Self Organisation and Hierarchical Concept Representation in Networks of Spiking Neurons

Timothy Rumbell

School of Computing and Mathematics

Plymouth University

A thesis submitted in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy

February 2013

Abstract

Self Organisation and Hierarchical Concept Representation in Networks of Spiking Neurons

Timothy Rumbell

The aim of this work is to introduce modular processing mechanisms for cortical functions implemented in networks of spiking neurons. Neural maps are a feature of cortical processing found to be generic throughout sensory cortical areas, and self-organisation to the fundamental properties of input spike trains has been shown to be an important property of cortical organisation. Additionally, oscillatory behaviour, temporal coding of information, and learning through spike timing dependent plasticity are all frequently observed in the cortex. The traditional self-organising map (SOM) algorithm attempts to capture the computational properties of this cortical self-organisation in a neural network. As such, a cognitive module for a spiking SOM using oscillations, phasic coding and STDP has been implemented. This model is capable of mapping to distributions of input data in a manner consistent with the traditional SOM algorithm, and of categorising generic input data sets. Higher-level cortical processing areas appear to feature a hierarchical category structure that is founded on a feature-based object representation. The spiking SOM model is therefore extended to facilitate input patterns in the form of sets of binary feature-object relations, such as those seen in the field of formal concept analysis. It is demonstrated that this extended model is capable of learning to represent the hierarchical conceptual structure of an input data set using the existing learning scheme. Furthermore, manipulations of network parameters allow the level of hierarchy used for either learning or recall to be adjusted, and the network is capable of learning comparable representations when trained with incomplete input patterns. Together these two modules provide related approaches to the generation of both topographic mapping and hierarchical representation of input spaces that can be potentially combined and used as the basis for advanced spiking neuron models of the learning of complex representations.

Contents

List of Figures

List of abbreviations

aEIF	Adaptive exponential integrate and fire
AMPAR	AMPA receptor
BAM	Bidirectional associative memory
BCM	Bienenstock-Cooper-Munro learning rule
BMU	Best matching unit
BSF	Brader-Senn-Fusi learning rule
EIF	Exponential integrate and fire
EPSP	Excitatory postsynaptic potential
IF	Integrate and fire
IPSP	Inhibitory postsynaptic potential
LIF	Leaky integrate and fire
LFP	Local field potential
LTD	Long term depression
LTP	Long term potentiation
MDS	Metric multidimensional scaling
NMDAR	NMDA receptor
PSP	Postsynaptic potential
SOM	Self organising map
STDP	Spike timing dependent plasticity

Acknowledgements

I would like to thank my supervisors, my mum, dad, brother, sister, brother-in-law and niece, and all of my friends.

Authors declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

Work submitted for this research degree at Plymouth University has not formed part of any other degree either at Plymouth University or at another establishment.

Relevant scientific seminars and conferences were regularly attended at which work was often presented. One paper has been accepted for publication in a refereed journal. One paper is under review for publication in a refereed journal.

Signed: _____

Date: _____

Publications :

Rumbell, T.H., Barnden, J., Denham, S. & Wennekers, T. *Emotions in autonomous agents: comparative analysis of mechanisms and functions.* Autonomous Agents and Multi-Agent Systems **2012 , 25(1), 1-45**

Rumbell, T., Denham, S.L. & Wennekers, T. *A spiking self-organising map combining STDP, oscillations and continuous learning* IEEE Transaction on Neural Networks and Learning Systems **2012, submitted**

Posters and conference presentations :

2012:

Cognition Institute opening, *University of Plymouth, UK* Poster presentation: **A spiking self-organising map combining STDP, oscillations and continuous learning**

2nd UK Neuroinformatics Node Congress, *University of Manchester, UK* Poster presentation: **A spiking self-organising map combining STDP, oscillations and continuous learning**

2010:

Bio-Inspired Cognitive Systems 2010 *Universidad Politécnica de Madrid, Spain* Oral presentation: **Emotion mechanisms in autonomous agents**

Word count for the main body of this thesis: 39,535

Chapter 1

Introduction

Information is processed in the brain at several levels of abstraction. Lower levels assimilate information from the environment through the sensory apparatus. The motor systems also operate at a lower level, facilitating interactions with the environment. Cognitive processes occur at a higher level than sensory-motor processing. These processes may involve representations of the fundamental properties of the external environment and the integration of current sensory input with existing memories.

Various types of process meet the criterion of being in some way more abstract than sensory-motor processes. Prominent among these are perception, memory, attention, language, and intelligence (Fuster 2003). Traditional cognitive neuroscience approaches to investigating these functions have considered each of them as a separate mechanism, maybe even a separate module, with specific neural correlates. The network model of cognition suggests that brain modules are not holders for the entirety of a single psychologically separate cognitive function, but are likely to be smaller in scope, with cognitive functions distributed and intermingled between them.

Generic cortical properties underpin the functional properties of cortical modules. Structures such as cortical columns, microcircuits, the connectivity patterns within and between them, and the dynamics of activity in individual neurons and groups of neurons all contribute to generating functional networks. The information processing in these networks will depend on a combination of these structural properties and mechanisms of learning and adaptation that are present within them.

1.1 Aims

The current research aims to implement generic components of cognitive functions in networks of spiking neurons, taking inspiration from the information processing methods present in the brain. The general idea is that complex cognitive systems can be built from modular components, and that these modules are not complete cognitive units in their own right. Instead, they are intended to be used as building blocks of larger networks capable of multiple complex functions. The primary goal is to demonstrate potential methods that can be used to generate modules of this type, taking inspiration from a combination of biological and existing computational architectures. As these modules are intended as generic components, a focus will be placed on learning and the adaptation of particular network structure to a variety of data and types of input/output relationship.

Of course, many types of module will undoubtedly be required within a network capable of a full suite of cognitive processing. As such, this thesis cannot aim to provide the complete suite of cognitive modules required, and will limit its scope to design of a single type of module. This module will be based on the self-organising map, which, as will be demonstrated, represents a strong candidate for a main type of generic cortical module.

The approach to achieving this goal is to capture functional properties of cortical networks using artificial neurons that in turn rely on the generic properties of cortical neurons and network structures. This approach links the goal of cognitive science, aiming to generate representations of the general functional properties of the brain and mind, with that of computational neuroscience, aiming to artificially recreate observable phenomena in the brain through the use of biologically plausible neuron and network models.

1.2 Contributions

The specific, scientific and technical contributions made by this thesis are:

- A spiking neuron implementation of the self-organising map algorithm that incorporates processing mechanisms inspired by cortical information processing: oscillatory behaviour, temporal coding, and spike timing dependent plasticity. This model improves on

existing similar models by enabling continuous processing of input and allowing ongoing adaptation.

- An analysis of the representational capacities of a spiking self-organising map model that greatly exceeds prior similar analyses.
- A feasibility simulation for the implementation of such a model in analogue spiking neuron hardware.
- An extension to the spiking self-organising map model that incorporates temporal coding of binary input and distinct sources of lateral excitation and inhibition within the map layer, enabling hierarchical representations of concepts in the form of binary object-feature relations.
- A set of parameter manipulations that alter the hierarchical level at which learning and recall of concepts is enabled, as well as demonstrating support for an equivalent learning when training with incomplete input data.

This pair of neural modules incorporate functional properties of biological networks and fulfill useful computational roles, making valuable contributions across several domains. Previous investigation into implementations of this type lack either generalisable topographic mapping properties of the self-organising map or make use of mechanisms that have no biological basis. The combination of important generic computational properties and the practical application of functional insights into cortical processing means that the work in this thesis provides unique solutions to important problems in artificial intelligence.

1.3 Thesis Structure

The thesis can be partitioned into three main sections. The first of these provides an overview of the biological inspiration and artificial neural network methods used throughout the work. The second focuses on the self-organising map algorithm and the implementation of this utilising the principles introduced in the first section. The final section extends work on the model from the second section to demonstrate the learning of hierarchical concept representations within

such a neural module. Specifically, the chapters are as follows.

In chapter 2 a survey of biological neural function is presented. This begins with an introduction to how neurons process information, and how learning occurs in biological neurons. It continues with a review of several key components of cortical networks that represent the inspiration behind the functional processes designed into the current models. Specifically, these are cortical connectivity, oscillatory activity, and temporal coding of information. The chapter finishes with brief reviews of cortical map formation and higher-order object representation in the brain.

In chapter 3 a survey of artificial neural network methods is presented. Common artificial implementations of neurons are introduced, with a focus on spiking neurons and related mechanisms, such as artificial representations of synaptic properties and neural plasticity.

Chapter 4 provides an introduction to the self-organising map. The traditional algorithm is described, and several uses of and modifications to the algorithm are discussed. Relevant extensions to the algorithm are then highlighted, primarily the artificial recreation of specific cortical regions using self-organising networks, and the existing implementations of the algorithm that use spiking neurons.

The second section consists of chapter 5. The spiking self-organising map model is described here, with full details of the implementation and parameter testing. Results are then provided that confirm the robustness and reliability of map formation. These consist of the network's learning in response to two-dimensional input data, particular datasets representing categorical information, and an alternate description of the model designed for implementation on analogue spiking neuron hardware.

The third section consists of chapter 6. Extensions to the existing model are introduced here that alter functional aspects of the self-organising map to make it capable of representation of concept hierarchies consisting of sets of binary feature-object pairings. Results demonstrate the ability of the network to learn and recall at multiple hierarchical levels through the modification of particular network parameters.

Finally, a discussion is presented in chapter 7 regarding the relationship between each of these

models and existing work in the fields of artificial intelligence and cognitive neuroscience. The implications for generating modular implementations of cognitive processes in networks of spiking neurons in the future is also discussed.

Chapter 2

Neural network biology

Information is processed in the brain through communication between neurons. The combination of ways in which neurons operate, networks are structured, and brain structures are modified by experience lead to the full range of functions and thoughts that the mind is capable of. For artificial intelligence methods to gain from knowledge of how complex biological systems operate it is necessary to have an understanding of the fundamental principles involved in generating those systems. This chapter will briefly describe the fundamentals of processing in the brain, by introducing how neurons function, communicate with each other, are organised, and learn.

2.1 Neurons

Neurons typically consist of four components (Kandel et al. 2000). First, the cell body (soma) is the metabolic centre of the cell. Second, a tubular axon protrudes from the soma. The axon is the part of the cell most responsible for transmitting nerve impulses to other neurons. The axon branches before terminating in the third component, the presynaptic terminals. These terminals meet with other neurons and transmit the signal by releasing neurotransmitter into the synapse, which is the gap between cells. The side from which the signal originates is the presynaptic side, and the signal's target on the opposite side of the synapse is the postsynaptic side. Finally, a branched set of protrusions from the soma, the dendrites, are the primary components through which the cell receives presynaptic signals. They spread out to potentially contact and receive information from many presynaptic cells.

A neuron can also be described in terms of its functional, rather than morphological, components. Neurons can be schematically represented as having four functional components: input,

integration, trigger and output (Kandel et al. 2000). The input component is the incoming electrical charge, which would usually be the combination of signals received through the dendrites. The integrative component is the resultant change in the membrane potential in the soma (see section 2.1.1). The trigger component is a mechanism that evaluates the state of the neuron and triggers an appropriate response in the form of an action potential (see section 2.1.2). Finally, the output component is represented by the release of chemical neurotransmitter at the presynaptic terminal, which determines the nature of the signal received by the postsynaptic neuron.

2.1.1 Membrane potential

The membrane potential is a value representing the concentration of ions within the cell relative to the extracellular concentration. Typically a neuron has a higher concentration of negatively charged molecules within the soma, which leads to a negative membrane potential at rest, known as the resting potential. With no stimulation from other neurons, a neuron is normally in a state of equilibrium at this value. The negatively charged molecules involved are amino acids and proteins and the positively charged molecules are sodium (Na^+) and potassium (K^+) ions. The membrane is selectively permeable to K^+ ions, with the Na^+ concentration kept low by an ion channel pump, meaning at rest there are vastly more K^+ than Na^+ ions present within the cell (Kandel et al. 2000).

The membrane potential is measured relative to the extracellular charge (effectively zero), and is usually at around -65 mV. This value represents a baseline that changes in membrane potential can be measured relative to. The membrane has both capacitive and resistive properties, such that an instantaneous increase in the level of a constant input current will lead to a slowly rising membrane potential. A purely resistive component would result in an instant rise, and a purely capacitive component would result in a linear rise. Likewise, a step decrease in the level of input current will result in the membrane potential gradually decreasing to the resting potential, with faster leakage at first that gradually slows (Kandel et al. 2000).

2.1.2 Action potentials

The input component essentially reduces the magnitude of the membrane potential through an influx of positive ions. This increase in voltage gradually opens an increasing number of voltage gated Na^+ channels, and then at a certain threshold, around 10 mV above the resting level, many voltage gated Na+ ion channels are open, making the membrane much more permeable to Na+ ions. In turn, this leads to a large change in membrane potential, as positive ions flood into the cell.

In the aftermath of the action potential, the membrane potential is usually transiently hyperpolarised by K^+ ion channels remaining open and allowing K^+ to exit the cell for longer than the action potential duration, as the ionic balance returns towards its resting state. The excitability of the membrane is also diminished through a residual inactivation of Na^+ channels. The neuron goes through an absolute refractory period, lasting approximately 1 ms, in which no second action potential can be generated. The neuron then goes through a relative refractory period, usually lasting for an additional 3-4 ms although maintaining an effect on the firing probability of the next spike for up to 20 ms (Gray 1967), in which stronger input than normal is required to generate an action potential.

The cell membrane is most sensitive to voltage changes, and has a lower threshold for voltage-gated ion channel activation, at the base of the axon, a site known as the axon hillock. This site is essentially the trigger point for an action potential. The potential travels from this point along the axon, which is often insulated against leakage along its length, and regenerates the potential at interval sites, preventing degradation.

2.1.3 Synapses

An action potential propagates along the axon to the presynaptic terminals. These terminals are adjacent to a small inter-cellular gap known as the synaptic cleft, on the opposite side of which is the postsynaptic neuron. An action potential arriving at the presynaptic terminal causes voltage gated calcium (Ca^{2+}) channels in the membrane to open, such that positive ions flow into the presynaptic neuron. This Ca^{2+} influx causes release of neurotransmitter from the presynaptic

terminal into the synaptic gap. The neurotransmitter molecules 'dock' with receptors on the postsynaptic membrane, causing receptor channels to open, which allows ions to flow into the postsynaptic cell. As such, the membrane potential of the postsynaptic cell is altered, resulting in what is known as a postsynaptic potential (PSP).

Excitatory neurons, such as pyramidal cells in the cortex, commonly release glutamate amino acid neurotransmitters into the synapse, which result in the opening of both Na^+ and K^+ ion channels. These molecules depolarise the postsynaptic cell in an excitatory response known as an excitatory postsynaptic potential (EPSP). Inhibitory neurons commonly release GABA amino acid neurotransmitters into the synapse, which gate chlorine (Cl^-) ion channels, allowing negative ions into the postsynaptic cell. This action hyperpolarises the membrane potential, generating what is known as an inhibitory postsynaptic potential (IPSP). Excitatory actions make the postsynaptic neuron more likely to generate an action potential, and inhibitory actions make it less likely to.

An individual synaptic response, caused by the arrival of an action potential at just one presynaptic terminal, can be characterised as an α-function, reaching a peak amplitude between 0.1 and 3 ms (Jack et al. 1971), but decaying much more slowly. The duration and amplitude of synaptic responses are highly variable between neurons and synapses (Jack et al. 1971). Also, incoming synaptic responses can interact with each other, with the history of previous synaptic responses modulating the profile of an additional response. These modulations are related to the morphological structure of the neurons, with the ordering of synapses on the dendritic spine playing a crucial role in the effects that are caused by signal combinations (Branco et al. 2010; Rumsey and Abbott 2006). As such, a post-synaptic neuron does not necessarily always compute a linear sum of its inputs. The depolarisation of the membrane potential caused by the combination of all incoming PSPs is what normally leads to generation of an action potential. In essence, a neuron combines its inputs at a given time, and makes a simple all or nothing decision about whether or not to send an output signal of its own.

2.2 Learning

A major facet of learning in the brain is the ability of synaptic efficacies to be adjusted such that a presynaptic action potential leads to either a stronger or weaker postsynaptic potential. This change is made if the presynaptic neuron should have its ability to contribute to the firing of the postsynaptic neuron increased or decreased. Various mechanisms affect synaptic efficacies, or weights, over a variety of timescales. This section will examine those mechanisms in terms of two broad timescales, short- and long-term. A commonality between all of the types of plasticity described here is that they are activity dependent: activity is required in at least one of the pre- or post-synaptic neurons for synaptic weight changes to occur.

2.2.1 Short term plasticity

Short term plasticity is generally on a millisecond timescale, but some forms can last for up to several minutes (Zucker and Regehr 2002). Plasticity that temporarily enhances the PSP resulting from a presynaptic action potential is called facilitation. The counterpart of facilitation is depression, which decreases the amplitude of PSPs generated by successive action potentials. Both facilitation and depression can be caused by either individual pairs of successive action potentials, or by a continuous incoming spike train over a longer time period.

Paired pulse depression is observed in most synapses after an action potential in a neuron follows a previous action potential in that neuron by less than $\sim$ 20ms (a 20ms interstimulus interval) (Citri and Malenka 2008). It is likely to result from a transient depletion of vesicles that are ready to release their neurotransmitter at the presynaptic terminal. Less neurotransmitter is able to be released following the second action potential, so the PSP is weaker. Paired pulse facilitation is seen in synapses over longer interstimulus intervals than paired pulse depression, between 20-500ms. This is likely to be primarily caused by leftover Ca^{2+} in the presynaptic terminal increasing the amount of neurotransmitter release caused by the second action potential. Both of these effects result from changes to the amount of neurotransmitter released following an action potential, so which effect is seen is determined by the recent history at the synapse, as well as the natural neurotransmitter release tendencies of a particular presynaptic terminal. Terminals that are less likely to release a lot of neurotransmitter normally tend to be facilitated

by paired pulses, and those that are more likely to release a lot of neurotransmitter tend to be depressed (Dobrunz and Stevens 1997).

Trains of stimuli can also lead to either facilitation or depression of a synapse. In this case the effects tend to be slightly longer lasting, and they need a longer timescale to occur, to allow for multiple spikes (200ms to 5s). A synapse can be facilitated following high frequency stimulation (10-200Hz); this is again thought to be due to build up of Ca^{2+} in the presynaptic terminal enhancing the probability of neurotransmitter release (Citri and Malenka 2008). Depression resulting from trains of stimuli is also primarily related to depletion of presynaptic, release-ready vesicles of neurotransmitter, but can occur at some synapses that do not deplete after a pair of pulses. Some synapses are likely to have a rapidly replenishing pool of release ready vesicles, which requires prolonged high frequency activity to deplete.

The type of short term plasticity present at a synapse determines some of the functional characteristics of that connection. Facilitating synapses will transfer high frequency bursts more reliably than low frequency, and depressing synapses will transfer low frequency bursts more reliably than high frequency. This quality makes facilitating synapses act as high pass filters, and depressing synapses act as low pass filters (Abbott and Regehr 2004). Modulating the initial probability of neurotransmitter release can switch a synapse from one mode to the other.

2.2.2 Long term plasticity

Changes in synaptic efficacy can occur that endure for hours or days. Long term plasticity can come in the form of an increase, long term potentiation (LTP), or decrease, long term depression (LTD) in the connection strength. While short term changes alter the filtering properties of synapses, the enduring connection strengths, determined by LTP and LTD, are considered to be a fundamental source of information storage in a neural network (Martin and Morris 2002; Fuster 2003).

LTP raises the synaptic efficacy by increasing the number of glutamate receptors on the postsynaptic membrane (Citri and Malenka 2008). Two major types of glutamate receptor are involved in the mechanism controlling this effect: NMDA receptors (NMDARs) and AMPA receptors (AMPARs). AMPARs are permeable to positive ions, Na^+ and K^+, and provide most of the

inward current generating the EPSP. NMDARs are normally blockaded by Magnesium (Mg) molecules at the resting membrane potential. As the membrane depolarises through positive ions entering via AMPARs, the Mg blockade of NMDARs is released, and Na^+ can also enter the cell through NMDARs. However, unlike AMPARs, NMDARs are also permeable to Ca^{2+}. A large calcium concentration increase in a dendrite leads to a signalling cascade within the cell involving a calcium-dependent protein kinase, CaMKII, which in turn promotes the incorporation of new AMPARs into the postsynaptic side of the synaptic gap. In this manner, the unblocking of NMDARs generates the neurotransmitter receptor changes that represent LTP.

LTD, in contrast, lowers synaptic efficacy by reducing the number of postsynaptic AMPARs. The mechanism is related to that of LTP, in that LTD also requires an increase in postsynaptic calcium concentration caused by activation of NMDARs. LTD is induced by a modest calcium concentration increase, whereas LTP requires a larger increase, above a threshold. A small calcium increase leads to preferential activation of particular proteins that lead to the endocytosis of AMPARs from the membrane, reducing synaptic efficacy (for detailed reviews of the LTP and LTD processes, see (Citri and Malenka 2008; Sjöström et al. 2008; Feldman 2009)).

The size of the active zone (the section of membrane able to hold receptors and receive signals from the presynaptic terminal) on the postsynaptic site of the synapse is correlated with the number of AMPARs within that active zone. Morphological changes have been reported to accompany the long term induction of LTP, and it seems likely that remodelling of the size of the pre and postsynaptic active zones is an important factor in the long term maintenance of synaptic changes.

LTP and LTD are both activity dependent processes. The main experimental condition for the induction of long term plasticity is repeated, external, electrical stimulation of a presynaptic cell. When this stimulation is at a low rate (1Hz), LTD is induced; at a high rate, LTP is induced (Sjöström and Nelson 2002). However, the relative timing of pre- and postsynaptic spikes has also been shown to play a crucial role in the direction of weight change. Repeated single spike pairings at very low frequencies have been shown to elicit both LTP and LTD. LTP is induced if the presynaptic spike precedes the postsynaptic spike, and LTD is induced if the postsynaptic

spike precedes the presynaptic spike (Abbott and Nelson 2000). This important phenomenon is known as spike timing dependent plasticity (STDP).

The STDP effect can be explained by the contribution of postsynaptic action potentials to the membrane potential. An action potential in the postsynaptic neuron occurs when the membrane potential at the base of the axon reaches a threshold; the glutamate channels are more sensitive in this part of the neuron, so the threshold is lower there, so action potentials will initialise at that point. When an action potential occurs, the membrane depolarisation backpropagates to the dendrites, and to each postsynaptic active zone (this is known as a backpropagating action potential). Depolarisation of the dendrites will be strongest immediately after the coincidence of pre- and postsynaptic action potentials. At this time, NMDA receptors will be unblocked most easily, and the largest influx of calcium will take place (Caporale and Dan 2008), with large calcium concentrations being the key to triggering LTP. LTD is explained by assuming that a backpropagating action potential leads to an afterdepolarisation potential, which, when coinciding with the PSP, leads to only a moderate calcium influx, triggering LTD.

In STDP, LTP and LTD are found within temporal windows. The duration of these windows has been found to be on the order of tens of milliseconds (Bi and Poo 1998; Markram et al. 1997; Abbott and Nelson 2000). Figure 2.1 shows example LTP and LTD windows, generated from observed changes of synaptic efficacy. However, the exact nature of these temporal windows varies widely across brain region, cell types, synapse types, and conditions (Sjöström et al. 2008). Importantly, STDP effects also depend on frequency (Sjöström and Nelson 2002) in a similar way to traditional LTP and LTD, with depression dominating at low frequencies and potentiation at high frequencies. It makes sense that the exact nature of STDP is malleable and varied, given that it relies on transient levels of a dynamic molecular concentration in the presence of continuously fluctuating opening and closing of ion channels.

In summary, LTP and LTD, and in particular STDP, can be viewed as the molecular mechanisms demonstrating the Hebbian postulate in action, that causal links between neurons should be strengthened, and remembered. The next section will introduce some of the aspects of the mammalian cortex that interact with these types of learning mechanisms to generate functional

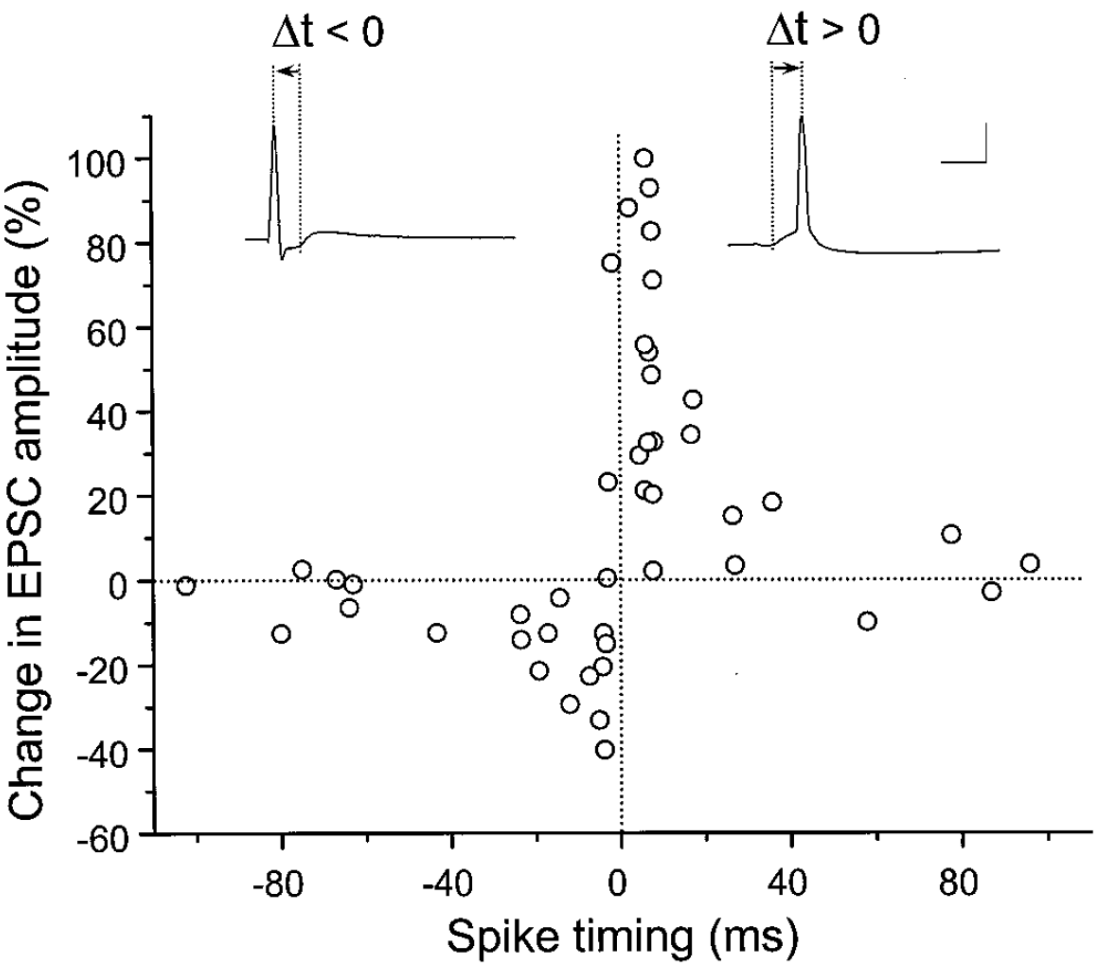


Figure 2.1: Critical window for the induction of potentiation and depression, as observed by Bi and Poo (1998). Here, EPSC stands for excitatory postsynaptic current, and Δt is the temporal lag between the pre- and post-synaptic spike. A negative Δt indicates a postsynaptic spike ahead of a presynaptic spike, resulting in a negative change in EPSC amplitude. A positive Δt indicates a presynaptic spike ahead of a postsynaptic spike, resulting in a positive change in EPSC amplitude.

networks in the brain.

2.3 Cortical features

The functionality present in a given neural network will depend on a combination of the properties of individual neurons (such as those outlined in section 2.1) and the connections between those neurons. The connectivity patterns possible in a network are potentially unlimited, and yet a selection of properties recur throughout the cortex of mammals. Mammalian cortex is crucial to the processing of sensory information, memory storage, and various cognitive functions. The structures present in the cortex underlying these functions could potentially represent a set of building blocks for the fulfillment of these various types of processing. This section will examine the extent to which particular properties of network connectivity and function can be identified as common to a wide range of cortical areas, and therefore can be considered to underpin a selection of cognitive processes.

The human cortex consists of neurons arranged into six distinct layers, with vertical interconnectivity across these layers throughout (Dayan and Abbott 2000). Horizontally, the cortex can be laid out as a 2D sheet, mapped to a set of densely intraconnected communities, or modules, linked through hub nodes, which are highly connected with several individual communities (Sporns 2011; Hagmann et al. 2008). At the largest scale the tightly grouped clusters of connections are entire cortical areas. A notably similar structure exists at the lowest levels, with clusters of neurons bound together in microcircuits, collections of microcircuits constituting cylindrical cortical columns, and groups of columns representing cortical subareas (Breakspear and Stam 2005).

2.3.1 Connectivity

The nature and range of connectivity patterns within the cortex varies across the different layers, with characteristic patterns seen for particular layers throughout the cortical sheet. Generally speaking, the axons of excitatory neurons in layer 6 project in a feedforward manner up to layer 4, and layer 4 neurons project to layer 3. Layer 3 neurons exhibit more variety, with intralayer lateral projections, and feedback projections to layer 5. In layer 5, excitatory connections project

both forward to layer 3 and back to layer 6. Similar projection patterns have been seen in many well studied cortical areas, such as cat and primate visual cortex, primate auditory and motor cortex, as well as a slight variant in rodent barrel cortex (see (Douglas and Martin 2004) for review). The computational functions of this set of vertical projections are thought to include modulation of responses to centre-surround activity, contrast variance, as well as modulation of subthreshold activity (see (Raizada and Grossberg 2003) for review).

Intralaminar connections are less common within layers 4-6, with most excitatory cells terminating in higher layers, in a feedforward pattern. A significant preference for vertical connections is seen throughout these layers (Stepanyants et al. 2008). As such, lateral connections are more commonly seen within layers 2 and 3; indeed, the number of potential inputs is roughly 3 times higher for a layer 2 or 3 neuron than for neurons in lower cortical layers (Stepanyants et al. 2008), with around 70% of a layer 2 or 3 cell's excitatory input being derived from lateral locations (Binzegger et al. 2009).

Several studies have attempted to map the range and strength of horizontal connections within these layers. Boucsein et al. (2011) used photostimulation, in which short pulses of UV-light are used to uncage glutamate channels and cause neuronal activity, to evaluate physiological parameters of horizontal connections, determining a maximum connection probability of 0.15 between neuron pairs, decaying exponentially with distance between the pair, with a constant of 330 µm. Recordings in neuron pairs in slices from rat and cat visual cortex have been reported (Markram et al. 1997; Thomson et al. 2002; Holmgren et al. 2003; Song et al. 2005; Lefort et al. 2009), with connection probabilities of less than 0.1 found at distances up to 200 µm. Conversely, (Bureau et al. 2004) demonstrated high connection probabilities in layer 2 or 3 of rat barrel cortex, ranging from roughly 1.0 to 0.13 at distances from 0 µm to 400 µm respectively. There is a consensus across these studies, and others, that connection probability decreases with distance, and that the majority (around 75%) of connections are 'local', within approximately 500 µm (reviewed in Voges et al. (2010)).

Lateral connections at distances greater than 500 µm tend to be made in "patches" (Voges et al. 2010; Muir et al. 2011). This means that, while the connection probability for neurons at greater

distances (up to 8mm) is low, it is above average at particular regions situated around a spatial centre. This makes sense in terms of efficiency, in that an axon may terminate in one place, making contact with a cluster of neighbouring efferent neurons. The number, size and distance of these patches is unclear, and certainly variable, with around 25% of layer 2 or 3 neurons lacking patchy connections altogether (Voges et al. 2010).

Approximately 20% of cortical neurons are inhibitory interneurons (Isaacson and Scanziani 2011). The afferent connection profiles of these neurons is markedly different from that of excitatory neurons (Fino and Yuste 2011). Inhibitory neurons receive inputs from a large fraction of neurons within around 100 µm, and can inhibit over 50% of excitatory cells in the same region. An analysis of expected numbers of synapses between neuron pairs in layer 4 suggests that inhibitory to excitatory connections remain dense in an area 2 to 3 times larger than excitatory to inhibitory connections (Stepanyants et al. 2008), although inhibitory neurons also receive input from long range excitatory sources via nearby axons. It has been found that up to 51% of synapses from excitatory cells are onto inhibitory targets (Bock et al. 2011). This was caused, despite the lower preponderance of inhibitory cells in cortex, by individual axons making multiple synaptic connections with inhibitory dendritic targets. Electron microscopy has revealed that the input preference of excitatory neurons does not determine their likelihood of connecting to a given inhibitory neuron; instead spatial neighbourhood determines the likelihood. However, some inhibitory neurons have also been shown to have their own input preferences (Isaacson and Scanziani 2011; Hirsch et al. 2003). Inhibitory neurons that pool excitatory input are likely to have a role in mediating the activity of nearby excitatory cells (Bock et al. 2011), with prominent likely roles including the sharpening of tuning and oscillatory synchronisation (Isaacson and Scanziani 2011).

A model of the general functions of the cortex, based on the above connection profiles, can be proposed (Douglas and Martin 2004). Layer 2 or 3 patches essentially receive a sample of information from both feedforward and lateral sources. These excitatory cells can then collectively participate in a soft winner-takes-all feature resolution process, mediated by dense lateral inhibition. The patterns established at this stage are fed back to layer 5 neurons, which con-

duct another kind of soft winner-takes-all competition, generating output to drive sub-cortical processes. The winner-takes-all-like function in layer 2 or 3 represents cooperative interaction within and between patches of the cortical sheet to find an interpretation consistent with inputs across the lower levels of processing. This model is tentative and, while it takes account of major physiological observations, presenting them in a way that can be utilised by theoreticians, much investigation remains to be done into the detailed organisation of specific cortical regions and circuits (Douglas and Martin 2004).

The following sections briefly investigate other cortical properties related to the interactions between excitatory and inhibitory neurons, and what the functional outcomes of these mechanisms look like.

2.3.2 Oscillations

Oscillatory activity can occur in the brain when pools of neurons synchronise their firing, which results in more spikes within a certain temporal window, followed by fewer spikes in the subsequent temporal window. This type of behaviour is seen at various rates in a variety of brain locations. It is usually recorded through the time course of a local field potential (LFP), which measures the average electrical potential across an area, representing average neural activity in that region. This section will introduce some important types of oscillatory behaviours and their functions, and describe some of the underlying mechanisms generating this activity (for a detailed review see (Wang 2010)).

Oscillations can be categorised through their approximate rate, with each category typically occurring in different locations and serving a different purpose. Theta rhythms (4-8 Hz) are often observed in the hippocampus (Buzsáki 2002), and have also been found in the cortex during working memory tasks (Jensen and Tesche 2002). Beta rhythms (15-30 Hz) have been observed in primary motor cortex (Swann et al. 2009), where they generally occur immediately prior to the onset of a motor movement, and so are thought to be related to inhibitory control in the motor system. They also occur in processing in olfactory cortex, (Perez-Orive et al. 2002). Gamma rhythms (30-100 Hz) have been observed in sensory (Gray et al. 1989; Singer and Gray 1995; Traub et al. 1996) and motor cortex (Schoffelen et al. 2005), and in hippocampus

(Csicsvari et al. 2003).

Recurrently coupled neurons can be synchronised by each other in one of two distinct ways, with a particular characteristic of the neurons in question determining which of these ways is more likely to play a role in the synchronisation. Normally, a presynaptic spike will excite a postsynaptic neuron, encouraging the postsynaptic neuron to fire, advancing the phase of firing of the postsynaptic neuron. However, some neurons, known as type II neurons (Wang 2010), actually exhibit a delay to the phase of firing in response to an excitatory presynaptic input that arrives soon after a previous postsynaptic response. In a type II neuron, a depolarising action potential activates a restorative current, like a voltage activated potassium current, which then decays while the membrane returns to its resting state. This decay, and therefore eventual return to equilibrium for the membrane, is delayed by a brief depolarisation during the after-hyperpolarisation phase (e.g., an incoming EPSP), eventually delaying the postsynaptic spike. Oscillatory behaviour in pairs of this type of neuron can be caused when one neuron fires, exciting a second neuron, advancing the phase of firing of that neuron, which then fires, and the recurrent connection causes an EPSP in the first neuron. This EPSP can come soon after the initial spike in the first neuron, delaying the decay of the restorative current, delaying its firing. The outcome is synchronised firing, followed by a synchronised lull in activity.

Neurons that do not display this delay in firing phase, type I neurons, are much less likely to be synchronised by mutual excitation, as recurrent connectivity always advances the spike time. The more likely source of synchronisation between these neurons is mutual inhibition (Ermentrout 1996; Bartos et al. 2007). In this case, while the reciprocally connected neurons excite each other, their firing also excites reciprocally connected inhibitory neurons, which, as a result of the excitatory activity, fire just after the excitatory cells. This type of phase relationship is shown in figure 2.2. The excitatory neurons are inhibited by the IPSPs, and prevented from firing for a duration, generating an oscillatory rhythm. This effect is known as an excitatory-inhibitory feedback loop. The magnitude of inhibition during one oscillation instantaneously affects the oscillation frequency, with a stronger inhibition resulting in a longer time until the next oscillation (Atallah and Scanziani 2009). The excitatory-inhibitory feedback effect can

be so pronounced that Cobb et al. (1995) showed experimentally in slices of hippocampus that a single inhibitory cell could synchronise spiking activity of pyramidal cells. This type of loop mechanic can be simulated and the effect manipulated through parameter alterations, with the duration of inhibition affects the synchronisation tightness in random networks, with faster decay of IPSPs leading to tighter synchrony (Börgers and Kopell 2003).

The mechanisms mentioned above identify some key principles underlying the common observation of synchronised oscillatory firing in brain networks. Individual neuronal membrane properties, types of cells in networks, and network connectivity contribute to these mechanisms. The relationship between excitatory and inhibitory activity is an important generator of oscillatory behaviour. The next section describes the relationships between spikes seen within a single oscillation, and begins to identify the importance of the relationship between excitation, inhibition and oscillations in terms of how information is processed.

2.3.3 Temporal coding

The generation of oscillatory behaviour in neurons through an excitatory-inhibitory feedback loop has an important functional implication. Rhythmic phases of inhibition ensure that a neuron can only respond to input within a limited window. Within each oscillation, the most excited cells will respond earliest as inhibition wanes, exciting inhibitory neurons again, and in doing so silencing weakly excited cells (Fries et al. 2007). This means that the phase of a spike within an oscillation can potentially play a crucial role in communicating information about a stimulus. It has been demonstrated in rat somatosensory cortex that some neurons respond only to a specific stimulus byy generating a strong output, and thus all of the information they can encode is converted into an output spike rate. Other neurons respond with differing temporally structured responses to a variety of stimuli, and so require the temporal properties of their output to be taken into account for the information content of their response to be captured. In the former case, no information about the stimulus is lost by assuming the output is temporal coded. However, in the latter case assuming that the information is encoded in output spike rate alone loses almost all of the information (Panzeri et al. 2001). The reliable conversion of information initially encoded in firing rates into synchronised temporal information has been

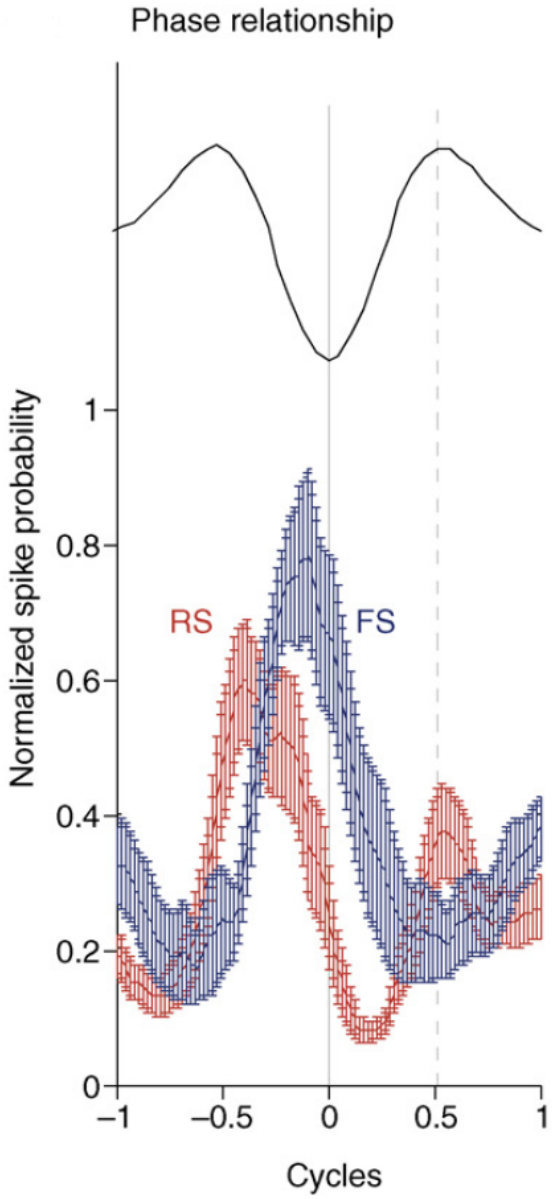


Figure 2.2: Average normalised spike probability for excitatory pyramidal cells (red) and inhibitory interneurons (blue) in rat hippocampus. Firing likelihood increases noticeably earlier in the phase for excitatory cells than inhibitory cells. This phase relationship is important for generation of oscillations and phasic coding of information in type I neurons. Reproduced from Fries et al. (2007).

demonstrated in rodent cortex (Reyes 2003). The direct encoding of tonic excitation in the phase of firing has been demonstrated to be highly accurate within theta oscillations, but was not possible within gamma oscillations due to a lack of temporal precision of neuronal firing (McLelland and Paulsen 2009). However, the phase code was shown to be significantly more efficient than an equivalent rate code (McLelland and Paulsen 2009).

In rodent somatosensory cortex, Panzeri et al. (2001) demonstrated that 44% more information can be conveyed about stimulus location through spike timing of single neurons than with a rate-based coding scheme. Spike trains from barrel cortex were analysed, and indeed it was found that 83% of the total information contained in the spike train was contained in the timing of the first spike upon each stimulus presentation. Additionally, Ghazanfar et al. (2000) showed that decreasing the resolution of the temporal code in barrel cortex, while retaining all rate information, significantly reduced the performance of a classification task using spike train data. Jittering of spikes within the train also degraded performance, suggesting that relative phase relationships between spikes within a pattern encode information about the stimulus. Subsequently, Foffani et al. (2004) generalised these results from barrel cortex to further areas of rat primary somatosensory cortex with a study involving the forelimb region. Identification of stimulus location decreased significantly with an increase in bin size from 5ms to 20ms. By taking just the first spike in each neuron, 94% of the performance from the full spike pattern was retained. In an analysis of spike coincidence in primate primary visual cortex, Maldonado et al. (2008) showed that coincidence rate is significantly higher than would be expected for a brief duration after stimulus onset, making the very first spikes preferentially synchronised. These examples suggest that stimuli with a temporal component (visual), and stimuli that are represented by both continuous (forelimb) and non-continuous (whiskers) spatial information can all be represented using a phase code. Representation using a phase code is plausible not only for temporal components of stimuli, but also for spatial information.

Processing in the human visual system furthers the claim that non-temporal variables are temporally coded in sensory cortex. The visual processing component of classification tasks occurs very rapidly, within 150ms of stimulus presentation (Thorpe et al. 1996; Maldonado et al. 2008).

This leads to the conclusion that, due to the number of synaptic integration steps between presentation of a stimulus to the retina and selective response in cortical areas, information must be transmitted using a coding scheme involving 0, 1 or 2 spikes at each stage (Thorpe and Imbert 1989). A theory that attempts to explain visual processing within this constraint is proposed by Van Rullen and Thorpe (2002), using a rank coding in which the firing order of neurons encodes information from downstream layers. This model involves a series of saliency to latency transformations, with the best representatives of a stimulus firing earliest. At the earliest stages of visual processing cells encode relative contrast within their receptive field. Continuous variables, such as contrast, can be represented in temporal sequences through this saliency to latency transformation. Additionally, investigation into primate V1 reveals that saccadic eye movements cause modulation in the amplitude of LFP oscillations, which in turn increases the level of phase-locking in early spikes within an oscillation (Ito et al. 2011). This suggests that the effect of properties of the sensory apparatus helps to enhance the role of temporal processing, contributing further evidence that spiking within individual oscillations encodes stimuli.

Temporal coding has also been shown to play a role in processing at a higher level than initial sensory assimilation. In primate facial recognition tasks, neurons in inferior temporal (IT) cortex respond to primate faces up to 15ms earlier than they respond to non-primate faces (Kiani et al. 2005). Peak information transmission rate occurs 100-200ms earlier for global categories (general face shapes) than for fine categories (facial expressions), indicating different temporal stages of processing for face recognition tasks (Sugase-Miyamoto et al. 2011). In monkey prefrontal cortex, short-term memory of separate objects has been shown to lock to separate phases of the LFP gamma oscillation (Siegel et al. 2009).

