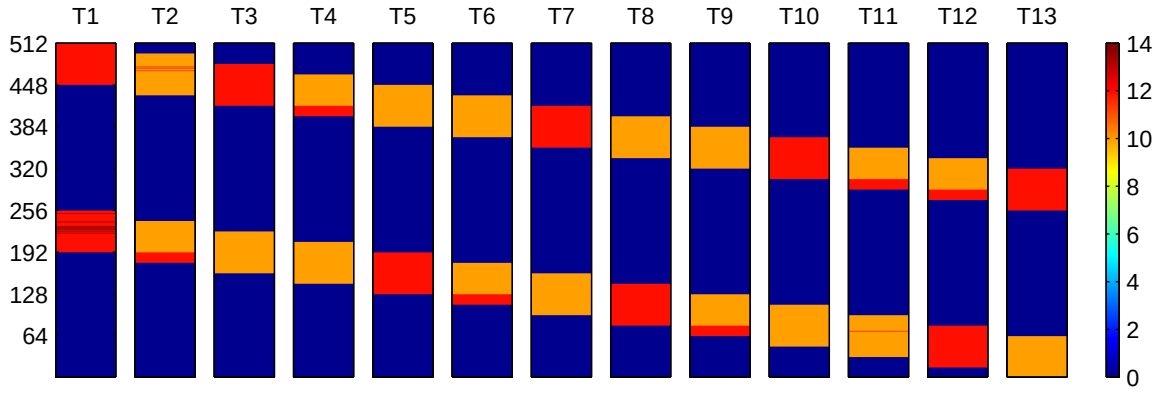
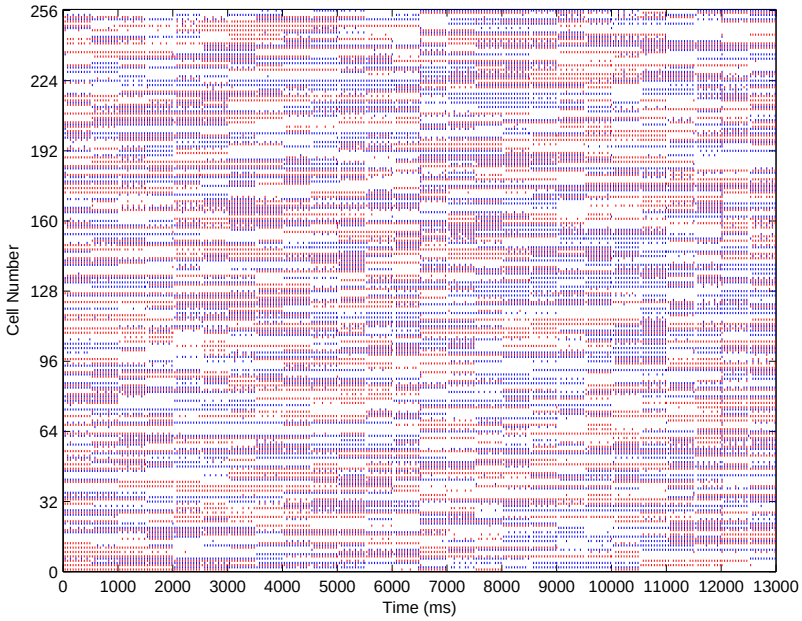
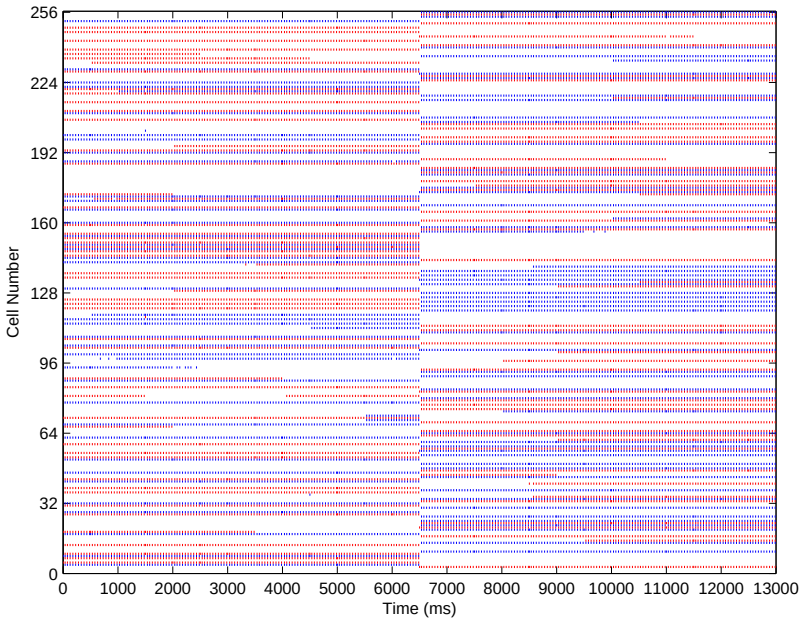


Figure 5.5: Firing rate plots of the compound training stimulus across all transforms. On the basis of the firing rate alone (indicated by the colour of heat map in Spikes/s) it is very difficult for the next layer of neurons to distinguish between the transforms of each of the testing stimuli when presented together.



(a) Before Training



(b) After Training

Figure 5.6: Raster plots of the output layer neurons during presentation of each transform of each stimulus before and after training. Before training (a), the output cells respond randomly to transforms of each stimulus. After training (b), the majority of output cells have become object selective and transformation-invariant.

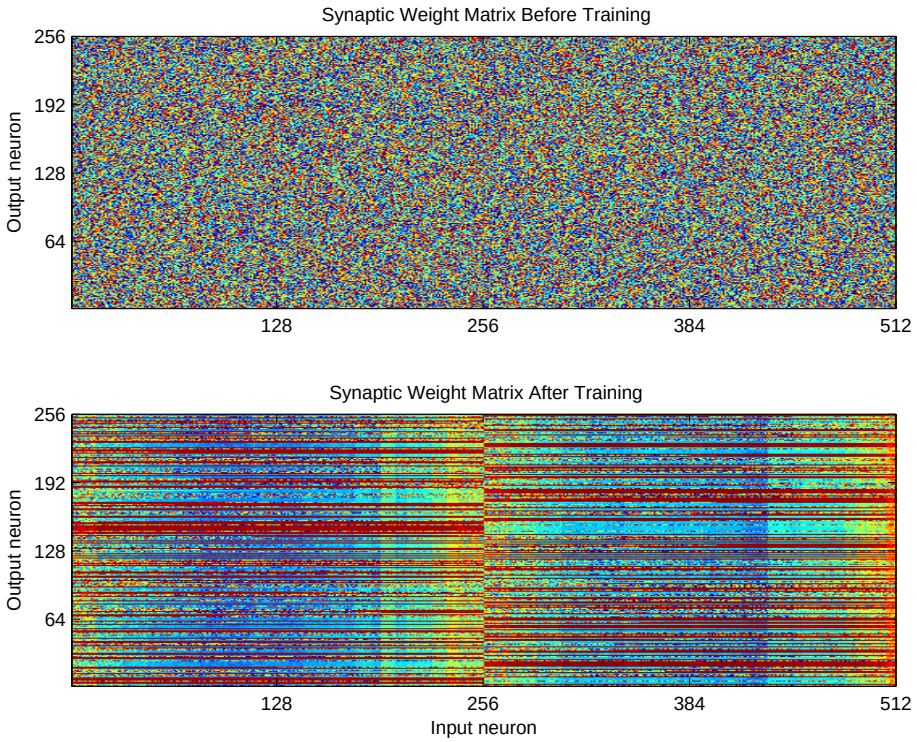
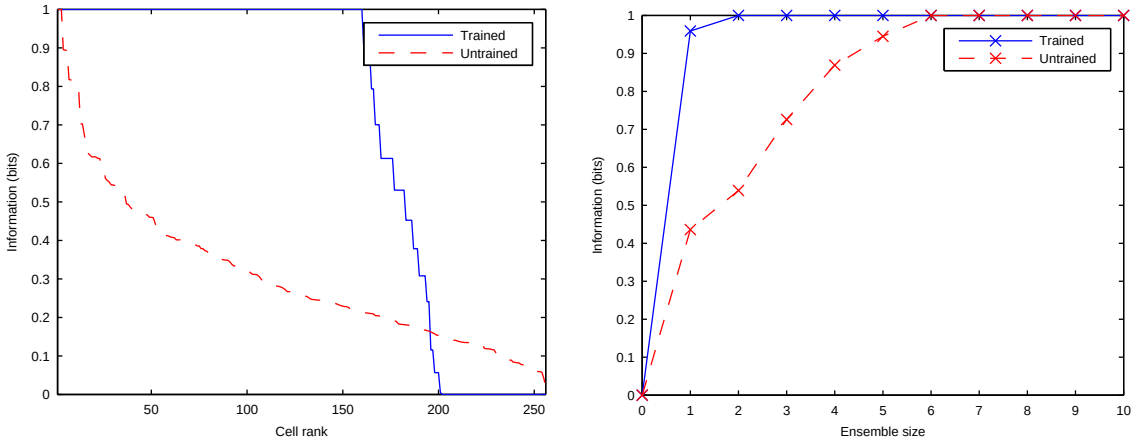


Figure 5.7: Excitatory feed-forward synaptic weight matrices (before and after training). The strength of the synaptic weights are indicated by the colour (red being high and blue being low). Before training (top) the weight matrix is random (unstructured). After training (bottom) there are strong connections from all the inputs representing all transforms of one stimulus to particular output neurons (indicated by horizontal red stripes) and the same for the input neurons representing transforms of the other stimulus to different output neurons.

training, particular output neurons (shown on the y-axis) can be seen to have striations of large weight values extending across, for example, the first half of the input layer (x-axis) corresponding to all transforms of the first stimulus, while the second half of the input layer have formed strong feed-forward synapses with other output neurons which have come to represent all transforms of the second stimulus.

As a quantitative measure of the network's ability to learn to form transform-invariant representations of each stimulus, information analysis plots of the output layer are given in Figure 5.8. The plot of the single-cell information measure shows that a large proportion of output cells transmit the maximum information ($1\,bit$) across all transforms of the stimuli, meaning that they unambiguously signal which stimulus is being presented. The multiple-cell measure plot shows that both stimuli are represented in the output layer which also attains the maximum information level.



(a) Single Cell Information (b) Multiple Cell Information

Figure 5.8: Information plots of output neurons for the trained and untrained network with two simultaneous training stimuli. For the trained case, the single cell information measure (a) shows that almost all cells in the output layer are maximally informative in discriminating between the stimuli and the multiple cell information measure (b) shows that both stimuli's transforms have been learnt by the network.

5.3.3 Gradient of strength

During preliminary explorations of the parameter space, earlier simulations revealed the gradient in lateral excitatory connections to be a crucial element in generating antiphase relationships in inputs presented to the network, further investigation was needed to understand how the spread of such connections relative to the size of the stimuli relates to the network's ability to learn transform-invariant representations.

To assess the network performance across a range of parameter values, an 'information score', ι_κ was calculated. The single-cell information of each neuron to each stimulus was calculated and the number of cells counted which conveyed at least 95% ($\kappa = 0.95$) of the theoretical maximum (in this case $0.95\,bits$). The minimum number of such cells was then found, computed across the stimuli, s, as a measure of how much information the network conveyed about all transforms of the least well represented stimulus (see Equation 5.6).

$$\iota_\kappa = \frac{\min_s \left| \{ I_{c,s} \geq \kappa \cdot \log_2 N_S \}_c \right|}{C} \tag{5.6}$$

Here, $I_{c,s}$ is the amount of information conveyed by a particular output cell, c, about a particular stimulus, s, according to the single-cell information measure, N_S is the number of stimuli and C is the number of cells in the output layer. Although this measure is derived solely from the single-cell information measure, taking the minimum

proportion of cells across all stimuli means that non-zero values of ι_κ indicate that all stimuli are represented, fulfilling the role of the multiple-cell information analysis.

If the radius of excitation grows too large relative to the spacing between the stimuli, then two stimuli are encouraged to fire together, leaving the postsynaptic neurons unable to distinguish between their synchronized activity. If the radius becomes too small relative to the size of the stimuli, the input representation becomes fragmented as not all features of the object are synchronized by the lateral excitatory connections and only partial views (or a subset of transforms) are learnt about by postsynaptic neurons. In summary, for large σ_{ElE} specificity suffers and for small σ_{ElE} invariance learning suffers.

Here we used the parameters from the optimal simulation (Table 5.2, with $\Delta g_{ElE} = 20$) and varied σ_{ElE} from 0 to 256. We use the same measure of network performance as for the previous simulations varying the STDP time constants and the results of running ten different random seeds are shown in Figure 5.9, with the mean information score plotted as points and the standard error of the mean over the ten random seeds plotted as whiskers.

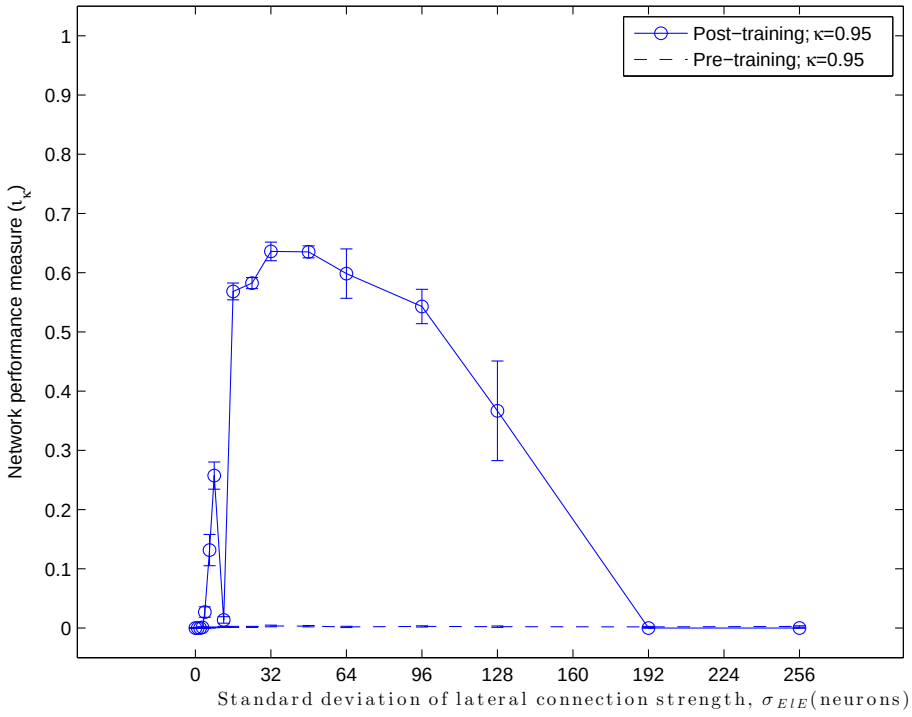


Figure 5.9: Network performance (ι_κ) plotted against the standard deviation of the Gaussian spread of strength of the lateral excitatory connections.

From this analysis, the optimal standard deviation of the lateral excitatory con-

nection profile was found to be $\sigma^*_{EIE} = 32$, (half the width of each stimulus and $1/8$ of the inter-stimulus distance from their centres). Network performance (as measured by the information score) was also found to be more tolerant to larger spreads of the weight profile than to smaller values of σ_{EIE} as shown by the relatively steep decline on the left side of Figure 5.9. The optimal value was used in subsequent simulations throughout this chapter.

The problem of synchronizing independent objects with a large radius of excitation may be alleviated with more realistic inputs and architecture. In V1, cells sensitive to a particular orientation are laterally connected to other cells of similar orientation through excitatory synapses providing a means of contour integration. This would mean that separate objects could be brought closer together in the field of view without their firing becoming synchronized provided that their edges are not too closely aligned. This predicts that the excitatory radius should be large relative to the radius spanned by the cortical representation of an object.

5.3.4 Temporal specificity in learning

In the simulations presented so far, the alternation of input representations on successive gamma cycles allows the output layer to exploit learning through Spike-Time Dependent Plasticity (STDP) — a temporally sensitive learning rule. It is hypothesized that if the form of STDP is made less specific (such that it begins to resemble a firing-rate based learning rule) then the advantage of the self-organized perceptual cycles in the inputs will be lost, as the learning rule will no longer be specific enough for a given population of output cells to learn about a particular input without interference from the other. In other words, the alternation of the input representations will effectively be too fast for output cells to learn about one or the other discriminatingly. In this case, the different input representations will be associated onto the same output cells.

To test this, further simulations were run with a range of STDP time constants (τ_C and τ_D), both symmetrical ($\tau_C = \tau_D$) and asymmetrical ($\tau_C = 3/5 \cdot \tau_D$). To summarize the network performance, the information score, ι was calculated as before according to Equation 5.6. The results of these simulations are presented in Figure 5.10 for asymmetrical time constants (with a larger LTD time window) and Figure 5.11 for symmetrical time constants.

As the STDP time constants deviate from the default values, $15\,ms/25\,ms$, the network performance can be seen to deteriorate (although with symmetrical learning windows, the optimal time constant was found to be longer at $50\,ms$). When the

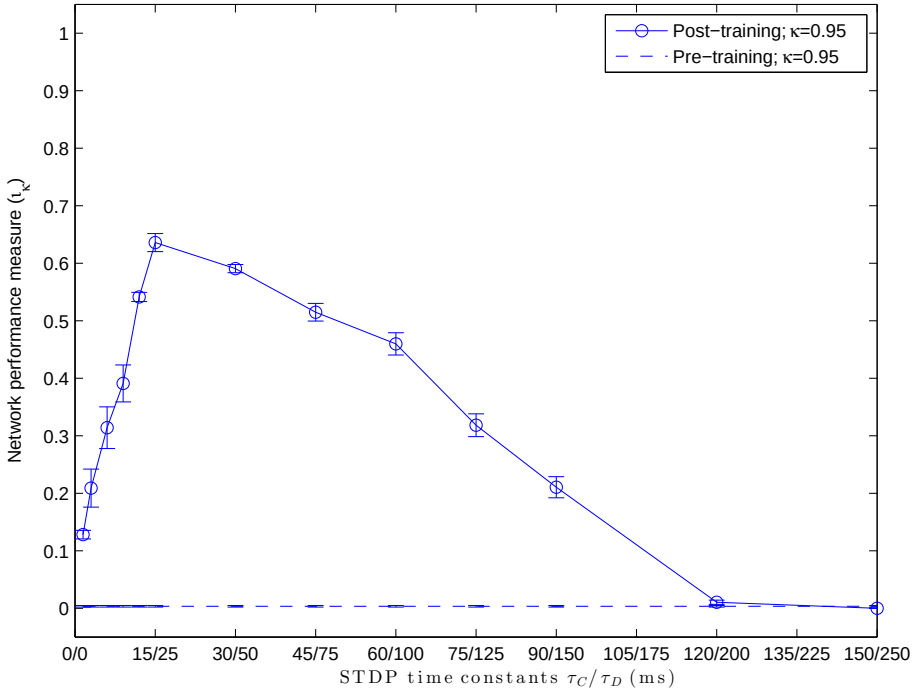


Figure 5.10: Network performance (ι_κ) plotted against varying STDP time constants with a constant ratio of $\tau_C = {}^3/_5\tau_D$.

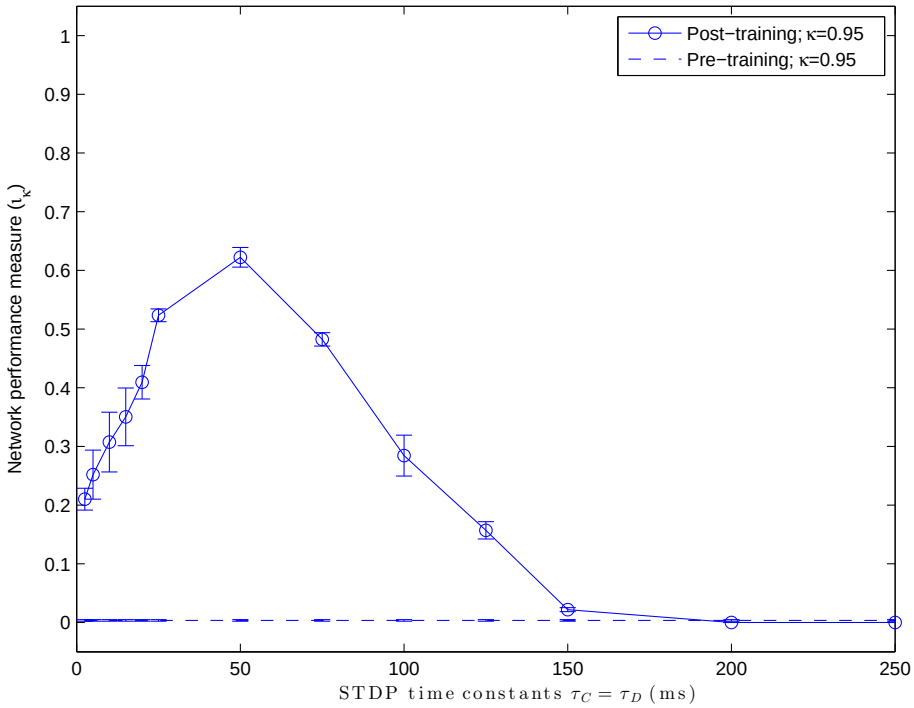


Figure 5.11: Network performance (ι_κ) plotted against varying STDP time constants with symmetrical temporal learning windows where $\tau_C = \tau_D$.

learning time constants are shortened, network performance is reduced since only partial fragments of the stimuli are synchronized within a more restrictive time window and so associating all transforms of a particular stimulus together becomes a more difficult task. By lengthening these time constants, the network performance also decreases (although more gradually) as the oscillations of both stimuli start to experience both LTP and LTD.

From inspecting the output layer cell response properties (not shown) for simulations with STDP time constants of $150\,ms$ or longer, it was confirmed as hypothesized that the learning rule is no longer temporally specific enough for separate sets of output cells to learn about each stimulus independently, and instead one set of output cells tend to form which are invariant to both stimuli.

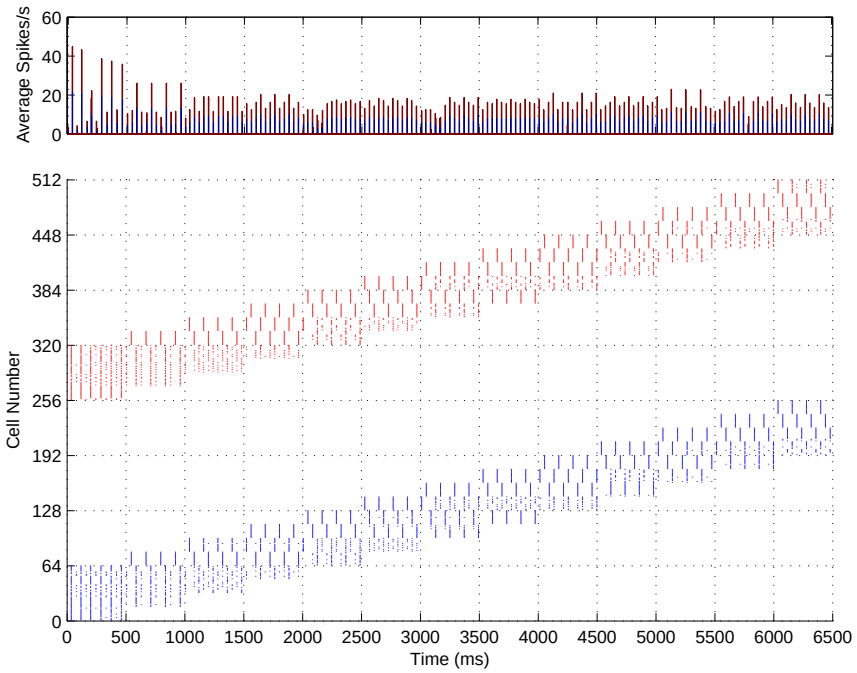
5.3.5 Lateral connections

In order for features of the same object to generate synchronized firing in the input cells which represent them and yet desynchronize the firing between more distant populations of cells representing different objects, a gradient of lateral excitation is necessary, such that cells close together are more mutually supportive than cells further apart.

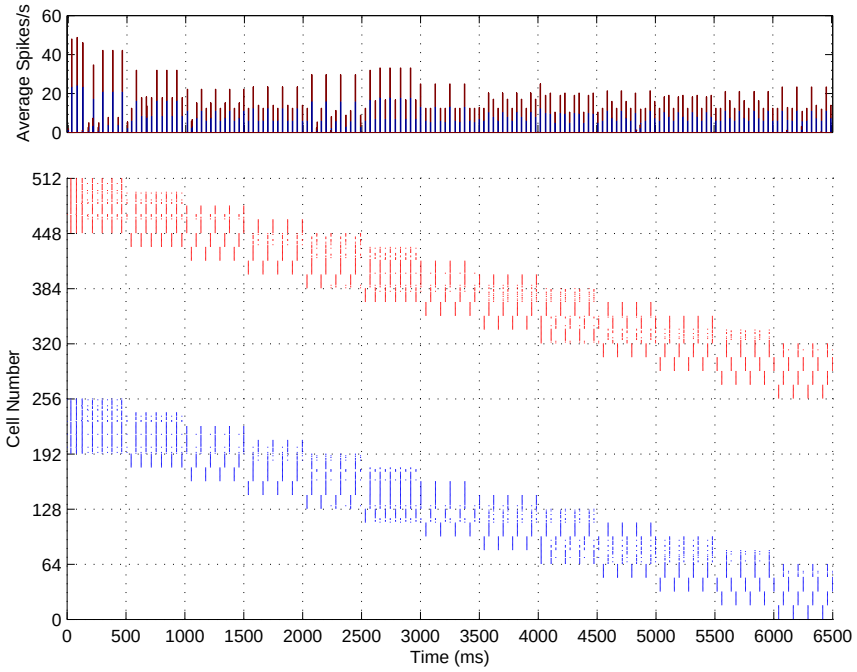
To test this, the same simulation was rerun under two different conditions; firstly with no lateral excitatory connections in the input layer and secondly, keeping the lateral excitatory connections but flattening the Gaussian profile of their strength (i.e. all lateral excitatory connections were of the same efficacy irrespective of the distance between the neurons) and retuning the strength of inhibition to prevent saturated firing through positive feedback.

In Figure 5.12 we present input rasters from a typical simulation with no lateral excitatory connections in the input layer (Figure 5.12a) and one with a flat strength profile and retuned lateral connection strength (Figure 5.12b). It is evident in each plot that not only have the perceptual cycles between the input stimuli disappeared, but also that there is no synchronization of the features of the inputs themselves, but instead disorganized volleys of spikes with no useful structure.

Consequently, with no temporal structure in the spike timings of cells representing the transforms of the input stimuli, the network does not manage to discriminate between the two inputs and hence the single and multiple cell information measures (not shown) are no better than the untrained network (essentially no improvement on random feed-forward connectivity). This suggests the need for a distance-based gradient in the profile of lateral connection strengths in order to form perceptual



(a) Before Training



(b) After Training

Figure 5.12: PSTH (top row) and input rasters (bottom row) for a network with no lateral excitatory connections (top pair) and a network with a flat synaptic efficacy profile in all its lateral connections (bottom pair). In each case, the spike volleys representing each stimulus are disorganized, with no synchronization within a stimulus and no desynchronization between stimuli.

cycles in the input layer as a basis for learning separate object representations in the output layer.

5.3.6 Cell Firing Rate adaptation

Adaptation was found to be a crucial element in generating antiphase representations between input stimuli, also known as 'perceptual cycles' (Miconi and VanRullen, 2010). Without cell firing-rate adaptation, the entire population of excitatory input cells was synchronized by the action of the inhibitory interneurons as can be seen in Figure 5.13 (which may be contrasted with the input layer raster plot of Figure 5.3 showing both stimuli being represented in antiphase cycles). Without a mechanism of self-inhibition and the effects of cell membrane noise and random initialization to help randomly select an initial winner to begin the oscillations, the two populations would continue to fire as one. As such, it was much harder to train the network, as without the dynamic of perceptual cycles, both stimuli would typically be associated onto the same output neurons.

This was also found to be the case for a wide range of parameters, including the strength of the lateral connections, the standard deviation of the distribution of their strengths, (σ_E), and the strength of excitatory to inhibitory connections and inhibitory to excitatory connections. Without the time-varying degree of competition (provided by the self-inhibiting effects of cell firing-rate adaptation in this case), the perceptual cycles can no longer be formed and so are not observed in the results.

In other work, fixed delays, including $E \rightarrow I$ or $I \rightarrow E$ (Miconi and VanRullen, 2010), postsynaptic potential decay rate (Choe and Miikkulainen, 2003) and dynamic-threshold variants of cell firing-rate adaptation (Stoecker et al., 1996; Reitboeck et al., 1993; Choe and Miikkulainen, 1998) have been used to achieve the same effect.

The information can still be high however, since with enough training epochs and the randomizing effect of neuronal noise in the cell membrane potentials, each stimulus should be represented on average as much as any other, and so provide sufficient exposure to each input for separate populations of output cells to learn about each stimulus. This is a fundamental property of competitive neural networks but one which requires extensive training and so can not account for the ability of the visual system to learn from a single exposure to a stimulus.

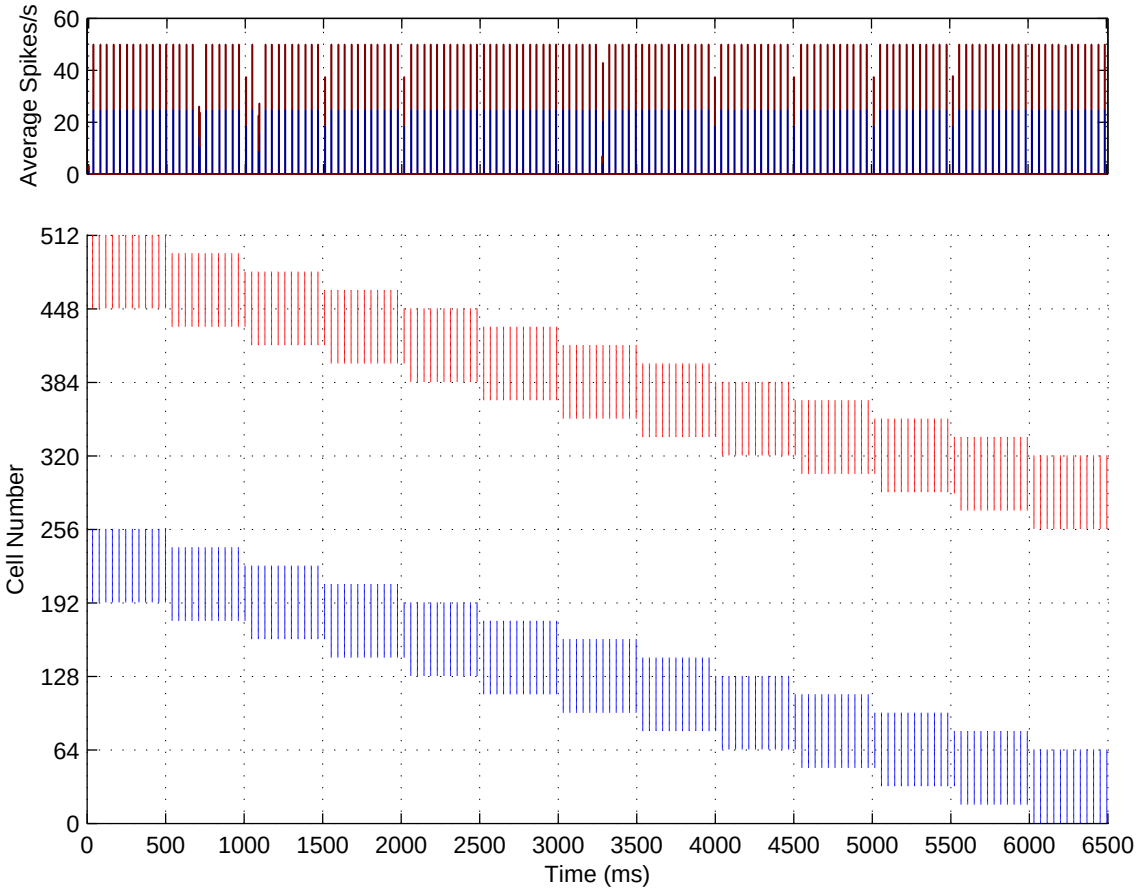


Figure 5.13: Input layer raster plot with no cell firing-rate adaptation. Very quickly, both populations of input neurons start firing together and remain synchronized throughout the epoch.

5.3.7 Capacity

While presenting pairs of stimuli to the network during training is an advance on presenting stimuli in isolation, there is still much scope for more biologically realistic and therefore improved ecological validity of the simulations. In the following simulations we aim to investigate the capacity of the network by presenting larger numbers of stimuli simultaneously during training. The size of the network and the number of transforms remain the same but as the number of stimuli doubles, the size of the stimuli and the shift between transforms both halve to 32 and 8 neurons respectively in the case of four stimuli.

Using the same simulation paradigm as described above (except for the changes necessary to accommodate four stimuli as discussed), the network was trained with all our stimuli presented simultaneously translating across their portions of the input

layer. The PSTH and raster plot of the input layer with four stimuli is shown in Figure 5.14. It can be seen from these plots that the four populations of input neurons have organized themselves into internally synchronized volleys which are out of phase with respect to the spikes from neurons representing the other three stimuli. This is qualitatively very similar to the case with two stimuli presented during training except with the principle difference that the volleys for a particular stimulus fire once every four cycles (as opposed to every two cycles).

