# MECHANISMS OF SHORT TERM AND LONG TERM MEMORY IN CORTEX: NEURAL FIELDS AND SYNAPTIC PLASTICITY

by

Dominic Standage

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
July 2007

*Your file Votre référence*
*ISBN: 978-0-494-31493-7*
*Our file Notre référence*
*ISBN: 978-0-494-31493-7*

DALHOUSIE UNIVERSITY

To comply with the Canadian Privacy Act the National Library of Canada has requested that the following pages be removed from this copy of the thesis:

Preliminary Pages

Examiners Signature Page (pii)

Dalhousie Library Copyright Agreement (piii)

Appendices

Copyright Releases (if applicable)

# Table of Contents

## List of Figures

## Abstract

Computational models have a long and fruitful history in neuroscience, addressing the information-theoretic properties of neural substrates and making predictions to guide physiological and psychological enquiry. Computational analyses and simulations have been extensively applied to the study of memory and are rightly considered core methods of this broad research field. Here, we study memory at two levels of analysis, investigating neural representations in cerebral cortex and synaptic plasticity.

Neural field models have been used to study numerous cortical regions and functions. Their use is thus compatible with the view that cortex is, in general, architecturally and mechanistically uniform, regardless of the function of specific cortices. In this regard, we use neural field models to investigate interactions within and between regions of cortex. We demonstrate parameter regimes that correlate with experimental findings on the active maintenance of short term memories and the spatial distribution of selective visual attention, making predictions for further experimental research.

Innumerable computational models show that learning results from activity-dependent changes in synaptic strength. In this regard, we extend earlier analyses of correlations between pre- and post-synaptic spike timing to the case of highly correlated spikes, showing a unique relationship between the contributions of spike timing and spike rate to plasticity. We also demonstrate that the common use of weight-dependent terms in spike-time dependent learning rules is not supported by experimental data, and we show that the data on which these terms are based may reflect experimental artifacts. Our research addresses important questions about cortical processing, and our results are compatible with theories of cortex in which attention and novelty-detection emerge from activity-dependent plasticity between hierarchically-arranged, bidirectionally-connected cortical regions.

## Acknowledgements

On completing my thesis, I owe a debt of gratitude to a number of people. I thank Thomas for his tirelessness as a teacher and supervisor, and for the innumerable hours we've spent discussing the brain and how (we think) it works. I hope our discussions have been as much fun for him as for me. I thank Ray, Alan and Stefan for their help and for the manner in which it was given. Their impact on my studies has been huge and I'm lucky to have had their interest, enthusiasm and expertise. I thank Mike for his time, effort and patience during completion of my studies. I thank my parents for their continual encouragement and for helping me believe that I wasn't following some sort of pipe dream. I thank Genevieve for her understanding. I thank Henry and Madeleine for comprehending that Dad was up to something a bit difficult, and for just being Henry and Maddy. Most of all, I thank Selina for keeping it all together. Thanks for being somewhere you didn't want to be, doing things you didn't want to do, and for coming to like it after all. We did it.

# Chapter 1

# Introduction

The brain is an intricate set of interacting, heterogenous structures, the functions of which are mediated by an equally intricate set of physiological processes. The role of computational methods in neuroscience is to determine how information processing arises from these myriad details. Our methods are applied at the levels of physiology and psychology, but also serve to mediate these two ends of the spectrum of neuroscience. In this regard, it is not enough to correlate anatomy and physiology with cognition and behaviour. To understand brain function, we must determine *how* anatomy and physiology mediate these psychological phenomena.

Biophysical explanations of psychology are most generally addressed by the field of cognitive neuroscience. Among computational methods, abstract systems-level models are sometimes grouped under the more general heading of cognitive science, such as connectionist models [102] that do not address (or very vaguely address) the biological mechanisms underlying their function. These models have been influential in cognitive psychology because they make predictions about cognitive function that guide experimental research. At their level of abstraction, detailed biology misses the point. At the other end of the spectrum, more detailed, biophysical models have a long history in neuroscience, perhaps best exemplified by the pioneering work of Nobel Laureates Alan Hodgkin and Andrew Huxley [56]. Such generic categories will of course overlap. Regardless of their level of abstraction, models are only useful if they make predictions. In the work below, we use computational methods to study memory and memory-serving processes at a number of levels. In all cases, our investigations serve biophysical, mechanistic explanations of psychological function.

## 1.1 Memory as a Biological and Psychological Phenomenon

Memory refers to the acquisition, storage and retrieval of information. The brain achieves these three aspects of memory by associating neural representations with one

another. In this way, organisms learn the causal relationships of the world by exposure to events, continually adjusting and augmenting their internal representations when learned associations do not match reality. Virtually all views of learning share this common thread, from simple conditioning paradigms to cognitive theories [118].

A common mechanism does not imply a single, monolithic memory system. Such a view was dispelled long ago by observations of spared and impaired learning abilities in patients with hippocampal damage [97]. Modern theories of learning typically divide memory into multiple psychological systems supported by different neural substrates. From a psychological perspective, different systems learn different kinds of information, performing their functions for different lengths of time. From a biological perspective, memory systems are interconnected neural structures suited to particular information processing demands, either storing information or participating in its storage in other structures. Similarities between these perspectives are presumed to reflect the same phenomena at different levels of analysis, where observations from one perspective serve to validate those of the other [63].

Virtually all experimental paradigms for the study of memory reveal specialized components serving different functions. Even basic, associative conditioning involves multiple memory systems, as the cerebellum, hippocampus and amygdala are known to contribute differentially to classically conditioned responses [118]. Such components have been extensively studied, with computational methods especially prominent in explaining the information-theoretic properties of biological structures [82, 7]. A multiple memory systems framework, however, does not imply interactions are limited to those between clearly defined components. To begin with, the anatomical and systemic boundaries between memory systems are blurry. For example, the hippocampus and MTL are often considered together as a system supporting declarative memory, but interactions between structures within this system are extremely complex and include cooperative and competitive relationships [63]. In cortex, bottom-up, top-down and lateral interactions occur at every processing stage, providing the foundation for information processing in this memory system. Cortical interactions and the processes that shape them are the focus of the work below.

## 1.2 Short Term and Long Term Memory in Cortex

In the present context, I draw a fundamental distinction between short term and long term memories. By short term memory (STM), I simply refer to active neural representations. The activity of a group of neurons represents something, so the memory lasts as long as the representation. By long term memory (LTM), I refer to the encoding of information in the strength or *weight* of synapses. LTM's may last different lengths of time, but regardless of their longevity, learned weights facilitate the retrieval of previously active neural representations. That is, LTM's serve retrieval of STM's.

Cortex is a hierarchical, bidirectional memory system. Detailed, low-level information flows uphill from sensory cortices and is combined hierarchically as these representations converge on higher cortical regions. For instance, in visual processing, V1 neurons respond to basic features such as edges, which are combined to represent increasingly complex forms until objects are represented in inferior temporal cortex (IT). The formation of LTM's ensures that features that regularly occur together are represented by cell assemblies.

A bottom-up description of cortical processing captures compositionality, but it does not capture the neural mechanisms enabling expectations and attention. For these phenomena to emerge, top-down processing is required. Connectivity in cortex is highly bidirectional. Not only do inputs converge on higher cortices from lower cortices, but higher regions project back down to lower regions. These feedback connections are just as abundant as compositional pathways. The formation of LTM's occurs in both directions, ensuring that higher cortical representations favour activation of the components that define them. For instance, when an object is represented in IT, it projects to V4, V2 and V1 where activity represents increasingly detailed components of the object. Top-down activity thus represents learned expectations that prime bottom-up features for detection, consistent with 'biased competition' models of attention [33], discussed in Chapter 3.

While the function of cortex is specialized by region, its structure is remarkably uniform. This observation has lead to the belief that cortical regions process information in the same way, regardless of their respective functions [89, 132]. This belief is supported by numerous experiments showing functional adaptation in cortex [16].

Consistent with these observations, in Chapters 2 and 3, we use the same class of model to simulate processing in prefrontal cortex (PFC) and posterior parietal cortex (PP) respectively. The former addresses the active maintenance of STM's. The latter addresses the influence of PFC on PP. The context of Chapter 3 is provided by a model of visual attention [31, 26] that is consistent with theories of cortex more generally [21, 44], in which attention is an inherent property of layered, bidirectional processing. According to this view, cortical STM's compete with each other by virtue of bottom-up, top-down and lateral interactions.

### 1.2.1 Centre Surround Neural Field Models and Short Term Memory

The model we use in Chapters 2 and 3 is most generally labeled as a centre surround neural field (CSNF) model. CSNF's have been used to model information processing in a number of brain regions, including head direction cells in the limbic system [136], place fields in the hippocampus [10, 100, 116], and the initiation of saccadic eye movements in the superior colliculus [122]. Their applicability to such a wide variety of processing tasks suggests they capture a fundamental property of neural computation.

In Chapters 2 and 3, we use CSNF's to model cortex. CSNF's are commonly used in cortical simulations, including the modelling of hypercolumns in sensory cortex [51], decision fields in frontoparietal cortex [23], active maintenance of STM's in PFC [25, 121] and topographic maps in PP [31]. Within the scope of cortical modelling, CSNF's capture the properties of a small patch of cortex representing some feature space, the size of which is thought to be about $1mm^2$ [51]. In this regard, hypercolumns in lower sensory cortices are no different than small patches of higher sensory and association cortices that mediate more complex feature spaces. This view is consistent with Wilson and Cowan's (1973) proposal that different parameterizations of CSNF's foster dynamic regimes suited to specific cortical functions [131], with Grossberg's use of CSNF's in Adaptive Resonance Theory (ART) [46, 47, 21], and with the model of selective visual attention by Deco et al. [31, 26], providing the context for Chapter 3. It is also consistent with the view that a common algorithm operates throughout cortex.

Our CSNF model is a fully connected network with a shift invariant Gaussian

weight profile and activity-dependent inhibition. Under this configuration, the weight between any two nodes is a Gaussian function of their distance. We subtract a constant value from this positive-valued weight matrix, yielding local excitatory weights and distal inhibitory weights (from the perspective of any node). Nodes thus support each other locally and inhibit each other distally. This weight profile may be criticized on the grounds that connections in the brain tend to have the opposite arrangement, but the model captures the *effect* of local cooperation and distal competition. This effect may be achieved in several ways, including nonselective inhibition in response to selective excitation [31, 21] and laminar circuitry [48].

CSNF's can operate in several parameter regimes [6], but an important consequence of our configuration is that a single region of activity will dominate the network following transient input. This *winner-take-all* (WTA) property is not necessarily the case for other weight profiles. Under our configuration, the long range inhibition provided by Gaussian weights (minus the constant of inhibition) serves distal, competitive interactions throughout the field. CSNF's are most commonly used in this WTA manner, clearly appropriate for the modelling of hypercolumns, decision fields and head direction cells, among their uses cited above.

With respect to terminology, please note that we refer to our model as a continuous attractor neural network (CANN) in Chapters 2 and 3 because each node in the network corresponds to one of a quasi-continuous manifold of point attractors. A description of attractor dynamics is beyond the scope of this document, but can be found in a number of books and articles, including [119, 55]. Additionally, we refer to an active region of our model as an 'activity packet' in Chapter 2 and an 'activity bubble' in Chapter 3. These terms refer to the same thing.

### 1.2.2 Stabilization of Short Term Memories in Centre Surround Neural Field Models

The symmetric, shift-invariant weight profile of our model poses a problem. Any noise in the weights leads a to drift of the activity bubble, corresponding to an inaccuracy of the STM. Because information processing in the brain is noisy, a perfectly symmetric, shift-invariant weight profile is unrealistic. A number of studies have addressed this issue [136, 75, 121] and several means of stabilization have been proposed, including

intrinsic cellular mechanisms [19] and the voltage dependence of NMDA receptors [75, 121]. Stabilization, however, comes at the expense of the WTA property of CSNF models, crucial to their use in so many domains.

In Chapter 2, we address the relationship between stabilization and the WTA property of the model. This work can be viewed from two perspectives. As a model of a small patch of cortex, we require CSNF's to support non-drifting STM's and to implement a WTA function. At a higher level of abstraction, multiple representations in a CSNF model can represent a number of interacting STM's.

### 1.2.3 Interactions in Prefrontal Cortex, Posterior Parietal Cortex, and Attentional Processing

The active maintenance of STM's in strategic tasks is sometimes referred to as working memory (WM). I prefer to think of WM as a kind of workspace, not unlike the role of random access memory in digital computers. According to this old analogy, LTM's equate with read-only memory, accessed and buffered during performance of a task. Interim calculations required for successful performance of the task are also cached in the buffer, and may be subject to long term storage. Not surprisingly, WM tasks require active maintenance of STM's. These STM's may represent transient sensory stimuli, task instructions, goals, calculations and so forth, but in each case, the information is required for completion of the task.

Numerous experiments correlate pre-frontal cortex (PFC) with active maintenance of STM's in WM tasks. PFC is reciprocally connected to sensory, motor and association cortices, as well as a number of subcortical structures, so is well positioned to buffer different types of information and to influence different types of processing. Regions within PFC are also highly interconnected. In this regard, the long range inhibition in our model allows us to equate activity bubbles in a single network with STM's in different regions of PFC, mimicking the interaction between coupled networks. This approach was taken by Trappenberg (2003) [121], where the number of STM's in a stabilized CSNF model was shown to be consistent with the observed constraints of WM. With respect to the influence of PFC on processing in other cortical regions, these constraints have important implications. It is well known that receptive fields broaden at increasingly higher levels of cortical processing, creating greater

interference between representations. As described above, higher level representations propagate downward to their contributing features, giving them an advantage in their ongoing competition. That is, lateral competition at higher cortical levels influences feature processing at lower levels [31, 21]. This fundamental property of cortical processing provides an explanation for Desimone and Duncan's (1995) [33] biased competition model of attention, as demonstrated by Deco and colleagues [31]. PFC is particularly influential because well-positioned, sustained STM's have a competitive advantage.

In Chapter 3, Deco's model provides the context for our investigation of the influence of PFC on the spatial distribution of selective visual attention (SVA). In this case, STM's in PFC are equated with task instructions and foreknowledge of stimulus characteristics in laboratory experiments [8, 90]. We simulate PP with a CSNF model, showing that the ability to actively maintain STM's in PFC can explain psychophysical and electrophysiological evidence for noncontiguous allocation of SVA.

## 1.3 Synaptic Plasticity and Long Term Memory

I have stated that LTM's serve retrieval of STM's, but the relationship between these two categories of memory is not unidirectional. Long term memory is the result of activity-dependent change in synaptic strength, or *synaptic plasticity*, as originally proposed by Hebb [54]. STM's thus serve the encoding of LTM's. We demonstrate this fundamental learning principle in Chapter 2, training the recurrent weights of our CSNF model with a Hebbian rule.

According to Hebb's postulate, when pre-synaptic activity contributes to post-synaptic spiking, the strength of synapses mediating this activity increases. The corollary of this hypothesis, also anticipated by Hebb, is that negatively correlated pre- and post-synaptic spiking leads to a decrease in synaptic strength. Plasticity is now well established in the forms of long term potentiation (LTP) [15] and long term depression (LTD) [79]. Early experiments showed LTP and LTD as a function of the rate of pre-synaptic stimulation, where high- and low-rate stimulation were shown to yield potentiation and depression respectively. Subsequent experiments using *pairing* protocols showed LTP as a function of pre- and post-synaptic activity [80, 28]. These protocols provide a more direct model of Hebbian learning than pre-synaptic rate

protocols, though they manipulate post-synaptic depolarization and not spiking per se.

More recent experiments demonstrate plasticity as a function of the timing of pre- and post-synaptic spiking [72, 81, 99]. When pre-synaptic stimulation is repeatedly paired with a post-synaptic spike, the direction of synaptic change (potentiation or depression) typically depends on the order of pre- and post-synaptic onset. When pre-synaptic stimulation precedes a post-synaptic spike (pre-post), synapses are potentiated. When the post-synaptic spike comes first (post-pre), synapses are depressed. The magnitude of synaptic change depends on the latency of pre- and post-synaptic stimulation in either direction, within a time window on the order of tens of milliseconds [99, 109, 41].

STDP has made a huge impact in the neuroscience community, largely because the temporal asymmetry of STDP data captures the causality of neural firing. That is, STDP is truly Hebbian. Learning rules based on these data have become a staple of the computational community. In simulations with spiking neurons, each pre- and post-synaptic spike pairing provides a potential 'event' for the enaction of these rules. Furthermore, the asymmetry of STDP rules readily lends itself to sequence learning, believed to underlie important aspects of hippocampal [82, 73, 59, 74, 129, 78, 5, 38] and cortical [13, 3, 132] function.

A known problem for Hebbian rules is that weights 'runaway' without a means to limit their growth. Unbounded weights are not only biologically unrealistic, but they lead to unconstrained post-synaptic activity and the consequent breakdown of information processing [52]. Several approaches to this problem appear in the literature, including activity-dependent plasticity thresholds [14, 117] and synaptic scaling [108]. More commonly, Hebbian rules include a weight-dependent term, where plasticity is not only a function of neural activity, but also of initial weight.