This section has summarised evidence that temporal coding plays an important role in neural information processing. Compared to a neural code based on rate alone, more information can be transmitted and processing can permeate a layered structure more rapidly. There are intuitive ties between readily observable oscillatory behaviour, the mechanisms underlying those oscillations, and a coding scheme based on spike phases within those oscillations. Finally, there are multiple examples available of neural representations of non-temporal variables by tempo-

ral sequences of spikes, although at higher rates of oscillation these codes are more likely to be based only on the spike order, rather than on the specific phase of firing (McLelland and Paulsen 2009). Section 2.4 will investigate common methods of representing information spatially in the brain.

2.4 Neural maps

A common feature across sensory cortex is the map-like arrangement of neurons in terms of their response properties (Kaas and Catania 2002). Neurons with similar stimulus preferences tend to be found near to each other. More importantly, the neuron's input stimulus preferences are topographically arranged; transitioning through a continuum of input stimuli, the neurons that respond best to those stimuli will be found on a spatial continuum. Discontinuities in the sensory receptor domain are reflected in septal regions of cortex, sometimes connecting a series of individual maps.

A clear example of this is found in somatosensory cortex. Cortical columns in somatosensory cortex are selectively responsive to touch signals from particular regions of the skin surface. The columns are arranged such that they form a complete topographic representation of the body (Kandel et al. 2000). The area of cortical surface that represents a sensory area is proportional to the degree of innervation in that area of the sensory apparatus, i.e. the cortical area reflects the density of sensory neurons in the skin, rather than skin area. A particularly well defined and studied region of somatosensory cortex is the barrel cortex in rodents, mapping signals from whiskers (Fox 2008). Here, spatial relationships of representations on the cortical surface mirror clearly the spatial arrangement of whiskers on the face. A septal region between each barrel reflects the discrete nature of the sensory receptor. A similar arrangement is apparent in the nose representation for the star-nosed mole, where each nasal appendage is represented in a cortical area separated from adjacent regions by a less specific section (Catania and Kaas 1996).

Physical location of different receptor cells in the sensory apparatus may be responsible for the form of the representations (Kandel et al. 2000). Cortical neurons that respond to adjacent areas of skin are adjacent to each other, and respond to highly overlapping input patterns. Also, studies into the plasticity of representations suggest that somatosensory neurons receive input from

a wide area of the sensory surface. While an area of skin is anaesthetised, the receptive field of cortical neurons representing that specific area grows rapidly, encompassing sensory regions outside of the anaesthetised area (Calford 2002). Neurons show, within minutes, responses to areas that previously they did not respond to. One explanation for this phenomenon is that, when an area of skin is anaesthetised, a tonic source of inhibition stemming from that area is removed, allowing a neuron to be activated by sensory apparatus that it may have received a less strong excitatory input from previously (Calford 2002). This is a useful functional mechanism, in that when an area of sensory apparatus no longer provides any information, the associated area of cortex can adapt to represent a useful region of the sensory apparatus, while still providing a high level of specificity under normal conditions. Indeed, further evidence supports the idea that somatosensory neurons receive input from a much wider range of afferent projections than the fine-grained nature of cortical representations would suggest (Rausell et al. 1998; Varga et al. 2011).

Visual cortex reveals a striking set of overlapping input stimulus preference continuums (Miikkulainen et al. 2004). Firstly, retinotopic maps of visual field position preference can be identified (Wandell 2011; Kolster et al. 2010). If the visual cortex is flattened onto two dimensions, neurons that have a preference for a central position in the visual field are found at the centre of that space. Preferences for increasing eccentricity in the visual field are found at increasingly eccentric locations in cortex. Throughout this retinotopic map, there are also bands of ocular dominance (Casagrande and Kaas 1994; Horton and Hocking 1996). Areas for each region of the visual field contain neurons that prefer each eye over the other. These regions are adjacent, and formed into bands, such that a continuous preference for one eye across an area of the visual field lies adjacent to a band reflecting a continuous preference for the other eye across the same area of the visual field. Retinotopic discrimination, a continuous variable, retains a continuous representation, while being striped with discrete regions for eye preference, a discrete variable. Furthermore, across each pair of adjacent ocular dominance bands within a particular section of the retinotopic map, lie columns with distinct preferences for the full range of contrast orientations (Hubel et al. 1978; Blasdel and Salama 1986; Kandel et al. 2000). The columns are formed into a continuous order in terms of the shift in axis of preferred orienta-

tion, with the full orientation range revolving in order around areas known as 'pinwheels' (Ohki et al. 2006). A complete cycle of orientations exists across approximately every 0.75mm in primate visual cortex. The columns consisting of a full set of orientations across a pair of ocular dominance bands are known as a 'hypercolumn', with activity in a hypercolumn reflecting the properties of a particular region of the visual field. Moreover, the patchy lateral connections that extend horizontally from layer 3 of the cortex have been shown to span approximately the width of one hypercolumn, preferentially terminating in regions tuned to a similar orientation (Gilbert 1989; Bosking et al. 1997; Sincich and Blasdel 2001). A close relationship has been shown between patchy structures within cortex and cortical responses encoding a value across the surface, suggesting a prominent role of the lateral connection profile in map establishment (Muir et al. 2011). Preference for temporal frequency of stimulus, on the other hand, was not organised within the map hierarchy, instead being distributed uniformly across visual cortex (Khaytin et al. 2008). This result suggests that cyclic and binary variables like orientation and ocular dominance are more readily mapped than a scalar variable like temporal frequency in visual cortex, demonstrating that in brain maps in general it is possible for different types of variable to be mapped to different types of representation.

Other sensory cortical areas show topographic mappings of stimulus properties. Cat auditory cortex features a tonotopic map-like representation of the cochlea, with a full range of frequency preferences across a continuous section of cortex (Merzenich et al. 1975). A gustotopic map exists in the primary taste cortex of rodents, where each taste category is mapped to a specific cortical region (Chen et al. 2011). In this case there is no smooth transition across sensory features, with taste categories mapped to discrete, specific spatial locations. Moreover, the peripheral afferents in the case of taste are distributed across the tongue, with more abstract response properties determining the mapping to cortical neurons; there is no underlying spatial organisation in the periphery guiding the spatial mapping in cortex. In the olfactory bulb the precise mapping of odorants to specific neurons governs the formation of a topographic map of relationships between odor receptors (Wang et al. 1998).

It is also possible to extend the map principle to non-sensory cortical regions. It has been shown

that the organisation of motor cortex in monkeys follows topographic properties (Graziano and Aflalo 2007). The two-dimensional cortical representation is consistent with a complex, high dimensional input space featuring somatotopic, action category, and spatial information. In multisensory temporal association cortex there is evidence for a modality map, with neurons preferring a particular sensory modality having a close spatial proximity, and being spatially segregated from neurons with differing modality preference, often with bimodal neurons mixed between the groups (Dahl et al. 2009). Neurons in macaque inferior temporal (IT) cortex (area TE) show a positive correlation between stimulus preference among objects of intermediate complexity (e.g. a set of various flat and textured geometrical shapes) and spatial distance (Tamura et al. 2005).

The structural and developmental mechanisms underlying ordered representations in the cortex are not yet precisely known (Honey et al. 2010). Total input to cortical neurons comes in part from afferent, feedforward sources, and in part from lateral sources within the layer. Interestingly, Tan et al. (2011) demonstrated that the average input to a given V1 neuron from both excitatory and inhibitory sources share the same orientation preference. This means that the maximal level of inhibition to a region occurs when receiving an input that also generates the maximal level of inhibition, suggesting a role for inhibition of gain control or a response sharpening "iceberg" effect (whereby subthreshold responses are broader in selectivity than firing responses) in a learned cortical circuit. An analysis of frequency responses in auditory cortex also suggests balanced excitation and inhibition, with the amplitudes of inhibitory and excitatory responses sharing a linear correlation (Wu et al. 2011), and that inhibitory input stems from a broader range of frequencies than excitatory input (Wu et al. 2008), further sharpening the response properties. Temporal advance of the excitatory signal over the inhibitory signal has been shown for both visual and auditory cortical neurons (Tan et al. 2011; Wu et al. 2011), lending more support for the role of response enhancement for feedforward inhibition relative to excitation.

The nature of modulation provided by inhibition within a network is shaped in part by the receptive fields of the inhibitory neurons. In general, inhibitory neuron receptive fields can be tuned

either more widely than excitatory neuron receptive fields, called a lateral inhibition profile, or equally tuned with excitatory neuron receptive fields, called a co-tuned profile (Shapley et al. 2007; de la Rocha et al. 2008). Lateral inhibition could result from: inhibitory cells being more sensitive to input; more presynaptic inputs converging onto inhibitory than excitatory cells; a broader lateral spread of inhibition within the cortex (Levy and Reyes 2011).

Lateral connections in cortex, as already discussed (see section 2.3.1), include several distinct profiles operating in combination. Excitatory connections decrease in likelihood with distance from a neuron, but also exhibit a higher probability in particular patches at long range. Inhibitory connections are local and dense. In terms of the overall effect of these interacting profiles at long-range, there is evidence suggesting that either long-range excitation or long-range inhibition is possible (Miikkulainen et al. 2004). Centre-surround studies have indicated that lateral inhibition is dominant at long range (Grinvald et al. 1994). A small, oriented, high contrast grating presented as a visual stimulus leads to a level of activation in an area of cortex representing that orientation and visual field location. If the stimulus is enlarged but centred at the same location, such that surround areas are activated also, then the response of the central area is dampened, suggesting that the overall impact of surrounding areas on the central area is lateral inhibition (Grinvald et al. 1994). Studies have also indicated that the same connectivity profile can have an excitatory effect under different conditions, for example when the neurons are less activated by using low contrast stimuli (Sceniak et al. 2001). It has been suggested that combinations of feedforward and lateral excitation and inhibition profiles are capable of altering the excitation/inhibition ratio in a way that emphasises a local centre/surround neighbourhood, resulting in a form of lateral competition despite highly overlapping excitatory and inhibitory profiles (Adesnik and Scanziani 2010).

Evidence from the developing visual cortex of cats (Callaway and Katz 1990) indicates that the lateral profile in cortex grows rapidly in the first week after birth, and organises into an adult form within 6 weeks. This suggests that a general lateral connection profile is established quickly, and partially controls the development of feedforward connections (Miikkulainen et al. 2004). Hardwired chemical gradients are known to control the initial routing of axons (Mor-

timer et al. 2009). The initial guidance provided by these processes provides a starting point for an activity driven map refinement once synaptic contacts are made (Rosa 2002). This combination of processing stages helps to explain the consistent properties between individuals regardless of input during developmental stages. Studies have examined the extent to which representations can be reorganised as a result of ongoing experience (see (Kaas 1991) for review). If two fingers of a monkey are surgically connected such that they are always used together, significant overlap develops between the representations of those fingers, which persists even after separation (Clark et al. 1988). In barrel cortex, rapid map plasticity is seen very early in life (Fox 2002), lesions of primary afferents affect development of barrels, and changes in whisker use alter receptive fields (Van der Loos and Woolsey 1973). Additionally, representations can change in response to preferential whisker usage (Fox 2002), but the mechanisms driving this competition are not certain (Feldman and Brecht 2005). Further ideas about the mechanisms likely to contribute to cortical ordering include minimal wiring requirements (Buzsáki et al. 2004). Young et al. (2007) ran simulations indicating that the cortical plasticity observed in the brain would not emerge from plasticity dependent on the coincidence of spikes regardless of the exact temporal relationship between pre- and post-synaptic spikes, but in fact relies on STDP.

The evidence presented in this section suggests a set of features underlying the presence of map formation in sensory cortex. Neighbouring neurons share common afferents but have slight differences in input preference, potentially through initial routing of axons according to hardwired chemical gradients. This may be caused by hard-wired properties of early cortical development. The lateral connection profile provides a combination of local clustering of similar response properties based on short-range excitation, and sharpening of response properties and competition based on inhibition. Long-term plasticity then alters connection strengths to tune response properties to match the representation generated by the cortical structure, strengthening the representation.

2.5 Higher order representation

The inferior temporal (IT) cortex of the macaque monkey has been the subject of substantial investigation into the representation of complex visual stimuli. The processing of a visual stimulus through extrastriate cortex prior to an IT response (see (Orban 2008) for detailed review) consists of a series of responses from neurons selective to different features of the input, such as orientation and direction. However, the response properties of IT neurons are less well known, and appear to depend on particular complex combinations of the shapes identified at lower levels. They tend to lack the visuotopical organisation of earlier areas and feature a combination of categorical and continuous representation, with specific areas representing complex categories like faces and body parts, and wider regions representing other features, particular object attributes, continuously (see (Tompa and Sáry 2010) for review).

The outcome, in terms of object representation, of this combination of categorical and continuous response properties is that a collection of columns is activated by a stimulus (Tsunoda et al. 2001). The columns participating in this activation pattern are thought to indicate which features are present in the stimulus. However, while 42% of the time removing features from a stimulus reduces the number of active columns to a subset of the original pattern, 58% of the time removing features leads to additional column activations. This result suggests that a feature-based representation must involve a combination of both active and inactive columns.

The response of IT neurons to visual images constitutes a data set from which those images can be categorised. Within the population examined, both category-specific neurons, that exclusively respond to one category, and non-category specific neurons, that respond to images from more than one category, were present. Removing the category-specific neurons from the analysis, which constituted about 25% of the neurons present, did not reduce the classification ability of the responses (Thomas et al. 2001). This result suggests that broadly tuned, non-category-specific neurons play a crucial role in object recognition. These neurons are thought to be more responsive to complex feature sets (Thomas et al. 2001), although the exact nature of the response properties remains unknown.

Inhibitory neurons in IT were also found to be selective to specific stimuli. In contrast to

sensory cortical areas, however, they were found to be more likely to connect with neurons with dissimilar overall stimulus preferences (Tamura et al. 2004). The same research also established that stimulus preferences are more highly correlated between adjacent columns than distant ones, indicating small clusters of preferences for similar, although not necessarily identical, input features.

In a study into responses of IT neurons to faces (Hirabayashi and Miyashita 2005), cell pairs responsive to faces exhibited a higher level of firing synchrony to correctly aligned face objects than to cell pairs responsive to jumbled face objects. These responses emerged around 300ms after simulus onset, a rapid enough response to be involved in stimulus recognition, in which time firing rates did not reveal responses consistent with information about feature configurations. This indicates a discriminability for feature configurations through correlated activity. The same discriminability was not present within firing rates. The role of the time-course of responses is also emphasised in Akrami et al. (2009), in which it is reported that the early response to a stimulus produces a linear representation of stimulus properties, and the latter part represents a categorical response. Tamura and Tanaka (2001) also showed an effect of temporal order on stimulus selectivity, demonstrating that the late portion of the response 240ms after stimulus onset, was more selective for stimulus than the early part.

Kiani et al. (2007) demonstrated hierarchical category structure in firing patterns of IT cortex. Through the use of a very large stimulus set they show responses to stimuli from a broad range of categories that contains a hierarchical structure intuitively to a human observer. The population response to these stimuli was shown to reflect this structure. Responses to animate and inanimate objects were in general negatively correlated (Pearson's r value of up to -0.31), and responses to pairs of stimuli from the same intuitive category were generally positively correlated (Pearson's r value of up to 0.54). The hierarchical structure created from the distribution of population responses could not be generated from simple low-level visual properties of the stimuli, suggesting that the categorisation relies instead on abstract features. Further evidence for a hierarchical population response is demonstrated by (Kriegeskorte et al. 2008), where similar category trees were found in both monkey and human IT.

2.6 Summary

This chapter has introduced some of the biological phenomena that generate neural activity. An introduction of neuron structure, membrane potentials, action potentials and synaptic transmission was provided in section 2.1. Section 2.2 briefly described some of the biological properties underlying the changing of synaptic efficacy. Several prominent features found throughout the cortex were discussed in section 2.3. These are connectivity properties, oscillatory activity and temporal representation. In section 2.4, evidence was provided for the ubiquitous nature of map-like response properties throughout sensory cortex. Finally, section 2.5 briefly introduced some of the response properties of neurons in higher order cortical areas.

The elements introduced in this chapter are the main inspiration behind the computational modelling approaches described in chapter 3, which provides an introduction to the computational neuroscience methods used later in the thesis. The biological phenomena described here also represent the inspiration behind the self-organising map model, discussed in chapter 4. Additionally, they play an important role in the development of the computational models described in chapters 5 and 6.

Chapter 3

Neuron Models

Artificial neurons are individual computational elements based on the properties of biological neurons. These properties can be computationally modelled at a wide variety of levels of abstraction. Detailed models incorporate models of molecular concentrations and the properties of individual ion channels. Abstract models use the principle of computing with a network of simple units without attempting to explicitly model particular biological features. Different amounts of realism are relevant for different modelling purposes, and it is not always the case that a more detailed model will generate better answers (Ibarz et al. 2011). Shortcomings of adding detail include: more detailed models require more parameters, which need to be accurately mapped from recordings, which is not always possible; a fully detailed model will be exactly as hard to understand as an actual neuron; and extra detail dramatically increases computational requirements. It is necessary to gauge the required level of abstraction such that the functional properties relevant to a particular goal are retained and unnecessary detail is removed.

Models from several levels of abstraction are relevant to the current work. Generally speaking, artificial intelligence approaches to cognitive mechanisms involve traditional approaches to modelling of individual neurons, which use less biologically accurate models. Spiking neuron models attempt to replicate the sub-threshold membrane dynamics, action potentials, and post-synaptic potentials found in biological neurons, so are less abstract than traditional approaches. A primary aim of the current work is to explore the implementation of generic cognitive mechanisms using these more realistic model neurons.

All of the models described here use the same basic structure. At each time step, they receive a vector of input values, each of which is multiplied by a weight vector (representing synaptic

strength). This input is applied, through some function, to an internal variable, and this internal variable is converted, through a transfer function, into an output value. These stages map approximately to the four functional components mentioned in section 2.1. The degree to which each stage of this process is able to model observed measurements in real neurons indicates the level of abstraction from biology that a model is at.

This chapter will describe several neuron models that use different computational methods to represent the various components of biological neurons. First, common implementations of the membrane potential, action potential and post-synaptic potential will be introduced. Then the modelling of adaptation and learning in artificial neurons is discussed.

3.1 Traditional neuron models

Among the simplest, most abstract artificial neuron models is the threshold neuron, introduced by McCulloch and Pitts (1943). In this model the input component is an input vector of continuous or discrete values, the integrative component is a transfer function applied to a weighted sum of those inputs, the trigger component is a threshold for that function, and the output component is a binary output set to 1, representing the neuron becoming 'active', if a threshold is met, and 0 if the amount of input is insufficient to reach the threshold. Variants of this model use an alternative transfer function, either non-linear, such as the sigmoid function, or linear, to produce real-valued output. In these cases the output value is no longer either active or non-active, but graded, representing an 'activity level'. Output values are analogous to firing rates in biological neurons. The binary case can be viewed as representing a neuron that either fires spikes at a low or high rate (0 or 1), while other transfer functions represent other mappings of input current to firing rate. This type of neuron model is shown in figure 3.1.

Another type of neuron used in artificial intelligence algorithms is the type used in the self-organising map (SOM) algorithm (Kohonen 2001), sometimes called a 'distance neuron'. In this case, the transfer function involves a comparison of the weights to current input values. A closer relationship between the two, commonly measured by Euclidean distance, results in a greater activity level. In the case of the SOM, the output component is represented through a winner-takes-all type of competition, in which the neuron with the highest activity level is

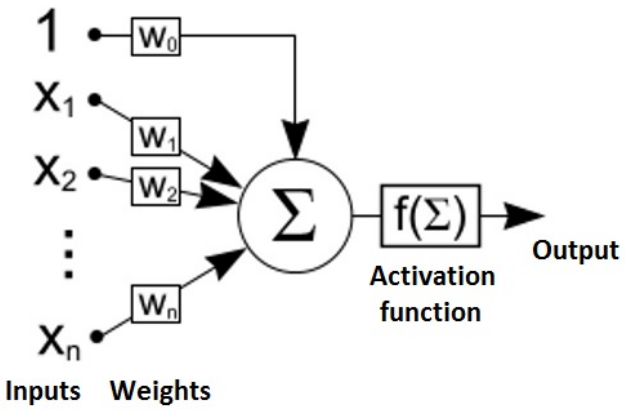


Figure 3.1: Illustration of the McCulloch-Pitts neuron model. The input values are multiplied by the weights, summed, and an activation function, which can be a threshold as in the original McCulloch-Pitts (1943) formulation, or a linear or non-linear function, to produce an either binary or continuous output value respectively.

labelled as the winning unit. Distance neurons could, however be used in other neural networks, with the input and integrative components determined by the distance to activity level transfer, and the trigger and output components acting as in the threshold neuron.

Artificial neurons at this level of abstraction are inspired by some of the properties of biological neurons without attempting to accurately represent them. They are suitable for models that make use of simple processing units in a network, but a more detailed neuron model is needed to capture some of the functional properties of biological neural networks. Sub-threshold membrane dynamics, action potentials, PSPs and temporal coding of information are functional properties not available to traditional neuron models.

3.2 Spiking neuron models

Spiking neurons are model neurons that communicate through action potentials. The primary means of coding information becomes the timing of spikes. They model the dynamics of the membrane potential through differential equations, and output becomes a temporally restricted pulse rather than a value analogous to a rate or activation level. Different spiking neuron models cover different levels of biological abstraction, with the level determined by the complexity and

accuracy of dynamics for the membrane, threshold, action potential and PSPs.

3.2.1 Conductance-based models

Hodgkin and Huxley (1952) recorded electrical pulses along an axon of a neuron. From these recordings they produced a set of equations governing the dynamics of the membrane capacitance, as well as conductance parameters for the various individual ion channels. This model has been shown to be capable of accurately reproducing a variety of biophysical variations in the membrane potential, such as the large rise resulting in an action potential, and the absolute and relative refractory periods. Parameters in this model may be given physiological meaning and assigned using measured values (Standage and Trappenberg 2005). Despite a high level of accuracy, the computational complexity of modelling an individual neuron using multiple simultaneous differential equations prevents it from being applicable in many neural network modelling situations. Other complex conductance-based models include modified versions of the Hodgkin-Huxley model, such as the Wang-Buzsáki model (Wang and Buzsáki 1996), which adapts parameters to produce behaviour replicating activity seen in hippocampal inhibitory interneurons, as well as more complex models incorporating the effects of additional ion channels (Wilson 1999).

3.2.2 Integrate and fire models

Integrate and fire (IF) neuron models were first postulated by Lapicque in 1907 (see Brunel and van Rossum (2007*b*) for an English translation, and see also (Abbott 1999; Brunel and van Rossum 2007*a*)). These models simplify the modelling of the membrane dynamics by reducing the sub-threshold membrane potential calculation to a single linear differential equation

$$\begin{aligned} \tau_m \frac{dV}{dt} &= I(t), \\ \text{if } V >= \theta \text{ then spike and reset } V &= V_R \end{aligned} \tag{3.1}$$

where V is the membrane potential, τ_m is the membrane time constant and $I(t)$ is input current

from pre-synaptic neurons at time t. Input is integrated into the membrane potential over time, eventually causing the potential to cross a threshold, θ, at which point an action potential is generated and the membrane potential is reset to the resting potential, V_R. The action potential is not described explicitly, but is represented by an instantaneous value of 1, indicating presence of a spike at that particular time. A common variant on this model is the leaky integrate and fire model (LIF), which adds a leaking term to equation (3.1), such that the membrane potential will leak towards its resting value over time:

$$\tau_m \frac{dV}{dt} = I(t) - g_L\,(V - V_L),$$
$$\text{if } V >= \theta \text{ then spike and reset } V = V_R \tag{3.2}$$

Each neuron has a membrane potential V that increases by integrating current input $I(t)$, and 'leaks' towards a resting potential of V_L when there is no input arriving from its afferent synapses. τ_m represents the time constant of change in V and g_L represents leak conductance. If the membrane potential is greater than a firing threshold, θ, then an action potential is generated and the membrane potential is reset to the reset potential, V_R. The LIF model exhibits spiking behaviour that mirrors the behaviour of several categories of biological neurons: "tonic spiking", "class 1 excitable", and "integrator" (Izhikevich 2004).

An extension to the LIF model that allows for a greater variety of spiking behaviour is the exponential integrate and fire (EIF) model (Fourcaud-Trocmé et al. 2003), which is an LIF variant, incorporating a non-linear, exponential component for intrinsic spike generation as the membrane potential approaches the threshold. Using an intrinsic spike generation mechanism allows for parameters to adjust the behaviour of the membrane close to the threshold, which is what makes possible spiking regimes that the LIF cannot enter (Izhikevich 2004; Fourcaud-Trocmé et al. 2003). The EIF model uses the equation

$$\tau_m \frac{dV}{dt} = I(t) - g_L(V - V_L) + g_L \exp\left(\frac{V-\theta}{\Delta_T}\Delta_T\right),$$
$$\text{if } V >= \theta \text{ then spike and reset } V = V_R, \tag{3.3}$$

where Δ_T is a slope factor determining the rate of change in the membrane potential as it approaches the firing threshold. The membrane potential diverges towards infinity as it crosses the threshold, intrinsically simulating an action potential.

An EIF neuron can also incorporate adaptation of the membrane potential to a train of stimuli through the addition of another current, resulting in the adaptive exponential integrate and fire (aEIF) model (Brette and Gerstner 2005), according to

$$\tau_m \frac{dV}{dt} = I(t) - g_L(V - V_L) + g_L \exp\left(\frac{V-\theta}{\Delta_T}\Delta_T\right) - g_A(t),$$
$$\text{if } V >= \theta \text{ then } \begin{cases} \text{spike and reset } V = V_R \\ g_A = g_A + A. \end{cases} \tag{3.4}$$

Here, g_A represents the level of the adaptation current, which increases by A with each action potential and decays to 0 according to

$$\tau_A \frac{dg_A}{dt} = -g_A, \tag{3.5}$$

where τ_A is the time constant of adaptation. This extension depresses the responsiveness of the membrane over time, such that firing rate slows gradually as g_A increases. The LIF model can be extended with the exponential component, the adaptation component, or both. IF models can also include a refractory period by locking the membrane potential to a fixed value for a time after firing. These additions result in a set of models that incorporate various different firing regime capabilities. Addition of the adaptation component to the regular LIF model facilitates

further types of firing behaviour, including "regular spiking" (Liu and Wang 2001). The aEIF model is capable of a wider range of firing regimes, approximating those seen in biology.

Another model with similar capabilities is the Izhikevich model (Izhikevich 2003), which is a two-dimensional model of the membrane potential, consisting of a pair of equations

$$\frac{dV}{dt} = 0.04V(t)^2 + 140 - w(t) + I(t) \tag{3.6}$$

$$\frac{dw}{dt} = a(bV(t) - w(t)) \tag{3.7}$$

with the additional after-spike resetting condition of

$$\text{if} V >= \theta \text{ then} V = c \text{ and} w = w + d \tag{3.8}$$

Both the Izhikevich and aEIF models are capable of reproducing a full range of more than twenty forms of spiking (Izhikevich 2004). This, of course, involves the addition of extra dimensions to the LIF model, reducing computational efficiency. The type of spiking that a neuron generates is largely determined by a choice of parameters set throughout a simulation. This means that if a particular spiking regime is not functionally useful or required in a given simulation then there is no real advantage to be gained in utilising a more complex model. As such, the choice of neuron model depends on the nature of the network being simulated and the purpose of the simulation.

3.2.3 Post-synaptic potentials

The output of any IF model is an action potential, or spike. The effect of this afferent spike on the membrane potential of efferent neurons is simulated by modelling a PSP. The simplest form of PSP model is a δ-function, transmitting an input current at one simulation time step, when the afferent neuron spikes. The effect on efferent neurons will be a jump in membrane potential at the time the spike arrives.

To better capture the temporal properties of biological PSPs, an α-function can be used (Des-

texhe et al. 2003), such as

$$P_s = \frac{P_{max} t}{\tau_s} \exp(1 - \frac{t}{\tau_s}), \tag{3.9}$$

where the presynaptic release begins at $t = 0$, the peak value of P_s occurs at $t = \tau_s$, and the function decays with a time constant τ_s (Dayan and Abbott 2000). A PSP of this nature can also be simulated with the pair of equations

$$\tau_r \frac{ds_1}{dt} = (s - s_1) \tag{3.10}$$

$$\tau_f \frac{ds_2}{dt} = (s_1 - s_2) \tag{3.11}$$

where s is a binary value representing instantaneous presence or absence of a presynaptic spike, s_1 is an internal state variable, s_2 is the α-function output, and τ_r and τ_f are time constants for the rise and fall duration of the response.

3.2.4 Integration and coincidence detection

Spiking neurons can perform different functions, with regard to analysing and making decisions about their input, depending on the choice of time constant parameters. A pertinent example is the difference between integration and coincidence detection behaviour in a LIF neuron depending on the relationship between the membrane time constant and PSP duration (Paugam-Moisy 2006). A longer PSP relative to the membrane time constant increases the likelihood that an incoming spike will contribute to a neuronal response; effectively the post-synaptic neuron is integrating the majority of its inputs into a spike response. Conversely, a shorter PSP relative to the membrane time constant decreases the likelihood that an incoming spike will contribute to a neuronal response. The post-synaptic neuron is now acting as a coincidence detector, as multiple spikes in a short temporal span will be required for the post-synaptic neuron to reach threshold before the current from a given pre-synaptic spike leaks away.

3.3 Learning

Learning in biological networks occurs when the ability of one neuron to affect another is altered. In artificial neural networks, this is commonly accomplished through the imitation of short or long term plasticity (see section 2.2). These phenomena, sensitive to shifts in molecular concentration and neurotransmitter availability, can also be abstracted to various degrees to produce analogous outcomes.

3.3.1 Short term plasticity

Short term plasticity, in the forms of facilitation and depression, alters the efficacy of a synapse by an amount that depends on the recent history of pre-synaptic action potentials at a synapse (see section 2.2.1). This can be achieved by incorporating factors for facilitation (F) and depression (D) into the postsynaptic response amplitude (A), via

$$A = P_s F D, \tag{3.12}$$

with P_s the instantaneous α-function output from equation (3.9) (or s_2 from equation (3.11)) (Abbott and Nelson 2003). The dynamics of F and D are determined by equation pairs 3.13 and 3.14, and 3.15 and 3.16 respectively:

$$\tau_F \frac{dF}{dt} = 1 - F, \tag{3.13}$$

$$\text{if presynaptic spike, then } F = F + f, \tag{3.14}$$

$$\tau_D \frac{dD}{dt} = 1 - D, \tag{3.15}$$

$$\text{if presynaptic spike, then } D = dD. \tag{3.16}$$

The time constants τ_F and τ_D represent recovery rates of the facilitation and depression processes, respectively, and the parameters f (where $f > 0$) and d (where $d < 1$) represent the onset rate of facilitation and depression, respectively. This approach ensures that facilitation

increases synaptic efficacy and depression reduces synaptic efficacy, while parameter selection allows synaptic preference for either regime (Abbott and Nelson 2003).

3.3.2 Long term plasticity

Information stored in an artificial neural network is encoded in the weights, or connection strengths, between units of that network. Learning in neural networks is a process of changing these weights, so that the desired type of information is represented in the output, given the input. Traditional artificial intelligence approaches to neural networks make use of a variety of learning rules for training the network. Many of these are supervised learning methods, including back-propagation methods. In this type of learning, the amount of error present in the network output relative to a teacher signal is accounted for by adjusting connection weights within the network until the error is minimised.

3.3.3 Hebbian learning

A major consideration for the current work is the use of unsupervised learning, as seen in the brain in the form of long-term synaptic plasticity. The majority of unsupervised learning rules for neural networks are extensions of the idea of Hebbian learning (Hebb 1949). Hebb proposed that simultaneous correlated activity in both members of a connected pair of neurons will strengthen the connection between the two. LTP in this type of learning relies on statistical correlations within the input to learn the constituent parts of an input pattern. The Hebbian LTP idea can be extended to include LTD, which in this case occurs in situations where the activity in a pair of neurons is uncorrelated, with one neuron firing when another is not. If a high activity level in a presynaptic neuron regularly contributes to a high level of activity in a postsynaptic neuron, then that activity is interpreted as causal and should be enhanced. If a high level of activity in either the pre or postsynaptic neuron is present independently of a high level of activity in its partner, then there is not a causal link between the two, and the connection can be disregarded.

Hebbian learning can be formulated as being dependent on either rate or spike timing. Rate dependent plasticity can be used with both non-spiking and spiking neuron models. It gener-

ally relies on a covariance of the firing rates of pre- and post-synaptic neurons to establish the Hebbian relationship. Spike timing dependent plasticity can only be used with spiking neuron models. It relies upon spike ordering in pre- and post-synaptic neurons to modify connection strengths.

3.3.4 Rate dependent plasticity

The most basic form of Hebbian learning only describes LTP, by increasing weights when activity levels in pre- and post-synaptic neurons are coincident. The incorporation of LTD in Hebbian rules for rate dependent plasticity takes inspiration from the observation that LTD dominates at low activity levels and LTP dominates at higher activity levels (Sjöström and Nelson 2002). This can be formalised as a covariance rule

$$\tau_w \frac{dw}{dt} = u(v - \theta_v), \tag{3.17}$$

where w is synaptic weight, τ_w is a time constant for weight change, u and v are pre- and post-synaptic neuron firing rates respectively, and θ_v represents a threshold on the firing rate in the postsynaptic neuron below which LTD dominates. This rule follows the observations that a low firing rate leads to LTD and a high firing rate leads to LTP (Sjöström and Nelson 2002), but does not require any postsynaptic activity to induce LTD. The Bienenstock-Cooper-Munro (BCM) rule (Bienenstock et al. 1982) modifies equation (3.17) by incorporating a dependence on the postsynaptic firing rate:

$$\tau_w \frac{dw}{dt} = vu(v - \theta_v). \tag{3.18}$$

The BCM rule was originally introduced as part of a model of development of centre/surround receptive field development and ocular dominance maps in visual cortex (Bienenstock et al. 1982).

3.3.5 Plasticity rules for spiking neurons

There are several types of learning rules that are only applicable for use with spiking neuron models. These rules rely on information relating to the relative timing of action potentials in pre- and post-synaptic neurons. Two of these are introduced here, a rule for STDP relying solely on the time differential between action potentials, and a rule introduced by Brader et al. (2007) (the Brader-Senn-Fusi (BSF) rule) that uses a decaying trace that indicates recent spiking activity, combined with membrane potential thresholds for LTP and LTD.

The computational form of STDP attempts to capture the temporal requirements for plasticity seen in biological neurons (see 2.2.2). A mathematical implementation of the STDP rule uses the following equation, with $\Delta t = t_{post} - t_{pre}$ (Song et al. 2000; Bi 2002; Izhikevich and Desai 2003; Bush et al. 2010):

$$f(\Delta t) = \begin{cases} A_+(1-\frac{1}{\tau_+})^{\Delta t} & \text{if } \Delta t > 0, \\ -A_-(1-\frac{1}{\tau_-})^{-\Delta t} & \text{if } \Delta t \leq 0. \end{cases} \tag{3.19}$$

A_+ and A_- are both positive and determine the maximum amount of synaptic strengthening and weakening that can occur, respectively. τ_+ and τ_- are time constants determining the range of time in which synaptic strengthening and weakening will occur, respectively.

A common STDP variant, known as 'multiplicative' STDP, involves an additional function to alter the magnitude of weight change by the current value of the weight. Multiplicative STDP can be formulated as:

$$f(\Delta t) = \begin{cases} F_+(w_{ij})A_+(1-\frac{1}{\tau_+})^{\Delta t} & \text{if } \Delta t > 0, \\ -F_-(w_{ij})A_-(1-\frac{1}{\tau_-})^{-\Delta t} & \text{if } \Delta t \leq 0. \end{cases} \tag{3.20}$$

Here F_+ represents a function that tends towards zero as w_{ij} increases and tends towards one as w_{ij} decreases, such that a larger weight will lead to a reduction in the weight change caused by LTP, and vice versa. Conversely, F_- represents the inverse function, tending towards one

as w_{ij} increases, and towards zero as w_{ij} decreases, such that a smaller weight will lead to a reduction in the weight change caused by LTD, and vice versa. Simple options for F_+ include $1 - w_{ij}$ and $\exp -w_{ij}$, and for F_- simply w_{ij} can be used (Bush et al. 2010). Other alternatives include $(1 - w)^{\mu}$ for F_+, and αw^{μ} for F_- (Gütig et al. 2003), where μ is the power of the weight dependence and α is a constant expressing a possible asymmetry between the scales of potentiation and depression.

The 'additive', non-weight dependent STDP rule can lead to bimodal final weight distributions (Song et al. 2000). As a weight becomes strong, it leads to an increased likelihood of a presynaptic spike resulting in a postsynaptic spike, which in turn leads to further potentiation of the weight. The opposite phenomenon also applies, with weakened weights leading to further depression. These factors mean that it is necessary to place bounds on the maximum and minimum synaptic weight values. Multiplicative STDP counteracts this by decreasing the magnitude of potentiation if a connection is already strong, and decreasing the magnitude of depression if a connection is weak (Gütig et al. 2003; Bush et al. 2010). This effect can lead to a unimodal final weight distribution (van Rossum et al. 2000). It is possible to stabilise weights at mid-points in the possible weight range, with the combination of weight dependence factor and fraction of time that a presynaptic spike is followed by a postsynaptic spike being the determining factors in the final weight. It has also been demonstrated that the average weight change found in rate-based rules such as the BCM rule (see equation (3.18)), which are based on observations of activity requirements for plasticity in cortex, can be replicated under certain conditions by uncorrelated firing using STDP (Izhikevich and Desai 2003; Bush et al. 2010).

The BSF rule modifies weights based on the current postsynaptic depolarisation along with a calcium variable $C(t)$, which is a function of postsynaptic spiking,

$$\tau_C \frac{dC(t)}{dt} = -C(t) + J_C \sum_i \delta(t - t_i), \tag{3.21}$$

where J_C is the contribution of a spike, τ_C is the time constant, and the sum is over postsynaptic spikes at time t_i. Each action potential generated by a neuron contributes an amount

proportional to J_C to the $C(t)$, which decays with the time constant τ_C towards a resting value of 0. Weight modifications are calculated upon arrival of a presynaptic spike, according to the pair of conditional equations:

$$w = w + a \text{ if } V(t_{pre}) > \theta_V \text{ and } \theta_{up}^l < C(t_{pre}) < \theta_{up}^h \tag{3.22}$$

$$w = w - b \text{ if } V(t_{pre}) \leq \theta_V \text{ and } \theta_{down}^l < C(t_{pre}) < \theta_{down}^h, \tag{3.23}$$

where a and b are jump sizes, θ_V is a voltage threshold ($\theta_V < V_\theta$) and the θ_{up}^l, θ_{up}^h, θ_{down}^l and θ_{down}^h are thresholds on the calcium variable. In the absence of a pre-synaptic spike or if the above conditions are not satisfied then w drifts towards one of two stable values

$$\frac{dw}{dt} = \alpha \text{ if } w > \theta_w \tag{3.24}$$

$$\frac{dw}{dt} = -\beta \text{ if } w \leq \theta_w, \tag{3.25}$$

where α and β are positive constants and θ_X is a threshold on the internal variable. The synaptic connection strength w is in this rule limited between w_{min} and w_{max}.

The BSF (Brader et al. 2007) rule is an example of a learning rule that uses parameters local to individual neurons to establish the direction and magnitude of weight change at a synapse. In this case the calcium variable represents the postsynaptic calcium concentration that is thought to play a crucial role in STDP. This rule is based on timing of spikes in the pre- and post-synaptic neurons, but uses internal variables to control the weight change, rather than the relative timing itself.

3.4 Summary

This chapter has introduced the fundamental computational techniques that are commonly used to recreate the properties of biological neurons introduced in sections 2.1 and 2.2. The equations provided are abstractions of complex physical processes that nevertheless capture important functionality. These methods will be used in chapters 5 and 6 as the basis for the generation of artificial neural networks that aim to reproduce some of the functional features described in sections 2.3, 2.4 and 2.5. The next chapter will introduce the artificial intelligence algorithm used as the basis for the spiking neuron model described in chapter 5, the self-organising map.

Chapter 4

Self-Organising Maps

This chapter will introduce the neural network algorithm known as the self-organising map (SOM), or Kohonen map. The SOM represents an abstraction of the frequently observed brain maps described in section 2.4. As such, the functional properties captured by the SOM can be thought of as being general functional properties relevant to large regions of the cortex. A description of the SOM algorithm and the functional implications of this will be provided in this chapter, followed by a discussion of modifications to the algorithm that change its properties in various ways. Implementations of the SOM in networks of spiking neurons will be introduced, ahead of the next chapter, which describes the refinements made to these existing implementations for the purpose of the current work.

Topologically ordered spatial representations of features can be found in various sensory cortical areas (Kaas and Catania 2002; Wandell 2011), such as ocular dominance bands (Casagrande and Kaas 1994; Horton and Hocking 1996) and orientation maps (Blasdel and Salama 1986; Hubel et al. 1978) in cat and primate primary visual cortex, auditory cortex in cats (Merzenich et al. 1975), primary taste cortex (Chen et al. 2011), the olfactory bulb (Wang et al. 1998) and the barrel cortex in rodents (Woolsey et al. 1975; Fox 2008), and the somatosensory cortex of primates (Friedman et al. 2004). Known as cortical feature maps, these regions are thought to largely develop during a critical period, the length of which varies between brain regions (Hensch 2004). Throughout development they accrue several distinctive properties, such as disruptions that reflect actual discontinuities in the sensory periphery and disproportionate representation of early-developing portions of the sensory periphery (Kaas and Catania 2002). The relationship between areas of the sensory periphery is reflected in the relationship between physical areas of cortex that are tuned to represent those areas (Feldman and Brecht 2005), and

the physical substrate of representation is capable of reorganisation in response to a change in input properties (Merzenich et al. 1983; Calford 2002).

4.1 SOM algorithm

The self-organising map (SOM) is a neural network algorithm inspired by the organisational structure of feature maps in the brain (Kohonen 2001), and can be summarised as follows:

1. The network consists of an ordered grid of output/map neurons, each of which has a set of associated 'weights', one per dimension in the input data.

2. An input pattern selected at random from the current data set is presented to the network.

3. The map neuron with weights closest to the current piece of input data is selected as the best matching unit (BMU), and 'wins' the competition.

4. The weights of the BMU and all nodes within a certain vicinity of the BMU, set by a neighbourhood function, have their weight values moved towards the values of the current piece of input data.

5. Steps 2 to 4 are repeated iteratively until a set number of training steps has been reached. If applicable, the neighbourhood size and learning rate are adjusted to enforce convergence.

This neural network makes use of the distance neuron model, mentioned in section 3.1, in which the transfer function establishes the activity level of a neuron through a distance comparison between the connection weights and the values of the actual input data. The neighbourhood function encourages neurons close to the BMU to adjust their weights in the same direction as the BMU does, a function analogous to a higher activity level in neurons close to the BMU than those further away. This is in turn analogous to a lateral connection profile consisting of strong local excitation and strong distant inhibition. Neurons near to the BMU are excited regardless of their own preference for the input pattern, and neurons far from the BMU are not activated regardless of their input preferences. Competition (to determine the BMU) occurs,

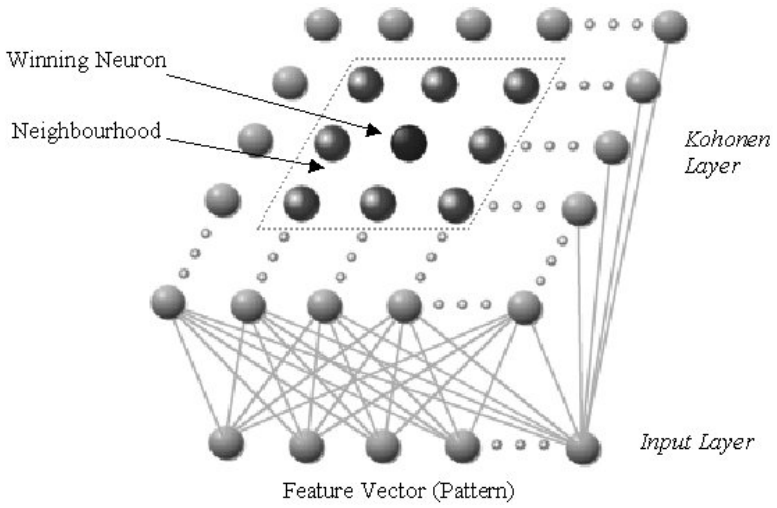


Figure 4.1: Network structure of the self organising map. Output neurons are shown as spheres, with grey lines indicating weight values between each input dimension and each output neuron. The neuron with the closest weight values to the actual input values is the winning neuron, and the neighbourhood width indicates the area around the winner within which other neuron's weights are also drawn towards the actual input values.

after which lateral connections dominate the activity profile, effectively creating a finite sized bubble of activity in the output map for each input pattern (Kohonen 2001). The general network structure is shown in figure 4.1.