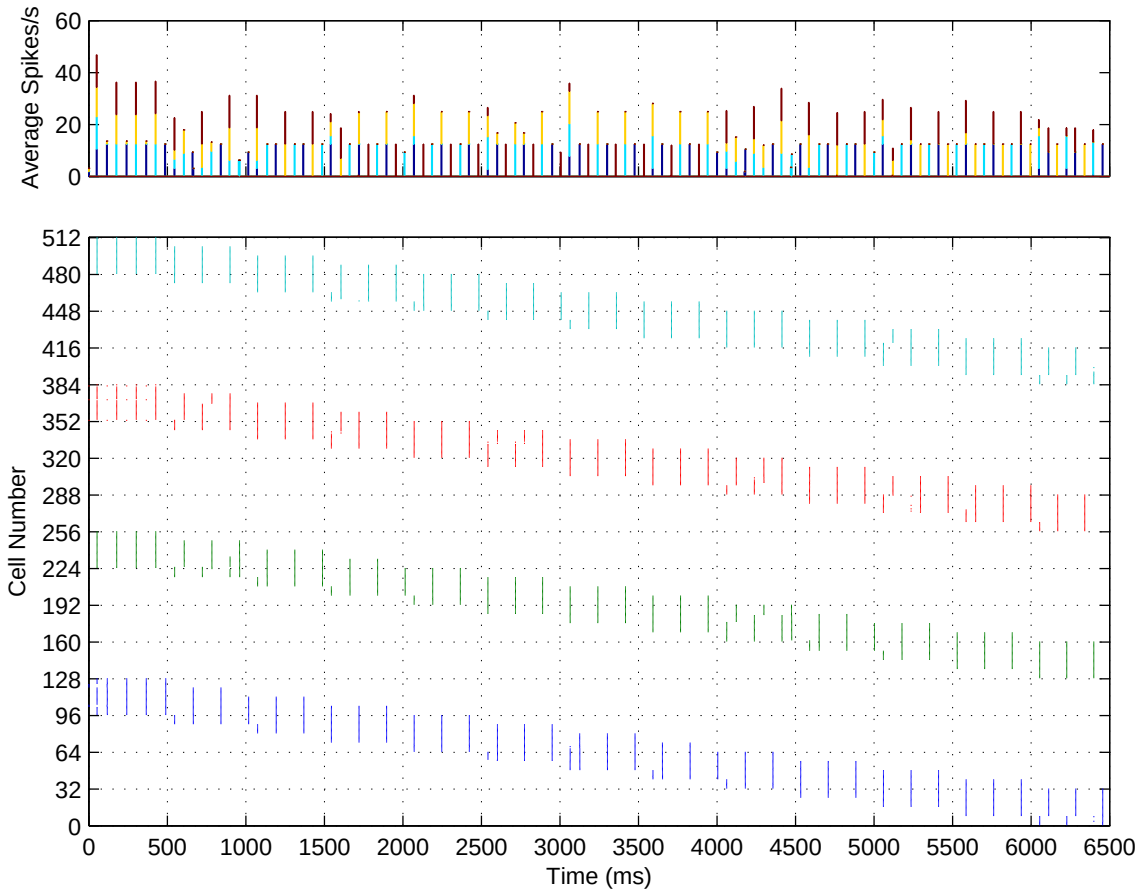
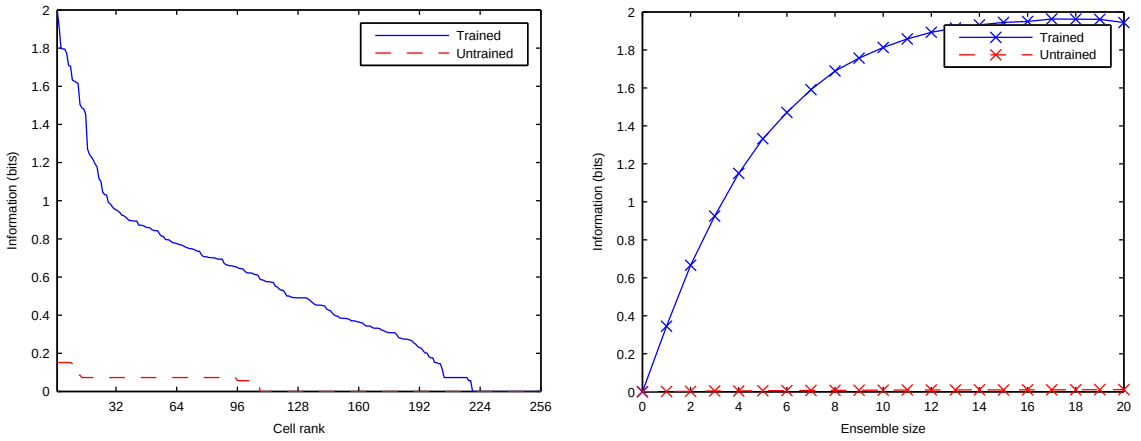


Figure 5.14: Post-stimulus time histogram of activity in the input layer separated according to stimuli (Top) with the spike raster of the input layer for the simultaneous presentation of four stimuli over all transforms (Bottom). The first stimulus consists of 32 neurons and its transforms are represented by the first (contiguous) quarter of the input layer (neurons 1–128) over which it gradually moves, while transforms of the second, third and fourth stimuli occupy the subsequent quarters of the input layer.

The Information plots for four stimuli (Figure 5.15) show that a few neurons in the output layer have been trained to convey almost the theoretical maximum amount of single-cell information possible which has risen to *two bits* ($\log_2(N)$, where N is

the number of stimuli, which in this case is four), rather than one *bit* in the case of two stimuli. This means that these cells have been able to learn to unambiguously signal which stimulus is being presented from any of its transforms despite always experiencing all four stimuli together during training. Furthermore, the cells in the output layer can collectively identify each of the four stimuli across nearly all transforms, as indicated by the multiple cell information measure nearly reaching the theoretical maximum of two *bits*.



(a) Single Cell Information (b) Multiple Cell Information

Figure 5.15: Information plots of the trained and untrained network with four simultaneous training stimuli. For the trained network, the single cell information measure (a) shows that a few cells in the output layer are highly informative in discriminating between the stimuli, and the multiple cell information measure (b) shows that all four stimuli's transforms have been learnt by the network.

The autocorrelations for each of the four populations of input neurons representing each of the four stimuli are plotted in Figure 5.16. They each show a high correlation repeated approximately every 175 ms implying that at least one population of input cells fires approximately every 45 ms in a repeating cycle.

5.4 Discussion

This chapter has investigated how a model of the primate ventral visual system may develop neurons which are selective to a particular *individual* stimulus and respond invariantly to it over a set of learned transforms, even though the network is only exposed to visual scenes containing *multiple* stimuli. Previous work has investigated this issue using rate-coded neural networks whereby the precise times of

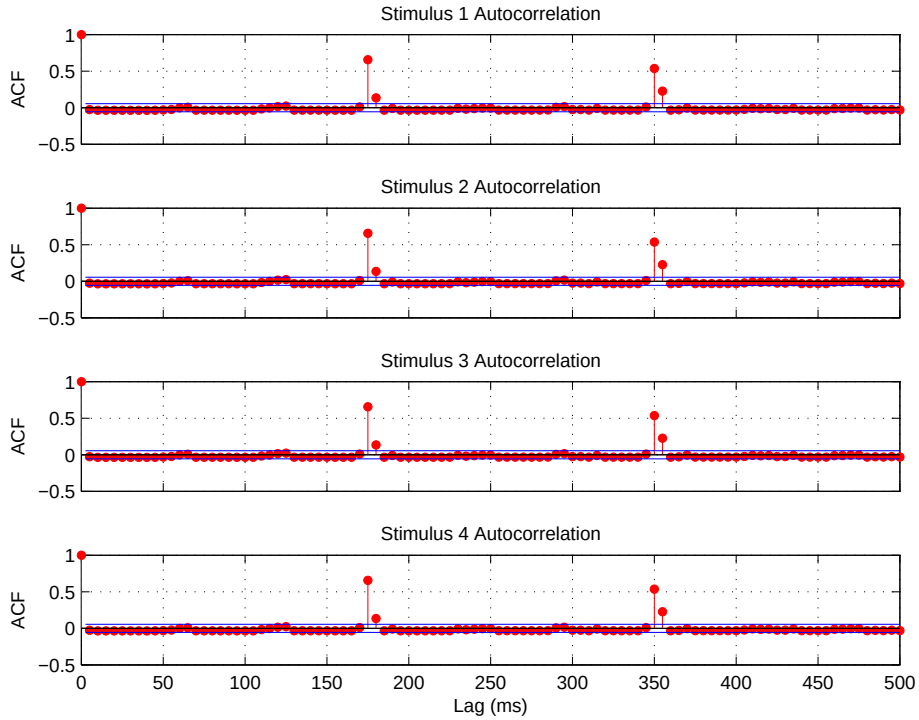


Figure 5.16: Autocorrelations for each of the four populations of input neurons representing each stimulus. The autocorrelations are significant (above the blue line) approximately every 175 ms giving them a combined period of approximately 45 ms, meaning that each stimulus is represented every fourth gamma cycle.

action potentials are not simulated explicitly but replaced by a temporal average. To overcome the 'superposition catastrophe' of associating the simultaneously presented stimuli together onto the same output neurons through a rate-based Hebbian learning rule, these earlier studies had to present the stimuli in different combinations on different learning trials (Stringer and Rolls, 2008; Stringer et al., 2007), or the stimuli had to be shown moving independently of each other during training (Tromans et al., 2012). In the presented work of this chapter, it has been demonstrated how by using a spiking neural network, (in which the times of the action potentials are explicitly modelled) output neurons can learn separate representations of visual stimuli which are always seen moving together in lockstep during training but which are separated in space.

Importantly, this model incorporates a SOM architecture within the input layer with cell firing-rate adaptation. The effect is that the short-range (exponentially declining) excitatory lateral connections help to synchronize localized clusters of input neurons (representing a stimulus). Conversely, the long-range (global) inhibition and firing-rate adaptation push the firing of one cluster out of phase with respect to the

other, causing them to oscillate through time due to the delayed self-inhibition effect of the firing-rate adaptation. Section 5.3.1 demonstrates the resulting effect which is to synchronize the firing of neurons representing a particular stimulus and desynchronize them with respect to the ensemble representing the other stimulus.

The alternating representations in the input layer which arise from the detailed properties of the network may then form the basis for learning about the stimuli individually. When combined with the temporal specificity of STDP in the feed-forward excitatory connections, this temporal separation between the oscillating input representations allows for different pools of output neurons to learn about each stimulus separately. Furthermore, this input layer dynamic persists as the stimuli transform (translate across the input layer) such that the output neurons also build transformation-invariant representations of the individual stimuli through the CT learning mechanism (Stringer et al., 2006; Evans and Stringer, 2012) as shown in Section 5.3.2.

In agreement with earlier work, the ability of the input layer to form the perceptual cycles of the individual stimuli (when presented simultaneously) was found to be be critically dependent upon a mechanism of delayed self-inhibition (Choe and Miikkulainen, 2003; Miconi and VanRullen, 2010) — in this case, cell firing-rate adaptation.

In addition to cell firing-rate adaptation, Section 5.3.5 demonstrated the importance of the lateral excitatory connectivity in generating perceptual cycles between the input representations. A SOM architecture is often taken to mean lateral connectivity with short-range excitation with long-range inhibition (Miikkulainen et al., 2005) which was modelled here through a gradual weakening of the excitatory lateral connections with increasing distance (according to a Gaussian profile) and uniform strength, fully laterally connected inhibitory interneurons (representing a long-range or global inhibitory mechanism). By flattening the profile of these lateral excitatory connections or removing them altogether, the perceptual cycles were extinguished and the input representations became disorganized and unable to facilitate transformation-invariant learning of separate object representations in the feed-forward connections of the next layer.

The importance of the lateral connectivity was further explored in Section 5.3.3, where the standard deviation of the lateral excitatory connection strength, σ_{ElE}, was systematically varied to assess the effect upon network performance. It was found that if this parameter was too small, then not all 'features' of a particular stimulus were synchronized — in other words, the lateral excitatory connections were unable

to promote coherence of the stimuli (intra-stimulus synchronization). Alternatively, with too large a spread of *ElE* strength, the neurons representing features of both stimuli are encouraged to fire in phase with each other, abolishing the perceptual cycles found with intermediate values. At each extreme, the disruption caused to the input representations had a negative impact on network performance, meaning separate transformation-invariant representations were less likely to form.

Furthermore, the ability of this model to produce output representations which are both selective to a particular stimulus and invariant across its transforms was found to be critically dependent upon a number of key properties. In section 5.3.4 the temporal specificity of the STDP learning rule was explored through systematically varying the time constants of the LTP and LTD time windows. If the time constants were too short, the output neurons were unable to learn transformation-invariant stimulus representations successfully as the learning rule became too sensitive to the timing jitter of the input spike volleys. Conversely, if the STDP time constants became too long, there was not enough temporal specificity to isolate the potentiation of a particular output neuron to just one stimulus and both stimuli became associated onto the same output neuron.

The great topmost sheet of the mass, that where hardly a light had twinkled or moved, becomes now a sparkling field of rhythmic flashing points with trains of travelling sparks hurrying hither and thither. The brain is waking and with it the mind is returning. It is as if the Milky Way entered upon some cosmic dance. Swiftly the head-mass becomes an enchanted loom where millions of flashing shuttles weave a dissolving pattern, always a meaningful pattern though never an abiding one; a shifting harmony of sub-patterns.

—Charles S. Sherrington

6 Axonal Delays

6.1 Introduction

Neurons in isolation are computationally very limited. The incredible cognitive properties of the brain come from their ability to communicate with, on average, 7000 other neurons. Spikes are propagated between neurons via an electro-chemical cascade along their axons and then a diffusion of neurotransmitter across the synaptic cleft (see Kandel et al., 2000). Keeping the delay (and metabolic cost) of neural communication to a minimum is important for rapid processing of information and generation of appropriate behavioural responses. Several strategies have arisen through the course of evolution to expedite communication, including the development of very thick axons found in the squid and wrapping axons in myelin sheaths found in mammals. However, axons may extend between cortical regions millimetres or centimetres apart, meaning there may be a significant delay in conducting these messages. This chapter aims to explore the computational implications for the formation of transformation-invariant representations of such axonal conduction delays, firstly in terms of robustness of performance with the standard learning mechanisms and secondly to investigate any additional computational benefits which delays may confer.

6.1.1 Cortical conduction delays

Recordings of axonal conduction delays in the mammalian brain vary across a considerable range. In early antidromic recordings in the rabbit cortex, latencies between

approximately 1 and 33 ms were measured from V1 to connected descending corticofugal neurons of layer 6 (Swadlow, 1992). In examining connections between the Locus coeruleus and the cerebral cortex of anaesthetized squirrel monkeys, even larger ranges of delays have been found, for example 29–132 ms in the axonal connections with the somatosensory cortex (Aston-Jones et al., 1985).

Overall, a considerable range of axonal conduction delays have been found throughout the mammalian brain. Specifically, the primate visual system has also been found to have pronounced differences within and between cortical layers. Feed-forward and feed-back projections between V1 and V2 in the monkey were found to have comparable conduction velocities of around 3.5 m/s resulting in median latencies of 1.1 and 1.25 ms respectively (Girard et al., 2001).

Interestingly, antidromic latencies from MT to V1 were recorded to be 1–1.7 ms, making them very similar to those from V1 to V2 despite the longer distance traversed between MT and V1 (Movshon and Newsome, 1996). Similarly, latencies from the thalamus to the cortex were found to be remarkably constant in mice at approximately 2 ms despite the wide variation in travelling distance (Salami et al., 2003). Histochemical staining revealed the degree of myelination to cause the variation in conduction velocities between different regions (Salami et al., 2003), potentially due to differences in the internodal distances between nodes of Ranvier (Carr and Konishi, 1990). The intriguing consequence is that interconnected areas in more distant locations in the brain are effectively brought into the same 'neural time-zone', enabling different types of processing to occur simultaneously in different specialized cortical regions.

In contrast to the impressively rapid flow of activity between layers, the lateral axons within a cortical area are typically unmyelinated, making their conduction velocities slower than myelinated axons by approximately an order of magnitude at around 0.1 m/s (Girard et al., 2001). This implies that processing of information may occur on several time scales, with myelinated feed-forward connections subserving very rapid communication between the layers and lateral connections modifying this processing, perhaps even before the feed-forward processing begins by manifesting top-down influences such as attention. Given the way in which evolution has uncovered a mechanism to equalize the conduction latencies between various regions of the brain by compensating for longer travelling distances through varying the conduction velocity over approximately an order of magnitude (Salami et al., 2003), it is reasonable to consider that there may be a functional role in leaving lateral connections unmyelinated and hence another order of magnitude slower.

Furthermore, the different time scales of processing seem to be properties preserved over the life of the neural circuit. The axonal conduction delays measured between a given pair of neurons have been found to be reproducible with sub-millisecond precision making them highly reliable as a feature of inter-neural communication (Swadlow, 1985). Even over distances of $10\,cm$, only a few microseconds of jitter are introduced (Rieke et al., 1999), a precision and reliability without which we could not detect phase differences corresponding to interaural time differences of less than $10\,\mu s$ (Rieke et al., 1999).

6.1.2 Computational functions of delays

One interesting computational use of axonal conduction delays is in the localization of sounds. The Barn owls (*Tyto alba*), as a nocturnal predator, relies upon its exceptional sense of hearing to locate its prey. The neural mechanism for this relies upon phase-locked spikes in coincident delay lines from each ear in the brainstem *nucleus laminaris* (Carr and Konishi, 1988) and the ability to detect time difference thresholds as low as $1\,\mu s$ (Rieke et al., 1999). Here, conduction delays change systematically with depth (tonotopic), providing a neuronal map of interaural time differences between the arrival of phase-locked pairs of spikes which may be detected by laminaris cells acting as coincidence detectors (Carr and Konishi, 1988). This was later modelled to show how the temporal precision of the system can exceed the time scale of the neuronal time constants (for example τ_m) and it was also demonstrated how a model of STDP could match delay lines appropriately through unsupervised learning to achieve the necessary degree of temporal coherence in spike arrival times (Gerstner et al., 1996).

In an analogous way to how delay lines in the *nucleus laminaris* of barn owls transforms a temporal code into a spatial code, the combination of STDP with synaptic delay lines has also been proposed as a mechanism for transforming temporal patterns of activity to be stored as memories encoded in the spatial patterns of synaptic efficacies distributed over a network of neurons Bi and Poo (1999). In this respect, the polysynaptic pathways of such interconnected networks may be regarded as delay lines with millisecond temporal precision and ranges of delays in proportion to the size of the network. Conduction delays, in this case traversing circuits of several synapses, provide alternative (variously circuitous) routes by which spikes may converge upon the same postsynaptic neurons. Since STDP is sensitive to the relative timing of pre- and postsynaptic action potentials, some synapses will be strengthened upon the early arrival of a spike through a 'short' route, while others at the end of a longer route will be weakened by the action potential arriving after the postsynaptic spike.

On the basis that spike timing encodes information, such a system thereby provides a mechanism for converting and storing this temporal information into distributed spatial information through time-sensitive synaptic modifications (Bi and Poo, 1999).

More complex dynamic interactions between axonal delays and STDP have also been considered and simulated, such as the idea of 'polychronization', which is defined as time-locked but not synchronous firing patterns with millisecond precision (Izhikevich, 2006). In simulations of Izhikevich spiking neural networks (Izhikevich, 2003, 2004), incorporating an STDP learning rule and axonal conduction delays led to spontaneous self-organization in neuronal groups, even in the absence of correlated inputs (Izhikevich et al., 2004). These groups compete with each other for prolonged survival, mediated by STDP, which tends to potentiate connections where presynaptic spike emission times match their axonal conduction delays, and suppress assemblies which impede this. Such neural assemblies are a robust phenomenon, may persist for long periods of time (Izhikevich et al., 2004) and far exceed the number of neurons in the network, since neurons may be shared between groups without ambiguity (due to their different firing orders in different groups). The stability and abundance of such polychronous neural circuits therefore has important implications for the memory capacity of networks featuring axonal conduction delays and STDP (Izhikevich, 2006).

A comparative study of variations in the Trace Learning rule found that significantly better network performance was achieved (i.e. higher information content in the output neurons) when the trace term used the previous time-step's trace ($\overline{y}^{\tau-1}$) and not the present trace activity ($\overline{y}^{\tau}$), described in Equations 1.10–1.11 (Rolls and Milward, 2000). Adding delays in the feed-forward connections may act in an analogous way, enhancing the network performance under a 'Trace-like' learning paradigm of orthogonal stimulus transforms and long feed-forward synaptic time constants.

6.1.3 Conditions for perceptual cycles

As explored in Chapter 5, temporally segmenting scenes with multiple stimuli enables a spiking neural network to avoid the superposition catastrophe which otherwise poses a significant problem for rate coded models of the visual system when exposed to natural visual scenes of several objects. In the earlier simulations, such perceptual cycles were found to occur through the spatial separations of stimuli together with a Gaussian profile in lateral excitatory weights.

Other investigations have found that robust synchronization of assemblies of neurons representing a particular stimulus ('feature linking') and the generation of an antiphase relationship between competing (input) representations may also

be achieved through axonal delays (Miconi and VanRullen, 2010; Nischwitz and Glünder, 1995). In general four regimes have been studied and shown to account for synchrony/desynchrony in a single population of principal cells (Nischwitz and Glünder, 1995): (1) mutual excitatory connections without delays cause synchrony (quickly); (2) mutual excitatory connections with delays cause desynchrony; (3) excitatory to inhibitory connections without delays cause desynchrony; and (4) excitatory to inhibitory connections with delays cause synchrony (slowly) as summarized in Table 5.1. Figure 6.1 (from Nischwitz and Glünder, 1995) illustrates the ways in which the delays and polarity of the synapses interact to produce either synchronous or asynchronous firing between two neurons.

In their analysis of conditions for synchronous firing among a pair of (leaky integrate-and-fire) principal cells, Nischwitz and Glünder (1995) determined that with mutually excitatory coupling, no transmission delay lead to synchrony whereas adding a $4\,ms$ conduction delay lead to 'counter-phase' firing (with synchronization reappearing for delays which are approximately multiples of the interspike interval of an uncoupled unit). They also ascertained that if instead the coupling was inhibitory, then the opposite dynamics were produced with respect to the presence/absence of transmission delays — that is, with no delay, the counter-phase state is attractive (and the synchronous state is repellent) whereas with a $4\,ms$ delay, the synchronous state is attractive (but interestingly, so is the counter-phase state if it happens to exist).

In more recent work Miconi and VanRullen (2010) applied these dynamics to show how a spiking neural network can segment a visual scene by organizing the input representations of different objects into counter-phase representations they called 'perceptual cycles'. They used a fixed delay to a globally connected inhibitory interneuron (rather than direct coupling through inhibitory synapses) with lateral excitatory connections between principal cells with similar orientation preferences to support contour integration. Crucially, their model incorporated an after-hyperpolarization current to model cell firing-rate adaptation. The excitatory lateral coupling between cells of similar orientation preference (without delays) helped to synchronize the firing of smooth contours which were most likely to be the same object. Initially the delayed inhibitory connections tended to synchronize the whole population in agreement with the two cell fixed point analysis of Nischwitz and Glünder (1995), however, the more targeted excitatory connections and cell firing-rate adaptation meant that populations of neurons representing different objects (which did not share the synchronizing lateral excitatory stimulation) settled into a counter-phase oscillation.

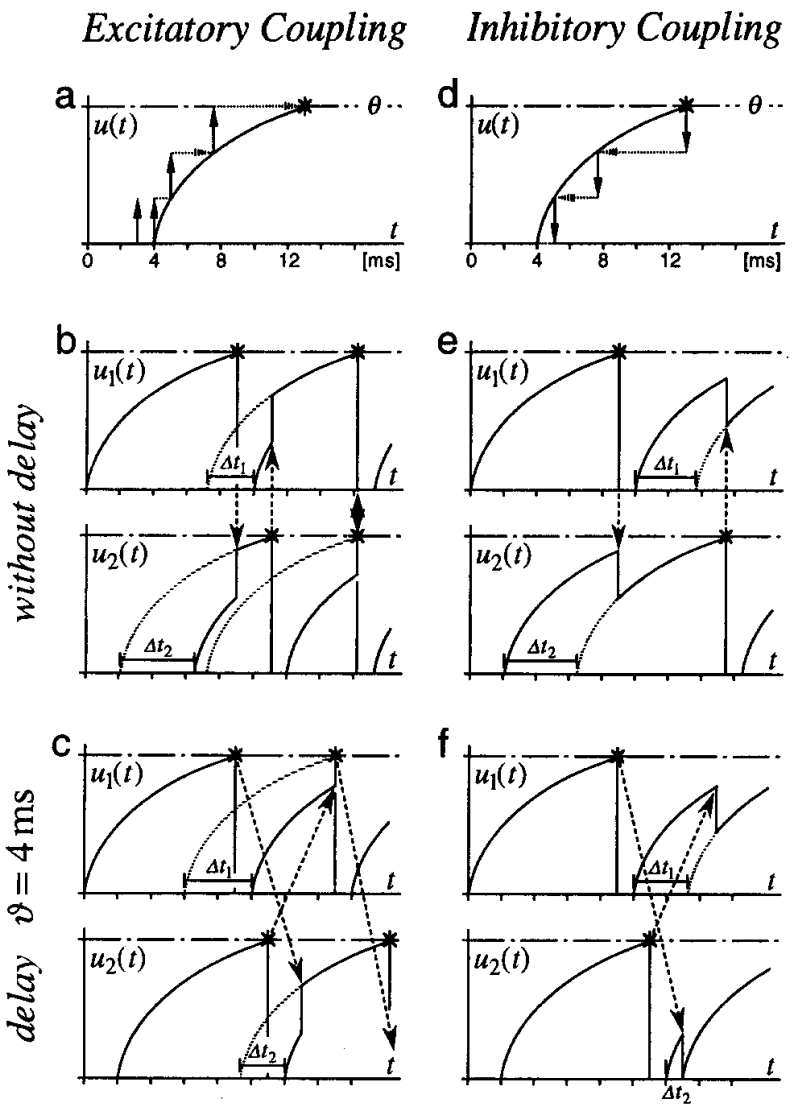


Figure 6.1: Four types of δ-impulse coupling between two idealized 'leaky integrate-and-fire' units and their synchronization behaviour. The units' integration characteristic and threshold θ with time shifts of their spike triggering for various step-like changes of the potential, (a & d). Examples of first interactions between excitatory coupled units without and with transmission delay that produce synchronous and counterphase firing respectively, (b & c). Examples of first interactions between inhibitory coupled units without and with transmission delay that produce counterphase and synchronous firing respectively, (e & f). Reproduced from Nischwitz and Glünder (1995), Figure 1 with kind permission from Springer Science and Business Media, Copyright © 1995, Springer.

In this chapter the simulations complement this earlier study by investigating delays in the excitatory connections (rather than inhibitory connections). By considering the two-neuron system of Nischwitz and Glünder (1995), we extend the ideas to a continuous spread of excitatory axon delays modelling the more realistic scenario of a fixed conduction speed and range of distances across a layer (that is delays proportional to Euclidean distance). When the distances are small between input neurons representing the same stimulus, this is expected to produce synchrony among this population of neurons, corresponding to the case of $E \leftrightarrow E$ with no delay (as some finite but small delay tends to 0). Conversely, cortical distances between neurons representing different stimuli will be much larger, and therefore have a larger conduction delay which should therefore encourage a counter-phase relationship between the two competing populations.

In summary, the hypothesis for this work is that if lateral axonal conduction delays are proportional to Euclidean distances in a laterally connected layer, this should result in synchrony over short distances approximating the situation of $E \leftrightarrow E$ with 'no' delay, and desynchrony between more distant features (such as another object) as in the case of $E \leftrightarrow E$ connectivity with a conduction delay. Furthermore, the dynamics resulting from this model's architecture should form perceptual cycles which enable a second layer of principal cells to learn through STDP in their feed-forward connections to form separate transformation-invariant representations of each stimulus (despite no distinction between the input populations representing the two stimuli in terms of their firing rates).

In the first section of this chapter's results, the original CT and Trace learning simulations are revisited with fixed or distributions of delays in the feed-forward excitatory axons (EfE) to investigate the basic consequences of axonal conduction delays on these two mechanisms in a highly idealized way. In the second section of results, delays are implemented in the lateral excitatory connections (ElE) to explore their effects upon the formation of perceptual cycles as a means for temporally segmenting multiple objects within a visual scene.

6.2 Methods

6.2.1 Model Architecture

In this chapter, the effect of axonal conduction delays are investigated, first in feed-forward (EfE) excitatory connections and then in lateral (ElE) excitatory connections. In line with earlier simulations, a slightly different model architecture is used for each

section with the basic model in Chapter 3 corresponding to the results presented here on feed-forward delays and the architecture used in Chapter 5 corresponding to this chapter's explorations of lateral delays. Importantly however, in the section on lateral delays in this chapter, it is the lateral *delays* which are based upon Euclidean distance rather than the lateral *synaptic efficacies* (as was the case in Chapter 5).

Both sections use two layers of excitatory neurons connected by feed-forward plastic synapses which are modified through ten training epochs according to a spike-time dependent (STDP) learning rule described in Section 2.3.1 by Equations 2.27–2.29 (Perrinet et al., 2001). As in Chapter 3, each layer consists of 400 excitatory neurons for the feed-forward delay simulations and as in Chapter 5, the model consists of two layers, comprising 512 input and 256 output neurons for the lateral delay simulations. Each layer of the model has its own pool of fully connected shunting inhibitory interneurons, where the number of inhibitory cells is exactly 1/4 of the number of excitatory neurons in the corresponding layer to provide a means of competition between the principal (excitatory) cells.

In the feed-forward delay simulations, each layer features all-to-all connectivity as described in Table 3.1. For lateral delay simulations, excitatory neurons within each layer were connected to other excitatory neurons according to the probability $p(ElE) = 0.5$ (disallowing self-synapses) with the synaptic weights Δg_{ij}^{ElE} initialized uniformly to fixed efficacies (changes in conductance) of $1.25\,nS$ per spike (whereas in the original work of Chapter 5, they decreased with increasing distance according to a Gaussian profile). Similarly, EI and IE synapses were not plastic but instead initialized to $\Delta g^{EI} = \Delta g^{IE} = 5\,nS$. The distance between two excitatory neurons within a layer, d_{ij}, is calculated as in the previous work (Equation 5.4) thus implementing periodic boundary conditions to give the input layer a ring topology. The axonal conduction delays in the lateral connections are then calculated according to Equation 6.1 where Δt_{ij} is the conduction delay in transmitting a spike from neuron j to neuron i and v_a is the axonal conduction speed.

$$\Delta t_{ij} = \frac{d_{ij}}{v_a} \tag{6.1}$$

Note, this embodies a simple linear relationship between distance and delay, unlike the exponential decrease in synaptic efficacy with increasing distance modelled in the original simulations of Chapter 5.

For convenience and to avoid specifying distances arbitrarily in *neurons* (and speeds in *neurons/s*), the delay between the centres of the two stimuli was specified, from which the conduction speed was calculated based upon the spatial scale ($S_{\{x,y\}}$)

of the largest layer as shown in Equation 6.2. Given the periodic boundary conditions of the model and the arrangement that the two stimuli were evenly spaced upon the input layer, the conduction delay between them was equal to the maximum lateral delay in the ring-shaped layer, Δt_{max}, (in the case of two stimuli) or equivalently, half the delay between the two ends of the equivalent layer without periodic boundary conditions.

$$v_a = \frac{\sqrt{S_x^2 + S_y^2}}{2 \cdot \Delta t_{max}} \tag{6.2}$$

6.2.2 Neuron model description

The same conductance-based leaky integrate-and-fire neurons (with cell-membrane noise) were used throughout all simulations in this chapter with the default parameters given in Table 2.2 and Table 2.3 mostly taken from Troyer et al. (1998). In the simulations with lateral delays and the presentation of multiple stimuli during learning, cell firing-rate adaptation was incorporated into the model as before (Chapters 5 and 7), described in Equations 5.1–5.3, whereas this feature was not used in the simulations with feed-forward delays (in keeping with the work of Chapter 3).

The simulations investigating feed-forward excitatory axonal delays re-examine both the CT and Trace learning mechanisms and so have appropriate excitatory synaptic time constants of $\tau_g^{EfE} = \{2, 150\}\, ms$ respectively. In contrast, only the CT learning mechanism was investigated with lateral delays, and so the feed-forward synaptic time constant was fixed at $2\, ms$ as in Chapters 5 and 7 (Troyer et al., 1998).

6.2.3 Stimuli and Training

Stimuli are represented by injecting tonic current into particular discrete sets of input layer neurons for each stimulus. The contiguous blocks of neurons which represent a particular transform of a stimulus are moved across that stimulus's section of the input layer through time, representing successive transforms of that stimulus. For all simulations, the untrained networks were first tested in a 'pre-training' phase by presenting all transforms of all stimuli sequentially, to provide an untrained baseline response from which to gauge the effects of the learning process during the subsequent training phase (by comparison with the responses from the final 'post-training' phase).

In the feed-forward delay simulations, stimuli are presented individually during training. When CT learning is used, each transform of each stimulus consists of

56 contiguous neurons shifting by 12 neurons per transform for a total of 13 transforms per stimulus. This provides a large overlap between consecutive transforms of approximately 78.6% for the CT learning mechanism to utilize. When the Trace learning mechanism is used (employing longer excitatory synaptic conductance time constants), transforms of each stimulus are represented by a block of 20 neurons and are orthogonal to one another, meaning they shift by 20 neurons per transform for a total of 10 transforms per stimulus. By default, the transform presentation period during training in these simulations was $100\,ms$.

For simulations of lateral conduction delays, the two stimuli are presented simultaneously translating across the input layer in a randomly chosen direction during training, for $500\,ms$ per transform, as in Chapter 5. Each stimulus was represented by 64 neurons with each transform shifting by 16 neurons across a total of 13 transforms, thus giving an overlap of 75% between contiguous transforms for the operation of the CT learning mechanism. The presentation of all transforms of the compound stimulus (consisting of the pair of individual stimuli) constituted one epoch of training, with the training phase consisting of ten epochs in total.

During the testing phases for both sets of simulations, transforms of each stimulus are always presented individually in sequence (for both feed-forward and lateral delay simulations) with the cell dynamic variables being reset between transforms in order to measure the network's responses to each transform of each stimulus in isolation. The transform presentation period was $250\,ms$ for feed-forward delay simulations and $500\,ms$ for lateral delay simulations. After presenting all transforms of all stimuli, the responses were saved and analysed.

6.3 Results

The first sets of results revisit the simulations of Chapter 3 with the addition of feed-forward axonal conduction delays. The results obtained with the modified form of the STDP rule which has been used throughout this thesis are also compared to those obtained with the original form which considered only the spike emission time rather than the arrival time at the synapse, thus neglecting any axonal conduction times (Perrinet et al., 2001).

The second set of results presented here have axonal conduction delays in the lateral connections and a multiple stimulus presentation paradigm. The work here explores how these delays instead of a gradient in the lateral weight profiles may

organize the input representations into perceptual cycles such that output neurons may learn to form independent transformation-invariant representations.

6.3.1 CT learning with Delays

The simulation results presented in this section investigate the effects of implementing axonal conduction delays on the original transformation-invariance results with the CT learning mechanism. The network performance was first explored with uniform conduction delays, which was then extended to consider a Gaussian distribution of axonal conduction delays by varying $\sigma_{\Delta t}$ around a fixed mean.

6.3.1.1 Uniform conduction delays

The STDP model used throughout this thesis was based largely upon Perrinet's original formulation, however the equations were modified to include an axonal delay term for effects upon the synapse specific variables (C_{ij} and Δg_{ij}). This change was in accordance with the original experimental data (Bi and Poo, 1998) and is based upon the rationale that synaptic effects due to action potentials (for example, vesicle release) can not occur until the action potential has propagated from the cell body to that particular dendrite.

Due to this modification, the presynaptic spike event is determined by the time that the spike arrives at the *synapses* rather than the time that the presynaptic cell body depolarizes. As such, any given synapse implementing this form of STDP is 'agnostic' with regard to transmission delays in the axons leading up to it. It therefore follows that the modifications in synaptic efficacy occurring due to this modified form of STDP should not be affected by fixed uniform axonal delays. This prediction was borne out by the computer simulations shown in Figure 6.2 which plots the network performance measure based upon the output neurons single cell information[1] (Equation 5.6) against a range of fixed axonal conduction delays.

It is clear from the analysis of network performance before and after training (Figure 6.2) that the information conveyed by the firing rates of neurons in the output layer has increased after training and does so fairly uniformly across the full range of axonal conduction delays explored $[0, 50]\,ms$. As such, constant feed-forward axonal conduction delays have no effect upon the network performance after CT learning

[1]The single-cell information of each neuron to each stimulus was calculated and the number of cells counted which conveyed at least 95% of the of the information limit (in this case $0.95\,bits$). The minimum number of such cells was then found, computed across the stimuli, s, as a measure of how much information the network conveyed about all transforms of the least well represented stimulus, expressed as a proportion.