Hebbian rules with a weight-dependent potentiation term are supported by several experiments showing a dependence of LTP on initial synaptic strength [99, 30, 87], including one STDP study [99]. The asymptotic consequences of weight-dependent STDP rules have been extensively studied, but this body of work is limited to the case of uncorrelated or weakly correlated pre- and post-synaptic spiking. This state of affairs is surprising, given that Hebbian learning results from correlated pre- and

post-synaptic activity. In Chapter 4, we address this shortcoming, fitting two weight-dependent STDP rules to data and determining their asymptotic consequences for correlated spiking. Our two rules capture different assumptions about weight-dependent STDP data.

With respect to plasticity, we also address the relationship between spike rate and spike timing in Chapter 4. This relationship is an important one. It is widely believed that different plasticity protocols address a common physiology. When rules derived from data do not accurately predict plasticity under other protocols, they not only identify gaps in our knowledge, but they also indicate the conditions under which laboratory protocols may serve to model activity in the brain. In cases where no data exist for comparison, predictions call for future experiments.

By fitting our rules to data, we also address the biological realism of weight-dependent STDP rules as they commonly appear in the literature [66, 128, 49, 17]. While there is a general consensus that STDP is not fully understood as it relates to spike trains more complex than simple pairings at low frequency, the literature gives the general impression that weight-dependent STDP is a *fait accompli* for this simplified case. Chapter 4 shows that the biological realism of these rules is questionable. We extend this work in Chapter 5, demonstrating several problems for weight-dependent STDP rules in simulations of associative learning and identifying issues for the interpretation of weight-dependent plasticity data more generally. In Chapter 6 (unpublished), we show that these data may be confounded by the effects of populations of synapses.

In Chapter 7, I summarise Chapters 2 to 6 and discuss some implications of this work for models of cortical memory and the interdependent roles of Hebbian learning, attention, and competitive interactions in cortex.

# Chapter 2

## Multi-Packet Regions in Stabilized Continuous Attractor Networks

Thomas P. Trappenberg and Dominic I. Standage, Neurocomputing, 65-66:617-622, 2005

## Abstract

Continuous attractor neural networks are recurrent networks with center-surround interaction profiles that are common ingredients in many neuroscientific models. We study realizations of multiple non-equidistant activity packets in this model. These states are not stable without further stabilizing mechanisms, but we show they can exist for long periods. While these states must be avoided in winner-take-all applications, they demonstrate that multiple working memories can be sustained in a model with global inhibition.

## 2.1 Introduction

Wilson and Cowan [131] derived a description of the population dynamics of neurons with excitatory and inhibitory pools coupled with center-surround interaction profiles. They identified various dynamical regimes in these networks and speculated that these regimes might map to different brain areas, including thalamic nuclei, visual neocortex, and prefrontal cortex. Continuous attractor neural networks (CANNs) are now common ingredients in models of information processing in the brain and are regarded as the principle model of cortical hypercolumns [51], place and head direction cells in the limbic system [136], and working memory [25].

The CANN model is most often used in a parameter regime where a single activity packet (also called bubble or bump) can be sustained without external input. The model then implements a winner-take-all function, which is appropriate for the modelling of place fields and feature representations in hypercolumns. Samsonovich [103] and Battaglia and Treves [10] have extended the model to multiple feature spaces, and our group has shown that many activity packets can be sustained simultaneously when the model is augmented with biologically realistic stabilization mechanisms [115]. Multiple bubbles can also be sustained within a single topographic feature map with such stabilization mechanisms [121], which is relevant to the modelling of working memory. Here, we limit our study to a single feature space, and study the dependence of multiple activity packets on the amount of activity dependent global inhibition and the strength of the stabilization. We also show that multiple bubbles can be sustained for a considerable length of time without stabilization.

## 2.2 Methods

We consider a basic recurrent rate model with $N$ nodes, though corresponding networks with spiking neurons have similar properties. The time evolution of the membrane state $u_i$ of a node with index $i$ is given by

$$\tau \frac{\mathrm{d}u_i(t)}{\mathrm{d}t} = -u_i(t) + \sum_j w_{ij} r_j(t) \Delta x + I_i^{\mathrm{ext}}(t), \tag{2.1}$$

where $\tau$ is a time constant, $I_i^{\mathrm{ext}}$ is the external input applied to the network, $\Delta x = 2\pi/N$ is a scale factor, and $r_i$ is a rate that is related to $u_i$ by sigmoidal gain function

$g(u) = 1/(1 + \exp(-\beta(u - \alpha)))$ with a slope parameter $\beta = 0.1$ and firing threshold $\alpha$. The weight matrix $\mathbf{w}$ is determined in a learning phase with Hebbian learning, $w_{ij} \propto \sum_\mu r_i^\mu r_j^\mu$ on patterns with index $\mu$. Such recurrent models are often studied after training on random patterns, resulting in networks with discrete attractors. In contrast, we study this model trained with well organized Gaussian patterns, where each pattern is centered around a different node in the network, $\mu = 1, ..., N$. This results in an excitatory Gaussian weight matrix with width $\sigma_w = \sqrt{2}\sigma_\mathrm{r}$,

$$r_i^\mu = A_\mathrm{r} \mathrm{e}^{-((i-\mu)\Delta x)^2/2\sigma_\mathrm{r}^2} \quad \rightarrow \quad w_{ij}^\mathrm{ex} = A_\mathrm{r}\sqrt{\pi}\sigma_\mathrm{r} \mathrm{e}^{-((i-j)*\Delta x)^2/4\sigma_\mathrm{r}^2}, \tag{2.2}$$

which is then augmented by an inhibition constant $C$ describing the activity dependent inhibition of an inhibitory pool of neurons and scaled by global strength constant $A_\mathrm{w}$,

$$w_{ij} = A_\mathrm{w}\left(\frac{1}{A_\mathrm{r}\sqrt{\pi}\sigma_\mathrm{r}} w_{ij}^\mathrm{ex} - C\right). \tag{2.3}$$

A well known problem in CANN models is that noise in the weight matrix leads to a drift of the activity packet [124]. It has been argued that drift slows down with increasing network sizes [136, 25], and activity dependent bistabilities in the excitability of neurons have also been shown to stabilize activity packets [19]. We implement the stabilization by a change of the threshold,

$$\Delta\alpha = \alpha_0 \Theta(u), \tag{2.4}$$

where $\Theta(u)$ is the Heaviside function. This stabilization mechanism is sufficient to sustain neural activity after transient external stimuli without further (excitatory or inhibitory) support by other nodes in the neural layer [75]. A network without the lateral connections typical of CANN models, however, cannot implement the competition between stimuli that is essential to much of the brain processing for which CANNs were proposed. Here we follow our previous speculation that competition is the basis for the limited capacity of working memory [121], and study the dependence of the strength of the stabilization on multiple simultaneous activity packets in CANN models.

## 2.3 Results

First, we consider the model without stabilization. Appropriate values for $C$ and $A_\mathrm{w}$ must be chosen to sustain an activity packet following transient input. This is

illustrated in Figure 2.1A which shows the maximum node activity within the activity packet at time $t = 100\tau$ for various values of $C$ and $A_w$. If inhibition is too weak then the entire network becomes active; too much inhibition shuts off all network activity. Additionally, the scaling parameter $A_w$ must be strong enough to sustain an activity packet, but this constraint does not apply when a threshold activation function is used, as in the studies by Amari [6].

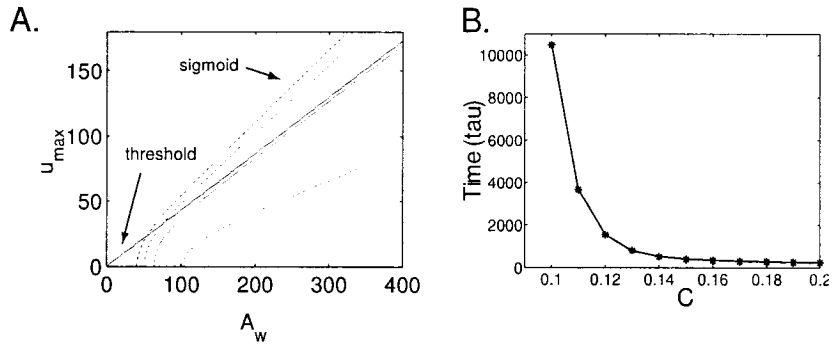


Figure 2.1: (A) The maximal value of node activity $u$ as a function of weight scaling factor $A_w$ for different values of inhibition constant $C$. The curved solid line shows results with an inhibition constant of $C = 0.4$ for simulations with sigmoidal gain function $g(u) = (1 + \exp(-0.1u))^{-1}$. The straight solid line shows results from simulations with a threshold gain function $g(u) = \Theta(u)$ with otherwise unchanged parameters. Results from simulations with the sigmoidal gain function and different equidistant values for inhibition constant $C$ are shown as dashed lines, from $C = 0.2$ (top) to $C = 0.7$ (bottom). (B) Time ($\tau$) required for two activity packets to merge as a function of global inhibition ($C$) in the 1000 node network with input activity around nodes 100 and 580.

Amari's analysis shows that a single activity packet is a stable attractor state in CANN networks [6], however, considerable time can be required before two or more externally driven activity regions merge or until one activity packet becomes dominant. The time required for a single maximum to emerge from two activated regions is shown in Figure 2.1B. The activity profile $\mathbf{u}$ was measured at each iteration of $dt/\tau$, where only two changes in direction of this curve determined a one-bubble state in the periodic network. The network was initialized with binary input bands around nodes 100 and 580, where one input band was 1% stronger than the other to brake the symmetry of input activity.

An example of 4 meta-stable asymmetric activity packets is shown in Figure 2.2A

for a simulation with parameters $\sigma_r = 2\pi/80$, $A_w = 300$, and $C = 0.05$. No stabilizing bistability was implemented in this simulation. The results of a simulation with slightly increased inhibition constant, $C = 0.08$, are shown in Figure 2.2B. Inhibition is now strong enough to introduce sufficient competition in the network such that one activity packet disappears within the time of $t = 100\tau$ following removal of the external stimulus.

We now consider the network with stabilization, and refer to the area where $n$ equidistant activity packets do not considerably decay within time $t = 100\tau$ as an $n$-phase. The results of several simulations with varying values of $a_0$ and $C$ are summarized in Figure 2.2C for different numbers of initial activity bands. Stabilizing more than 4 activity packets with the current parameters is impractical.

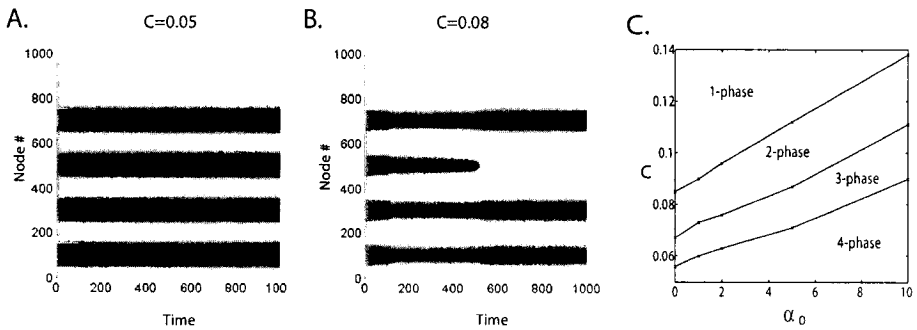


Figure 2.2: (A) Network activity over time for a simulation with asymmetric external input at nodes 100, 300, 500, and 700 in a 1000 node network until $t = 100\tau$ with inhibition constant $C = 0.05$. (B) Corresponding simulations for $C = 0.08$. (C) Transitions between $n$-phases dependent on inhibition $C$ and threshold adjustment $\alpha_0$. The transitions between the $n$ and $n+1$ phases were thereby studied with $n+1$ of the original 4 input bands.

## 2.4 Discussion and Conclusion

The results of the simulations shown in Figure 2.2 seem to contradict theoretical proof by Laing et al. [70, 20] that without stabilizing mechanisms, more than one asymmetric activity packet cannot be sustained in the CANN model. However, our results simply show that multiple asymmetric bubbles can be sustained for considerable periods of time.

Figure 2.2C indicates that the effect of inhibition in the model is approximately inverse to that of stabilization. Increasing stabilization (increasing $\alpha_0$) can lead to the stabilization of multiple activity packets, while increasing competition (increasing $C$) can destroy the existence of multiple activity packets. While this relationship appears linear, there are other parameters that effect stabilization in our model. Firstly, in the infinite time limit, we know there can be only one stable activity packet in the $\alpha_0 = 0$ limit. The relationship must therefore deviate considerably from the linear case for longer simulation times close to the $\alpha_0$ limit. Secondly, the width of the weight profile and the strength of connectivity effects the number of activity packets sustained by the network, here modelled as $\sigma_r = 2\pi/80$ and $A_w = 200$ respectively. A wider weight profile and weaker connectivity serve to impede stabilization. Increasing the asymmetry of activation bands will further diminish stability of multi-packet solutions. Thirdly, the spatial discreteness inherent in numerical simulations has a stabilizing effect on the model. This effect increases with the sharpness of the activity packet profile, as only a Gaussian profile can be moved continuously with constant support over an equidistant lattice. This effect is directly related to strength parameter $A_w$, as increasing $A_w$ saturates the activity packet profile.

While much recent attention has been paid to stabilizing activity packets in CANN models, little research has focused on balancing stabilization such that the model retains its winner-take-all functionality. Strong stabilization effectively partitions the network into a series of local networks with winner-take-all characteristics similar to networks with short range inhibition [127]. As such, it is possible to stabilize large numbers of activity packets under the CANN model. Alternatively, the use of low levels of global inhibition permits the co-existence of a small number of activity packets for finite periods in the absence of stabilization. We speculate that unlike models with short range inhibition, global inhibition in the CANN model parallels the effect of interaction between brain areas, limiting the number of simultaneous activity packets in accordance with the limited capacity of working memory.

# Chapter 3

## Modelling Divided Visual Attention with a Winner-Take-All Network

Dominic I. Standage, Thomas P. Trappenberg and Raymond M. Klein, Neural Networks, 18(5-6):620-627, 2005

## 3.1 Addendum to Standage, Trappenberg and Klein (2005)

In this paper, we model the spatial distribution of selective visual attention with a centre surround neural field (CSNF) model, here referred to as a continuous attractor neural network (CANN). From the perspective of hierarchical, bidirectional cortical processing, this paper addresses the interactions between cortical regions, where top-down and bottom-up activity converge on higher sensory cortex. Specifically, we model posterior parietal cortex (PP), where a topographic map of the visual field is hypothesized to determine the location of visual attention [123]. The context of this work is provided by the work of Deco *et al.* [31] who instantiate the biased competition hypothesis of attention [33], where representations compete at all processing levels. Competition is believed to be necessary due to limited processing resources, and is biased at lower cortices by activity propagating downward from higher cortices (described in Chapter 1).

In the context of hierarchical, bidirectional memory, biased competition is unavoidable. Bottom-up learning provides composite representations at higher levels of the cortical processing hierarchy. Top-down learning provides expectations. Higher level representations thus propagate activity to the features than define them, providing a bias in the ongoing competition at the lower processing level. In this regard, we do not differentiate the spatial dimension from other feature dimensions. In the spatial dimension, the widening of receptive fields at higher levels of visual processing reflects spatial composition, just as complex objects are composed of other kinds of features. Described in this way, attention is an emergent property of hierarchical, bidirectional processing and its 'allocation' is the outcome of cortical interactions.

The majority of experimental evidence supports the theory that visual attention is a unitary phenomenon, limited to one region of space at a time. Here, we model two experiments showing a more flexible allocation of visual attention. At a minimum, to provide support for this kind of flexible attentional processing, measurements of visual attention must be recorded at three contiguous locations. If visual attention is spatially contiguous, then measurements at the middle location will be as high as those at the outer locations. If measurements are higher at the outer locations than in between, then to the degree that these measurements differ, visual attention may be said to be divisible under the conditions of the experiment.

Awh and Pashler [8] provided psychophysical evidence for the noncontiguous allocation of attention. Their general experimental procedure was as follows, where differences between the specific experiments of their study are described below. Subjects were shown five consecutive screens on a computer monitor, where a 25x25 grid determined the layout of a fixation point and target and distracter locations. The grid itself was not seen. The first screen contained a dot at the centre of the grid and subjects were instructed to fixate their gaze at this location. Two 'equals' signs (=) were flashed either side of fixation on the second screen, offset by a single grid square, and subjects were instructed to attend to these cued locations. On the third screen, 2 digits appeared at the attended locations on 80% of trials (valid trials). On the remaining 20% of trials, one of the digits appeared in between the attended locations and the other was offset from this location by two grid squares, perpendicular to the direction of the cues (invalid trials). The subjects' task was to identify the 2 digits. The first three screens of their study are depicted in Figures 3.4 and 3.5. The fourth screen showed a mask of 'hash' symbols (#) and the fifth screen showed a dot at each of the 23 locations that did not contain a digit, and a question mark at each digit location (postcues).