Throughout learning a SOM gradually maps statistical correlations present in a set of data onto a simple, low-dimensional, topological output representation. The discovery of features and feature relations in a complex space is also a goal of principal component analysis (Hertz et al. 1991), which aims to discover the principal dimensions of variation in a dataset, finding the dominant eigenvectors of that data. The SOM performs a dimensionality reduction, with each of the dimensions in the output map representing a dimension of major variance in the higher dimensional dataset. The combination of winner-takes-all competition and a neighbourhood function for learning allows generated representations to be sparse, orthogonalised, and analogous to the representations developed by clustering algorithms (Rolls and Deco 2002). Learning in the SOM is unsupervised, making it useful in a variety of situations and easily modified to suit a variety of purposes.

The basic SOM implementation is able to represent the principal relations between input vectors in a lower dimensional space, and as such is suitable for use in a wide variety of applications. These include problems that are related to those solved in the brain by sensory processing systems, such as visual, auditory and tactile data analyses that benefit from a dimensionality reduction (Miikkulainen et al. 2004). The two-dimensional SOM can be folded over a three-dimensional object to create a surface representation by conforming to data points from that surface. In particular, this technique can be used to model the surface of brain areas (Chuang et al. 2005, 2007). The standard SOM has been successfully utilised to find solutions to a variety of categorisation problems (Foody 1999; Godin et al. 2005). Additionally, analysis of contextual and feature-based data has been conducted using the conventional algorithm (Ritter and Kohonen 1989).

4.1.1 SOM quality analysis

The quality of a SOM can be quantitatively assessed using a metric that demonstrates the extent to which the topography of the input space is preserved in the learned output map. Polani (2002) suggested a pair of criteria that any measure would need to fill in order to be considered useful for evaluating map quality:

1. It should quantify the process of self-organisation during training, i.e. its value should on average increase or decrease to indicate a reduction in mapping error.
2. It should measure the quality of embedding of the input space into the output manifold, i.e. an obviously better representation should be indicated by a "better" value.

Without a quantification method, SOM quality can be assessed through visual inspection, and an obviously better representation would be one that can be seen through this method. For example, a one-dimensional map snaking across a two-dimensional space is a poorer representation of an evenly distributed two-dimensional input data space than a two-dimensional map with a regular, rectangular organisation in that input space.

Various SOM quality measures exist that fulfil these requirements, each taking different quantities from the SOM into account. The choice of quality measure depends on the appropriateness

of the measure to a given situation. In the following equations, N is the number of input patterns, I_i is an actual data vector for input pattern i, and w_i is the BMU for an input pattern i.

Quantisation error: quantisation error (QE) is simply the average distance from all possible input vectors to the output unit that most accurately represents them, taken as Euclidean distance between the input vector and the weight vector for that distance neuron (Beaton et al. 2010):

$$QE = \sum_{i=1}^{N} d(I_i, w_i)/N, \tag{4.1}$$

where $d(I_i, w_i)$ represents the distance between the actual input vector I_i and the weight vector for the BMU for pattern i, w_i. This measure fluctuates depending on the number of output neurons in such a way that a larger output map is always favoured by the metric.

Topographic error: topographic error (TE) is an error value based on retrieving the best and second best matching units for an input pattern. If they are not neighbours in the output grid, an error value is incremented (Beaton et al. 2010):

$$QE = \frac{1}{N} \sum_{i=1}^{N} \begin{cases} 0, & \text{if } w_i \text{ and } w2_i \text{ are adjacent.} \\ 1, & \text{otherwise.} \end{cases}, \tag{4.2}$$

where $w2_i$ is the second best matching unit for input pattern i. This measure assesses the amount of tangling in an output map, but lacks any measure of the actual mapping of inputs to outputs.

Metric multidimensional scaling: metric multidimensional scaling (MDS) can be used to assess the ability of a SOM to map an input data set can be assessed by checking the topographic mapping error of the output map given the input data. For a map with no mapping error, the relative distance between any pair of data in the input space is the same as the relative distance between the locations activated by that pair of input data in the output map.

$$E_{MDS} = \sum_{i=1}^{N} \sum_{j<i} (F(i,j) - G(M(i), M(j)))^2 \tag{4.3}$$

where N is the number of input patterns, $F(i, j)$ represents the actual dissimilarity of the pair of input patterns i and j (measured as Euclidean distance), and $G(M(i), M(j))$ represents the dissimilarity between the locations in the output map representing patterns i and j (measured as Euclidean distance), where $M(i)$ and $M(j)$ are the locations of the winning nodes in layer v for input patterns i and j respectively Goodhill and Sejnowski (1997). The value of E_{MDS} represents how well the final network mapping preserves the topology of the input data set.

The QE, TE and MDS measures cover three fundamental types of SOM quality measure. QE analyses the relationship between individual inputs and the output nodes representing them; TE analyses whether the output representation has managed to untangle itself from an initial random organisation; and MDS reveals the quality of the representation of the input space by the output space. Among other prominent measures of SOM quality are: the topographic product (Bauer and Pawelzik 1992), which measures neighbourhood preservation between input and output spaces and can assess the appropriateness of the dimensionality of the map (which will not be relevant to the spiking neuron SOM implementation described in section 5 as, being inspired by cortical maps, it will not vary from a two-dimensional structure); the topographic function (Villmann et al. 1997), which assesses the neighbourhood relations between the receptive fields of the output neurons; and the CQoCO measure (Beaton et al. 2010), which effectively combines quantization error and topographic error with a third measure, indicating the extent to which the output map represents vectors that are not part of the actual input space, into a single value encapsulating all three of these SOM features.

4.1.2 SOM variants

For the tasks described above, as well as many other specific applications, the SOM algorithm does not require modification to act as a valuable analytical tool. However, many modifications to the algorithm have been made, to adapt the SOM for particular applications (Kohonen et al. 1996), to make the algorithm more efficient or even more general (Keith-Magee 2001), or to increase the relevance of the results to particular areas, for example to increase the biological plausibility of the algorithm (Miikkulainen et al. 2004). Common sources of modification include the specification of learning rate and neighbourhood function, dynamic modification

of network properties, hardware implementations, and changes to the neuron model. Some approaches to these areas will be outlined below.

4.1.3 Decay schemes

In the conventional SOM the learning rate and neighbourhood function follow decay schemes throughout training. Kohonen (2001) suggests that a convex and wide neighbourhood function will generate an initial global mapping and after this stage the neighbourhood can be reduced to increase the accuracy of the final map. A comparable reduction in the learning rate throughout training helps to avoid local minima convergence in earlier stages of learning through an initially large learning rate, while maintaining the map structure later in training when finer changes are made. Selection of these decay schemes has been investigated in several SOM variants. Keith-Magee (2001) investigated optimal neighbourhood reduction for the condition of ongoing learning, where learning rate is fixed throughout training. Experiments showed that a useful and versatile function for establishing neighbourhood decay is achieved with the following equations:

$$\bar{w}_{ij} = (1+a)G(||i-j||,r) - aG(||i-j||,br), \tag{4.4}$$

in which the lateral connection strength $\bar{w}_{ij}$ between output neurons at locations i and j in the grid is determined by (4.4), where a represents the magnitude of the negative component of the function, b determines the decay of the negative component of the function, r determines the radius of the positive component of the function, and the function G is a Gaussian function of the distance between i and j.

The traditional SOM formulation includes a decaying neighbourhood width over time to produce a more finely tuned output mapping. An appropriate decay function for the width (r) of this lateral connection kernel is established, through a series of experiments (Keith-Magee 2001), as a step function with a filter to smooth the step function over time:

$$r(t) = X - \frac{X - X'}{1 + (\sqrt{2} - 1)((T/t)^{2n})}. \tag{4.5}$$

where $r(t)$ gives the value of r to use in (4.4) at training step t in the simulation, X and X' are values of r at the start and end of training respectively, T represents the value of t that the 'step' is centred around, and n is the order, or amount of smoothing, of the 'smoothed step' function. Compared with classic linear decay schemes and a non-smoothed step decay, use of (4.5) results in an accurate output mapping being reached more quickly Keith-Magee (2001). Additionally, identical values for X and X' results in no neighbourhood decay, facilitating a simple transition between regimes.

Other work has investigated time invariant learning rate schemes (Cho and Seok 1998; Mulier and Cherkassky 1994; Haese 1998; Berglund and Sitte 2006). The use of a constant learning rate, combined with a binary term increasing the amount of learning probabilistically for the winning neuron in a training step can produce results comparable with the original SOM formulation (Cho and Seok 1998). The boost of learning for the BMU helps to encourage the fine-grained differentiation of input preference that is seen in the case of decaying learning rates in the original SOM formulation. Learning rate can also be based on the distance from the input to weight vector (Berglund and Sitte 2006), in which case a greater difference represents a worse fitting of the map to the current piece of input data, which leads to selection of a higher learning rate to compensate. Additionally, it has been suggested that the optimal learning rate scheme is one in which every training step contributes an equal amount to the final map configuration (Mulier and Cherkassky 1994). Using a linear learning rate decay, the later training steps contribute a much greater amount to the final map than earlier training steps. Calculating the expected contribution of a particular training step to the final map, including the contribution of the current neighbourhood state to this variable, allows for less manipulation of the random selection order of input data on final map configuration. Another approach is to vary learning rate individually for each output neuron (Haese 1998). This approach requires that the utility of a neuron's current weight arrangement is measured at each training step, and the learning rate for each neuron set accordingly, with better representatives of the input data requiring a lower

learning rate, effectively being frozen in place.

4.2 Alternate SOM structures

The standard two-layer, feedforward structure of the SOM allows for topographic organisation to occur. Alterations can be made to this structure to make the SOM suitable for alternate implementations. Three general types of these alternatives are particularly relevant to the current work: hardware implementations, biologically inspired implementations, and spiking neuron implementations. Prominent examples of these are introduced in this section.

4.2.1 Hardware SOMs

Some research has investigated the possibility of implementing the SOM algorithm in modern hardware. The digital phase-locked loop (DPLL) has been used as a method of implementing the SOM on field programmable gate array (FPGA) hardware (Hikawa 2005). This approach takes advantage of the properties of a DPLL, which is able to compare the phase of a variable frequency oscillator output signal with an input reference signal through a phase detector, and control the frequency of the oscillator through the output from the phase detector. Hikawa (2005) show that a network of these is able to function as a SOM when combined with an update pulse signal acting as a neighbourhood function, by sending pulses that allow the frequency of an oscillator to change. These pulses are sent more frequently for the winning node, allowing that node to adjust its frequency by a larger amount, and are sent with decreasing frequency for other nodes depending on distance from the winner. The network organisation into a two-dimensional input pattern is demonstrated.

4.2.2 Biologically inspired SOMs

The original SOM algorithm is based on feature maps found in the cortex, although significant modifications to the initial formulation are required to reproduce the properties of brain maps in a SOM. The visual cortex in particular is a common candidate for representation using a SOM. The two-dimensional arrangement of the retinotopic map in visual cortex is simply represented in the standard two-dimensional SOM, the folding of which to encompass a three-dimensional input space produces patterns strikingly similar to the layout of ocular dominance columns in

the visual cortex (Miikkulainen et al. 2004). While the functional properties of cortical maps can be approximated by the output of the traditional SOM, the nature of input and learning in biological systems appears markedly different. As such a variety of modifications to the algorithm attempt to increase the biological plausibility of the network.

The LISSOM model (see Miikkulainen et al. (2004) for review) recreates the full range of overlapping feature representations found in the visual cortex through several modifications to the SOM idea. The SOM is related to the structure of the visual system in this model through the identification of each output neuron as a cortical column, representing the six vertical layers in cortex, and each input neuron as a patch of retinal ganglion cells. The neurons are activity-based, rather than distance-based. Activity levels in the input layer are transmitted to the output layer via a limited width receptive field of excitatory connections. Lateral connections in the output layer effectively recreate a neighbourhood function by being excitatory at short range and inhibitory at long range. This effect is in keeping with the overall lateral profile seen in visual cortex when using a high contrast stimulus (Grinvald et al. 1994) (see section 2.4). Output neuron activity levels combine both feedforward and lateral connections, and a Hebbian learning mechanism trains both types of connection. Adaptation of the lateral weights represents the most marked departure from the traditional SOM algorithm, resulting in patchy connections at long range between neurons that have developed similar input pattern preferences (Bednar and Miikkulainen 2006). The long range connections are inhibitory, and therefore act as a form of competition between regions that are likely to be responsive to similar inputs, resulting in a form of activity level modulation between areas excited by an input pattern. Initially this model produced maps with preference for stimulus orientation. Developments to the model include an additional retina layer to learn ocular dominance patterns, an LGN layer with on and off centre response types and a range of delays along with moving visual stimuli, to learn direction preferences. These modifications encourage mapping of the set of complex overlapping properties present in biology (Bednar and Miikkulainen 2006) (see figure 4.2).

LISSOM has also been used as the basis for the development of a model of direction preference development in the barrel cortex (Wilson et al. 2010). In this example the input map, repre-

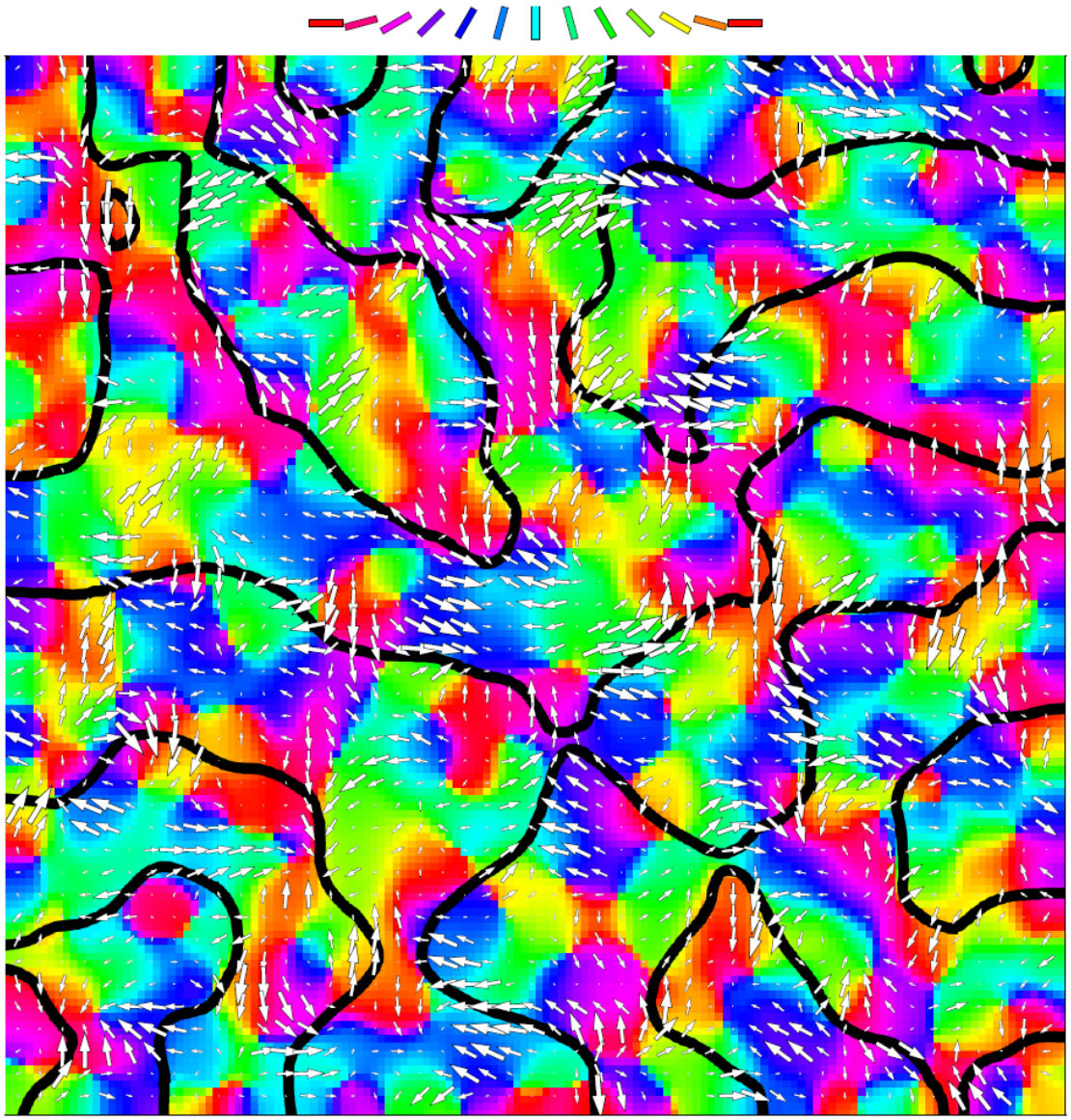


Figure 4.2: LISSOM model output response preferences. Each square in the figure represents a neuron in the output layer of a trained LISSOM model. The colour scale represents orientation preference, the thick black lines represent divisions between ocular dominance regions, and the arrows represent direction of movement preference, with the arrow direction showing the preferred direction and the size of the arrow representing the preference magnitude. When trained using oriented, moving Gaussian patterns with different brightnesses in each eye, LISSOM develops overlaid orientation, ocular dominance, and direction maps simultaneously. Reproduced from Miikkulainen et al. (2004).

senting laminar layer 4 of the rodent somatosensory cortex, is divided into a grid representing 'barrels', with each neuron within a barrel only receiving input from one whisker, and responding preferentially to a randomly assigned direction of whisker deflection. In the output layer, representing layer 2 or 3 of the rodent somatosensory cortex, the grid structure remains, with receptive fields ensuring that feedforward activity propagates only within the barrel. Lateral connections are again distance dependent, with local excitation and distant inhibition. Input consists of the input barrels being excited as if some of the whiskers had been deflected in similar directions. This activity then propagates and establishes a pattern in the output layer, and Hebbian learning occurs, as in the LISSOM model. A somatotopically aligned map of whisker deflection direction is formed, with the full range of directions represented in each barrel, and neurons in approximately the same location within each barrel having the same direction preference. Again, lateral connections form primarily to neurons in other regions of the map (in this case this means neurons connected to other barrels, rather than neurons connected to other parts of the visual field) that have similar input preference.

A similar two-layer barrel cortex model was described in (Kremer et al. 2011). The input layer, again representing layer 4, is functionally the same in this model, except that the neurons have a preferred direction and respond with a firing rate according to a Gaussian tuning curve around that direction. Layer 2 or 3 in this case is different, with each barrel in layer 2 or 3 being composed of IF neurons in an approximately 4:1 ratio of excitatory to inhibitory. The neuron model includes adaptation, a refractory period, β-function PSPs (difference of two exponential functions), and sparse synaptic connections based on barrel cortex connectivity profiles (Lübke and Feldmayer 2007). Lateral connectivity profiles were Gaussian functions of distance in number of barrels, with excitatory connections having a standard deviation of 0.4 barrels and a maximum connections probability (at zero distance) of 15%, to both other excitatory and inhibitory neurons, but with connections to inhibitory neurons three times stronger. Inhibitory connections were only made to excitatory neurons, with a standard deviation of 0.2 barrels and a maximum connection probability (at zero distance) of 100%. Weight adaptation was performed through STDP, as in Song et al. (2000). The lateral connection profile in this case leads to a slightly different map formation than that seen in the LISSOM model, with the dense

local inhibition causing adjacent excitatory neurons to develop differing direction preferences. However, this model uses approximately three times as many neurons as the LISSOM-based model above, and when direction preferences are averaged across an area the maps appear qualitatively similar.

The models described in this section use the principles of the SOM algorithm to learn functional properties of cortical regions. These models make use of imitation sensory input and alter the input layer of the SOM such that the properties of the sensory apparatus are captured. The changes made in the LISSOM model, switching from distance neurons to activity-based neurons and the incorporation of a lateral synaptic connection profile over a neighbourhood function, also present alternate methods of implementing properties equivalent to those in the SOM in a more realistic way. Additionally, the barrel cortex model introduces an output layer of spiking neurons, and a large increase in network size that allows more accurate connection properties to generate cortical map structure. The next section (section 4.2.3) will examine attempts to replicate the functional properties of the traditional self-organising map using networks of spiking neurons.

4.2.3 Spiking SOMs

Ruf and Schmitt (Ruf and Schmitt 1998) proposed a self-organising map implemented with IF neurons, using synaptic plasticity based on spike-timing. This network consists of an input layer, u, which encodes an input pattern, with fully interconnected feedforward connections to a second layer, v, which has fully interconnected lateral connections. The lateral connection profile is a monotonically decreasing function, with local excitation graduating to distant inhibition.

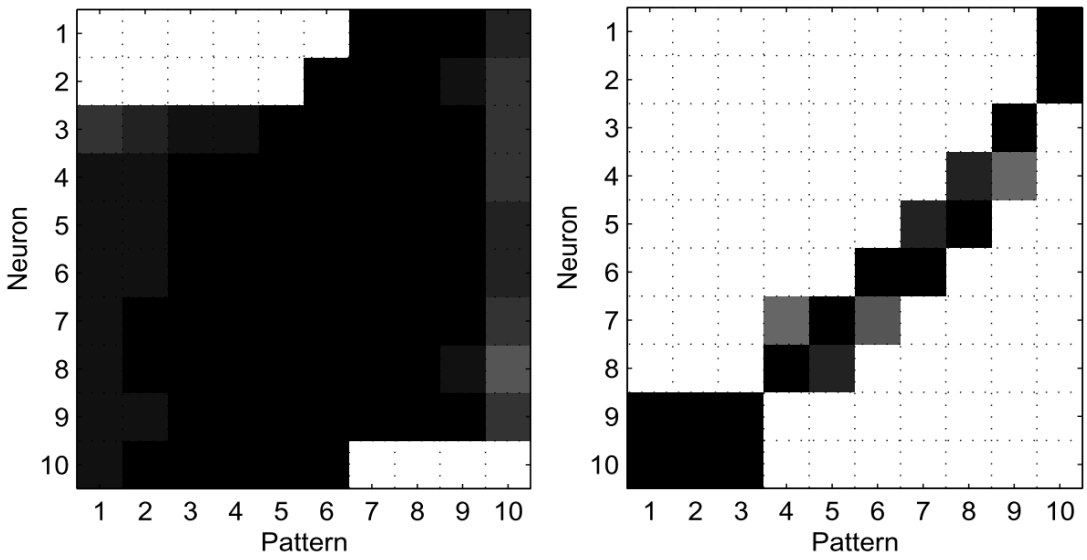
Initially an input pattern s is selected and used to drive the input layer u neurons. Layer u neurons are tuned to prefer regular incremental locations, uniformly distributed from within the range of values that a dimension in s can take. Each neuron v_j in layer v receives feedforward inputs from this output of each neuron u_i in layer u, with weight w_{ij}, as well as lateral input from each neuron v_k, $k \neq j$, with weight $\bar{w}_{kj}$. From the presentation of an input, each v_j starts to compute $\sum_i w_{ij} s_i$, where s_i represents the instantaneous presence or absence of a spike in u_i with

the first v_j to fire representing the output neuron whose weight vector best matches the input encoding, thereby becoming the best matching unit. A competitive learning rule is applied to this winning unit to move each w_{ij} towards the input pattern, given by

$$\Delta w_{ij} = \eta \frac{T_{out} - t_j}{T_{out}}(s_i - w_{ij}), \tag{4.6}$$

where η is the learning rate, t_j is the firing time of a neuron j, and T_{out} is a reference time window for learning to occur. The firing time of the winning neuron is not affected by the lateral connections in v. In the self-organising map algorithm, a neighbourhood function is used to graduate the amount of learning occurring in other competitive neurons based on their distance from the BMU. Two mechanisms are employed in this spiking version to accomplish this: the lateral connections in v alter the firing times of other neurons in v by either excitation or inhibition depending on their distance from the unit in v that has just fired; and a temporal window is placed after the firing of the first output neuron, such that learning is altered for subsequent neurons based on their firing time within that window. For the immediate neighbour of the BMU, BMU+1, $\bar{w}_{kj}$ will be relatively strongly excitatory, so the firing time of BMU+1 will move closer to the firing time of the BMU. In this way the combination of lateral connections and $\frac{T_{out}-t_j}{T_{out}}$ represent the neighbourhood function, as BMU+1 will learn an input pattern more similar to that learnt by the BMU than it would have otherwise, due to its more rapid firing time, and it will learn that input pattern more strongly as it fires sooner within the T_{out} window. A remote neuron from the BMU will be inhibited by lateral connections, so will fire later within the T_{out} window, so will learn less than it would have without the lateral interaction. In keeping with the normal SOM algorithm, the neighbourhood function and learning rate are both reduced throughout training; for the neighbourhood function in this case the lateral connections are all lowered at every training step, so that the size of the neighbourhood is reduced by a general increase in lateral inhibition.

The network was tested using one- and two-dimensional input patterns. In the one-dimensional case an input value is selected from ten inputs uniformly distributed in the range [0,1], and in the two-dimensional case the value for each input dimension was chosen randomly from the same



(a) (b)

Figure 4.3: Output firing times for testing of the Ruf and Schmitt (1998) network using one dimensional input, where a darker square indicates an earlier firing time, and a white square indicates no firing. (a) shows networ firing prior to learning; (b) shows the organised network, with a progression through input patterns represented in the network output by an equivalent progression through the network output space. Reproduced from Ruf and Schmitt (1998).

range. The one-dimensional output layer consisted of ten competitive neurons, and the formation of a linear map was found, assessed by visual inspection of the firing times of neurons, with adjacent neurons usually reacting the most quickly to adjacent input patterns (see figure 4.3). The two-dimensional output layer consisted of a 5×5 grid of neurons, and the formation of the map was assessed using a measure for quantifying the neighbourhood preservation known as "metric multidimensional scaling" (MDS) (Goodhill and Sejnowski 1997), showing that the relative neighbourhood distortion (i.e. the difference between winning nodes for two input patterns in the output layer compared with the actual difference between those two input patterns) reduced by around 80% as a result of training.

Variants on the Ruf and Schmitt (1998) model have been implemented elsewhere. One approach replaces the learning rule with a rule that relies on a pre-existing optimal weight derivation function to select weight changes (Panchev and Wermter 2001). Instead of relying on the firing time of the BMU as a reference time for the learning function, the difference between pre- and post-synaptic firing is applied as input to a function that describes the ideal weight of

a connection for a neuron with that firing time difference. The disadvantage of this approach is that exhaustive trial experiments are required prior to learning to establish the optimal weight function for the current network setup (Panchev and Wermter 2001). Simulation results suggest that trained network functionality is comparable with that shown in (Ruf and Schmitt 1998). Another approach reports similar results to Ruf and Schmitt (Svetlik 2006), but replaces the integrate-and-fire model with the 'JASTAP' model (a variant on integrate-and-fire, with biologically relevant action potentials, which the model in Ruf and Schmitt (1998) lacks).

Other SOM-like networks have been implemented using spiking neurons. In Sala et al. (1998) a two layer SOM structure similar to that of Ruf and Schmitt (1998) made up of 'MacGregor' neurons (MacGregor 1987) is used to test a pair of Hebbian learning rules, one with learning based on current strength of pre-synaptic potentials and the other based on temporal spike correlations in input and output layers. It is demonstrated that, with an appropriate lateral connection neighbourhood, either of these learning approaches can result in output space segregation that is related to properties of the input space. However, properties of the conventional SOM such as smooth mapping of input to output space and categorisation are not demonstrated.

In Veredas et al. (Veredas et al. 2008) a three-layer feedforward network of integrate-and-fire neurons with a STDP-like long-term potentiation window is used to produce a self-organised map of orientation preference, given appropriate receptive field shape and input properties. This map does not feature lateral connections or direct competition between neurons, instead relying on those receptive fields and input properties for the self-organisation to occur. As such, it is likely that an organised output map will only result from a limited range of inputs.

The LISSOM model has been modified to incorporate spiking neuron properties (Choe and Miikkulainen 1998). The model uses LIF neurons with relative and absolute refractory periods, as well as membrane potential adaptation. Standard Hebbian learning with normalisation to implement LTD is used, that is dependent on average spiking rates rather than spike timing. The feedforward connections from input to output layer learn, then, in the same way as the standard LISSOM model. Spiking neuron properties are only exploited with regards to the output activity and lateral interactions. After the learning phase for afferent connections, with input

consisting of individual oriented bars at locations in the visual field, input consists of several oriented bars simultaneously presented to the retinal input layer. Lateral connections feature short-term plasticity (implemented by a large increase in learning rate) and adapt based again on coincidental firing rates. The short term adaptation encourages synchronous firing in regions of the output layer, each responding to a particular segment of the input pattern, implementing an interesting form of feature binding. In this case, however, the detailed temporal properties of spike trains are not utilised.

4.3 Summary

The SOM algorithm in its traditional form represents a computational method that is inspired by generic cortical function, and as such is abstracted to a large degree from the biological mechanisms that generate similar functions within the cortex. The utility and importance of this algorithm have been introduced in this chapter, as well as an introduction to the wealth of variants that have been invented, each capable of performing slightly different analytical roles. Models that explicitly attempt to capture observed biological phenomena have been discussed in more detail, and section 4.2.3 has introduced the previous research that has attempted to use more biologically plausible spiking neuron models to capture the computational properties of the SOM algorithm.

None of this previous research, however, has tackled all of the key features of cortical information processing established in section 2.3. There is no existing model using spiking neurons and a biologically relevant learning rule, with temporal coding within oscillations, capable of continuous learning and reorganisation. Furthermore, there is a notable lack of validation that spiking neuron SOM models recreate computational properties of the original SOM algorithm. The next chapter describes and analyses an original model that fully addresses all of these problems.

Chapter 5

A spiking self-organising map

5.1 Introduction

Self-organisation to the fundamental properties of input spike trains has been shown to be an important feature of cortical organisation (see section 2.4). There are several mechanical building blocks that appear to be common across cortical regions, and contribute significantly to the style of information processing present in the cortex (see section 2.3). The self-organising map represents an important artificial intelligence algorithm, a frequent tool for data analysis, and is often modified to fit a variety of uses (see section 4). The SOM algorithm has been adapted to represent biological input features, capturing the qualities of sensory cortices (see section 4.2.2), and initial attempts have been made to implement a SOM using spiking neuron methods (see section 4.2.3). However, no successful implementation exists that makes use of spiking neurons, features of cortex for processing, and is capable of representation of generic datasets, capturing the computational properties of the original SOM algorithm.

The model described in this section is a two layer network of integrate-and-fire neurons similar to the model proposed of Ruf and Schmitt (1998). In both models, firing in the first layer is a temporal sequence encoding the value in the actual input space. A neuron in the second layer responds as a result of this firing, winning the competition by virtue of firing sooner than other second layer neurons. Further firing in the second layer is influenced by lateral connections, representing the neighbourhood. Neurons close to the winning unit fire sooner due to excitatory lateral influence, and neurons further from the winning unit fire later due to inhibitory lateral influence.

The current model differs from that of Ruf and Schmitt (1998) by incorporating realistic post-

synaptic potentials, spike-timing dependent plasticity, and inhibitory control of the input layer to generate oscillatory behaviour (facilitating continuous input presentation), while allowing for continuous, on-going learning and stable neighbourhood size. The mechanisms controlling each of these aspects of the network will be detailed in this section. Parameters for the equations introduced below are listed in table 5.1, in section 5.3.

Simulations were conducted using custom made C software. A simulation timestep of $0.1ms$ was used throughout.

5.2 Methods

5.2.1 Neuron model

Leaky integrate-and-fire neurons were used for all neurons in the network presented, modelled by the equation:

$$\tau_m \frac{dV}{dt} = I(t) - V + g\eta(t),$$
$$\text{if } V >= \theta \text{ then spike and reset } V = 0. \tag{5.1}$$

Equation (5.1) is equivalent to equation (3.2) (on page 39) with the value of V_L set to 0, representing a resting membrane potential of 0, and the addition of a noise term $g\eta(t)$. Each neuron has a membrane potential V that increases by integrating current input $I(t)$, and 'leaks' towards a resting potential of 0 when there is no input arriving from its afferent synapses. The membrane potential time constant, τ_m, is set to $10ms$ for all neurons in the model. A spike is generated whenever a neuron's membrane potential reaches a firing threshold θ, which varies according to the layer to which the neuron belongs (see sec. 5.2.3 below). Neurons are also subject to a Gaussian white noise process η (mean = 0, variance = 1), which is scaled by a factor g; $g = 0$ (i.e. no noise injection) for the majority of testing, but the robustness of the network output to membrane potential noise is tested by varying g in section 5.3.3.

All connections between neurons are modelled synapses, with weight w_{ij}, bounded between 0 and w_{max} (which varies according to the location of the neurons being connected). They transmit post-synaptic potentials (PSPs) to the post-synaptic neuron whenever a spike is generated in the pre-synaptic neuron. These PSPs are modelled as α-functions, using equations (5.2) and (5.3):

$$\tau_r \frac{ds_1}{dt} = s - s_1 \tag{5.2}$$

$$\tau_f \frac{ds_2}{dt} = s_1 - s_2 \tag{5.3}$$

where s is a binary value representing instantaneous presence or absence of a presynaptic spike, s_1 is an internal state variable, s_2 is the α-function output, and τ_r and τ_f are time constants for the rise and fall duration of the response. The time course of s_2, the α-function output, as a result of a presynaptic spike is shown in figure 5.1. Time constants are set independently for each layer according to the requirements for PSP duration for that layer (see table 5.1), but the ratio of τ_r:τ_f is always set at 1:5. This ratio was determined empirically to generate PSPs with a peak at approximately $2\tau_r$, to allow for straightforward selection of time until peak PSP values for different PSP durations in the model, and that resemble the time course of biologically observed PSPs (Dayan and Abbott 2000).

Input current at time t, $I(t)$, is calculated using equation (5.4):

$$I(t) = \sum_j w_j s_{2j}(t) \tag{5.4}$$

where w_j represents the connection weight (or synaptic efficacy) between neuron j (presynaptic) and the current neuron (postsynaptic), and $s_{2j}(t)$ represents the current α-function output from neuron j.

5.2.2 Learning

The learning rule used in the spiking SOM of Ruf and Schmitt (1998) makes use of some artificial features. Neurons have access to a global time stamp, which allows the gap between the

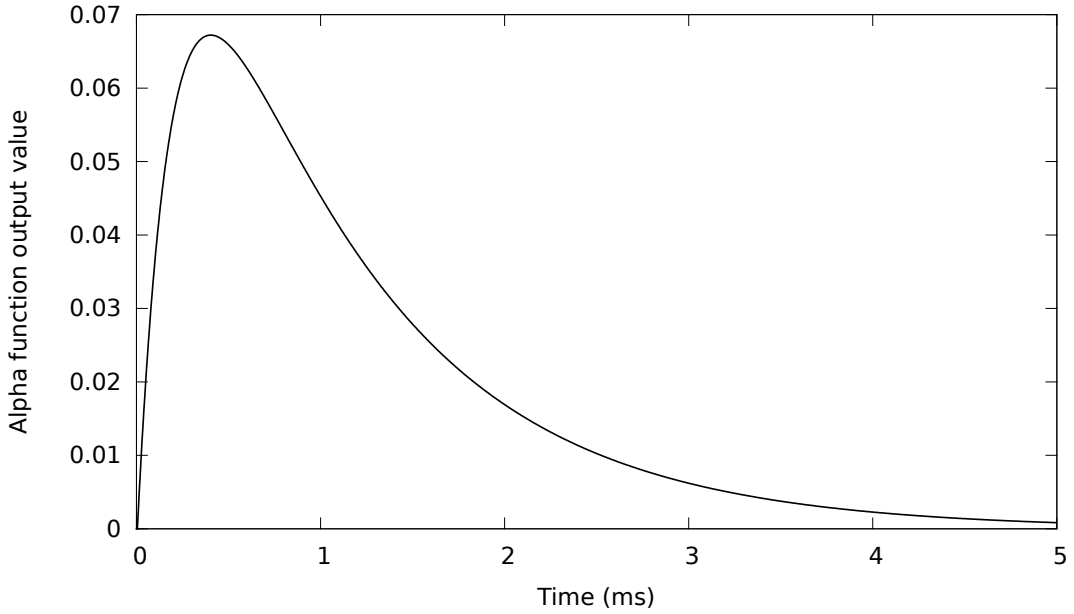


Figure 5.1: The time course of the s_2 value in equation (5.3) in response to $s = 1$ (a presynaptic spike) at time $t = 0$. The values on the vertical scale are multiplied by the synaptic weight value to determine the contribution of a synapse to the postsynaptic neuron's input current at a given time step. Time course calculated using the parameters for $u \rightarrow v$ synapses shown in table 5.1.

firing time of the best matching unit and firing time of the current neuron to be calculated. Also, the actual input value is compared with the current synaptic weight to determine the weight change, meaning that the synapse somehow has knowledge of the actual input patterns. Replacing this learning rule with a standard spike-timing dependent plasticity (STDP) rule removes these issues, providing a basis for learning that is more biologically plausible (Song and Abbott 2001) and more robust due to reliance only on local information to which each neuron already has access.

STDP (Song et al. 2000) provides a function for long-term potentiation (LTP) or depression (LTD) of synapses based on the time difference Δt, measured in ms, between a single pair of pre- and post-synaptic spikes, in neurons i and j respectively, according to $W_{ij}(t+1) = W_{ij}(t) + f(\Delta t)$. A linear multiplicative rule for LTD and exponential multiplicative rule for LTP are used, according to Bush et al. (2010):

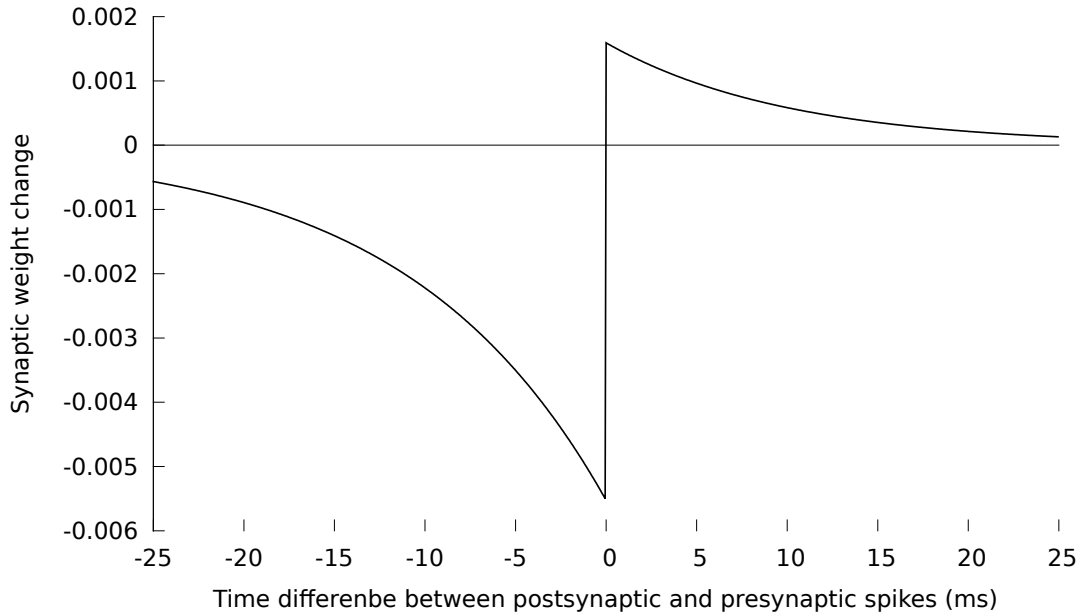


Figure 5.2: The STDP window generated from equation (5.5), with $f(\Delta t)$ plotted against Δt. The magnitudes of potentiation and depression shown are maximal; the depression window represents weight change when $w_{ij} = w_{max}$ and the potentiation window represents weight change when $w_{ij} = 0$. Time course calculated using the learning parameters for $u \to v$ synapses shown in table 5.1.

$$f(\Delta t) = \begin{cases} \exp(-w_{ij})A_+(1-\frac{1}{\tau_+})^{\Delta t} & \text{if } \Delta t > 0, \\ -w_{ij}A_-(1-\frac{1}{\tau_-})^{\Delta t} & \text{if } \Delta t \leq 0. \end{cases} \tag{5.5}$$

A_+ and A_- are both positive and determine the maximum amount of synaptic strengthening and weakening that can occur, respectively. τ_+ and τ_- are time constants determining the range of time in which synaptic strengthening and weakening will occur, respectively. The specific values these four variables are set to, along with the motivation for differing forms of the rules for LTD and LTP, are discussed in section 5.3.2, which details an extensive parameter search conducted to optimise learning. The learning window generated by equation (5.5) is shown in figure 5.2.

5.2.3 Network Structure

The spiking SOM network structure is shown schematically in figure 5.3. Conventional instantiations of the SOM receive input, in the form of numerical values for each dimension in the

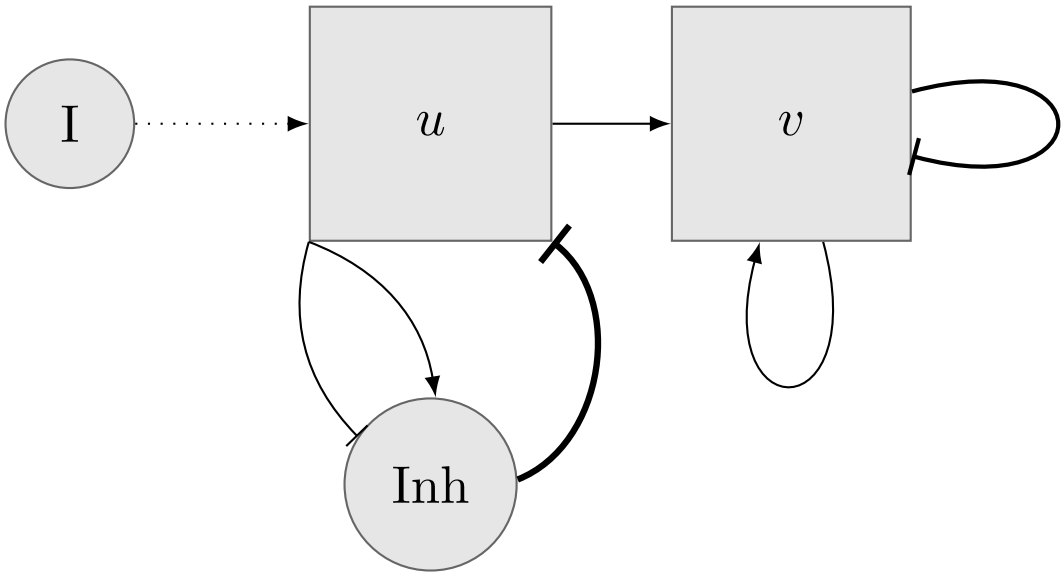


Figure 5.3: Spiking SOM network structure: actual input value I is encoded in spike times for nodes in layer u; the inhibitory unit causes oscillations, allowing continuous input to be presented to u (see section 5.2.4 for details); feedforward connections from u to v drive firing in v; early firing in v determines the location of output activity in v through lateral (neighbourhood) connections (see section 5.2.5 for details).

input data set, directly into the SOM neurons. In the spiking version these values (represented by node *I*) feed into a bank of neurons within an intermediate input layer, u. The actual input values are converted into a temporal spike sequence within each bank through the use of an inhibitory mechanism, described in section 5.2.4. This spike sequence then drives the SOM layer, v, through all-to-all feedforward synaptic connections. There are no connections from layer v back to layer u, or from layer v to the inhibitory mechanism in layer u. All-to-all lateral synaptic connections in v implement the neighbourhood function that helps to generate the self-organising process.

5.2.4 Input encoding within oscillations

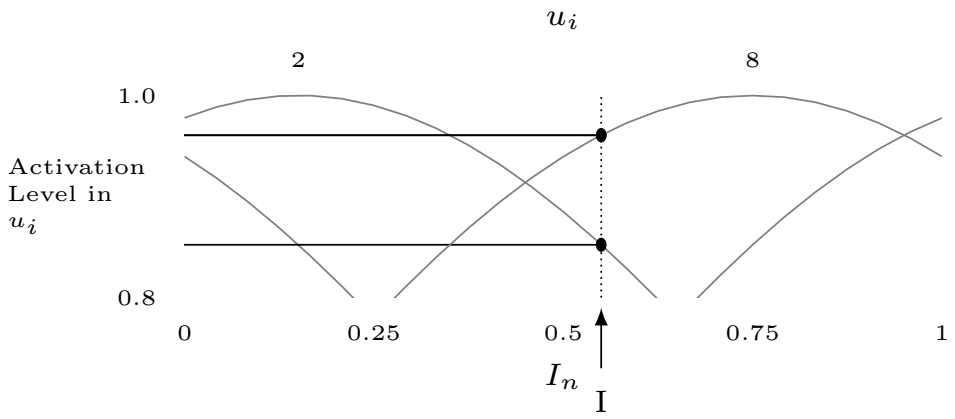
Input to this model is in the form of an m-dimensional vector of real numbers; each dimension I_n of this vector needs to be encoded in the firing of neurons in layer u. This can be achieved by representing each I_n in the actual input pattern with a bank of neurons u_n from layer u. Each neuron u_i in that bank is tuned to be centred around a point from within the range of values that I_n can take (Bohte et al. 2002). A Gaussian function is used such that the closer the actual value I_n is to the tuned value of u_i, the higher the activation to u_i, as shown in figure 5.4. The

use of integrate-and-fire neurons means that u_i with higher activation levels will reach threshold and fire faster than u_i with low activation levels. This creates a unique yet structured temporal pattern of spikes for each value that I_n can take. Using a bank of neurons to represent each input value holds a significant advantage over representing individual patterns with the firing of single neurons. The relative temporal structure of multiple spikes can represent precise, fine-grained differences between actual input values in a continuous, which would be lacking if only one neuron fired for each region of the input space. If only one neuron is stimulated by each input pattern, a single spike fires, effectively segmenting the input space into regions rather than creating a unique structure for each specific point in that input region.