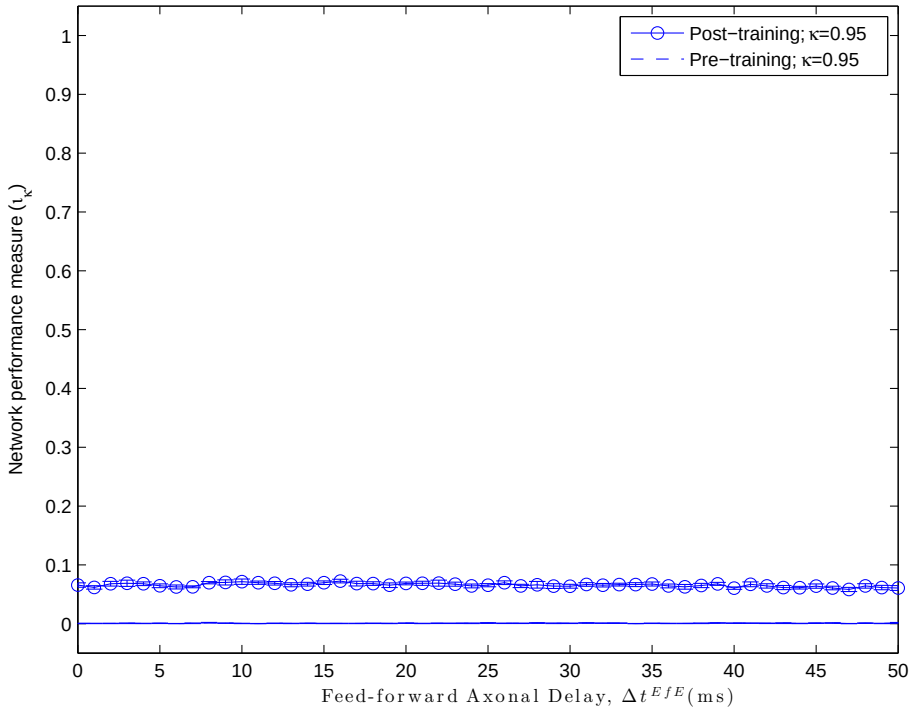


Figure 6.2: The information analysis summary (with the corresponding pre-training values plotted with broken lines) for a range of fixed feed-forward axonal conduction delays for CT learning. Each value of Δt_{EfE} was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean. Due to the modified STDP learning rule used through these simulations, the performance of the network is not appreciably affected by constant, uniform axonal delays.

when trained with the modified STDP learning rule used throughout this thesis (Section 2.3.1).

For comparison, the same model was run with the same parameter range using the unmodified (original) form of STDP (Perrinet et al., 2001) which, like many spiking neural networks without axonal conduction delays (e.g. Song et al., 2000; Masquelier et al., 2008; Hennequin et al., 2010), updates the synapses on the basis of the presynaptic spike emission times (in the cell body) rather than their arrival times at the particular synapse in question. Essentially, the change made to revert back to the original STDP form was to remove the Δt_{ij} terms from Equations 2.27 and 2.29 and therefore the corresponding terms in the finite difference equations. However, the delay term in the synaptic conductance model (Equation 2.19) was unchanged, meaning there was still an axonal conduction delay in effecting the excitatory impulses at each synapse.

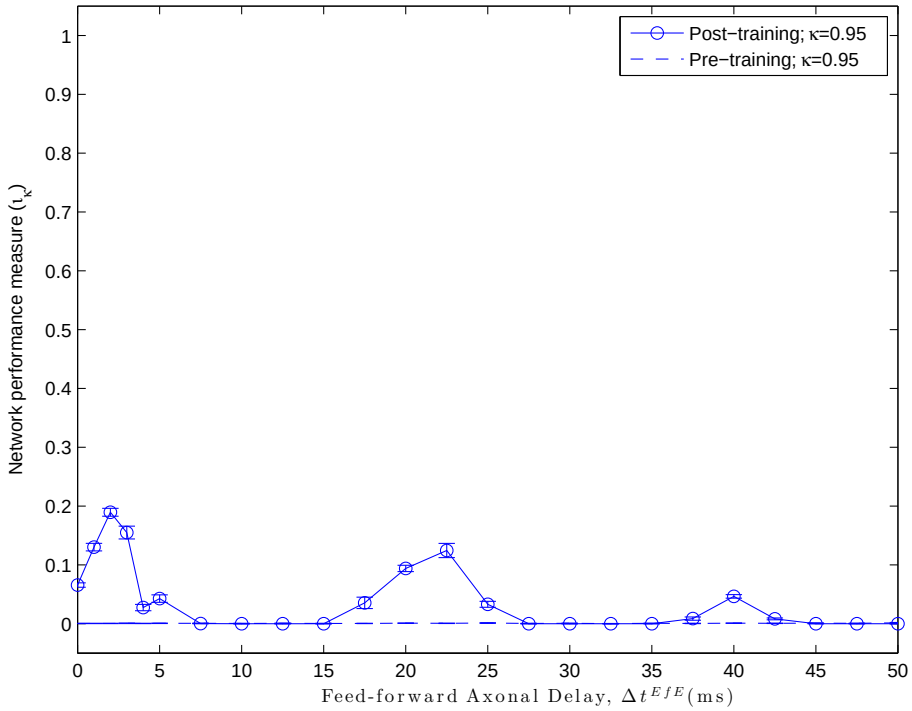


Figure 6.3: The information analysis summary (with the corresponding pre-training values plotted with a broken line) for a range of fixed feed-forward axonal conduction delays for CT learning with the original (delay-independent) STDP model. Each value of Δt_{EfE} was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean.

The network performance measure is plotted in Figure 6.3 for a range of constant axonal delays, (which delayed only the timing of changes in synaptic conductances, not the plasticity variables). In contrast to the simulations with the modified STDP form where axonal conduction delays are explicitly featured, (Figure 6.2) the network performance is not constant across the range of delays. Instead, the performance starts off at the same level (since 0-delay is the same with either model) and then declines as the delay increases. This is as expected since the changes in the presynaptic plasticity variable C_{ij} and the changes in activity causing postsynaptic spikes (EPSPs) become further apart in time. Interestingly however, the performance increases again in the region of approximately every $20\,ms$ before declining again.

From examination of the input layer activity, the reason for the periodic fluctuations in network performance becomes clear. The spiking activity of the input layer excitatory cells is very regular as shown by the raster plot and PSTH of Figure 6.4. The volleys of spikes representing a stimulus transform are synchronized and repeat

approximately every $20\,ms$. This regular period of activity is confirmed by the Fourier spectrum (periodogram) in Figure 6.5 which shows a dominant frequency at approximately $51\,Hz$. The consequence of this regular periodicity is that while the delay increases from 0–20 ms, the increase in C_{ij} becomes further apart from any subsequent postsynaptic spike occurring due to the delayed changed in the excitatory conductances at the synapses. At approximately $20\,ms$ (with a tolerance of a few milliseconds beyond) the postsynaptic spike (caused by the first volley of presynaptic spikes) now occurs just after the second volley of presynaptic spikes, which therefore have high values of C_{ij} and therefore lead to strong positive changes in synaptic efficacy.

In general, these resurgences of network performance can be seen to occur when the axonal conduction delays match (multiples of) the period of the input layer oscillations. At delays of $\Delta t_{ij} = f \cdot \rho\,ms$ (where $f \in \{0, 1, 2, ..., n\}$ and in this case, the oscillation period in the input layer, $\rho \approx 20\,ms$), the high values of C_{ij} due to presynaptic spikes at those times coincide with the arrival of delayed presynaptic spikes (generated $20\,ms$ earlier). These presynaptic spikes then cause postsynaptic spikes moments later and subsequently lead to positive weight changes. These weight changes are, however, acausal (in the respect that the spikes which change the presynaptic plasticity variables leading to LTP are not the same spikes which caused the postsynaptic spikes) and so violate Hebb's original postulate. The network performance improves because of the repetition of input spike volleys, although there may also be a Trace effect beginning to aid the learning process. If learning were on the basis of only one pairing of pre- and postsynaptic spikes per transform, or if the pairings were sufficiently spread out such that the period between volleys was much longer than the axonal conduction delays then these artefactual improvements would disappear and network performance would decline with increasing delay and stay low.

Another curious aspect of Figure 6.3 is that the performance with a $2\,ms$ delay is better than when using the modified STDP rule (Figure 6.2). This is likely to be a small Trace effect beginning to operate.

6.3.1.2 Gaussian distribution of conduction delays

From earlier simulations, it is clear that synchrony of the ensembles of neurons representing a particular stimulus is key for the CT mechanism to work. The inhibitory interneurons in the input layer encourage these neural assemblies to fire at the same time but have no further effect upon the subsequent spike arrival time at their efferent synapses, which is dependent on the conduction delay for the particular

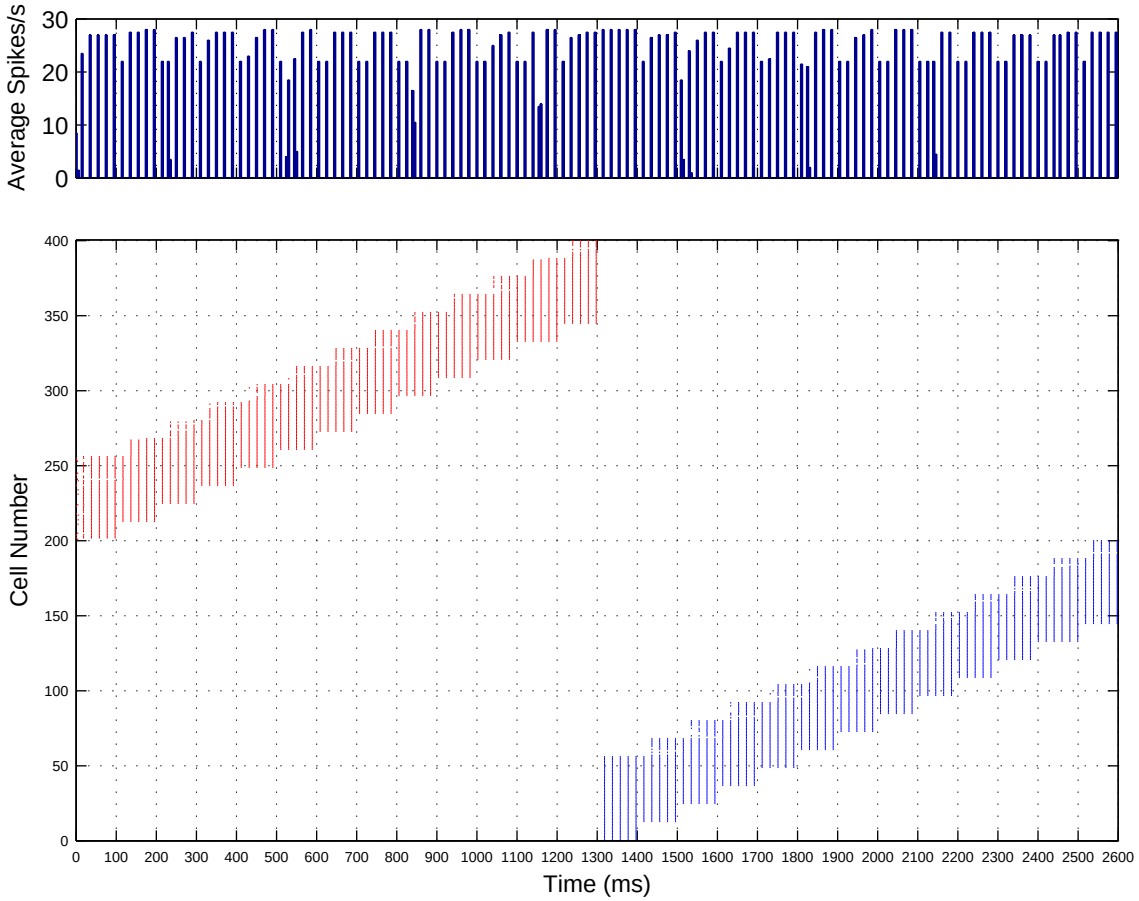


Figure 6.4: Input layer activity during training with individually presented stimuli. In the raster plot (Bottom) it can be seen that the two stimuli are presented sequentially, translating across their halves of the input layer. Stimulus 1 (shown in blue) occupies neurons 1–200 while Stimulus 2 (shown in red) occupies neurons 201–400. Both stimuli can be seen to have tightly synchronized volleys of spikes for each of their transforms repeating approximately every $20\,ms$. The post-stimulus time histogram (Top) also illustrates the regularity of the input layer spiking activity.

axon in question. If axonal conduction delays are drawn randomly from a Gaussian distribution, $\sim \mathcal{N}(\overline{\Delta t}, \sigma_{\Delta t})$, then there will be a spread in the spike arrival times at the postsynaptic neurons. The inhibitory interneurons in the output layer would also act to encourage synchronized firing of the excitatory cells in that layer, but are expected to struggle to overcome the effects of an increasingly large spread of spike arrival times. The spread of spike arrival times at the excitatory output cells would therefore mean that some would arrive before the postsynaptic cell fired meaning those synapses would undergo LTP, while others would arrive after the postsynaptic spike and so undergo LTD.

When this situation is coupled with STDP, it has been shown to effect the useful

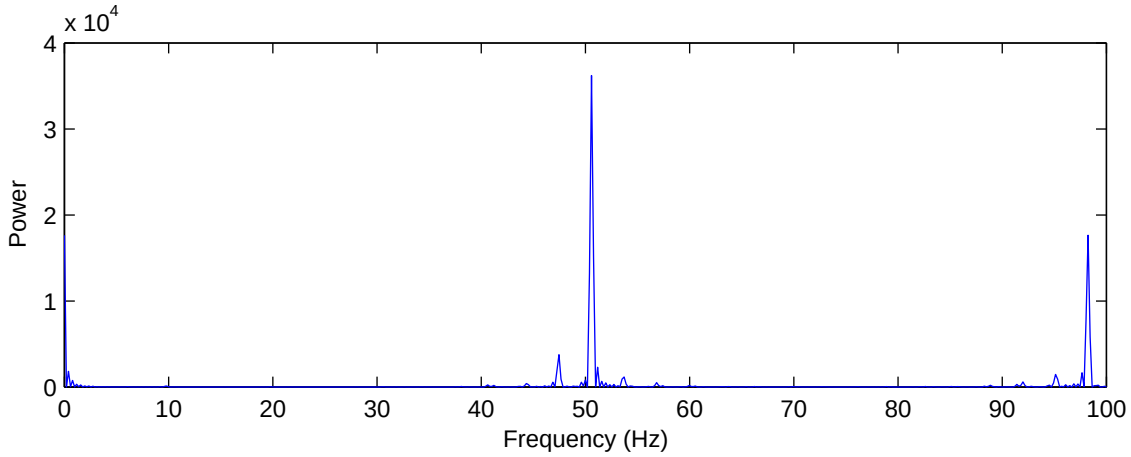


Figure 6.5: Input layer Fourier spectrum during training. It can be seen that the input layer activity is fairly constant at a frequency slightly higher than $50\,Hz$ as indicated by the large peak in the power at that frequency. The activity is fairly regular too, since there is little power in other parts of the spectrum.

property of reducing the latency of processing (Song et al., 2000), as inputs which arrive before the postsynaptic spike will be strengthened and further enhanced with repetitions of the stimulus, encouraging the postsynaptic cells to fire even earlier. While this is advantageous for increasing the speed of information processing, particularly in extremely fast feed-forward networks such as the ventral visual stream (Thorpe et al., 1996), if the spread of spike arrival times becomes too large, there may not be enough early arriving presynaptic spikes for each transform to cause the postsynaptic neurons to fire in a 'toy' network such as this. It is therefore expected that an increasingly large spread of axonal conduction delays would lead to degraded performance of the network.

In this section of simulations, the same 'CT paradigm' was used as in the previous section, however repeated simulations were run with axonal delays drawn from a Gaussian distribution with a fixed mean, $\overline{\Delta t} = 10\,ms$, and a systematically varied range of standard deviations, $\sigma_{\Delta t}^{EfE} = [0, 10]\,ms$. All simulations over this range were repeated for a total of ten random seeds and the mean performance measure plotted with the standard error of the mean across random seeds displayed by the error bars. In these and all further simulations, the modified (delayed) form of STDP was used in keeping with all other chapters unless explicitly stated otherwise.

The results are graphed in Figure 6.6 and, as hypothesized, show a clear degradation of performance with an increasing spread of axonal conduction delays. The proportion of perfectly invariant cells in the output layer has dropped to practically 0 (the level of

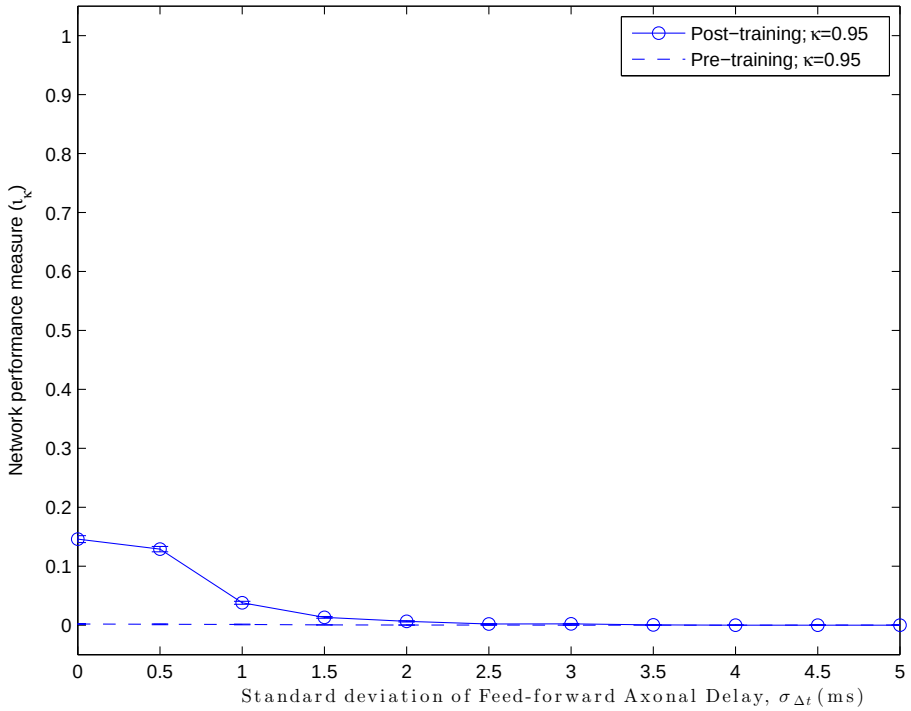


Figure 6.6: The information analysis summary (with the corresponding pre-training values plotted with broken lines) for a range of standard deviations of a Gaussian distribution of feed-forward axonal conduction delays for CT learning. Each value of Δt_{EfE} was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean. CT learning performs best with synchronized input spike volleys, which are disrupted by increasingly large spreads of conduction delays, resulting in a decrease in network performance.

the untrained network) with a standard deviation of approximately $2.5\,ms$. Although, reasonably good performance is still obtained with a small spread of conduction delays up to around $1\,ms$. CT learning thus has limited tolerance to variability in axonal delays between layers.

6.3.2 Trace learning with Delays

In this section, the same one-layer feed-forward model is used from the Trace simulations of Chapter 3 featuring longer excitatory synaptic time constants ($\tau_g^{EfE} = 150\,ms$) and orthogonal sets of transforms in order to investigate the effects of conduction delays on learning transformation-invariant representations with the Trace learning mechanism. As with the previous CT learning section, results are first presented from a uniform distribution of conduction delays and then a Gaussian distribution.

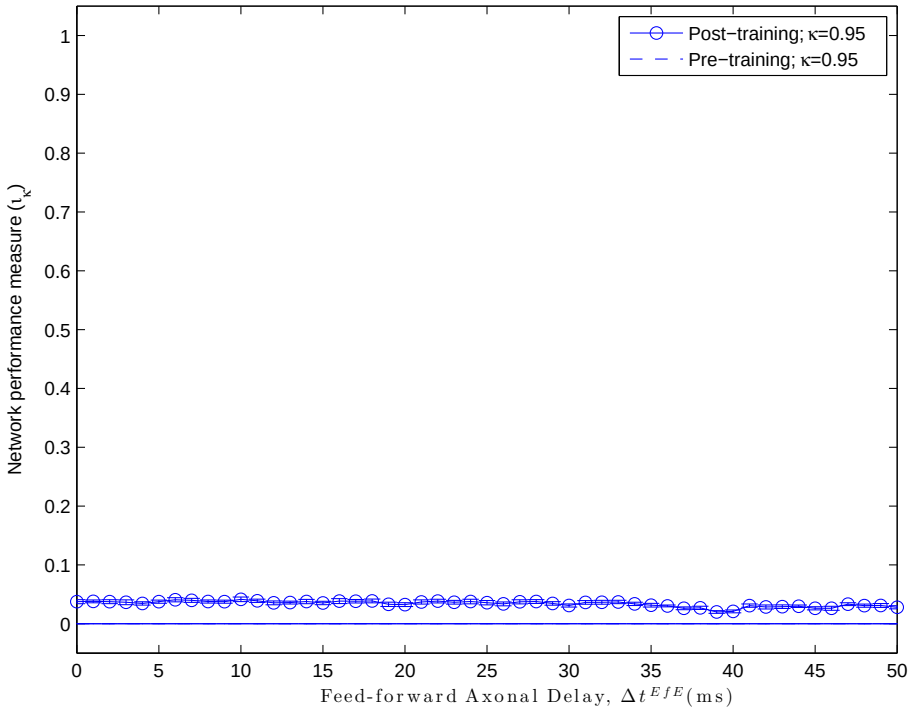


Figure 6.7: The information analysis summary (with the corresponding pre-training values plotted with broken lines) for a range of fixed feed-forward axonal conduction delays for Trace learning. Each value of Δt_{EfE} was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean. Like CT learning, Trace learning is largely unaffected by a constant conduction delay for the purposes of forming transformation-invariant representations.

6.3.2.1 Uniform conduction delays

By the same rationale as for considering fixed delays with a CT learning paradigm, the fixed axonal conduction delays should have no significant effect on the trained network performance with a Trace learning simulation paradigm. This hypothesis was tested by simulating a range of fixed axonal conduction delays from 0 to 50 ms for ten different random seeds. The results, as expected show no degradation (or enhancement) of trained network performance across the range of conduction delays tested and are shown in Figure 6.7.

As with the CT learning mechanism, constant feed-forward axonal conduction delays have not appreciably affected the ability of the Trace learning mechanism (namely long synaptic conductance time constants) to successfully train the network to form transformation-invariant representations.

The same simulations were rerun (using the same random seeds and parameters)

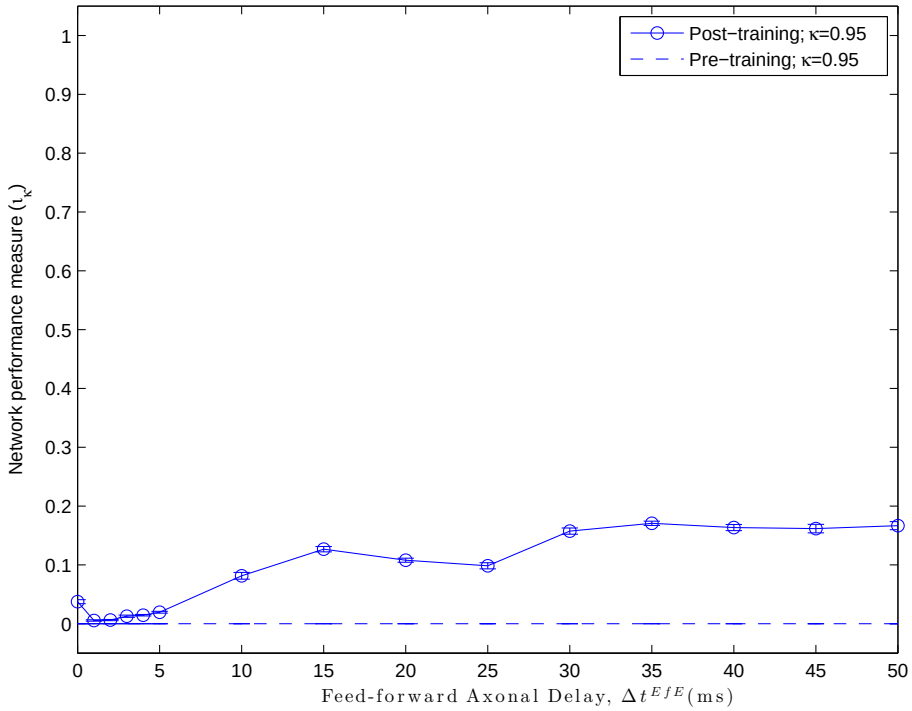


Figure 6.8: The information analysis summary (with the corresponding pre-training values plotted with broken lines) for a range of fixed feed-forward axonal conduction delays for Trace learning with the original (delay-independent) STDP model. Each value of Δt_{EfE} was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean.

after changing the STDP model to the unmodified form (by removing axonal conduction delays in the STDP equations as before). The results are presented in Figure 6.8 which shows how the network performance varies with axonal conduction delays from 0 to 50 ms.

Unlike, the CT learning mechanism, the Trace learning mechanism's performance does not deteriorate with increasing delay when using the original STDP model. On the contrary, the performance can in fact be seen to improve in Figure 6.8 as the delays become longer. Since the Trace learning mechanism used in this work relies upon long excitatory synaptic conductance time constants (τ_{EE}), the presynaptic spikes are effectively low-pass filtered and the changes they cause in the postsynaptic cell's membrane potential are prolonged and smoothed out through time, helping the neurons representing a recently presented transform become associated onto the same postsynaptic cells as those representing the present transform. As such, the Trace learning mechanism is much less sensitive to the precise timing of spikes and so does

not perform worse as the delay increases. The delay actually enhances the Trace effect since the presynaptic spikes which cause the postsynaptic spikes become further apart from the time of the synaptic updates, helping to associate events further apart in time, which in these simulations, are different transforms of the same stimulus.

6.3.2.2 Gaussian distribution of conduction delays

While a CT learning paradigm in a spiking neural network performs best with short plasticity time constants and tightly synchronized volleys of spikes from ensembles of input neurons, Trace learning operates best in a less temporally specific manner (both with respect to the synaptic conductance time constants and the plasticity time constants). The exact timing of the spikes is therefore less crucial to the self-organizing behaviour under this regime since the longer synaptic time constants serve to smooth out the effects of presynaptic spikes in order to maintain a recent history of activity over a longer time window. By being more tolerant to precise spike timings, it therefore follows that a Trace learning paradigm should be more tolerant to wider spreads of axonal conduction delays than the CT learning mechanism.

In preliminary CT simulations with a Gaussian delay distribution, it was found that network performance would begin to improve again for larger standard deviations. This was found to be artefactual however, caused by the clipping of the delay distribution at zero (since negative delays are impossible) which resulted in a growing population of zero-delay axons which could be enough to successfully train the network. Since the Trace learning mechanism was believed to be more robust to a spread of axonal conduction delays, the mean of the distribution was increased to $50\,ms$ (and the stimulus training presentation period was increased to $500\,ms$ to compensate for the correspondingly lesser effective training time) in order to avoid this artefact.

It is clear from Figure 6.9 that, as predicted, the Trace learning mechanism is far more robust than CT learning for distributions of axonal conduction delays. This is as expected since the Trace learning mechanism in this model relies upon long time constants for synaptic conductance in the feed-forward excitatory connections, which effectively 'smooths' the effects of action potentials over time, thus making the learning process less sensitive to their precise times of occurrence.

6.3.3 Delays in lateral connections

Having examined the effects of delays in feed-forward connections, this section investigates the effects of axonal conduction delays in lateral excitatory connections. The

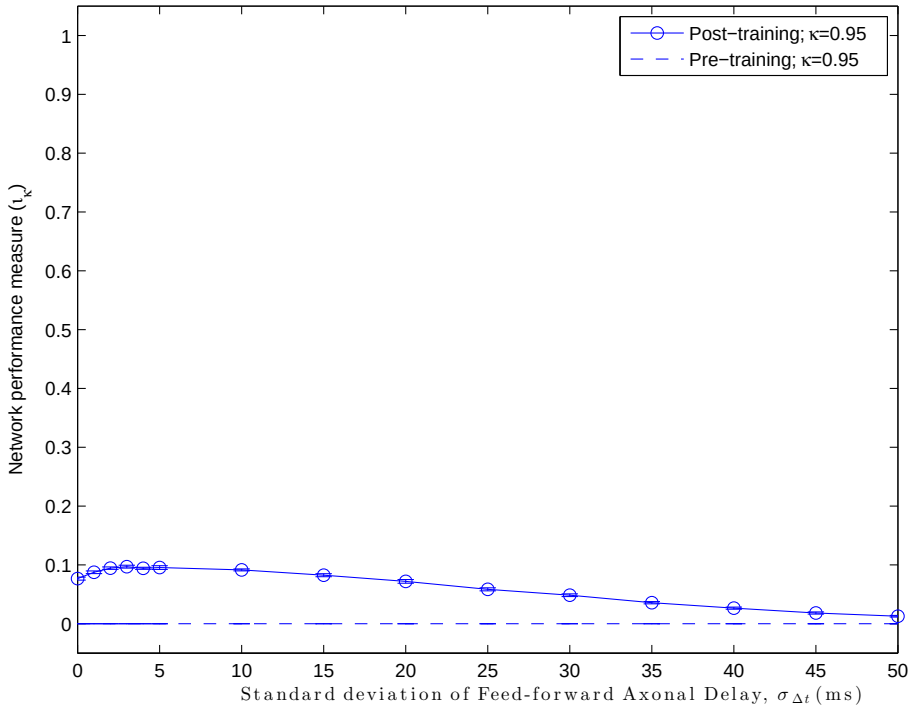


Figure 6.9: The information analysis summary (with the corresponding pre-training values plotted with broken lines) for a range of standard deviations of a Gaussian distribution of feed-forward axonal conduction delays for Trace learning. Each value of Δt_{EfE} was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean. Trace learning is shown to perform best with a small spread of conduction delays but is a highly robust mechanism for extremely large spreads of conduction delays as demonstrated by the slow decline of network performance with increasing $\sigma_{\Delta t}^{EfE}$.

aim of this section is to investigate the potential role which delays proportional to distance may have within a layer in order to organize input representations of multiple objects into perceptual cycles. As such, the model presented here uses mostly the same parameters and multiple stimulus presentation training paradigm as was used previously in Chapter 5. The main difference is that the lateral excitatory synaptic efficacies are fixed rather than decreasing exponentially with distance, and axonal conduction delays are introduced into the lateral excitatory connections instead, which are proportional to the Euclidean distance between the neurons (that is, a constant conduction speed is assumed).

In preliminary simulations, the strength of these fixed uniform lateral conductances was explored. For stronger initial lateral excitatory weight scaling, the input representations are still organized into perceptual cycles but when only one stimulus

is presented during testing, neurons in other parts of the input layer (which normally represent the other stimulus) are also stimulated. This 'ghosting' effect therefore reduces the apparent information in the output neurons for this parameter set (as neurons representing the other stimulus are encouraged to fire while one is being presented). This problem could also be alleviated however by lowering the strength of feed-forward connections or otherwise effectively raising the firing threshold for the output neurons.

If the scaling of the lateral connection strengths became too low, the lateral excitatory connections became ineffective and unable to organize the input layer representations into perceptual cycles. As a balance between these two ends of the scale, the strength of the fixed uniform lateral conductances was set at $1.25\,nS$ throughout the following simulations. During training, transforms of the two stimuli were presented simultaneously moving in lock-step for $500\,ms$ per transform (but were presented individually during testing for the same duration).

6.3.3.1 Input layer representations

In Figure 6.10, transforms of the two stimuli can be seen presented together during training, shifting across the input layer. It can be seen from the raster plot (Bottom) that the neurons representing the transforms of a particular stimulus emit synchronized volleys of spikes periodically. This 'internal coherence' of a particular transform is due to the globally connected inhibitory interneurons and the relatively short axonal conduction delays. The PSTH (Top) however, illustrates that while the neurons of both stimuli may start firing in global synchrony due to their equal and simultaneous stimulation, as hypothesized, they quickly settle into antiphase oscillations, or perceptual cycles, resulting in interleaved stimulus representations through time.

The inferences made from the raster plot and PSTH are confirmed by examination of the auto- and cross-correlations of the input layer neurons' activity in Figure 6.11. The autocorrelations of both stimuli (Figures 6.11a and 6.11c) show a clear oscillation through time with a period of $50\,ms$. From examining the cross-correlation between the two stimuli (Figure 6.11b) it is clear that these oscillations occur in opposite phase to one another, as the peaks of significant positive correlation occur with the same period but after an initial shift of half the period ($25\,ms$). To be specific, peaks can be seen in the cross-correlation at $\pm\{25, 75, 125, 175, 225\}\,ms$ which correspond to $\pm\rho(1/2 + f)\,ms$ where $f \in \{0, 1, 2, 3, ..., n\}$ and $\rho = 50\,ms$.

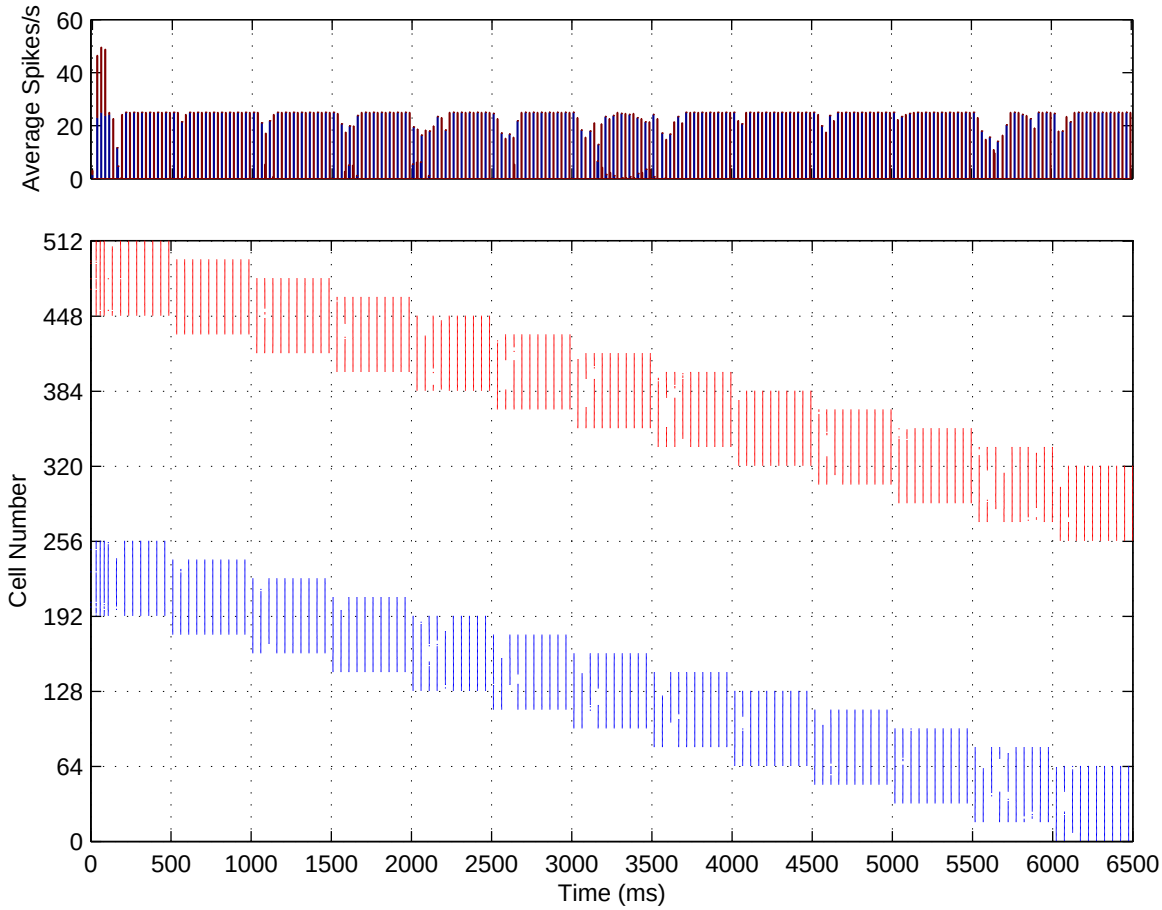
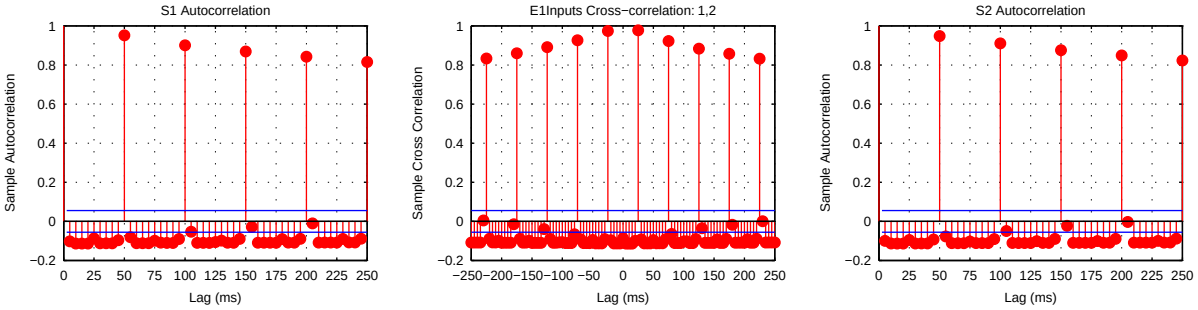


Figure 6.10: Input layer activity during training with two simultaneously presented translating stimuli. In the raster plot (Bottom) it can be seen that the two stimuli are presented simultaneously, translating across their halves of the input layer. Stimulus 1 (shown in blue) occupies neurons 1–256 while Stimulus 2 (shown in red) occupies neurons 257–512. Both stimuli can be seen to have tightly synchronized volleys of spikes for each of their transforms. Furthermore, the post-stimulus time histogram (Top) shows that the volleys of spikes representing the two stimuli oscillate in antiphase perceptual cycles.