Experiment 1 showed 23 distracter letters on the third screen, accompanying the 2 target digits. Thus, distracters appeared over the entire non-target grid, illustrated in Figure 3.4. They used a full-field noise mask on the fourth screen, so all grid squares were covered by a 'hash' symbol. In their Experiment 4, they removed all the distracter letters. In their Experiment 4a, they further removed the 'hash' symbols at all non-target locations, illustrated in Figure 3.5. These manipulations of their experimental paradigm addressed a central question posed by their study: does visual attention enhance processing at attended locations or does it suppress noise at unattended locations? By removing the distracter letters in Experiment 4, they aimed 'to preclude the need for attentional gating'. The removal of the non-target mask in Experiment 4a aimed to further identify the role of noise in attentional processing.

To measure visual attention, Awh and Pashler recorded the accuracy of their subjects in identifying the two digits. In Experiment 1 (with noise) they found that subjects performed significantly better at the cued locations (valid trials) than at the invalid locations. In the present context, the location of primary interest is the

invalid location between the two cues. Their measure of visual attention was higher at the outer locations than at the intervening location, satisfying our general criteria for experimental evidence for divided attention.

In Experiment 4 (removal of distracters) the difference between the subjects' accuracy at the outer and middle locations was significantly reduced. In Experiment 4a (removal of the non-target mask) subjects performed as well at the middle location as at the surrounding cued locations. Indeed, they performed slightly better. Awh and Pashler thus concluded that their subjects were able to allocate their visual attention non-contiguously in space, and that this flexible allocation of attention may reflect the suppression of noise at distracter locations, rather than the enhancement of processing at target locations.

Our simulations do not consider the fourth and fifth visual screens used in this study. We model the realtime interaction of top-down (volitional) and bottom-up (sensory) inputs to PP, governed by the network dynamic, and we assume that their masking and postcue screens invoke the requirement of trace activity for the identification of the digits, in addition to object-recognition processing. Activity in PP is hypothesized to determine the location of selective visual attention, not short term memory and object recognition. Our simulations are also limited to the case of horizontal target locations, not vertical locations. In our model, there is no difference between these two cases, but the results of Awh and Pashler described here pertain to the former case only. These authors attribute differences between these two cases to interhemispheric and intrahemispheric interactions.

Where Awh and Pashler provide psychophysical evidence for the division of visual attention over hundreds of milliseconds, Müller *et al.* [90] provide electrophysiological evidence for a more sustained division of attention ($\sim$ 3s). They recorded steady state visual evoked potentials (SSVEP's) mapped to four noncontiguous locations in the visual field. The SSVEP is an electroencephalographic (EEG) measurement in response to a flickering visual stimulus. In the present context, this measurement has two key properties. Firstly, the EEG waveform has the same frequency as the flicker rate. Secondly, the waveform is amplified when the location of the stimulus is attended. In their experiment, a rectangle at each location flickered at a different rate, providing an EEG signature for its corresponding location.

Within the flickering rectangles, a randomised sequence of symbols was flashed in synchrony, depicted in Figure 3.1. Subjects were instructed to attend to two of the four locations and to press a button when the same symbol occurred simultaneously at the attended locations. Importantly, the frequency of the changing symbols was high enough to rule out the possibility that subjects were switching their attention from one location to the other, according to generally accepted time constraints for reallocation of visual attention. Clearly, on trials when subjects were instructed to attend to two locations either side of an intervening location, our general criteria for experiments investigating the division of visual attention are again satisfied.

Additionally, to strengthen their results, Müller *et al.* recorded target detection rates (TDR's) at each of their four locations in the visual field. They found that SSVEP amplitudes and TDR's were significantly lower at locations between surrounding attended locations. Indeed, for both electrophysiological and psychophysical measurements, the relative difference between measurements at the attended and unattended locations in these 'split' trials was similar to measurements in 'adjacent' trials, where the attended locations were next to each other in the hemisphere opposite to the unattended locations.

Given the description of CSNF models in Chapters 1 and 2, our model makes several predictions *a priori*. We can expect enhanced processing at attended locations *and* suppression of noise at unattended locations, as investigated by Awh and Pashler, due to the local cooperation and distal competition in the model. That is, enhancement and suppression are inextricably linked in the CSNF model. Furthermore, in the context of biased competition, the role of a topographic map in PP is to integrate bottom-up and top-down signals and to propagate activity downward to lower visual processing areas, enhancing processing at regions in space mapped by PP. Once again, enhanced processing in these regions will suppress activity in other regions by lateral inhibition.

We can also expect a contiguous allocation of attention in the case of transient inputs, due to the winner-take-all nature of the network, bearing in mind our results in Chapter 2 showing that the time required for a winner to emerge is parameter-dependent. Chapter 2 also shows that the model can sustain a bubble indefinitely following transient input, a feature commonly used to model the active maintenance

of representations in prefrontal cortex (PFC). In this way, top-down, volitional signals from PFC may be expected to determine the flexible allocation of visual attention in the present framework of biased competition. Our simulations assume these signals are maintained in PFC, representing the locations to which Awh and Pashler and Müller *et al.* instructed their subjects to attend. Finally, our model does not make quantitative predictions about human performance data. Rather, we make qualitative predictions based on the premise that activity propagating downward from PP biases the competition between representations of stimuli in lower visual cortices.

## Abstract

Experimental evidence on the distribution of visual attention supports the idea of a spatial saliency map, whereby bottom-up and top-down influences on attention are integrated by a winner-take-all mechanism. We implement this map with a continuous attractor neural network, and test the ability of our model to explain experimental evidence on the distribution of spatial attention. The majority of evidence supports the view that attention is unitary, but recent experiments provide evidence for split attentional foci. We simulate two such experiments. Our results suggest that the ability to divide attention depends on sustained endogenous signals from short term memory to the saliency map, stressing the interplay between working memory mechanisms and attention.

## 3.2 Introduction

Attention is an old concept in psychology correlated with enhanced processing of objects or regions in space [98]. While attention is a multi-modal phenomenon [22, 135], the majority of research has focused on selective visual attention (SVA). The limited capacity of the visual system necessitates a mechanism to select stimuli from the visual field, and Tsotsos pointed out that attention solves the complexity problem of sensory processing [125].

A distinction can be drawn between pre-attentive and attentive visual processing [91]. Pre-attentive processing refers to bottom-up (BU) feature saliency of visual stimuli whereby items that differ from their surroundings 'pop out' to the viewer. Attentive processing refers to top-down (TD) influences on perception of stimuli determined by object and locational bias such as task instructions or foreknowledge of stimulus characteristics. Determining saliency, then, is both a BU and TD requirement, and computational models of SVA include maps that integrate BU salience across object features [67], TD bias [123], and the interplay of both [134, 31].

Koch and Ullman (1985) [67] provide a neural network model of SVA in which topographic feature maps are integrated by a winner-take-all (WTA) saliency map of BU stimuli. In their model, inhibiting the selected location causes a shift to the next most salient location. Wolfe (1994) [134] builds on Neisser's pre-attentive/attentive distinction [91], integrating BU and TD saliency criteria in his Guided Search model. Treisman (1998) [123] provides a model of spatial attention to solve the Binding Problem, in which a TD saliency map determines object features selected for further processing, and suggests parietal cortex as the biological correlate of her 'master' map. Deco et al. (2002) [31] use inhibition to mediate BU and TD influences in an instantiation of Duncan and Humphreys' biased competition model [36], simulating saliency in posterior parietal cortex (PP) with a Continuous Attractor Neural Network (CANN). Spatial saliency in PP interacts with BU feature maps to converge on a winning location. See Shipp (2004) [107] and Itti and Koch (2001) [57] for a review of these and other models.

There is long-standing debate about the distribution of SVA. Many cognitive models propose a unitary focus of attention, likened to a roving spotlight over the visual field [98]. Variants of the spotlight metaphor include gradient [34, 69] and zoom

lens [39] models, suggesting that attention may be a graded phenomenon, attenuated around a central focus. A large body of evidence supports such unitary models [98, 84], but several more recent experiments have provided evidence for non-contiguous allocation of SVA [50, 8, 90].

Here, we study how split attention can be achieved by a dynamic implementation of a WTA map. Despite their WTA nature, CANNs are able to account for split attention when network dynamics facilitate long transition states between regimes [120] and when dominated by sustained inputs [111]. We simulate the experiments of Muller et al. (2003) [90] with a 1-dimensional (1D) CANN model. We build on simulations presented in Standage et al. (2005a) [113] that use a narrow weight profile, facilitating steeply sloped regions of activity that occupy a small portion of the network. Because we don't know the size of the active region of PP and its relation to coordinates in the visual field, we run similar experiments with a wide weight profile, resulting in activity that spans the majority of the network. We demonstrate that the ability of the model to account for divided attention doesn't depend on fine tuning this network parameter.

We simulate two experiments by Awh and Pashler (2000) [8] with a 2-dimensional (2D) CANN model, demonstrating how the model accounts for their finding divided attention in one experiment and unitary attention in the other. Preliminary simulations in 1D are reported in Standage et al. (2005a) [113]. Our simulations are consistent with their experimental findings, but our model offers an alternative conclusion.

## 3.3 Methods

In 1D and 2D simulations, we use a fully connected recurrent rate model with $N$ nodes, where $N = N_x N_y$. We model only PP from the model by Deco et al. (2002) [31]. WTA is implemented by local cooperation and long distance competition in the laterally connected network. The average state $u_i$ of a node with index $i$ is given by

$$\tau \frac{\mathrm{d}u_i(t)}{\mathrm{d}t} = -u_i(t) + \sum_j w_{ij} r_j(t)\ a + I_i^{\mathrm{ext}}(t), \tag{3.1}$$

where $\tau$ is a time constant, $I_i^{\text{ext}}$ is external input to the network, $a = 2\pi/N_x$ is a scale factor, and $r_i$ is a normalized square of $u_i$ given by

$$g(u_i) = \frac{{u_i}^2}{1 + \frac{1}{2}\, a \sum_j {u_j}^2}. \tag{3.2}$$

We use this normalization through divisive normalization (shunting inhibition) to force more biologically realistic smooth (Gaussian) bubbles [32].

The weight matrix $\mathbf{w}$ is determined by a shifted Gaussian function

$$w_{ij} = A_{\text{w}}\, e^{-d^2/2\sigma_w^2} \; - \; C \tag{3.3}$$

between node $i$ and node $j$ where $d$ is given by

$$d = \sqrt{d_x^2 + d_y^2} \tag{3.4}$$

$$d_x = min(|i_x - j_x|a, 2\pi - |i_x - j_x|a) \tag{3.5}$$

$$d_y = min(|i_y - j_y|a, 2\pi - |i_y - j_y|a), \tag{3.6}$$

and $i_x$ and $i_y$ are the x and y components of node $i$, $d_y = 0$ in the 1D case, $C$ is an inhibition constant describing the activity dependent inhibition of an inhibitory pool of neurons, and $A_{\text{w}}$ is a scale factor.

The external input $I_i^{\text{ext}}$ is Gaussian shaped around input location $j$, determined by

$$I_i^{\text{ext}} = e^{-d^2/2\sigma_{ext}^2} \tag{3.7}$$

where $d$ is given by Equation 3.4.

In 1D, $N_x = 100$, $N_y = 1$ ($N = 100$). In 2D, $N_x = N_y = 30$ ($N = 900$). In all simulations, $C \in \{0.1, 0.3\}$, $A_{\text{w}} = 10$, $t = 1$, $\tau = 10$, and $\sigma_w = 1.2$ (1D) and 1.3 (2D), $\sigma_{ext} \in \{0.3, 0.5\}$.

We classify our inputs along exogenous (exo) and endogenous (endo) dimensions. Exo inputs refer to neural responses to stimuli, here representing visual cues. Endo inputs refer to voluntary control of attention, here representing task instructions in behavioural studies. Exo and endo inputs thus correspond to BU and TD signals respectively.

Simulations are run with transient and sustained inputs. We equate network activity with SVA. Because transient inputs elicit WTA behaviour in CANN models, we

start by demonstrating one-bubble attractor states as models of a unitary attentional focus. Transient input stimuli are the norm in biological networks, as evidenced by high firing rates at stimulus onset followed by lower rates when stimuli are sustained in experimental settings. In the exo case, this initial burst of activity serves as input to higher cortical areas such as PP. Sustained firing after transient stimulation is a property of highly specialized neural assemblies [43], and as such is the exception among biological networks, not the norm. Where we simulate exo stimuli as sustained inputs, the stimuli being modelled are spatially static, rapidly changing symbols. We interpret these changes as providing continual 'refreshment' of neural representations due to novelty effects [24]. We interpret sustained endo inputs as STM representations of task instructions in PFC.

We use a Gaussian shaped input profile to approximate typical tuning curves of neurons, so their firing profiles are well approximated by smooth curves. In the case of transient input, the specific shape of localized input is unimportant because the network dynamic dominates after cessation of input. With sustained input, a Gaussian input profile leads to a good approximation of a Gaussian output profile, achieving the biological realism of our input profile described above.

Finally, we compare our CANN model of SVA to one with no lateral interaction, modelling the latter by simply adding together its Gaussian inputs.

### 3.4 Simulations

Muller et al. (2003) [90] provide evidence for sustained division of visual attention by recording steady state visual evoked potentials (SSVEP) while subjects viewed a horizontal array of four stimulus elements following instructions to attend to two locations. On separate blocks of trials, subjects attended to adjacent and separated positions. The SSVEP is the electrophysiological response in visual cortex to a rapidly flickering stimulus, and has been shown to increase in amplitude when attention is paid to the location of the stimulus [90]. They found that SSVEPs were lower at the location between separated targets in a detection task. Additionally, they showed that split locations were attended just as well as adjacent locations in their experiment.

We model these experiments in Simulations 1 and 2, however, we widen the network weight profile from our earlier work, increasing $\sigma_w$ from 0.4 to 1.2 and $C$ from

0.1 to 0.3. This change results in an increase in the width of a stable post-stimulus bubble from $\sigma = 4 * a$ to $\sigma = 8 * a$, demonstrating that our findings are robust in this respect.

Awh and Pashler (2000) [8] use a partial report procedure to test subjects' ability to divide spatial attention. Subjects viewed a 5x5 array of alpha-numeric characters containing 23 letters and 2 digits. Subjects fixated a central location before the presentation of two cues, either side of fixation, indicating the probable location of the digits. The character array was subsequently presented, and the subjects' task was to identify the digits. During eighty percent of trials, digits appeared at the cued (valid) locations. During the remaining twenty percent of trials, digits appeared either side of fixation in the orthogonal direction. Thus, on invalid trials, one of the digits appeared directly between the cued locations. Performance at the cued and intervening locations was compared. To the extent that SVA can be divided, subjects should perform better at the cued locations than in the middle. If division of attention were perfect, performance on the two unattended locations would be equal. Subjects' ability to divide SVA was found to depend on the presence of a subsequent noise mask, but the removal of array noise alone was sufficient to significantly reduce division of attention, regardless of subsequent masking. We model this work in Simulation 3.

### 3.4.1 Simulation 1

Adopting Müller's terminology, we refer to the locations of stimuli as 1, 2, 3 and 4, where 1 is the left-most location and 4 is the right-most location (Figure 3.1A). A 1+2 trial refers to trials in which subjects were instructed to direct their attention to locations 1 and 2, a 2+4 trial refers to instructions to attend to locations 2 and 4, and so forth for other combinations of the four locations.

In these transient-input trials, we give the network exo and endo inputs for 300 iterations of $dt/\tau$, simulating the changing symbols in the visual field and task instructions respectively. This input activity is followed by 300 iterations without either source of input. These iterations are sufficient for the network to stabilize under both dynamic regimes (both during and after input). In adjacent trials (1+2 and 3+4) and split trials (1+3 and 2+4), network activity merges into a single winning bubble between target locations, predicting a unitary focus of attention. These results are

shown in Figure 3.1.

In comparison to our earlier study, the wider activity profile predicts a more even distribution of attention once two bubbles merge into one. Specifically, in the 1+2 trial, activity at locations 1 and 2 is 67% and 90% of maximum activity respectively, compared to 21% and 37% in Standage et al. (2005a) [113]. The wider activity profile also effects split trials. The bubble drifts into the area between attended locations, shown in Figure 3.1F and G. Both WTA effects conflict with Müller's findings.

### 3.4.2 Simulation 2

Network configuration and the shape and location of inputs is identical to Simulation 1. Exo and endo inputs are sustained simultaneously for 500 iterations, sufficient for the network to stabilize.

Under sustained inputs, our model replicates Müller's findings in split trials, as network activity is greater at locations 1 and 3 than in between (Figure 3.2B). Two distinct bubbles are also seen in adjacent trials (Figure 3.2A) suggesting that Müller's subjects may have divided their attention between adjacent locations. Because Müller et al. didn't test subjects' attention between adjacent stimuli, this effect doesn't conflict with their results. Having found similar results with a narrow weight profile in Standage et al. (2005a) [111], results here show that sustained inputs dominate the network regardless of the width of its weight profile.

To achieve a single bubble in adjacent trials, we increase the overlap between representations of input stimuli from $\sigma_{ext} = 0.3$ to $\sigma_{ext} = 0.5$. The model predicts divided attention in split trials (Figure 3.2D, but no longer in adjacent trials (Figure 3.2C).