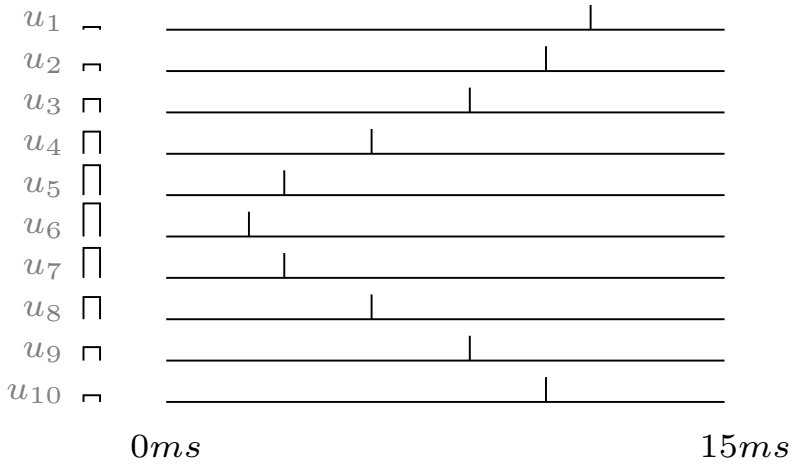
Continuous stimulus presentation is an important feature in constructing a versatile and general network, allowing network operation to be ongoing, with no need for discretisation of temporal aspects, such as automatic resetting of the network state at each training step. Using the current method of input encoding, continuous input presentation provides constant activation to layer u. This disrupts the temporal representation of each input pattern: the first u_i to fire will begin integrating input again first after resetting, and fire sooner in the next cycle of firing. This quickly desynchronises the input neurons from one another.

This problem is remedied through an inhibitory neuron, Inh_u, which responds to firing in u with a strong, slow inhibitory PSP back to all neurons in u. The inhibition depresses the membrane potential of all u_i after firing, establishing an approximate baseline for the effects of activation from I, creating a close to identical repetition of the temporal sequence. Inhibition in response to excitation creates an oscillatory behaviour, with periods of firing across the layer followed by periods of inhibition, ahead of the next batch of spikes.

The inhibitory neuron receives input from each u_i through both an excitatory and an inhibitory synapse, as shown in figure 5.5(a). The excitatory synapse has a relatively long time constant, making it slow, and the inhibitory synapse has a relatively short time constant, making it fast, as shown in figure 5.5(b). The overall effect of the pair of the connections on the membrane potential of Inh_u is an initial dip, followed by a recovery into the positive region. Combined with a resting potential fractionally below the firing threshold, the effect is that the membrane



(a) Translation of tuning curves into activation levels



(b) Translation of activation levels into spike times

Figure 5.4: The generation of a temporal sequence from a value in an input dimension and a bank of neurons tuned to different points within that dimension. (a) shows the generation of an activation value for each neuron (u_i) from within the bank of input neurons representing an actual input dimension (I_n). An example input value, at arrow I, of 0.55 is shown by the vertical dotted line. Activation values are established by the vertical point at which this dotted line crosses the tuning curve for a (u_i) neuron. Example tuning curves are shown for neurons 2 and 8, tuned to values of 0.15 and 0.75 respectively, which intersect the input value at points denoted by circles. The vertical locations of the horizontal lines from these circles then represent the activation levels that the input value generates in neurons 2 and 8. The tuning curves wrap from one edge of the input space to the other, indicating a circular tuning to the input space. (b) demonstrates the type of output firing sequence that results from this type of input encoding. The vertical height of the bars shown next to each u_i represents the activation level generated from the intersection of a given input value and the neuron's tuning curve. When a LIF neuron receives constant activation at a particular value, that activation value is translated into a spike time, with neurons receiving higher activation values generating earlier spikes than those receiving lower activation values. The time scale shows the approximate duration in which input layer activity occurs in the current model.

potential of Inh_u will stay sub-threshold as long as there are spikes in u within a reasonably short time of each other, and the membrane potential of Inh_u is guaranteed to reach threshold as soon as there is a sufficient gap in activity in u. Membrane potentials in u_i are reset when there is a large gap in firing, or when the pattern ends.

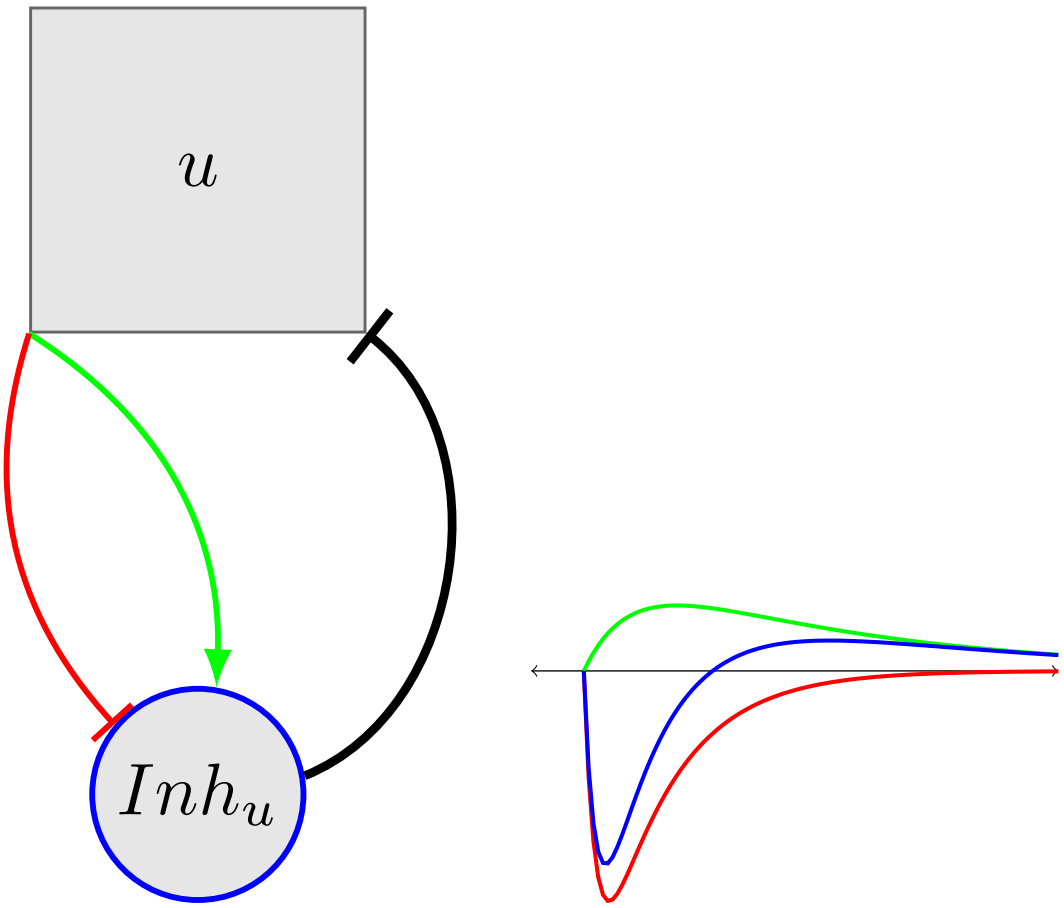
In summary, oscillations are induced in the layer through excitatory firing followed by inhibitory firing, as in a classic excitatory-inhibitory feedback loop (Wang 2010). Oscillations of this type allow input neurons to be constantly excited and maintain a reliable firing pattern for a given input. In turn, this allows for a stimulus, or input datum, to be continuously presented to the network, resulting in a versatile and reliable input coding mechanism.

5.2.5 Neighbourhood function

Self-organisation in the spiking SOM is produced through the use of a lateral interaction profile (analogous to a neighbourhood function), and STDP (see section 5.2.2). Learning in the spiking SOM occurs when an output node fires in response to the input sequence; in particular, learning of the current input values is strongest when an output neuron fires soon after the start of the input sequence, causing greater strengthening of the afferent synapses from nodes that better represent the actual input values. Lateral synaptic connections in the output layer v send excitatory signals to neurons that are within a certain distance and inhibitory signals to more distant neurons. This lateral profile encourages neurons within a spatial region to fire and discourages neurons outside of that region from firing. The effect is compounded by neurons very near to the winning neuron firing very soon after and sending similar signals laterally to a similar region of the map.

The current spiking SOM model uses equations (4.4) (on page 57) and (4.5) (on page 58) to establish lateral synaptic weights. As introduced in section 4.1.3, these equations generate a lateral connection profile similar to a Laplacian kernel, and the use of a 'step' function for decay over time can result in more accurate mappings than classic linear decay schemes (Keith-Magee 2001). Equations (4.4) and (4.5) are reproduced here for convenience:

$$\bar{w}_{jk} = (1+a)G(||i-j||,r) - aG(||i-j||,br),$$



(a) (b)

Figure 5.5: Function of the inhibitory circuit in layer u; (a) shows the inhibitory (red flat-headed line) and excitatory (green arrowed line) synapses from each node in u to Inh_u; (b) shows the time course of the post-synaptic potential from both of these synapses (red and green time courses respectively), together with the combined effect on the membrane potential of Inh_u (blue time course)

$$r(t) = X - \frac{X - X'}{1 + (\sqrt{2} - 1)((T/t)^{2n})}.$$

For a 10x10 grid of neurons in layer v, parameter searching reveals that setting a, b, X and X' all to 3.0 provides a lateral connection profile capable of topographic map formation. Identical values for X and X' lead to a constant r in equation (4.4); this was found to be capable of topographic map formation, demonstrated in section 5.3 below, although it is possible that more accurate topographic mappings can be obtained using a larger X and smaller X' (Keith-Magee 2001). This decision was made to ensure continuous learning in the output map, which has been identified as a goal of the current system. Decay of neighbourhood size throughout training was utilised for results in sections 5.3.6 and 5.3.7, however, and in these cases the step function in equation (4.5) was used to modify the lateral weights.

5.2.6 Self organisation

The inhibitory current generating an oscillation, the temporal coding of each input dimension in spike times, the neighbourhood function, and multiplicative STDP all contribute to the self-organisation of the output map. At the start of an oscillation the input neurons have depressed membrane potentials due to inhibition from the previous oscillation. Membrane potentials increase through constant input current and early spikes within an oscillation indicate neurons that represent the actual input well. This first part of the input firing pattern generates a spike in the output layer from the neuron best matching the input firing. This is followed by the firing of spatially local output neurons due to lateral activity, all before firing of neurons in the input layer that are relatively poor representatives of the actual input. STDP causes the synaptic connections from neurons in the early part of the input pattern to be strengthened, and the later part of the input pattern to be weakened, for output neurons within the neighbourhood of the winning neuron. This mechanic is illustrated schematically in figure 5.6.

A multiplicative form of STDP helps to ensure that weights reach a stable point proportional to how often an input neuron fires before an output neuron relative to how often it fires after an output neuron. Output neurons will respond for actual input values that are a distance away

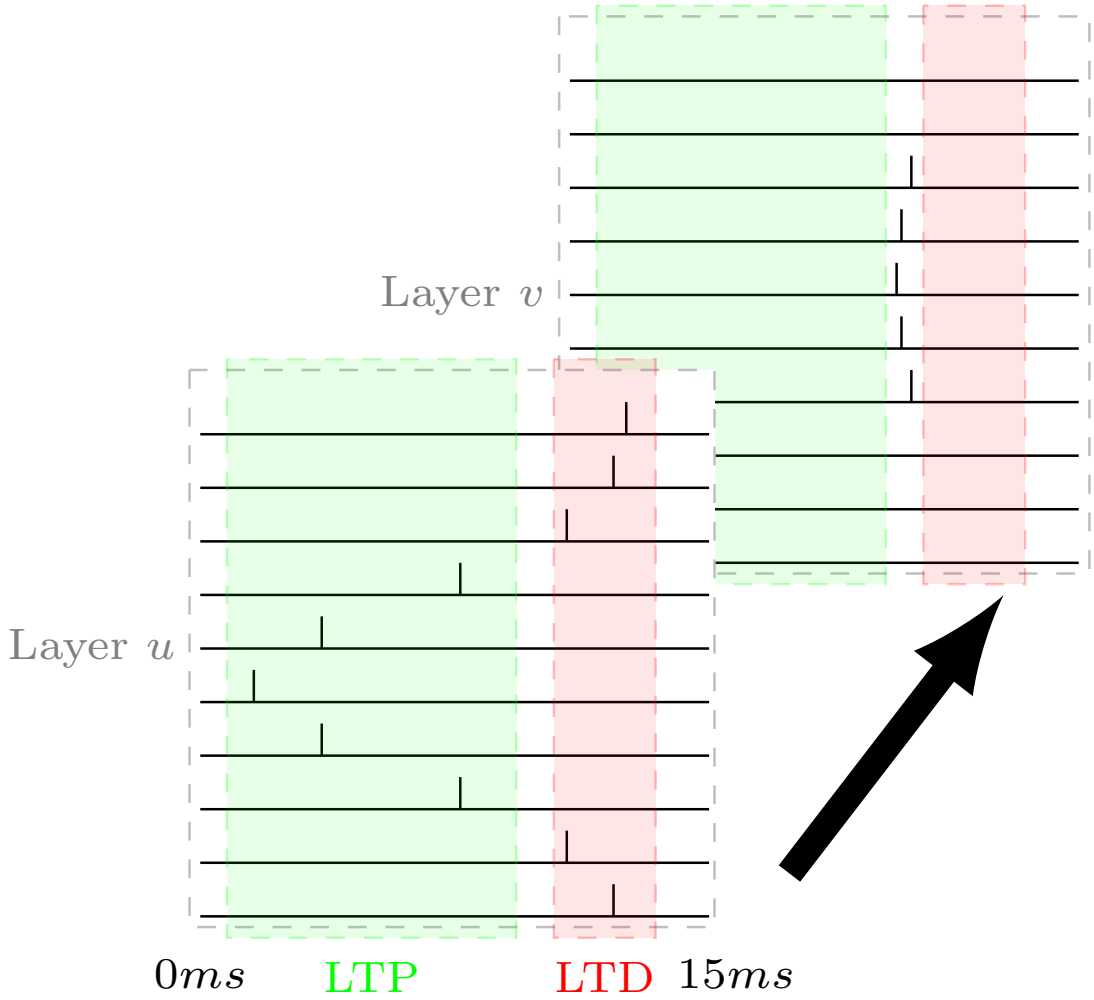


Figure 5.6: The mechanics of self-organisation in the spiking SOM. A spike sequence in a bank of neurons in layer u represents the actual input value, with early firing neurons being well tuned to the actual value, and late firing neurons being poorly tuned to the actual value. The black arrow represents all-to-all feedforward synaptic connections from u to v. At some point in the firing of the pattern in u, a neuron in v fires, winning the competition and becoming the BMU. Neighbouring neurons in v are caused to fire within close temporal proximity. The gap between the LTP and LTD boxes in layer u represents the time at which firing in v occurs relative to the firing in u. Synapses from any neurons that have fired before that point are strengthened (the LTP box) and synapses from any neurons that fire after that point are weakened (the LTD box).

from their preferred input value, due to lateral excitation. As such, weights will be increased at synapses from input neurons that normally fire after the output neuron in the output neuron's preferred input pattern. Synaptic weights will be decreased from input neurons that normally fire before the output neuron in the output neuron's preferred input pattern. These changes are weight dependent, so for a certain weight value a few instances of depression balance with a greater number of instances of excitation, and vice versa, creating stability. This stability means that there can be differentiation between winning output neurons for input patterns; adjacent output neurons will prefer similar sets of input neurons, so stable weights between the maximum and minimum are important in determining which of several neurons with similar preferences reaches threshold first.

Figure 5.6 also shows that the spike for the layer u neuron that is tuned closest to the actual input value comes earlier in the phase than other layer u spikes that also fall in the potentiation window by arriving earlier than firing in layer v. This results in a larger weight increase for connections from neurons that represent the actual input value slightly less accurately, in a single oscillation. This does not, however, result in a greater weight for this connection in the long term, due to the dependence of the weight on the proportion of times that the neuron fires before the layer v region relative to the number of times it fires after the layer v region. For a given layer v neuron, the layer u neuron to which that layer v neuron can be considered best tuned is the layer u neuron that is part of all of the input patterns to which the layer v neuron responds.

The formation of a topological map stems, as in the conventional SOM algorithm, from having a local neighbourhood within which neurons endeavour to imbue other neurons with their input preferences. In the conventional case, the SOM organises around actual input values. In the spiking case, the SOM organises around spike times from neurons that respond with varying firing times when stimulated by those actual input values.

5.2.7 Quality of map formation metric

Several measures to assess SOM mapping quality are introduced in section 4.1.1. The MDS measure was determined to be the most suitable method for assessing the performance of the

spiking SOM. The QE and TE metrics rely on locating the output nodes in the input domain to judge the map quality, and the current network does not contain an obvious and accurate source of evaluating this quality in the way that the weights of the traditional SOM provide. The MDS metric utilises a feature of the spiking SOM model, the grid locations of output nodes. Furthermore, the evaluation of topographic error results in a generally comparable evaluation. It is desirable to be able to make comparisons from the spiking SOM model to the performance of other potential spiking neuron models that involve some form of topographic representation. Some of the more complex measures, such as topographic product or CQoCO, are specifically tailored to address the organisation of a self-organising map, and as such would not be useful for comparison with the output of spiking neuron networks that address topographic representation of an input space, without specifically following the SOM algorithm. Due to the likelihood of the current model having a closer method of operation to other spiking neuron models than to the traditional SOM algorithm, it is more useful to ensure the metric used here enables comparison between spiking neuron models, than it is to provide additional detail to the quality of SOM mapping. The MDS metric meets the criteria set out by Polani (2002) as being important for a good map quality evaluation (see section 4.1.1) and evaluates the topographic mapping error from input to output space, which is the primary motivation for the development of the current network.

The MDS metric describes the ability of an output mapping to represent an input data set by checking the topographic mapping error of the output map given the input data. For a map with no topographic mapping error, the relative distance between any pair of items in the input space is the same as the relative distance between the locations activated by that pair of input data in the output map. The MDS metric is calculated according to equation (4.3) on page 56, repeated here for convenience:

$$E_{MDS} = \sum_{i=1}^{N} \sum_{j<i} (F(i,j) - G(M(i), M(j)))^2$$

where N is the number of input patterns, $M(i)$ and $M(j)$ are the grid locations of the winning nodes in layer v for input patterns i and j respectively, $G()$ is the distance between the pair

of grid locations relative to the total grid size (measured as Euclidean distance), and $F(i, j)$ represents the actual dissimilarity of the pair of input patterns i and j (measured as squared Euclidean distance) (Goodhill and Sejnowski 1997). The value of E_{MDS} represents how well the final network mapping preserves the topology of the input data set. The most accurate mapping achieved is one in which relative distances between patterns in the input space are reflected exactly by relative distances between neurons representing those patterns in the output space, resulting in a minimum E_{MDS} value of 0. The least accurate mapping seen in practice is one in which all input patterns result in activation of the same location in the output map; the E_{MDS} value for this situation will vary depending on the distribution of input patterns. This test is used by Ruf and Schmitt (1998) to analyse their spiking SOM, with final E_{MDS} values of under 20% of the starting value being reported. Section 5.3 will use mean values for E_{MDS}, with the final summed E_{MDS} value divided by the total number of pairs of input patterns compared, to give a value that can be compared regardless of the number of patterns in the input space.

5.3 Results

This section covers the results of testing conducted to confirm the behaviour of the spiking SOM implementation. A measure of the quality of map formation is introduced in section 5.2.7, to be used to interpret the rest of the testing results. The parameters used throughout testing are described in section 5.3.1, and a parameter search on the variables involved in equation (5.5) is described in section 5.3.2, to determine the range of values that result in good map formation. Section 5.3.3 demonstrates the robustness of learning under the chosen parameters in the presence of noise. The spiking SOM has been tested in common scenarios used to test the conventional SOM. As such, the response of the spiking SOM to evenly distributed, randomly selected, two-dimensional input data is analysed in section 5.3.4. Section 5.3.4 also features an analysis of map formation in response to unevenly distributed and changing two-dimensional input spaces. Finally, the results of categorisation tests carried out with the spiking SOM are reported in sections 5.3.5, 5.3.6 and 5.3.7.

5.3.1 Network parameters

The testing conducted in sections 5.3.2 and 5.3.4 made use of standardised parameters for the network, shown in table 5.1. In summary, each dimension n in the input I was associated with a bank u_n of 10 neurons in layer u. The value of n was set to 2 for two-dimensional input. The preferential values of the neurons in each u_n were equally spaced between 0.05 and 0.95. Gaussian tuning curves around these preferential points were used for calculation of the activation values, with distance calculations including circular wrapping from 1 to 0.

Layer v, the SOM layer, was initialised with 100 neurons, arranged in a 10x10 grid through the lateral connection weights. The radius r in equation (4.4) was set to 3.0, and the distance between neurons in the layer were calculated with toroidal structure. Values of $\tau+$ and $\tau-$ in equation (5.5) were set to 11ms and 10ms respectively. This width of learning window approximately matches the temporal width of a network oscillation, leading to negligible influence on learning of spikes from within neighbouring oscillations.

At the start of each training step an input value was determined by selecting randomly from 10 values for each dimension, equally spaced between 0.05 and 0.95, making a total of 100 input patterns from within the two-dimensional input space. A training step lasted through 5 oscillations of the network (approximately 125ms) before the input pattern was changed. The network was allowed to learn for 4000 training steps.

Quality of map formation was assessed using mean E_{MDS} for each pair of input patterns. Given that for the two-dimensional case both the input and output space wrap toroidally, this situation results in a maximum mean E_{MDS} value of $\frac{1}{6}$.

The maximum connection strength values for synapses with presynaptic neurons in layer u are varied for the categorisation tests conducted in sections 5.3.6 and 5.3.7. These datasets contain four and nine input dimensions, resulting in 40 and ninety neurons in u, respectively. As such, the value of $w_{max}^{u \to v}$ is scaled down to 1.5 in section 5.3.6, and 0.7 in section 5.3.7, and the $w_{max}^{u \to Inhu}$ values to 0.4 in section 5.3.7.

Table 5.1: Network parameters for all simulations described in section 5.3

(A) Neuronal parameters, used in (5.1) and (5.4)							
Δt	$\tau_m^{u,v}$	$\tau_m^{Inh\,u}$	V_{rest}	θ^u	$\theta^{Inh\,u}$	θ^v	g
0.1ms	1ms	0.5ms	0.0	0.5	0.01	1.0	0.0

(B) Synaptic parameters, used in (3.10) and (3.11) for different synapse types									
$u \to v$		$u \to Inhu_{exc}$		$u \to Inhu_{inh}$		$Inhu \to u$		$v \to v$	
τ_r	τ_f	τ_r	τ_f	τ_r	τ_f	τ_r	τ_f	τ_r	τ_f
0.2ms	1.0ms	0.4ms	2.0ms	0.2ms	1.0ms	1.0ms	5.0ms	0.1ms	0.5ms

(C) Maximum magnitudes of synaptic connection strength				
$w_{max}^{u\to v}$	$w_{max}^{u\to Inhu_{exc}}$	$w_{max}^{u\to Inhu_{inh}}$	$w_{max}^{Inhu\to u}$	$w_{max}^{v\to v}$
2.2	1.0	1.0	100.0	1.0

(D) Neighbourhood parameters, used in (4.4) and (4.5), for layer v			
a	b	X	X'
3.0	3.0	3.0	3.0

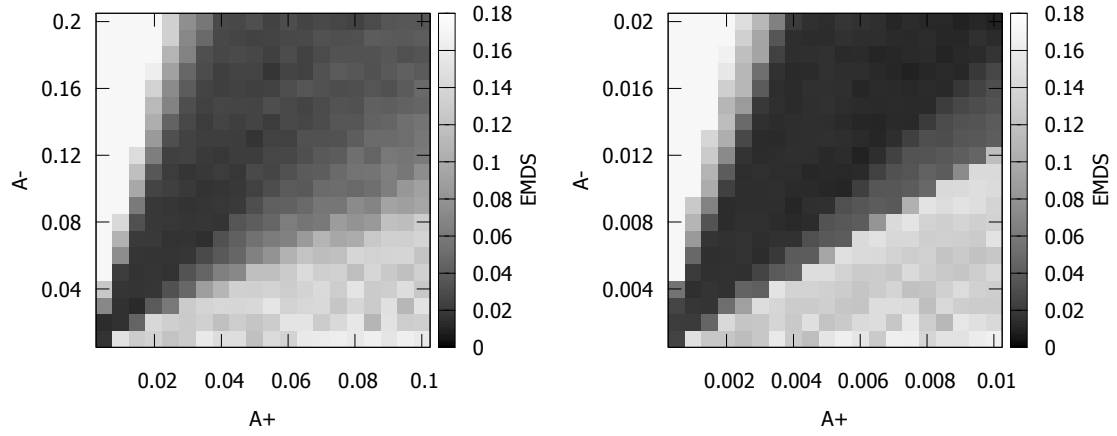
(E) Learning parameters, used in (5.5)			
A_+	A_-	τ_+	τ_-
0.0016	0.0055	11ms	10ms

5.3.2 Learning parameter analysis

Parameter testing was conducted to establish suitable values for the maximum and minimum weight change parameters, A_+ and A_- from equation (5.5). A test of a parameter set consisted of 30[1] randomly initialised maps, trained using the method described in section 5.3.1, with an average normalised E_{MDS} value taken at the end of training used to gauge the quality of map formed with those parameters. Coarse- and fine-grained searches were conducted, the results of which are shown in figures 5.7(a) and 5.7(b) respectively.

The fine-grained search results establish that, for a range of A_+ values up to 0.01, and A_- values up to 0.02, a ratio between 1:2.5 and 1:3.5 of A_+ to A_- will result in good map formation. The coarse-grained search result establishes that there is little performance degradation up to A_+ values of 0.045 and A_- values of 0.11, meaning that large weight changes relative to the

[1] One training run took approximately 20 minutes on a single core of a Beowulf computer cluster (AMD Opteron processor 254, 2792 MHz), and as such the number of trials and parameter granularities used was limited by the available computational resources.



(a) Coarse-grained search

(b) Fine-grained search

Figure 5.7: Parameter search results for A_+ to A_- values: normalised final E_{MDS} values averaged from 30 trials are presented, for ranges of A_+ and A_- values at low resolution (a) and at high resolution (b). Shade indicates the averaged E_{MDS} value, with dark representing low error (high quality mapping) and light representing high error (low quality mapping).

maximum weight can still result in map formation. Only 500 training steps were used in the coarse-grained simulation results, but the high learning rates involved result in a fluctuating error value after this point, rather than increased convergence of error values.

Figure 5.8 shows the progression of E_{MDS} values throughout training for several important locations in the parameter space. Figures 5.8(a), 5.8(b) and 5.8(c) show the difference in map development attained by effectively changing the learning rate, maintaining the same ratio of magnitudes between potentiation and depression. It can be seen that with a higher learning rate comes a slightly quicker decrease in mapping error, but that for the remainder of training the mapping quality fluctuates more dramatically. These fluctuations are caused by the map altering more in relation to statistical properties of very recent input patterns in the high learning rate condition.

Figure 5.8(d) shows the degradation of map quality for an A_+:A_- ratio of 1:20. Depression dominates, and weights are gradually lowered until activity in u no longer evokes any spikes in v. If no neuron wins the competition for any input pattern, there is no distance between

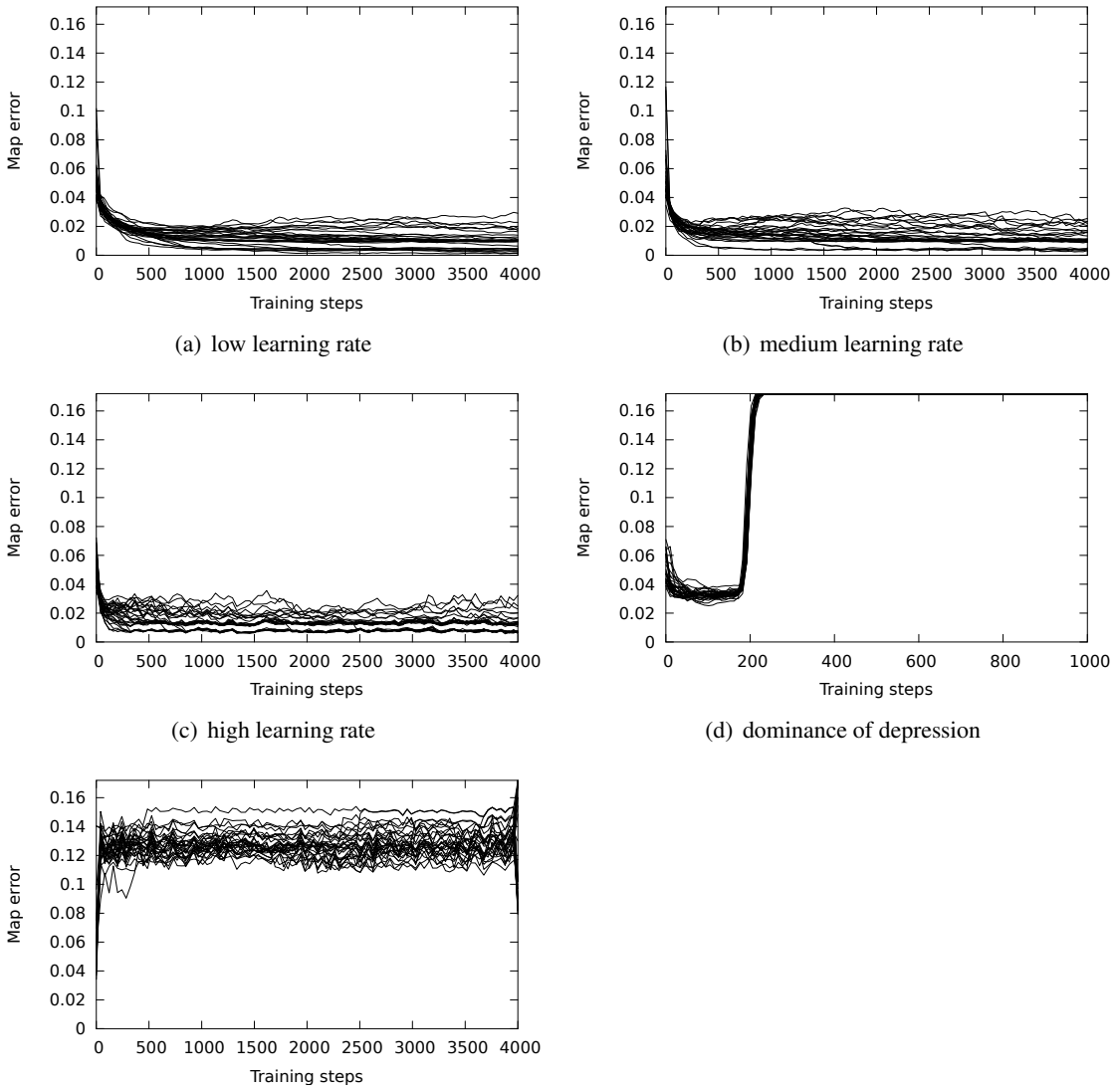
winning nodes for any input pattern, so a maximum error value is reached. This outcome is characteristic of all A_+ to A_- ratios smaller than 1:6, but will result more quickly if there is a greater dominance of depression.

Progression of error values throughout training for the opposite situation, a dominance of potentiation, is shown in figure 5.8(e), with an A_+ to A_- of 4:1. In this regime, weights for less preferred input patterns remain strong, eventually resulting in one output region of the map dominating for all input patterns. Wild fluctuations are seen in the error value; this is caused by slightly different neurons, still close to the dominant region, winning the competition for different input patterns. A brief change in winning neuron can result in a temporary large variation in error value. This outcome is characteristic of all A_+ to A_- ratio greater than 1:1, with an increasing dominance of potentiation leading to decreasing amounts of fluctuation in error value throughout training.

For the following simulations, an A_+ value of 0.0016 and an A_- value of 0.0055 will be used. These values are situated within the acceptable ratio of these parameters, and represent a low learning rate compared to the maximum acceptable rate. A low learning rate results in a more stable output mapping, so is preferable in testing the types of map that are formed.

5.3.3 Robustness to noise

The noise scaling factor g was tested for 11 values between 0.0 and 1.0, with the same value used for both layers u and v, to analyse the robustness of the SOM formation to variable spike times. These values of g resulted in spike time variations in the input pattern of up to around 3ms. Figure 5.9 shows E_{MDS} values, averaged over 30 trials, for training with variation in g. The average values of E_{MDS} ranged from 0.008 to 0.020 for values of g up to 0.5, and between 0.090 and 0.146 for g from 0.6 to 0.9, showing that noise in the neuron model does not prevent the learning mechanism from picking up the statistical correlations present in the input data until a value of g greater than 0.5.



(a) low learning rate

(b) medium learning rate

(c) high learning rate

(d) dominance of depression

(e) dominance of potentiation

Figure 5.8: Time course of E_{MDS} values for 5 points in the parameter space, representing: (a) low learning rate ($A_+ = 0.002$, $A_- = 0.007$); (b) medium learning rate ($A_+ = 0.004$, $A_- = 0.014$); (c) high learning rate ($A_+ = 0.02$, $A_- = 0.07$); (d) dominance of depression ($A_+ = 0.0005$, $A_- = 0.01$); and (e) dominance of potentiation ($A_+ = 0.02$, $A_- = 0.005$). Each figure shows 30 lines, one line per trial in the parameter search.

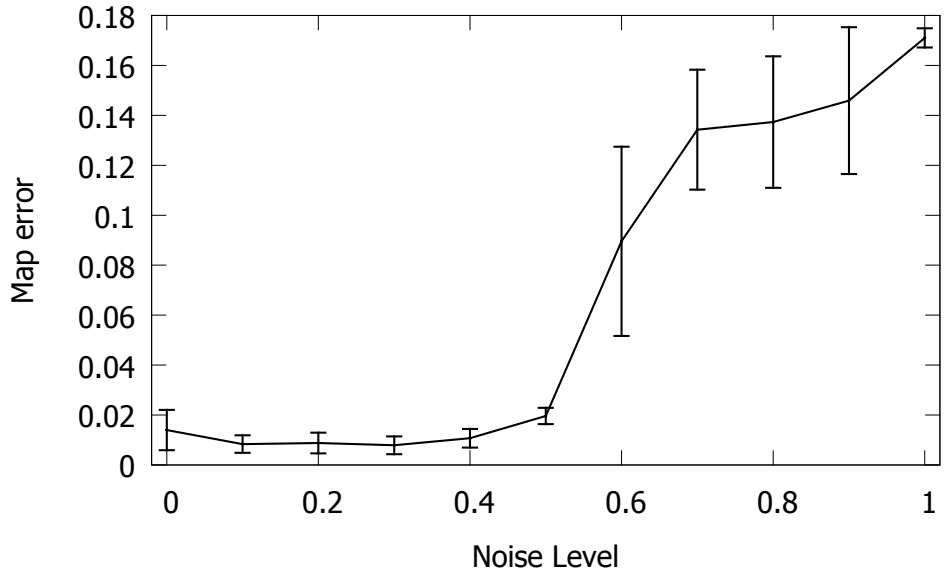


Figure 5.9: Average E_{MDS} value after training plotted against the value of variable g, magnitude of noise in the neuron model. Average E_{MDS} values were taken over 30 trials and error bars represent standard deviation of E_{MDS} values. Accuracy of the output mapping stays high until noise levels reach a critical point around 0.5, with accuracy of the final mapping ending up close to the maximum error value as g reaches 1.

5.3.4 2D input

A test often applied to SOM algorithms is to present random samples from a range [0,1] in two dimensions as input to the network, and test the ability of the output map to organise itself into an appropriate formation capable of representing this input data.

An average final normalised E_{MDS} value of 0.00554 was achieved over 64^2 trials (standard deviation = 0.00483). The final weight matrix for the feedforward connections from input to output neurons is shown in figure 5.10; it can be seen that neurons in one row of the output layer are strongly connected to a specific range of input neurons in one dimension, while varying their connection strength to input neurons in the other dimension uniformly across the row (with the inverse pattern seen within and between columns). Figure 5.11 displays the output layer neurons as circles, positioned in the input space according to the input value to which they respond most quickly (i.e. are best tuned to). Nearest neighbour connections, determined by a maximal lateral excitatory weight on the connection between a pair of neurons, are indicated by connecting

[2] Again, number of trials for computation of this value was dependent on available computing resources.

Figure 5.10: Final *u* (y-axis) to *v* (x-axis) synaptic weights after training with 2-dimensional data (light represents a strong connection and dark represents a weak connection; the top half of the graphic is the bank of input neurons encoding the first dimension and the bottom half is the bank encoding the second dimension; and each ten steps along the x-axis represents a row of output neurons, then the next ten represent the next adjacent row, and so on): the connections to the output layer vary in one dimension across an individual row (with consistency throughout the map), and vary in the other dimension across the rows, encoding the current input as a location in the output map in an organised way.

lines. This preferential tuning is initially random, and throughout training organises to mirror the inputs received. It is worth noting that 5.11(c) and 5.11(d) both show the network state after a good topographic mapping has been achieved, but the position of the neurons is variable; this is because the network is learning at the same rate throughout training in this example, and will morph slightly depending on the distribution of the most recent input data.

The spiking SOM is also capable of generating a map representation skewed to fit the input distribution. To test this, the map was trained using six distributions of two-dimensional input data. In each distribution the likelihood of either dimension being drawn from the range [0,0.5] was altered to a value in the range of [0,0.5] at increments of 0.1. A likelihood of 0.5 represents an even distribution across the two-dimensional space, and a likelihood of 0 represents a distribution entirely in the quadrant of the two-dimensional space between 0.5 and 1.0 in both dimensions. Within each half of each dimension, the distribution is even across the range. The representation of the output nodes in the input space is shown in figure 5.12, as in figure 5.11. In figure 5.12(b) - figure 5.12(e) the input distribution can effectively be split into quadrants of likelihood. At any training step the input pattern is least likely to be selected from the lower-left quadrant (between 0 and 0.5 in both dimensions), most likely to be selected from the upper-right quadrant (between 0.5 and 1 in both dimensions), with each of the remaining quadrants at an intermediate likelihood (between 0 and 0.5 in only one dimension). In figure 5.12(a) the likelihood of an input data point being from the upper-right quadrant is 1, and in figure 5.12(f)

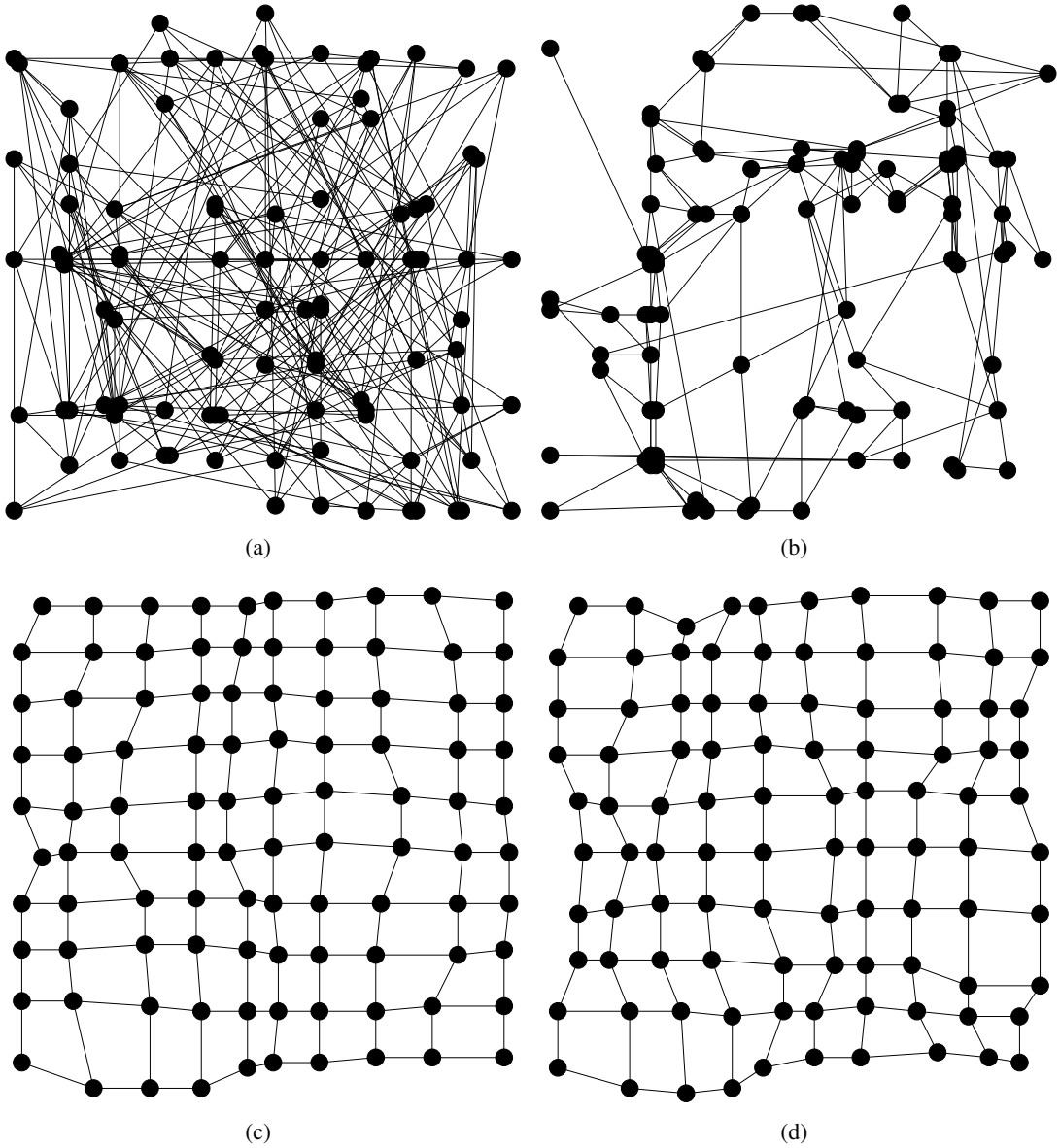


Figure 5.11: Representation of $u \rightarrow v$ feedforward weights in the 2-dimensional input space, where black dots represent output layer neurons and black lines represent nearest neighbour synaptic connections (input and output dimensions are toroidal, but nearest neighbour edges to opposite sides of the input space are omitted): (a) shows the random starting distribution; after 200 training steps (b) the nodes begin to align to the input data; (c) and (d) show the trained map after 2800 and 3600 training steps respectively - learning is ongoing, so stochastic fluctuations in the distribution of recent input patterns are reflected by minor modulations in the map weights.

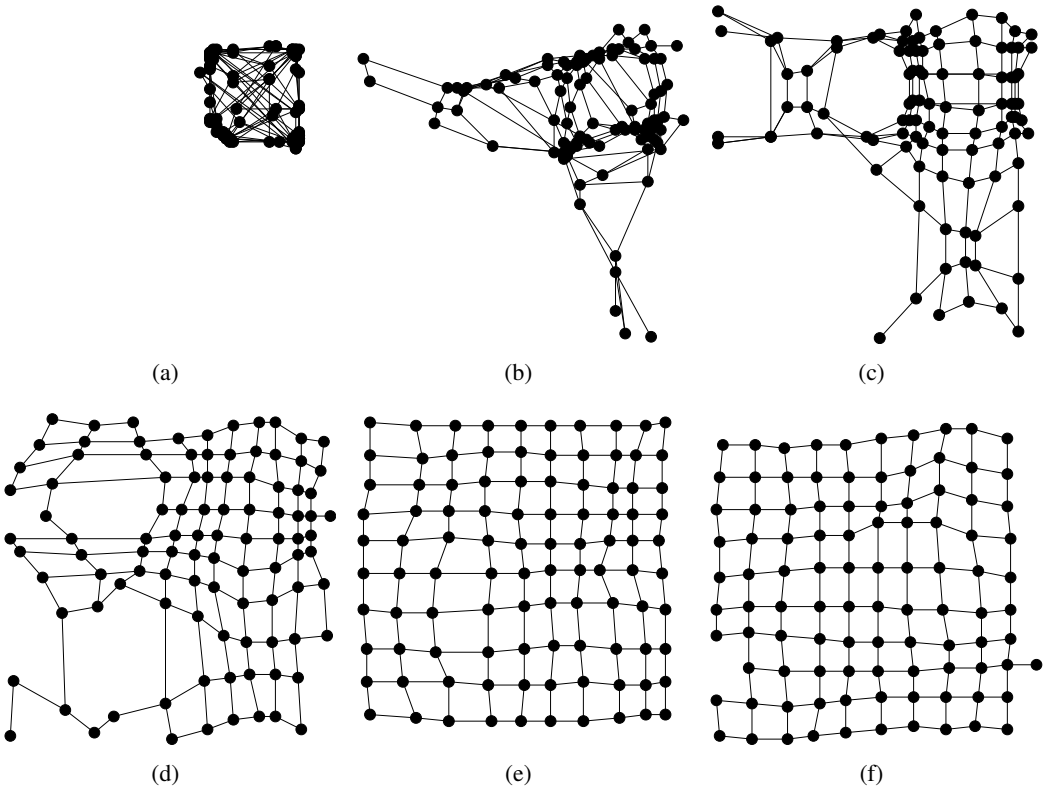


Figure 5.12: Representation of $u \rightarrow v$ feedforward weights in the 2-dimensional input space, where black dots represent output layer neurons and black lines represent nearest neighbour synaptic connections (input and output dimensions are toroidal, but nearest neighbour edges to opposite sides of the input space are omitted for clarity of the figure). (a), (b), (c), (d), (e), (f) show the final mapping for input distributions in which the probability in both dimensions of an input value being between 0 and 0.5 is 0, 0.1, 0.2, 0.3, 0.4 and 0.5 respectively.

the distribution is even across the entire two-dimensional space. Figure 5.13 demonstrates that map formation on average results in a good representation of the actual distribution of input data by the proportion of nodes in the final output mapping that represent a given quadrant of the input space.