6.3.3.2 Learning transformation-invariant representations

While the raster plot and activity correlations show a clear structure to the organization of the input representations, when only firing rates are considered, this structure is obscured. Figure 6.12 shows the input layer activity summed over the duration of the stimulus presentation period during training and demonstrates that when only spike counts (firing rates) are considered, it is very difficult to distinguish between the two stimuli. As such, a temporally sensitive learning rule is required in the feed-forward



(a) Autocorrelation: S_1 (b) Cross-correlation (c) Autocorrelation: S_2

Figure 6.11: Input layer cross-correlation and autocorrelations for each stimulus after training. The Autocorrelations both show significantly periodic firing with a period of approximately $50\,ms$ (the blue lines are the approximate upper and lower 95% confidence bounds, assuming that time series are completely uncorrelated). Only positive lags are shown for the autocorrelation functions since they are symmetrical about 0. Correspondingly, the Cross-correlation demonstrates that these oscillations are out of phase with one another, as shown by the lack of positive correlation at $0 - lag$ and the positive correlations every $50\,ms$ but shifted by $25\,ms$ as expected for interleaved volleys of spikes from each stimulus.

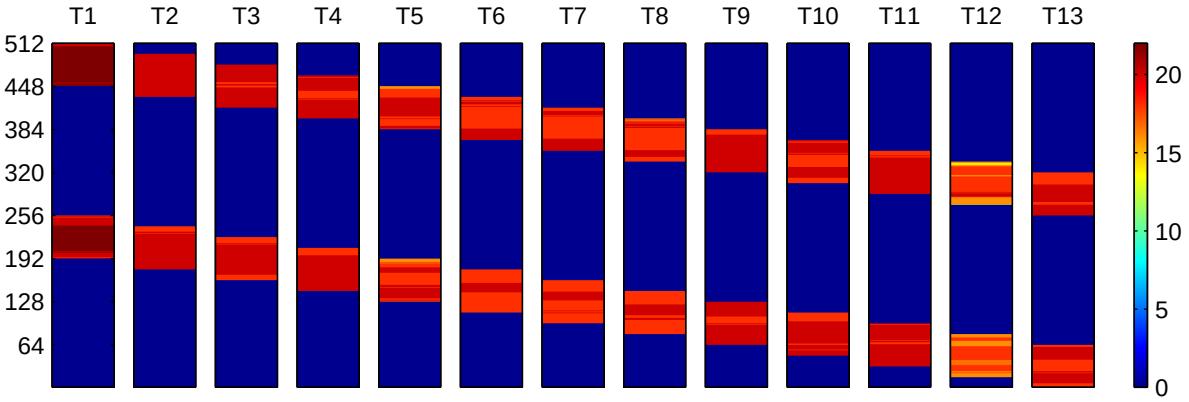


Figure 6.12: Firing rate plots of the input layer during training with two simultaneously presented stimuli. When presenting both objects as a compound stimulus translating across the input layer, it is clearly very difficult for a rate-based learning rule to differentiate between each object in order to form separate output transformation-invariant representations of each object.

connections, such as STDP, in order to avoid the superposition catastrophe (which rate-based learning rules are prone to) and thereby form separate representations of the two stimuli in the subsequent layer.

After presenting the two stimuli simultaneously translating together in a random direction for ten epochs ($1000\,ms$ per transform), separate transformation-invariant

representations were formed in the output layer. The raster plots of the output layer (Figure 6.13) demonstrate how neurons which are initially responsive to a random selection of transforms for each stimulus become selective to a particular stimulus across all its transforms.

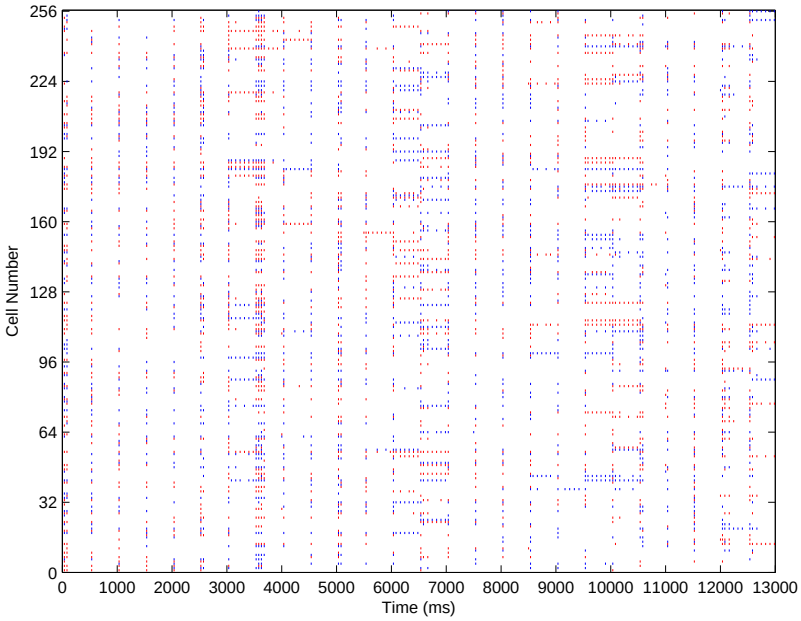
Examining the feed-forward synaptic weight matrices before and after training (Figure 6.14) reveals the structure developed through learning which leads to these stimulus-specific transformation-invariant representations. From an initially random distribution of synaptic efficacies, clear red horizontal lines emerge extending continuously through either the first or second half of the input layer neurons, representing weights from all neurons representing the features of all transforms of a particular stimulus onto a particular output neuron (y-axis) which have undergone LTP. Conversely, the weights from the other half of the input layer onto the same output neuron are typically blue, signifying that they have undergone LTD.

As a quantitative measure of network performance, the standard information analyses were performed as detailed in Section 2.9.4. The results are plotted in Figure 6.15 and show that for most of the output neurons, the organization of the input representations achieved through the lateral axonal delays has enabled them to form fully transformation-invariant representations of the stimuli, as shown by the single cell information measure attaining the maximum $1\,bit$ for over 150 neurons (Figure 6.15a). The subsequent multiple cell information measure based upon the five most informative cells for each stimulus shows that randomly drawing two from this ensemble conveys the maximum level of information meaning that both stimuli are represented invariantly across their transforms in the output layer.

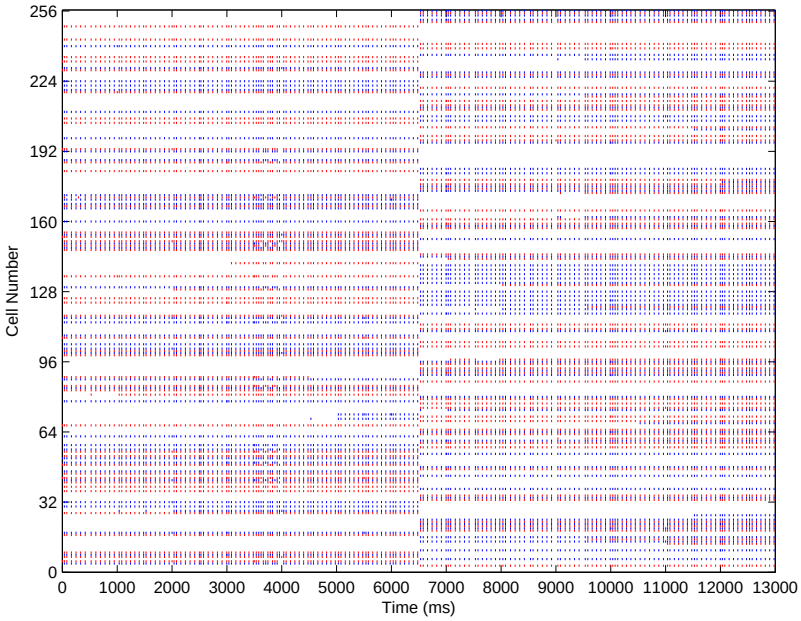
6.3.3.3 Inter-stimulus Delay

While previous theoretical work and the simulations above have shown that delays may lead to perceptual cycles when multiple stimuli are presented to a spiking neural network, the magnitude of the delays may be important in forming the perceptual cycles and so was systematically explored by varying the inter-stimulus delay, Δt_{IS}.

The most effective delay for producing the perceptual cycles in the input representations was expected to be equal to half the period of oscillation, since this would mean that the excitatory stimulation would arrive from the midpoint of one stimulus at the midpoint of the other (and therefore encourage it to fire) at exactly half-way between the two volleys of the first stimulus. This is a symmetrical relationship, such that each stimulus would encourage the other to fire at half-way between its volleys of spikes, which should therefore lead to perceptual cycles.



(a) Before Training



(b) After Training

Figure 6.13: Output layer raster plots before and after training with two simultaneously presented stimuli. The cell responses to the 13 transforms of Stimulus 1 (0–6500 ms) and Stimulus 2 (6500–1300 ms) are random to transforms of each stimulus before training. After training, many of the cells respond invariantly and selectively to all transforms of a particular stimulus.

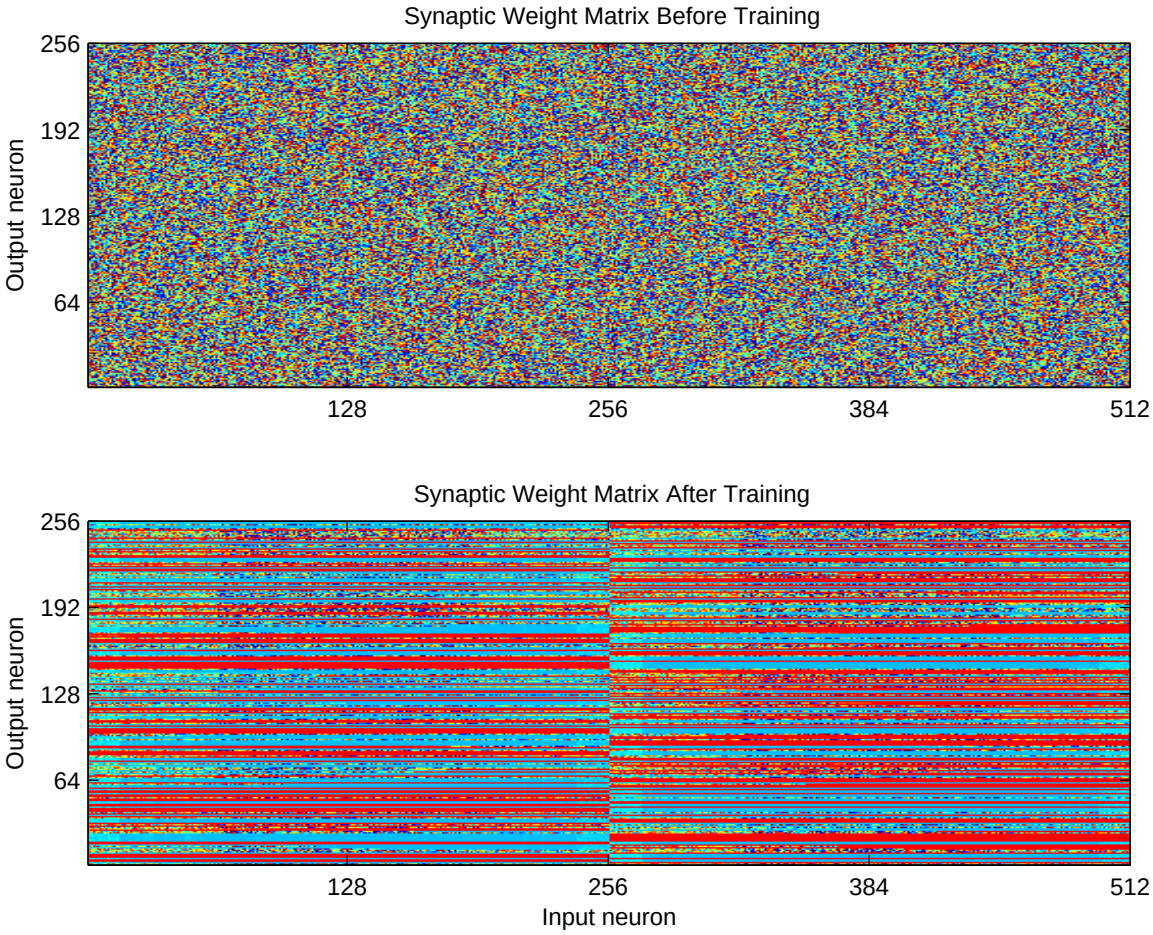
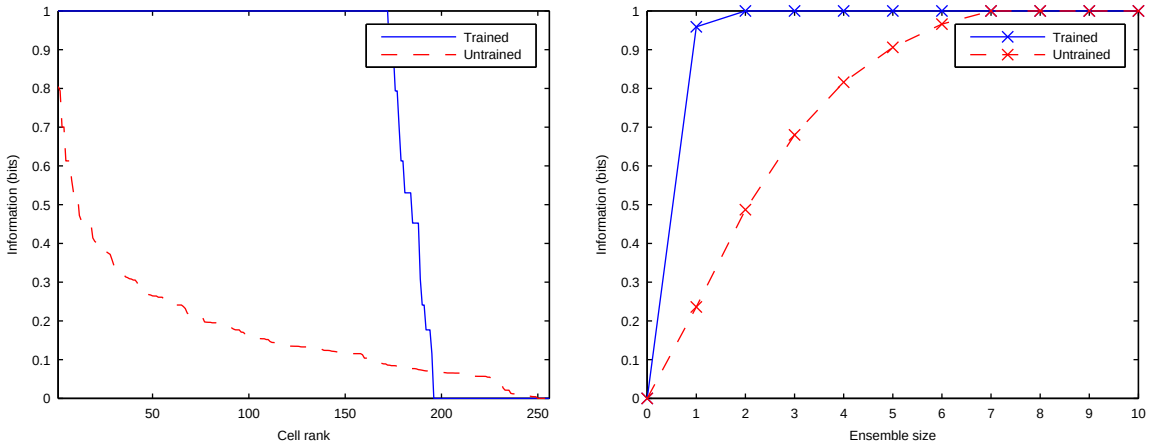


Figure 6.14: Excitatory feed-forward synaptic weight matrices (before and after training). Before training (Top) the weight matrix is unstructured due to the random initialization of the weights. After training (Bottom), there are strong connections from all the input neurons representing all transforms of Stimulus 1, shown by horizontal red stripes of large weights from input neurons 1–256 and likewise for Stimulus 2, with horizontal red stripes from input neurons 257–512 to different output neurons.

Figure 6.16 plots the network performance measure against a range of values of the inter-stimulus delay, $\Delta t_{IS} = [0, 60]\, ms$. While network performance was very high in the region of $\Delta t_{IS} = 25\, ms$ as expected, the optimal inter-stimulus delay was found to be $24\, ms$ after ten random seeds, slightly shorter than predicted. This may be because a small amount of time is required to depolarize the postsynaptic cells through the increase in excitatory synaptic conductances and so the lateral impulses have to arrive slightly ahead of each midpoint in the cycles, that is $\rho(1/2 + f)\, ms$, in order that the postsynaptic cells (of the other stimulus) can fire at the mid-point and establish a more precisely timed perceptual cycle.



(a) Single cell information measure

(b) Multiple cell information measure

Figure 6.15: Information plots of output neurons for the trained and untrained network. After training, the single cell information measure shows that most of the output layer neurons have become maximally informative in signalling a particular stimulus invariantly across all its transforms. The multiple cell analysis shows that different neurons among the most informative cells have learnt to respond to both stimuli invariantly, conveying the maximum level of information with a small ensemble of output neurons.

6.4 Discussion

In this chapter, simulations have demonstrated some of the computational implications of axonal conduction delays with respect to the process of learning to form transformation-invariant representations in a spiking neural network model of the primate ventral visual stream. The first section was devoted to investigating the consequences of augmenting the model with conduction delays in the excitatory feed-forward axons while the second section explored the use of lateral excitatory delays in providing a mechanism for temporally segmenting a visual scene composed of multiple objects.

It was found that both the CT and Trace learning mechanisms are unaffected by constant delays in the excitatory feed-forward connections when the axonal conduction delays are also incorporated into the STDP model (as used throughout this thesis). When the original form of STDP was used (Perrinet et al., 2001), where the presynaptic plasticity variables (C_{ij}) and synaptic weight updates occur upon the emission of the presynaptic spikes rather than when they arrive at the synapse at the end of their axons, the effects of axonal delays are rather different. Generally speaking, the network performance measure under a CT learning paradigm decreases with increasing axonal

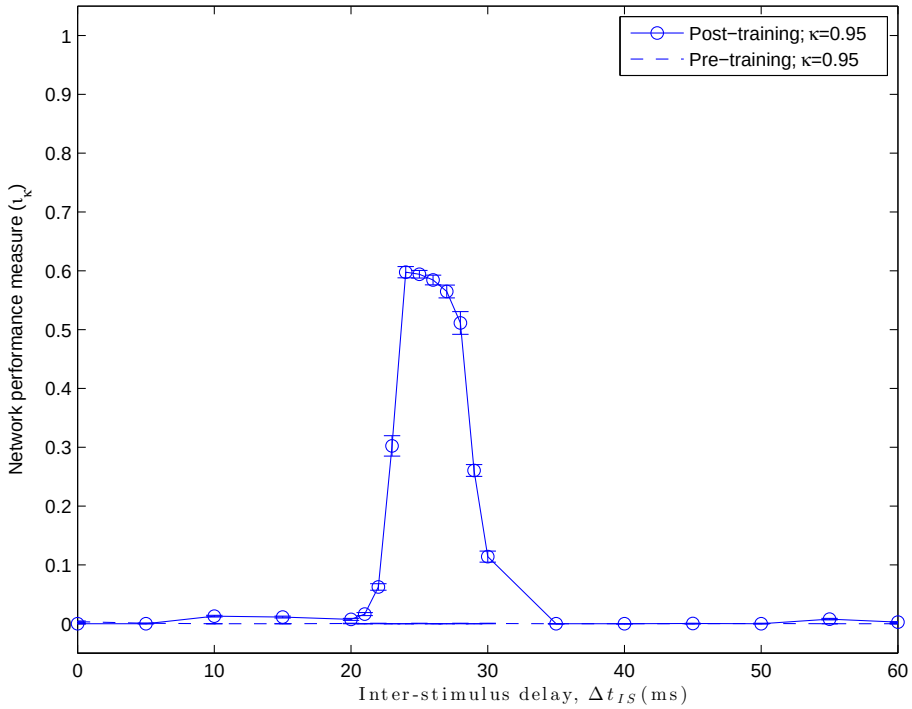


Figure 6.16: The information analysis summary (with the corresponding pre-training values plotted with broken lines) plotted against the maximum axonal conduction delay. Each value of Δt_{IS} was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean. The optimal result was found to occur at an inter-stimulus delay of $24\,ms$ across these random seeds, with some tolerance around this value, particularly for larger delays.

conduction delays. The exceptions occur when the delays (approximately) match the period of the input spike volleys such that synapses are strengthened in an artefactual (acausal) manner by the effects of earlier input spike volleys arriving at the synapses in time to cause the postsynaptic cells to fire while the C_{ij} variables are high due to the latest spike emissions.

When a Trace learning paradigm is used with the unmodified STDP model, network performance actually increases with longer conduction delays. The activity trace remains unchanged, as the increases in the excitatory conductances due to spikes are still delayed and still have a long time constant which maintains their activity across transforms. The long synaptic time constants smooth the activity over time, making the model less sensitive to the precise timing of the spikes. With the original STDP model however, the model has the additional benefit that the learning starts sooner (beginning on the emission of each spike) and so effectively has the benefit of more

extensive training periods.

The results obtained with these two forms of the STDP model are qualitatively very different in the effects they have upon the learning processes of the model. To avoid unrealistic synaptic learning when incorporating axonal conduction delays into a spiking neural network utilizing the CT learning mechanism, it is therefore important to also revise the STDP model appropriately.

Due to the same relative insensitivity to the time of spikes, the network performance is very robust to distributions in the feed-forward axonal conduction delays in a Trace learning paradigm, with some highly informative cells even up to standard deviations of $50\,ms$. This is in stark contrast to the CT learning simulations with distributions of delays where network performance is almost completely degraded with a standard deviation of $2\,ms$.

It is evident from the simulations presented that CT learning struggles to cope with distributions of conductions delays, since the scattered arrival of spikes at the synapses onto a particular postsynaptic neuron break the effect of the collective action achieved by the synchrony of presynaptic cell firing. However, this work was carried out in a one-layer model for simplicity. Therefore, mechanisms observed in multi-layer neural networks shown to help overcome noise and reduce timing jitter (resynchronize volleys) through successive layers were not available (Diesmann et al., 1999). As such, the problems in matching conduction delays for CT learning to operate successfully may be ameliorated in multi-layer models which would be an interesting direction for future work.

Another idea to explore would be to set each presynaptic neuron to make *multiple* contacts on each postsynaptic neuron, each with a different delay. The STDP learning rule may then select from this distribution of delays those which were most effective at driving the stimulus by matching them to the firing sequences (Rousselet et al., 2004) in a similar way to how polychronization arises (Izhikevich et al., 2004; Izhikevich, 2006)

In the second section, axonal conduction delays were incorporated into lateral connections with multiple simultaneous stimuli presented during training. This is in contrast to the previously presented work (in Chapter 5) where the perceptual cycles were formed through a Gaussian gradient in the strength of the lateral connections based upon the Euclidean distance between neurons. This model uses the natural assumption that the delays are proportional to the Euclidean distance with the consequence that nearby neurons with small lateral delays (representing features of the

same stimulus) are brought into synchrony, whilst those further apart (representing a different stimulus) are pushed out of phase.

By linking neurons together in proportion to the time delay between them, perceptual cycles may be reliably formed in order to segment simultaneously experienced stimuli. While previous work has demonstrated how latencies in feed-forward connections may help to bind the features of individual object representations when several objects are presented through the coordinated arrival time of spikes (Rousselet et al., 2004), the present work shows how lateral delays may serve the opposite function — essentially to 'unbind' competing representations within the same neural layer to enable separate representations to form in the next.

Through a systematic exploration of the scaling of the lateral delays, (Δt_{IS}) it was found that the perceptual cycles were formed most reliably (and hence network performance was highest) when the inter-stimulus delay approximately matched half the period of oscillation for each stimulus, such that the delayed excitation arrives at the neurons representing one stimulus and so encourages them to fire midway through the cycle of the neurons representing the other stimulus. For a fixed lateral conduction velocity and typical firing rate in the visual cortex, it may be possible to predict the minimum separation between two visual stimuli at which they are perceived as distinct objects.

As another potentially interesting direction for future research, lateral delays may be implemented along with lateral plasticity in the form of STDP. Here I would predict that this may lead to the natural formation of a SOM architecture with weaker lateral excitatory connections at greater separations. If neurons are firing synchronously within a neighbourhood of a neural layer with initially random lateral connections, a particular pattern will be established according to the period of firing and the distance (delay) between the neurons. Spikes travelling over a long distance which arrive after a postsynaptic spike will lead to LTD, weakening the more distant synaptic connections. Those spikes travelling a shorter distance however, will arrive before the postsynaptic cells fire, leading to LTP and strengthening synapses with closer neurons. The radius of the excitatory connections will then decrease until a minimum distance is reached through the same mechanism which leads to the reduction of latency in feed-forward plastic connections (Song et al., 2000), leaving a tail of weaker lateral connections which were not reinforced to the same extent. This may prove to be the principle by which the computational maps found throughout the early visual cortex arise as an epiphenomenon of STDP and the slow conduction speeds of unmyelinated lateral connections.

Axonal conduction delays are often omitted from models of the visual cortex, possibly in order to avoid moving away from point neurons and using much more complex and computationally expensive partial differential equations (PDEs), describing the evolution of the dynamic variables along the spatial extent of the neuron and its axons as well as through time. In a similar way to which the shape of an action potential is often neglected as an unnecessary level of detail, modelling the flow of ions along the axons and dendrites of a neuron may also be reasonably omitted when the purpose of the model is to generate accurate spike timings as the foundation for the phenomena of interest. In the implementation presented here and found elsewhere (Brette et al., 2007), the incorporation of axonal conduction delays without the computational expense of PDEs was achieved by modelling each axon as a spike queue (FIFO buffer), implemented as a circular array where spikes are stored for a particular delay period before they have an effect upon the synapses. This was found to be an adequate level of abstraction to explore the effects of axonal conduction delays upon the Trace and CT learning mechanisms and to provide an additional mechanism for the temporal segmentation of multiple stimuli with only a moderate increase in computational cost (see Section 2.4).

In conclusion, axonal conduction delays have been shown to be a potentially important feature for spiking neural network models of the ventral visual system, and should be carefully considered for inclusion, most notably for their ability to temporally segment stimuli in more natural visual scenes composed of multiple objects. In so doing, the form of STDP used should also be modified to incorporate the effects of axonal delays. A computationally cheap alternative to systems of partial differential equations is implemented and demonstrated to be sufficient to generate perceptual cycles between spatially separate objects. These input layer representations are then shown to be sufficient for forming separate transformation-invariant representations of each stimulus in the output layer as the stimuli translate across the input layer.

There is, next, what is commonly referred to as the binding problem, a critical problem for visual physiology. The problem is that of determining that it is the same (or a different) stimulus which is activating different cells in a given visual area or in different visual areas.

—Semir Zecki

7 Lateral Plasticity and Stimulus Categorization

7.1 Introduction

While the increasingly large receptive field sizes along the ventral visual stream may help to build cell responses with tolerance to stimulus size and position, they may simultaneously fail to preserve spatial information, leading to the 'binding problem' when multiple stimuli are seen together (Rousselet et al., 2004). Previous attempts to model the learning processes of the ventral visual system have been limited by the need to present stimuli individually during training (Fukushima, 1980, 1988; Wallis and Rolls, 1997) in order to avoid what is otherwise known as the 'superposition catastrophe' (von der Malsburg, 1999). While this single stimulus presentation has been a useful first step in understanding how temporal and spatial continuity can lead to view-invariant representations, several researchers have recognized how limiting this can be for real world applications or modelling the human visual system, where visual scenes typically involve multiple stimuli (Stringer et al., 2007; Stringer and Rolls, 2008).

Some have proposed that this approach is not lacking in ecological validity as the attentional spotlight means that we only perceive one object at a time, even though multiple objects may be present in our visual fields (Rolls and Deco, 2002). However, this begs the question of how the objects are segmented in the first instance (in order

for attention to select one out of the entire scene) and so requires further investigation.

Spratling (2005) produced a model where line orientations are learnt irrespective of their translation. This was achieved over two layers of neural network nodes, with the first layer receiving inputs from images of training stimuli and learning conjunctions of features within those images. The second region received its inputs from the first layer and learned disjunctions across the patterns of activity generated in the lower region. Interestingly, under some circumstances, the performance of the network was actually enhanced by the presence of multiple stimuli rather than stimuli presented in isolation.

Plausible solutions to the problem have come from considering not just the properties of the model, but also the statistics of the environment. By presenting different combinations of stimuli to a simple one-layer competitive neural network, it is able to separate out the stimuli and form individual representations despite only being trained on multiple objects (Stringer and Rolls, 2008). This process requires a sufficient degree of 'statistical decoupling' between objects however, (that is, each object is experienced with a sufficient number of other objects) which may not always be available to the observer.

Progressing this research, it was found that another mechanism requiring less extensive training was to show the network pairs of objects moving independently, for example, rotating at different speeds or in different directions (Tromans et al., 2012). Again, this relies upon a more realistic training environment rather than any additional properties of the model and it seems likely that there are other mechanisms in an ecologically valid visual scene for helping to segment and learn about objects independently.

Research conducted on these mechanisms has so far been with rate-coded neural network models. In Chapter 5, it was demonstrated how additional, biologically plausible properties of a *spiking* neural network may provide additional ways to overcome the superposition catastrophe by desynchronizing the volleys of spikes representing each stimulus with respect to each other through hard-wired lateral connections and cell firing rate adaptation in a competitive network.

In this chapter, learning is implemented in the lateral excitatory connections rather than hard-wiring a 'SOM' architecture, to investigate whether a single layer of laterally connected excitatory and inhibitory neurons is able to use these lateral connections to learn about the visual categories it is presented with. After an initial phase of encoding category information in the lateral connections, the network is then able to segment simultaneously presented novel stimuli through organizing their representations into

antiphase oscillations. It is then shown how this dynamic of 'perceptual cycles' (Miconi and VanRullen, 2010) may be used to learn transformation-invariant representations through modification of the feed-forward excitatory connections as the stimuli translate across the input layer.

The first section of research in this chapter explores how prior learning of category information in the excitatory lateral connections allows the network to push representations of two previously unseen stimuli from different categories out of phase with respect to one another (based upon the prior learning about their categories) and several of the key factors affecting such one-layer dynamics.

During the initial training phase, the network is presented with 'exemplar stimuli' from each of two groups where the exemplars are composed of a fixed number of neurons (representing input 'features'), drawn randomly from their respective prototype pools of neurons. The neurons of a particular exemplar are then activated with an external current such that each exemplar is presented to the network individually. Here it is assumed that stimuli from the same category share a proportion of their features in common with each other and that members of different categories have far fewer features in common. In the idealized simulations of this work, there are no shared features between different categories.

As the exemplars are presented to the network, the coactivity of the features of a category leads to their association through the Spike-Time Dependent Plasticity (STDP) learning rule in the excitatory lateral connections. This first stage serves to strengthen these initially weak (zero change in conductance) connections between members of the same category and so models an early stage of visual development where such implicit information is encoded through the prototype effect (Posner and Keele, 1968).

After building up the lateral connections in this way, the network is tested by presenting a combination of two novel exemplars, representing a simple visual scene composed of two objects. It is hypothesized that the effect of training the lateral connections will be that the two stimuli are segmented through antiphase oscillations, that is to say, the neural firing representing the features of a particular exemplar will be synchronized with respect to the other features of the same exemplar (due to the strengthened lateral connections between features of the same category) and desynchronized with respect to the features of the other exemplar (due to the weak connections between features of different categories).

The second section here augments the network to incorporate a second layer of neurons with excitatory, plastic feed-forward connections between the pyramidal cells

of each layer. The aim of this work is to investigate how such higher layers may exploit the input layer dynamics formed from prior learning about categories in order to segment a visual scene composed of multiple stimuli and learn transformation-invariant representations of them as they move in lockstep across the input layer.

A similar training paradigm is used in the first phase of training for the lateral connections, except that the input layer is extended and the individually presented stimuli translate across it. This allows the network to build strong lateral connections between neurons representing the different transforms (that is, locations) of stimuli in the same category, without strengthening connections between transforms of different categories.

During the second phase of training, the lateral excitatory synaptic weights are prevented from further modification and learning commences in the excitatory feed-forward connections utilizing the same STDP learning rule used to train the lateral connections in the prior phase. A pair of novel stimuli are presented together to the network in the same portion of the input layer, which are shifted progressively across the input layer in lockstep.

The hypothesis tested is that the antiphase oscillations generated by the learning in the plasticity of the input layer's lateral excitatory connections will allow the output neurons to learn to respond selectively to either one (but not both) of the two novel stimuli across most or all of their transforms. This should be possible due to the temporal specificity of the STDP learning rule whereby significant LTP occurs only if the input spike precedes the output spike within a narrow time window. Due to the self-organized dynamics of the input layer representations in which two stimuli are out of phase, this should occur for one particular stimulus at a time, while the other stimulus spike volley should fall outside of this LTP time window (and instead come after the output spikes, thus making them subject to LTD). This should allow transformation-invariant representations to form of the translating stimuli individually through the CT learning mechanism, as detailed in Chapter 3.

7.2 Methods

7.2.1 Model Architecture

In the first set of simulations described in Section 7.3.1, the network was a single layer of 512 excitatory leaky integrate-and-fire pyramidal neurons with plastic lateral excitatory connections initialized to zero strength. Unlike the simulations of Chapter 5, these lateral connections were modified through learning rather than imposing a

fixed 'Mexican Hat' profile upon them. The pyramidal cells featured cell firing-rate adaptation mediated by calcium-gated potassium currents as described in Section 5.2.2. There was also a separate pool of 128 inhibitory interneurons with fixed strength lateral connections to and from each of the excitatory cells. All synapses were conductance-based and plastic synapses (both lateral, ElE and feed-forward EfE) featured a form of spike-time dependent plasticity as described in Section 2.3.1 (Perrinet et al., 2001).

In the second set of simulations with translating stimuli described in Section 7.3.2, the same basic architecture was used but an additional output layer of 64 excitatory pyramidal cells was incorporated. Importantly, the addition of an output layer also meant that there were feed-forward excitatory connections from pyramidal cells in the input layer to pyramidal cells in the output layer (EfE). These feed-forward connections were modified during training using the same STDP learning rule used in the excitatory to excitatory lateral connections but initialized randomly and uniformly in the range $[0, g_{Max}]$. Each of the two layers of pyramidal cells had their own pool of inhibitory interneurons in a ratio of 4:1 ($E : I$). The input layer excitatory pyramidal neurons were arranged into a spatial configuration of 32 neurons wide by 16 neurons deep (instead of being one-dimensional as before) to accommodate translating stimuli more easily, while the output layer was arranged in an 8×8 configuration.

7.2.2 Neuron model description

The neurons used in this work are conductance-based leaky integrate-and-fire neurons, with Gaussian white noise added to the cell membrane potential with zero mean and standard deviation $\sigma = 0.015 \cdot \left(\Theta - V^H\right)$ as used by Masquelier et al. (2009) and cell firing-rate adaptation as described in Chapter 5. The cellular parameters used were the default values described in Table 2.2 from Troyer et al. (1998) and the default STDP parameters used are as detailed in Table 2.3 (Troyer et al., 1998; Perrinet et al., 2001).

This system of differential equations describing the dynamics of the cell bodies, synaptic conductances and synaptic plasticity are discretized with a Forward Euler numerical scheme and simulated with a numerical time-step Δt of $0.02\,ms$.

7.2.3 Lateral connectivity

Excitatory neurons within each layer were connected to other excitatory neurons according to the probability $p(ElE) = 0.5$ (disallowing self-synapses) with the synaptic weights Δg_{ij}^{ElE} initialized to zero and changed during the course of training according

to the STDP learning rule described in Section 2.3.1 (Equations 2.27–2.29). The excitatory neurons were fully connected to the inhibitory neurons and vice-versa within each layer, however unlike the ElE synapses, EI and IE synapses were not plastic but instead initialized to $\Delta g^{EI} = \Delta g^{IE} = 5\,nS$.