Because network output so closely resembles the shape of sustained inputs, we investigate the contribution of the network dynamic to the output profile. In the extreme case, complete neglect of the network dynamic reduces the model to a simple addition of Gaussian (AOG) input curves. To test if an AOG provides a model of CANN behaviour under sustained inputs, we measure the reduction or 'dip' in activity between bubbles at different distances between inputs, comparing it to peak activity in the bubbles. Correspondingly, we measure the height of the midpoint between two Gaussian curves as a function of the distance between them, comparing it to their

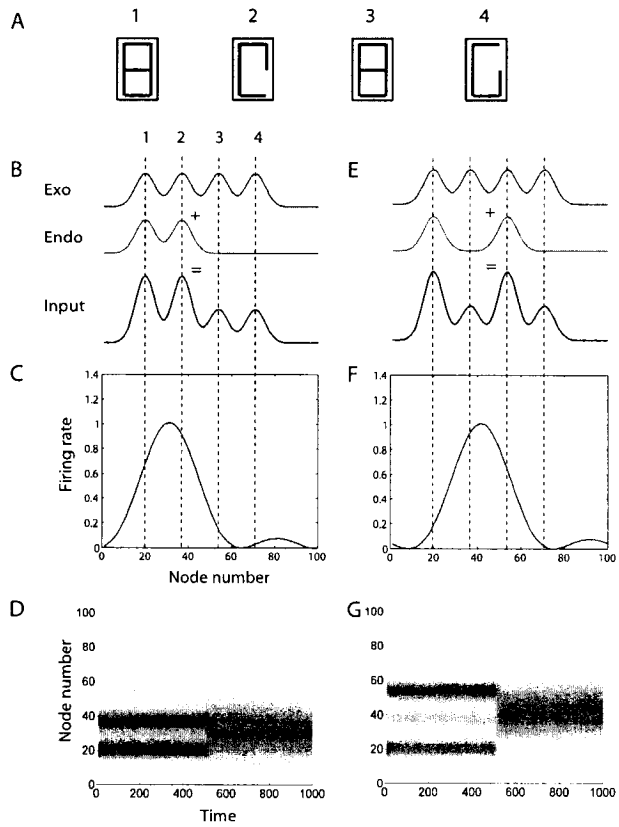


Figure 3.1: (A) Subjects in Müller's experiment attended two of four horizontal locations, here labelled 1, 2, 3 and 4. Rectangles were flashed at different rates, creating SSVEPs. Random sequences of five symbols were provided at all locations. Subjects' task was to report simultaneous occurrence of a target symbol at two attended locations. The figure depicts a 1+3 trial with target symbol '8'. (B) Nodes 20, 37, 54 and 71 correspond to locations 1, 2, 3 and 4 respectively. Exo inputs are applied to all locations. Endo inputs are applied to locations 1 and 2 only. Combined exo and endo input activity shown on bottom. Gaussian width factors $\sigma_w = 1.2, \sigma_{ext} = 0.3$, constant of inhibition $C = 0.3$. Dashed vertical lines run through target locations. (C) Stable bubble following transient input. The bubble is centred on node 31, reflecting the merge between locations 1 and 2. (D) Network activity over time. Input is stopped after 500 iterations, followed by transition to a one bubble (merged) state by approximately 550 iterations. (E) All parameters are identical to B except endo inputs are applied to locations 1 and 3. (F) Stable bubble following transient input, centred on node 42 (location 3). (G) Network states over time. A merged bubble is stable by approximately 550 ms.

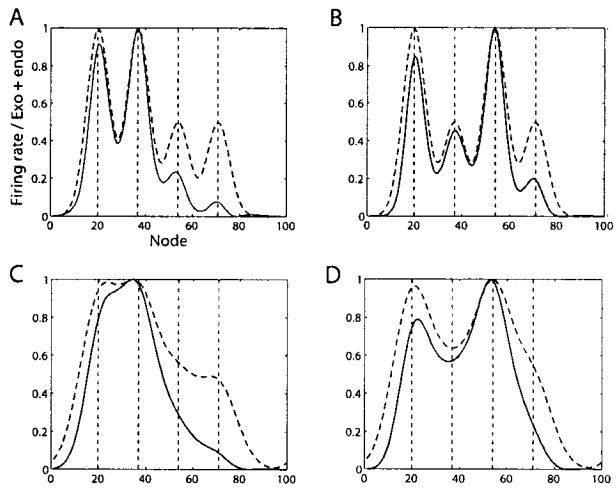


Figure 3.2: Network configuration as described in Figure 3.1B. Activity conforms to the sustained input profile, sharpened by lateral inhibition. For clarity, input (dashed) and output (solid) are normalized to 1. Dashed vertical lines show target locations. (A) 1+2 trial, $\sigma_{ext} = 0.3$ (B) 1+3 trial, $\sigma_{ext} = 0.3$. (C) 1+2 trial, $\sigma_{ext} = 0.5$. Wider input profile abolishes divided attention in adjacent trials. (D) 1+3 trial, $\sigma_{ext} = 0.5$. Under sustained wide inputs, the CANN model still predicts divided attention in split trials.

maxima. The solid line in Figure 3.3A represents the CANN model. The dotted line represents AOG. Both curves predict unitary attention when targets are spatially proximal ($\approx$ 10 nodes) and divided attention between more distant targets ($\approx$ 30 nodes). In between, the slight difference between curves reflects the effect of lateral inhibition, reminiscent of Mountcastle's two-point discrimination [88]. In contrast, the effect of local excitation is not evident in the figure, as the onset of a divided activity is not right-shifted for the CANN curve.

Figure 3.3A shows results of our analysis for only one value of $\sigma_{ext}$. Figure 3.3B shows how the AOG curve in A depends on the width of Gaussian inputs. The same effect is observed for the CANN. Our model predicts that attention cannot be divided at close distances, but without a means to map $\sigma_{ext}$ to physical parameters, we do not predict specific distances over which attention may be divided.

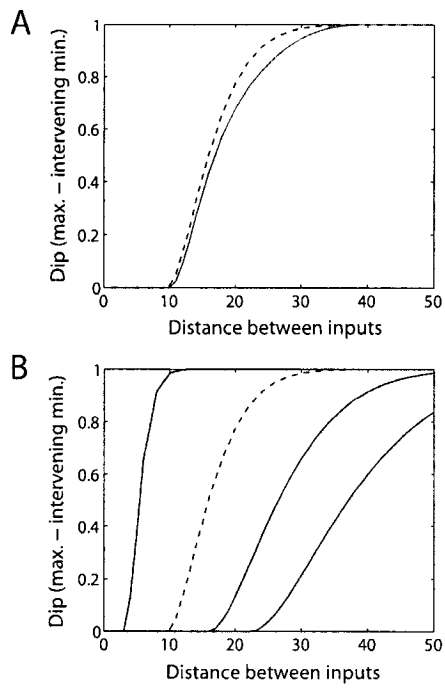


Figure 3.3: (A) Dip between CANN bubbles (max. - intervening min.) plotted against the distance between peaks (solid line). $\sigma_w = 0.8, \sigma_{ext} = 0.3, C = 0.3$. Dip between peaks of summed inputs, plotted against distance between them (dashed line). (B) AOG distance vs. dip (as in A). From left to right, $\sigma_i = 0.1, 0.3, 0.5, 0.7$. Dashed line shows curve in A.

### 3.4.3 Simulation 3

We simulate Awh and Pashler's experiments in 2D. In keeping with our 1D simulations above, we use a wide weight profile in our 2D model ($\sigma_w = 1.3, C = 0.1$). The effects of a sustained input profile are similar to the 1D case. In the 2D CANN model, the number and location of inputs are arranged to reflect Awh and Pashler's experimental conditions (described in Figure 3.4). Inputs are centred on every fourth node (horizontally and vertically) in a square bounded by nodes (7,7) and (23,23) in the 900-node 2D network. Exo inputs at these locations represent the character array in Awh and Pashler's Experiments 1, 4 and 4a. In all trials, endo inputs centred on nodes (11,19) and (19,19) represent subjects' attention to the probable location of the digits. In valid trials, digits are represented by exo inputs centred on these nodes. In invalid trials, digits are represented by exo inputs centred on nodes (15,11) and (15,19). Subjects' fixation is represented by an exo input centred on node (15,15). In these simulations, we model Awh and Pashler's experiments only as far as the presentation of the character array. The effect of noise masks and subsequent identification of target digits presumably involve STM and object recognition processing not included in our model.

To model Awh and Pashler's Experiment 1, we provide an exo input to the fixation point for 500 ms. This exo signal is then accompanied by endo inputs to target locations for 750 ms. Finally, exo inputs are centred on all character locations and endo inputs are continued for 118 ms, where $dt/\tau = 1$ ms. These iterations model the duration of input screens in Awh and Pashler's study.

Our results replicate those of Awh and Pashler's Experiment 1. The sustained endo inputs dominate the network, facilitating the activity profile shown in Figure 3.4C.

Our simulation of Awh and Pashler's Experiments 4 and 4a uses the same network configuration and width and duration of inputs as Experiment 1, except the character array is provided for only 62 ms, as in Awh and Pashler's Experiment 4.

We simulate removal of non-target characters during valid trials by removing all exo inputs except at attended locations. Similarly, we simulate removal of non-target characters during invalid trials by limiting exo inputs to the middle and far locations. Results of our invalid trials (Figure 3.5B) coincide with those of Awh and

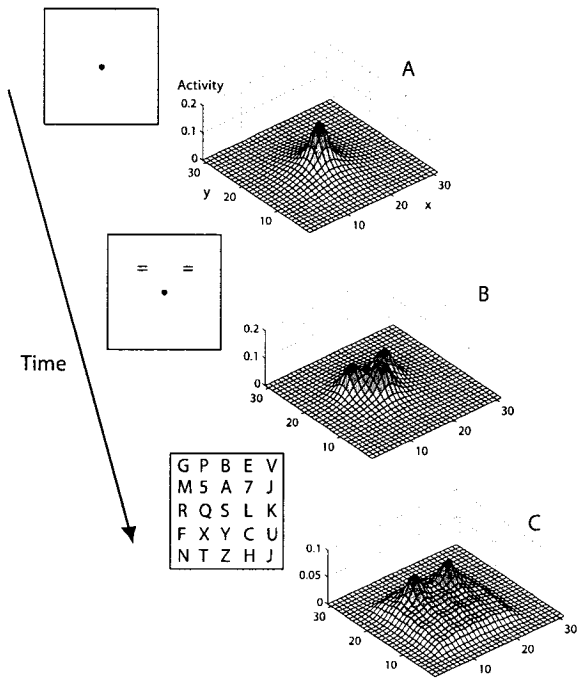


Figure 3.4: Simulation of Awh and Pashler's Experiment 1. Subjects fixate on the central dot for 500 ms, then fixate on the dot *and* attend to the 'equals' signs for 750 ms before presentation of a 5x5 character array for 118 ms. On 80% of trials, digits appear at the attended locations (valid trials). On invalid trials, digits appear at the locations shown in Figure 3.5B. Here, the full character array is presented on all trials. The 2D CANN's response to corresponding input signals is shown on the right.

Pashler. The middle location is more active than the attended locations. Results of our valid trials (Figure 3.5A) paint a different picture. The absence of letter noise reduces competition between stimuli in the character array, so that activity at attended locations is more cleanly divided than in our simulation of Experiment 1. Our interpretation of these results is that the probe stimulus may have been responsible for the reduction of divided attention in Awh and Pashler's Experiment 4 and its abolition in their Experiment 4a.

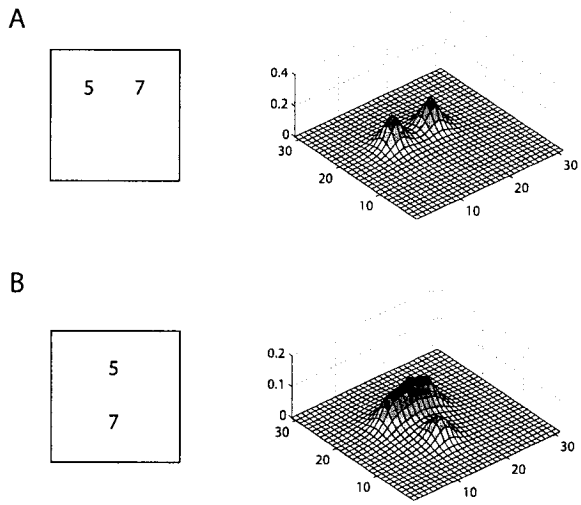


Figure 3.5: Simulation of Awh and Pashler's Experiments 4/4a. Awh and Pashler's partial character arrays are shown on the left (valid array top, invalid array bottom). The 2D CANN predicts divided attention on valid trials (top right). On invalid trials (bottom right), the model predicts unitary attention over the target and middle locations, with less attention focused on the 'far' vertical location. This output coincides with Awh and Pashler's results on invalid trials.

## 3.5 Discussion

In our simulations of Müller's experiments, we address a parametric issue raised in Standage et al. (2005a) [113]. By increasing the width of the model's weight profile, and comparing network output with similar experiments in Standage et al. (2005a) [113], we show that under transient input, the model outputs a flatter bubble, predicting that the magnitude of subjects' attention in Müller's adjacent trials was more evenly distributed across attended locations than predicted in Standage et al.

(2005a) [113]. This input/output paradigm still supports a unitary gradient model, but with a much less extreme slope.

In Simulation 2, we show that under sustained inputs, network output capitulates to input, regardless of the width of our weight profile. Under this paradigm, we no longer need the CANN to account for divided attention. A simple AOG model suffices when inputs are Gaussian shaped. In this case, our ability to model divided SVA is mediated by the width and proximity of input activity. Because inputs to PP are highly pre-processed, coming from PFC and V4 in Deco's model (in addition to less pre-processed input from V1), we believe these inputs may be characterized by considerable overlap, given the expansion of receptive fields in hierarchical processing. We conjecture that the overlap between integrated object representations may increase as a function of their number of common features. For example, neural representations of a red circle and a red square may overlap more than those of a red circle and a blue square, and attention may be more difficult to divide between them.

Simulation 2 predicts that attention should be easier to divide as foci become more distant (within a reasonable visual area). This prediction is parameter-dependent, as a stiffer gain function would still predict divided attention, but in all-or-none fashion. With a Sigmoid gain function, the model no longer resembles an AOG. These predictions could be tested by a probe stimulus between adjacent locations in Müller's experiment, and the addition of a 1+4 trial. Such experiments are important to further constrain computational models of attention. Additionally, although Müller et al. recorded target detection rates (TDR) in their experiment, the subjects' task was to detect simultaneous occurrence of the target symbol at the attended locations only. TDRs at adjacent locations were no better than at split locations (indeed, they were slightly worse), but without testing subjects' ability to detect simultaneous occurrence of target symbols at *any* two locations while subjects attended to two specific locations, TDRs provide no direct psychophysical evidence of divided attention. For example, we don't know that TDRs would be better at positions 1 and 3 during a 1+3 trial than at position 2. We believe that adding a task requiring subjects to detect simultaneous occurrence of a target symbol at any two locations during the same trial block would strengthen Müller's conclusions.

A parameter largely unexplored in these simulations is the strength of connectivity

in the network. By greatly reducing input strength in comparison to the network's connection strength, connectivity dominates sustained input and the WTA nature of the model re-emerges. It is also likely that the strengths of exo and endo inputs are not equal. If we assume a model that tends to WTA in the general case, that is, that attention is unitary 'by default', and that strong endo inputs are able to override this tendency in unusual cases, then perhaps exo inputs to the model should be weak and wide, and endo's should be strong and narrow. In this regard, our model suggests that an understanding of the modulation of signals from PFC to PP is crucial to understanding SVA.

Awh and Pashler found the ability to divide SVA was greatly reduced following removal of noise surrounding target stimuli. We believe the model can account for this result if their probe stimuli on invalid trials dominated voluntary attention. That is, in the absence of competition from letter noise in the character array, exo signals at the invalid locations dominated subjects' attention. In valid trials, exo and endo signals were directed to the same locations, so attention may have been divided when Awh and Pashler weren't testing for it, only to be unified by the testing procedure.

Simulations 2 and 3 show that the CANN model is able to account for divided SVA under sustained inputs. As such, we believe that divided SVA may be possible for as long as endo and/or exo signals are provided to PP, and that differences in behavioural findings may reflect differences in experimental conditions rather than subjects' ability to divide their attention. The nature of these conditions is largely unexplored. This conjecture echoes that of Schneider [104] that different experimental paradigms may facilitate measurements of different attention-related phenomena.

Because we interpret endo inputs as subjects' representations of task instructions in WM, our model predicts that interference with STM should abolish split attention in both Müller's and Awh and Pashler's experimental conditions. This prediction could be tested in a dual task paradigm. By equating task instructions in behavioural studies with STM representations in WM, and by modelling these representations as sustained endo inputs to the CANN, we revisit the relationship between WM and attention. Attention has often been cited as the primary constraint on WM capacity [27], but here we view WM representations as the driving force behind attention.

Our focus has been on stable attractor states in this paper. Transitions between

dynamic regimes tend to be rapid, and given the large number of parameters that effect the model, stable states provide a better foundation for our simulations. As we show in Standage et al. (2005a) [113], parameter adjustments effect transitions between regimes. A possible explanation of the findings of Müller et al. and Awh and Pashler is that divided SVA corresponds to the transition between two-bubble and one-bubble states in a WTA model. Thus, subjects may only be able to divide attention during these meta-stable states; given sufficient time for the network to settle, subjects may be unable to divide their attention.