Another feature of the topographic mapping shown by some SOM algorithms that attempt continuous learning is the ability of the network to reconfigure to a new input distribution midway through training, after a mapping has been established to an existing input distribution. This was tested by training the network using only 75% of the input data space, leaving out the quarter of the input space square covered by values of greater than 0.5 in both dimensions. This reduced data set was used for 2000 training steps, then the full range of input data, including

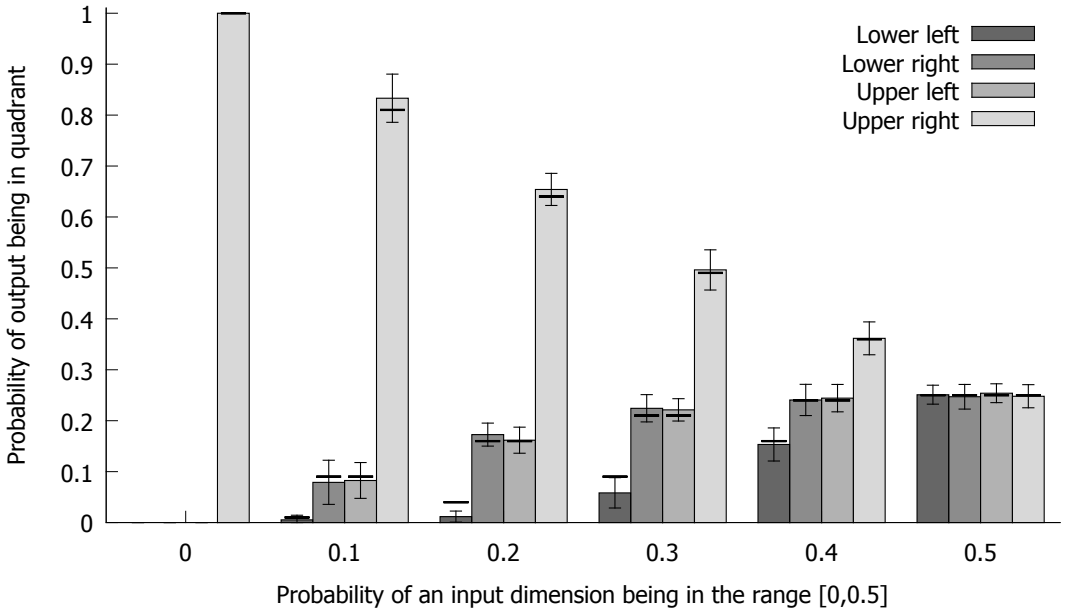


Figure 5.13: The probability of an output node in a trained map representing each quadrant of the input space, plotted against the likelihood of an input dimension having a value within the range [0,0.5]. Bar height represents average over 30 trials; error bars represent standard deviation; black horizontal marks represent the actual probability of an input datum being within a quadrant of the input space; bar colour represents which quadrant of the two-dimensional space neurons were found in, according to the key.

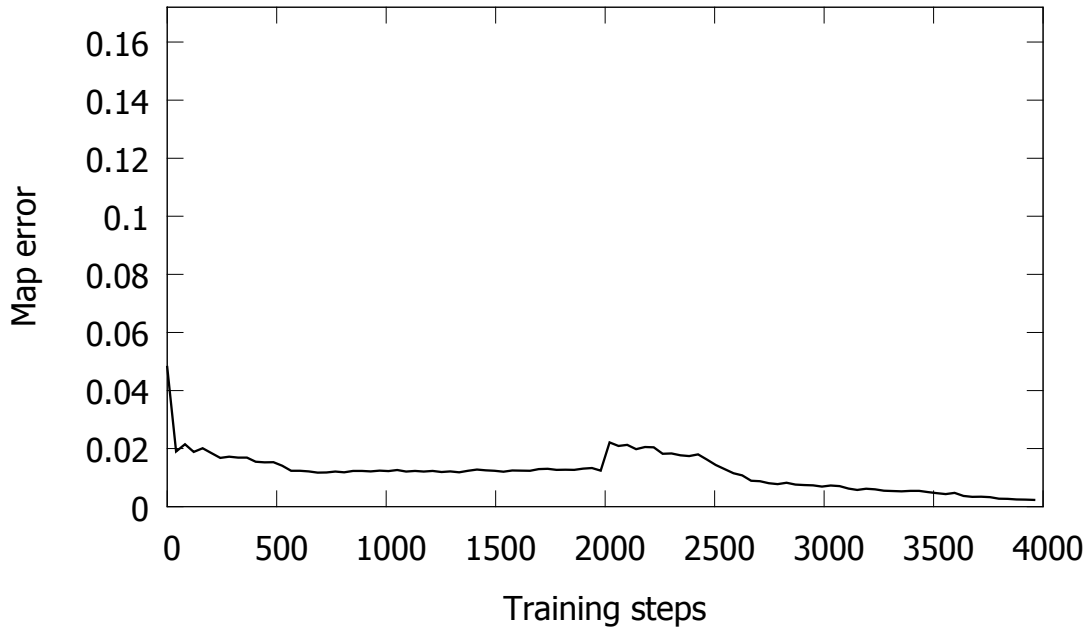


Figure 5.14: Typical evolution of average E_{MDS} value for 2-dimensional input: the error in the output map reduces up to the halfway point in training, at which point the input distribution is expanded; the error in the map jumps up as the error measure is now relative to the new distribution; the error then decreases again as the map adjusts to the new distribution.

the previously omitted quarter, was used for a further 2000 training steps. The organisation of the output network over time to this change in input data set is shown in figure 5.15.

The evolution of the topographic error (figure 5.14) shows that the network adjusts to the initial input range as normal, but settles at a slightly higher E_{MDS} value, most likely due to the discrepancy between shape of input space and shape of lateral connections in the output layer. The extra data is introduced half way through training, resulting in a spike in E_{MDS} value, as the output mapping is no longer suitable for the input data, and the topographic error is then reduced to a lower value as the output map reorganises to the new data. The final weight matrix for the feedforward synapses is qualitatively identical to the one obtained when training using the full data range from the start (figure 5.10). This simulation demonstrates that the map has learnt to represent the initial input distribution, but when a new distribution is presented the map is capable of adjusting appropriately.

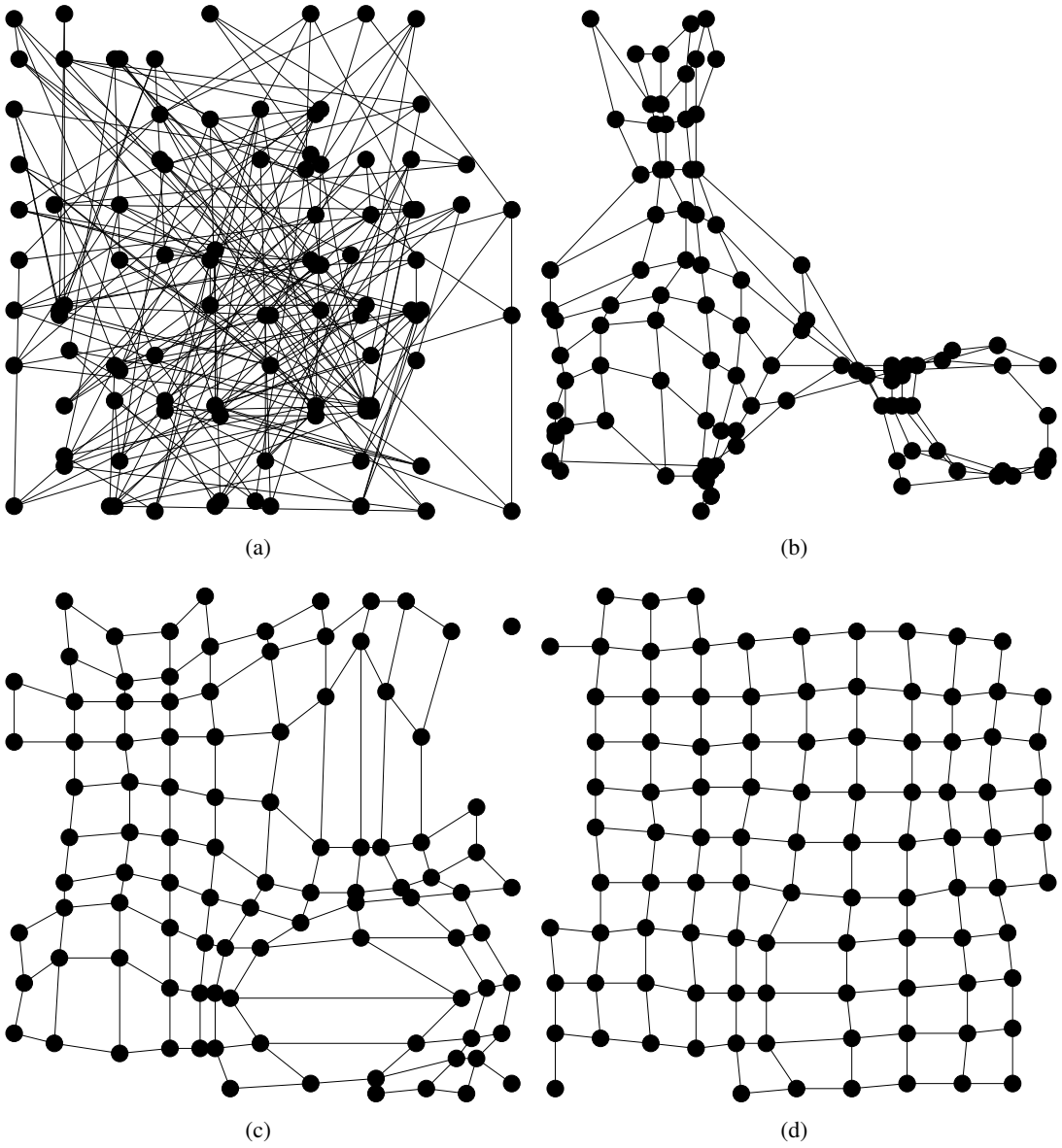


Figure 5.15: Representation of $u \rightarrow v$ feedforward weights in the 2-dimensional input space (arranged as per figure 5.11): (a) shows the random starting distribution; after 1600 training steps with the partial input data (b) the nodes are aligned to the input data space, leaving a gap in the input space from which no training examples have yet been received; (c) shows the expansion of the map to the newly increased range of input data after 800 training steps with the full input distribution; and (d) shows the trained map after 2000 training steps with the full data range - the map has adjusted to the new input data (neurons on the map border are alternating from one side of the figure to the other because the toroidal nature of the input node's tuning functions means that input values of 0 and 1 are essentially identical, and the row or column that the neurons are a part of is lined up along the 0-1 divide, with some neurons placed on one side and some on the other.)

5.3.5 Categorisation

SOMs can cluster input patterns, creating a specific spatial location that is activated by incoming members of a specific category. If there are category divisions in the input data, nodes in the output layer will respond more reliably to one category of input than to others. A trained SOM can therefore be used as a categorisation tool by assigning each output node a category to represent based on whether that node fires reliably for one particular category. The capacity of the spiking SOM to be used as a categoriser in this way has been tested through training with two datasets commonly used for assessing the categorisation ability of a system. This capacity is demonstrated to provide evidence that this type of functionality is possible with the network, and that organisation can occur to datasets that contain relatively distinct cateogires within a high-dimensional space.

In the examples in sections 5.3.6 and 5.3.7 five-fold cross validation is performed: a dataset is split into 5 chunks and the network is initialised and trained 5 times, using a different set of 4 chunks as training data and 1 chunk as testing data each time, so that in total all data points are used for testing once. Each training phase lasted for 4000 training steps with random selection of input pattern after each 5 oscillations of the network, as described in section 5.3.1. At the end of a training phase the output nodes were designated as representing a category based on the input category to which they responded most frequently during training. The testing patterns were then presented to the trained network one at a time, and the output activity recorded. Lateral connections were still used in this testing phase, so multiple neurons in the output layer fired for each testing pattern, in an area with width determined by the lateral connection profile at the end of the training phase. The testing pattern was categorised by the network as belonging to the category to which the highest number of output neurons firing in response to that input pattern had been designated as representing. If more of the output neurons firing in response to a testing pattern had been designated with that pattern's category during the training phase than any other single category, then the pattern was considered to have been correctly classified.

5.3.6 Iris dataset

The first dataset used to test the categorisation performance of the spiking SOM was the Iris dataset (Fisher 1936). This dataset is of sizes of flowers of the Iris plant; it consists of three categories each with 50 members, and each data point has 4 values, petal length, petal width, sepal length and sepal width. One of these categories "Iirs Setosa Canadenisis", is fairly distinct from the other two, "Iris Virginica" and "Iris Verisicolor".

For the current purpose, the values for each dimension were normalised in the range [0,1], and the spiking SOM model was initialised as in section 5.3.4 but with 40 input neurons, making one bank of 10 for each dimension. This increase in number of input neurons meant that the feedforward connection strength from layer u to layer v had to be reduced; the value of $w_{max}^{u\to v}$ was scaled down to 1.5. Remaining parameters remained identical, with the exception of the X and X' values in equation (4.5), controlling the evolution of the width of the neighbourhood function. For continuous learning these values are identical, meaning no change in lateral connection strengths over time. The standard neighbourhood parameters were used as one condition in the categorisation performance tests (X=3.0 and X'=3.0). However, output map quality can potentially be improved by starting with a large neighbourhood and reducing it throughout training. This approach can obtain a globally ordered topology initially, and refine details later on. This regime was used as a second condition in the categorisation tests, with X set to 4.0 and X' to 2.5. These values are in numbers of neurons, so for a 10×10 map a radius of greater than 4.0 is the majority of the map.

Categorisation accuracy, averaged over 9 trials, was 87.8% (standard deviation = 1.3%) in the without neighbourhood reduction condition, and 90.9% (standard deviation = 1.7%) in the with neighbourhood reduction condition. Figure 5.16 shows the confusion matrix from this categorisation task, demonstrating that all of the incorrect classifications are a result of a slight merging of two of the three categories. Table 5.2 shows these results in comparison with the results achieved for other categorisation algorithms using this dataset. The spiking SOM categorises better than Matlab implementations of the k-Means and SOM algorithms (Bohte et al. 2002) (although it is unclear the parameters used and extent of parameter searching conducted to achieve

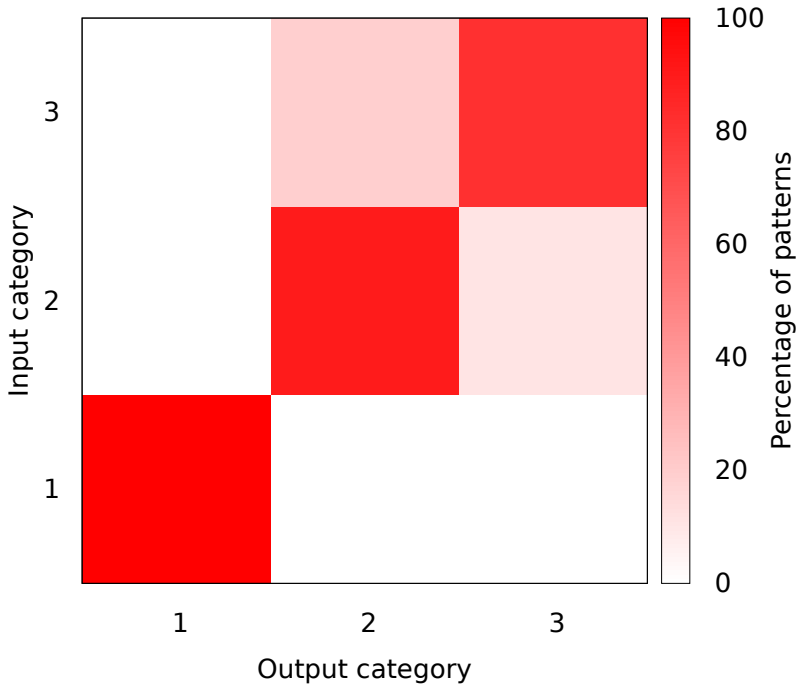


Figure 5.16: Confidence matrix for classification using the iris data set. Rows show input category and columns show the output category determined by analysis of the network responses. The colour scale shows the percentage of input patterns from a category that are classified as each category by the network output. Category 1 is perfectly classified, and 2 and 3 are well classified, with slight confusion between them.

Table 5.2: Clustering accuracy (%) of spiking neuron and non-spiking neuron algorithms for the iris dataset. The current approach is shown as Spiking SOM with neighbourhood reduction (NR) and Spiking SOM without NR.

Non-spiking	% acc.	Spiking	% acc.
k-Means (Bohte et al. 2002)	88.6	RBF (Bohte et al. 2002)	92.6
SOM (Bohte et al. 2002)	85.33	SpikeProp (Wu et al. 2006)	96.1
Matlab BP (Wade et al. 2010)	95.5	SWAT (Wade et al. 2010)	95.3
Matlab LM (Wade et al. 2010)	95.7	RBF (Gueorguieva et al. 2006)	89
TEST (Yoon and Lee 1999)	91.7	SNN_{Bako} (Bako 2010)	83.4
		Spiking SOM with NR	**90.9**
		Spiking SOM without NR	**87.8**

these results), the FPGA implemented classification network of (Bako 2010), and the spiking neuron RBF network of (Gueorguieva et al. 2006).

Categorisation performance is slightly worse than several other networks: the spiking RBF model of Bohte et al. (2002), the SpikeProp model of Wu et al. (2006), the SWAT model of Wade et al. (2010), the TEST algorithm (Yoon and Lee 1999), and Matlab implementations of the backpropogation and Levenberg-Marquardt training algorithms (Wade et al. 2010); it is worth noting, however, that these are designed specifically for data classification purposes and do not feature the topographical ordering properties of the SOM model.

5.3.7 Wisconsin Breast Cancer dataset

The second dataset used for categorisation testing was the Wisconsin Breast Cancer dataset (WBCD), consisting of 683 samples from 2 categories (444 benign and 239 malignant tumours), with 9 measures of features of cytology. Each of the 9 measures is a discrete value from [1,10], converted into a value in the range [0,1] for the current purpose, and represented using a bank of 10 input neurons, meaning that layer u consisted of 90 neurons in this case. As such the value of $w_{max}^{u\to v}$ is scaled down to 0.7, the value of $w_{max}^{u\to Inhu}$ to 0.4, and the X and X' values were adjusted to 3.5 in the without-neighbourhood-reduction condition, to account for only having 2 categories occupying the 10×10 output map; all other parameters remained identical to those used in section 5.3.6.

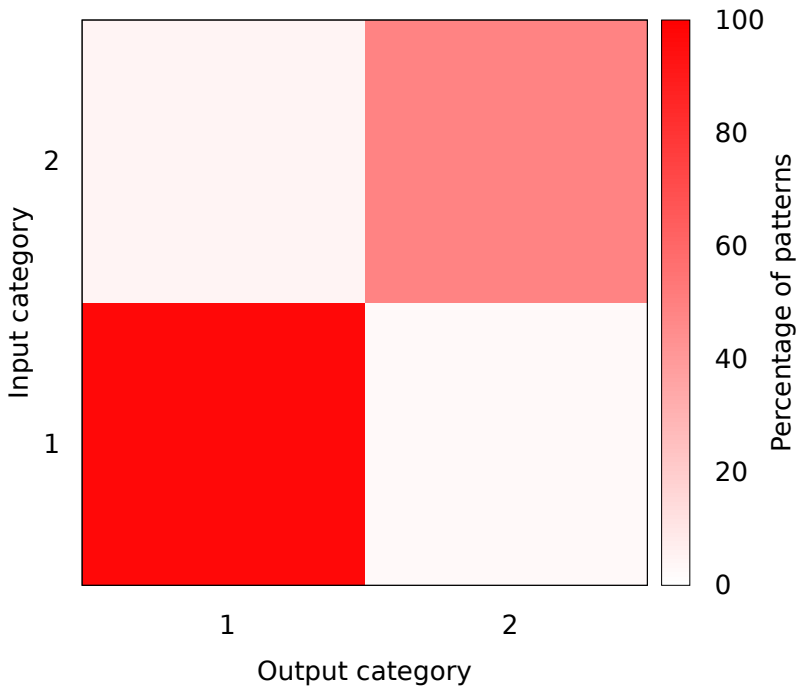


Figure 5.17: Confidence matrix for classification using the Wisconsin breast cancer data set. Rows show input category and columns show the output category determined by analysis of the network responses. The colour scale shows the percentage of input patterns from a category that are classified as each category by the network output. The two categories are well classified.

Table 5.3: Clustering accuracy (%) of spiking neuron and non-spiking neuron algorithms for the WBC dataset.

Non-spiking	% acc.	Spiking	% acc.
Matlab BP (Wade et al. 2010)	96.3	SpikeProp (Wu et al. 2006)	97
Matlab LM (Wade et al. 2010)	96.7	SWAT (Wade et al. 2010)	95.3
		SNN_{Bako} (Bako 2010)	89.5
		Spiking SOM with NR	**96**
		Spiking SOM without NR	**97**

Categorisation accuracy, averaged over 8 trials, was 96.4% (standard deviation = 0.4%) in the without neighbourhood reduction condition and 97.0% (standard deviation = 0.1%) in the with neighbourhood reduction condition. Figure 5.17 shows the confusion matrix for this categorisation task, demonstrating good categorisation of both categories present in the dataset. Again, these results are compared with the categorisation accuracy of other algorithms using this dataset, shown in table 5.3. The spiking SOM again outperforms the FPGA categorisation algorithm implementation of Bako (2010), and achieves a very similar level of accuracy to models that have been designed specifically for clustering and categorisation operations.

5.3.8 Micro-Doppler dataset

The third dataset used for categorisation testing was a micro-Doppler dataset. The initial data was collected from experiments in which 3 participants each performed 5 activities twice in front of a sonar recording device (Zhang and Andreou 2008). The raw sonar output was filtered through 100 spatiotemporal receptive field (STRF) filters and normalised to the range [0,1], resulting in a 100 dimension input vector for each pattern. As such, this dataset consisted of 30 samples from 5 categories (2 participants performing each activity: fast cycling, slow cycling, skating, running and walking), with 100 measures of STRF activity throughout the pattern. Each of the 100 measures (a value from between 0 and 1) was represented using a bank of 10 input neurons, meaning that layer u consisted of 1000 neurons in this case. As such the value of $w_{max}^{u \to v}$ is scaled down to 0.06, the value of $w_{max}^{u \to Inhu}$ to 0.05, and the X and X' values were set to 3.0; all other parameters remained identical to those used in section 5.3.6.

Three-fold cross-validation, rather than five-fold, was used for this input dataset, due to the

Table 5.4: Clustering accuracy (%) of spiking neuron and non-spiking neuron algorithms for the micro-Doppler dataset.

Non-spiking	% acc.	Spiking	% acc.
Bootstrap (Garreau et al. 2011))	80	**Spiking SOM without NR**	**72**
K-means (Dura-Bernal et al. 2011)	95		

combination of reduced number of input stimuli and increased number of categories relative to the iris dataset and WBCD; using three-fold cross-validation ensured that there was more than one input pattern in a testing set for each fold. Categorisation accuracy achieved was 72.3%. Only a without neighbourhood reduction condition was tested in this case. Again, these results are compared with the categorisation accuracy of other algorithms using this dataset, shown in table 5.4. The spiking SOM categorisation is worse than other categorisation attempts performed on this dataset (Dura-Bernal et al. 2011; Garreau et al. 2011). However, these other attempts do not use a generic spiking neuron network that is not specifically designed for a categorisation task of this nature. The confusion matrix of the spiking SOM classification output, shown in figure 5.18, shows that activity category 4 is frequently mistaken for category 1. The inability of the spiking SOM to detect the differences between these categories from among the 1000 input spike trains results in overall decreased accuracy at this task. Classification for the other categories remains strong.

A potential method for improving categorisation accuracy in cases with an increased number of categories is to increase the size of the network relative to the width of the neighbourhood. This could result in more space within the output layer for a finer granularity of category separation to form. However, without altering the network parameters (aside from the feedforward connection strengths) from those used to produce SOM-like results, acceptable categorisation performance is found.

5.3.9 Initial testing for potential hardware implementation

Special purpose, analogue spiking neuron hardware has been investigated in recent years (Indiveri et al. 2011). The fundamental difference between the analogue principles of neural processing and the digital principles of traditional computing mean that huge computational power

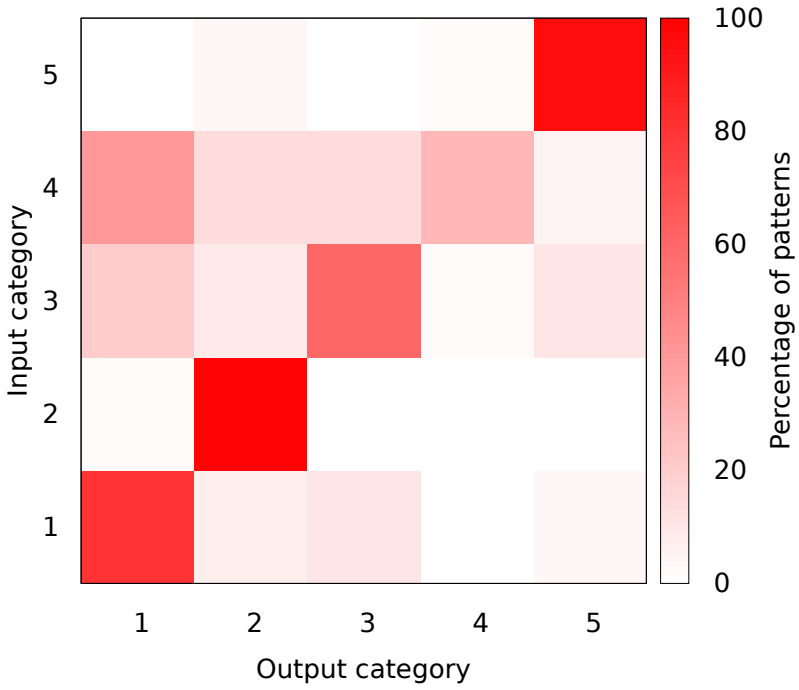


Figure 5.18: Confidence matrix for classification using the iris data set. Rows show input category and columns show the output category determined by analysis of the network responses. The colour scale shows the percentage of input patterns from a category that are classified as each output category. Categories 1, 2 and 5 are well classified, category 3 gets confused with category 1 in a fifth of cases, and category 4 is often confused with other categories, lowering the classification accuracy for this dataset.

is required to enable real-time[3] performance of a spiking neural network. This has motivated the field of neuromorphic engineering, constructing analogue hardware in which the speed of the neurons and their connections is independent of their quantity. The development of network implementations that perform useful, non-trivial functions on these platforms is crucial to continuing to improve the hardware. With this in mind, simulations were carried out to examine the possibility of constructing a version of the spiking SOM on an analogue chip.

The hardware has 128 IAF neurons, and 128 $\times$ 32 synaptic circuits. Each synaptic circuit contains 2 excitatory, 2 inhibitory and 28 excitatory plastic synapses. Additionally, synaptic circuits from unused neurons can be redirected to other neurons if fewer than 128 neurons are used in the network. The plastic synapses are capable of learning using the BSF rule (see section 3.3.5). This hardware has been used in a model of thalamo-cortical processing for auditory stimuli (Sheik et al. 2012), which was able to learn and selectively respond to the dynamic features of auditory stimuli. Also, a classification task was performed on binary input patterns represented by either high or low firing rates across 60 input synapses, with a good accuracy level achieved using this hardware (Mitra et al. 2009).

Software simulations were conducted, using the same spiking SOM model described previously, although with several modifications to take account of the constraints imposed by the limitations of the planned hardware platform. To simplify the network for initial testing on the hardware, one input dimension was used, with a one dimensional output map. This involved a reduction in the number of neurons in u to 10, all of which were tuned to respond to the same input dimension. Input neuron receptive field tuning remained as in the two-dimensional model. The output layer was also reduced to 10 neurons, with neurons located on a one dimensional output grid, reflected in the lateral connection weights. The parameters relating to lateral connection initialisation (see equations (4.4) (on page 57) and (4.5) (on page 58)) were set as: $a = 5$, $b = 7$, $X = 3$, $X' = 3$, with grid distances between output neurons calculated using a ring structure. Feedforward connection strengths between the input and output layers were initialised with the same Gaussian noise as in the two-dimensional model, but with stronger maximum weight

[3]The current spiking SOM model, set up for two input dimensions, runs approximately three times slower than real-time; performance is greatly reduced for an increased number of input dimensions.

($w_{max}^{u\to v} = 4.0$, $w_{max}^{u\to Inhu} = 2.0$), to make up for the reduced number of units in u driving v. Network parameters not mentioned were identical to those in table 5.1.

The largest change to the function of the spiking SOM enforced by the nature of the hardware is a change to the learning rule. Here the STDP rule used previously (see section 5.2.2) is replaced by the BSF rule (see section 3.3.5), as described by equations (3.21) to (3.25) (on pages 47 and 48). The BSF rule is designed to be able to replicate rate-based neural plasticity by potentiating a synapse when, at the time of presynaptic firing, the postsynaptic membrane potential is above a threshold (indicating sufficient presynaptic activity) and a calcium concentration measure is within a high range (indicating a recent history of postsynaptic activity). A synapse is depressed when presynaptic firing coincides with low postsynaptic membrane potential and a calcium concentration within a low, non-zero range. This learning rule poses a fundamental problem for the type of self-organisation described in section 5.2.6. The input neuron best representing the actual input pattern fires ahead of the other input neurons, early in the phase, at which time the output neurons will have membrane potentials close to zero. As such, connections from early firing input neurons to firing output neurons will never be potentiated, and the self-organisation fails.

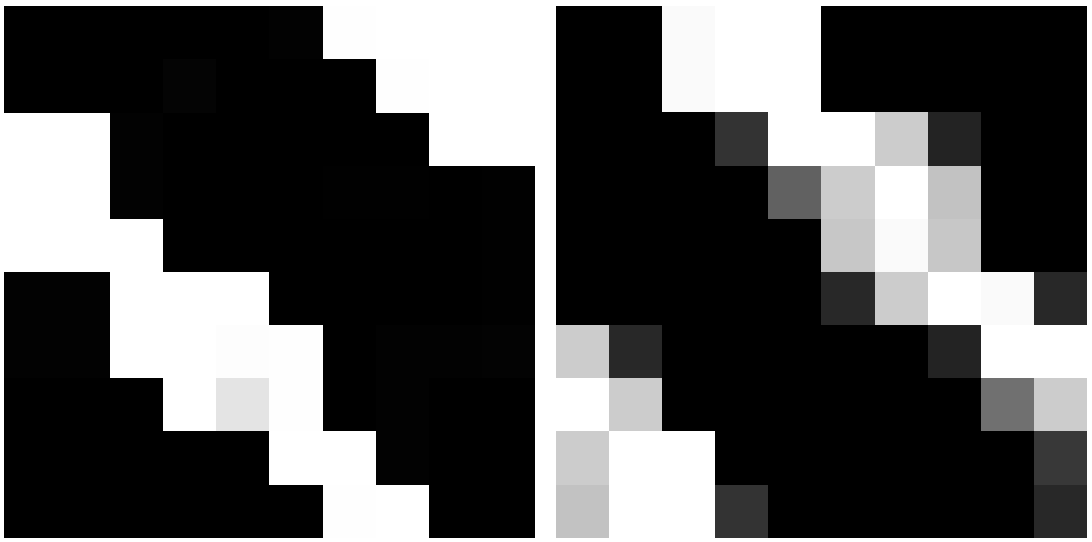
A solution to this problem is to set the threshold value θ_V (from equations (3.22) and (3.23)) to below the resting potential, 0, and above the after-spike reset potential, -1. This means that synapses from all early firing neurons will be potentiated by all output neurons firing for the input pattern. Input neurons that are poor representatives of the actual input values will fire after the output firing. This will therefore occur while the firing output neuron's membrane potentials are hyperpolarised, causing synapses from these input neurons to be depressed. The calcium variable prevents neurons without a recent history of postsynaptic firing from receiving changes to their synaptic weights, ensuring that, despite the low threshold for potentiation, weights do not increase for neurons that are not winning the current competition. This aspect of the rule coincides well with the established spiking SOM training protocol of presenting an input pattern across multiple network oscillations in succession (see section 5.3.1). Upwards and downwards drift of synaptic weights, as in equations (3.24) and (3.25) was not necessary in the current

context. BSF rule parameters were, then, established as: $\theta_V = -0.05$, $a = 0.0016$, $b = 0.006$, $J_C = 1.0$, $\tau_C = 10ms$, $\theta^l_{up} = 1.0$, $\theta^h_{up} = 10.0$, $\theta^l_{down} = -10.0$ and $\theta^h_{down} = 1.0$. These calcium rule parameters ensure that successive post synaptic spikes in the most recent two oscillations are required for potentiation to occur.

Training of the network was conducted as detailed in section 5.3.1, with 4000 training steps, each consisting of the presentation of one input pattern for 5 oscillations of the network. An average final normalised E_{MDS} value of 0.00813 was obtained across 9 trials. The maximum likely E_{MDS} value for this input data set, if every input pattern produced a response from the same output node, is 0.0944. The obtained average value represents an average E_{MDS} reduction through training of 83% relative to the values for the random initial weights at the start of training. Figure 5.19 shows examples of the final weight matrix and output layer firing. After training, a transition through the input data space is met with a representative transition through the output layer. Figure 5.20 shows a typical example of the time course of the E_{MDS} value throughout training. As in the two-dimensional spiking SOM, the error present in the mapping for the simulated spiking hardware SOM quickly reduces to a fraction of its starting value, then remains stable while training continues despite there being no reductions in learning rate or neighbourhood width. This result demonstrates the plausibility of a potential implementation of the spiking SOM on analogue spiking neuron hardware. The principles of self-organisation through spike timing, oscillatory activity, and a lateral neighbourhood established for the spiking SOM still result in reliable output map formation, despite changes to the network structure and scale, and the inclusion of an entirely different learning rule.

5.4 Summary

The spiking SOM implementation described in this chapter is unique, making use of a set of general cortical properties, implemented with spiking neuron methods, to produce results that are demonstrably qualitatively similar to those produced by the original SOM algorithm. The model, described in section 5.2, demonstrates how a network of spiking neurons can process information in a manner mimicking some important aspects of the flow of information in cortical networks. Oscillations within a layer result in a natural segregation of continuous input stimuli,



(a) Firing times of output neurons (b) Weight matrix from input to output

Figure 5.19: Example results of 1-dimensional spiking SOM for hardware: 5.19(a) Firing times of neurons in v (rows) for input patterns (columns); 5.19(b) Weights between neurons in u (columns) and nodes in v (rows). Lighter squares represent earlier firing times in 5.19(a) and stronger weights in 5.19(b).

while transmitting relevant stimulus properties to the subsequent layer through a temporal code within each phase of the oscillation. The oscillatory nature of the input encoding results in oscillating periods of activity and inactivity within an output layer, while recurrent lateral connections result in a winner-takes-all functionality. This combination of temporally transmitted information and competition are a good fit for learning through a biologically plausible STDP rule. The outcome of this network structure is the reliable generation of a two-dimensional topographic representation of the input space.

In section 5.3.4 the robustness of this map formation is demonstrated. With no need for learning rate or neighbourhood function decay schemes, the spiking SOM is capable of organising to two-dimensional input with a final topographic mapping error that is just a few percent of the error present in a random initial map. This mapping is representative of the input distribution, with maps trained on data skewed away from an even distribution across the input range displaying a skewed output representation. The stability of learning rate and neighbourhood size ensure that a reorganisation of the output representation is possible at any point in response to a change in the nature of the input distribution. Of course, the map could potentially be frozen

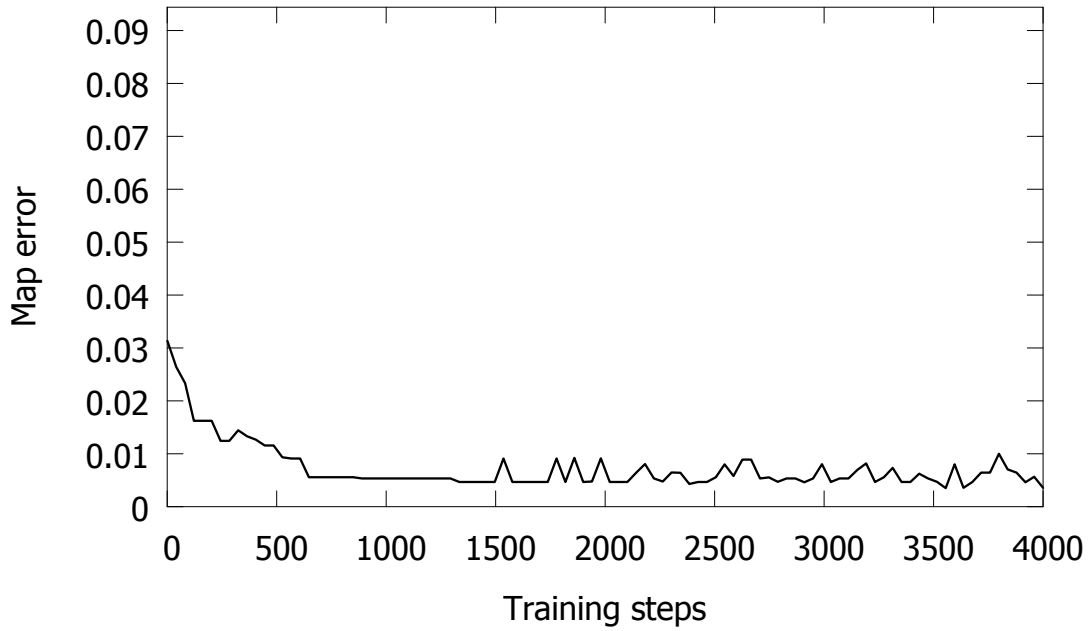


Figure 5.20: Typical time course of E_{MDS} values for 1-dimensional spiking SOM for hardware. E_{MDS} value reduces within the first 1000 training steps, then fluctuates throughout the remainder of training.

at any point in the learning process by reducing the learning rate to zero, with intermediate learning rates resulting in more map stability, such that a changed input distribution would need to be present for a longer period of time before the output representation would adapt.

In sections 5.3.6, 5.3.7 and 5.3.8 the effectiveness of the spiking SOM as a classifier for datasets containing categorical information is demonstrated. In general, categorisation performance is on a par with other spiking neuron network models. However, the majority of the models included in the performance comparison are designed with the task of classification as the main goal of the algorithm. The spiking SOM also incorporates the general representational properties described in the above paragraphs, and as such its ability to rival specifically designed categorisation models at a categorisation task is impressive. Finally, in section 5.3.9, a potential hardware implementation of the spiking SOM is discussed, with some preliminary results demonstrating the feasibility of this endeavour. A continuation of this work, leading to an actual implementation, represents a potentially important step in the development of analogue spiking neuron hardware, as it would be the first implementation of its kind of a model with general

topographic mapping and categorisation properties. In turn, this could lead to removal of some software imposed computational constraints on the network, resulting in the implementation of fast, large SOMs of spiking neurons, the investigation and analysis of which might be a highly fruitful enterprise in the field of artificial intelligence.

A goal of the current work is to present a model capable of topographic map formation, but also capable of representing the structure of conceptual information. Chapter 6 extends the model in the current section to include encoding of binary input patterns and differential levels of lateral excitation and inhibition through the introduction of separate excitatory and inhibitory neurons within the output layer. These changes allow for the storage and retrieval of hierarchical data sets with the generic spiking SOM framework.

Chapter 6

Hierarchical category representation

6.1 Introduction

The aim of this chapter is to produce a model that is capable of representing hierarchical conceptual information. As discussed in section 2.5, representation of high-order information, for example visual object categories in inferior temporal cortex, appears to have components of both continuous and categorical representational structure in the brain (Tompa and Sáry 2010). Neurons respond to categories with an inherent hierarchical structure (Kiani et al. 2007). Population response patterns for one stimulus are more highly correlated with response patterns for a stimulus that is a member of the same sub-category than for a less related stimulus. The clusters formed from response patterns in inferior temporal cortex resemble categories that seem intuitively to be present to humans.

Evidence from biology also suggests that some of these high-order representations are formed from input that represents the presence or absence of features. Kiani et al. (2007) found that the category structure present in population output responses of neurons in IT cortex could not be accounted for by low-level features of a visual stimulus, but instead must have resulted from visual information processing after V1 that is likely to extract information about the features present in a stimulus. Additionally, it has been demonstrated in a study of the representation of information about external events in the mouse hippocampal populations that neural assembly responses can be interpreted as a binary code representing information about situational features (Lin et al. 2006).

This suggests that a binary set of feature-object pairings, as seen in formal contexts from the field of formal concept analysis (FCA) (Wille 1982; Wolff 1993) may be similar to some of the

types of input used to form object categories in the brain. A conceptual structure in FCA, known as a context, consists of binary object-feature relationships, whereby a feature that is present for a given object is denoted by a 1, and a 0 is used where a feature is not present. In this domain, a concept is the set of object-feature relations such that for all features and objects selected, all features are shared between all objects. Table 6.1 (introduced in section 6.2.5, below) shows an example input data set of this nature. In terms of this context, an example concept is the relation between features 1, 2 and 4 and objects 1 and 2. The three features are shared by the two objects, each of which is itself a subconcept of this superconcept.

Formal concepts can be represented by bidirectional associative memories (BAMs) (Kosko 1988), in which a neural layer represents objects, another represents features, and the weight matrix between them is set identically to the context matrix (Bělohlávek 2000). Upon presentation of a stimulus consisting of some set of present features, the threshold is set such that the feedforward connections to the other layer will only be able to activate neurons for which all of those features have weights of 1. In this way, only the objects that share all of the input features presented to the network will be activated (Rajapakse and Denham 2005). This also works in the other direction, selecting a set of just the features that are shared by a set of input objects. Figure 6.1 shows this process in effect in a BAM storing a context of data relating to the solar system. The attributes of each object (planet) are stored as positive weights between the object and attribute; activation of an object results in activation of all of the attributes present for that object on the forward step, then on the reverse step all of the objects sharing those attributes are activated.