7.2.4 Stimuli and Training

Stimuli were represented by injecting tonic current into spatially separate pools of input-layer neurons. Contiguous blocks of neurons within these pools were gradually shifted across the input layer to represent successive overlapping transforms of each stimulus which may be associated by Continuous Transformation learning (Stringer et al., 2006). In the initial simulation with two stimuli, a stimulus consisted of 64 neurons which was presented in 13 locations (transforms), with a shift of 16 neurons between each adjacent transform giving an overlap of 75%.

During testing phases, all stimuli were presented individually to measure how the network responded specifically to each stimulus, however during training, the stimuli were presented together, such that the network never learnt about them in isolation. The untrained network was first tested in a 'pre-training' phase by presenting all transforms of all stimuli sequentially, each for a cue period of $500\,ms$ to provide a baseline behaviour to contrast with the effects of learning in the feed-forward synapses.

After saving the network outputs and resetting the dynamic variables for each cell and synapse, the training phase began where all transforms of all stimuli were presented for $500\,ms$ per transform. For combinations of stimuli, the direction of shift between transforms was randomly chosen but with each stimulus in the presented pair shifting in the same direction. The presentation of all transforms of all combinations (e.g. pairs) of stimuli constituted one epoch of training, and the training phase consisted of five epochs in total.

Once the network had been trained and the dynamic variables (except synaptic weights) reset, the 'post-training' testing phase was simulated in an identical way to the pre-training testing phase. The final outputs were then saved and analysed.

7.3 Results

In the first section of results, we explore the formation of perceptual cycles in a single layer of laterally connected excitatory and inhibitory neurons. In general, the lateral excitatory to excitatory connections (ElE) are trained through exposure to examples of each group of stimuli presented individually and then tested with a

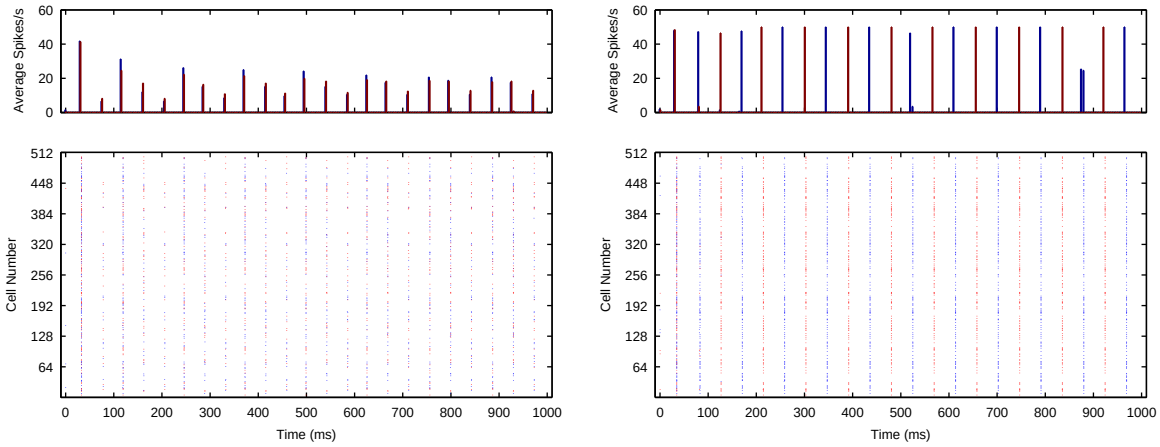
novel exemplar from each category presented simultaneously. The network is then analysed to determine whether it was able to perceptually organize the novel stimuli by representing each alternately through time in antiphase oscillations.

In the second section, this work is extended to the case of translating stimuli which are presented individually and shifted (in lockstep) across the input layer during training of the lateral connections. A second output layer of excitatory and inhibitory neurons was added to the network with plastic feed-forward synapses from the excitatory neurons in the input layer to the excitatory neurons in the output layer (EfE), which were trained on the novel exemplars translating *together* across the input layer. The network was then tested by presenting each novel stimulus *individually* translating across the input layer. The network's output neurons were then analysed for their ability to recognize each novel stimulus across all of its transforms using the information analysis techniques described previously in Sections 2.9.4 and 5.3.3 (Equation 5.6).

7.3.1 Single layer representations

The first results presented are a demonstration of how plasticity in lateral excitatory connections enables temporal segmentation by antiphase oscillation of novel exemplars belonging to previously seen categories. This is evident in the raster plots of Figure 7.1. Prior to training, when presented with two testing exemplars simultaneously (one from each group), the excitatory neurons from both groups exhibit global synchronization driven by the effects of the inhibitory interneurons (Figure 7.1a). After training the plastic ElE connections on ten exemplars from each of two groups, the same combination of novel exemplars now cause the two populations of excitatory neurons representing each group to synchronize their firing with respect to other neurons in their group but fire in opposite phase with respect to neurons in the other group (Figure 7.1b). This can also be observed in the histograms where the two colours denote the binned average spike rates for the two groups of neurons (Figure 7.1, top row).

The self-organization through synaptic modification can be further seen in the auto- and cross-correlations shown in Figure 7.2. Before training, the neurons of both groups fire with a regular period of approximately 40–45 ms as evidenced by the regular significant peaks in the autocorrelations of Figure 7.2a. The cross-correlation before training exhibits the same periodicity as the autocorrelations and also has a large positive correlation at zero-lag, indicating that both groups of neurons are firing in phase with one another. After training however, the period of the

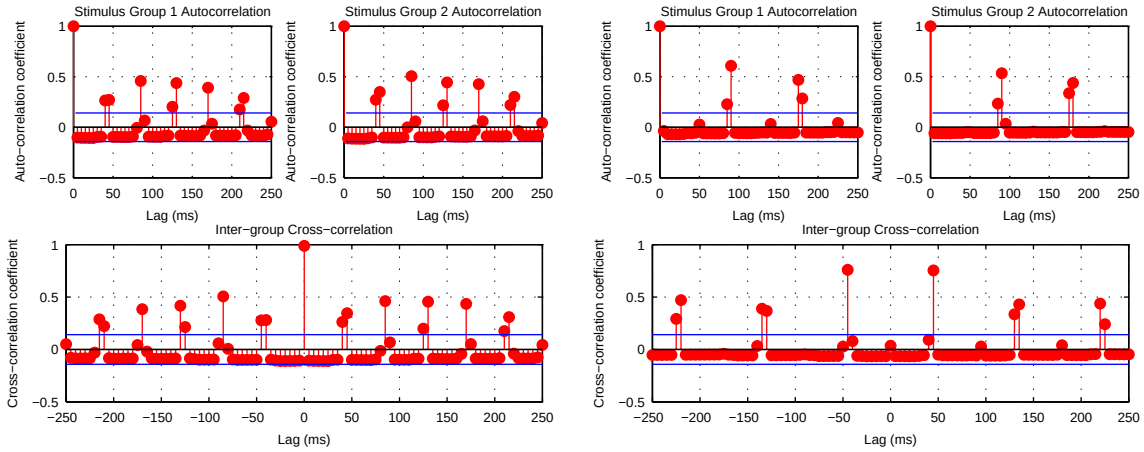


(a) Before training (b) After training

Figure 7.1: Spike rasters before and after training the excitatory lateral connections. Prior to learning, both populations are rapidly synchronized through the action of the fully connected inhibitory interneurons (global inhibition). After learning, it can be seen that each population of neurons is internally synchronized and desynchronized with respect to the other population.

autocorrelations for each stimulus group can be seen to have doubled to approximately 85 ms, demonstrating that each group still has periodic (oscillatory) firing but now each fires at half the frequency of the untrained case (Figure 7.2b). Additionally, the cross-correlation also now shows double the period between significant correlation peaks and no longer has a significant correlation at zero-lag but at $\pm$ 40–45 ms lag, indicating that neurons of each stimulus group are no longer firing synchronously with the other group, but are firing on alternative cycles of approximately 22 Hz (while synchronized with members of their own group).

From examining the weight matrices, the reason for this organization of the input representations is made clear (as shown in Figure 7.3). Before training (Figure 7.3a), there is no structure in the lateral excitatory–excitatory lateral connections as all synaptic efficacies (Δg_{ij}^{ElE}) were initialized to 0. After training, some of the weights can be seen to have increased (Figure 7.3b) as expected from the effects of global inhibition and stimulation of the principal cells representing particular features of their group causing them to fire in synchronous volleys. After reordering the rows and columns (representing the post- and presynaptic excitatory cells) of the post-training weight matrix according to group membership (Figure 7.3c), the structure built in the lateral connections is made apparent. Since the features drawn from a particular group are always experienced together during exposure to the ten exemplars of that

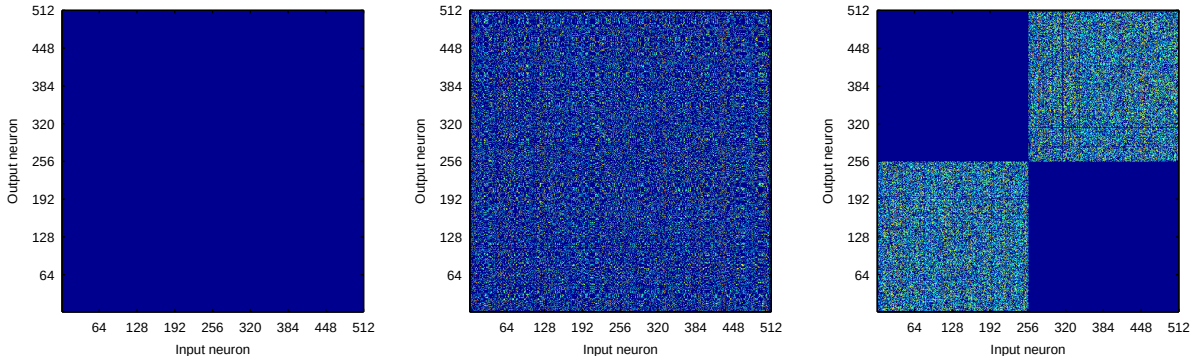


(a) Before training (b) After training

Figure 7.2: Autocorrelations for each group (top row) and Cross-correlation between groups (bottom row) before and after training (the blue lines are the approximate upper and lower 95% confidence bounds, assuming that time series are completely uncorrelated). Only positive lags are shown for the autocorrelation functions since they are symmetrical about 0. Before training, neurons from both groups fire synchronously, (as indicated by the strong positive cross-correlation at zero-lag) with a regular period of approximately 40–45 ms. After training however, the period of firing for each group has doubled to approximately 85 ms with each group firing in opposite phase to one-another as indicated by the very small zero-lag cross-correlation.

group during training, strong connections form between these features of, for example Group 1, shown in the sorted lower left quadrant of the post-training weight matrix (Figure 7.3c). The same is true for the stimulus features of Group 2 which can be seen to have mutually strong lateral connections in the upper right quadrant.

Since these strengthened lateral connections are effectively shaped through the course of training according to the statistics of the co-occurrence of visual features, the connections are strengthened between neurons representing features from the same group. Then when multiple stimuli are presented together after training the lateral connections (even without spatial separation), neurons representing the features of one group act in a mutually supportive way to synchronize the firing of neurons representing other visual features of the same group. Together with the competitive effects of inhibitory interneurons and cell firing-rate adaptation, this leads to the generation of perceptual cycles when novel exemplars with visual features characteristic of their particular groups are presented after training, as demonstrated in Figures 7.1 and 7.2.



(a) Before training (b) After training (c) After training (sorted)

Figure 7.3: Excitatory lateral synaptic weights (Δg_{ij}^{ElE}). Before training, the lateral weights (synaptic efficacies) have no structure as they are uniformly initialized to 0. After training, some of the weights have increased due to the synchronized firing of input representations. After sorting the weights according to stimulus group membership, the structure of the weight matrix is made clear. The stimulus features of Group 1 have formed strong connections with the other features of their group exclusively (bottom left quadrant) while the features of Group 2 have made strong connections with the other features of their group.

7.3.1.1 Strength of lateral connections

The key function of the networks of input neurons investigated here is to synchronize the neurons representing the features of one group and desynchronize them with respect to the others, allowing them to alternate in an antiphase relationship dubbed 'perceptual cycles' (Miconi and VanRullen, 2010). To quantify the network's ability to self-organize in this way, we apply the same measure as Miconi and VanRullen which calculates correlations in spiking activity between the two (or more) populations representing each group and between subpopulations of the same group for the 'between-group' and 'within-group' measures respectively.

During the testing phase (with synaptic plasticity turned off) the network is presented with the two novel stimuli simultaneously (one from each group) for 1000 ms. The spikes from neurons of each group are put into 10 ms bins and those bins containing less than a total of ten spikes were excluded to prevent quiescent periods from degrading the performance measure (Miconi and VanRullen, 2010). The correlation between this frequency series is then calculated with Spearman's rank method (Miconi, 2012, personal communication), yielding the between-groups correlation measure. Similarly, each group's population of neurons are divided in two and the same procedure is repeated for the two halves to obtain a within-group correlation for each group which

is then averaged across all groups to give the final within-groups measure.

For these simulations, the standard one-layer network was used and tested over a range of strengths of the lateral excitatory connections, from 0.05 to 500 nS. This range of parameters was repeated for a total of ten random seeds to gauge the effects of statistical variation and the *intergroup* and *intragroup* synchrony measures were calculated as described. The means across the random seeds were calculated and plotted with error bars representing the standard error of the means.

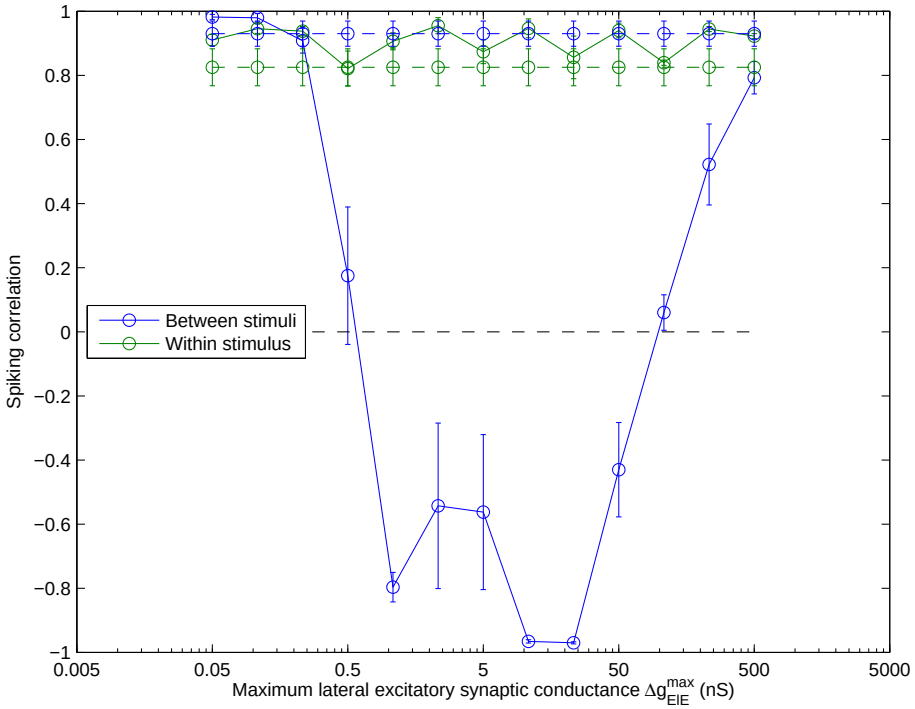


Figure 7.4: The spiking correlation measure within and between groups (with the corresponding pre-training measures plotted with broken lines) for a range of lateral excitatory conductance strengths. Each conductance strength was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean. The antiphase representations are a robust phenomenon over approximately two orders of magnitude (note the logarithmic scale of the x-axis).

It can be seen from Figure 7.4 that while the synchrony measures within and between the stimuli are very similar before training, the effect of learning in the lateral connections is to make the correlation between stimuli strongly negative (antiphase), whilst maintaining a high within-group correlation measure, over a broad range of conductances of approximately two orders of magnitude. When the conductances become too low, the synaptic efficacies are not strong enough to synchronize the

neurons belonging to the same stimulus group and so the effect tends to resemble that of no training. When the conductances are too high, the effect on the correlation measure is similar (converging the two measures in global synchrony) but for a different reason. Very strong conductances saturate the firing rates of the neurons, overwhelming the inhibitory interneurons and firing-rate adaptation such that all excitatory neurons in the whole layer are firing near their upper limit and hence are in global synchrony.

In much previous work, inhibitory interneurons have been found to play an important role in synchronizing volleys of excitatory cells. Earlier research not shown here suggests that modelling inhibitory cells as hyperpolarizing makes the system very sensitive to the strength of inhibition, requiring careful retuning for different simulation conditions. Shunting inhibitory interneurons have been found to produce more consistent effects with respect to network oscillations and excitatory synchrony and so the effects of variation in the inhibitory conductance strength Δg_{IE} are presented here.

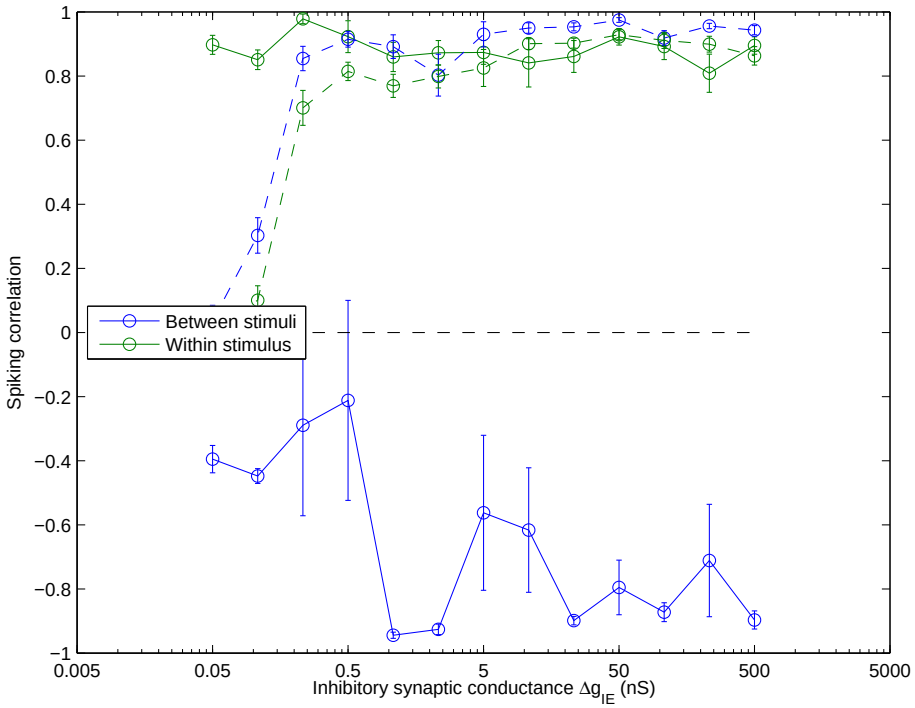


Figure 7.5: The spiking correlation measure within and between groups (with the corresponding pre-training measures plotted with broken lines) for a range of lateral inhibitory conductance strengths. Each conductance strength was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean. The antiphase representations are a robust phenomenon over four orders of magnitude (note the logarithmic scale of the x-axis).

It can be seen that before training, the network loses its ability to synchronize the volleys of excitatory action potentials for inhibitory conductances at approximately 10% or less of the strength of excitatory conductances. Through the training regime however, the lateral excitatory conductances become sufficient to form the antiphase input representations (albeit with less consistency than at higher values of inhibitory conduction strength). Figure 7.5 also illustrates that for very large inhibitory to excitatory conductances, the network still performs well, demonstrating that the strength of these synapses is not important for achieving perceptual cycles in this experimental paradigm (with zero-delay excitatory–excitatory axons), although inspection of the raster plots does reveal that the period of oscillations is extended in this case.

7.3.1.2 Adaptation

While in previous work, delayed inhibition has been used to generate perceptual cycles (Miconi and VanRullen, 2010), the mechanism of self-inhibition which allows for competing representations to alternate is, in this model, provided by cell firing-rate adaptation. The parameters governing the firing-rate adaptation are therefore key to the self-organization of the network and thus parameter variations are explored here in the same way.

The first parameter governing the cell firing-rate adaptation explored here is the calcium decay time constant τ_{Ca}. This adaptation time constant determines the time course over which the intracellular calcium dissipates and hence how quickly the neuron recovers from the additional membrane leakage (adaptation current) caused by the calcium-gated potassium channels. With a short time constant, the calcium clears quickly and hence the potassium current acting against perturbations from resting potential quickly returns to nothing.

Figure 7.6 shows the effect of varying the adaptation time constant τ_{Ca} from $5\,ms$ to $5000\,ms$ on the spiking correlation measures. The means of the results of ten randomly initialized simulations are plotted along with the standard error of the mean indicated by the whiskers on each datum. The results demonstrate that the formation of antiphase input representations is robust over approximately two orders of magnitude of variations in the adaptation time constant (from $\sim$ 20–2000 ms).

Inspection of the spike rasters reveals that with very short time constants, the firing-rate adaptation is too fleeting for the neural groups to self-inhibit, so all neurons fire synchronously. As the time constant becomes too large, the groups do fire in antiphase but are so strongly adapted (inhibited) by their calcium-mediated potassium currents that the oscillations occur with a very low frequency. As such, there may

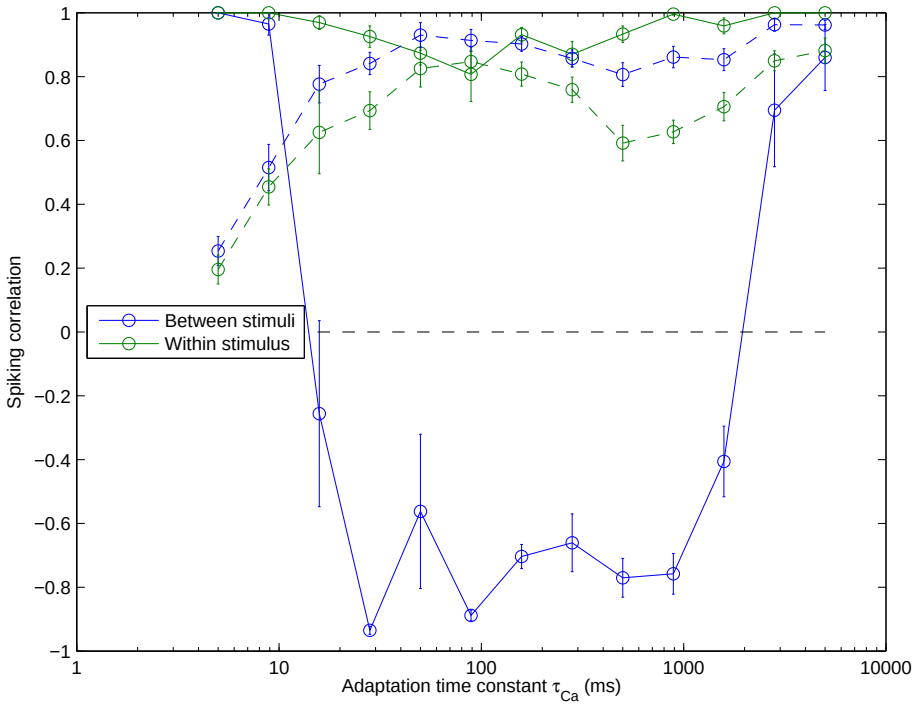


Figure 7.6: The spiking correlation measure within and between groups (with the corresponding pre-training measures plotted with broken lines) for a range of adaptation time constants. Each value of τ_{Ca} was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean. The antiphase representations are a robust phenomenon over approximately two orders of magnitude (note the logarithmic scale of the x-axis).

only be one volley per group in the entire 1000 ms period (after the initial inter-group synchronized volleys) which leads to very low performance as determined by the spiking correlation measure.

The second adaptation parameter explored here is the spike-triggered potassium conductance increase Δg_K which is mediated by intra-cellular calcium concentration α_{Ca} as described in Equation 7.1. Larger potassium conductances mean that large potassium currents leak through the cell membrane, meaning that the cell membrane potential is more resistant to larger perturbations from its resting potential (either depolarizing or hyperpolarizing) making the neuron more difficult to excite to its firing threshold. Conversely, smaller potassium conductances mean that the self-inhibiting effects of the adaptation (potassium) current are weaker so the firing rate does not decrease as much during constant (super-threshold) excitatory stimulation.

$$\Delta g_K = \alpha_{Ca} \cdot g_{AHP} \tag{7.1}$$

In a similar way to the analysis of the adaptation time constant, the spiking correlation measure was calculated for a range of potassium current increments (Δg_K) from 0.6–60 nS. Figure 7.7 shows the effect of varying this parameter on the spiking correlation measure within and between stimulus groups.

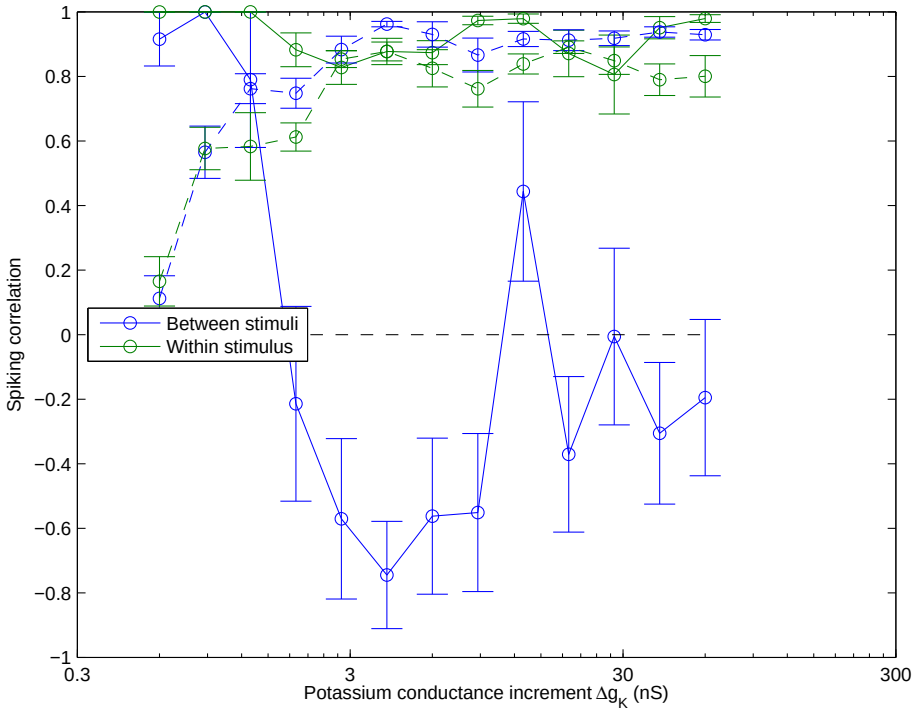


Figure 7.7: The spiking correlation measure within and between groups (with the corresponding pre-training measures plotted with broken lines) for a range of potassium current increments. Each value of Δg_K was simulated ten times with different random seeds with the data points representing the mean values across random seeds and the whiskers indicating the standard error of the mean. The antiphase representations are a robust phenomenon over approximately an order of magnitude (note the logarithmic scale of the x-axis).

As may be expected, for very low values of Δg_K, the effect of the adaptation current is weakened to the extent that the degree of adaptation is negligible, much like the case of very short adaptation time constants. Correspondingly, the raster plots in such conditions show that the neurons from both stimulus groups are firing synchronously (as the dominant mechanism becomes the synchronizing force of the inhibitory interneurons). With very large potassium current increments, the firing-rate adaptation becomes very strong, slowing the spike rate for both populations of neurons. The network is still reasonably capable of forming antiphase representations, since qualitatively there is still a self-inhibitory mechanism (unlike when it becomes too

small as to be negligible). However, quantitatively, the effects of adaptation become so strong as to reduce the period of oscillations, effectively reducing performance due to the ever fewer number of cycles to measure within the 1000 ms testing period.

7.3.2 Learning with antiphase input representations

Having explored the conditions under which alternating antiphase input representations may arise through lateral synaptic plasticity in a single layer model, this section extends this work to a two layer model incorporating plastic feed-forward connections between the excitatory cells in the input layer and those in the output layer. The working hypothesis is that the spike-time sensitive learning rule in the feed-forward connections will be able to exploit the time gaps between the volleys of spikes from each stimulus, even as they translate, in order to learn about one stimulus without interference from the other. Thus, although the two translating stimuli are presented to the network simultaneously, without spatial separation and moving in lockstep, the input layer is expected to use the richer dynamics of spike timings to represent the stimuli in alternate cycles through a mechanism not captured by rate-coded models such that separate transformation-invariant representations of each stimulus may form in the output layer.

The training regime was adjusted for the multiple layer model so that stimuli translated across the input layer for a total of five transforms, such that while consecutive transforms are significantly overlapping in order for the CT effect to operate, the first and the last transforms of a particular stimulus are completely orthogonal. The training period also consisted of two phases, the first of which trained the input layer lateral excitatory connections exclusively (preliminary training phase) whilst the second trained only the excitatory feed-forward connections between the layers (secondary training phase). This two stage training represented the early learning through exposure to exemplars of the categories and then more recent learning of transformation-invariant representations to specific novel exemplars.

During the preliminary training phase, ten training stimuli from each group were presented individually translating across the input layer for ten epochs while the stimulus groupings were learnt in the lateral excitatory connections. During the secondary training phase, the novel stimuli were presented as a pair translating across the input layer for another ten epochs while the excitatory feed-forward connections were trained. The network was then presented with each of the novel stimuli individually translating across the input layer in order to test and record the output layer neurons' responses to them in isolation for each of their transforms. The responses were then

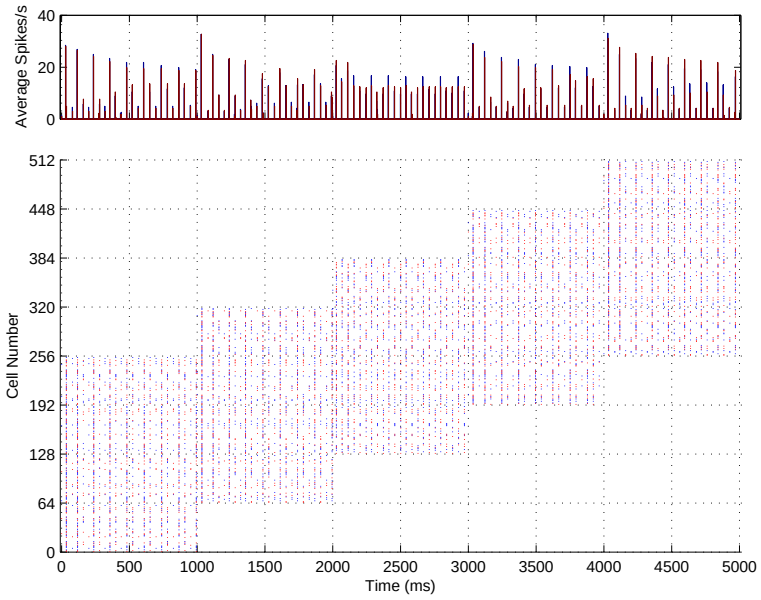
analysed for stimulus selectivity and transformation-invariance. An example where this was achieved through training is presented below.

The effects of the preliminary training can be seen in Figure 7.8. The figure shows the spike rasters of the input layer neurons as a novel pair of stimuli from the two stimulus groups are presented together translating across the input layer before training (Figure 7.8a) and after training (Figure 7.8b). Similarly to the single layer simulations presented earlier, the training changes the input volleys from disorganized fragments of each stimulus firing synchronously (Figure 7.8a) to the antiphase oscillations of each stimulus ensemble in turn (Figure 7.8b). The dynamic of perceptual cycles is maintained as the stimuli translate together across the input layer such that they are still temporally separate despite their concerted movement — the Gestalt prägnanz law of 'common fate'.

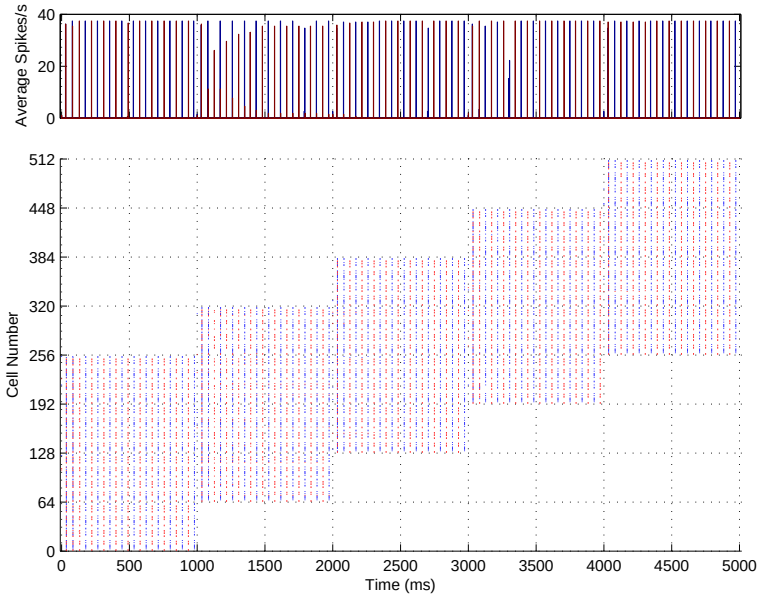
Having established perceptual cycles in the combined novel stimuli which form the inputs to the network, the feed-forward connections between the layers of excitatory cells (EfE) are able to exploit this organization (temporal separation of stimuli) during the secondary training phase with spike-time dependent plasticity. Thus, even though the network was presented with both novel stimuli simultaneously translating across the input layer, many output neurons have learnt to respond discriminatingly to one stimulus exclusively, across all of its transforms as shown in Figure 7.9. In the raster plot of the output layer's excitatory neurons, each transform of the first stimulus is presented in turn for $1000\,ms$ (accounting for the first $5000\,ms$) followed by each transform of the second stimulus (accounting for the second $5000\,ms$).

The raster plot of the output layer (Figure 7.9) shows that almost all the cells have become selective for one stimulus or the other, and respond invariantly across all transforms of their preferred stimulus, firing continuously for either 0–5000 ms or 5000–1000 ms. The cell firing rate plots of Figure 7.9 depict the same information in an alternative form where the transform presentation periods are discretized as individual plots and the firing rates within those periods are represented by the heat map. It can be seen that the same patches of output cells remain active (shown in red) across all transforms of one stimulus, and then are quiescent (shown in blue) during presentations of the other stimulus. The two groups of output cells can be seen to be almost mutually exclusive between the two stimuli.

Presenting each stimulus individually during testing allows for the specific responses to each particular stimulus to be measured, so that the information conveyed by the firing rates may be quantified and is presented in Figure 7.10. The single cell information measure of Figure 7.10a shows that almost all neurons convey the



(a) Before training



(b) After training

Figure 7.8: Spike rasters of input layer principal cells before and after training the excitatory lateral connections. A novel pair of stimuli from each group are presented together translating across the input layer. Prior to learning, both populations (representing each of the novel exemplars) are rapidly synchronized through the action of the fully connected inhibitory interneurons (global inhibition). After learning, it can be seen that each population of neurons is internally synchronized and desynchronized with respect to the other population.