## 3.6 Conclusions

The model of SVA by Deco et al. (2002) [31] implements a saliency map in PP with a CANN network. This instantiation of biased competition [36] integrates BU and TD influences in a biologically realistic computational architecture. Our simulations test this promising model's ability to explain behavioural and physiological evidence on the spatial distribution of SVA.

Our results demonstrate that CANNs provide a model of spatial attention in PP capable of explaining divergent experimental findings. With transient inputs, the model's WTA nature predicts a unitary attentional focus. With sustained inputs, the model accounts for divided SVA. As such, our predictions depend on the nature of exo and endo signals in attentive phenomena. Here, the use of sustained inputs replicates the findings of Muller et al. (2003) [90] and some of the findings of Awh and Pashler (2000) [8]. Where Awh and Pashler interpret unitary attention on invalid trials as demonstrating unitary attention on valid trials, we believe subjects' attention may have been divided on valid trials, only to be unified by their probe stimulus.

The interplay between WM and SVA is paramount to our model. If divided attention is facilitated by STM representations providing endo inputs to PP, then disruption of STM should abolish divided attention. We believe further research in this area would improve our understanding of the relationship between WM and attention.

**Acknowledgment**

This work was supported in part by the NSERC grant RGPIN 249885-03.

# Chapter 4

## Computational Consequences of Experimentally Derived Spike-Time and Weight Dependent Plasticity Rules

Dominic Standage, Sajiya Jalil and Thomas Trappenberg, Biological Cybernetics, 96(6):615-623, 2007

The original publication is available at www.springerlink.com

http://www.springerlink.com/content/41327604730vg63u/fulltext.pdf

## 4.1 Addendum to Standage, Jalil and Trappenberg (2007)

The present chapter marks a change in the focus of my research. Chapters 2 and 3 use neural field models to investigate mechanisms of cortical short term memory, focusing on the active maintenance of neural activity and attentional processing respectively. As described in Chapter 1, neural field models capture the properties of a small region of cortex mediating a feature space, such that activity in the network represents a feature value (or values) from that space. On a larger scale, cortex is a hierarchy of bidirectionally-connected regions, where the strengths of connections between regions are determined by neural activity. Activity-dependent change in synaptic strength, or synaptic plasticity, is the focus of the next three chapters.

Earlier, in Chapter 2, we used a Hebbian covariance role to learn the internal connections of our neural field model. The rule extracts the statistics of input activity, yielding a shift-invariant weight-profile from shift-invariant inputs. In this case, the statistics of input activity are described by the firing rates of nodes in the network, where nodes are representative of populations of neurons with similar tuning curves. At this level of description, positive correlations in the variance of pre- and post-synaptic rates yield an increase in weights, and negative correlations yield a decrease in weights. In the following three chapters, we address the effect of spike timing on synaptic strength. As above, the role of Hebbian learning is to extract the statistics of pre- and post-synaptic activity. We also consider the dependence of plasticity on the initial strength of a synapse, a common method for limiting weights under rate- and timing-based Hebbian rules.

In the present chapter, we study the asymptotic consequences of pre- and post-synaptic spike timing for learning rules fit to experimental data. These 'consequences' are theoretical, quantifying the expected strength of a synapse for a given (unchanging) statistical profile of spiking activity. Such steady state activity should not be expected in the brain, but allows an analysis of the implications of the data for a common form of learning rule. While the paper shows asymptotic weights as a function of spike rate, our rules do not extract pre- and post-synaptic rate correlations. Here, pre- and post-synaptic spike rates are perfectly correlated in all cases.

The principle finding of this paper is that asymptotic weights decrease with increasing rate with correlated pre- and post-synaptic spike timing. We thus show a

novel form of synaptic scaling, offering a potential mechanism for regulation of post-synaptic activity during and after learning. This effect would appear to conflict with data from rate-based experiments, where high-rate pre-synaptic stimulation yields potentiation. These experiments have not measured post-synaptic spike rates, however. The scaling shown here also conflicts with rate-based learning rules such as the one used in Chapter 2. With the benefit of hindsight, the effect should not be surprising. According to spike time dependent plasticity (STDP) data, potentiation results when a post-synaptic spike repeatedly follows a pre-synaptic spike (pre-before-post). Conversely, depression results when a pre-synaptic spike repeatedly precedes a post-synaptic spike (post-before-pre). Because the magnitude of synaptic change is reduced with increasing spike latency, in the case of repeated, time-locked pre-before-post spikes, by increasing our spike rate, we simply shorten the post-before-pre time window, resulting in more depression. Whether this kind of scaling occurs *in vivo* requires further investigation.

## Abstract

We present two weight- and spike-time dependent synaptic plasticity rules consistent with the physiological data of Bi and Poo (1998). One rule assumes synaptic saturation, while the other is scale free. We extend previous analyses of the asymptotic consequences of weight-dependent STDP to the case of strongly correlated pre- and post-synaptic spiking, more closely resembling associative learning. We further provide a general formula for the contribution of any number of spikes to synaptic drift. Asymptotic weights are shown to principally depend on the correlation and rate of pre- and post-synaptic activity, decreasing with increasing rate under correlated activity, and increasing with rate under uncorrelated activity. Spike train statistics reveal a quantitative effect only in the pre-asymptotic regime, and we provide a new interpretation of the relation between BCM and STDP data.

## 4.2 Introduction

Change in synaptic efficacy is believed to underlie learning and memory and has long been established in the forms of long term potentiation (LTP) [15] and long term depression (LTD) [79]. Experimental and theoretical work on plasticity has addressed the dependence of plasticity on pre-synaptic firing rates [14, 35, 64], the timing [72, 81, 99] and interaction [109, 11, 41, 58, 130, 42, 106] of pre- and post-synaptic spikes, and initial synaptic strength or *weight* [99, 30, 87, 130]. Theoretical studies have further investigated the relationship between rate based and spike-time dependent plasticity (STDP) frameworks [58, 17] and the conditions under which STDP rules transduce correlations among pre-synaptic spike trains into correlations between pre- and post-synaptic activity [61, 66, 68, 49] as required by Hebb's postulate [54].

Here, we focus on weight-dependent STDP rules. In Section 4.3, we derive parameters for two weight-dependent STDP rules from experimental data. We calculate asymptotic weights resulting from these rules in Section 4.4. Where earlier studies of the asymptotic consequences of STDP rules consider uncorrelated or weakly correlated pre- and post-synaptic spike trains [61, 66, 110, 128, 101, 49, 17], we extend these analyses to the case of strongly correlated spikes, where pre-synaptic activity 'repeatedly and persistently takes part in firing' the post-synaptic cell, as proposed by Hebb. Our analysis shows that the means of equilibrium weight distributions are principally determined by the correlation and rate of pre- and post-synaptic spiking, where weights decrease with increasing rate in the correlated case. In Section 4.5, we show that our qualitative results do not depend on our choice of spike train statistics or correlation model. Furthermore, we derive a general formula for the contribution of any number of individual spikes to synaptic drift, proving that our results do not depend on a specific implementation of spike interactions, in contrast to the interpretation of Izhikevich and Desai (2003). We end Section 4.5 by showing a novel instance of rate-based BCM curves [14] under STDP. These curves emerge when temporal constraints prevent weights from reaching asymptotic values at lower spike rates.

## 4.3 Weight Dependence of STPD

Weight-dependent plasticity has been shown by several groups [99, 30, 87, 130], but only Bi and Poo (1998) have done so under the STDP pairing protocol. Not only did they use the same protocol in their weight- and spike-time-dependent experiments, but they controlled spike timing in their weight-dependent experiment, allowing us to relate these two data sets. We therefor derive our weight-dependent plasticity rules from their data.

For simplicity, our analysis of weight- and spike-time-dependent plasticity assumes these two factors are independent. A learning rule of this form may be written as

$$\Delta w_{\{p,d\}} = k f_{\{p,d\}}(w) e^{-c_{\{p,d\}}\Delta t}, \tag{4.1}$$

where $\Delta w_{\{p,d\}}$ is the change in weight for potentiation (index $p$) or depression (index $d$), $\Delta t = t_{post} - t_{pre}$ is the difference between the times of post-synaptic ($t_{post}$) and pre-synaptic ($t_{pre}$) firing, and $c$ parameterizes the timescale of the plasticity window. The experimental data are commonly shown in relative terms (%) whereas our formulations express absolute changes in synaptic strength.

We consider two forms of the weight dependent factor $f$. Bi and Poo (1998) hypothesized a log-linear rule by drawing a line through these data in the semi-logarithmic plot. This hypothesis yields

$$f_{\{p,d\}}(w) = (a_{\{p,d\}} - b_{\{p,d\}} \log w) w \; \theta(w), \tag{4.2}$$

where $\theta(w)$ is the Heaviside function, keeping weights positive, and the parameters $a$ and $b$ differ for potentiation and depression. A maximum weight implicit in this rule agrees with evidence for saturable synapses [96, 95]. We contrast the above rule with a power rule of the form

$$f_{\{p,d\}}(w) = a_{\{p,d\}} w^{b_{\{p,d\}}}. \tag{4.3}$$

A power law more closely approximates the data plotted on a log scale, and, unlike the log rule, imposes no maximum weight. While synapses must surely have intrinsic limits, we include this rule for comparison with the limited case. We hereafter refer to Equation 5.1 with weight-dependence determined by Equations 5.2 and 4.3 as the Log and Power rules respectively.

The constants in Equations 5.2 and 4.3 were determined by first fitting the weight-dependent STDP data shown in Figure 4.1B. The fitted curves for the Log rule are shown as solid lines alongside the potentiation data (circles) and depression data (stars) of Bi and Poo (1998). The equivalent curves for the Power rule are shown as dashed lines. The fits to the depression data are nearly indistinguishable. The fits to the potentiation data agree similarly with the data, but the intrinsic limit in the Log rule generates a marked difference for large weight values.

### 4.3.1 Detailed Fitting Procedure

Bi and Poo (1998) controlled spike timing in their weight-dependent experiment by limiting their pre-before-post (LTP) time interval to $5 < \Delta t < 15$ms and limiting their post-before-pre (LTD) interval to $3 < \Delta t < 30$ms. For the weight-dependent fit, we replaced these intervals with their midpoint values ($\Delta t_p = 10, \Delta t_d = 17.5$). We then fit their spike-time dependent data (Figure 4.1B) to determine the remaining parameter $c$ in Equation 5.1, capturing the time course of spike-time dependence for each of potentiation and depression. Because of the wide range of initial synaptic strengths in Bi and Poo's spike-time dependent experiment ($30 < w < 500$pA) we assumed that for a given $\Delta t$, the largest relative changes in weight represent synapses with the smallest initial values. Thus, consistent with their initial weights, we assign an initial weight $w = 30$pA to the data showing the largest STDP values, and only include these data in our time-dependent fit determining $c$. The resulting parameter values are $a_{\{p,d\}} = \{208, -54\}$, $b_{\{p,d\}} = \{26.4, 3.5\}$, $c_{\{p,d\}} = \{0.054, 0.042\}$ for the Log rule and $a_{\{p,d\}} = \{431, -59\}$, $b_{\{p,d\}} = \{0.4, 0.1\}$, $c_{\{p,d\}} = \{0.039, 0.043\}$ for the Power rule. We further assume that each of the 60 pre-before-post pairings in the experiment contributed equally to the overall synaptic change. This assumption is common in computational studies [128]. A deviation from this linear assumption results in an altered learning rate that does not effect the means of equilibrium weight distributions. As the above parameters are deduced by fitting the percentage data, we include a factor of 100 in the learning rate to yield a fractional scale. The learning rate used in our analyses and simulations is therefore $k = 1/6000$ unless otherwise stated.

Fitted curves for the Power rule are shown as solid lines in Figure 4.1A for our

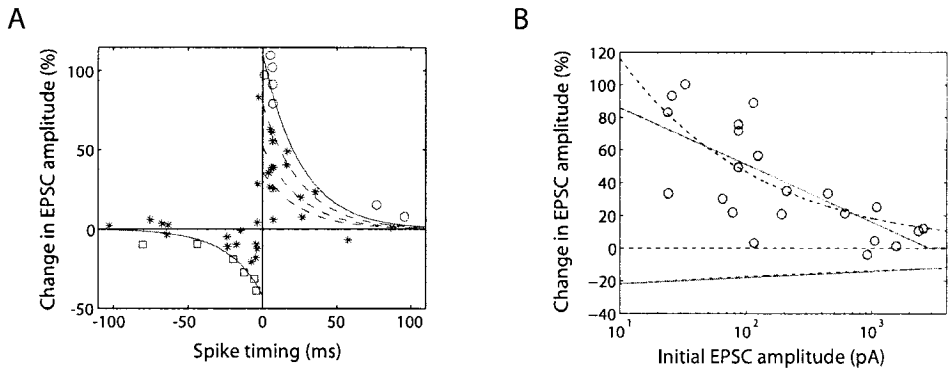


Figure 4.1: (A) Fit of the Power rule to Bi and Poo's (1998) spike-time dependent data for estimates of different initial weights $w$. Circles and squares represent $w = 30$ for potentiation and depression respectively. Solid curves show fits to these data. Dashed curves for the potentiation data show fits for $w = \{70, 200, 500\}$pA top to bottom. The Log rule leads to similar fits. (B) Log and power fits to Bi and Poo's weight-dependent STDP data. A log fit imposes a maximum synaptic weight where the upper solid line meets the x-axis. A power fit (dashed curve) imposes no such maximum. These two fits are nearly indistinguishable for the LTD data.

estimates of the data representing initial weights of 30pA (open symbols). The large scatter in the figure is commonly interpreted as noise, but we interpret it according to the weight-dependence shown in Figure 4.1B. For potentiation, we include dashed lines representing initial weights set to 70, 200 and 500 pA respectively (top to bottom). The fit of the Log rule leads to qualitatively similar plots.

## 4.4 Equilibrium Weights for Uncorrelated and Time-Locked Pre- and Post-Synaptic Poisson Spike Trains

While the above plasticity rules are deterministic, they yield a stochastic drift of weights with stochastic spike trains. If this drift is driven by a novel correlation between pre- and post-synaptic spiking, weights will undergo considerable change from an initial random state. When weights have been driven for long periods by pre- and post-synaptic spiking with a given statistical profile, they begin to fluctuate around a mean value where the average potentiation equals the (negative) average depression. In this section, we calculate the means of these equilibrium weight distributions for the Log and Power rules. The drift of weights can be calculated with Fokker-Planck

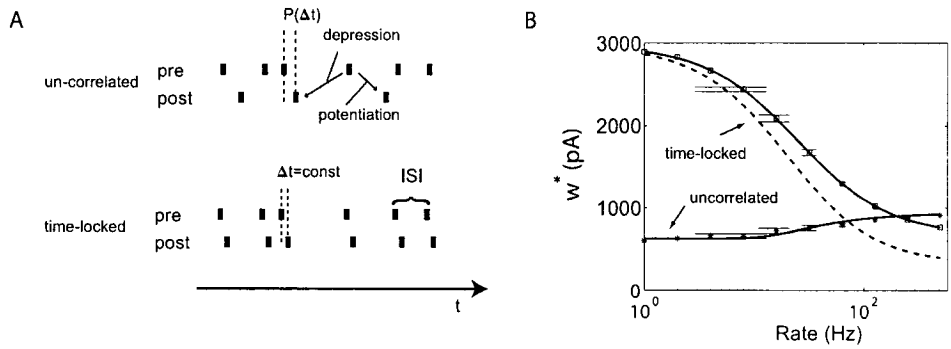


Figure 4.2: (A) Illustration of uncorrelated and time-locked spike trains. (B) Equilibrium weights as a function of spike rate under the Log rule for analytic (curves) and numeric (symbols with error bars) calculations. For time-locked spiking, the solid and dashed curves correspond to $\Delta t = 4$ and $\Delta t = 10$ms respectively. In numeric simulations, weights were averaged over 5000 trials following an equilibrating 5000 trials.

mean field theory [61, 66, 128, 101, 17], but we are concerned with the means of equilibrium distributions and adopt the simplified methodology of Izhikevich and Desai (2003). We refer to these asymptotic values as equilibrium weights $w^*$.

First, we present our basic analysis for the cases of uncorrelated and time-locked pre- and post-synaptic spike trains (see Figure 4.2A) where these spike trains are Poisson distributed [9] (exponential inter-spike intervals) and where only *nearest neighbour* spikes contribute to plasticity. Under nearest neighbour STDP, each pre-synaptic spike triggers LTP with the next post-synaptic spike and triggers LTD with the previous post-synaptic spike, as described by Izhikevich and Desai (2003). In Section 4.5, we show that our qualitative results do not change for partially correlated spike trains or for other spike train distributions or spike interactions.