The BAM learns these hierarchical structures through a 'one-shot' learning rule, in which each of the weight values is inserted into the weight matrix. A more natural learning strategy would be to gradually learn the correct associations between features that make specific objects in an unsupervised fashion, through the analysis of the relational properties of the data over many trials. Of course, the SOM is capable of this, and work has investigated the utility of incorporating a self-organising map layer as the 'object' layer of the BAM (Chartier, Giguère and Langlois 2009; Chartier, Giguère, Langlois and Sioufi 2009). The network consists of a layer

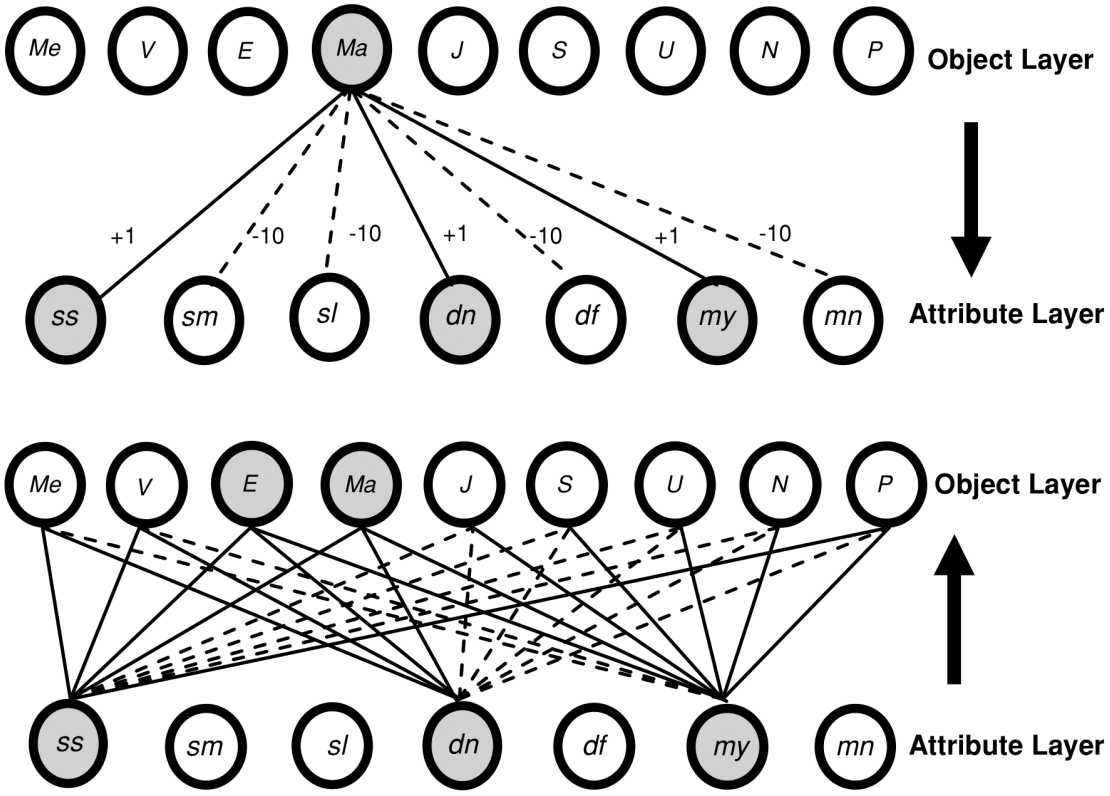


Figure 6.1: The process leading to activation of concepts in the BAM. The example presented here contains a context representing the solar system. The object layer consists of one node for each planet. The attribute layer consists of one node for each feature that a planet object can possess, which are size small, medium or large, distance near or far, and moons yes or no. Connections between the layers are either a small positive value (solid lines) or a large negative value (dotted line). The top half of the figure shows that activation of the Mars object causes, through the positive connecting weights, activation of the size small, distance near, and moons yes attributes. Attributes with negative weights are not activated. The bottom half of the figure shows the second step in the algorithm, which, due to the strong negative weights, results in activation only of objects that possess all of the same features, in this case activating the Earth object. From Rajapakse and Denham (2005).

of neurons representing input features, fully connected to an output layer, arranged through a lateral interaction profile in a two-dimensional grid; effectively a SOM layer. However, there are also feedback connections from the SOM output layer to the input nodes. The weights are random initially, and learn through Hebbian association. A spatial response in the output layer indicates the object selected by the set of input features. However, it is unclear whether a full set of concepts can be learned from this model; indeed, partial input objects are completed by the associative feedback mechanism from the output layer, leading to pattern completion, rather than sub-concept representation. The links between this model and the BAM for FCA context representation suggest that if a SOM layer is capable of generating output representative of the hierarchical structure of its input, this could be a valuable model for extension to representation of full formal contexts.

Other neural network models have attempted to utilise cell assemblies (CAs) to capture the hierarchical composition of conceptual relations (Huyck 2007; Wennekers 2009). A CA is a collection of neurons that co-activate, such that the excitation of one section of the CA will eventually ignite the rest, resulting in pattern completion, a kind of category selection similar to that demonstrated in the model of Chartier et al. (2009). When neurons are members of multiple CAs, the overlap naturally generates a hierarchical relational structure between the categories represented. This property is exploited in (Huyck 2007) in a network that learns, through a Hebbian rule, to respond to a set of binary input features with a particular CA activation in an output layer. It is shown that the correlations between the output representations of objects reflects the amount of feature overlap between those objects. Additionally, a super-concept CA forms in the output layer, representing the features that are common to all of the input objects.

Other models attempt to capture the hierarchical subconcept-superconcept relation between concepts. Among older, symbolic models of concept formation is the ‘CLASSIT’ algorithm (Gennari et al. 1989). Interestingly, this approach makes use of continuous features, essentially adding prototype objects to the representation, each situated at a mean location in the input space, and representing some range of feature values. There are links between this approach and the type of representation that might be seen in classification tasks in the SOM, with output

neurons having prototype values for a range in the space of the input dimensions, and the combination of prototypes for clusters of neurons delineated by sharp changes in the map representing particular categories.

A model that provides an interesting insight into hierarchical category formation is the model of Kim et al. (2008), a competitive two-layer network featuring recurrent lateral inhibition and Hebbian learning, which is used for category learning. In this work the number of categories formed in the output layer of the network emerge as a consequence of the level of impact of excitatory activity upon the neurons causing the lateral inhibition. Greater levels of inhibition lead to a finer granularity of category formation; an input data set is more finely categorised, with smaller regions of the input space represented by separate small regions of the output space. In the case where the input data are binary feature vectors, a finer granularity of category means a smaller number of output neurons represent a smaller collection of input features. This suggests that altering the responsiveness of inhibition in a neural layer presents a method of adjusting the hierarchical level of the concept learned or recalled by a network.

The model described in this chapter is based primarily on the self-organising map model described in chapter 5. Changes to the method of input encoding are made to account for a binary input vector, with a data set of binary object-feature relations being used for the new model. Differences in the structure of the output layer help to facilitate the development of a hierarchical relationship between representations while still maintaining the topographic organisational capabilities of the spiking SOM.

The following sections describe learning and recall in the modified network using input data sets consisting of binary object-feature relations. The results demonstrate the extent to which hierarchical structure is represented in the learned network, the conditions in which it can be learned, and the relationship of this type of learning to existing models of binary feature-object representation, as well as some indication of the relationship between this model and cortical representation. The new model is described in section 6.2. The training and testing methods, the input data sets used to train the model, and the parameters used in the simulations are also described here. Section 6.3 provides analysis of how the trained network represents a

hierarchy. Modifications to the network parameters that allow for either learning or recall at specific hierarchical levels are also presented here, as well as a demonstration that similar results can be achieved using noisy partial input patterns.

6.2 Methods

For the storage and retrieval of hierarchical information from the network, several key changes are made to the model described in chapter 5, introduced below. The functional relationship of these changes is discussed in section 6.2.4, below. Parameters for all equations relating to the model are listed in table 6.2. In spite of these changes, important computational properties of the spiking SOM are retained, such that the new model is capable of a qualitatively similar topographical representation.

6.2.1 Neuron model

The LIF neuron model is retained, modelled by equation (5.1) (on page 70) as in the spiking SOM model. However, in this case the model also features an after-spike relative refractory period, modelled through a transient increase in the firing threshold, according to

$$\tau_\theta \frac{d\theta_j}{dt} = 1 - \theta_j + as_j, \tag{6.1}$$

where s_j is a binary variable representing the instantaneous presence or absence of a spike in neuron v_j, θ_j is the threshold value in neuron v_j, τ_θ is the time constant for decay of the membrane threshold, and a is a constant representing a sufficient multiplier to render the threshold impossible to reach from subsequent PSPs.

6.2.2 Input encoding

The input data for the current model consists of sets of binary feature relationships, similar to the object-feature pairings found in a formal context in FCA (Bělohlávek 2000). As such, input to the new model at each training step consists of an n-dimensional vector of binary numbers. The banks of neurons representing each number in the input vector from the spiking SOM model are removed here, and instead each dimension I_n can be represented by a single neuron.

Each I_n in the input vector could simply be applied as an input current to its associated input neuron u_n. However, assuming equal membrane potentials in all u_i at the initiation of the input, this would result in simultaneous firing for each u_n representing an I_n that equalled one, and no firing for all other u_n. For the STDP rule and topographic mapping properties to be retained from the spiking SOM model, a phase-of-firing coding in which good representatives of the input pattern fire early and poor representatives fire late is essential (see figure 5.6). To facilitate this type of temporal representation of input, the value of the noise term g in equation (5.1) was set to 0.6, causing spike times to be jittered by up to $\sim$4 ms. Additionally, the activation value for all u_n representing zeroes in the input was set to 0.88. Coupled with a firing threshold of 0.5 for all u_n, these values lead to a firing pattern in which the u_n representing ones in the input pattern fire relatively early, and the u_n representing zeroes in the input pattern fire relatively late (between 0 and 8 ms after the latest firing for a u_n representing an I_n value of 1). An example firing sequence resulting from this is shown in the layer u section of figure 6.2.

6.2.3 Output layer

It is shown in (Kim et al. 2008) how the modification of responsiveness of separate populations of excitatory and inhibitory neurons in a layer can lead to the manipulation of the granularity of category formation in both learning and recall. The existing spiking SOM model, however, simplifies the actual cortical anatomy by considering the overall lateral effect of a neuron in a layer. In this case, both excitatory and inhibitory efferents exist exiting from each neuron. To investigate the effects of altering the ratio of excitatory to inhibitory excitability within the layer, the output layer of the new model consists of separate populations of excitatory and inhibitory neurons.

The model retains a 10×10 regularly spaced grid of 100 neurons, now all purely excitatory. A 5×5 grid of 25 purely inhibitory neurons is introduced. Both types of neurons are located on the same grid in the output space, so distances between both populations are calculated according to location within the same grid structure. The inhibitory neurons are interlaced such that each inhibitory neuron is located in the output layer space at a point between four excitatory neurons. The inhibitory neurons are also equally spaced, effectively having twice the distance

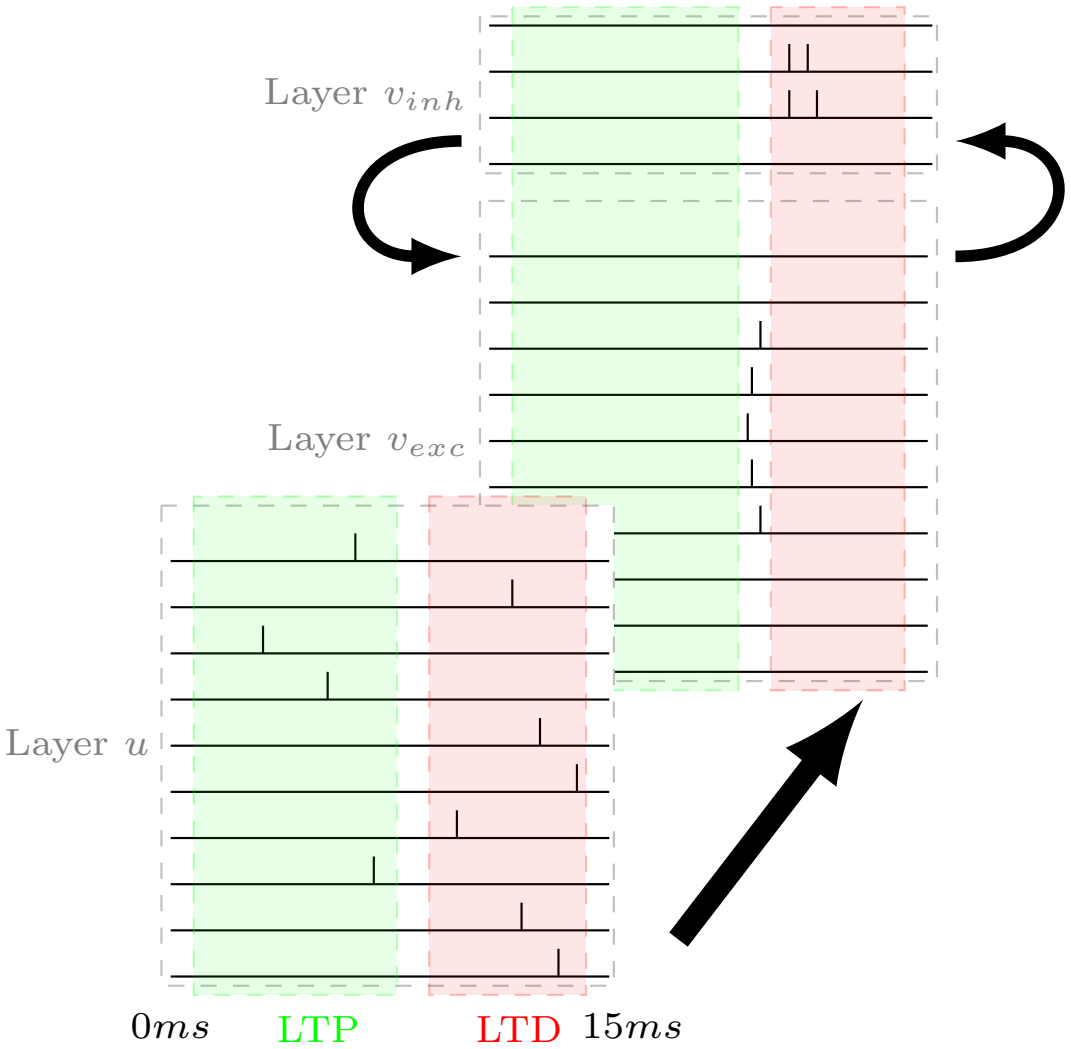


Figure 6.2: The mechanics of learning in the hierarchical representation network. A spike sequence in a bank of neurons in layer *u* represents the binary input vector, with early firing neurons representing values of 1 in the actual input object and late firing neurons representing values of 0. The lower black arrow represents all to all feedforward synaptic connections from *u* to *v*. At some point in the firing of the pattern in *u*, an excitatory neuron in *v* fires, winning the competition. Neighbouring neurons in *v* are caused to fire within close temporal proximity. This excitatory activity activates inhibitory neurons in *v*, shown by the upper right black arrow. Inhibitory firing then feeds back to the layer *v* excitatory neurons, stifling firing in the layer, shown by the upper left black arrow. The gap between the LTP and LTD boxes in layer *u* represents the time at which firing in *v* occurs relative to the firing in *u*. Synapses from any layer *u* neurons that have fired before that point are strengthened (the LTP box) and synapses from any layer *u* neurons that fire after that point are weakened (the LTD box).

between each pair of inhibitory neurons as there is between each pair of excitatory neurons, due to there being fewer inhibitory neurons spread over the same space. Excitatory neurons have synapses with both excitatory and inhibitory neurons in the layer, with the presence of a connection being determined by the distance within the layer between the two neurons. The same applies to inhibitory neurons. The lateral connection weight profile is based on distance in the output layer, with shorter range connections being excitatory overall, and longer range connections being inhibitory overall, as in the spiking SOM model.

To recreate the spiking SOM weight profile with two different types of neurons there is no overlap between the distance dependent profiles of excitatory and inhibitory connections. The maximum distance for an excitatory connection is determined by the same variable that determines the minimum distance for an inhibitory connection, r. As such, the synaptic weight $\bar{w}_{jk}$ between neurons at locations i and j in the grid for connections from excitatory and inhibitory neurons are determined by equations (6.2) and (6.3) respectively:

$$\bar{w}_{jk} = w_{max}^{v_{exc}} \frac{r-d}{r}, \text{ for } d < r \tag{6.2}$$

$$\bar{w}_{jk} = -w_{max}^{v_{inh}} \frac{d-r}{1-r}, \text{ for } d > r \tag{6.3}$$

Here, $w_{max}^{v_{exc}}$ and $w_{max}^{v_{inh}}$ represent the maximum magnitude of excitatory and inhibitory weights respectively, d is the distance between the neurons as a fraction of the maximum possible distance within the layer, toroidal wrapping accounted for, and r is a radius for the grid distance for the maximum distance for excitatory connections and the minimum distance for inhibitory connections.

Both excitatory and inhibitory neurons receive complete all-to-all feedforward connections from the input layer. The connection strengths are initialised randomly as in the spiking SOM model (see section 5.3.1). However, the connection strengths to inhibitory neurons are initially slightly lower than those to excitatory neurons (see table 6.2 for specific parameters), to prevent inhibitory firing in the layer ahead of excitatory firing (which would suppress all output layer

activity in a region). Connections from the input layer to output inhibitory neurons are also not subject to learning.

6.2.4 Functionality

The goal of the changes to the network presented above is to enable the learning of a binary input pattern through an alternative method of input encoding, while also introducing changes to the output layer that allow for manipulation of the learning and recall of hierarchical patterns.

The process through an oscillation in this model resembles that of the spiking SOM model. Neurons that are good representatives of the actual input pattern fire early. Feedforward connections between the layers cause the output neuron with the most appropriate weights to fire first. After that point, more neurons fire in the input layer, which are poor representatives of the actual input pattern. Noise in the membrane potential model in the input neurons causes firing times to be spread across a range, regardless of the fact that all nodes in the pattern take a value of 1 or 0 as an input activation level. This temporal sequence allows steady integration of PSPs in the output neurons, generating a spike fastest in the best representative in the output layer.

The network retains the topographic mapping properties of the spiking SOM through the use of a qualitatively similar lateral connection weight profile. Short range excitatory connections encourage the firing of neurons in a neighbourhood around that winner, both excitatory and inhibitory. The local inhibitory neurons prevent neurons from outside of that neighbourhood from firing, but have no effect on neurons within the neighbourhood. Excitatory neurons within the neighbourhood in turn stimulate each other further. Unchecked, this would generate run-away firing through reciprocal excitation with no inhibition within the neighbourhood. The incorporation of a refractory period for the excitatory neurons counteracts this by raising the firing threshold sufficiently to prevent a second firing of an excitatory neuron until the lateral excitatory PSPs have diminished. Inhibitory neurons are not subjected to threshold adaptation, and so may fire multiple times within an oscillation. The excitatory neuron membrane potential has a lower limit, however, so these multiple inhibitory spikes are not disruptive, and in fact aid in resetting excitatory membrane potentials to a consistent level ahead of the next oscillation.

	Features														
Objects	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0
2	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0
3	1	1	0	0	1	0	0	0	0	1	0	0	0	0	0
4	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0
5	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0
6	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0
7	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0
8	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1

Table 6.1: Simple hierarchical data set. All patterns share feature 1, which represents the top hierarchical level. Features 2 and 3 are shared by exclusive halves of the input objects each, meaning that they represent the second hierarchical level. The remaining features divide the objects into exclusive halves again, generating two more hierarchical levels.

6.2.5 Sample input pattern

A simple hierarchical structure can be formed from three binary features and two input data, if only one feature is shared by those data, and the other two features are only held by one of the data each. In this case, the presence of the shared feature between both data constitutes a superset, and the differentiation of the two data through the other features places them in separate subsets. Extending this approach, an extra layer of the hierarchy will split each of those subsets into two further subsets, with four new features added. Adding an additional eight features results in a simple hierarchy of four levels, consisting of fifteen features split into eight individual objects. This data set is shown in 6.1, and will be used as the main example data set for testing the hierarchical learning and recall capabilities of the network.

Each input pattern contains exactly four 1s. Accordingly, the maximum feedforward weight is set such that, when weights have been fully potentiated, an output neuron will fire after exposure to four input spikes that match the fully potentiated weights.

6.2.6 Network parameters

Table 6.2 shows the parameters used for the new network. The input and output layers, u and v respectively, were setup as described in section 6.2.1. The radius r in equations (6.2) and (6.3) was set to 0.32, and the distances between neurons in the layer were calculated with toroidal

(A) Neuronal parameters					
Δt	$\tau_m^{u,v}$	τ_m^{Inhu}	V_{rest}	θ^u	θ^{Inhu}
0.1ms	1ms	0.5ms	0.0	0.5	0.01
θ^v	tau_θ	g_u			
1.0	0.3	0.6			
(B) Synaptic parameters					
$u - v$		$u \rightarrow Inhu_{exc}$		$u \rightarrow Inhu_{inh}$	
τ_r	τ_f	τ_r	τ_f	τ_r	τ_f
0.5ms	2.5ms	1.0ms	5.0ms	0.3ms	1.5ms
$Inhu \rightarrow u$		$v_{exc} \rightarrow v$		$v_{inh} \rightarrow v$	
τ_r	τ_f	τ_r	τ_f	τ_r	τ_f
0.6ms	3.0ms	0.5ms	2.5ms	0.25ms	1.25ms
(C) Connection parameters					
$w_{max}^{u \rightarrow v_{exc}}$	$w_{max}^{u \rightarrow v_{inh}}$	$w_{max}^{u \rightarrow Inhu_{exc}}$	$w_{max}^{u \rightarrow Inhu_{inh}}$	$w_{max}^{Inhu \rightarrow u}$	$w_{max}^{v_{exc} \rightarrow v}$
4.8	3.8	1.0	1.0	200.0	0.5
$w_{max}^{v_{exc} \rightarrow v_{exc}}$	$w_{max}^{v_{inh} \rightarrow v_{exc}}$	$w_{max}^{v_{inh} \rightarrow v_{inh}}$			
2.0	2.0	2.0			
(D) Neighbourhood parameters					
r %					
0.32					
(E) Learning parameters					
A_+	A_-	τ_+	τ_-		
0.002	0.01	10ms	10ms		

Table 6.2: Network parameters for all simulations described in section 6.3. (A) Neuronal parameters, used in equations (5.1), (5.4) and 6.1 (B) Synaptic time constants, used in equations (5.2) and (5.3) for different synapse types. (C) Maximum magnitudes of synaptic connection strength (D) Lateral connection profile parameters, used in equations (6.2) and (6.3), for layer v. (E) Learning parameters, used in equation (5.5)

structure. This means that excitatory connections cover a radius of 32% of the network width, and inhibitory connections only occur outside of this radius.

Values of $\tau+$ and $\tau-$ in equation (5.5) (on page 73) were both set to 10ms. Again, this width of learning window approximately matches the temporal width of a network oscillation, leading to negligible influence on learning of spikes from neighbouring oscillations.

The parameters for the current model differ from those used in the spiking SOM (see table 5.1) in several ways. Synaptic time constants for feedforward connections between u and v, as well as for lateral connections in v are longer than in the spiking SOM model. This is to account for the greater potential spread of firing times for input nodes; there are effectively pre-defined

firing times for neurons in the spiking SOM input layer, but here there may be larger gaps between spikes. Time constants for the inhibitory chopping circuit in the input layer are also adjusted to account for this increase in variability. Connection weight parameters are scaled to the number of input neurons, with the maximum feedforward weights between u and v increased to account for a decrease in input layer size, and also a decrease in the number of early spikes in the layer that are representative of the actual input value. Learning parameters are skewed in favour of depression, to account for the variability in input neuron firing times causing output neurons to be more likely to fire after a spike from a non-representative input neuron.

6.2.7 Training and testing

The network was trained on the input data set for 4000 training steps, as in the training for the spiking SOM model. At the start of each training step an input pattern was determined by selecting randomly from the patterns available in the given data set. A training step lasted through 10 oscillations of the network (approximately 250ms) before the input pattern was changed. The ability of the network to capture the hierarchical structure present in the sample input data set can be assessed using several methods outlined below.

Learning a hierarchy

Network output was first tested after training with the 8 main input objects, at the lowest hierarchical level. The predicted outcome for this testing is that, if the network is capable of representing a particular level of the hierarchy, network output for presentation of a stimulus from that level of the hierarchy should correlate with the sum of network output from all individual members of that superset. For example, in the current data set (see table 6.1), the superset consisting of only objects 1 and 2 would consist of features 1, 2 and 4, but not 8 or 9. The prediction states that the output from presenting features 1, 2 and 4 only to the network should correlate with the sum of the output from presenting objects 1 and 2 (which do include features 8 and 9, respectively).

In the testing phase the output activity of the trained network was recorded over 100 trials with each of the original input patterns, along with the supersets from each hierarchical level. The

	Features														
Patterns	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
5	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0
6	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0
7	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
8	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0
9	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0
10	1	1	0	0	1	0	0	0	0	1	0	0	0	0	0
11	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0
12	1	0	1	0	0	1	0	0	0	0	0	1	0	0	0
13	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0
14	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0
15	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1

Table 6.3: The simple hierarchical data set with all supersets of the original 8 objects included, to generate 15 testing patterns. 7 additional testing patterns are added to the training data set shown in table 6.1. Each of these new patterns consists of just the features that constitute a hierarchical level, with pattern 1 representing the top hierarchical level by only consisting of feature 1, patterns 2 and 3 representing the second hierarchical level by also incorporating features 2 and 3 respectively, and so on, such that all three levels above the lowest represented.

original 8 input patterns are paired into 4 supersets, by removing any 1s present for features 8 to 15 from the patterns. This renders objects 1 and 2 indistinguishable, and so on, resulting in 4 further distinct input patterns. The same principle can be applied to features 4 to 7 to generate 2 further patterns, and again to features 2 and 3 after that, to generate one final pattern corresponding to the superset of which all of the original 8 patterns are members. This results in a total of 15 patterns presented to the network throughout testing, shown in table 6.3.

Learning supersets using subset input

Hierarchical structure may exist within an input data set to a depth below that which a particular network configuration is capable of learning. The prediction in this case is that the excitability of neurons in the output layer should play a role in determining the maximum depth of hierarchy that will be learned by a network. If output neurons are more excitable they will fire earlier, requiring firing of fewer input neurons to activate. This will essentially shift the LTP and LTD windows in figure 6.2 to the left. As a result, less features will have their weights potentiated,

resulting in the learning of less fine-grained distinctions between input patterns, and a loss of information at lower hierarchical levels.

This prediction was tested by varying the firing threshold of layer *v* neurons. Training was conducted as above, using the original 8 objects as the training data set, and using different firing thresholds in the training phase for different trials. The threshold ranged from 0.2 to 0.8 of the original threshold used in the previous training, at intervals of 0.2, with 10 trials conducted for each value. Testing then consisted of the recording of average network output, as above, for the 8 original input objects, with the expected outcome that lower thresholds (more excitable neurons) result in more homogenised output activity patterns.

Recalling supersets using subset input

A further test of the extent to which hierarchical structure is represented in the network is to examine whether the network can produce a response representative of a higher hierarchical level when presented with a more specific input. This type of behaviour is examined in Kim et al. (2008) and is achieved through the use of a top down input to modify the responsiveness of inhibitory neurons in the output layer. As such, it is predicted that similar behaviour will be present in the current model.

Reducing the responsiveness of the inhibitory neurons in layer *v*, relative to the responsiveness of the excitatory neurons, should represent a method of recalling output patterns within a single oscillation that more accurately approximate a superset representation. Lower responsiveness will result in an inhibitory neuron firing later in the phase relative to the excitatory neurons. This late response allows time for excitatory neurons to fire that would otherwise not receive enough input to reach threshold before becoming inhibited. After the firing of the winning excitatory neuron, reciprocal lateral excitatory connections become the dominant connections within the network. In this scenario, later firing of inhibitory neurons simply results in a wider radius of neighbourhood activity around the winning unit. Neurons sharing a superset with the actual input object will already have received input from several of their preferred input neurons, priming them for firing. These neurons are also, due to the topographic mapping properties of the layer, relatively close to the winning neuron. As such, the extra time taken for inhibitory

firing results in firing for, on average, neurons that represent the remainder of a superset of the actual input object.

Testing of this prediction was conducted on the network after training as above, using the original 8 objects as the training data set. In the testing phase, the levels of two variables were modified, $w_{max}^{v_{exc} \rightarrow v_{inh}}$, the connection strength from excitatory neurons in v to inhibitory neurons in v, and $w_{max}^{v_{inh} \rightarrow v_{inh}}$, the connection strength reciprocally from inhibitory neurons in v to other inhibitory neurons in v. The former variable controls the phase lag between responses of the excitatory and inhibitory populations, with a reduction resulting in an increased lag. The latter variable is reduced so that approximately the same number of inhibitory neurons fire overall in an oscillation, after reducing excitatory strength in those inhibitory neurons. The expected outcome is that reducing the inhibitory responsiveness will result in an output representation that more accurately reflects expected firing patterns for a higher level of the hierarchical structure than that on which the network was trained.

Partial input patterns

Exposure to all of the features that constitute a particular object at every presentation of that object is not a reasonable criterion to be necessary for learning of a context. Conceptual representations may be built up over time from multiple exposures to parts of objects that reveal different information about the feature combinations present. The current model should be capable of reproducing all of the above predictions using training patterns that have features removed stochastically from the actual input objects.

This was tested through further simulations, replicating all of the training and testing methods described in this section, but with an alternate input pattern selection process, as follows. In a training step an input pattern was selected at random, as normal. Each feature in the pattern is then subject to a 0.25 chance of removal, and replacement with a zero, resulting in an average of 1 feature being absent from each training pattern. The prediction in this case is that, despite a random amount of features being absent from each pattern in the training process, the network will be able to produce results that are equivalent to those resulting from the testing processes described above.

The next section will present the results of the manipulations to the network parameters described in this section. First, the learning of the hierarchical structure present in the input data set will be demonstrated. Second, the effect of firing threshold for the membrane potential in the output layer neurons on the structure of the learned patterns will be demonstrated. Third, the effect of the responsiveness of inhibitory neurons in the output layer on the structure of output during recall will be demonstrated. Fourth, the above results will be replicated using input patterns with stochastic feature removal. The combination of this experimentation will provide an overall framework for learning and recall of specific desired aspects of a hierarchical structure in the current network.

6.3 Results

The following sections detail the results of learning in the modified network using input data sets consisting of binary object-feature relations. Four separate predictions were tested here, as described in section 6.2.7. The results demonstrate the extent to which hierarchical structure is represented in the learned network and the conditions in which it can be learned.

6.3.1 Learning a hierarchy

The final weight matrix after training with the hierarchical data set (table 6.1) is shown in figure 6.3. The learned weight matrix approximately matches up with the input data. The original input structure can be seen across the weights, with ranges of output neurons across the two-dimensional output space having weights that correspond to individual objects in the input space.

The testing phase consisted of the presentation to the trained network of the 15 input patterns shown in table 6.3, as described in section 6.2.7. The result of this testing is shown in figure 6.4. Each square shows a representation of the excitatory neurons in the output layer, colour coded by how often they fire in response to an input pattern. Testing pattern 1 shows the largest superset, with just one input feature present, feature 1. Patterns 2 and 3 are the second layer of the hierarchy, and patterns 4 to 7 are the third layer of the hierarchy. The final 8 patterns shown, 8-15, are the original 8 training patterns at the lowest level of the hierarchy. It can be seen that

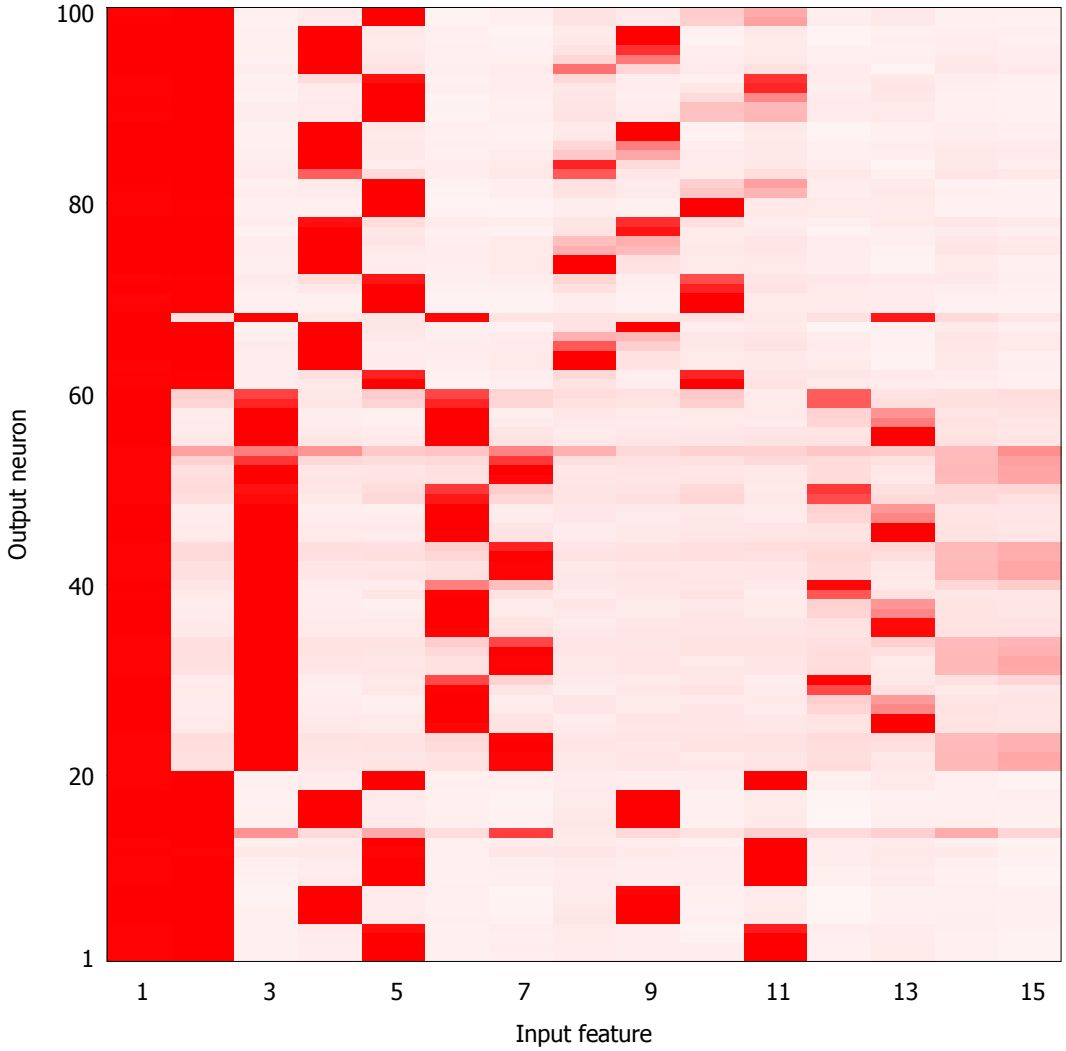


Figure 6.3: Weight matrix for connections from the 15 input neurons (columns), representing 15 features of the input pattern, to the 100 excitatory neurons (rows) in layer v. Colour shows weight, with a darker rectangle representing a stronger weight. All output neurons respond to feature 1, which is included in all input patterns that the network has seen. Appropriately diminishing number of output neurons have strong weights to further input features, depending on the level of hierarchy that the feature resides at. Each column of 100 output neurons is grouped by rows in the layer v output space; the first ten weights are to neurons in the first row of the output layer, the second ten weights are to neurons in the second row of the output layer, and so on.

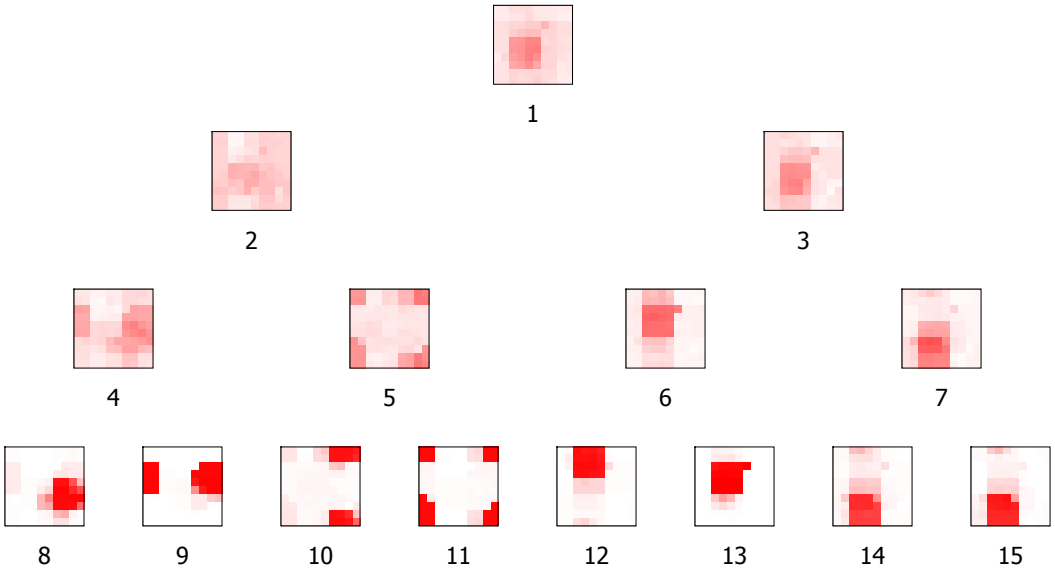


Figure 6.4: Percentage of 100 testing trials on the same learned network for each input pattern for which each excitatory output layer neuron fired. Colour represents percentage values, with darker showing a higher value. Patterns are numbered according to the input pattern used to generate that output activity pattern; input pattern represents the highest hierarchical level, containing only feature number 1; 2 and 3 represent the next level, also containing features 2 and 3; and so on through the input data hierarchy.

these 8 patterns have the most disjoint network responses, with patterns at increasingly higher levels of the hierarchy having 'blurred' network responses.

The relationship between the 'observed' and 'predicted' network activity for supersets was established. 'Observed' activity is the actual output for an input pattern representing a superset. 'Predicted' activity is the sum of the output for all immediate members of that superset (i.e. the patterns exactly one level below in the hierarchy that also contain all of the features of the superset). For each superset, the observed firing likelihoods for all excitatory nodes in v (see figure 6.4) for that input were normalised as a vector and compared with the normalised predicted firing likelihoods for that superset, using Pearson's linear correlation coefficient, r, with significance values of p.

The training and testing phases were conducted for 20 separate trials, and the average correlations between observed and predicted patterns are shown in figure 6.5. Red represents a positive

correlation and blue a negative correlation, with the magnitude of the correlation represented by colour intensity, such that white represents no correlation at all. Predicted activity for patterns 8 to 15, at the lowest level of the hierarchy, are simply taken as identical to the observed patterns, as there are no sub-patterns for their activity to be predicted; correlation values along the central diagonal for these patterns are therefore 1. Adjacent pairs of patterns (e.g. 8 and 9) have slightly positively correlated output (mean r = 0.32), as these have the most overlap, differing by only 1 feature. Pairs of patterns that share only 2 common features (e.g. 8/9 versus 10/11) have slightly negative correlations in output (mean r = -0.14), with more negative correlations (mean r = -0.22) between pairs that share only 1 common feature (e.g. 8 to 11 versus 12 to 15). This result shows the ability of the network to map similarities between actual input patterns to the degree of similarity between resultant output patterns, demonstrating the retention of topographic mapping characteristics in the current network. The lateral connection profile ensures that responding neurons in the output layer are grouped in an approximate circle, so greater overlap between response patterns must be through adjacent regions of activity.

The correlations displayed between observed and predicted activity for patterns 1 to 7 are using predicted values, as described above. The central diagonal line of strong correlations indicates that the predicted output activity for a superset correlates strongly with the observed activity for that superset. The weaker values around that diagonal indicate that predicted output activity for a second, different superset correlates less well with the observed activity for the first superset. Again, pairs of patterns that are different by more features than they are similar exhibit negative correlations (mean r = -0.37, mean p = 0.023) between predicted and observed activity (e.g. the predicted activity for pattern 4 correlates negatively with the observed activity for pattern 7). No real correlation (mean r = -0.08, mean p = 0.341) is seen between the predicted and observed activity for pairs that differ by just one feature at this level (e.g. predictions for pattern 5 and observations of pattern 4).

A more detailed depiction of the correlations at the second level of the hierarchy is shown in figure 6.6. Each cluster of bars shows how well the predicted activity for each of the patterns at that hierarchical level correlates with the observed activity for one pattern. The height of the

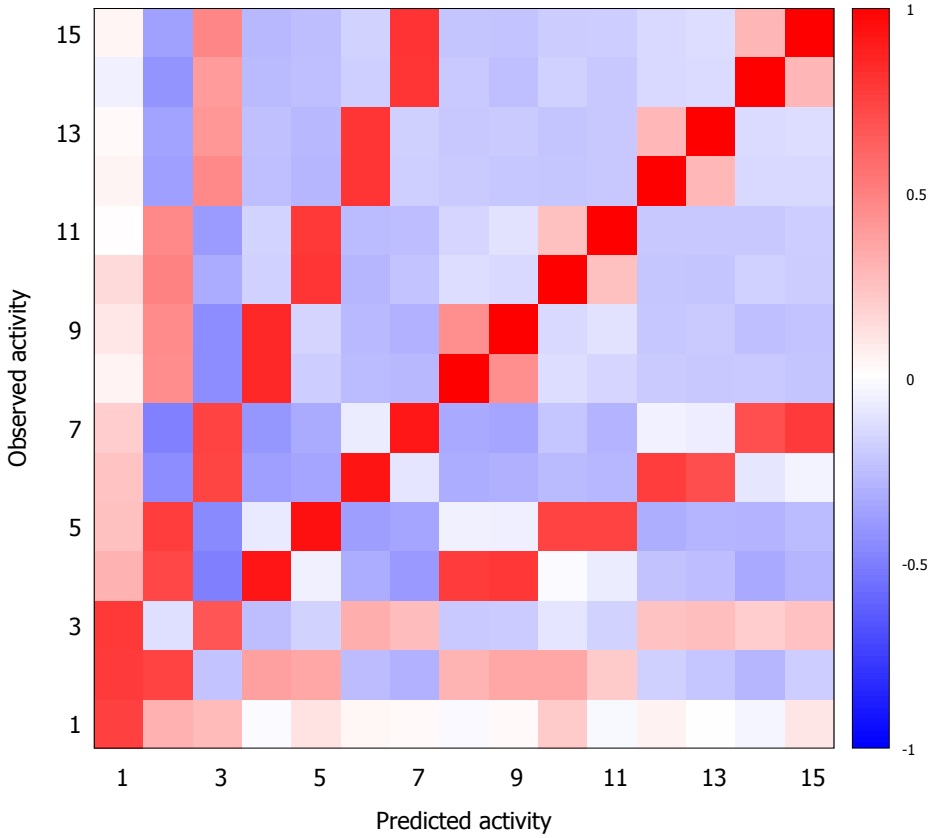


Figure 6.5: Correlation values between observed firing patterns in the network and predictions for firing patterns at higher levels based on those observations. Predicted activity is the sum of the output for all immediate members of that superset. Predicted activity for patterns 8 to 15, which are not supersets, is just the observed activity for that input pattern. The main observation is that observed activity when using a superset as input correlates with the sum of observed activity for all of the immediate subpatterns of that superset. For more detail see the text.

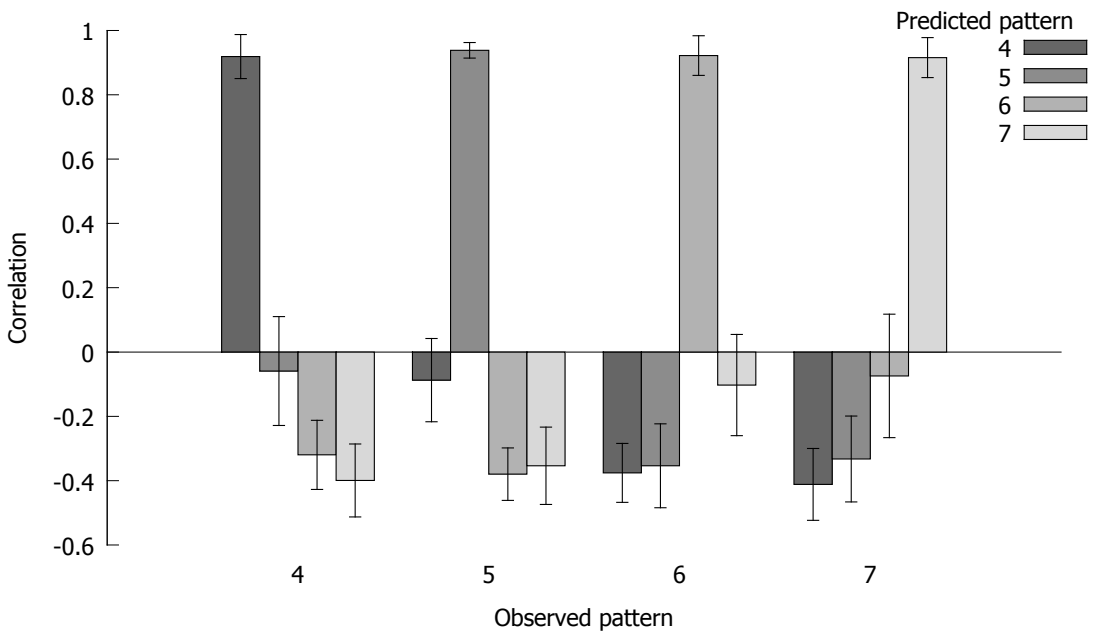


Figure 6.6: The correlation values between observed firing patterns for input patterns at the second level of the hierarchy and predicted second level firing based on observed activity for the subsets of those patterns. High correlations are seen between observed supersets and the sums of output activity from the subpatterns that make up that superset. Anti-correlated output activity is seen between input patterns that share fewer than half of the features of the subpatterns used to make a prediction.

bar represents the mean correlation and the error bars represent standard deviations. It can be seen that, for each observed pattern, the predictions generated from observations of activity of the members of that superset are correlated significantly more highly with the observed output than the predictions generated by activity from members of other supersets are. Importantly, the four autocorrelations are highly significant in their own right (mean $r = 0.92$, mean $p < 0.001$), demonstrating that output generated by members of a superset is a reliable predictor of the output generated by the superset itself.

Interestingly, the predictions for patterns at a hierarchical level several levels below a given superset also have positive correlations with observed activity for that superset, and negative correlations with observed activity for a superset for which they are not a member (see, e.g., predictions for patterns 8 to 15 compared with observations for patterns 2 and 3). This indicates that the hierarchical structure of the output is preserved throughout network responses to input patterns from different hierarchical levels down the full depth of the hierarchy. As such,

predicted activity for any pattern is never correlated negatively with observed activity for the superset pattern 1, of which all other 14 patterns shown are members.