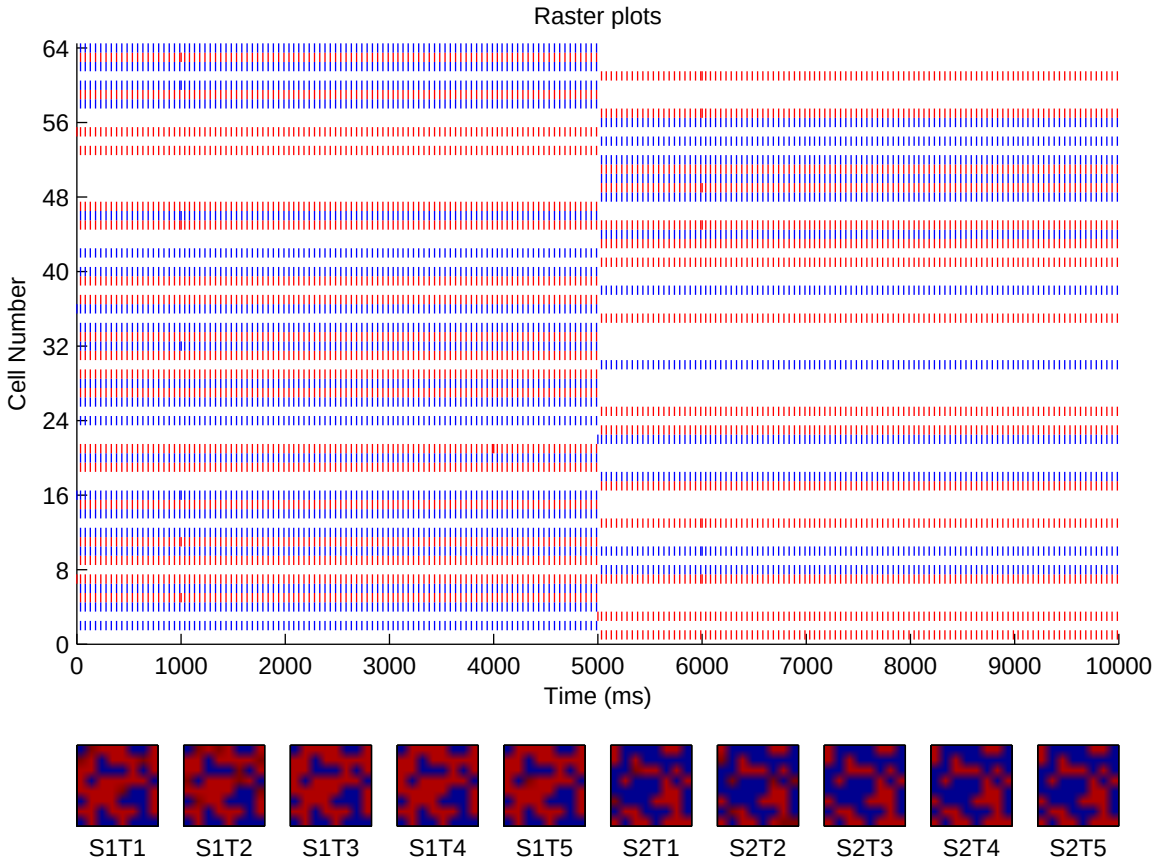
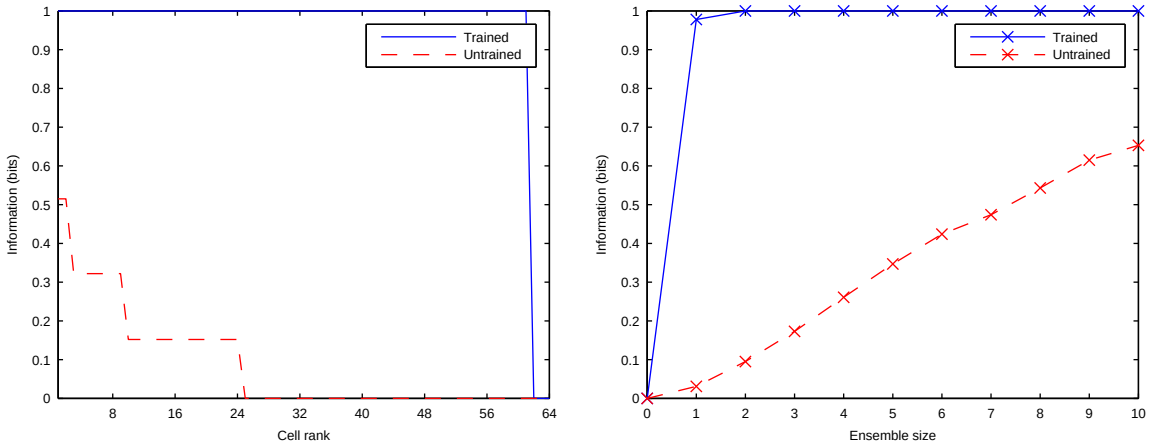


Figure 7.9: Raster plot of the output layer (top) and firing rate plots (bottom) for each transform of each stimulus in the output layer after training. Some excitatory cells in the output layer have become responsive to each transform of Stimulus 1, presented over the first 5000 ms, while other excitatory output cells have become responsive to all transforms of Stimulus 2, presented across the second 5000 ms.

theoretical maximum 1 bit of information meaning that they are perfectly able to distinguish between the two stimuli across all of their transforms. The multiple cell information measure of Figure 7.10b shows that of these trained output neurons, some have become responsive to each of the stimuli, that is, both stimuli are represented.

7.3.2.1 Excitatory feed-forward synaptic strength

While the input representations had self-organized appropriately, the strength of the feed-forward excitatory synapses required some tuning to ensure that there was enough activity to stimulate the network's output layer, but not so much as to saturate the postsynaptic neurons' firing rates. The feed-forward synaptic conductance strengths were initialized randomly, uniformly distributed in the range $[0, \Delta g_{EfE}^{max}]$ and updated (within the same range) according to the same spike-time dependent



(a) Single Cell Information (b) Multiple Cell Information

Figure 7.10: Information plots for the trained and untrained network. The single cell information measure (a) shows that almost all excitatory neurons in the output layer convey the maximum quantity of information for distinguishing between the two stimuli, ($1\,bit$) across all transforms. The multiple cell information measure (b) confirms that both stimuli are represented rather than just one, conveying the maximal quantity of information with an ensemble size of just two neurons.

learning rule used for modifying the lateral excitatory connections. To explore the effects of varying the feed-forward synaptic conductance and find an optimal value for Δg_{EfE}^{max}, a parameter exploration was conducted over the range 1.25–3.75 nS. Network performance was evaluated as the lowest proportion of output cells with at least 95% of the theoretical maximum quantity of information ($\log_2(N)$) to either stimulus, denoted ι_κ (see Equation 5.6) and is shown across the range of synaptic efficacies in Figure 7.11.

From the mean of ten random seeds, an excitatory feed-forward synaptic strength of $3\,nS$ was found to be optimal for training the output layer across all five transforms. The results of one realization (particular random seed) of this parameter are presented previously in Figures 7.8–7.10. For lower values of the feed-forward conductance strength, the input layer was unable to stimulate output neurons adequately (or at all) such that there were no spike responses to record and hence no information in the quiescence. For larger values, too much activity was propagated making it more difficult to distinguish the responses of the output neurons to each stimulus from one another as the firing rates approached saturation point.

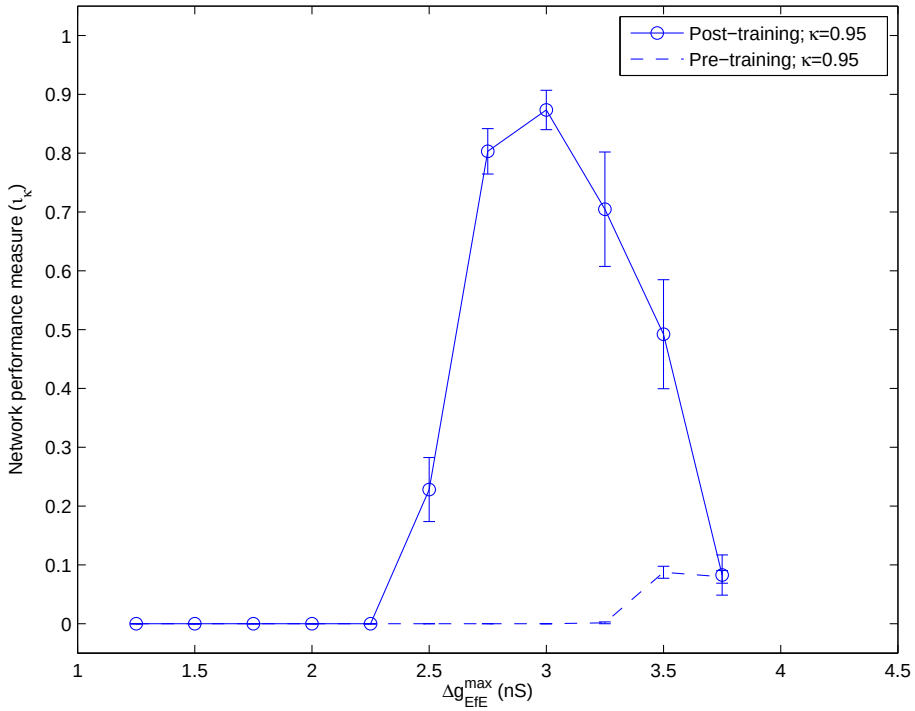


Figure 7.11: Mean network performance (with standard error of the mean denoted by the whiskers) across ten random seeds before and after training against varying the maximum strength of the excitatory feed-forward synapses (Δg^{max}_{EfE}). The optimal network performance (evaluated as the lowest percentage of output cells with at least 95% of maximum information to either stimulus) can be seen to be achieved at $\Delta g^{max}_{EfE} = 3\,nS$.

7.3.2.2 Time constants of synaptic plasticity

Forming antiphase input representations is an important mechanism by which output neurons may learn to form transformation-invariant representations of individual stimuli when multiple stimuli are presented simultaneously. Although, populations of excitatory neurons representing each stimulus receive the same tonic current stimulation over the same time course, the oscillations of the perceptual cycles they form may leave enough of a time gap between representations of each stimulus for a postsynaptic output neuron to learn about a particular stimulus without interference from the others.

In order for this mechanism to work, the excitatory feed-forward synapses require a suitably spike-time sensitive learning rule such that only the input volley immediately preceding an output neuron's spike causes potentiation, and not prior volleys from the representations of other stimuli. If the form of spike-time dependent plasticity used to modify these EfE synapses is made less temporally specific, the oscillations

may then occur on a time scale too short for postsynaptic excitatory neurons to learn about each stimulus in isolation causing input neurons representing transforms of multiple stimuli to be associated onto the same output neurons. Conversely, if the time constants of synaptic plasticity are too small (and so the time windows for the STDP are too short) the spread of spikes within a volley may become too large for the earliest firing neurons to become potentiated onto the same output neuron as those firing later which eventually cause the postsynaptic neuron to fire.

As an illustration of how the shape of the STDP time windows are affected by the time constants, the percentage change in synaptic weight ($^{\Delta g_t}/_{\Delta g_0}\%$) is shown in Figure 7.12 for three different values of the initial synaptic weight (Δg_0) for τ_C/τ_D pairings of 3/5, 15/25 and $45/75\,ms$. Note how the long term depression (LTD) when $t_{post} - t_{pre} < 0$ is constant for variations in Δg_0 in accordance with Perrinet et al. (2001).

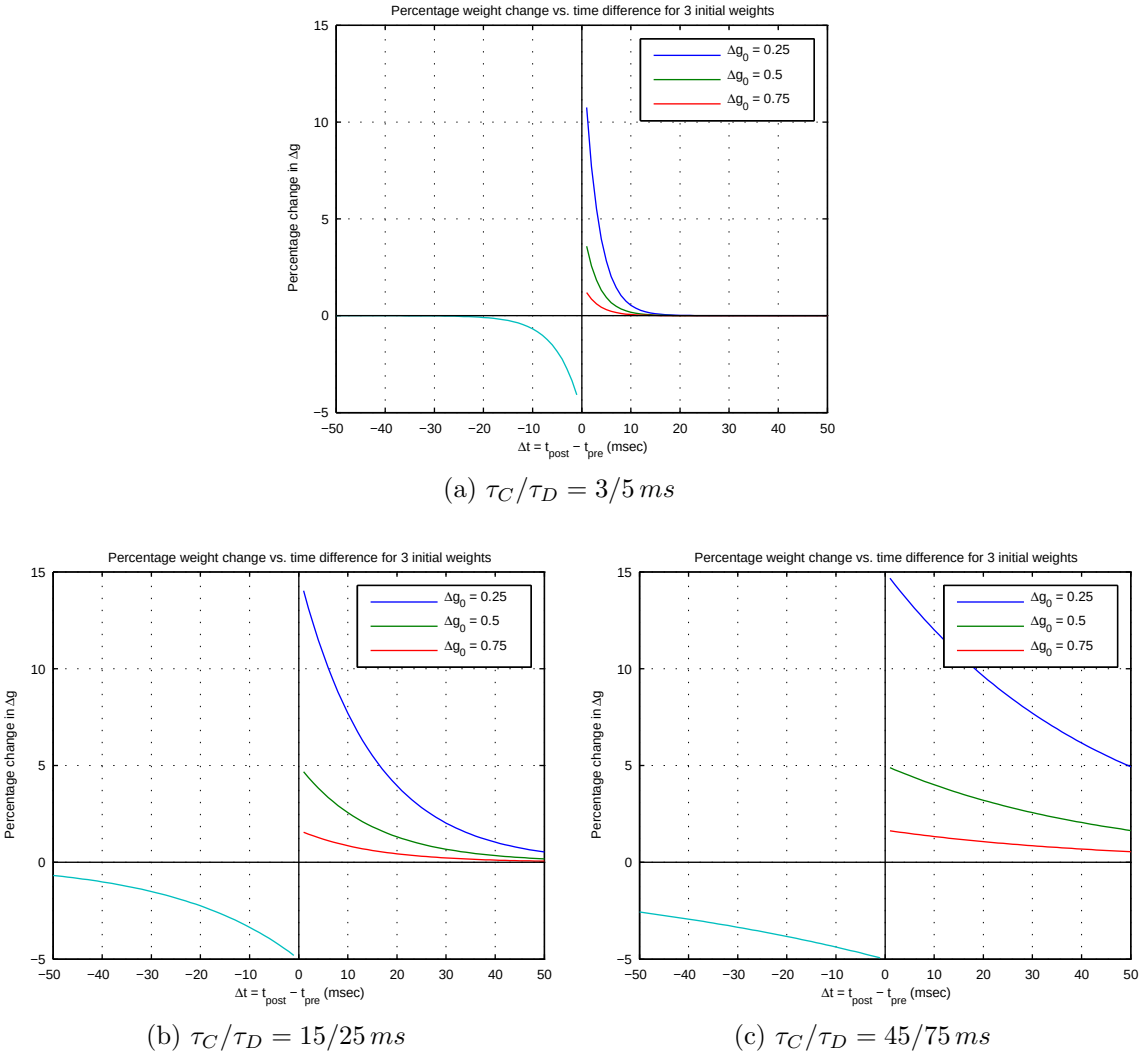
To test these hypotheses regarding the time constants of the STDP model used (τ_C and τ_D) these parameters were systematically varied while maintaining a constant ratio of $^{\tau_C}/_{\tau_D} = 0.6$. Figure 7.13 shows the results of simulations varying the STDP time constants from 3/5 to $150/250\,ms$ for ten different random seeds showing the mean network performance (and standard error of the mean) across these random seeds.

As expected, the network's performance is quite sensitive to the STDP time constants, achieving good performance in the region of 12/20 to $45/75\,ms$.

Rerunning the same simulations but with symmetric time constants ($\tau_C = \tau_D$) resulted in slightly lower performance overall and also longer time constants for optimal performance over the same ten random seeds at $50\,ms$ as shown in Figure 7.14.

7.4 Discussion

Here it has been demonstrated how introducing plasticity into excitatory lateral connections enables them to encode information about the category to which stimuli belong, through associating together co-occurring features within each category. Once modified through exposure to several category members, these lateral connections were then able to segment a visual scene composed of two novel exemplars (one from each category) by synchronizing the features within a particular stimulus and desynchronizing each stimulus representation with respect to the other — a dynamic known as 'perceptual cycles' (Miconi and VanRullen, 2010), as investigated in Section 7.3.1. In a similar way to previous work, a mechanism of delayed self-inhibition was necessary



(a) $\tau_C/\tau_D = 3/5\,ms$

(b) $\tau_C/\tau_D = 15/25\,ms$

(c) $\tau_C/\tau_D = 45/75\,ms$

Figure 7.12: Spike-time dependent learning rule function for three pairs of time constants over a time window of $\pm 50\,ms$. The change in synaptic efficacy or "weight", Δg is expressed as a percentage of the initial synaptic efficacy, Δg_0 and is given for three different initial values. Note that in this form of STDP, the percentage weight decrement does not depend upon the initial value.

to achieve organization of the input representations' perceptual cycles, which in this case was cell firing-rate adaptation (rather than fixed delays), which is a common property of cortical neurons.

By augmenting the network with a second output layer of excitatory neurons linked by feed-forward, plastic synaptic connections, Section 7.3.2 shows how the perceptual cycles generated in the input layer by the previously learned category training could be exploited through a spike-time sensitive learning rule in order to learn separate transformation-invariant representations of each new exemplar in the output layer

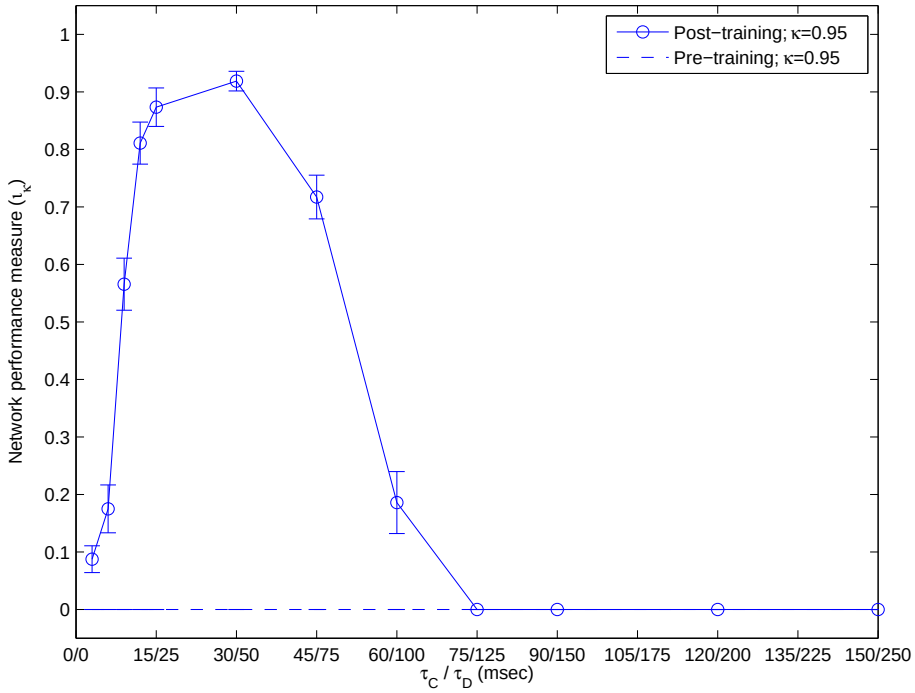


Figure 7.13: Mean network performance (with standard error of the mean denoted by the whiskers) across ten random seeds before and after training plotted against varying the STDP plasticity time constants (τ_C, τ_D). The optimal network performance (evaluated as the lowest percentage of output cells with at least 95% of maximum information to either stimulus) can be seen to be achieved at $\tau_C/\tau_D = 15/25\,ms$.

through modification of the feed-forward connections. The formation of perceptual cycles (antiphase oscillations of the input representations) and subsequent network performance in learning independent and invariant stimulus representations in the output layer was found to be robust to a wide range of several parameters.

In contrast to the earlier work of Chapter 5, where the synchronization of a visual stimulus along with the desynchronization between different stimuli was achieved through spatial separation and a Gaussian profile of fixed lateral excitatory connection strengths, the same organization of the input representations was achieved through encoding prior experience of stimulus categories in the initially weak lateral excitatory connections. This meant that the stimuli did not need to be physically separated, exhibit independent motion or be statistically decoupled by presentation with many other stimuli in order to segment the visual scene in such a way as to allow the subsequent layer of principal cells to form independent transformation-invariant representations of each stimulus through modifications of the feed-forward synaptic efficacies between the layers. By replacing the fixed lateral SOM connections of Chapter 5 with the

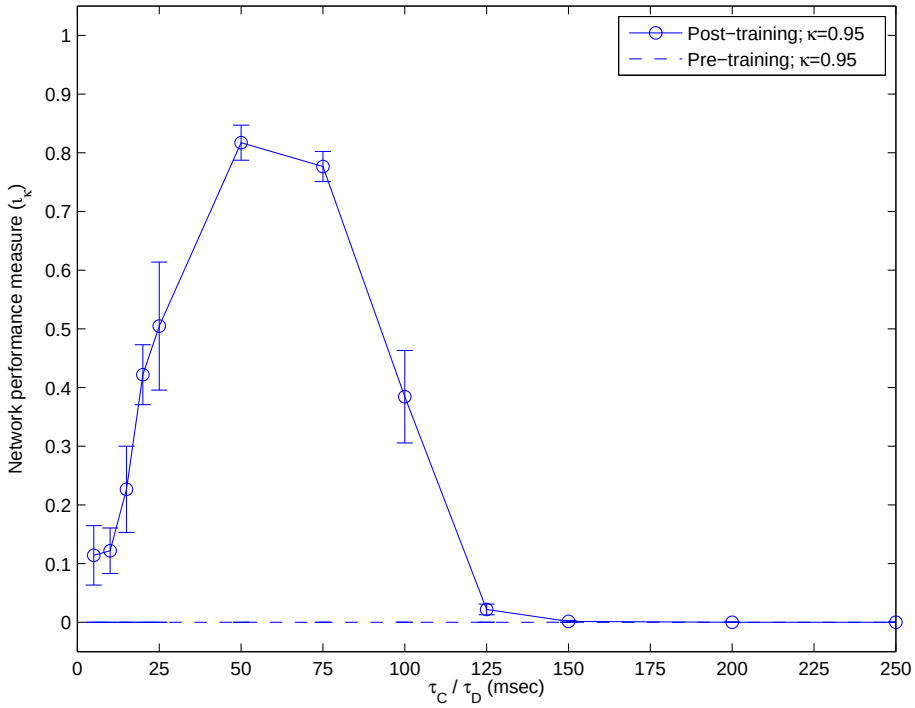


Figure 7.14: Mean network performance (with standard error of the mean denoted by the whiskers) across ten random seeds before and after training plotted against varying the STDP plasticity time constants symmetrically ($\tau_C = \tau_D$). The optimal network performance (evaluated as the lowest percentage of output cells with at least 95% of maximum information to either stimulus) can be seen to be achieved at $\tau_C = \tau_D = 50\,ms$.

plastic ones incorporated in this chapter, the model gains the additional abilities as follows:

(i) Learn about categories of objects (by associating together the features common to their exemplars) and encode this information in the lateral connections between the excitatory neurons.

(ii) Utilize prior category knowledge encoded in the lateral connections to temporally segment novel exemplars of previously experienced categories when presented together.

(iii) Exploit the temporal segmentation in the input layer with STDP in the feed-forward excitatory connections to simultaneously learn independent transformation-invariant representations of novel translating stimuli.

The categories of stimuli used in these simulations were artificial and highly idealized in the respect that there were no features in common between them. In the

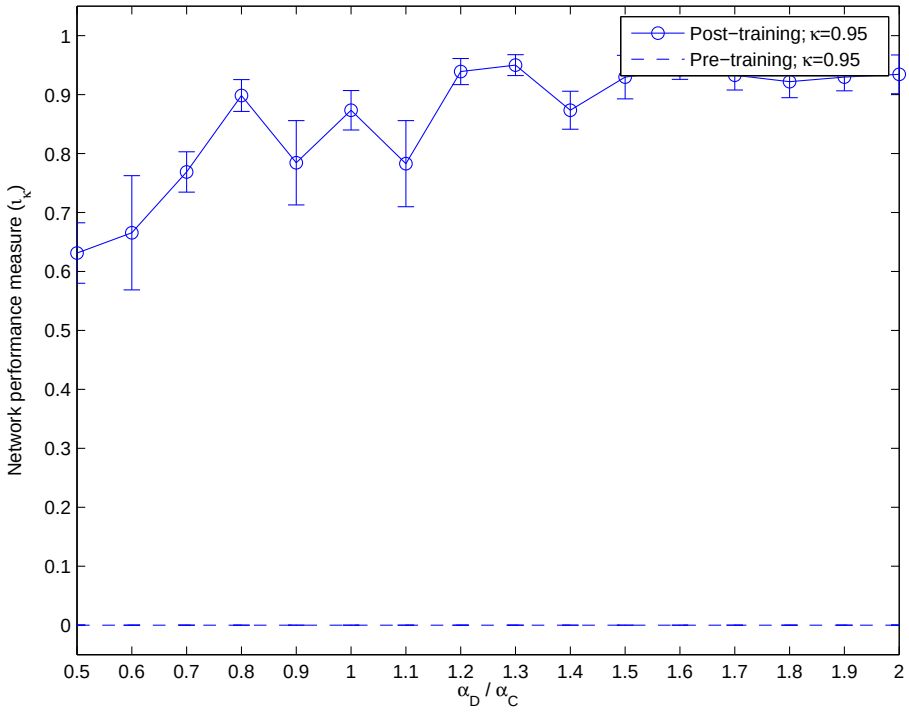


Figure 7.15: Mean network performance (with standard error of the mean denoted by the whiskers) across ten random seeds before and after training plotted against varying the STDP plasticity α constants ratio (α_D/α_C). Good network performance (evaluated as the lowest percentage of output cells with at least 95% of maximum information to either stimulus) can be seen to be achieved for ratios of approximately 0.8 and above.

visual system however, groups or categories of stimuli may share several features with other groups, which if present in particular exemplars of each group and constituting a large proportion of the visible features, would tend to bring the stimuli into synchrony with one another. With more natural stimuli however this is not expected to be a problem, since during early visual experience, the statistical properties of the features would naturally form categories according to the frequency with which they are associated, a consequence of which would be that the more prototypical features of a group would be most strongly associated together while those features also found in members of other groups would be relatively weak and therefore not exert a strong enough influence to synchronize the two exemplars in the presence of the more canonical features. Furthermore, other mechanisms of differentiation may be available such as spatial separation or independent movement which would help to temporally segment a natural scene. This would be interesting to explore in a more detailed model using filtered images as stimuli.

The same processes demonstrated here to encode group identity in lateral connections may be considered in another way. If we consider the visual world to be more commonly composed of continuous contours rather than discontinuous, abrupt changes in orientation, low level visual features with similar orientation preferences would more commonly be coactive than those sets with perpendicular orientation preferences. Such statistics, combined with the lateral plasticity as described, would naturally lead to lateral connections between similar receptive field (orientation) preferences as are found in the visual cortex of several species (Malach et al., 1994). Such learned connections would then help to segment partially occluded stimuli, based for example, on their texture or continuation of edges and would provide an explanation for perceptual effects such as illusory contours.

The purpose of computing is insight, not numbers.
—Richard W. Hamming

8 Discussion

The experimental chapters of this thesis are now complete so this chapter is devoted to considering this body of work in its entirety and discussing the wider implications of the work so far presented for the process of object recognition in the primate visual system. First, the findings of each experimental chapter are summarized, highlighting the novel contributions of the work and drawing out some of the overarching themes which have emerged from the modelling investigations. These insights are related back to the learning processes occurring in the brain and some suggestions are provided for future investigations.

8.1 Summary of Investigations and Findings

The starting point for this thesis was to investigate the extent to which the Hebbian learning mechanisms, of CT (Stringer et al., 2006) and Trace learning (Földiák, 1991), may be implemented with a more biologically accurate spike-time dependent learning rule in a spiking neural network for learning to recognize objects in a model of the ventral visual stream.

Both the Trace and CT learning mechanisms have been previously applied successfully to this problem in rate-coded models to form transformation-invariant representations with both temporal continuity (Wallis and Rolls, 1997; Wyss et al., 2006; Michler et al., 2009; Isik et al., 2012) and spatial continuity (Stringer et al., 2006; Perry et al., 2010) of the stimuli respectively. The simulations conducted in Chapter 3

explored for the first time how these learning paradigms may be successfully adapted to a spiking neural network featuring STDP in its feed-forward excitatory synapses.

It was shown that with appropriate modification of the stimulus presentation and changing the excitatory synaptic conductance time constant, τ_g^{EfE}, (modelling the dynamics of the postsynaptic ion channels), the model could operate in two dissociable modes to learn translation-invariant representations. In particular, a short synaptic time constant and high degree of spatial overlap between transforms allowed the model to operate in a manner akin to CT learning, whereas a long synaptic time constant facilitated a form of Trace learning, where the network could associate together completely non-overlapping (orthogonal) views of the stimuli onto the same output neurons.

While both mechanisms were found to be robust to the order of transform presentation within a stimulus block, Trace learning suffered when the transforms of two stimuli were interleaved as expected from a violation of the assumption of temporal continuity of stimuli within natural scenes (Földiák, 1991; Wallis and Rolls, 1997; Sprekeler et al., 2007; Li and DiCarlo, 2008, 2010, 2012). The CT learning mechanism was found to require tight synchrony of the input volleys and temporally specific STDP time constants, whereas Trace learning was more robust to jitter and performed better with longer STDP time constants.

Closely following the simulations of simple stimuli used to investigate the transformation-invariance learning in Chapter 3, the aim of Chapter 4 was to demonstrate that the same principles apply when using more realistic three-dimensional stimuli. Through using a combination of the Trace and CT learning mechanisms, it was demonstrated that the network could be trained to recognize both a cone and a cube as they translated across the input layer. Furthermore, it was demonstrated how additional sorts of transformation-invariance, such as view invariance, may be learnt through the same mechanisms by training the network to recognize the same stimuli independently of their view as they rotated in depth.

Some novel problems arose through using filtered greyscale images, such as the difficulty in representing particular transforms of low contrast with respect to the grey background. In a fuller model incorporating details of colour vision, the use of colour stimuli would avoid this problem, as the borders could also be detected according to hue rather than just contrast. While this is slightly artefactual and not a major challenge to the biological plausibility of such a model, it does highlight some of biological details which have been neglected and thus might be fruitfully incorporated in future revisions of the model.

A central question in understanding the process of biological object recognition is how to form separate representations of objects when natural visual scenes consist of multiple objects. Since previous work has largely concentrated on presenting stimuli in isolation (Wallis and Rolls, 1997; Stringer et al., 2006; Riesenhuber and Poggio, 1999; Masquelier and Thorpe, 2007), in this respect these models are lacking in ecological validity. Rate coded models in particular suffer in this respect, as a rate-based model of Hebbian learning predicts that stimuli presented together will be associated together leading to a superposition catastrophe (von der Malsburg, 1999). This is a theme addressed in several ways over the course of Chapters 5–7.

In Chapter 5, the model was augmented with a lateral connectivity architecture similar to the standard Self-Organizing Map (Kohonen, 1982; Miikkulainen et al., 2005) to address the question of how the extensive lateral connectivity of the visual system may aid learning to recognize objects. Based upon previous analyses of laterally connected spiking neurons (Nischwitz and Glünder, 1995; Miconi and VanRullen, 2010) the architecture and neuronal properties of the network were expected to enable competing stimulus representations to push each other out of phase. In particular, the hypothesis tested was that with such antiphase oscillations (perceptual cycles) between input layer representations, a time-sensitive STDP learning rule in the feed-forward synapses would be able to learn separate stimulus representations in the output layer as they translated across the input layer.

The key properties of the network were found to be a mechanism of self-inhibition, consistent with previous work (Choe and Miikkulainen, 2003; Miconi and VanRullen, 2010) — in this case, cell firing-rate adaptation — and a gradient in the excitatory synaptic weights (which decrease exponentially with distance). Variations in these properties were investigated with respect to how they help the network to synchronize features represented by proximal neurons and desynchronize them with respect to features represented by distal neurons and therefore temporally segment the input representations in a way suitable for transformation-invariant learning with STDP.

Following from the conclusions of Chapter 3, specifically that CT learning requires tight synchronization of spikes in an input volley (as slightly later spikes fall into the STDP rule's LTD window) and that Trace learning is less temporally specific (with respect to input synchrony and the STDP time constants), only the CT learning mechanism was explored in these simulations, as a Trace learning paradigm would undermine the dynamics of the temporal segmentation of stimuli.

The work showed how the spatial separation of the stimuli is effectively converted into a temporal separation (through the lateral weight gradient and firing-rate adapta-

tion) in a way that allows separate translation-invariant representations to be formed even when both stimuli are always presented together (with no statistical decoupling) and moving in lockstep. This is an advancement on previous work in rate-coded networks with multiple stimulus presentation paradigms, where either ecologically implausible training regimes were required to statistically decouple the objects (Stringer et al., 2007; Stringer and Rolls, 2008), or independent motion (Tromans et al., 2012).

The power of neurons lies in their ability to communicate with one another through their synapses. Their connections may be very short or extend across millimetres of cortex (Miikkulainen et al., 2005) and so the delay in conducting their action potentials along their axons (with or without myelination) varies accordingly from the order of one millisecond (Girard et al., 2001) to the order of tens (Swadlow, 1992) or even hundreds of milliseconds (Aston-Jones et al., 1985). Following from these experimental data, Chapter 6 investigated the role of axonal conduction delays, firstly in the feed-forward axons of a variant of the model used in Chapter 3, and secondly as a replacement for the Gaussian profile of lateral excitatory weights used in Chapter 5.

In using a constant axonal conduction delay for the feed-forward connections, both the CT and Trace learning mechanisms were unaffected in their capacity to learn transformation-invariant representations in the output layer when the STDP rule also incorporated the delay term. This was a novel change to the original formulation of Perrinet et al. (2001), which like many other models of STDP, neglects axonal delays (Song et al., 2000). When the unmodified version of STDP was implemented in the model, there was a pronounced and different effect between the CT and Trace learning regimes as the axonal conduction delays were increased. These were argued to be artefactual however and that when implementing axonal delays in such a model, the STDP rule should also be revised accordingly.

When Gaussian distributions of delays in the feed-forward connections were investigated, another marked difference between the CT and Trace learning mechanisms was discovered, whereby the network performance of CT learning decreased rapidly with the increasing standard deviation of the distribution, whereas Trace learning was far more robust, as expected from their respective differences in sensitivity to the synchrony of the inputs.

The second section considered the computational benefits of axonal conduction delays (rather than viewing them as a potential difficulty for the model to overcome) by demonstrating that they can replace the role of the gradient in the excitatory lateral connection strengths to temporally segment multiple stimuli through the dynamics of perceptual cycles. The delays in this section were proportional to the Euclidean

distance between the neurons giving a graduation of delays and thereby extending the simple two neuron case of Nischwitz and Glünder (1995) and the single neural layer of Miconi and VanRullen (2010) which each involved a fixed delay. It was predicted that the optimal delay between the stimulus representations was half the period of oscillation, such that the wave of excitation arrives midway through the cycle of the afferent neurons, encouraging their targets to fire at that point. The results however indicated that a delay $1\,ms$ lower was very slightly better for network performance, which may be due to the time required for the synaptic conductances to affect the postsynaptic membrane potential.

In the final body of experimental work, Chapter 7 investigates a third solution to the problem of separating representations of multiple translating stimuli. Rather than imposing a fixed lateral weight structure or attempting to match the lateral delays to the (half) period of oscillation, plasticity is introduced into the lateral connections and a preliminary training phase is introduced to allow them to self-organize through exposure to exemplars of categories. During this preliminary phase, knowledge of the categories is built up in the lateral connections which is then found to enable the network to temporally segment a visual scene composed of two novel exemplars.

While initially the effects of 'early' exemplar exposure are investigated in a single layer of excitatory neurons (and inhibitory interneurons), the second section augments this layer with another and adds feed-forward plastic connections between them, and translating stimuli are presented to the input layer. This model is then used to demonstrate that category learning in the lateral connections allows the output layer to form independent translation-invariant representations for previously unseen, stimulus representations. In addition to the lack of statistical decoupling and independent movement, in contrast to the work of Chapters 5 and 6, the use of distributed stimulus representations meant that spatial separation was not required to temporally segment the stimuli.