### 4.4.1 Equilibrium Weights for Uncorrelated Poisson Spiking

The average synaptic increase (LTP) for each pre-synaptic spike is given by

$$\langle \Delta w_p \rangle = \int_0^\infty p_p(\Delta t) \Delta w_p d\Delta t, \tag{4.4}$$

where $p_p(\Delta t)$ is the probability density of a post-synaptic spike following a pre-synaptic spike with time lag $\Delta t$. Similarly, the average depression for each pre-synaptic event is given by

$$\langle \Delta w_d \rangle = \int_0^\infty p_d(\Delta t) \Delta w_d d\Delta t, \tag{4.5}$$

where $p_d(\Delta t)$ is the probability density of a post-synaptic spike preceding a pre-synaptic spike with time lag $\Delta t$. Thus, independent pre- and post-synaptic Poisson spike trains have an average depression and potentiation of

$$\begin{aligned} \langle \Delta w_{\{p,d\}} \rangle &= f_{\{p,d\}}(w) \int_0^\infty r e^{-(c_{\{p,d\}}+r)\Delta t} d\Delta t & (4.6) \\ &= f_{\{p,d\}}(w) \frac{r}{c_{\{p,d\}}+r}, & (4.7) \end{aligned}$$

where $r$ is the rate of the spike trains. We only consider the case where $r_{\text{pre}} = r_{\text{post}} = r$.

An equilibrium weight $w^*$ is reached when the average potentiation equals the (negative) average depression. For the Log rule this value is given by

$$w^* = \exp \frac{a_p(c_d + r) + a_d(c_p + r)}{b_p(c_d + r) + b_d(c_p + r)} \tag{4.8}$$

and for the Power rule by

$$w^* = \left(-\frac{a_p}{a_d} \frac{c_d + r}{c_p + r}\right)^{\frac{1}{b_d - b_p}}. \tag{4.9}$$

Equilibrium weights for independent pre- and post-synaptic Poisson spike trains are shown as a function of rate in Figure 4.2B for the Log rule. Symbols represent the results of simulations where weights were averaged over 5000 spike pairings following 5000 equilibrating pairings.

### 4.4.2 Equilibrium Weights for Time-Locked Poisson Spiking

The above analysis of uncorrelated pre- and post-synaptic spikes is relevant if events represented by pre-synaptic firing are *not* associated with a post-synaptic response. In contrast, associative learning is achieved if a neuron becomes responsive to (correlated with) a pre-synaptic spike pattern. Studies have shown that STDP rules capture correlations among input spikes driving a model neuron [66, 110], as synapses mediating correlated pre-synaptic activity learn to provide the strongest, fastest [110] and most precisely timed [66] inputs to the post-synaptic cell. Alternatively, plasticity may 'piggyback' other sources of activity driving pre- and post-synaptic neurons.

In this section, we assume associations have been formed by one or both of these mechanisms, labelled self-organisation and associativity respectively under the terminology of Hasselmo [53]. We investigate the ongoing effect of these associations on the asymptotic strength of weights for rules grounded in Bi and Poo's data [99].

For analytic simplicity and to illustrate the limiting case for informative spike timing, we consider the case where a post-synaptic spike is triggered with a short but fixed delay $\Delta t$ following a pre-synaptic spike. We do not suggest that a single pre-synaptic spike should drive a post-synaptic neuron in this way. Rather, our pre-synaptic spikes represent the activity of one of many inputs from pre-synaptic neurons participating in an established association. We later relax this condition by varying the probability of a time-locked post-synaptic spike in Section 4.5.1.

Time-locked pre- and post-synaptic spikes result in an ongoing potentiation of weights given by

$$\langle \Delta w_p \rangle = f_p(w) e^{-c_p \Delta t} \tag{4.10}$$

for each pre-synaptic event. However, every pre-synaptic event can also trigger depression in conjunction with a previous (uncorrelated) post-synaptic spike (Equation 4.6). An equilibrium weight $w^*$ for correlated pre- and post-synaptic spikes under the Log rule is therefore given by

$$w^* = \exp \frac{a_p e^{-c_p \Delta t}(c_d + r) + a_d r}{b_p e^{-c_p \Delta t}(c_d + r) + b_d r} \tag{4.11}$$

and for the Power rule by

$$w^* = \left(-\frac{a_p}{a_d}\frac{1}{r} e^{-c_p \Delta t}\ (r + c_d)\right)^{\frac{1}{b_d - b_p}}. \tag{4.12}$$

The rate-dependence of equilibrium weights for time-locked pre- and post-synaptic Poisson activity under the Log rule is shown in Figure 4.2B, where the solid and dashed lines represent analytic solutions for $\Delta t = 4$ms and $\Delta t = 10$ms respectively. Symbols represent corresponding numeric simulations. The figure shows that equilibrium weights decrease with increasing spike rates in the strongly correlated Poisson case, but quantitatively, these values exceed biologically realistic values.

## 4.5 Equilibrium Weights for Alternative Conditions

In the previous section, we determined equilibrium weights for the Log and Power rules under the specific conditions of Poisson-distributed pre- and post-synaptic spike trains, uncorrelated and time-locked pre- and post-synaptic spikes, and nearest neighbour spike interactions. Here, we show in Section 4.5.1 that partially correlated pre- and post-synaptic spike trains interpolate between the extreme cases of uncorrelated and time-locked pre- and post-synaptic spikes. We show in Section 4.5.2 that our results do not qualitatively depend on Poisson-distributed spike trains. In Section 4.5.3, we show that our results do not depend on nearest neighbour spike interactions, in contrast to the claims of Izhikevich and Desai (2003). Finally, in Section 4.5.4, we demonstrate learning under a finite number of spike pairings, and discuss how BCM-like curves [14] are generated by the Log and Power rules.

### 4.5.1 Equilibrium Weights for Partially Correlated Pre- and Post-Synaptic Poisson Spiking

The cases of uncorrelated and time-locked pre- and post-synaptic spike trains define the extreme cases of possible spike train relations. Here we extend this analysis to some more realistic cases with partially correlated pre- and post-synaptic activity. We discuss two examples of correlation models, providing a biological interpretation of each. In both models, we again consider Poisson pre-synaptic spiking.

**Correlation model 1:** Post-synaptic spikes are generated with a probability given by constant $p$ with a fixed time delay $\Delta t = 4$ms. In the $(1-p)$ cases where there is no time-locked post-synaptic spike the time delay is distributed exponentially. This methodology approximates the biological case where pre-synaptic input generates a post-synaptic spike with a finite probability, but otherwise the post-synaptic neuron emits Poisson background activity. Figure 4.3A shows mean weights over 10,000 pairings following 10,000 equilibrating pairings under the Log rule for $p = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. Low values of $p$ lead to rate-dependent curves qualitatively similar to the perfectly time-locked case ($p = 1$) where lower means reflect the lower likelihood of correlated pre- and post-synaptic spikes.

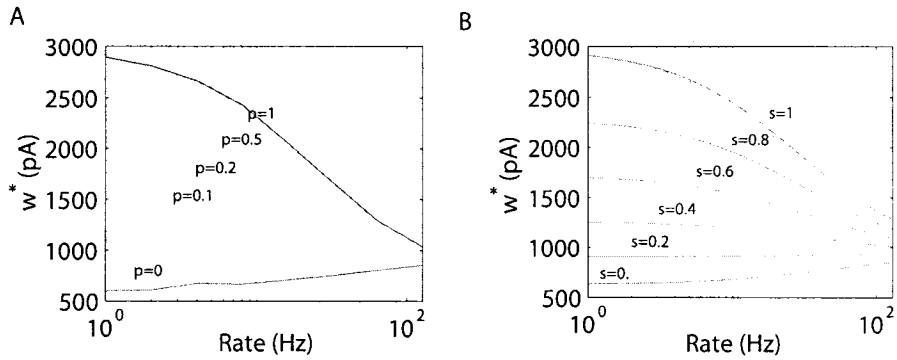


Figure 4.3: (A) Correlation model with varying probability $p$ of a time locked post-synaptic spike. Weights were averaged over 5000 trials following 5000 equilibrating trials. (B) Correlation model with varying exponential distribution for different correlation parameters $s$ where $\Delta_t = r/(1-s)$. The dashed line shows the time looked case discussed in section 4.4.2 with $\Delta t = 1$ms.

**Correlation model 2:** We consider the case where correlations between pre- and post-synaptic spikes are expressed by an altered probability of a post-synaptic spike within a time $\Delta t$ of a pre-synaptic event. Specifically, we consider an exponential distribution of delay times $\Delta t$, where the decay parameter is modulated for different correlations between pre- and postsynaptic spikes. The alteration of the probability density of $\Delta t$ is accomplished by parameter $s$ in Equation 4.13. The mean delay $\Delta t$ is set to be $(1-s)\lambda$, where $\lambda = 1/r$ is the inverse of the spike rate. For $s = 0$, this parameter yields the uncorrelated case discussed above. When $s$ tends to 1, we expect a post-synaptic spike with a very short average time delay. The average potentiation (and similarly for depression) and the corresponding equilibrium weights are given by

$$\Delta w_p = (a_p - b_p \log w) w \int_0^\infty \frac{r}{1-s} e^{-(\frac{r}{1-s}+c_p)\Delta t} d\Delta t \tag{4.13}$$

$$\Leftrightarrow w^* = \exp \frac{a_p(1/((1-s)c_p+1) + a_d(1/(c_d+r))}{b_p(1/((1-s)c_p+1) + b_d(1/(c_d+r))}. \tag{4.14}$$

Analytic equilibrium weights for this correlation model are plotted in Figure 4.3B for different values of $s$, where $s = 0$ is equivalent to the uncorrelated case in Figure 4.2B and $s = 1$ is close to the time looked case with $\Delta t = 1$ms, shown with a dashed line in the figure. Both correlation models show that partial pre- and post-synaptic correlations interpolate the extreme cases discussed in Section 4.4.

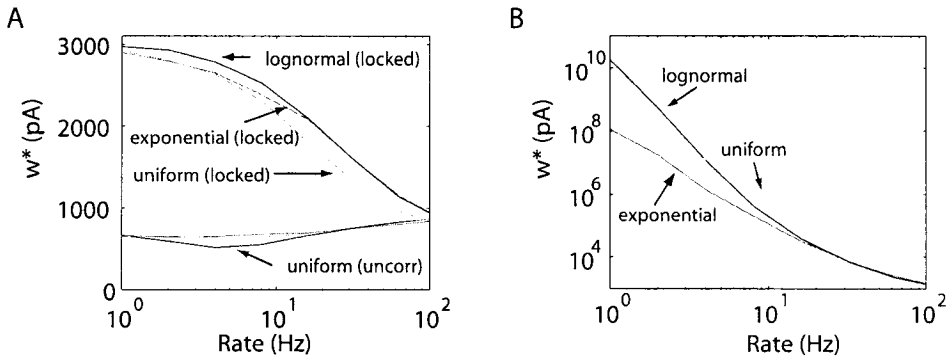


Figure 4.4: Equilibrium weights as a function of rate for pre- and post-synaptic spike trains with different ISI distributions. (A) Time-locked (top) and uncorrelated (bottom) spiking with the Log rule. (B) Time-locked spiking with the Power rule.

### 4.5.2 Equilibrium Weights for Spike Trains with Non-Poisson Statistics

We have thus far shown results for Poisson spike trains in our analysis, but similar derivations can be made for other spike distributions. While the specific values of equilibrium weights for a given rate depend on the distribution model, their qualitative dependence on rates and the correlation between pre- and post-synaptic spiking does not change. Examples from simulations for several ISI distributions are shown in Figure 4.4. Figure 4.4A shows the cases of uncorrelated and time-locked pre- and post-synaptic spikes under the Log rule. Figure 4.4B shows the time-locked case under the Power rule. The dependence of equilibrium weights on rate is similar for these distributions, but could possibly differ for very different distributions such as those corresponding to bursting behaviour.

### 4.5.3 Equilibrium Weights Beyond Nearest Neighbour Interactions

In the nearest neighbour case, we only consider the first post-synaptic spike following a pre-synaptic spike for potentiation and the first pre-synaptic spike following a post-synaptic spike for depression. While the first spike makes the greatest contribution to plasticity due to the decaying exponential term in STDP rules, it is possible that the sum of subsequent spikes has a pronounced effect on synaptic strength, as argued by Izhikevich and Desai (2003).

For Poisson spike trains, we use the following method to analytically calculate

the average contribution to potentiation of the $n$-th spike following a specific pre-synaptic spike (the method is the same for depression, where pre-synaptic spikes follow a specific post-synaptic spike). The first post-synaptic spike is expected to occur on average at $\Delta t_1 = 1/r$ for uncorrelated pre- and post-synaptic activity, and at $\Delta t_1 = \text{const}$ in the time locked case, and is weighted by $e^{-c_p \Delta t_1}$. The second spike, which we expect on average at $\Delta t_2 = \Delta t_1 + 1/r$, contributes less to potentiation because $e^{-c_p \Delta t_1} < e^{-c_p \Delta t_2}$. To calculate the average potentiation, we determine the density function of the $n$-th spike by convolving the density functions of all random variables in the sum. This convolution can be done analytically for Poisson spike trains by independently summing exponentially distributed random variables with equal mean $\lambda$. The resulting random variable is gamma distributed with mean $\lambda = 1/r$ and parameter $n$,

$$p(\Delta t_n) = \frac{(\Delta t/\lambda)^{n-1} e^{-\Delta t/\lambda}}{\lambda \Gamma(n)} \tag{4.15}$$

where $n$ is the number of spikes considered. Thus, the average potentiation for uncorrelated pre- and post-synaptic spike trains is given by

$$\langle \Delta w_p \rangle = k f_{\{p,d\}}(w) \sum_{n=1}^{N} \int_0^\infty r^n (\Delta t)^{n-1} \frac{e^{-(r+c_p)\Delta t}}{\Gamma(n)} d\Delta t \tag{4.16}$$

and similarly for depression. Using the definition of the Gamma function,

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \tag{4.17}$$

we can evaluate this integral as

$$\langle \Delta w_p \rangle = k f_{\{p,d\}}(w) \sum_{n=1}^{N} \left( \frac{r}{r + c_p} \right)^n, \tag{4.18}$$

generalizing equation (4.7) to multiple spike contributions. Thus, equilibrium weights for the uncorrelated case are given for the Log rule by

$$w^* = \exp \frac{a_p \sum_{n=1}^{N} (c_d + r)^n + a_d \sum_{n=1}^{N} (c_p + r)^n}{b_p \sum_{n=1}^{N} (c_d + r)^n + b_d \sum_{n=1}^{N} (c_p + r)^n}, \tag{4.19}$$

which reduces to equation (4.8) for $n = 1$. Furthermore, since

$$\sum_{n=1}^{\infty} x^n = \frac{1}{x - 1} \quad \text{for } x < 1, \tag{4.20}$$

where $x = r/(r+c_{\{p,d\}})$, the all-to-all case for uncorrelated spike trains is independent of $r$,

$$w^*(n=\infty) = \exp\frac{a_p c_d + a_d c_p}{b_p c_d + b_d c_p}, \tag{4.21}$$

in agreement with [61]. This value is also equivalent to the $r = 0$ limit for $n = 1$ (eq. 4.8).

A similar evaluation for the time-locked case yields

$$w^* = \exp\frac{a_p \Delta P + a_d \Delta D}{b_p \Delta P + b_d \Delta D} \tag{4.22}$$

with

$$\Delta P = e^{-c_p \Delta t}(1 + \sum_{n=2}^{N} \frac{r^{n-1}(n-2)!}{(c_p + r)^{n-1}\Gamma(n-1)}) \tag{4.23}$$

and

$$\Delta D = \sum_{n=1}^{N} \frac{r^n (n-1)!}{(c_p + r)^n \Gamma(n)}. \tag{4.24}$$

Equilibrium values for interaction models with different $n$ (*nearest-n* interactions) are shown in Figure 4.5 for the time-locked and independent Poisson cases. These curves correspond to values of $n = \{1, 2, 4, 10, 50, 100\}$ (from top to bottom) although results between $n = 50$ and $n = 100$ become indistinguishable. As expected, an increasing number of neighbouring spikes will only influence equilibrium values at high rates due to the exponential decay of the STDP time window. Furthermore, increasing the number of neighbours $n$ quickly converges to an asymptotic value. We reached machine precision around $n = 150$.

### 4.5.4 Convergence to Equilibrium: BCM-Like Curves in the Pre-Asymptotic Regime

The analysis above concerns the asymptotic regime, where equilibrium weights correspond to the infinite limit of pairings, and simulations approximate this limit with many thousands of pairings. Experiments in vitro are temporally constrained in that a plasticity-inducing stimulus is applied for a short time, after which synaptic responses are measured and compared to pre-stimulus measurements. While the duration of plasticity-inducing stimuli typically varies across protocols and the nature of experiments, the number of repetitions is typically on the order of 10 to 100 for STDP

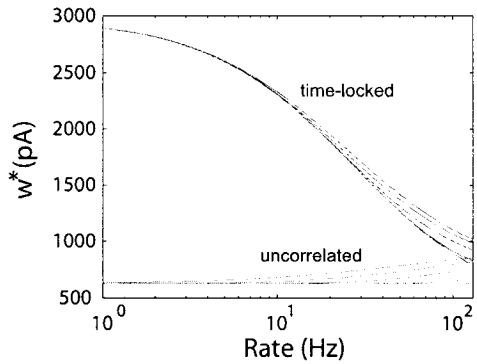


Figure 4.5: Equilibrium weights under *nearest-n* interactions for time-locked and uncorrelated Poisson spike trains. In each case, curves correspond to values of $n = \{1, 2, 4, 10, 50, 100\}$ (top to bottom). Curves between $n = 50$ and $n = 100$ become indistinguishable and approximate the infinite (asymptotic) case.