The trained output layer contains regions that have learned to respond to each of the patterns at the lowest level of the hierarchy. If an incomplete pattern, representing a superset of input objects, is presented to the network, all of the neurons with strong weights to features that are present will become excited by the strong input. These neurons are in fact the neurons that respond to input objects that are members of that superset. The early input spikes in an oscillation are reduced in number, due to the reduced number of features present in a superset compared to the full objects. At this point, noise becomes the determining factor in the decision of which output neuron fires first, with a substantial advantage held by those neurons excited by the early spikes, representing present features. Once a best matching unit is chosen, the lateral connection profile initiates firing in an output region; this region will represent an individual member of the superset. Effectively, the network has no information determining exactly which learned object is present based on the features shown, only that it must be one from a particular set. At this point, the next random piece of information the network receives determines the response of the layer, and a random member of the superset is selected. Repeating this experiment multiple times results in the observed firing for a superset approximating the average firing of the member objects.

6.3.2 Learning supersets using subset input

The second prediction to be tested was the prediction that lowering the output neuron firing threshold throughout training would result in the learning of merged representations at lower hierarchical levels, meaning that the network output only represents supersets of the input it has been trained with. Figure 6.7 shows the average responses for networks trained using each firing threshold, in response to the 8 original input patterns. As the threshold decreases, responses to different input patterns homogenise. A threshold at 80% of the original value (bottom row of 6.7) results in responses to input objects being identical for pairs of patterns that have a superset one level of the hierarchy above, but still substantially different for pairs of patterns that do not. Essentially there is no longer a differentiation in output firing for two distinct objects, as

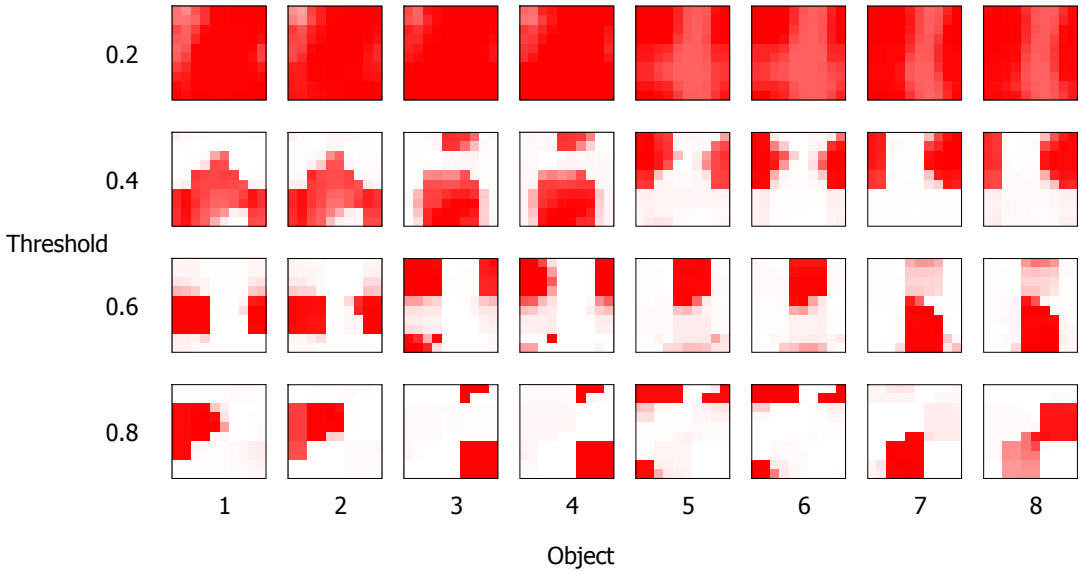


Figure 6.7: Average firing over 100 testing trials for each of the 8 original input patterns for 4 networks trained at different levels of output neuron threshold. Colour represents percentage values, with darker showing a higher value. Results are demonstrated for training with thresholds at values of 0.2, 0.4, 0.6 and 0.8 of the original threshold, while training with the 8 original input objects, patterns 8 to 15, at the lowest hierarchical level. A low threshold results in homogenised activity, effectively representing only the highest hierarchical level. Heterogeneity of network output patterns increases as the firing threshold is increased, with output pattern separation occurring in line with the hierarchical structure.

long as they share a sufficient number of characteristic features (75% in this case). Lowering the threshold increases the gap between input objects and output representation in terms of the number of hierarchical levels that are represented. A threshold at 20% of the original value merges output representations to the extent that only 2 distinct representations remain, which even have significant overlap between them. The second hierarchical level is the maximum that is represented in the output map, despite training using input objects incorporating the full hierarchical structure.

Learning to different levels of the hierarchical structure by modifying the firing threshold is enabled by the use of STDP. A low threshold results in output layer firing at an early phase. Fewer input neurons will have fired by this point, meaning that fewer connections will be strengthened and more will be weakened. At a low threshold the output firing occurs at a point in the phase before the input neurons representing 1s in the input pattern have all fired. The stochastic com-

ponent to the input neuron firing means that a random selection of neurons representing 1s for a given input pattern will have fired ahead of output firing. Input features that indicate larger supersets at a higher hierarchical level are present in a randomly selected input firing pattern more often than those unique to individual objects at the lowest level of the hierarchy (e.g. feature 1 in the current data set is present in 100% of randomly selected input patterns, features 2 and 3 are present in 50% of patterns, etc.). As a result, connections from features indicating larger supersets will be potentiated a larger percentage of the time than features indicating subsets. In this way, a change in firing threshold leads to a change in weight distribution in favour of features indicative of high hierarchical levels relative to those indicative of low hierarchical levels.

6.3.3 Recalling supersets using subset input

The third prediction tested was that reduced responsiveness of inhibitory neurons in the output layer will result in output activity representing a superset of the presented input pattern, despite the network learning using only those subset input patterns in the training phase. The recall of superset patterns after learning with members of those supersets has been demonstrated in the current network through the recording of averaged output activity over multiple trials (see section 6.3.1). A less consistent network response indicates a higher level of the hierarchy, although the inconsistency will only be within areas of the output map that represents lower levels in the correct subsets of the hierarchy. A focus of the current work has been to process information within a single oscillation. A single oscillation of the network in response to a superset does not provide any indication that the superset has been provided as input rather than a lower level object.

The network was trained using the 8 lowest level input patterns, as described in section 6.2.7. Figure 6.8 shows average firing after 100 testing trials for each of the 8 input objects in the hierarchy, at three different levels of inhibitory responsiveness. The top row shows the response with the values set to the same level as during learning. Seven unique output patterns emerged in this case, with the final two input patterns being merged in this particular training trial. The middle row shows recall using the same trained network and input patterns, but with the in-

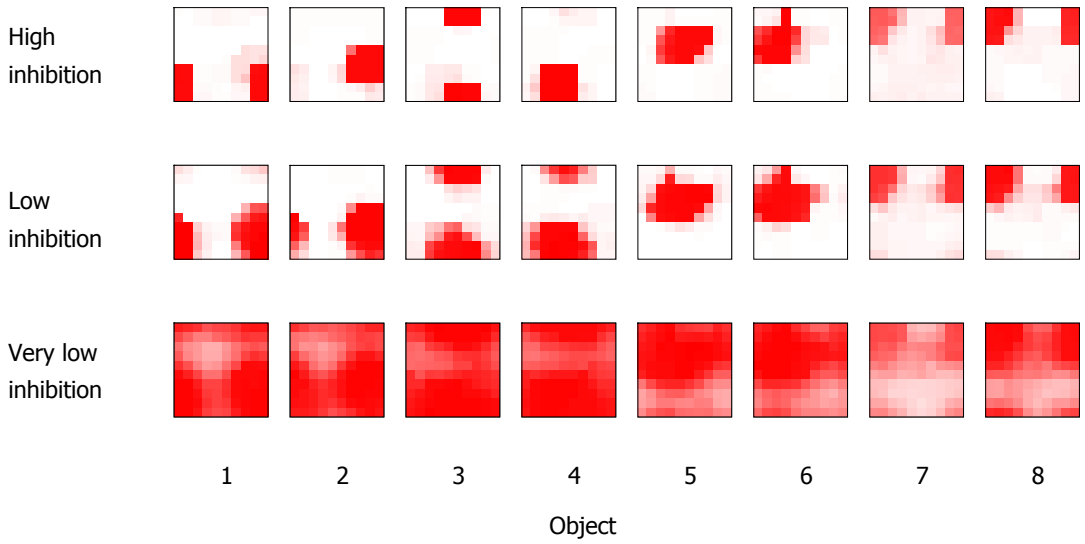


Figure 6.8: Percentage of 100 trials for each input pattern that each excitatory output layer neuron fired for. Colour represents percentage values, with darker showing a higher value. Testing has been conducted on one trained network, trained using the 8 original input objects at the default (high) level of inhibitory responsiveness. The top row shows that the network has learned distinct firing patterns for the original input objects, using the default level of inhibitory responsiveness. The middle row shows the network output when using 0.3 of the original level of inhibitory responsiveness; distinct firing patterns merge, thereby representing a higher hierarchical level. With a very low level of inhibitory responsiveness (0.1 of original) the network output represents a higher hierarchical level again.

hibitory responsiveness reduced to 0.3 of the original value. Only four distinct output patterns are present, with each pair of input patterns merged into one output region. The bottom row shows network response with inhibitory responsiveness divided by 10. Clear distinctions between any input patterns are no longer visible, and the network output may be representing the top level of the hierarchy with this duration of phase lag for inhibitory firing.

Analysis was conducted to examine the accuracy with which the firing in one oscillation recreates the firing that would be predicted for a superset, given the observations of firing for members of that superset. Reduction in inhibitory responsiveness should result in an output firing pattern that recreates the sum of output firing for members of a superset, if that superset is being represented by the new output pattern. Figure 6.9 displays example output vectors from another trained network, for the 8 input objects at the lowest level of the hierarchy, in three separate

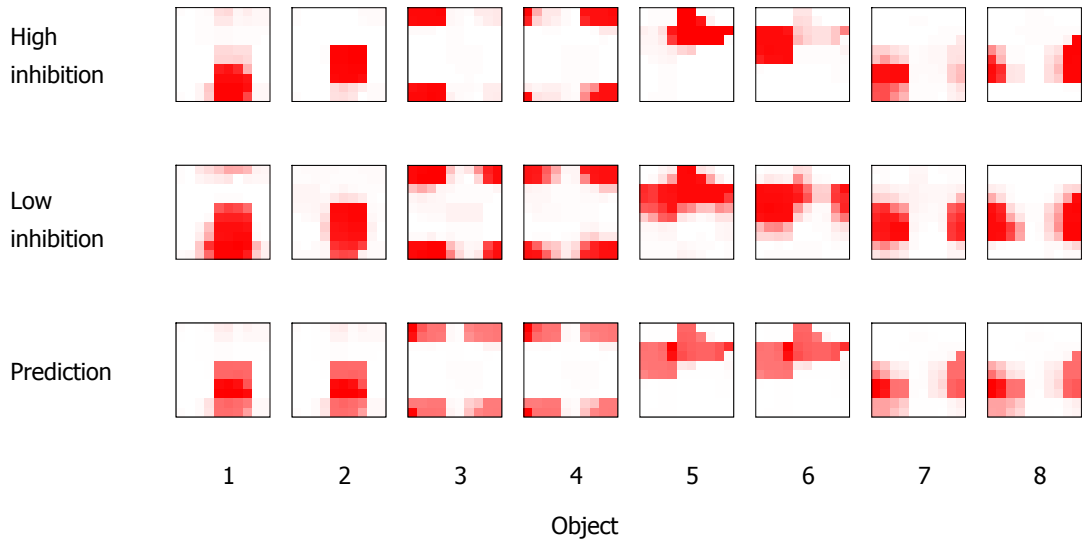


Figure 6.9: Percentage of 100 trials for each input pattern that each excitatory output layer neuron fired for. Colour represents percentage values, with darker showing a higher value. Testing has been conducted on one trained network, trained using the 8 original input objects at the default (high) level of inhibitory responsiveness. The top and middle rows show the same effect as in figure 6.8. The bottom row shows predicted output activity from summing members of supersets for each pair of objects; "low inhibition" output more accurately reflects these predictions than "high inhibition" output.

conditions. The bottom row shows the predicted network response to the second hierarchical level, from the sums of responses to all objects that are subsets of the 4 second level patterns. The middle row shows the firing likelihoods in the output layer in response to each input pattern when the network is tested using reduced inhibitory responsiveness (divided by 0.3), and the top row shows the response when testing using the default inhibitory responsiveness (divided by 1.0). The figure demonstrates the same merging of sub-categories when there is a temporal lag in inhibitory firing seen in figure 6.8. Comparisons between each of the top two rows and the bottom row shows that with reduced inhibitory responsiveness the firing likelihood in any given oscillation more accurately reflects the predicted likelihood pattern for a subset one hierarchical level above the actual input object.

This result is confirmed through a t-test for significant difference between mean correlation values across 20 trials. The mean correlation between the predicted firing likelihood patterns

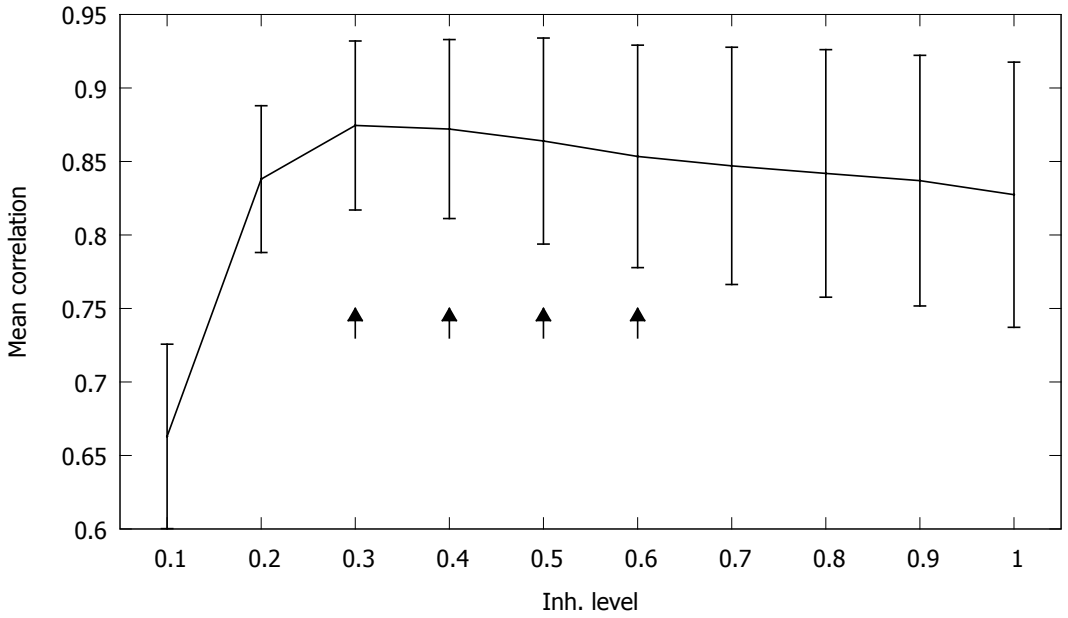


Figure 6.10: Mean correlations between observed network activity in response to the 8 lowest level input patterns, and predictions of activity for the 4 supersets at the next level up in the hierarchy. Mean and standard deviations (error bars) are shown of 20 trials each for 10 levels of inhibition relative to the default level at which the predictions were made. Arrows indicate significant ($p < 0.05$) increases in level of correlation above that seen with the default level of inhibition. A significant increase in the ability of the network response to represent the sums of subpatterns when presenting the network with one of those subpatterns is seen for a range of decreased inhibitory responsiveness levels.

for supersets at the second level of the hierarchy and the observed responses to the 8 lowest level input patterns when using the default level of inhibitory responsiveness is 0.827, significantly less ($p < 0.001$) than 0.875 when using a lowered inhibitory responsiveness level. A parameter test was conducted to identify the range of inhibition levels resulting in increased correlation of observed activity in response to the lowest hierarchical level with predicted activity at the next level up the hierarchy from the activity measurements using the default level of inhibition. The 20 simulation runs were carried out for 10 different inhibition levels, scaling the default level of inhibition that the network had learned at by values in the range [0.1,1.0] in increments of 0.1. The results of this testing are shown in figure 6.10. Significant increases ($p < 0.01$) in correlations are seen for scalings of the default inhibition value between 0.3 and 0.6.

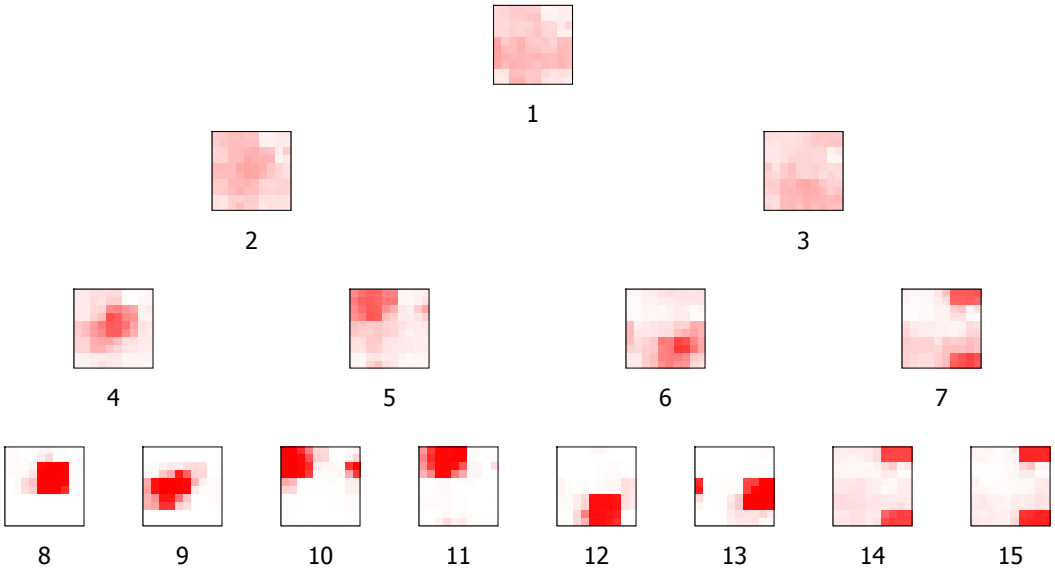


Figure 6.11: Recreation of figure 6.4, but for partial input patterns generated through stochastic removal of input pattern features. Network output is still capable of representing higher levels of the hierarchy than that used for training after features are removed from the training objects.

6.3.4 Results are replicable using partial input patterns

The current model is capable of reproducing all of the results described above using an input pattern selection method that temporarily removes features stochastically from the actual input objects. This fourth prediction was tested by replicating all of the above simulations while using the alternate input pattern selection method described in section 6.2.7. Figure 6.11 shows the average output firing for each output layer node after training, as in section 6.3.1.

Training was also conducted with multiple firing threshold levels in layer v, as described in section 6.3.2. Threshold modification remains capable of altering the number of levels of hierarchical structure installed in the network with partial input patterns, as demonstrated in figure 6.12.

Furthermore, figure 6.13 demonstrates that recall of supersets at multiple levels is still possible through modification of the inhibitory responsiveness in layer v, as in section 6.3.3.

Learning is slow and based on accumulation of coincidences over time, through multiple pre-

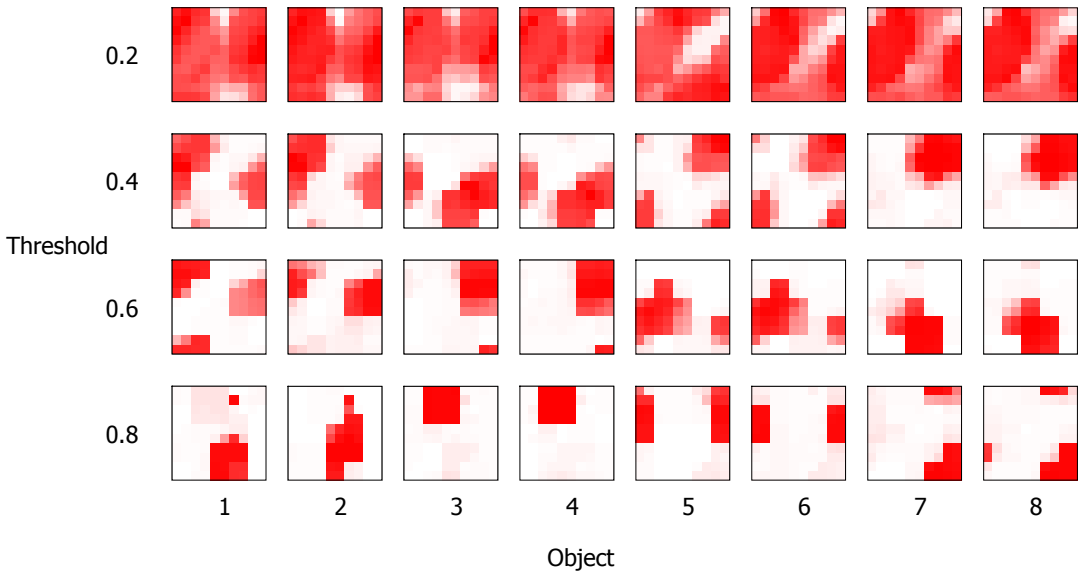


Figure 6.12: Recreation of figure 6.7, but for partial input patterns generated through stochastic removal of input pattern features. Network output still becomes more distinct in response to distinct input objects when training with increased firing thresholds after features are removed from the training objects.

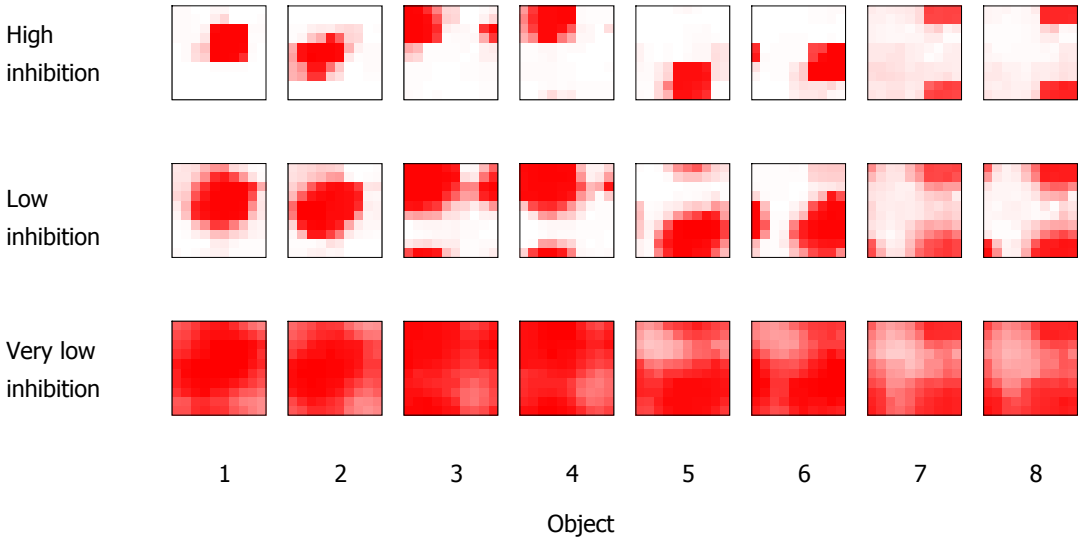


Figure 6.13: Recreation of figure 6.8, but for partial input patterns generated through stochastic removal of input pattern features. Supersets can still be recalled using subset input by reducing the level of inhibitory responsiveness after features are removed from the training objects.

sentations of input patterns. Pairs of input neurons that fire frequently together at an early phase will on average both have their weights potentiated for an output neuron that represents them, without necessarily having to fire together on every possible occasion.

Of course, an increase in the likelihood of an input feature being absent will inevitably lead to degradation of the learned patterns at a certain point, due to their being on average a lack of simultaneous activation of a pair of input neurons that are actually related in the original input data. A prediction concerning this is that representation of the lowest levels of the hierarchy will be the first to degrade as the likelihood of a missing feature increases. The input features discriminating additional sub-categories at low hierarchical levels occur less often than those discriminating super-categories at higher levels. This should result in a relative lowering of the average synaptic weights for features at the foot of the hierarchy, resulting in output regions that fire in response to more than one distinct member of a superset.

6.4 Summary

The model described in this chapter is based on the spiking SOM model of chapter 5, but features several modifications to make it capable of the learning of a hierarchical data set and the manipulation of that learned representation. A method of coding for binary input patterns that is compatible with the use of STDP and the nature of learning present in the spiking SOM is introduced to the model. The output layer features separate excitatory and inhibitory neurons that are reciprocally interlinked in a grid structure, and a relative refractory period has been added to the excitatory neuron model to prevent mutual excitation causing runaway activity.

These changes allow the model to learn the hierarchical structure present within a binary set of feature-object relations, as demonstrated by the results in section 6.3.1. Positive correlations are seen between average population responses to pairs of patterns that share the majority of their features, with negative correlations seen between population responses to pairs that share few features. Sums of population responses for subsets of a particular superset pattern correlate extremely highly with the population output in response to that superset.

Modifications to the excitability of the neurons in the output layer and to the ratio of excitabil-

ity between excitatory and inhibitory neurons in this layer allow the nature of the learned or recalled representation to be altered. Reducing the firing threshold for all output layer neurons during training results in less discriminable population activity patterns in response to distinct input patterns. Responses to input patterns merge in such a way that higher levels of the hierarchy are represented by the output. Reducing the level of inhibitory responsiveness during recall results in population activity patterns that represent supersets of the patterns presented as input. A potential combination of adjustments to the threshold during training and the inhibitory responsiveness during recall allows the level of the hierarchy that the network learns and recalls at to be customised.

Additionally, the model does not need to be trained using complete input objects and is capable of learning representations equivalent to those described above using patterns with missing features. This stochastic feature removal, along with the stochastic determination of firing times in the input layer, demonstrates the resilience of the network to noisy input and shows the importance of an approach that can make use of statistical correlations in the input data over multiple trials.

Chapter 7 will highlight the significance of the results obtained from the networks described in both this chapter and in chapter 5.

Chapter 7

Discussion

The aim of this thesis (see section 1.1) was to implement spiking neuron networks for generic cognitive functions. Specifically, the main cognitive function tackled was the formation of neural maps, a feature of cortical processing found to be generic across several sensory cortical areas (see section 2.4). The implementation of this feature was extended through the creation of a network capable of recreating some of the hierarchical object representation properties of higher-level sensory cortical areas (see section 2.5). Importantly, a key principle of the approach taken by the current work was to ensure that several common cortical properties were utilised (see section 2.3). A further principle was to demonstrate that useful computational principles of existing artificial intelligence algorithms for these types of cognitive processing were maintained in their more biologically relevant spiking neuron implementation (see chapter 4).

These aims have been achieved through the models described and analysed in chapters 5 and 6. This chapter will discuss the relation of both models to existing literature in spiking neuron modelling, other neural network modelling, neural network biology, and artificial intelligence applications. Finally, the importance of these models as building blocks for the generation of a potential suite of cognitive modules will be discussed, and some directions for future research will be provided.

7.1 Spiking self-organising map model

The spiking SOM model (see chapter 5) provides a valuable contribution to the field of computational neuroscience models of cognitive processes. It is the first reliable, robust spiking neuron implementation of the generic SOM algorithm that incorporates observable properties

of cortical function.

7.1.1 Relation to existing models

The current spiking SOM model was initially based on the model of Ruf and Schmitt (1998), but in the current form represents an improvement on and advancement of that work in multiple ways. The incorporation of continuous input and oscillatory firing means that the current network does not need resetting at any point. The change to STDP for learning and introduction of α-function PSPs both contribute to an improvement in the biological plausibility and performance of the model. Additionally, testing has established the robustness of the new approach to changes in the learning parameters, noise level, and input data set.

Spiking SOM variants were introduced in section 4.2.3. Two networks describe improvements made to the model of Ruf and Schmitt (1998), in the form of a learning rule with a rule that relies on a pre-existing optimal weight derivation function to select weight changes (Panchev and Wermter 2001), and biologically relevant action potentials (Svetlik 2006). The modified learning rule relies on prior knowledge of the optimal weight distribution, and no attempt is made to model a biologically plausible learning mechanism, local to individual neurons, such as STDP. The neuron model change made in the model of Svetlik (2006) is similar to one change made in the current spiking SOM, in that realistic PSPs are introduced. However, this model does not make any of the other improvements made in the current model, retaining the other aforementioned drawbacks of the Ruf and Schmitt (1998) model. Furthermore, none of these three models attempt to demonstrate the robustness and reliability of resultant map formation. They completely lack an analysis of topographical mapping to changing input distributions, as well as the categorisation performance presented here. The failure to demonstrate the retention of the computational properties of the traditional SOM algorithm is also a major problem with other related networks. In particular, the SOM-like network of Sala et al. (1998) demonstrates that, using an appropriate lateral connection structure, either of two Hebbian learning rules can result in a segregation of the output space that appears representative of input space segregation. The current spiking SOM model demonstrates greater behavioural depth, moulding the output space in a way that topographically maps the input space, with categorical data resulting in

representative output clusters.

Several other models focus on implementations of a spiking SOM that replicate properties of specific sensory cortical areas. The current model is untested with structure and input that is related to a specific sensory modality. A somatosensory cortex model is constructed by Cios et al. (2004), and a visual cortex model is constructed by Veredas et al. (2008), both of which use spiking neuron networks organised hierarchically through multiple SOM-like layers. A limitation of the current spiking SOM model is the difficulty in connecting multiple SOMs together into a feedforward hierarchical structure. The lateral connection profile causes the representation in the output layer to be spatial, with a very narrow temporal window, relative to the input representation. This means that the same system of self-organisation will not work if that SOM layer output is taken as input to an additional downstream SOM. Spreading the firing in the SOM layer into a temporal code (for example by adjusting synaptic time constants) would cause self-organisation to fail in the initial SOM, as neighbouring neurons in the output layer need to fire in close temporal proximity to ensure their weights are adjusted towards the same point. Of course, multiple independent SOMs could each activate their own spatial patterns, but with a temporal sequence between the SOMs, which could in turn become input to a downstream SOM layer, creating a kind of multi-modal association or integration of various spatially coded features; this is not, however, exactly the kind of hierarchical construction demonstrated in the above mentioned research. The model of Veredas et al. (2008) in particular demonstrates reliable formation of receptive fields and orientation selectivity. Replication of these results would not be possible in the current model if a similar network structure was used.

However, other spiking neuron models of cortical map formation do not rely on a hierarchy of multiple SOMs. The spiking neuron version of the LISSOM model (Choe and Miikkulainen 1998) makes use of spiking neurons in a two-layer model of visual cortical mapping. Average activity, or firing rates of the neurons, rather than temporal properties, are used throughout, providing no insight into the potential roles of phasic coding and STDP in modelling sensory cortex. Also, the self-organisation of direction preference in barrel cortex has been modelled using integrate-and-fire neurons and STDP (Kremer et al. 2011). This model incorporates bio-

logically plausible lateral connection profiles, and is capable of learning to represent direction preferences for 25 whiskers, each with their own neural network, interconnected in a square grid. Topographic structure is achieved both within and between barrel regions, with input as a temporal sequence of spikes representing movement of a bar through the whisker field. Competition within each barrel through an effectively similar lateral neighbourhood to the current spiking SOM model (although not similar in scale; the barrel cortex model is much larger than the spiking SOM) ensures regions representative for the full range of direction preference arranged in a topographic order. Additional long-range connections between barrels cause a particular geographic region of one barrel to learn the same direction preference as the same geographic region in a neighbouring barrel, such that regular, temporally ordered activity across the learned network results from a stimulus moving in one direction through the whisker field.

In summary, each of the above models suggest ways in which spiking neuron models can facilitate self-organisational network properties, but none of them represents a solution that incorporates the ideas of temporal coding of input data through relative phases of spikes within oscillations, continuous presentation of input data points to the network, and learning via STDP simultaneously. In addition, none of these models present a network structure possessing self-organisational and classification properties comparable with the traditional SOM algorithm. The spiking SOM in its current form does not produce representations of specific properties of sensory cortical areas, but none of the models capable of this are capable of mapping a wide variety of input data as in the current model and the conventional SOM. Additionally, some models suggest that modification primarily to the input mechanisms and scale of the current spiking SOM that may enable representation of such cortical feature maps (Miikkulainen et al. 2004; Kremer et al. 2011).

7.1.2 Relation to biology

The SOM algorithm is itself partially inspired by the topographic feature maps found across sensory cortex (Kohonen 2001). The relation of the current model to these maps is briefly discussed in section 7.1.1. This section will focus on the relationship between the underlying mechanisms present in the spiking SOM model and the cortical features described in section

2.3: lateral connection profiles, oscillations and temporal coding, along with STDP.

Lateral connections within cortex are most prominent within layer 2 or 3 (Binzegger et al. 2009; Stepanyants et al. 2008). Lateral connections are assumed by the SOM algorithm to be crucial to self-organisation, and indeed the presence of a lateral neighbourhood in the current spiking SOM is a fundamental principle underlying topographic map formation (see section 5.2.6). A key distinction between any cortical region with self-organising properties and the current model is the scale involved. Long range cortical connections are found to be patchy in nature (Voges et al. 2010; Muir et al. 2011); the current model does not incorporate this type of profile, exhibiting an all-to-all connectivity, decaying smoothly with distance from excitatory to inhibitory. Local connectivity, of the dense nature modelled here, is usually found within 0.5mm (Voges et al. 2010). As such, one complete spiking SOM output layer can be considered to be maximally representative of this approximate size of cortical region. Given the value of 90,000 neurons per cubic millimetre of cortical tissue (Schüz and Plam 1989), and the approximate cortical depth of 0.25mm for layers 2 or 3 (Voges et al. 2010), along with the 0.5mm lateral width (1mm diameter), the portion of cortex putatively represented by the spiking SOM would contain roughly 225 times the number of neurons and synapses modelled here. As such, each neuron in the spiking SOM layer would represent only an average of the response properties of many neurons in a small region. However, it has been shown that cortical feature maps appear most clearly when average responses in regions are taken, and that neighbouring neurons are merely likely to share similar response properties (Kremer et al. 2011; Chen et al. 2011; Wang et al. 1998). This coarse approximation of a small cortical region represents a reasonable assumption. With this in mind, the spiking SOM model demonstrates that a local lateral connection profile capable of combining excitation within a neighbourhood with competition within a small cortical area can contribute significantly to self-organisation of input preferences within that area. This is a reasonable fit with the type of dense local connectivity seen in small cortical regions.

The oscillatory regime in the spiking SOM model is generated by the feedforward-feedback excitatory-inhibitory circuit in the input layer of the model. This is the type of inhibition most

commonly seen in type I neurons (Bartos et al. 2007; Wang 2010), where mutual inhibition is required to produce an oscillatory effect. Oscillations are important to the segmentation of continuously presented input stimuli in the model, and this method of induction of oscillations is compatible with the standard IF neuron model. As seen in biological networks, an increase in the magnitude of the layer u inhibitory PSPs would result in a decrease in the oscillation frequency of the network (Atallah and Scanziani 2009). Also, the duration of inhibition affects the synchronisation tightness in random networks, with faster decay of IPSPs leading to tighter synchrony (Börgers and Kopell 2003). This effect would also be present in layer u of the current network; if an IPSP lasts longer, then the recovery from that IPSP will be slower, resulting in greater distribution of firing times of the layer u neurons. These methods represent biologically plausible options for altering the distribution of input firing, essentially changing the phase code, if necessary.

Neurons that are good representatives of the input data fire early within the phase, resulting in a temporal code that complements STDP learning, as described in section 5.2.6. This is an example of saliency to latency transformation resulting in a phase-coded firing sequence. The reliable conversion of rate-based information into phase-coded information within oscillations can be seen in the brain (McLelland and Paulsen 2009; Panzeri et al. 2001), along with examples of originally non-temporal information coded in temporal properties of spike trains (Foffani et al. 2004; Maldonado et al. 2008). The spiking SOM model presents a concrete example of combining oscillatory activity and phase coding with STDP and competition in the downstream layer to learn the information provided by the saliency to latency transformation.

The use of phase-of-firing coding and an STDP learning rule has some limitations in the current context. Synaptic connections from neurons involved in the input pattern are weakened if the neuron fires late within the temporal sequence (i.e., after firing in the output layer), with relatively late firing being interpreted as a neuron poorly representing the actual current input value. However, a neuron with much lower input activity, representing the current input value much more poorly, would not fire at all for the current input; with STDP learning, synaptic connections from such neurons to the input layer are not weakened. This problem is not relevant if

the input neurons that have connections to an output map represent a narrow domain of actual input stimuli, in which case any stimulus from the associated domain presented to that bank of input neurons would generate some relatively high activity level (and therefore firing) for the entire bank of neurons, but with a higher activity for some input neurons than others.

7.1.3 Applications

The traditional SOM algorithm has a variety of applications across any domain in which data analytical techniques are often required. The spiking SOM implementation is capable of recreating the computational properties of the traditional SOM. There are some circumstances in which the use of a spiking neuron SOM might be more appropriate than the traditional SOM. First, spiking neurons function as independent units with no global properties shared between them, offering greater efficiency gains when parallelised. Second, the spiking SOM constitutes a suitable basic cognitive function for implementation on analogue spiking neuron hardware. These points are related in that individual analogue neurons represent a fully parallel spiking SOM implementation that can run in real time. Combined with the ability of the model to learn from continuous presentation of an input stimulus, this could make a suitable chip for real time learning of the general properties of an environment. A circuit of this nature might be ideal for installation in a robotic platform, operating as an independent module for unsupervised learning of continuous sensory data.

The inhibitory mechanism introduced to produce oscillatory firing with phase-of-firing coding from a continuously presented input activity level (see section 5.2.4) is a versatile and useful neural function in its own right. The primary difference between this mechanism and a conventional feedforward-feedback excitatory-inhibitory loop is that the current mechanism effectively detects the end of the input pattern, rather than charging through successive PSPs as an ordinary downstream neuron would. A drawback is that it is necessary to pre-determine the required size of a gap in the input pattern that will allow the inhibitory neuron to fire (by setting the PSP parameters *to* the inhibitory neuron from the input layer), and the length and magnitude of the inhibition, which controls the rate of oscillation (by setting the PSP parameters *from* the inhibitory neuron to the input layer). However, these values do not require rigorous fine-

tuning provided the range of input activity levels across the pool of input neurons is relatively low, meaning that, with no external interference, all neurons in the pattern will fire within a restricted time period, followed by a gap before the neuron with the strongest input will fire again.

7.2 Hierarchical concept learning model

The hierarchical concept learning model (see chapter 6) extends the spiking SOM by separating the output layer neurons into excitatory and inhibitory neurons, and adjusts the input layer for the case of a binary input data set. The learned representation of a data set containing a hierarchy of features is demonstrated to appropriately represent that hierarchical structure. Additional modifications can be made to output neuron parameters to adjust the nature of the learned representation in terms of its relation to the input hierarchy.

7.2.1 Relation to existing models

The model of Kim et al. (2008) features an output layer that is capable of categorising input patterns at varying degrees of granularity. This effectively imposes hierarchical structure on the input data set through the manipulation of network parameters, in particular the level of top down excitatory input into the output layer, which has the effect of varying the level of inhibitory responsiveness. An example of the result of this method is shown in figure 7.1. When using a strong top-down input to the output layer, the output layer population activity is lower on average and less correlated between input patterns that have less similarity. These responses merge into stronger population responses that are harder to distinguish when a weak level of top-down input is used.

The current model displays behaviour consistent with this result. As demonstrated in section 6.3.3, lowering the level of inhibitory responsiveness leads to a more diffuse network response, with output patterns merging between categories (see figure 6.8). A lower level of responsiveness of the inhibitory neurons is equivalent to reduced excitatory input across the inhibitory neurons in the layer, which potentially could come from a top-down source. A generally higher level of top-down excitation to these neurons would encourage them to fire more quickly in

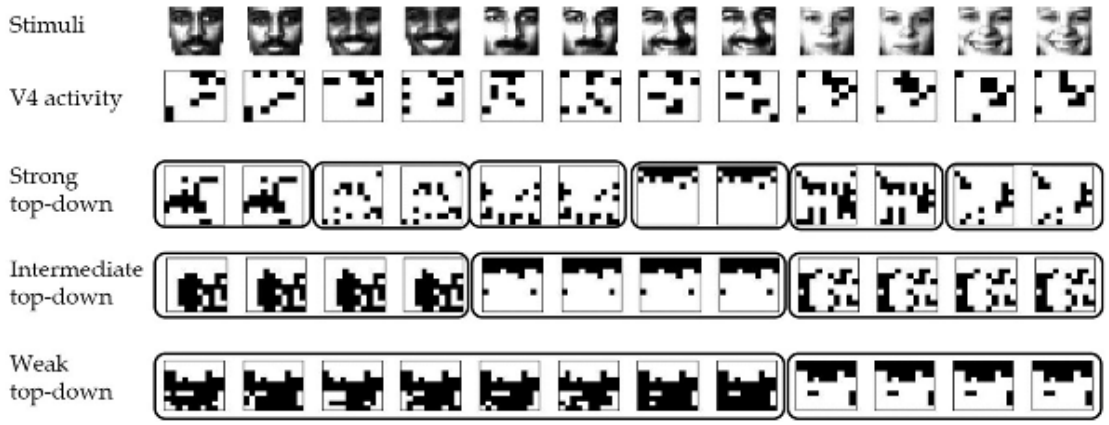


Figure 7.1: In the model of Kim et al. (2008), visual features (V4 activity) were extracted from images of faces (Stimuli), and used as input to the network. Strong, intermediate, and weak levels of top-down activity were applied to the output layer of the network, resulting in the merging of categories in terms of the population output activity. Reproduced from Kim et al. (2008).

response to excitatory firing in the layer, resulting in inhibition shutting down the spread of excitatory activity more quickly within an oscillation. In both examples, the outcome of the increased amount of excitatory firing across the output layer is that population responses merge in relation to the hierarchical structure apparent within the input data set.

A major difference between the two models, aside from the use of spiking neurons in the current model, is that lateral connections in the Kim et al. (2008) model are plastic. This results in the formation of the type of attractor states demonstrated to exhibit hierarchical responses when trained with binary object-feature relationships by other associative network models (Bělohlávek 2000; Huyck 2007; Wennekers 2009). In the current model, the effect of plastic lateral connections has not been investigated. The use of STDP and the current network model complicate the learning of this type of output layer structure. Whereas Hebbian learning in the above mentioned models will lead to the bidirectional strengthening of simultaneously activated neurons, using an STDP regime will potentiate the connection only in the direction in which presynaptic firing is found before postsynaptic firing. In fact, the reciprocal connection will be weakened through STDP, as one neuron has to fire after another in the output layer response of the current network. In general the role of STDP in the formation of neural maps is considered to be to learn feedforward weights in conjunction with a competitive or mediative lateral

connection profile (Song and Abbott 2001; Kepecs et al. 2002). Recurrent connections become secondary to feedforward connections as learning progresses, eventually resulting in columns defined wholly by their responses to input neurons (Kepecs et al. 2002).

Chartier et al. (2009) present a BAM model with a SOM layer incorporated. This model achieves a spatial map for input patterns consisting of binary feature sets, but the work does not address the extent to which hierarchical structure in the input data can be learned or recalled. The current work relates to this by also using a SOM-inspired output layer with some topographical mapping capabilities to learn binary input patterns. Through this relation, it can be suggested that the Chartier et al. (2009) model is also capable of the type of hierarchical representation demonstrated in the current network. The incorporation of feedback connections with plasticity enables the autoassociative properties of the BAM in this network. Again, the current model does not feature this type of connection, and the method for the learning of reciprocal connections of this nature is not obvious within the current STDP regime.

7.2.2 Relation to biology

Neuronal activity patterns in IT cortex in response to stimulus pairs that are relatively close in an intuitive hierarchical category structure have a relatively high correlation with each other (Kiani et al. 2007). A category tree assembled from neural activity distances resembles the intuitive hierarchical structure between the input stimuli. Additionally, negative correlations were observed between patterns that are intuitively only related at a high level of the hierarchy. The output patterns observed in the current model in response to input patterns from the hierarchical data set shown in table 6.1 exhibit similarly correlated activity, with positive correlations between responses to similar patterns, and negative correlations between responses to less similar patterns. This suggests that in both IT cortex and the current model, the responses of a population of cells represent both the individual categories and the relationship between the categories. The number of stimuli used in the study in (Kiani et al. 2007) is very large (>1000), and recordings of 674 neurons, evenly distributed across a relatively wide area in IT, were made. This is in contrast to 8 stimuli across an output layer of 100 neurons in the model case. Also, the stimuli were visual pictures, and although the number of high-level features present in these stimuli is

not enumerated it is likely to be greater than the 15 binary features contributing to the input data set in the model case. As such, these data cannot be treated as directly comparable.

However, the results of the model suggest that the nature of the relations between population responses in the brain may be a result of small patches of self-organisation to sets of binary input features. Indeed, a further finding of (Kiani et al. 2007) is that cells within 1mm of each other exhibit significantly stronger correlation between their stimulus and category response properties than cells with a >1mm distance between them. At distances beyond 1mm there is little change in the correlation values, indicating that cells with similar category selectivity make multiple small clusters distributed across a region. Similar spatial clustering of response properties has been found in other studies (Wang et al. 2000; Tompa and Sáry 2010), and indeed the 1mm radius found here to be the extent of spatial clustering is also the approximate distance up to which dense, local lateral connections are made (Voges et al. 2010). This is a further indication that the representational structure formed in cortex is the result of topographic organisation over areas established by dense local lateral excitatory and inhibitory projections.