The input layer dynamics generated in this model were found to be robust to wide ranges of several parameters, including the strength of inhibitory and excitatory connections and the adaptation model strength and time constant. The effect upon the subsequent training of the feed-forward connections was also explored, principally through systematically varying the parameters of the STDP model, which was found to be similarly robust to the earlier simulations of this thesis which involved multiple stimulus presentation paradigms.

Having summarized the main findings of this work, some of the overarching themes are discussed along with areas where the work could be improved.

8.2 Learning mechanisms

Insofar as it aimed to map the existing rate-coded models of CT and Trace learning onto a more biologically detailed model of STDP, Chapter 3 was novel and successful. To investigate these learning rules in a controlled way, each was studied in isolation by examining highly idealized simulation paradigms designed to eliminate one mechanism as far as possible. This therefore still leaves open the questions of whether a balance is struck between the two extremes in the brain or perhaps each is employed in different regions.

The results suggest that CT may be more effective with smooth transformations such as changes in scale or pose, whereas Trace may cope better with discontinuous changes such as 'jumps' in retinal positions due to saccades (Cox et al., 2005). There is likely to be an element of Trace learning throughout the early layers of the ventral visual stream at least, due to the continual changes due to saccades (roughly every 200–300 ms in humans) and micro-saccades an order of magnitude more frequent (VanRullen and Thorpe, 2002; Masquelier and Thorpe, 2007). According to the analysis of the principle of 'slowness', a learning window with a width of tens of milliseconds could filter out changes occurring on the scale of hundreds of milliseconds and thus STDP could become invariant to (micro-)saccades and small translations (Sprekeler et al., 2007). This postulate appears to be in agreement with physiological data for the STDP timing windows (Bi and Poo, 1998).

The differences in sensitivity to timing and synchrony consequently lead to the prediction that Trace learning (that is to say, longer excitatory synaptic time constants) may be a more prominent feature of lower visual areas where the changes are more abrupt. Conversely, CT learning may be adequate for building more complex invariances once lower levels of the visual system have already developed some tolerance to transformations and thus present more smoothly varying representations (with a greater degree of 'spatial overlap') to their higher-level targets.

However, if the Trace learning mechanism does prove to be more dominant in lower visual areas, where multiple representations compete, this may be a problem for the mechanisms of temporal segmentation explored in this thesis. The Trace learning mechanism would undermine the ability of the next layer on to learn about the objects independently, as activity would be carried over from one oscillation to the next through longer synaptic time constants causing the two objects to become associated (Cox et al., 2005; Li and DiCarlo, 2008). This problem may be ameliorated however, in a more detailed model of the architecture featuring small, convergent

receptive fields rather than the full feed-forward connectivity used throughout this thesis, which would help to 'window' spatially separate stimuli. More finely adjusted cell-firing rate adaptation in the postsynaptic neurons with respect to the frequency of oscillations may also help by desensitizing the postsynaptic neuron immediately after it fires in response to one stimulus, such that it does not respond to the next volley of spikes from the other stimulus but recovers in time for the next volley from the original stimulus.

Trace learning was not used for several chapters because it is harder to tune the parameters to obtain good network performance when learning transformation-invariant object representations, as the longer excitatory synaptic time constants generate more activity in the postsynaptic neurons. While this was manageable in feed-forward connections, where it resulted in high firing rates in the output layer, this was expected to be more problematic in the lateral excitatory connections due to the positive feedback of activity created by the recurrent connectivity.

In general, Trace learning is less dependent upon tightly synchronized volleys of spikes representing a particular stimulus as evidenced by its robustness to large variations in the spread of axonal conduction delays. Provided the temporal statistics of the visual environment hold whereby different views (transforms) of the same objects are more commonly close together in time than views of different stimuli (Földiák, 1991), this is a powerful mechanism through which STDP can associate different transforms of an object (including translations and rotations) into an invariant representation to support object recognition.

CT learning was generally found to be a robust mechanism given sufficient similarity between the transforms of a particular object, although it was found to be sensitive to the degree of synchronization within an ensemble of presynaptic neurons representing a particular percept (stimulus). In a sense, this makes it a more precise learning mechanism than the Trace rule, as the CT mechanism can harness the temporal precision of STDP to learn independently about the temporally segmented representations of individual objects in a natural visual scene.

8.3 Lateral connectivity

In keeping with previous work, lateral connections have been shown to help push two competing representations out of phase with respect to each other's oscillations (Nischwitz and Glünder, 1995; Miconi and VanRullen, 2010). Earlier studies were extended to show that this dynamic could occur from either a SOM-like gradient

of lateral excitatory connection strengths, axonal conduction delays or prior learning about category members. Importantly, it was shown that these dynamics are maintained as the oscillating stimuli translate across the input layer and that the perceptual cycles are a useful basis for a temporally specific STDP learning rule to build translation-invariant representations in the next layer of excitatory cells.

This addresses a fundamental problem in visual object recognition and a significant limitation of previous models which were trained with only one stimulus at a time (Wallis and Rolls, 1997; Riesenhuber and Poggio, 1999). Given the ubiquity of lateral connections in the visual cortex and the robustness of this phenomenon, it seems likely that at least one of these variants will be employed in the brain to temporally segment natural scenes consisting of multiple stimuli. Generating these dynamics and subsequently building invariant individual representations from them is a natural property to emerge from modelling individual spikes and as such, is a mechanism which could not be easily or parsimoniously implemented in prevailing rate-coded models of the ventral visual system. These findings suggest that rate-coded models constitute an inadequate level of description to fully capture and understand the learning processes involved in the primate visual system. The results therefore support the argument for needing spiking neural networks to adequately explain the dynamics of biological object recognition.

There may be further modelling benefits to combining some of these features in a spiking neural network which are not captured in rate-coded models. In the classic SOM, the learning rate and neighbourhood size are gradually decreased during learning. If plasticity with an STDP learning rule is combined with axonal conduction delays in the lateral connections, I would expect this unsupervised learning mechanism to become an emergent property of such a network. Since STDP has the useful function of reducing the latency of processing, (as studied in feed-forward networks) and reducing the weight changes near maximal synaptic efficacies in multiplicative forms, the same mechanism applied to lateral connections may therefore explain the emergence of computational maps in the visual cortex (Miikkulainen et al., 2005).

8.4 Visual environment

CT learning was enhanced by interleaving the stimuli during the original invariance learning investigations of Chapter 3. It was also found that learning in a multiple stimulus presentation paradigm (with a SOM, axonal conduction delays or plastic lateral connections) was better than these original demonstrations as was found to

be the case by other work Spratling (2005). Could these observations result from the same underlying principle?

The oscillations of different stimuli in perceptual cycles (as in Chapters 5, 6 and 7) or the interleaving of different stimuli's successive transforms when presented individually (as in Chapter 3) works in two ways. First, new features of the next transform of a particular stimulus emit spikes in phase with those representing the features in common with the previous transform, (since both groups of features are driven up at the same time due to the oscillations) making the learning updates more effective at associating all the features of a particular stimulus together onto the same output neurons. The second aspect to this training paradigm is that when volleys of spikes from a different stimulus are presented soon after the postsynaptic cells fire to the previous stimulus, then they are close together in time but in the acausal temporal order leading to LTD. This therefore helps to dissociate those output neurons from the immediately proceeding stimulus's input spikes, meaning that useful synaptic efficacy updates occur in *both* directions.

Following from the analysis of Song et al. (2000), this process may also be viewed from another perspective. The STDP learning rule is a competitive form of Hebbian learning which allows synapses from clusters of input neurons with synchronized action potentials to grow stronger while weakening other synapses that are not part of the cluster, while remaining insensitive to the degree of variability in the firing of the inputs (Song et al., 2000). This can be exploited by the CT learning mechanism to associate different transforms of the same object onto the same output neurons (provided there is enough overlap of features within the set of transforms). However, according to predictions of STDP simulations, the correlations should decay rapidly otherwise this trend disappears due to the larger LTD integral (Song et al., 2000) which may explain the overall better performance of the model in this work when presented with two stimuli and adapting neurons which generate antiphase oscillations (as in Chapters 5, 6 and 7). These perceptual cycles effectively serve to make the correlations decay very rapidly such that each 'cluster' of inputs representing one stimulus is potentiated onto a set of output neurons then rapidly silenced while another stimulus cluster potentiates onto a different set of outputs, with each cluster therefore avoiding the LTD effects which otherwise hamper learning in the paradigm of continuous activity due to a single stimulus (Chapter 3).

8.5 Limitations

Whilst care and effort were taken in designing the simulations and data analyses, there are several limitations of the work presented. This section details these limitations of the thesis which might usefully be addressed in future studies and provides advice for how to improve upon the approaches taken here to aid further progress in this field.

8.5.1 Network Performance

Throughout this thesis and a large body of previous work, cells in the output layer are assessed by means of the information conveyed by their firing rates with respect to the stimulus being presented (Tovée et al., 1994; Sugase et al., 1999; Rolls and Milward, 2000; Elliffe et al., 2002; Evans and Stringer, 2012). Reasoning that when the conditions for learning were optimal, a large proportion of the cells in the output layer would convey (near) the maximum level of information for a single cell led to the choice of ι_κ as a network performance measure (see Equation 5.6). Starting from a point in the parameter space with a high value of this metric (a large proportion of maximally informative cells) showed fairly clearly the effects of variations in the parameters of interest in terms of the network performance measure.

While this approach of considering changes in ι_κ with respect to particular parameters provided an insight into their effect upon the learning process, simply maximizing the proportion of cells that learn to encode the stimuli cannot be regarded as the aim of the biological system, making this performance measure slightly misleading. An individual must learn to recognize many objects and faces throughout its lifetime and it would therefore be inefficient to devote a large proportion of the neurons available to recognizing a single stimulus. Recently it has been argued that due to the large numbers and longevity of such memories, a system that learns to represent each stimulus with a very small number of reliable neurons would be better equipped to serve the demands of long-term encoding in the neocortex (Thorpe, 2011).

This would be more apparent if the model was trained using more stimuli (provided that the simulation run-time could be lowered sufficiently). Assuming an equal probability of presentation for each stimulus, the maximum amount of information which any cell could convey about the presented stimulus (the 'information ceiling') would rise logarithmically to $I_{max} = \log_2(N_S)$, where N_S is the number of stimuli. Network performance could then be more plausibly assessed by how close to the (potentially much higher) information ceiling cells climb after training, rather than the proportion of all output cells achieving a relatively modest target of 1 *bit*. Alternatively

if the network still produces cells at or near the information ceiling, the percentage of correct identifications (or classifications) may be plotted against the ensemble size (Hung et al., 2005, see, for example) as an alternative performance function.

With a shift from considering the proportion of output cells to reach the information ceiling, to achieving instead a very small number of reliable (maximally informative) cells, further alternative performance measures may be considered. If larger sets of stimuli are used, confusion matrices may be plotted to show, not only identification performance, but also the patterns of mistakes for direct comparison to human psychophysical data (Quian Quiroga and Panzeri, 2009). Another useful metric would be to graph the Receiver Operating Characteristic (ROC) curve, whereby the true positive rate is plotted against the false positive rate to assess recognition performance, as is commonly used in psychophysical studies and the evaluation of similar models with respect to benchmark machine vision algorithms (Serre et al., 2007).

Having achieved a small group of maximally informative cells, one such approach would be to then assess the network performance by reducing the number of epochs or the transform presentation periods in order to determine the minimum amount of training (in milliseconds) required for robust performance at that point in the parameter space, or equivalently plotting error rate against training extent (Krizhevsky et al., 2012). Alternatively, the area under the ROC curve could be plotted against the total training time to visualize how the efficiency of learning compared to other models and humans (Serre et al., 2007; Krizhevsky et al., 2012). Such a performance measure based upon the speed of *learning* would compliment previous investigations into the remarkably rapid speed of *recognition* (Thorpe et al., 1996; Hung et al., 2005; Afraz et al., 2006) achieved in a 'post-learning' visual system.

8.5.2 Temporal Analysis

In spiking neural networks time is explicitly modelled by virtue of encoding the precise occurrence of spiking events and in many clock-driven variants, the evolution of solution variables through time. This explicit treatment of time allows for a more detailed investigation of not only the temporal properties of input (and intermediate) representations, but also the time course of several waves of information which gradually transition from 'global categories' to 'fine categories' (e.g. ultimately identity and expression) in the responses of neurons recorded in both macaques and humans (Sugase-Miyamoto et al., 2011).

Intracranial field potentials recorded in epilepsy patients have shown that object category information can be reliably decoded from the visual cortex as early as $100\,ms$

post-stimulus, with performance robust to depth rotation and scale changes (Liu et al., 2009). The speed at which this stimulus-category information became available in the temporal lobe was offered as evidence in support of feed-forward theories of visual processing, commensurate with earlier work and theorizing (Thorpe et al., 1996; Hung et al., 2005; Afraz et al., 2006).

However, evidence from sliding temporal window analyses of the fMRI activity of face-processing areas of the macaque (Anterior Medial and Anterior Lateral) shows that the nature of representations changes substantially through time, tending to increase identity selectivity in a view-invariant manner (Freiwald and Tsao, 2010). Contrary to earlier theorizing about visual processing, this change with the passage of time occurs in a manner not directly reconcilable with a simple feed-forward model, suggesting additional recurrent mechanisms both within face patches and between face patches at different levels.

Earlier work using single cell recordings from Macaque Inferotemporal cortex with a $50\,ms$ sliding time window revealed a more detailed structure to the time course of information. The earliest part of the responses were found to convey information about stimulus category (i.e. Monkey face vs. Human face vs. Shape), with fine information concerning identity and facial expression beginning $51\,ms$ later (Sugase et al., 1999). The two curves for global and fine facial information were found to reach their peak transmission rates at 117 and $165\,ms$ post stimulus onset respectively.

These data would be interesting to model, and would potentially allow the function and onset time of recurrent connections to be elucidated. The existing studies have calculated the information by binning the neural responses, but in future work calculating the Fisher information (Equation 8.1) may be a more appropriate metric for examining the time course of information in continuous stimuli (Quian Quiroga and Panzeri, 2009). Specific hypotheses about their role and speed of operation could be tested by comparing the time courses of the information functions resulting from particular architectures and parameters to the observed neurophysiological data, thus helping to resolve long-standing debates concerning the direction and time-course of information flow in visual processing.

$$FI = \sum_{r \in R} P\left(r|s\right) \left[\frac{d}{ds} \log P\left(r|s\right)\right]^2 \tag{8.1}$$

In subsequent studies using a saccadic choice paradigm, human participants were found to be able to reliably saccade to the side of their visual field containing an animal in $120\,ms$ (Kirchner and Thorpe, 2006) or as little as $100\,ms$ when the target

stimulus is a face (Crouzet et al., 2010). A follow-up study provided evidence to account for the face recognition bias by showing that the information required to trigger these saccades could be extracted very early on in visual processing using low-level amplitude spectrum information in the Fourier domain (Crouzet and Thorpe, 2011), showing a clear drop in performance when the amplitude spectra were swapped between, for example, a face and a vehicle.

In a future model with more detailed inputs, such early diagnostic information should be seen in the first waves of spikes, providing a basis for very rapid stimulus-category decoding, followed soon after by waves conveying stimulus-identity and even facial expression. The formation of such a 'quick-and-dirty' mechanism for orienting the eyes to particularly salient stimuli could be modelled by adjusting the time spent looking at faces during learning to investigate the resulting sensitivities to their signature amplitude spectrum. Accordingly, one might predict that emphasizing another class of stimulus during learning with its own distinct amplitude spectrum may instead lead to particularly rapid and biased responses to that category.

8.5.3 Stimuli

In Chapter 4, the input layer of the network was a two-dimensional grid of hypercolumns, consisting of banks of Gabor filters, modelling the response properties of V1 neurons. This allowed the network to be presented with realistic three-dimensional shapes as stimuli and allowed greater freedom in terms of the type of transformations which could be simulated. Most simulations however were conducted with only two idealized, one-dimensional, abstract stimuli in order to save on the computational cost of the simulations, albeit at the expense of considerable biological detail. This section reviews some of the limitations specific to the stimuli used and suggests ways in which to advance future simulations in this respect.

8.5.3.1 Small numbers of stimuli

Much of the work in this thesis was conducted with only two stimuli, for the sake of computational tractability. This two-stimulus paradigm represents the minimum for testing the network's ability to discriminate, that is, the 'identification' component of the object recognition problem (DiCarlo and Cox, 2007). The learning problem difficulty was therefore largely attributed to the number of transforms of each stimulus which the network had to generalize across, that is, the 'categorization' component (DiCarlo and Cox, 2007). While it is this latter facet of the object recognition problem,

(exhibiting the generality across natural variations in the input), which is believed to be the computational crux and major difficulty for artificial approaches to visual object recognition (DiCarlo et al., 2012), it would be informative to further explore how the difficulty varies when straining the specificity component with greater numbers of stimuli.

Training and testing the network with greater numbers of stimuli would not only allow for more realistic performance measures by testing how close to the 'information ceiling' the output cells are able to reach as discussed, but also for more ecologically valid network inputs, in keeping with the huge numbers of recognizable objects in the natural world. Using greater numbers of stimuli would therefore allow for manipulation of the statistics of the visual environment, for example, to investigate the formation of perceptual categories based upon co-occurrence of visual features.

More generally, forming an efficient representation of the statistics of the world ('The Efficient Coding Hypothesis') is believed to be the driving force which explains the development and subsequent structure of the visual system (Simoncelli, 2003; Felsen et al., 2005; Karklin and Lewicki, 2009). Empirical studies suggest that the primate visual system is highly sensitive to these stimulus statistics, most likely with a non-trivial scaling (Ruderman and Bialek, 1994). It is also found to be constantly adapting to them, particularly in the specialized domain of face recognition (Webster et al., 2004). Furthermore, models of the spatio-temporal receptive fields of V1 neurons obtain much better fits with natural visual statistics than with synthetic gratings (David et al., 2004; Felsen and Dan, 2005).

It follows from Efficient Coding hypothesis that the primate visual system will likely have adapted to the characteristics of natural stimuli and so should similarly exhibit higher information content (and thus improved performance) under naturalistic stimulus conditions (Felsen and Dan, 2005). Thus, in order to more accurately model the system's response properties and performance, it is important to use more natural visual inputs, albeit in a controlled way to avoid misleading results (Pinto et al., 2008).

8.5.3.2 Feature similarity

With the exception of the filtered images in Chapter 4, the stimuli used throughout this thesis were orthogonal (non-overlapping) representations. An important aspect to consider for future research (but omitted in this work), was the role of feature similarity between stimuli. The simulations of Chapter 7 explored how novel exemplars from different (abstract) stimulus categories could be desynchronized with respect to each other following prior learning of category information which was encoded in the

lateral connections. The stimuli used for this however, were completely orthogonal (although not spatially contiguous). The stimuli used were approximating exemplars mutated from two category prototypes (for example, the prototypical cat and the prototypical dog). Naturally, there would be a much higher feature similarity within a category than between categories, however, the between category similarity was 0 and therefore unrealistic.

Previous psychophysical studies (Preminger et al., 2009) and modelling studies with attractor networks featuring recurrent Hebbian synapses (Blumenfeld et al., 2006; Bernacchia and Amit, 2007) have shown how initially different internal representations can be merged together through learning. According to these studies, when presenting a sequence of stimuli which are morphs between two end points, the two internal representations will gradually collapse due to the changes in the recurrent plastic connections. This effect was observed even with very gradual presentations (over several weeks) in humans, suggesting it may be a way in which for example, memories of a face are updated as that face ages (Preminger et al., 2009). However, presenting the stimuli in random rather than sequential order (Bernacchia and Amit, 2007), biasing the weights towards separate attractor basins (Blumenfeld et al., 2006) or implementing a novelty-facilitated learning rule (Blumenfeld et al., 2006) all counteracted the collapse of the attractor states, revealing several alternative stimulus dimensions to consider as a compliment to feature similarity.

These previous studies were conducted with rate-coded or binarized neurons and learning rules, leaving the question open as to how the principles may apply to spiking neurons with STDP. Nonetheless, these results indicate that the degree of feature sharing between two categories of stimuli is expected to have a significant effect upon the ability of the input neurons to desynchronize representations of exemplars from different categories. Furthermore, the internal clusters of exemplars which form (for example, encoded in the plastic lateral connections) are likely to be very different depending upon the feature similarity within and between different sets of exemplars. Similarly, the order in which examples are presented is also likely to play an important role (Bernacchia and Amit, 2007), whereby we may expect better and more persistent separation of the input representations if learning begins with the most typical exemplars of each category. Manipulating the inter-group feature similarity in a controlled and systematic way would therefore be an interesting and important dimension of the stimulus set to explore in order to further our understanding of how the ventral visual stream may categorize stimuli.

8.5.3.3 Noisy inputs

Recent work has shown that connections between spiking neurons trained with STDP can quickly find the beginning of repeating patterns of spikes embedded within random spiking activity (Masquelier et al., 2008), for example, cars in different lanes of a motorway seen through a silicon retina (Bichler et al., 2012). Furthermore, the strength of inhibitory interactions may be adjusted to limit the total activity and make the probability of responding to a random input for a trained neuron arbitrarily low. By remaining silent in the absence of the cell's preferred stimulus, its synaptic weight structure may remain unchanged for very long periods of time. The possibility of such highly selective neurons being formed rapidly in an unsupervised learning regime has led to the suggestion that a substantial percentage of cortical neurons may be 'Neocortical Dark Matter' (Thorpe, 2011).

An interesting future avenue of investigation would therefore be to examine the expected decreasing sensitivity to non-specific, noisy inputs throughout the course of synaptic learning in support of this conjecture. The benchmark simulations for comparison in this work have always been the responses of the network to the testing stimuli, including information analysis measures, (or the structure in the synaptic weights) *before training*. A useful alternative test case would be to compare the responses of the network (potentially before and after training) to noisy inputs which are not transforms of the test stimuli. It may also be necessary to augment the network with additional layers to allow for the sufficiently sparse firing of the hypothesized Grandmother cells in the output layer, when more realistic stimuli (or even a silicon retina) are used. In this way, some of the issues arising from the use of small numbers of stimuli may be ameliorated (by examining responses to 'stimuli' outside of the limited training set) and evidence may be provided to either confirm or refute this intriguing possibility of 'Neocortical Dark Matter' (Thorpe, 2011).

8.5.4 Linking the model to empirical data

Computational neuroscience provides a way to test ideas about how information processing algorithms may be implemented by neural tissue and systematically explore candidate explanations for neurophysiological and even psychological processes. While efforts are made to capture the essential details of the biological system without oversimplifying the issue, the models must still be compared to and constrained by empirical data for validation. In turn, modelling work should feed back into the empirical domain by framing hypotheses concerning the biological phenomenon of

interest and guiding experimentalists to collect data with which to directly test such hypotheses. It is therefore important to consider ways in which the models here may be linked to neurophysiological and psychological data in future work.

One area which has received some interest from experimentalists over recent years is that of Trace learning (Földiák, 1991), or equivalently 'Unsupervised Temporal Slowness Learning' (UTL) (Li and DiCarlo, 2008) and 'Slow Feature Analysis' (SFA) (Wiskott and Sejnowski, 2002). Trace learning is a heuristic hypothesized to be employed by the visual system, in which successive views of an object are associated together into a coherent representation through their temporal continuity (Földiák, 1991). Since object identity typically varies more slowly than contextual variables (Sprekeler et al., 2007), cells in the primate ventral visual system may use this unsupervised principle together with the spatio-temporal statistics of natural images (Simoncelli and Olshausen, 2001) to become selective for a particular object and insensitive to natural variations in the object's appearance (thereby achieving transformation-invariant visual representations).

In addition to modelling studies (Wallis and Rolls, 1997; Michler et al., 2009; Evans and Stringer, 2012) and a robotics instantiation (Wyss et al., 2006), this principle has been demonstrated in both psychophysical data from humans (Cox et al., 2005) and neurophysiological data from Macaques (Li and DiCarlo, 2008, 2012). These empirical studies from the DiCarlo laboratory carefully manipulated the statistics of the stimuli to reduce temporal continuity and thereby break invariance learning, thus providing further evidence that temporal continuity are used in the brain to learn transformation-invariant visual representations.

Similar results were also obtained when using natural stimuli to train a neural network model of the ventral visual stream. When HMAX was trained with a variant of the Trace rule on video recorded from a camera attached to a cat's head as it freely explored its environment, cell response properties naturally emerged throughout the hierarchy which mimicked those found in the primate ventral stream. These included Gabor-like simple cell sensitivities in the first layer and complex cell tunings (which pooled across similar orientations in different locations) in the next layer (Masquelier et al., 2007). When the frames of the video were randomized, the formation of complex (invariant) but not simple cells was impaired (Masquelier et al., 2007). More recently, the original data from the 'invariance disruption' experiments of Li and DiCarlo (2008) were modelled in HMAX (Isik et al., 2012), replicating the trends found in the neurophysiological data by instantiating Trace learning in a rate-coded model of the ventral stream.

In terms of Trace learning, one way in which the work of this thesis could guide experimental work would be in the examination of the time constants of the synaptic transmitter processes, for example the re-uptake or cleft clearance rate of candidate neurotransmitters, or the rate of closure of the ion channels which they open. If any of the particular species of neurotransmitters or ion channels (or even a number of them working collectively) were found to have the appropriate time-courses for natural spatio-temporal statistics (in the order of 100–150 ms), then these would be prime candidates for the neural implementation of the Trace learning mechanism. Such synaptic processes could then be selectively inhibited, for example through chemical blockage or genetic knock-out, in order to study the predicted impairment of transformation-invariant representation learning (Masquelier et al., 2007). As a more precise test of the theory, if these processes could instead be accelerated or decelerated in a more controlled way, then the optimal statistics of the inputs to match the new system dynamics should be relatively quickened or slowed accordingly.

Similarly, the work of this thesis would benefit from empirical studies concerning the operation of CT learning (Stringer et al., 2006). Whilst CT learning has been studied in both rate-coded (Perry et al., 2006, 2010; Tromans et al., 2012) and spiking neural networks (Evans and Stringer, 2012), to date we know of only one empirical study to offer evidence for the learning mechanism in humans (Perry et al., 2006). In the psychophysical studies of Perry et al. (2006), an enhancement to learning views of rotating 3-D shapes in a 'same/different' (delayed match to sample) task, both when the views of (five) different stimuli were presented sequentially, and crucially, also when they were interleaved between stimuli. This study therefore showed transformation-invariance learning in the absence of close temporal correlations between different views of the same object. A second follow-up experiment demonstrated that this effect was weakened when the rotational increments between views were increased, as expected of CT learning.

It would be instructive to augment these studies with a similar paradigm to that used by DiCarlo and colleagues (Cox et al., 2005; Li and DiCarlo, 2008) but focussing specifically on 'breaking' transformation-invariance learning solely through the CT mechanism. However, this is likely to be a difficult effect to isolate, since spatially similar views also typically occur close together in time and the work of Perry et al. (2006) suggests that Trace learning (temporal association) may be the dominant mechanism employed in the brain.

Several chapters of this thesis looked at the formation of antiphase input representations (or 'perceptual cycles') and their utility in allowing the next layer of

neurons to learn transformation-invariant representations of multiple stimuli presented simultaneously. With multi-electrode recording units continuing to develop, it should be possible to search for direct physiological evidence of the representation oscillations, as Singer and colleagues have done *in vivo* with simple oriented moving bars (Gray et al., 1989; Engel et al., 1991). If one stimulus could then be systematically morphed into looking more similar than the other, the two cycles found in the model should eventually collapse into one cycle with inter-stimulus synchronization.

As computers become more powerful (or the model becomes more efficient to simulate) and larger stimulus sets become more feasible to use, the model may be more directly compared to psychological performance data such as the number of images which may be learnt in a given time or presentation epochs (extent of training). This has been a standard way of evaluating the performance of computationally cheaper rate-coded (and connectionist) models at the cutting edge of machine vision research (Krizhevsky et al., 2012).

8.5.5 Advantages to rate-coded CT and Trace learning

In this section, the relative merits of the spiking implementations of the Trace and CT learning mechanisms over their rate-coded counterparts are discussed. It has been made clear that their rate-coded implementations are computationally much cheaper and so should be used in the absence of a compelling reason to model spiking neurons. However, the main thrust of this thesis has been to investigate forms of information processing only available in spiking neural networks which may impact the formation of transformation-invariant object representations and so the relevant findings are briefly reviewed here.

The model used to simulate the Trace mechanism with spiking neurons (Chapters 3–4) suggests a neurophysiologically measurable property which may support a temporal trace, namely the excitatory synaptic conductance time-constant. By lengthening this model parameter appropriately (typically to $150\,ms$), the network was able to form transformation-invariant representations based upon the temporal continuity of successive transforms of the same object. As such, there is no computational advantage to what has already been achieved in rate-coded models of the ventral visual system (Wallis and Rolls, 1997; Michler et al., 2009; Isik et al., 2012). However, the spiking model does demonstrate how the mechanism may work when spike times are explicitly considered rather than statistical averages, it proposes a candidate property of biological neurons where experimentalists may look for trace-like effects to measure, and an approximate order of magnitude for the parameter (relative to

the other constants of integrate-and-fire neurons). While offering no computational advantage then, the spiking implementation of Trace learning does offer a substantial advantage in terms of biological detail.

In contrast to the spiking Trace learning simulations, CT learning has a clear advantage when implemented with spiking neurons. In Chapters 5–7 it was shown how two sets of neurons representing simultaneously presented stimuli may have the same firing-rate but still facilitate the formation of separate transformation-invariant representations by virtue of having different spike timings. Rate-coded models would be insensitive to these dynamics and would suffer from the superposition catastrophe, whereas CT learning in a spiking neural network is able to overcome these failings in a multiple-stimulus presentation paradigm. As such, the spiking version of CT learning in these simulations may be taken as further evidence of the importance of the precision of spike timings in neural information processing (Bohte, 2004).

Finally, one further principle advantage of the spiking learning mechanisms used here is that they are framed in the more biologically accurate learning paradigm of STDP (for reviews see Bi and Poo, 2001; Dan and Poo, 2004, 2006; Caporale and Dan, 2008; Markram et al., 2011). This is an important advance over the learning regimes based upon statistical averaging given the mounting neurophysiological evidence for the importance of both spike ordering and spike timing in synaptic plasticity.

8.5.6 Methodology

While the primary purpose of this thesis was to understand the self-organizational behaviour of (a model of) the primate ventral visual system, a considerable degree of the challenge in this goal was in developing the model and appropriate methods to investigate the psychological phenomena of interest. It therefore seems appropriate to devote this section of the discussion to appraising the process of model development, note its shortcomings and suggest some future directions for improvement.

The model was written in the C Programming language (Kernighan and Ritchie, 1988) for the greater speed of execution and readily available open-source scientific libraries such as the GSL (Galassi et al., 2009) that this language offers in comparison to many alternatives. The low-level nature of this language, while providing flexibility and speed of execution, also began to slow development as it is verbose and difficult to encapsulate. As new features were added to the model and changes were made to incorporate more simulation paradigms, the code grew more complex and difficult to manage until even simple changes would require editing six or more separate sections of the source code, increasing the likelihood of forgetting or failing to carry the changes

through in one of the source files and thereby introducing a bug. If the model is to be used and further developed in the future, taking the time to rewrite it in C++ would provide almost all the benefits of C (albeit with a modest decrease in execution speed) but with the additional facilities to encapsulate code, make it less verbose (especially when creating and initializing new objects or structures) and therefore making it more readable, along with the advantages of an object-oriented programming language.

At the outset of the thesis, it was not known which features would ultimately need to be included in the model and so the sensible choice seemed to be to start with a basic model and add more features as and when they were required. In choosing a numerical scheme to simulate the systems of differential equations describing the neurons' behaviour, simplicity was therefore prioritized. A simple numerical method such as the Forward Euler scheme (Atkinson and Han, 2004) helped to speed up the processes of implementation and debugging so that new features could be introduced and evaluated quickly. The problem with the Forward Euler scheme is that it requires a comparatively very small time-step in order to achieve the same degree of accuracy as, for example a Runge–Kutta scheme, thus making it computationally expensive (Atkinson and Han, 2004). While this cost was ameliorated by parallelizing and optimizing the code, the upper limit on the size of the networks (in terms of what is computationally feasible given the current hardware) was still a limitation to the thesis, especially regarding the use of more ecologically valid filtered inputs for stimuli. Having arrived at a more stable feature set for the model neurons, the model could be given a significant computational boost by rewriting the core routines to implement a faster numerical scheme such as a 4^{th} Order Runge–Kutta scheme or the Parker–Sochacki method (Stewart and Bair, 2009). The time cost of implementing these more complex schemes would be rapidly recouped from much faster simulation run-times, assuming that the core neuron model would change relatively little.

As features have been incorporated into the neuron model, it has grown in complexity and also the time for execution. Initially, the Leaky Integrate-and-Fire neuron model was chosen for its reasonable degree of biological accuracy and very low computational cost, but as it was developed and added to throughout the course of this research, it has become increasingly appropriate to consider other models which may provide scope for far greater feature complexity with relatively modest increases in computational cost such as the Izhikevich model neuron (Izhikevich, 2003, 2004). It would therefore be a useful comparison to prototype such an alternative model neuron in order to evaluate its relative *required*-accuracy/cost ratio for further simulations.

In the years since starting this thesis, GPU programming has gained considerable traction in the scientific computing community and the GPU hardware and programming tools have improved greatly. It would therefore be advantageous to take time to re-organize the data structures and flow of execution to take advantage of this trend and potentially achieve huge increases in execution speed. This could also be harnessed along with the multi-core architecture of CPUs (depending on the specific details of the code to run) whereby multiple threads are run on each type of processing unit simultaneously.

8.6 Future work

Having reviewed and discussed the main themes of the thesis and the limitations of this work in investigating them, this section sets out some of the main areas considered to be worthy of further investigation.

8.6.1 Multiple layers

A common and plausible view of information processing in the ventral visual stream is that the same type of computations are performed in each layer and that from a common underlying principle and self-organizational process, increasingly complex and invariant responses are constructed as the hierarchy is ascended (Riesenhuber and Poggio, 1999). The approach taken in this work was commensurate with this view, in that studying a component of this structure (that is to say, a single weight layer) should provide insights into the processing occurring at each stage of the ventral stream and a qualitative understanding of the pathway as a whole through a detailed and controlled study of its components. Furthermore, since spiking neural networks are more computationally expensive than their rate-coded counterparts, networks had to be kept to a computationally tractable size. Together, this meant that the majority of the simulations were conducted with one (weight) layer models, where a progressive convergence of connectivity was substituted for full feed-forward connectivity.

While this did provide some useful insights into the operation of the ventral stream as a whole, a multi-layered model would represent greater computational power (Šíma and Orponen, 2003) and so potentially be capable of solving more difficult recognition tasks with closer qualitative and quantitative agreement to neurophysiological recordings of details such as onset latencies and the nature of stimulus preferences in intermediate layers. This would also allow for the implementation of diluted convergent

feed-forward connectivity and projections skipping steps in the hierarchy as found throughout the visual system (Felleman and Van Essen, 1991).

A further benefit of using a multi-layer architecture would be to explore the intermediate level representations which emerge and relate them to those found at intermediate stages of the ventral stream. This process however, would be best studied using more realistic three-dimensional stimuli with more natural movement (types of transformation).

8.6.2 Alternative Stimulus Transformations

Most straightforwardly, the simulations presented with the simple stimuli may be thought of as *translation*-invariance learning, as a contiguous patch of neurons shifted across a section of the input layer during training such that output neurons learn to selectively respond to a particular stimulus across most or all of its locations (transforms). However, this was an idealized abstraction of the problem of transformation-invariance learning in general. Suppose for instance, that the location of the input neurons is not important and that each instead is a particular V1 cell corresponding to a particular feature somewhere in the visual field. If the presented stimulus is rotating, a particular view (transform) will stimulate a particular assortment of such features. As the object rotates a small degree to another view, a new yet overlapping set of visual features will be stimulated which may be associated together through the CT or Trace learning mechanisms. Although the features would be spread across the retinotopic map in V1, it is the same principle as demonstrated in the work presented here.