[99, 41] and pairing protocols [96, 94] and 100 to 1000 repetitions for rate-based protocols [35, 64]. We now consider the effects of the Log and Power rules under similar conditions.

Figure 4.6A shows mean percentage weight changes following $n = 100$ spike pairings where weights were initialised in the middle range of possible values (700pA) and means were calculated over 100 trials. In this simulation, we use Poisson spike trains under the Log rule with nearest neighbour spike interactions. Results for uncorrelated pre- and post-synaptic spikes correspond to the case studied by Izhikevich and Desai (2003). While we can generate BCM-like curves in this limited case of finite pairings and uncorrelated spiking, the effect is small and not found in the time-locked case. Results for time-locked spikes resemble those in the asymptotic regime (eg. Figure 4.2B) though here, percentage weight change is reduced by several orders of magnitude.

We can, however, show BCM-like curves with different spike train statistics. Figure 4.6B shows that BCM-like curves resembling those measured by Dudek and Bear (1992) and Kirkwood and Rioult (1996) are produced with time-locked, periodic pre- and post-synaptic spiking ($\Delta t = \frac{1000}{2r}$ms). Results are shown for the Log rule. The solid line represents weight changes after 1000 pairings for a synapse initialised to 700pA. Performing the same experiment with 10,000 spike pairings results in the

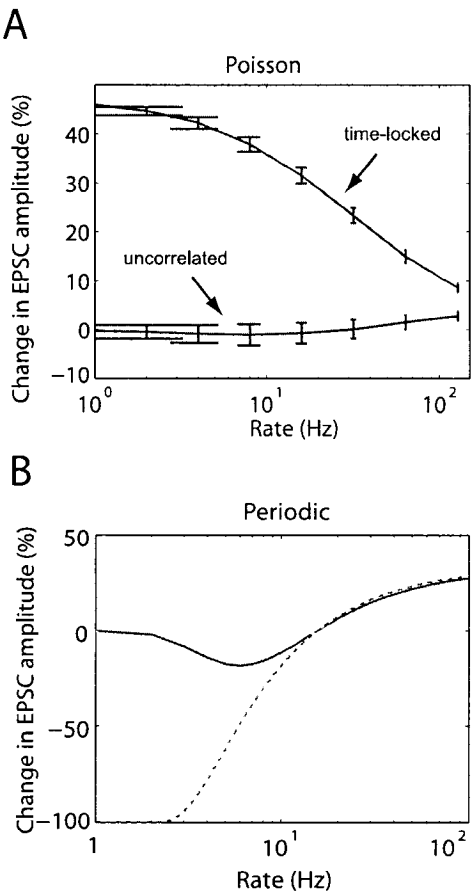


Figure 4.6: Mean percentage weight change under the Log rule for limited numbers of spike pairings. Weights were initialised to mid-range values (700pA) and means were calculated over 100 trials. (A) Results for time-locked and uncorrelated Poisson spike trains for 100 pairings at rate $r$ with nearest neighbour spike interactions. (B) Results for periodic, time-locked pre- and post-synaptic spike trains where pre- and post-synaptic spikes are 180 degrees out of phase. The solid line shows changes after 1000 pairings for a synapse initialised to 700pA. The same experiment with 10,000 spike pairings results in the dashed line. The dotted curve shows analytic equilibrium, corresponding to an infinite number of pairings.

dashed line. The dotted line corresponds to weight changes for an infinite number of pairings (equilibrium weights). As shown in the figure, the depression valley becomes deeper and shifts to the left with an increased number of pairings, while changes in the potentiation portion of the curve are much smaller. Our investigation shows that BCM-like curves may be produced by STDP rules in the pre-asymptotic regime, but these curves depend on the precise form of pre- and post-synaptic firing. Further experiments are required to investigate this and other possible explanations for BCM-like curves, as experiments investigating BCM [35, 64] have not controlled post-synaptic firing, crucial to the BCM hypothesis.

## 4.6 Discussion and Conclusions

Weight-dependent STDP rules are commonly used in modelling studies. We have presented two such rules with parameters fit to physiological data [99] and studied their consequences in both the asymptotic and pre-asymptotic regimes. Our analysis includes the case where synapses are driven by noisy spike trains with little or no correlation between pre- and post-synaptic spikes, as done in previous studies, and also the case of highly correlated pre- and post-synaptic spike trains more closely resembling the case of associative learning. We found that 'runaway' synapses [4] are still a problem for these rules, at least under parameters suggested by weight-dependent STDP data [99].

In the pre-asymptotic regime, we show that BCM-like curves [14] can be generated by weight-dependent STDP rules when low-rate activity prevents weights from reaching equilibrium values in finite time. While this effect is parameter dependent, it provides a novel instance of these curves and highlights the need for rate-based plasticity experiments that control post-synaptic spiking.

In the asymptotic regime, we find that for all spike train statistics considered, equilibrium weights for correlated spike trains decrease with increasing spike rate, a novel form of synaptic scaling. We prove this relationship for an arbitrary number of contributing spikes in the Poisson case, providing a general formula for the drift of potentiation and depression in the steady state. We further demonstrate this relationship for partially correlated Poisson spiking, showing that partial correlations

interpolate between the extreme cases of time-locked and independent pre- and post-synaptic activity.

Equilibrium weights for uncorrelated spike trains increase with rate for a finite number of contributing spikes. This increase approaches 0 in the infinite limit of spike contributions, showing rate independence in this specific case, consistent with the analysis of Kempter and Gerstner (1999). Correlated and independent equilibrium weights converge at around 100Hz. This effect suggests that low to intermediate rates provide a better regime for associative learning than high rates, or, stated differently, high rates may prevent weights from distinguishing between correlated and uncorrelated activity.

Under a rule imposing no maximum weight (the Power rule) synapses do not reach infinite values because depression balances potentiation, but the resulting equilibrium values under both rules (with and without maxima) are too large to be useful in a biologically realistic regime. This problem has an additional, unwanted consequence. Weight-dependent STDP rules implicitly assume that a synapse can span the entire range of values in Bi and Poo's weight-dependent STDP data [99], suggesting changes in synaptic efficacy of around $10,000\%$. No synapse in their experiments, however, changed in strength by more than around 100%. Many more pairings than the 60 of their protocol would be required to traverse this range, assuming their synapses could in fact be further strengthened. Alternatively, it is possible that the large variation in their initial weights (Figure 4.1B) reflects varying populations of synapses. Neurons in culture often make multiple post-synaptic contacts [29] and this possibility must be carefully addressed in future experiments on weight-dependent STDP.

## Acknowledgements

We thank Alan Fine and Stefan Kruger for helpful discussions. This work was supported in part by the NSERC grant RGPIN 249885-03.

# Chapter 5

## The Trouble with Weight-Dependent STDP

Dominic Standage and Thomas Trappenberg, The 2007 International Joint Conference on Neural Networks

## 5.1 Addendum to Standage and Trappenberg (2007)

In this chapter, we use the Log rule from Chapter 4 in a neural simulation. As described below, the simulation demonstrates that the common formulation of weight-dependent STDP rules is not supported by experimental data, and that rules consistent with the data lead to slow learning. Perhaps more importantly, we demonstrate that noise causes forgetting with STDP rules. We may conclude from this work that STDP rules are of limited utility without a means of protecting learned associations from noise. This conclusion speaks to the very nature of plasticity, assumed by STDP rules to occur in realtime, such that synaptic strengths are continually adjusted with ongoing spiking activity. If so, associations learned during periods of highly correlated pre- and post-synaptic spiking must be frequently 'refreshed'. Otherwise, background noise will lead all weights to noise-dependent values, regardless of prior learning.

I believe it is unlikely that plasticity works this way. If the role of plasticity is to capture correlations between pre- and post-synaptic spiking, it seems more likely that a statistical burden of proof should be established before a synapse changes in strength. This (possibly all-or-none) change may also reflect the level of correlation 'proved'. Positively correlated pre- and post-synaptic activity should therefor lead to potentiation, negatively correlated activity should lead to depression, and uncorrelated activity should not cause synaptic change. According to this hypothesis, asymptotic weights for uncorrelated activity, so extensively studied in the literature (see Chapter 4), may tell us little about plasticity and learning. It is worth noting that studies of asymptotic weights for STDP rules have only considered correlated pre- and post-synaptic spike *rates*. Plasticity is presumably a function of spike rate and spike timing correlations. Understanding the respective roles of spike rate and timing is of fundamental importance not just to plasticity and learning, but to information processing in the brain more generally.

I have stated that a burden of proof should be established before a synapse changes in strength. This proof does not preclude so-called one-shot learning in the hippocampus. One-shot learning need not refer to plasticity resulting from one instance of correlated pre- and post-synaptic activity, but rather, to plasticity resulting from one instance of an event to be learned. Hippocampal representations are believed to be repeated at theta frequency [59, 40], providing multiple instances of the to-be-learned

correlation.

I have also stated that the common formulation of STDP rules implies that synapses continually change in strength with ongoing spiking activity. This implication is most constructively considered a *prediction.* As such, it calls for experiments to test it. Systematically varying the number of pairings in STDP experiments would appear a reasonable first tack in this regard. A specific prediction of STDP rules with a global maximum weight (discussed at length in this chapter) is that repeated pre-before-post pairings at medium latency (say 40ms) will eventually lead to synapses as strong as those resulting from low latency pairings. Experiments are needed to test this prediction. If synapses do in fact determine the level of pre- and post-synaptic correlation before changing to reflect that correlation, the called-for experiment will not lead to equally strong synapses.

## Abstract

We fit a weight-dependent STDP rule to the classic data of Bi and Poo (1998), showing that this rule leads to slow learning in a simulation with an integrate-and-fire neuron. The slowness of learning is explained by an inequality between the range of initial weights in the data and the largest relative potentiation. We show that slow learning can be overcome with an increased learning rate, but that this approach leads to rapid forgetting in the presence of realistic levels of background spiking. Our study demonstrates that weight-dependent STDP rules, commonly used in neural simulations, have biologically unrealistic consequences. We discuss the implications of this finding for several interpretations of weight-dependent plasticity and STDP more generally, and recommend directions for further research.

## 5.2 Introduction

Activity-dependent change in synaptic strength, or synaptic plasticity, is widely believed to provide the basis for learning and memory, as originally proposed by Hebb [54]. Countless physiological data reveal Hebbian plasticity, and similarly vast numbers of neural simulations show associative learning from activity-dependent rules. While early experiments showed potentiation [15] and depression [79] of synaptic strength as a function of pre-synaptic firing rates, more recent experimental methods reveal similar changes based on the timing of pre- and post-synaptic activity [72, 81, 99]. Experiments showing spike-time dependent plasticity (STDP) reveal that the direction (potentiation or depression) and magnitude of synaptic change depend on the order and latency of pre- and post-synaptic activity, within a time window on the order of tens of milliseconds [99, 133].

Weight-dependent STDP rules provide the additional constraint that changes in synaptic strength or *weight* are a function of the initial strength of a synapse. Such rules have been widely studied [66, 128, 49, 17], largely because they provide a cap on synaptic strength while leading to asymptotic weight distributions that compare favourably with experimental data [128] and that are believed to preserve the statistics of input activity [49]. As such, weight-dependent STDP rules are now commonly used in modelling studies.

While several experimental studies have shown weight-dependent plasticity [99, 30, 87, 130], only Bi and Poo's (1998) study [99] has done so under the STDP pairing protocol. In recent work [112] we fit a weight-dependent STDP rule to the data of Bi and Poo, showing how asymptotic weights depend on the correlation and rate of pre- and post-synaptic activity. Here, we use this rule in a neural simulation, demonstrating that self-organisation of weights and associative learning are achieved, but that learning is slow. While an increased learning rate accelerates the learning process, it also leads to rapid forgetting in the presence of biologically realistic background spiking.

Our study shows that the common formulation of weight-dependent STDP rules [66, 62, 49] has biologically unrealistic consequences in rules fit to data. These consequences result from the assumption that weights may traverse the full range of initial values shown in weight-dependent studies (Bi and Poo's weight-dependent data are

shown in Figure 5.1B). This assumption is inconsistent with Bi and Poo's data, and also with weight-dependent data resulting from other plasticity protocols [30, 130]. Given the common inclusion of weight-dependent terms in learning rules, this issue is an important one, and poses questions about experimental methods, data and their interpretation.

## 5.3 A Weight-Dependent STDP Rule Fit to Data

For simplicity, we assume that weight- and spike-time-dependencies are independent. Under this assumption, the general form of a weight-dependent STDP rule may be expressed as

$$\Delta w_{\{p,d\}} = k f_{\{p,d\}}(w) e^{-c_{\{p,d\}}\Delta t}, \tag{5.1}$$

where $\Delta w_{\{p,d\}}$ is the change in weight ($p$ for potentiation, $d$ for depression), $\Delta t = t_{post} - t_{pre}$ is the time between pre- and post-synaptic firing, $c$ captures the timescale of the STDP window, and $k$ is a learning rate.

A log-linear fit to Bi and Poo's weight-dependent STDP data, hypothesized by Bi and Poo and reproduced in Figure 5.1B, yields the following form of $f(w)$:

$$f_{\{p,d\}}(w) = (a_{\{p,d\}} - b_{\{p,d\}} \log w) w \; \theta(w), \tag{5.2}$$

where $\theta(w)$ is the Heaviside function, keeping weights positive, and the parameters $a$ and $b$ are different for potentiation and depression.

Bi and Poo controlled spike timing in their weight-dependent experiment by limiting their spike latencies $\Delta t$. We used the midpoints of these latencies in our weight-dependent fit ($\Delta t_p = 10, \Delta t_d = 17.5$) and assumed that for a given $\Delta t$, the largest absolute values for $\Delta w/w$ represent synapses with the smallest initial weights. We therefor assign an initial weight $w = 30$pA to the largest STDP data (open symbols in Figure 5.1A) and only include these data in our spike-time fit for $c$. The resulting parameter values are $a_{\{p,d\}} = \{208, -54\}$, $b_{\{p,d\}} = \{26.4, 3.5\}$, $c_{\{p,d\}} = \{0.054, 0.042\}$. Because these parameters were fit to data expressed in percentage terms, and because we make the common assumption that each of the 60 pairings in Bi and Poo's experiment contributed equally to synaptic change [128], we use a learning rate of

$k = 1/6000$.

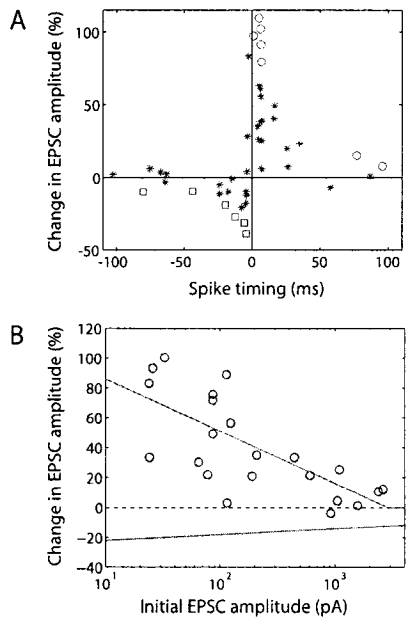


Figure 5.1: (A) Reproduction of Bi and Poo's (1998) spike-time dependent data. Circles and squares represent estimates of $w = 30$pA for potentiation and depression respectively. Stars represent estimates of $w > 30$pA. (B) Log-linear fit to Bi and Poo's weight-dependent STDP data, imposing a maximum weight where the upper solid line meets the x-axis. Open circles and solid stars show data obtained under potentiation (pre-before-post) and depression (post-before-pre) protocols respectively.

## 5.4 STDP and Spike Interactions

STDP rules are fit to data from experiments in which pairings of pre- and post-synaptic spikes occur at low frequency, typically 0.5-5Hz. The long inter-spike intervals (ISI's) of these spike trains effectively isolate spike pairings from the effects of previous and subsequent spikes. How to apply STDP rules to spike trains more realistic than temporally isolated pairings of pre- and post-synaptic spikes is the subject of great research interest [109, 11, 41, 58, 130, 42]. Here, we use a model we call *closest pair* interactions, where only the pair of pre-synaptic spikes closest to the surrounding post-synaptic spikes contributes to plasticity, as shown in Figure 5.2.

When exactly one pre-synaptic spike falls between two post-synaptic spikes, this model is equivalent to the nearest neighbour model suggested in [58]. These two

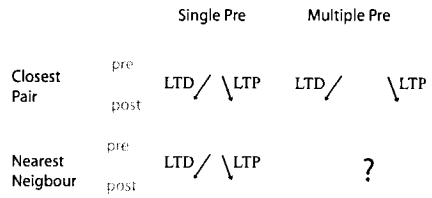


Figure 5.2: Cartoons depict spike interaction models described in the text. It is unclear how the *nearest neighbour* model extends to the case where multiple pre-synaptic spikes fall between two post-synaptic spikes. We define the *closest pair* model for this case.

interaction models are depicted in Figure 5.2. Because the analysis in [58] does not consider spike trains beyond this simplified case, we compare asymptotic weights calculated with their analytical method to asymptotic weights from simulations under closest pair interactions.