An interesting property of the cortical object representation is the non-additive mapping of features and non-monotonic stimulus preference shown by Wang et al. (2000) and Tsunoda et al. (2001). The addition of features to a stimulus does not necessarily result in an addition to the active regions in IT, and vice versa for a reduction in the features of a stimulus. This indicates that input objects may be represented by both active and inactive columns, resulting in a representation scheme that is more complex than a sum of features. The results of the current model do not account for this property completely, in that groups of neurons not activated by a wide stimulus will not be activated by a narrow stimulus. However, due to the hierarchical representation, the removal of features (leading to a higher hierarchical level) results in a greater number of active output neurons. As can be seen in 6.4, presentation of a more vague input pattern results in a less specific population response on average in the output layer. The model of Kim et al. (2008) achieves non-additive mapping of features, with output neurons capable of responding to a narrow stimulus without responding to a wide stimulus. The exact components of the model that enable this behaviour are not explicated, but it appears to emerge through

the use of a homeostatic process applied to activity levels across the layer. Combined with lateral synaptic plasticity, this encourages average activity levels across the output layer that are independent of the number of input features. The use of spiking neurons and the competitive nature of the self-organisation in the output layer of the current model make the implementation of this type of property difficult. However, a potential solution could be found through the use of a feedforward inhibition level relating to the level of feedforward excitation. A large amount of early excitatory firing in the input layer, indicative of the presence of many features, could also lead to a large amount of early inhibitory firing in the input layer, such that a response from the output layer is suppressed. This kind of mechanism would have serious ramifications for the nature of self-organisation and hierarchical learning in the network, and represents important research to conduct as the investigation of this type of network continues.

7.3 Future Work

Several general avenues for future work are suggested by the research presented in this thesis. These are: analogue spiking neuron hardware implementation of the developed models; adaptation of the spiking SOM to model specific sensory modalities; an analysis of the potential of the hierarchical representation model to fully learn arbitrary formal contexts; and the development of cognitive architectures using the models presented here as modular components.

7.3.1 Hardware implementation

A plausibility experiment for a hardware implementation of the spiking SOM model is conducted in section 5.3.9. The general principles of the output organisation generated by the spiking SOM were demonstrated in a virtual simulation of existing analogue spiking neuron hardware. It should, therefore, be possible to implement this reduced SOM model on the hardware in question. As such, this immediate goal represents a promising direction for continuation of this work.

The extension of this model to a two-dimensional SOM output layer, and further to the hierarchical representation model of chapter 6, present potentially greater difficulty. Constraints are imposed by the number of neurons and synaptic connections available on the chip, as well as

on which of these synapses can learn. Additional simulations would be required to establish the plausibility of these extensions. The possibility to interconnect multiple chips until sufficient number of neurons and connections of the correct type are reached also exists. It seems that this longer term goal is more difficult but still achievable. Certainly the development of these models on analogue spiking neuron hardware would represent a unique accomplishment in terms of the complexity and utility of the implemented models.

7.3.2 Sensory cortex models

A potential approach to modifying the current spiking SOM to model sensory cortex could take inspiration from the traditional SOM extensions featured in the development of the LISSOM model (Miikkulainen et al. 2004). First, the lateral connections self-organise at the same time as the feedforward connections to the output map. This helps related areas of the map excite each other at distance, generating long-range topographic properties. The LISSOM architecture is much larger in scale than the current spiking SOM. Second, a pair of two-dimensional neural grids are used as the input layer to the map, representing a pair of retinas, with inputs consisting of oriented Gaussian bars at various locations in the input space, with high correlations between the retinas. Third, multiple delay channels are introduced as input to the map, along with movement of the oriented input bar across the retinal fields. These modifications to the SOM algorithm, over and above the replacement of the traditional learning rule with a form of average-activity based Hebbian learning, result in formation of maps that capture the combined topographic representations of present retinotopic, orientation, ocular dominance and direction preference in visual cortex. The adaptation of these modifications for use in the spiking SOM model would make a promising starting point for the development of a general spiking SOM model replicating visual cortex.

7.3.3 Learning and recall of arbitrary contexts

The bidirectional associative memory is capable of the storage and recall of all concepts from any arbitrary context (Bělohlávek 2000). The establishment of such a capacity in a network similar to the current one could represent a significant cognitive module in spiking neurons. The sample data set used for testing throughout chapter 6 is fully learned by the network, although

responses become less distinct at levels higher than the level the network has been trained at. It was shown in section 6.3.3 that altering the ratio of excitability between excitatory and inhibitory neurons in the output layer can modify the population response such that an appropriate level of the hierarchy is represented in that response. This alteration can come from some source external to the network, and automatic modulation of this ratio would allow an appropriate network response to be selected within an oscillation. If a response representing a low level of the hierarchy is required, this would be indicated by a greater number of early spikes in the input layer, representing more 1s in the current input pattern. If a higher level is required, there will be less early spikes. This means that, within an oscillation, more time until an output layer spike is a good indication that an input is from a higher level. As such, a source of inhibition to the inhibitory output neurons that diminishes their sensitivity could potentially be based on elapsed time within an oscillation.

Plastic lateral connections may also be an important factor in the learning and recall of any context. Output neurons that are activated by an input pattern should not necessarily excite each other for a concept to be accurately recalled. The retention after learning of the current lateral connection profile means that reciprocal connections between pairs of output neurons will have identical strength. The drawbacks of plastic lateral connections using STDP in the current learning scheme have been outlined above, but the development of a learning rule that could enable this function could be crucial. Additionally, the likelihood of features being absent from input data will play a role in the extent to which a context can be learned. This aspect is discussed in section 6.3.4. Of course, the scale of the network and input data sets are also relevant. The BAM requires variable numbers of neurons to be capable of recall of arbitrary data, and it seems implausible that a relatively small output layer in the current network would be able to store a context with a relatively large number of features. The relationship between these scales remains to be investigated.

In summary, automatic learning of arbitrary contexts would require an analysis of network scales, and feature noise levels and lateral connection plasticity, and automatic recall of all concepts in that context would require an automatically modulated level of the ratio of excitatory

to inhibitory neuron excitability.

7.3.4 Cognitive modules

A key motivation for the current work is to generate a suite of modular components that perform generic cognitive processes. These should be related to relatively simple cortical mechanisms, perform computationally important roles, and eventually be combined into architectures capable of fulfilling a variety of cognitive functions. The spiking neuron SOM model and the hierarchical concept learning model presented in this thesis represent good candidates for a pair of such modules.

This pair of modules are constructed from a network that is abstracted from general cortical properties and results in functionality that is apparent in a variety of cortical processes. Modules of this nature may be interlinked throughout a sheet of neurons. A larger neural surface may host input connections from many channels. These modules only feature the output layer neurons and lateral connections relevant to a particular input domain. Several input domains can potentially exist alongside each other while maintaining the modularity of their processing; they may even be physically intermingled without interacting through their connections. Parallel representations of several input channels can be created through modules of this nature that lack lateral influence on each other, resulting in a direct spatial representation of each input domain for further downstream processing to interpret.

The modular nature of these individual networks can, however, be impinged through lateral interactions between them. Two potential implementations of this type of link are that some neurons may be shared between modules, or that synaptic connections may exist between neurons belonging to independent modules. In the former case, the combination of feedforward input connections from different domains could result in a neuron, or group of neurons, that respond to essentially a multi-modal combination of input properties. In the latter case, spatial formation of a map in one input domain would bias the spatial formation of the map from the other domain. This type of map formation is seen in the barrel cortex model of Kremer et al. (2011), with similar direction preference appearing in the similar relative location in the barrel due to lateral connections stretching one barrel in diameter. Both of these types of interaction

would have a significant impact on the hierarchical and categorical nature of the representations formed. The exact nature of the influence of these interactions relate to representations constitutes a crucial avenue for further research if the potential of modular mechanisms in generating cognitive processes is to be fully explored.

List of references.

Abbott, L. F. (1999), 'Lapicque's introduction of the integrate-and-fire model neuron (1907)', *Brain Research Bulletin* **50**(5-6), 303–304.

Abbott, L. F. and Nelson, S. B. (2000), 'Synaptic plasticity: taming the beast', *Nature Neuroscience Supplement* **3**, 1178–1183.

Abbott, L. F. and Nelson, S. B. (2003), Temporal dynamics of biological synapses, *in* M. A. Arbib, ed., 'The Handbook of Brain Theory and Neural Networks', 2nd edn, MIT Press, Cambridge, MA, pp. 1156–1158.

Abbott, L. F. and Regehr, W. G. (2004), 'Synaptic computation', *Nature* **431**, 796–803.

Adesnik, H. and Scanziani, M. (2010), 'Lateral competition for cortical space by layer-specific horizontal circuits', *Nature* **464**, 1155–1160.

Akrami, A., Liu, Y., Treves, A. and Jagadeesh, B. (2009), 'Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli', *Cerebral Cortex* **19**, 760–776.

Atallah, B. V. and Scanziani, M. (2009), 'Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition', *Neuron* **62**, 566–577.

Bako, L. (2010), 'Real-time classification of datasets with hardware embedded neuromorphic neural networks', *Briefings in Bioinformatics* **11**(3), 348–363.

Bartos, M., Vida, I. and Jonas, P. (2007), 'Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks', *Nature Reviews Neuroscience* **8**, 45–56.

Bauer, H. and Pawelzik, K. (1992), 'Quantifying the neighborhood preservation of self-organizing feature maps', *IEEE Transactions on Neural Networks* **3**(4), 570–579.

Beaton, D., Valova, I. and Maclean, D. (2010), 'CQoCO: A measure for comparative quality of coverage and organization for self-organizing maps', *Neurocomputing* **73**, 2147–2159.

Bednar, J. A. and Miikkulainen, R. (2006), 'Joint maps for orientation, eye, and direction preference in a self-organizing model of V1', *Neurocomputing* **69**, 1272–1276.

Berglund, E. and Sitte, J. (2006), 'The parameterless self-organizing map algorithm', *IEEE Transactions on Neural Networks* **17**(2), 305–316.

Bi, G. and Poo, M. (1998), 'Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type', *The Journal of Neuroscience* **18**(24), 10464–10472.

Bi, G.-Q. (2002), 'Spatiotemporal specificity of synaptic plasticity: cellular rules and mechanisms', *Biological Cybernetics* **87**, 319–332.

Bienenstock, E. L., Cooper, L. N. and Munro, P. W. (1982), 'Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex', *The Journal of Neuroscience* **2**(1), 32–48.

Binzegger, T., Douglas, R. J. and Martin, K. A. C. (2009), 'Topology and dynamics of the canonical circuit of cat V1', *Neural Networks* **22**, 1071–1078.

Blasdel, G. G. and Salama, G. (1986), 'Voltage-sensitive dyes reveal a modular organization in monkey striate cortex', *Nature* **321**, 579–585.

Bock, D. D., Lee, W.-C. A., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., Yurgenson, S., Soucy, E. R., Kim, H. S. and Reid, R. C. (2011), 'Network anatomy and in vivo physiology of visual cortical neurons', *Nature* **471**, 177–182.

Bohte, S. M., Kok, J. N. and La Poutré, H. (2002), Implementing position-invariant detection of feature-conjunctions in a network of spiking neurons, *in* 'Proceedings of IJCNN'2002', Honolulu, Hawaii.

Börgers, C. and Kopell, N. (2003), 'Synchronization in networks of excitatory and inhibitory neurons with sparse , random connectivity', *Neural Computation* **538**, 509–538.

Bosking, W. H., Zhang, Y., Schofield, B. and Fitzpatrick, D. (1997), 'Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex', *The Journal of Neuroscience* **17**(6), 2112–27.

Boucsein, C., Nawrot, M. P., Schnepel, P. and Aertsen, A. (2011), 'Beyond the cortical column: abundance and physiology of horizontal connections imply a strong role for inputs from the surround', *Frontiers in Neuroscience* **5**(32), 1–13.

Brader, J. M., Senn, W. and Fusi, S. (2007), 'Learning real-world stimuli in a neural network with spike-driven synaptic dynamics', *Neural computation* **19**, 2881–2912.

Branco, T., Clark, B. A. and Häusser, M. (2010), 'Dendritic discrimination of temporal input sequences in cortical neurons', *Science* **329**(5999), 1671–1675.

Breakspear, M. and Stam, C. J. (2005), 'Dynamics of a neural system with a multiscale architecture', *Philosophical Transactions of the Royal Society B* **360**, 1051–1074.

Brette, R. and Gerstner, W. (2005), 'Adaptive exponential integrate-and-fire model as an effective description of neuronal activity', *Journal of Neurophysiology* **94**, 3637–3642.

Brunel, N. and van Rossum, M. C. W. (2007*a*), 'Lapicque's 1907 paper: from frogs to integrate-and-fire', *Biological Cybernetics* **97**(5-6), 337–339.

Brunel, N. and van Rossum, M. C. W. (2007*b*), 'Quantitative investigations of electrical nerve excitation treated as polarization', *Biological Cybernetics* **97**(5-6), 34134–34139.

Bělohlávek, R. (2000), 'Representation of concept lattices by bidirectional associative memories', *Neural Computation* **12**, 2279–2290.

Bureau, I., Shepherd, G. M. G. and Svoboda, K. (2004), 'Precise development of functional and anatomical columns in the neocortex', *Neuron* **42**, 1–20.

Bush, D., Philippides, A., Husbands, P. and O'Shea, M. (2010), 'Reconciling the STDP and BCM models of synaptic plasticity in a spiking recurrent neural network.', *Neural Computation* **22**, 2059–2085.

Buzsáki, G. (2002), 'Theta oscillations in the hippocampus', *Neuron* **33**, 325–340.

Buzsáki, G., Geisler, C., Henze, D. A. and Wang, X.-J. (2004), 'Interneuron diversity series: circuit complexity and axon wiring economy of cortical interneurons', *Trends in Neurosciences* **27**(4), 186–193.

Calford, M. B. (2002), 'Dynamic representational plasticity in sensory cortex', *Neuroscience* **111**(4), 709–738.

Callaway, M. and Katz, C. (1990), 'Emergence cat striate and refinement cortex of clustered horizontal connections in cat striate cortex', *The Journal of Neuroscience* **10**(4), 1134–1153.

Caporale, N. and Dan, Y. (2008), 'Spike timing-dependent plasticity: a Hebbian learning rule.', *Annual Review of Neuroscience* **31**, 25–46.

Casagrande, V. A. and Kaas, J. H. (1994), 'The afferent, intrinsic, and efferent connections of primary visual vortex in primates', *Cerebral Cortex* **10**, 201–259.

Catania, K. C. and Kaas, J. H. (1996), 'The unusual nose and brain of the star-nosed mole', *Bioscience* **46**(8), 578–586.

Chartier, S., Giguère, G. and Langlois, D. (2009), 'A new bidirectional heteroassociative memory encompassing correlational, competitive and topological properties', *Neural Networks* **22**, 568–578.

Chartier, S., Giguère, G., Langlois, D. and Sioufi, R. (2009), Bidirectional associative memories, self-organizing maps and k-winners-take-all: uniting feature extraction and topological principle, *in* 'Proceedings of International Joint Conference on Neural Networks IJCNN '09', Atlanta, Georgia, pp. 503–510.

Chen, X., Gabitto, M., Peng, Y., Ryba, N. J. P. and Zuker, C. S. (2011), 'A gustotopic map of taste qualities in the mammalian brain', *Science* **333**, 1262–1266.

Cho, S. and Seok, J. (1998), 'Self-organizing map with time-invariant learning rate and its exponential stability analysis', *Neurocomputing* **19**, 1–11.

Choe, Y. and Miikkulainen, R. (1998), 'Self-organization and segmentation in a laterally connected orientation map of spiking neurons', *Neurocomputing* **21**, 139–157.

Chuang, C.-H., Cheng, P. E., Liou, M. and Chen, C.-Y. (2005), Constrained self-organizing map for the reconstruction of the brain lateral ventricle, *in* 'Proceedings of the 12th International Conference on Neural Information Processing', Taipei.

Chuang, C.-H., Cheng, P. E., Liou, M., Liou, C.-Y. and Kuo, Y.-T. (2007), 'Application of self-organizing map (SOM) for cerebral cortex reconstruction', *The International Journal of Computational Intelligence Research* **3**, 26–30.

Cios, K. J., Swiercz, W. and Jackson, W. (2004), 'Networks of spiking neurons in modeling and problem solving', *Neurocomputing* **61**, 99–119.

Citri, A. and Malenka, R. C. (2008), 'Synaptic plasticity: multiple forms, functions, and mechanisms', *Neuropsychopharmacology* **33**, 18–41.

Clark, S. A., Allard, T., Jenkins, W. M. and Merzenich, M. M. (1988), 'Receptive fields in the body-surface map in adult cortex defined by temporally correlated inputs', *Nature* **332**, 444–445.

Cobb, S. R., Buhl, E. H., Halasy, K., Paulsen, O. and Somogyi, P. (1995), 'Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons.', *Nature* **378**, 75–78.

Csicsvari, J., Jamieson, B., Wise, K. D. and Buzsáki, G. (2003), 'Mechanisms of gamma oscillations in the hippocampus of the behaving rat', *Neuron* **37**, 311–322.

Dahl, C. D., Logothetis, N. K. and Kayser, C. (2009), 'Spatial organization of multisensory responses in temporal association cortex', *The Journal of Neuroscience* **29**(38), 11924–11932.

Dayan, P. and Abbott, L. F. (2000), *Theoretical Neuroscience*, MIT Press.

de la Rocha, J., Marchetti, C., Schiff, M. and Reyes, A. D. (2008), 'Linking the response properties of cells in auditory cortex with network architecture: cotuning versus lateral inhibition', *The Journal of Neuroscience* **28**(37), 9151–9163.

Destexhe, A., Rudolph, M. and Paré, D. (2003), 'The high-conductance state of neocortical neurons in vivo', *Nature Reviews Neuroscience* **4**, 739–751.

Dobrunz, L. E. and Stevens, C. F. (1997), 'Heterogeneity of release probability, facilitation, and depletion at central synapses', *Neuron* **18**, 995–1008.

Douglas, R. J. and Martin, K. A. C. (2004), 'Neuronal circuits of the neocortex', *Annual Review of Neuroscience* **27**, 419–451.

Dura-Bernal, S., Garreau, G., Andreou, C., Georgiou, J., Wennekers, T. and Denham, S. (2011), Human action categorization using ultrasound micro-Doppler signatures, *in* 'Human Behavior Understanding - Second International Workshop, HBU 2011', Springer, Amsterdam, pp. 18–28.

Ermentrout, B. (1996), 'Type I membranes, phase resetting curves, and synchrony', *Neural Computation* **8**(5), 979–1001.

Feldman, D. E. (2009), 'Synaptic mechanisms for plasticity in neocortex', *Annual Review of Neuroscience* **32**, 33–55.

Feldman, D. E. and Brecht, M. (2005), 'Map plasticity in somatosensory cortex', *Science* **310**, 810–815.

Fino, E. and Yuste, R. (2011), 'Dense inhibitory connectivity in neocortex', *Neuron* **69**, 1188–1203.

Fisher, R. A. (1936), 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics* **7**(2), 179–188.

Foffani, G., Tutunculer, B. and Moxon, K. A. (2004), 'Role of spike timing in the forelimb somatosensory cortex of the rat', *The Journal of Neuroscience* **24**(33), 7266–7271.

Foody, G. M. (1999), 'Applications of the self-organising feature map neural network in community data analysis', *Ecological Modelling* **120**, 97–107.

Fourcaud-Trocmé, N., Hansel, D., van Vreeswijk, C. and Brunel, N. (2003), 'How spike generation mechanisms determine the neuronal response to fluctuating inputs', *The Journal of Neuroscience* **23**(37), 11628–11640.

Fox, K. (2002), 'Anatomical pathways and molecular mechanisms for plasticity in the barrel cortex', *Neuroscience* **111**(4), 799–814.

Fox, K. (2008), *Barrel Cortex*, Cambridge University Press.

Friedman, R. M., Chen, L. M. and Roe, A. W. (2004), 'Modality maps within primate somatosensory cortex', *Proceedings of the National Academy of Sciences* **101**(34), 12724–12729.

Fries, P., Nikolić, D. and Singer, W. (2007), 'The gamma cycle', *Trends in Neurosciences* **30**(7), 309–316.

Fuster, J. M. (2003), *Cortex and Mind: Unifying Cognition*, 1st edn, Oxford University Press, New York.

Garreau, G., Nicolaou, N., Andreou, C., D'Urbal, C., Stuarts, G. and Georgiou, J. (2011), Computationally efficient classification of human transport mode using micro-Doppler signatures, *in* '45th Annual Conference on Information Sciences and Systems (CISS)', pp. 1–4.

Gennari, J. H., Langley, P. and Fisher, D. (1989), 'Models of Incremental Concept Formation', *Artificial Intelligence* **40**, 11–61.

Ghazanfar, A. A., Stambaugh, C. R. and Nicolelis, M. A. (2000), 'Encoding of tactile stimulus location by somatosensory thalamocortical ensembles.', *The Journal of Neuroscience* **20**(10), 3761–3775.

Gilbert, D. (1989), 'Columnar specificity of intrinsic horizontal connections in cat visual cortex and corticocortical', *The Journal of Neuroscience* **9**(7), 2432–2442.

Godin, N., Huguet, S. and Gaertner, R. (2005), 'Integration of the Kohonen's self-organising map and k-means algorithm for the segmentation of the AE data collected during tensile tests on cross-ply composites', *NDT & E International* **38**(4), 299–309.

Goodhill, G. J. and Sejnowski, T. J. (1997), 'A unifying objective function for topographic mappings', *Neural Computation* **9**, 1291–1303.

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**, 334–337.

Gray, P. R. (1967), 'Conditional probability analyses of the spike activity of single neurons', *Biophysical Journal* **7**(6), 759–77.

Graziano, M. S. A. and Aflalo, T. N. (2007), 'Mapping behavioral repertoire onto the cortex', *Neuron* **56**(2), 239–251.

Grinvald, A., Lieke, E. E., Frostig, R. D. and Hildesheim, R. (1994), 'Cortical point images and long-range lateral interaction in primary visual cortex of Macaque monkey', *The Journal of Neuroscience* **14**, 2545–2568.

Gueorguieva, N., Valova, I. and Georgiev, G. (2006), 'Learning and data clustering with an RBF-based spiking neuron network', *Journal of Experimental & Theoretical Artificial Intelligence* **18**(1), 73–86.

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.

Haese, K. (1998), 'Self-organizing feature maps with self-adjusting learning parameters', *IEEE Transactions on Neural Networks* **9**(6), 1270–1278.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J. and Sporns, O. (2008), 'Mapping the structural core of human cerebral cortex', *PLoS Biology* **6**(7), 1479–1493.

Hebb, D. O. (1949), *The Organization of Behavior*, John Wiley & Sons Inc., New York.

Hensch, T. K. (2004), 'Critical period regulation', *Annual Review of Neuroscience* **27**, 549–579.

Hertz, J., Krogh, A. and Palmer, R., eds (1991), *Introduction to the Theory of Neural Computation*, Vol. 1, Addison-Wesley.

Hikawa, H. (2005), 'FPGA implementation of self organizing map with digital phase locked loops', *Neural Networks* **18**, 514–522.

Hirabayashi, T. and Miyashita, Y. (2005), 'Dynamically modulated spike correlation in monkey inferior temporal cortex depending on the feature configuration within a whole object', *The Journal of Neuroscience* **25**(44), 10299–10307.

Hirsch, J. A., Martinez, L. M., Pillai, C., Alonso, J.-M., Wang, Q. and Sommer, F. T. (2003), 'Functionally distinct inhibitory neurons at the first stage of visual cortical processing', *Nature Neuroscience* **6**(12), 1300–1308.

Hodgkin, A. L. and Huxley, A. F. (1952), 'A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes', *Journal of Physiology, London* **117**, 500–544.

Holmgren, C., Harkany, T., Svennenfors, B. and Zilberter, Y. (2003), 'Pyramidal cell communication within local networks in layer 2/3 of rat neocortex', *The Journal of Physiology* **551**(1), 139–153.

Honey, C. J., Thivierge, J.-P. and Sporns, O. (2010), 'Can structure predict function in the human brain?', *NeuroImage* **52**, 766–776.

Horton, J. C. and Hocking, D. R. (1996), 'Anatomical demonstration of ocular dominance columns in striate cortex of the squirrel monkey', *The Journal of Neuroscience* **16**(17), 5510–5122.

Hubel, D. H., Wiesel, T. N. and Stryker, M. P. (1978), 'Anatomical demonstration of orientation columns in macaque monkey', *The Journal of Comparative Neurology* **177**, 361–380.

Huyck, C. (2007), 'Creating hierarchical categories using cell assemblies', *Connection Science* **19**(1), 1–24.

Ibarz, B., Casado, J. M. and Sanjuán, M. A. F. (2011), 'Map-based models in neuronal dynamics', *Physics Reports* **501**, 1–74.

Indiveri, G., Linares-Barranco, B., Hamilton, T. J., van Schaik, A., Etienne-Cummings, R., Delbruck, T., Liu, S.-C., Dudek, P., Häfliger, P., Renaud, S., Schemmel, J., Cauwenberghs, G., Arthur, J., Hynna, K., Folowosele, F., Saighi, S., Serrano-Gotarredona, T., Wijekoon, J., Wang, Y. and Boahen, K. (2011), 'Neuromorphic silicon neuron circuits', *Frontiers in Neuroscience* **5**, 1–23.

Isaacson, J. S. and Scanziani, M. (2011), 'How inhibition shapes cortical activity', *Neuron* **72**, 231–243.

Ito, J., Maldonado, P., Singer, W. and Grün, S. (2011), 'Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes', *Cerebral Cortex* **21**, 2482–2497.

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. and Desai, N. S. (2003), 'Relating STDP to BCM', *Neural Computation* **15**(7), 1511–1523.

Jack, J. J. B., Miller, S., Porter, R. and Redman, S. J. (1971), 'The time course of minimal excitatory post-synaptic potentials evoked in spinal motoneurones by group Ia afferent fibres', *Journal of Physiology* **215**, 353–380.

Jensen, O. and Tesche, C. D. (2002), 'Frontal theta activity in humans increases with memory load in a working memory task', *European Journal of Neuroscience* **15**, 1395–1399.

Kaas, J. H. (1991), 'Plasticity of sensory and motor maps in adult mammals', *Annual Review of Neuroscience* **14**, 137–167.

Kaas, J. H. and Catania, K. C. (2002), 'How do features of sensory representations develop?', *BioEssays* **24**, 334–343.

Kandel, E. R., Schwartz, J. H. and Jessell, T. M., eds (2000), *Principles of Neural Science*, 4th edn, McGraw-Hill, New York.

Keith-Magee, R. (2001), Learning and development in Kohonen-style self-organising maps, PhD thesis, Curtin University of Technology.

Kepecs, A., van Rossum, M. C. W., Song, S. and Tegner, J. (2002), 'Spike-timing-dependent plasticity: common themes and divergent vistas.', *Biological Cybernetics* **87**, 446–458.

Khaytin, I., Chen, X., Royal, D., Ruiz, O., Jermakowicz, W. J., Siegel, R. M. and Casagrande, V. A. (2008), 'Functional organization of temporal frequency selectivity in primate visual cortex', *Cerebral Cortex* **18**(8), 1828–1842.

Kiani, R., Esteky, H., Mirpour, K. and Tanaka, K. (2007), 'Object category structure in response patterns of neuronal population in monkey inferior temporal cortex', *Journal of Neurophysiology* **97**, 4296–4309.

Kiani, R., Esteky, H. and Tanaka, K. (2005), 'Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces', *Journal of Neurophysiology* **94**, 1587–1596.

Kim, Y., Vladimirskiy, B. B. and Senn, W. (2008), 'Modulating the granularity of category formation by global cortical states', *Frontiers in Computational Neuroscience* **2**, 1–14.

Kohonen, T. (2001), *Self-Organizing Maps*, third edn, Springer-Verlag, Berlin Heidelberg New York.

Kohonen, T., Oja, E., Simula, O., Visa, A. and Kangas, J. (1996), 'Engineering applications of the self-organizing map', *Proceedings of the IEEE* **84**(10), 1358–1384.

Kolster, H., Peeters, R. and Orban, G. A. (2010), 'The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors', *The Journal of Neuroscience* **30**(29), 9801–9820.

Kosko, B. (1988), 'Bidirectional associative memories', *IEEE transactions on Systems, Man and Cybernetics* **18**(1), 49–60.

Kremer, Y., Léger, J.-F., Goodman, D., Brette, R. and Bourdieu, L. (2011), 'Late emergence of the vibrissa direction selectivity map in the rat barrel cortex', *Journal of Neuroscience* **31**(29), 10689–10700.

Kriegeskorte, N., Mur, M. and Bandettini, P. (2008), 'Representational similarity analysis - connecting the branches of systems neuroscience', *Frontiers in Systems Neuroscience* **2**, 1–28.

Lefort, S., Tomm, C., Sarria, J.-C. F. and Petersen, C. C. H. (2009), 'The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex', *Neuron* **61**, 301–316.

Levy, R. B. and Reyes, A. D. (2011), 'Coexistence of lateral and co-tuned inhibitory configurations in cortical networks', *PLoS Computational Biology* **7**(10), 1–14.

Lin, L., Osan, R. and Tsien, J. Z. (2006), 'Organizing principles of real-time memory encoding: neural clique assemblies and universal neural codes', *Trends in Neurosciences* **29**(1), 48–57.

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**, 25–45.

Lübke, J. and Feldmayer, D. (2007), 'Excitatory signal flow and connectivity in a cortical column: focus on barrel cortex', *Brain Structure and Function* **212**, 3–17.

MacGregor, R. J. (1987), *Neural and Brain Modeling*, Academic Press.

Maldonado, P., Babul, C., Singer, W., Rodriguez, E., Berger, D. and Grün, S. (2008), 'Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images', *Journal of Neurophysiology* **100**, 1523–1532.

Markram, H., Lübke, J., Frotscher, M. and Sakmann, B. (1997), 'Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs', *Science* **275**, 213–215.

Martin, S. J. and Morris, R. G. M. (2002), 'New life in an old idea: the synaptic plasticity and memory hypothesis revisited', *Hippocampus* **12**, 609–636.

McCulloch, W. S. and Pitts, W. H. (1943), 'A logical calculus of the ideas immanent in nervous activity', *Bulletin of Mathematical Biophysics* **5**, 115–133.

McLelland, D. and Paulsen, O. (2009), 'Neuronal oscillations and the rate-to-phase transform: mechanism, model and mutual information', *The Journal of Physiology* **587**(4), 769–785.

Merzenich, M. M., Kaas, J. H., Wall, J., Nelson, R. J., Sur, M. and Felleman, D. (1983), 'Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation', *Neuroscience* **8**, 33–55.

Merzenich, M. M., Knight, P. L. and Roth, G. L. (1975), 'Representation of cochlea within primary auditory cortex in the cat', *Journal of Neurophysiology* **38**(2), 231–249.

Miikkulainen, R., Bednar, J. A., Choe, Y. and Sirosh, J. (2004), *Computational Maps in the Visual Cortex*, Springer, New York.

Mitra, S., Fusi, S., Indiveri, G. and Member, S. (2009), 'Real-time-classification of complex patterns using spike-based learning in neuromorphic VLSI', *IEEE Transactions on Biomedical Circuits and Systems* **3**(1), 32–42.

Mortimer, D., Feldner, J., Vaughan, T., Vetter, I., Pujic, Z., Rosoff, W. J., Burrage, K., Dayan, P., Richards, L. J. and Goodhill, G. J. (2009), 'A Bayesian model predicts the response of axons to molecular gradients', *Proceedings of the National Academy of Sciences* **106**(25), 10296–10301.

Muir, D. R., Da Costa, N. M. A., Girardin, C. C., Naaman, S., Omer, D. B., Ruesch, E., Grinvald, A. and Douglas, R. J. (2011), 'Embedding of cortical representations by the superficial patch system', *Cerebral Cortex* **21**, 2244–2260.

Mulier, F. and Cherkassky, V. (1994), 'Learning rate schedules for self-organizing maps', *Proceedings of the 12th IAPR International Conference on Pattern Recognition* pp. 224–228.

Ohki, K., Chung, S., Kara, P., Hübener, M. and Bonhoeffer, T. (2006), 'Highly ordered arrangement of single neurons in orientation pinwheels', *Nature* **442**, 925–928.

Orban, G. A. (2008), 'Higher order visual processing in macaque extrastriate cortex', *Physiological Reviews* **88**, 59–89.

Panchev, C. and Wermter, S. (2001), Hebbian spike-timing dependent self-organization in pulsed neural networks, *in* 'Proceedings of World Congress on Neuroinformatics', pp. 378–385.

Panzeri, S., Petersen, R. S., Schultz, S. R., Lebedev, M. and Diamond, M. E. (2001), 'The role of spike timing in the coding of stimulus location in rat somatosensory cortex', *Neuron* **29**, 769–77.

Paugam-Moisy, H. (2006), Spiking neuron networks: a survey, Technical report, IDIAP.

Perez-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I. and Laurent, G. (2002), 'Oscillations and sparsening of odor representations in the mushroom body', *Science* **297**, 359–365.

Polani, D. (2002), Measures for the Organization of Self-Organizing Maps, *in* U. Seiffert and L. C. Jain, eds, 'Self-Organizing Neural Networks', Physica-Verlag, Springer, Berlin, Heidelberg, pp. 13–44.

Raizada, R. D. S. and Grossberg, S. (2003), 'Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system', *Cerebral Cortex* **13**(1), 100–113.

Rajapakse, R. and Denham, M. (2005), 'Fast access to concepts in concept lattices via bidirectional associative memories', *Neural Computation* **17**, 2291–2300.

Rausell, E., Bickford, L., Manger, P. R., Woods, T. M. and Jones, E. G. (1998), 'Extensive divergence and convergence in the thalamocortical projection to monkey somatosensory cortex', *The Journal of Neuroscience* **18**(11), 4216–4232.

Reyes, A. D. (2003), 'Synchrony-dependent propagation of firing rate in iteratively constructed networks in vitro', *Nature Neuroscience* **6**(6), 593–599.

Ritter, H. and Kohonen, T. (1989), 'Self-organizing semantic maps', *Biological Cybernetics* **254**, 241–254.

Rolls, E. T. and Deco, G. (2002), *Computational Neuroscience of Vision*, Oxford University Press.

Rosa, M. G. P. (2002), 'Visual maps in the adult primate cerebral cortex: some implications for brain development and evolution', *Brazilian Journal of Medical and Biological Research* **35**, 1485–1498.

Ruf, B. and Schmitt, M. (1998), 'Self-organization of spiking neurons using action potential timing', *IEEE Transactions on Neural Networks* **9**(3), 575–578.

Rumsey, C. C. and Abbott, L. F. (2006), 'Synaptic democracy in active dendrites', *Journal of Neurophysiology* **96**(5), 2307–2318.

Sala, D. M., Cios, K. J. and Wall, J. T. (1998), 'Self-organization in networks of spiking neurons', *Australian Journal of Intelligent Information Processing Systems* **5**(3), 161–170.

Sceniak, M. P., Hawken, M. J. and Shapley, R. (2001), 'Visual spatial characterization of macaque V1 neurons', *Journal of Neurophysiology* **85**(5), 1873–1887.

Schoffelen, J.-M., Oostenveld, R. and Fries, P. (2005), 'Neuronal coherence as a mechanism of effective corticospinal interaction', *Science* **308**, 111–113.

Schüz, A. and Plam, G. (1989), 'Density of neurons and synapses in the cerebral cortex of the mouse', *The Journal of Comparative Neurology* **283**(4), 442–455.

Shapley, R., Hawken, M. and Xing, D. (2007), The dynamics of visual responses in the primary visual cortex, *in* P. Cisek, T. Drew and J. F. Kalaska, eds, 'Progress in Brain Research', Vol. 165, Elsevier B.V., pp. 21–32.

Sheik, S., Coath, M., Indiveri, G., Denham, S. L., Wennekers, T. and Chicca, E. (2012), 'Emergent auditory feature tuning in a real-time neuromorphic VLSI system', *Frontiers in Neuroscience* **6**(17), 1–11.

Siegel, M., Warden, M. R. and Miller, E. K. (2009), 'Phase-dependent neuronal coding of objects in short-term memory', *Proceedings of the National Academy of Sciences* **106**(50), 21341–21346.

Sincich, L. C. and Blasdel, G. G. (2001), 'Oriented axon projections in primary visual cortex of the monkey', *The Journal of Neuroscience* **21**(12), 4416–4426.

Singer, W. and Gray, C. M. (1995), 'Visual feature integration and the temporal correlation hypothesis', *Annual Review of Neuroscience* **18**, 555–586.

Sjöström, P. and Nelson, S. B. (2002), 'Spike timing, calcium signals and synaptic plasticity', *Current Opinion in Neurobiology* **12**, 305–314.

Sjöström, P., Rancz, E. A., Roth, A. and Hauser, M. (2008), 'Dendritic excitability and synaptic plasticity', *Physiological Reviews* **88**, 769 – 840.

Song, S. and Abbott, L. F. (2001), 'Cortical remapping through spike timing-dependent plasticity', *Neuron* **32**, 1–20.

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.

Song, Y., Han, M. and Wei, J. (2005), 'Stability and Hopf bifurcation analysis on a simplified BAM neural network with delays', *Physica D: Nonlinear Phenomena* **200**, 185–204.

Sporns, O. (2011), 'The human connectome: a complex network', *Annals of the New York Academy of Sciences* **1224**, 109–125.

Standage, D. I. and Trappenberg, T. P. (2005), Differences in the subthreshold dynamics of leaky integrate-and-fire and hodgkin-huxley neuron models, *in* 'Proceedings of the 2005 IEEE International Joint Conference on Neural Networks (IJCNN '05)', pp. 396–399.

Stepanyants, A., Hirsch, J. A., Martinez, L. M., Kisvárday, Z. F., Ferecskó, A. S. and Chklovskii, D. B. (2008), 'Local potential connectivity in cat primary visual cortex', *Cerebral Cortex* **18**, 13–28.

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**(141), 1–8.

Svetlik, M. (2006), 'Self-organizing maps with spiking neuron model JASTAP'.

Swann, N., Tandon, N., Canolty, R., Ellmore, T. M., McEvoy, L. K., Dreyer, S., DiSano, M. and Aron, A. R. (2009), 'Intracranial EEG reveals a time- and frequency-specific role for the right inferior frontal gyrus and primary motor cortex in stopping initiated responses', *The Journal of Neuroscience* **29**(40), 12675–12685.

Tamura, H., Kaneko, H. and Fujita, I. (2005), 'Quantitative analysis of functional clustering of neurons in the macaque inferior temporal cortex', *Neuroscience Research* **52**, 311–322.

Tamura, H., Kaneko, H., Kawasaki, K. and Fujita, I. (2004), 'Presumed inhibitory neurons in the macaque inferior temporal cortex: visual response properties and functional interactions with adjacent neurons', *Journal of Neurophysiology* **91**, 2782–2796.

Tamura, H. and Tanaka, K. (2001), 'Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex', *Cerebral Cortex* **11**, 384–399.

Tan, A. Y. Y., Brown, B. D., Scholl, B., Mohanty, D. and Priebe, N. J. (2011), 'Orientation selectivity of synaptic input to neurons in mouse and cat primary visual cortex', *The Journal of Neuroscience* **31**(34), 12339–12350.

Thomas, E., Van Hulle, M. M. and Vogels, R. (2001), 'Encoding of categories by noncategory-specific neurons in the inferior temporal cortex', *Journal of Cognitive Neuroscience* **13**(2), 190–200.

Thomson, A. M., West, D. C., Wang, Y. and Bannister, A. P. (2002), 'Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2-5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling in vitro', *Cerebral Cortex* **12**, 936–953.

Thorpe, S. J., Fize, D. and Marlot, C. (1996), 'Speed of processing in the human visual system', *Nature* **381**, 520–522.

Thorpe, S. J. and Imbert, M. (1989), Biological constraints on connectionist modelling, *in* 'Connectionism in Perspective', 1st edn, Elsevier Science Ltd, pp. 63–92.

Tompa, T. and Sáry, G. (2010), 'A review on the inferior temporal cortex of the macaque', *Brain Research Reviews* **62**, 165–182.

Traub, R. D., Whittington, M. A., Stanford, I. M. and Jefferys, J. G. R. (1996), 'A mechanism for generation of long-range synchronous fast oscillations in the cortex', *Nature* **383**, 621–624.

Tsunoda, K., Yamane, Y., Nishizaki, M. and Tanifuji, M. (2001), 'Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns', *Nature Neuroscience* **4**(8), 832–838.

Van der Loos, H. and Woolsey, T. A. (1973), 'Somatosensory cortex: structural alterations following early injury to sense organs', *Science* **179**(4071), 395–398.

van Rossum, M., Bi, G. and Turrigiano, G. (2000), 'Stable Hebbian learning from spike timing-dependent plasticity', *Journal of Neuroscience* **20**(23), 8812–8821.

Van Rullen, R. and Thorpe, S. J. (2002), 'Surfing a spike wave down the ventral stream', *Vision Research* **42**, 2593–2615.

Varga, Z., Jia, H., Sakmann, B. and Konnerth, A. (2011), 'Dendritic coding of multiple sensory inputs in single cortical neurons in vivo', *Proceedings of the National Academy of Sciences* **108**(37), 15420–15425.

Veredas, F., Mesa, H. and Martínez, L. A. (2008), 'Imprecise correlated activity in self-organizing maps of spiking neurons', *Neural Networks* **21**, 810–816.

Villmann, T., Der, R., Herrmann, M. and Martinetz, T. (1997), 'Topology preservation in self-organizing feature maps: exact definition and measurement', *IEEE Transactions on Neural Networks* **8**(2), 256–266.

Voges, N., Schüz, A., Aertsen, A. and Rotter, S. (2010), 'A modeler's view on the spatial structure of intrinsic horizontal connectivity in the neocortex', *Progress in Neurobiology* **92**, 277–292.

Wade, J., McDaid, L., Santos, J. and Sayers, H. (2010), 'SWAT: a spiking neural network training algorithm for classification problems', *IEEE Transactions on Neural Networks* **21**(11), 1817–1830.

Wandell, B. A. (2011), 'The neurobiological basis of seeing words', *Annals of the New York Academy of Sciences* **1224**, 63–80.

Wang, F., Nemes, A., Mendelsohn, M. and Axel, R. (1998), 'Odorant receptors govern the formation of a precise topographic map.', *Cell* **93**, 47–60.

Wang, X.-J. (2010), 'Neurophysiological and computational principles of cortical rhythms in cognition', *Physiological Reviews* **90**, 1195–1268.

Wang, X. J. and Buzsáki, G. (1996), 'Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model', *The Journal of Neuroscience* **16**(20), 6402–6413.

Wang, Y., Fujita, I. and Murayama, Y. (2000), 'Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex', *Nature Neuroscience* **3**(8), 807–813.

Wennekers, T. (2009), 'On the natural hierarchical composition of cliques in cell assemblies', *Cognitive Computation* **1**(2), 128–138.

Wille, R. (1982), Restructuring lattice theory: an approach based on hierarchies of concepts, *in* I. Rival, ed., 'Ordered Sets', Reidel, Dordrecht-Boston, pp. 445–470.

Wilson, H. R. (1999), 'Simplified dynamics of human and mammalian neocortical neurons', *Journal of Theoretical Biology* **200**, 375–388.

Wilson, S. P., Law, J. S., Mitchinson, B., Prescott, T. J. and Bednar, J. A. (2010), 'Modeling the emergence of whisker direction maps in rat barrel cortex', *PloS One* **5**(1), e8778.

Wolff, K. E. (1993), 'A first course in formal concept analysis: how to understand line diagrams', *Advances in Statistical Software* **4**, 429–438.

Woolsey, T. A., Welker, C. and Schwartz, R. H. (1975), 'Comparative anatomical studies of the SmL face cortex with special reference to the occurrence of "barrels" in layer IV', *Journal of Comparative Neurology* **164**, 79–94.

Wu, G. K., Arbuckle, R., Liu, B.-H., Tao, H. W. and Zhang, L. I. (2008), 'Lateral sharpening of cortical frequency tuning by approximately balanced inhibition', *Neuron* **58**, 132–143.

Wu, G. K., Tao, H. W. and Zhang, L. I. (2011), 'From elementary synaptic circuits to information processing in primary auditory cortex', *Neuroscience and Biobehavioral Reviews* **35**(10), 2094–2104.

Wu, Q., McGinnity, T., Maguire, L., Glackin, B. and Belatreche, A. (2006), 'Learning under weight constraints in networks of temporal encoding spiking neurons', *Neurocomputing* **69**, 1912–1922.

Yoon, S.-Y. and Lee, S.-Y. (1999), 'Training algorithm with incomplete data for feed-forward neural networks', *Neural Processing Letters* **10**(3), 171–179.

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.

Zhang, Z. and Andreou, A. G. (2008), 'Human identification experiments using acoustic micro-Doppler signatures', *Proceedings of the Argentine School of Micro-Nanoelectronics, Technology and Applications, EAMTA 2008* pp. 81–86.

Zucker, R. S. and Regehr, W. G. (2002), 'Short-term synaptic plasticity', *Annual Review of Physiology* **64**, 355–405.