This was demonstrated in Chapter 4 through the use of realistic filtered three-dimensional shapes as stimuli, firstly through a combination of the Trace and CT learning mechanisms in the case of translating stimuli and then extending the same principles to show that invariant representations could be built up from rotating stimuli.

In the primate visual system, the feed-forward connectivity is more dilute than the idealized architecture used here (where the probability of synapsing with each afferent neuron in the preceding layer is set to one), so it follows that a neuron in V2 may not receive inputs from all the relevant features in the way that the model does. However, given the convergence of feed-forward connections in the primate visual system which lead to increasingly large receptive fields at higher layers (encompassing the entire visual field in some IT cells), it follows that the necessary features will likely come to be bound together onto the same postsynaptic neuron in a high enough visual area, with partial invariances (across subsets of transforms) achieved in intermediate layers.

A useful next step would therefore be to test an architecture with more diluted connectivity using realistic stimuli transforming in other ways, including changing size, lighting conditions or being partially occluded.

8.6.3 Feedback projections

It has been convincingly argued by considering the speed of the recognition process and the number of synapses between the retina and aIT, that feed-forward projections are sufficient for the basic task of object recognition (Thorpe et al., 1996; Hung et al., 2005). However, feed-back projections are found throughout the cortex in general and the visual system in particular (Felleman and Van Essen, 1991) and very recent evidence from a TMS based study of the ventral stream suggests that feedback may have a role (Koivisto et al., 2011). Authors have proposed likely roles for these feed-back projections including an attentional mechanism, incorporating prior knowledge and expectations to bias perceptions and refining 'perceptual hypotheses' to identify cluttered or ambiguous figures (DiCarlo and Cox, 2007). In recognition tasks like the aforementioned, which require more than just a glimpse, visual processing would be expected to operate on a longer time scale and so as the recognition task becomes harder, feed-back projections may be found to play an increasingly important role. Such feed-back projections would therefore be an interesting addition to the model to explore their role in more difficult recognition tasks.

8.6.4 Inhibition

Much of the work presented was originally produced with hyperpolarizing (subtractive) inhibitory interneurons. It was discovered however, that this model of inhibition made the network highly sensitive to the strength of inhibitory synapses (Δg_{IE}), which ultimately required a lot of tuning of this one parameter. The shunting (divisive) inhibition used throughout the presented simulations was found to give more robust results to variations along this parameter dimension and so cut down the time required for tuning the network dynamics. There is some evidence suggesting that a mixture of the two types of inhibitory interneuron coexist in the same cortical areas and confer additional information processing capabilities. It would therefore be interesting to explore this idea with a heterogeneous model.

8.6.5 The Neural Code

While the firing rate based information analysis has been a useful tool in understanding the ability of the model to form transformation-invariant representations, it is not without its limitations as discussed in Section 8.5. Notably, for instance, there is increasingly a wealth of evidence detailing the way in which the information conveyed by neurons or neural circuits changes over time (Sugase-Miyamoto et al., 2011; Pasupathy and Brincat, 2012).

It would therefore be an important advancement to develop an analysis which could calculate the information time-course to compliment the rate-based average information measures used in this thesis and previous work (Sugase et al., 1999). Not only would this provide a more precise instrument of analysis, but also allow the simulation results to be more directly and meaningfully compared to the biological data.

8.6.6 Spike-Time Dependent Plasticity

It may still be premature to reduce the phenomenological model of STDP, based only upon the relative timings of pre- and postsynaptic spikes, used here and commonly elsewhere to a more mechanistic implementation but there are other candidates to consider, modelling the interactions of triplets of spikes for example (Froemke and Dan, 2002; Pfister and Gerstner, 2006) or extending the learning paradigm to a model of STDP incorporating a global reward signal (Seung, 2003).

Simulation of natural spike trains should be considered (not just pair based interactions) especially when applying STDP to real world problems such as object recognition. This was tested in pyramidal cells from layer 2/3 of rat visual cortex by Froemke and Dan (2002) who concluded that synaptic modification depends not only on the interval between a pair of spikes but also on the timing of preceding spikes. This led to the proposal of an efficacy-based form of STDP whereby the efficacy of each spike in synaptic modification was suppressed by the preceding spike in the same neuron, meaning that the first spike in each burst is dominant. If biological evidence could be found for this, it would support Thorpe's rank order coding scheme (VanRullen and Thorpe, 2001, 2002).

One typical aspect of modelling work (including the present studies), which may be considered a shortcoming, is an artificial mode where synaptic weights are 'frozen' to prevent further modification in order to probe the responses of the network before and after a learning period. This however, is contrary to the evidence from studies

which through manipulation of stimuli, demonstrate that learning is a continuous process (Cox et al., 2005; Li and DiCarlo, 2008; McMahon and Leopold, 2012). While stopping STDP can be regarded as a useful way to understand the responses of the system without effecting any further change, it can also be seen as a work-around to avoid the problem of striking a balance between sensitivity and stability. Given the rapidity of learning to recognize an object in the primate visual system and the persistence of that ability (that is, the ability to recognize that and many other stimuli over a period of years and recall them without overwriting earlier memories), there must be a biologically plausible way of unifying these phases.

In the simple model of STDP commonly used in simulations, there is for example, no frequency of pairing dependency and so each time a pair of pre- and postsynaptic spikes occur close enough together in time, synaptic modification will occur and hence previously stored information will be vulnerable to erasure. A more detailed understanding of the precise biophysical mechanisms and critical factors involved in STDP may also enable modellers to eliminate this artificial schism between 'training' and 'testing' phases common to many simulation protocols by introducing additional thresholds such as dendritic depolarization, frequency of stimulation or additional neuromodulatory signals influenced by systems-level evaluation of the importance of the information (Lisman and Spruston, 2005).

Synaptic weight normalization is an important feature of rate-coded learning rules since typically the updates are unbounded. Multiplicative forms of STDP lead to an automatic stabilization of the distribution of the whole population of synaptic weights (van Rossum et al., 2000), potentially making weight normalization redundant insofar as the neurons will converge to stable firing rates. Two forms were implemented, (maintaining either the sum of presynaptic weights for each neuron or the sum of squares) but were found to make little difference. Under different conditions however with some of the different forms of STDP discussed, weight normalization might prove to be an important feature.

8.7 Conclusion

The aim of this thesis was to understand how a spiking neural network model of the primate ventral visual stream may learn to form transformation-invariant representations of stimuli. In particular, the work addressed the question of how this might be achieved with a biologically realistic form of spike-time dependent plasticity and how our existing understanding of the unsupervised rate-coded Hebbian learning

rules might map onto STDP. The investigations also set out to uncover and analyse additional mechanisms for solving the problem of object recognition which may not be available to simpler rate-coded models.

In addressing these research questions, the simulations presented have offered several insights into the function of the ventral visual stream. A biologically plausible form of the Trace learning mechanism was proposed, (implemented through lengthening the decay of the excitatory conductances), and found to be successful in facilitating the formation of transformation-invariant representations with both abstract and realistic stimuli. The CT learning mechanism was also shown to hold in a spiking neural network, relying upon the same core principle of environment statistics, namely sufficient similarity between transforms of a stimulus.

It has also been demonstrated that by using a spiking neural network to model the precise times of action potentials, that an additional mechanism for separating out individual objects may be employed. This relies upon temporal segmentation of the input representations through antiphase oscillations which may be achieved through a gradient in lateral connectivity strength, axonal conductance delays, or prior learning of category features in lateral connections. Crucially, this mechanism also depends upon temporally specific spike-time dependent learning, which should also be a feature of future models.

The neural network models explored in this thesis are not complete models of the visual system, and do not attempt to model the full range of challenges their biological counterpart faces in recognizing objects. However, they have managed to shed some light on the learning and recognition processes that may be occurring at each stage of the visual system. There is some evidence to suggest that these principles will hold for more complex models, with the same mechanism operating throughout the hierarchy (Riesenhuber and Poggio, 1999). Future work has been proposed to build upon the findings of this thesis and further enhance our understanding of biological object recognition.

Bibliography

Abbott, L. F. and Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–1183.

Afraz, S.-R., Kiani, R., and Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature*, 442(7103):692–695.

Ahmed, B., Anderson, J. C., Douglas, R. J., Martin, K. A., and Whitteridge, D. (1998). Estimates of the net excitatory currents evoked by visual stimulation of identified neurons in cat visual cortex. *Cerebral Cortex*, 8(5):462–476.

Amit, D. J. and Brunel, N. (1997). Dynamics of a recurrent network of spiking neurons before and following learning. *Network: Computation in Neural Systems*, 8(4):373–404.

Aston-Jones, G., Foote, S., and Segal, M. (1985). Impulse conduction properties of noradrenergic locus coeruleus axons projecting to monkey cerebrocortex. *Neuroscience*, 15(3):765–777.

Atkinson, K. and Han, W. (2004). *Elementary Numerical Analysis*. John Wiley & Sons, Inc., 3rd (international) edition.

Bell, A. J. and Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338.

Bernacchia, A. and Amit, D. J. (2007). Impact of spatiotemporally correlated images on the structure of memory. *Proceedings of the National Academy of Sciences*, 104(9):3544–3549.

Bi, G.-q. and Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24):10464–10472.

Bi, G.-q. and Poo, M.-m. (1999). Distributed synaptic modification in neural networks induced by patterned stimulation. *Nature*, 401(6755):792–796.

Bi, G.-q. and Poo, M.-m. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience*, 24(1):139–166.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., and Warland, D. (1991). Reading a neural code. *Science*, 252(5014):1854–1857.

Bichler, O., Querlioz, D., Thorpe, S. J., Bourgoin, J.-P., and Gamrat, C. (2012). Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity. *Neural Networks*, 32(0):339–348.

Bishop, C. M. (1997). *Neural Networks for Pattern Recognition*. Oxford University Press, 5th edition.

Blumenfeld, B., Preminger, S., Sagi, D., and Tsodyks, M. (2006). Dynamics of memory representations in networks with novelty-facilitated synaptic plasticity. *Neuron*, 52(2):383–394.

Bohte, S. (2004). The evidence for neural information processing with precise spike-times: A survey. *Natural Computing*, 3(2):195–206.

Booth, M. C. A. and Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex*, 8:510–523.

Brette, R., Rudolph, M., Carnevale, T., Hines, M., Beeman, D., Bower, J., Diesmann, M., Morrison, A., Goodman, P., Harris, F., Zirpe, M., Natschläger, T., Pecevski, D., Ermentrout, B., Djurfeldt, M., Lansner, A., Rochel, O., Vieville, T., Muller, E., Davison, A., El Boustani, S., and Destexhe, A. (2007). Simulation of networks of spiking neurons: A review of tools and strategies. *Journal of Computational Neuroscience*, 23(3):349–398.

Brincat, S. L. and Connor, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron*, 49(1):17–24.

Brindley, G. S. and Lewin, W. (1968). The sensations produced by electrical stimulation of the visual cortex. *Journal of Physiology*, 196:479–493.

Brunel, N. and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation*, 11(7):1621–1671.

Burkitt, A. N. (2006a). A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biological Cybernetics*, 95(1):1–19.

Burkitt, A. N. (2006b). A review of the integrate-and-fire neuron model: II. Inhomogeneous synaptic input and network properties. *Biological Cybernetics*, 95(2):97–112.

Caporale, N. and Dan, Y. (2008). Spike Timing–Dependent Plasticity: A Hebbian learning rule. *Annual Review of Neuroscience*, 31(1):25–46.

Carlson, N. R. (2000). *Physiology of Behaviour*. Pearson, 7th edition.

Carr, C. and Konishi, M. (1990). A circuit for detection of interaural time differences in the brain stem of the barn owl. *The Journal of Neuroscience*, 10(10):3227–3246.

Carr, C. E. and Konishi, M. (1988). Axonal delay lines for time measurement in the owl's brainstem. *Proceedings of the National Academy of Sciences*, 85(21):8311–8315.

Caywood, M. S., Willmore, B., and Tolhurst, D. J. (2004). Independent components of color natural scenes resemble V1 neurons in their spatial and color tuning. *Journal of Neurophysiology*, 91(6):2859–2873.

Chandra, R., Dagum, L., Kohr, D., Maydan, D., McDonald, J., and Menon, R. (2001). *Parallel programming in OpenMP*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Choe, Y. and Miikkulainen, R. (1998). Self-organization and segmentation in a laterally connected orientation map of spiking neurons. *Neurocomputing*, 21(1-3):139–157.

Choe, Y. and Miikkulainen, R. (2003). The role of postsynaptic potential decay rate in neural synchrony. *Neurocomputing*, 52-54:707–712.

Clavagnier, S., Falchier, A., and Kennedy, H. (2004). Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive, Affective, & Behavioral Neuroscience*, 4(2):117–126.

Connor, C. E., Brincat, S. L., and Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, 17(2):140–147.

Cowey, A. and Stoerig, P. (1991). The neurobiology of blindsight. *Trends in Neurosciences*, 14(4):140–145.

Cowey, A. and Walsh, V. (2000). Magnetically induced phosphenes in sighted, blind and blindsighted observers. *NeuroReport*, 11(14):3269–3273.

Cox, D. D., Meier, P., Oertelt, N., and DiCarlo, J. J. (2005). 'Breaking' position-invariant object recognition. *Nature Neuroscience*, 8(9):1145–1147.

Crouzet, S. M., Kirchner, H., and Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, 10(4):1–17.

Crouzet, S. M. and Thorpe, S. J. (2011). Low level cues and ultra-fast face detection. *Frontiers in Psychology*, 2:1–9.

Dan, Y. and Poo, M.-m. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron*, 44:23–30.

Dan, Y. and Poo, M.-m. (2006). Spike Timing-Dependent Plasticity: From Synapse to Perception. *Physiological Reviews*, 86(3):1033–1048.

David, S. V., Vinje, W. E., and Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of V1 neurons. *Journal of Neuroscience*, 24(31):6991–7006.

Dayan, P. and Abbott, L. F. (2005). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press.

Delorme, A., Gautrais, J., VanRullen, R., and Thorpe, S. (1999). SpikeNET: A simulator for modeling large networks of integrate and fire neurons. *Neurocomputing*, 26-27:989–996.

Delorme, A., Perrinet, L., and Thorpe, S. J. (2001). Networks of integrate-and-fire neurons using Rank Order Coding B: Spike timing dependent plasticity and emergence of orientation selectivity. *Neurocomputing*, 38-40:539–545.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience*, 3:1–8.

DiCarlo, J. J. and Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11(8):333–341.

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3):415–434.

Diesmann, M., Gewaltig, M.-O., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature*, 402(6761):529–533.

Elliffe, M. C. M., Rolls, E. T., and Stringer, S. M. (2002). Invariant recognition of feature combinations in the visual system. *Biological Cybernetics*, 86(1):59–71.

Engel, A. K., König, P., and Singer, W. (1991). Direct physiological evidence for scene segmentation by temporal coding. *Proceedings of the National Academy of Sciences*, 88(20):9136–9140.

Evans, B. D. and Stringer, S. M. (2012). Transformation-invariant visual representations in self-organizing spiking neural networks. *Frontiers in Computational Neuroscience*, 6:1–19.

Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1–47.

Felsen, G. and Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience*, 8(12):1643–1646.

Felsen, G., Touryan, J., Han, F., and Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS Biology*, 3(10):e342.

Ferster, D. and Spruston, N. (1995). Cracking the neuronal code. *Science*, 270(5237):756–757.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3:194–200.

Földiák, P. (1993). The 'ideal homunculus': Statistical inference from neuronal population responses. In Eeckman, F. H. and Bower, J. M., editors, *Computation and Neural Systems*, chapter 9, pages 55–60. Kluwer Academic Publishers, Norwell, MA.

Freeman, J. and Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, 14(9):1195–1201.

Freiwald, W. A. and Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, 330(6005):845–851.

Fries, P., Schröder, J.-H., Roelfsema, P. R., Singer, W., and Engel, A. K. (2002). Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *Journal of Neuroscience*, 22(9):3739–3754.

Froemke, R. C. and Dan, Y. (2002). Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature*, 416(6879):433–438.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202.

Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Networks*, 1(2):119–130.

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M., and Rossi, F. (2009). *GNU Scientific Library Reference Manual*. Network Theory Limited, 3rd edition.

Gallant, J., Braun, J., and Van Essen, D. (1993). Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. *Science*, 259(5091):100–103.

Gerstner, W., Kempter, R., van Hemmen, J. L., and Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383(6595):76–78.

Gerstner, W. and Kistler, W. (2006). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press, 3rd edition.

Gerstner, W., Ritz, R., and van Hemmen, J. (1993). Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns. *Biological Cybernetics*, 69(5):503–515.

Girard, P., Hupé, J. M., and Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *Journal of Neurophysiology*, 85(3):1328–1331.

Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1):20–25.

Goodman, D. F. M. and Brette, R. (2008). Brian: a simulator for spiking neural networks in Python. *Frontiers in Neuroinformatics*, 2(5):1–10.

Gray, C. M., König, P., Engel, A. K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338(6213):334–337.

Gregory, R. L. (1966). *Eye and Brain: The Psychology of Seeing*. Weidenfeld and Nicolson, London.

Gross, C. G., Rodman, H. R., Gochin, P. M., and Colombo, M. W. (1993). Inferior temporal cortex as a pattern recognition device. In Baum, E., editor, *Computational Learning and Cognition*, chapter 3, pages 44–73. Society for Industrial and Applied Mathematics, Philadelphia.

Gütig, R., Aharonov, R., Rotter, S., and Sompolinsky, H. (2003). Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *The Journal of Neuroscience*, 23(9):3697–3714.

Hebb, D. O. (1949). *The Organization of Behaviour: A Neuropsychological Theory*. Wiley, New York.

Hecht, S., Shlaer, S., and Pirenne, M. H. (1942). Energy, quanta and vision. *The Journal of General Physiology*, 25:819–840.

Hegdé, J. and Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area V2. *The Journal of Neuroscience*, 20(5):RC61.

Hegdé, J. and Van Essen, D. C. (2007). A comparative study of shape representation in macaque visual areas V2 and V4. *Cerebral Cortex*, 17(5):1100–1116.

Hennequin, G., Gerstner, W., and Pfister, J.-P. (2010). STDP in adaptive neurons gives close-to-optimal information transmission. *Frontiers in Computational Neuroscience*, 4:1–16.

Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546.

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544.

Hosoya, T., Baccus, S. A., and Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, 436(7047):71–77.

Hubel, D. (1995). *Eye, Brain and Vision*. Scientific American Library.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591.

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

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1):215–243.

Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749):863–866.

Hyvarinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–430.

Isik, L., Leibo, J. Z., and Poggio, T. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. *Frontiers in Computational Neuroscience*, 6(37):1–7.

Ito, M., Tamura, H., Fujita, I., and Tanaka, K. (1995). Size and position invariance of neuronal response in monkey inferotemporal cortex. *Journal of Neurophysiology*, 73:218–226.

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6):1569–1572.

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5):1063–1070.

Izhikevich, E. M. (2006). Polychronization: Computation with spikes. *Neural Computation*, 18(2):245–282.

Izhikevich, E. M. and Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences*, 105(9):3593–3598.

Izhikevich, E. M., Gally, J. A., and Edelman, G. M. (2004). Spike-timing dynamics of neuronal groups. *Cerebral Cortex*, 14(8):933–944.

Jin, X., Lujan, M., Plana, L., Davies, S., Temple, S., and Furber, S. (2010). Modeling spiking neural networks on SpiNNaker. *Computing in Science & Engineering*, 12(5):91–97.

Johnson, S. (2001). *Emergence: The connected Lives of Ants, Brains, Cities and Software*. Penguin Books.

Jones, J. P. and Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M., editors (2000). *Principles of Neural Science*. McGraw-Hill, New York, 4th edition.

Karklin, Y. and Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*, 457(7225):83–86.

Keck, I. R., Nassabay, S., Puntonet, C. G., and Lang, E. W. (2005). A new approach to clustering and object detection with independent component analysis. In Mira, J. and Álvarez, J. R., editors, *Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach*, Lecture Notes in Computer Science, pages 558–566. Springer Berlin Heidelberg.

Kernighan, B. W. and Ritchie, D. (1988). *The C Programming Language*. Prentice Hall, Inc., 2nd edition.

Kirchner, H. and Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46(11):1762–1776.

Kobatake, E. and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, 71(3):856–867.

Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons*. Computational Neuroscience. Oxford University Press, New York.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69.

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., and Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *The Journal of Neuroscience*, 31(7):2488–2492.

Kreiter, A. K. and Singer, W. (1996). Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. *Journal of Neuroscience*, 16(7):2381–2396.

Kreuz, T., Haas, J. S., Morelli, A., Abarbanel, H. D. I., and Politi, A. (2007). Measuring spike train synchrony. *Journal of Neuroscience Methods*, 165(1):151–161.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Bartlett, P., Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing 25 (NIPS 2012)*, volume 25, pages 1106–1114.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1):37–68.

Kuwabara, N. and Suga, N. (1993). Delay lines and amplitude selectivity are created in subthalamic auditory nuclei: the brachium of the inferior colliculus of the mustached bat. *Journal of Neurophysiology*, 69(5):1713–1724.

Lehky, S. R. and Sereno, A. B. (2007). Comparison of shape encoding in primate dorsal and ventral visual pathways. *Journal of Neurophysiology*, 97(1):307–319.

Li, N. and DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, 321(5895):1502–1507.

Li, N. and DiCarlo, J. J. (2010). Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. *Neuron*, 67(6):1062–1075.

Li, N. and DiCarlo, J. J. (2012). Neuronal learning of invariant object representation in the ventral visual stream is not dependent on reward. *The Journal of Neuroscience*, 32(19):6611–6620.

Lisman, J. and Spruston, N. (2005). Postsynaptic depolarization requirements for LTP and LTD: a critique of spike timing-dependent plasticity. *Nature Neuroscience*, 8(7):839–841.

Liu, H., Agam, Y., Madsen, J. R., and Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, 62(2):281–290.

Liu, Y.-H. and Wang, X.-J. (2001). Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron. *Journal of Computational Neuroscience*, 10(1):25–45.

Logothetis, N. K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5(5):552–563.

Maass, W. and Bishop, C. M., editors (1999). *Pulsed Neural Networks*. MIT Press, Cambridge, MA, USA.

MacKay, D. J. C. (2003). *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, Cambridge.

Malach, R., Tootell, R. B. H., and Malonek, D. (1994). Relationship between orientation domains, cytochrome oxidase stripes, and intrinsic horizontal connections in squirrel monkey area V2. *Cerebral Cortex*, 4(2):151–165.

Markram, H., Gerstner, W., and Sjöström, P. J. (2011). A history of spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience*, 3(4):1–24.

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.

Masquelier, T., Guyonneau, R., and Thorpe, S. J. (2008). Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PLoS ONE*, 3(1):e1377.

Masquelier, T., Hugues, E., Deco, G., and Thorpe, S. J. (2009). Oscillations, phase-of-firing coding, and spike timing-dependent plasticity: An efficient learning scheme. *Journal of Neuroscience*, 29(43):13484–13493.

Masquelier, T., Serre, T., and Poggio, T. (2007). Learning complex cell invariance from natural videos: A plausibility proof. Technical Report 269, CBCL, MIT.

Masquelier, T. and Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology*, 3(2):e31.

McCormick, D. A., Connors, B. W., Lighthall, J. W., and Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology*, 54(4):782–806.

McCulloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 5(4):115–133.

McMahon, D. B. and Leopold, D. A. (2012). Stimulus timing-dependent plasticity in high-level vision. *Current Biology*, 22(4):332–337.

Mel, B. (1997). SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation*, 9:777–804.

Mel, B. W. (2000). In the brain, the model is the goal. *Nature Neuroscience*, 3:1183.

Michler, F., Eckhorn, R., and Wachtler, T. (2009). Using spatiotemporal correlations to learn topographic maps for invariant object recognition. *Journal of Neurophysiology*, 102(2):953–964.

Miconi, T. (2012). Personal Communication.

Miconi, T. and VanRullen, R. (2010). The gamma slideshow: object-based perceptual cycles in a model of the visual cortex. *Frontiers in Human Neuroscience*, 4:1–11.

Miikkulainen, R., Bedner, J. A., Choe, Y., and Sirosh, J. (2005). *Computational Maps in the Visual Cortex*. Springer.

Movshon, J. A. and Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *The Journal of Neuroscience*, 16(23):7733–7741.

Nageswaran, J., Dutt, N., Krichmar, J., Nicolau, A., and Veidenbaum, A. (2009a). Efficient simulation of large-scale spiking neural networks using CUDA graphics processors. In *International Joint Conference on Neural Networks, 2009. IJCNN 2009.*, pages 2145–2152, Atlanta, Georgia, USA.

Nageswaran, J. M., Dutt, N., Krichmar, J. L., Nicolau, A., and Veidenbaum, A. V. (2009b). A configurable simulation environment for the efficient simulation of large-scale spiking neural networks on graphics processors. *Neural Networks*, 22(5-6):791–800.

Nischwitz, A. and Glünder, H. (1995). Local lateral inhibition: a key to spike synchronization? *Biological Cybernetics*, 73(5):389–400.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609.

Op de Beeck, H. and Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology*, 426:505–518.

Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica Supplementum*, 13:6:1–102.

Palmeri, T. J. and Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, 5(4):291–303.

Panzeri, S., Senatore, R., Montemurro, M. A., and Petersen, R. S. (2007). Correcting for the sampling bias problem in spike train information measures. *Journal of Neurophysiology*, 98(3):1064–1072.

Panzeri, S. and Treves, A. (1996). Analytical estimates of limited sampling biases in different information measures. *Network: Computation in Neural Systems*, 7:87–107.

Pasupathy, A. (2006). Neural basis of shape representation in the primate brain. In Martinez-Conde, S., Macknik, S., Martinez, L., Alonso, J.-M., and Tse, P., editors, *Visual Perception Fundamentals of Vision: Low and Mid-Level Processes in Perception*, volume 154, Part A, pages 293–313. Elsevier.

Pasupathy, A. and Brincat, S. L. (2012). Population coding of object contour shape in V4 and posterior inferotemporal cortex. In Kriegeskorte, N. and Kreiman, G., editors, *Understanding Visual Population Codes — Towards a Common Multivariate Framework for Cell Recording and Functional Imaging*. MIT Press, Cambridge MA.

Perrinet, L. (2003). *Comment déchiffrer le code impulsionnel de la vision? Étude du flux parallèle, asynchrone et épars dans le traitement visuel ultra-rapide*. PhD thesis, Université Paul Sabatier, Toulouse, France.

Perrinet, L., Delorme, A., Samuelides, M., and Thorpe, S. J. (2001). Networks of integrate-and-fire neuron using rank order coding A: How to implement spike time dependent Hebbian plasticity. *Neurocomputing*, 38-40:817–822.

Perry, G., Rolls, E., and Stringer, S. (2010). Continuous transformation learning of translation invariant representations. *Experimental Brain Research*, 204(2):255–270.

Perry, G., Rolls, E. T., and Stringer, S. M. (2006). Spatial vs. temporal continuity in view invariant visual object recognition learning. *Vision Research*, 46(23):3994–4006.

Petkov, N. and Kruizinga, 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.

Pfister, J.-P. and Gerstner, W. (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *The Journal of Neuroscience*, 26(38):9673–9682.

Pinto, N., Cox, D. D., and DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS Computational Biology*, 4(1):e27.

Pinto, N., Doukhan, D., DiCarlo, J. J., and Cox, D. D. (2009). A high-throughput screening approach to discovering good forms of biologically inspired visual representation. *PLoS Computational Biolology*, 5(11):e1000579.

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., and Duchaine, B. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Current Biology*, 19(4):319–324.

Posner, M. I. and Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77(3):353–363.

Preminger, S., Blumenfeld, B., Sagi, D., and Tsodyks, M. (2009). Mapping dynamic memories of gradually changing objects. *Proceedings of the National Academy of Sciences*, 106(13):5371–5376.

Quian Quiroga, R. and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience*, 10(3):173–185.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107.

Reitboeck, H., Stoecker, M., and Hahn, C. (1993). Object separation in dynamic neural networks. In *IEEE International Conference on Neural Networks, 1993.*, volume 2, pages 638–641, San Francisco, CA, USA.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1999). *Spikes: Exploring the Neural Code*. Computational Neuroscience. MIT Press, Cambridge, Massachusetts.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2:1019–1025.

Riesenhuber, M. and Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, 3:1199–1204.

Riesenhuber, M. and Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, 12(2):162–168.

Rodieck, R. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5(12):583–601.

Rolls, E. T. and Deco, G. (2002). *Computational Neuroscience of Vision*. Oxford University Press, Oxford.

Rolls, E. T. and Deco, G. (2010). *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*. Oxford University Press.

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.

Rolls, E. T. and Treves, A. (1998). *Neural Networks and Brain Function*. Oxford University Press, Oxford.

Rolls, E. T., Treves, A., and Tovee, M. J. (1997a). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Experimental Brain Research*, 114(1):149–162.

Rolls, E. T., Treves, A., Tovee, M. J., and Panzeri, S. (1997b). Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *Journal of Computational Neuroscience*, 4(4):309–333.

Rousselet, G. A., Thorpe, S. J., and Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, 8(8):363–370.

Ruderman, D. L. and Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73(6):814–817.

Salami, M., Itami, C., Tsumoto, T., and Kimura, F. (2003). Change of conduction velocity by regional myelination yields constant latency irrespective of distance between thalamus and cortex. *Proceedings of the National Academy of Sciences*, 100(10):6174–6179.

Sereno, A. B. and Maunsell, J. H. R. (1998). Shape selectivity in primate lateral intraparietal cortex. *Nature*, 395(6701):500–503.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3):411–426.

Seung, H. (2003). Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron*, 40(6):1063–1073.

Shimazaki, H. and Shinomoto, S. (2007). A method for selecting the bin size of a time histogram. *Neural Computation*, 19(6):1503–1527.

Shimazaki, H. and Shinomoto, S. (2010). Kernel bandwidth optimization in spike rate estimation. *Journal of Computational Neuroscience*, 29(1):171–182.

Šíma, J. and Orponen, P. (2003). General-purpose computation with neural networks: A survey of complexity theoretic results. *Neural Computation*, 15(12):2727–2778.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*, 13(2):144–149.

Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1):1193–1216.

Smith, A., Singh, K., Williams, A., and Greenlee, M. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex*, 11(12):1182–1190.

Song, S., Miller, K. D., and Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3(9):919–926.

Spratling, M. W. (2005). Learning viewpoint invariant perceptual representations from cluttered images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):753–761.

Sprekeler, H., Michaelis, C., and Wiskott, L. (2007). Slowness: An objective for spike-timing–dependent plasticity? *PLoS Computational Biology*, 3(6):e112.

Stewart, R. and Bair, W. (2009). Spiking neural network simulation: numerical integration with the Parker–Sochacki method. *Journal of Computational Neuroscience*, 27(1):115–133.

Stoecker, M., Reitboeck, H. J., and Eckhorn, R. (1996). A neural network for scene segmentation by temporal coding. *Neurocomputing*, 11(2-4):123–134.

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.

Stringer, S. M. and Rolls, E. T. (2000). Position invariant recognition in the visual system with cluttered environments. *Neural Networks*, 13:305–315.

Stringer, S. M. and Rolls, E. T. (2008). Learning transform invariant object recognition in the visual system with multiple stimuli present during training. *Neural Networks*, 21(7):888–903.

Stringer, S. M., Rolls, E. T., and Tromans, J. M. (2007). Invariant object recognition with trace learning and multiple stimuli present during training. *Network: Computation in Neural Systems*, 18(2):161–187.

Sugase, Y., Yamane, S., Ueno, S., and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400(6747):869–873.

Sugase-Miyamoto, Y., Matsumoto, N., and Kawano, K. (2011). Role of temporal processing stages by inferior temporal neurons in facial recognition. *Frontiers in Psychology*, 2:1–8.

Swadlow, H. A. (1985). Physiological properties of individual cerebral axons studied *in vivo* for as long as one year. *Journal of Neurophysiology*, 54(5):1346–1362.

Swadlow, H. A. (1992). Monitoring the excitability of neocortical efferent neurons to direct activation by extracellular current pulses. *Journal of Neurophysiology*, 68(2):605–619.

Tanabe, M., Gähwiler, B. H., and Gerber, U. (1998). L-Type Ca^{2+} channels mediate the slow Ca^{2+}-dependent afterhyperpolarization current in rat CA3 pyramidal cells in vitro. *Journal of Neurophysiology*, 80(5):2268–2273.

Tanaka, K. (1996). Representation of visual features of objects in the inferotemporal cortex. *Neural Networks*, 9(8):1459–1475.

Tanaka, K. (2003). Columns for complex visual object features in the inferotemporal cortex: Clustering of cells with similar but slightly different stimulus selectivities. *Cerebral Cortex*, 13(1):90–99.

Tanaka, K., Saito, H., Fukada, Y., and Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, 66:170–189.

Tarr, M. J. and Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, 67(1-2):1–20.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582):520–522.

Thorpe, S. J. (2011). Grandmother cells, neocortical dark matter and very long term visual memories. *Journal of Vision*, 11(11):1243.

Thorpe, S. J., Delorme, A., VanRullen, R., and Paquier, W. (2000). Reverse engineering of the visual system using networks of spiking neurons. In *The 2000 IEEE International Symposium on Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva.*, volume 4, pages 405–408.

Tovée, M. J., Rolls, E. T., and Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. *Journal of Neurophysiology*, 72(3):1049–1060.

Tovée, M. J., Rolls, E. T., Treves, A., and Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology*, 70(2):640–654.

Treves, A. and Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples. *Neural Computation*, 7(2):399–407.

Tromans, J. M., Page, H., and Stringer, S. M. (2012). Learning separate visual representations of independently rotating objects. *Network: Computation in Neural Systems*, 23(1–2):1–23.

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.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., and Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, 6(9):989–995.

Tsodyks, M., Kenet, T., Grinvald, A., and Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*, 286(5446):1943–1946.

Underleider, L. G. and Mishkin, M. (1982). Two cortical visual systems. In Ingle, M. A., Goodale, M. I., and Masfield, R. J. W., editors, *Analysis of Visual Behavior*, pages 549–586. MIT Press, Cambridge, MA.

Ungerleider, L. G. and Haxby, J. V. (1994). 'What' and 'Where' in the human brain. *Current Opinion in Neurobiology*, 4(2):157–165.

Usher, M. and Donnelly, N. (1998). Visual synchrony affects binding and segmentation in perception. *Nature*, 394(6689):179–182.

van Hateren, J. H. and Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society B: Biological Sciences*, 265(1412):2315–2315.

van Hateren, J. H. and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society B: Biological Sciences*, 265(1394):359–366.

van Rossum, M. C. W., Bi, G.-Q., and Turrigiano, G. G. (2000). Stable Hebbian learning from spike timing-dependent plasticity. *Journal of Neuroscience*, 20(23):8812–8821.

VanRullen, R. and Thorpe, S. J. (2001). Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neural Computation*, 13(6):1255–1283.

VanRullen, R. and Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, 42(23):2593–2615.

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.

Webster, M. A., Kaping, D., Mizokami, Y., and Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, 428(6982):557–561.

Wiskott, L. and Sejnowski, T. J. (2002). Slow Feature Analysis: Unsupervised learning of invariances. *Neural Computation*, 14(4):715–770.

Wu, Q., McGinnity, T., Maguire, L., Belatreche, A., and Glackin, B. (2008). Processing visual stimuli using hierarchical spiking neural networks. *Neurocomputing*, 71(10-12):2055–2068.

Wyss, R., König, P., and Verschure, P. F. M. J. (2006). A model of the ventral visual system based on temporal stability and local memory. *PLoS Biology*, 4(5):e120.

Young, J. M., Waleszczyk, W. J., Wang, C., Calford, M. B., Dreher, B., and Obermayer, K. (2007). Cortical reorganization consistent with spike timing – but not correlation-dependent plasticity. *Nature Neuroscience*, 10(7):887–895.