To this end, we numerically approximate the means of equilibrium weight distributions (equilibrium weights $w^*$) for closest pair interactions, comparing them to analytic values for the nearest neighbour case. In the simulations, post-synaptic spikes were time-locked to Poisson-distributed pre-synaptic spikes at $\Delta t = 4$ms. Weights were averaged over 5000 pairings following 5000 equilibrating pairings at spike rates $r = \{1, 2, 4, 8, 16, 32, 64, 128\}$Hz. Analytic calculations for the nearest neighbour case with the same spiking statistics were calculated in [112]. The relative difference between numeric and analytic asymptotic weights for these spike interaction models is shown in Figure 5.3. There is a systematic difference that increases with increasing spike rate, but the relative difference is very small.

## 5.5 Simulations with a Neuron

To test the effectiveness of Equation 5.1 for associative learning, we drive an integrate-and-fire (IF) node with two groups of synapses. Before learning, synchronous input from one group is strong enough to sporadically drive the node in combination with background activity from all synapses. The other group cannot drive the node. We test whether repeated synchronous activation of the high-strength group allows self-organisation of these weights, and whether synchronous activation of both groups (following self-organisation) is sufficient for the low-strength group to 'piggyback' the

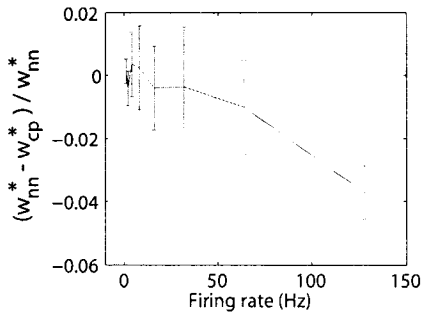


Figure 5.3: The relative difference between mean weights $w^*$ under *nearest neighbour* (analytic) and *closest pair* (numeric) spike interactions (see text). There is a very small (but significant) relative difference at high rate. Error bars show standard deviations.

high-strength group. If so, the low-strength group will have learned to drive the node by associativity.

### 5.5.1 Model and Parameters

The membrane potential of the IF node is described by

$$\tau_m \frac{dv(t)}{dt} = V_{rest} - v(t) + G_e(V_e - v(t)) + G_i(V_i - v(t)) \tag{5.3}$$

with membrane time constant $\tau_m = 20$ ms and reversal potentials $V_{rest} = -70$mV, $V_e = 0$mV and $V_i = -70$mV where subscripts $e$ and $i$ refer to excitatory and inhibitory potentials respectively. When the membrane potential crosses threshold $\Theta$, it is reset to $-60$mV with an absolute refractory period of 2mS. A relative refractory period is implemented by increasing $\Theta = -54$mV by $\gamma = 20$mV when the neuron fires, after which $\gamma$ decays exponentially with half width 10mS. Excitatory and inhibitory conductances $G_e$ and $G_i$ are described (as in [110]) by

$$\tau_e \frac{dG_e(t)}{dt} = -G_e(t) + \tau_e \Sigma_j \; g_j \; \delta(t - t_j^f) \tag{5.4}$$

and

$$\tau_i \frac{dG_i(t)}{dt} = -G_i(t) + \tau_i \Sigma_j \; g_j \; \delta(t - t_j^f), \tag{5.5}$$

where $\tau_e = \tau_i = 5$ms, $\delta$ refers to the Dirac delta function and $t_j^f$ is the time of firing of inputs mediated by conductance synapses $g_j$ (excitatory for $G_e$ and inhibitory for $G_i$).

### EPSC's as Conductances

There is no obvious way to translate Bi and Poo's EPSC measurements to synaptic conductances. To begin with, we do not know the amplitudes of corresponding EPSP's and so cannot determine conductances with Ohm's law. More problematic is the variation in these data ($\sim 25 - 2500$pA) and the size of the larger EPSC's. Because our objective is to study the implications these data for STDP rules, we cannot simply ignore the larger currents. Rather than use current synapses (independent of $v$), we use a linear transformation, allowing a range of 30-3000pA to correspond to 10-150pS. A maximum conductance of 150pS is used in [110]. Inhibitory synapses are not subject to plasticity and are uniformly allocated conductances of 500pS (also from [110]).

The simulation was run with 1000 excitatory synapses and 200 inhibitory synapses. We initialised the high-strength synapses (trainers) to uniformly distributed random values between 800 and 1200pA. The low-strength synapses (piggybackers) were initialised to 50pA, providing sub-threshold input. All other excitatory synapses were given uniformly distributed random values between 500 and 900pA. Poisson background firing with mean 10Hz was mediated by all synapses throughout the simulation. Synchronous inputs mediated by the trainers and piggybackers were given a uniformly distributed random jitter of 1-5ms, superimposed on the Poisson activity. Under our transformation to conductances, these weights are too weak for background activity to drive the node, but are able to elicit occasional firing when the trainers fire synchronously on top of background activity. Axonal delays $AD$ were assigned uniformly distributed random values (to the nearest integer) from $AD = 4 - 17$ms for excitatory synapses and $AD = 3 - 6$ms for inhibitory synapses. Equation 5.1 was applied under the closest pair spike interaction model.

### 5.5.2 Associative Learning under Equation 5.1

The IF node was first driven by Poisson activity alone, establishing a baseline membrane potential. This activity was followed by a period of self-organisation by the trainer synapses, driven by periodic synchronous activity at 10Hz. Periodic 10Hz activity provides a sufficiently long ISI to isolate neural activity in each cycle from the

effects of the previous cycle. Following this activity, the piggybackers were simultaneously activated with the trainers for the associative task, again at 10Hz. The node was subsequently driven by Poisson activity for comparison with the pre-training response to noise.

Figure 5.4A shows the trajectories of three trainer synapses and three piggyback synapses over the course of the simulation. The lower solid curve depicts a fast trainer synapse with a small initial weight [$AD$ = 4ms, $w(t = 0)$ = 800pA], the upper solid curve shows a slow trainer with a large initial weight [$AD$ = 17ms, $w(t = 0)$ = 1200pA] and the middle curve shows an intermediate case [$AD$ = 9ms, $w(t = 0)$ = 1000pA]. Early during self-organisation, the trainer group is not strong enough to regularly fire the node, but in combination with background activity, input synchrony generates enough post-synaptic firing for weights to increase until the group regularly drives the node.

Post-synaptic firing during early and late self-organisation is depicted by the upper and lower insets of the figure respectively. All three weights initially increase because they all contribute to post-synaptic activity. As faster weights gain in strength, EPSC onset of slower weights occurs after post-synaptic firing, subjecting them to LTD. As expected, this effect is first exhibited by the slow synapse and later the by mid-latency synapse. The former is effectively inactivated, with final value $w(t = 4 \cdot 10^5\text{ms}) = 2.4\text{pA}$. The latter assumes an intermediate value [$w(t = 4 \cdot 10^5\text{ms}) = 1009\text{pA}$] determined by an ongoing cycle of 'promotion' and 'relegation' to and from the effective input group. This cycle is governed by the interaction of background firing and post-synaptic activity, and the overlap between the jitter in synchronous input and axonal delay. The fast synapse reaches an asymptotic value close to the log rule cut-off.

The dashed curves in Figure 5.4A show the trajectories of three piggyback synapses. The arrow at the top of the figure shows onset of associative learning. At this point, these small synapses are climbing toward equilibrium weights determined by their background firing and the now-periodic post-synaptic firing generated by the trainer group. With the onset of associative activity, the fast piggyback synapse follows the fast trainer synapse toward an asymptotic weight, the mid-latency synapse assumes a value similar to that of the mid-latency trainer synapse, and the slow piggyback

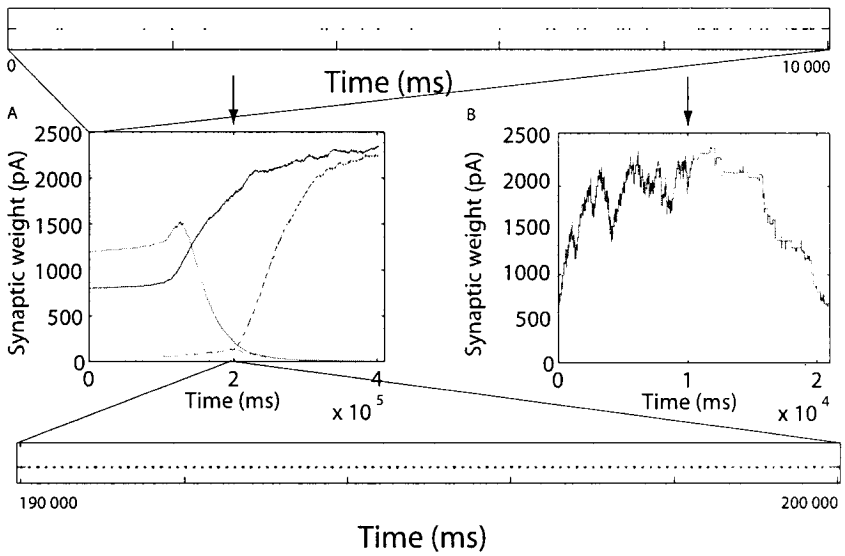


Figure 5.4: (A) Self-organisation and associativity under Equation 5.1 with an integrate-and-fire node. Curves show weights over time for three synapses with large initial weights (trainers, see text) and three synapses with small initial weights (piggybackers, see text). Curves to the left of the arrows show weights as the trainers self-organise during periodic synchronous activity (10Hz, uniform-distributed random jitter 1-5ms) on top of Poisson background activity (mean 10Hz). Arrows depict the onset of associative learning when the piggyback synapses fire synchronously with the trainers, also at 10Hz. A range of 30-3000pA is linearly compressed to 10-150pS for use by the node (see text). The upper solid curves show the weights of three trainers with axonal delays of $AD = 17$, $AD = 9$ and $AD = 4$ms respectively, top to bottom on the left of the figure. During self-organisation, all synapses increase in weight until the faster trainers drive the node without the slower ones, rendering slower synapses subject to depression. The same effect is seen for the piggyback synapses during associative learning. Self organisation and associativity are achieved, but learning is very slow with learning rate $k = 1/6000$. The fast ($AD = 4$ms) piggyback synapse is still converging to the log rule cut-off after 200,000ms (2000 pairings). The firing of the node during early and late self-organisation is shown by the top and bottom insets respectively. (B) An increased learning rate ($k = 1/200$) accelerates learning, but background noise soon causes forgetting.

synapse is effectively inactivated.

### Learning is Extremely Slow under Equation 5.1

Self-organisation and associativity have been achieved, as synchronous input by either group consistently drives the node. Furthermore, Equation 5.1 has selected the fastest weights within the groups and shown the ability to recruit and discard slower weights as and when needed, as shown in [110]. The mean membrane potential $\bar{v}$ in response to background noise has increased from $\bar{v} = -62.7$mV (standard deviation $sd = 1.1$) before learning to $\bar{v} = -60.7$mV ($sd = 1.2$) after learning, rendering the node more responsive to the timing of its inputs.

Unfortunately, the figure also shows that learning is extremely slow. The fast synapses are still experiencing net potentiation after 400,000ms. While 400s may not be considered slow on a behavioural timescale, our simulations are based on *in vitro* data. This interval accounts for around 3000 pairings for the trainers and 2000 for the piggybackers, approximately 50 times the number of pairings in Bi and Poo's experiments. Relating plasticity data to behavioural learning is fraught with difficulty [83] but the frequency of periodic, synchronous input in our simulations is within the theta range, consistent with the hypothesis that the theta rhythm provides an optimal pairing frequency for LTP during learning tasks [18].

## 5.5.3 Fast Learning Implies Fast Forgetting under Weight-Dependent STDP

To combat the slowness of learning shown in Figure 5.4A we raise $k$ from $k = 1/6000$ to $k = 1/200$. With this learning rate, weights begin to asymptote after around 60 pre- and post-synaptic pairings, the same number of pairings used in Bi and Poo's experiments [99]. It is unclear what a learning rate $k \neq 1/6000$ represents, but an accelerated learning rate is required due to the large variation in synaptic strength in Bi and Poo's weight-dependent data (Figure 5.1), allowing weights to make the transition from low to high values in a reasonable period of time.

Figure 5.4B shows the weight of a synapse mediating the same pre-synaptic spike statistics as the trainers did in the learning task, but over a much shorter timespan (10,000ms). The post-synaptic response is time-locked to $\Delta t = 2$ms after pre-synaptic

activity, again superimposed on Poisson spiking at 10Hz. Learning is stopped at time $t = 10,000$ms (top arrow), after which pre- and post-synaptic spike trains are given independent Poisson distributions at 10Hz. The synapse forgets the learned association within 10s.

## 5.6 Discussion and Conclusions

It is possible to parameterize weight-dependent STDP rules to accomplish various learning tasks, but our study shows these rules have unwanted consequences when fit to data. The main consequence is due to the range of initial weights in the only weight dependent STDP data available [99], shown in Figure 5.1B. Weight-dependent STDP rules implicitly assume that a synapse can span this entire range of values, suggesting changes in synaptic efficacy around $10,000\%$. No synapse in these experiments, however, changed in strength by more than around 100% (see Figure 5.1A,B).

This issue suggests several possibilities. One possibility concerns the standard computational interpretation of weight-dependent plasticity, in which there is an implicit assumption that weight-dependent data do not reflect saturated synapses. As an example, consider a synapse in Bi and Poo's weight-dependent experiment with an initial strength of 25pA. According to their data, this synapse would not have been potentiated by more than around 100% of control following their potentiation protocol. It would therefor be no larger than around 50pA after 60 pairings, much less than their largest initial weights. If all synapses have the same maximum, then for all but the largest synapses, an additional 60 pairings in Bi and Poo's experiment would have lead to weight changes on the same order as the 60 pairings they performed. The consequences of this assumption are far reaching. For instance, STDP data typically suggest a gradient of weight changes in either direction as a function of the time between pre- and post-synaptic spikes [99, 41]. In the case of potentiation, the standard interpretation dictates that given enough pairings, all weights will reach the same maximum, regardless of the latency between pre- and post-synaptic spikes.

Alternatively, Bi and Poo's potentiation data may reflect saturated synapses. If so, to be consistent with the data, weight-dependent STDP rules must implement intrinsic synaptic maxima rather than a global maximum. For instance, a synapse

with an initial weight of around 25pA should have an intrinsic limit of around 50pA, regardless of the size of the larger initial weights. We are unaware of any weight-dependent Hebbian rule that operates under this premise and the utility of such a rule is unclear. Certainly, large and small synapses would have different roles under this scheme.

These possibilities suggest that weight-dependent learning rules with a global maximum must be supported by weight-dependent data in which the largest relative change in weight is roughly equal to the relative difference between the largest and smallest initial weights. We may express this relationship as

$$max(\frac{w_i^{fin} - w_i^{init}}{w_i^{init}}) \approx \frac{w_{max}^{init} - w_{min}^{init}}{w_{min}^{init}}, \tag{5.6}$$

where $w_i^{init}$ and $w_i^{fin}$ represent the weight of a synapse $i$ before and after potentiation respectively, and $w_{max}^{init}$ and $w_{min}^{init}$ are the largest and smallest initial weights respectively. The data of Montgomery et al. (2001) [87], produced by a pairing protocol with no spike-time-dependence, are roughly consistent with this equation for synapses with initial weights less than 50pA.

The weight-dependent data of Debanne et al. (1999) [30] are intriguing in this regard. These authors paired pre-synaptic spikes with post-synaptic bursts in organotypic hippocampal slice cultures. Their CA3-CA1 data are reminiscent of Bi and Poo's in that the relative difference in initial weights is much greater than the normalized weight change, suggesting (as above) that these synapses have widely varying instrinsic maxima or that they have not reached saturation. Their CA3-CA3 synapses, however, appear to show the opposite effect (the left side of Equation 5.6 is greater than the right). Their combined data fit roughly with Equation 5.6. The degree to which plasticity data from different protocols may be related under computational rules is an open question and is currently receiving considerable attention [108, 58].

We show in Section 5.5.2 that a weight-dependent STDP rule fit to weight-dependent STDP data leads to slow learning. As shown in Section 5.5.3, speeding up learning with an increased learning rate leads to rapid forgetting in the presence of noise. Learning thresholds [105] and hysteretic weight dynamics [65] provide potential solutions to this problem, but our study focuses on the standard weight-dependent STDP formulation defined by Equation 5.1 and investigated in numerous

studies [66, 128, 49, 17]. An unexplored solution to the problem may be grounded in pre- and post-synaptic mechanisms if we (quite happily) abandon the assumption that the effects of spike pairings sum linearly. Consider the insertion and removal of AMPA receptors (AMPAR's) in the post-synaptic membrane, correlated with plasticity in hippocampal regions CA1 and CA3 [86]. It seems likely that potentiation should be harder to initiate as the number of AMPAR's increases. Pre-synaptically, plasticity is expressed at least in part by changes in the probability of transmitter release [114]. Weight-dependence is implicit in such a mechanism, where synapses with a low probability of release have greater scope for potentiation, and synapses with a high probability of release have greater scope for depression. A convolution of pre- and post-synaptic plasticity mechanisms could protect learned associations if pre-synaptic plasticity were to precede post-synaptic plasticity in both directions. Under such a scheme, AMPAR's may not be inserted in the post-synaptic density until a sufficiently high probability of release has been achieved, providing an initial resistance to post-synaptic learning. In reverse, AMPAR's may not be removed from the post-synaptic density until the probability of release is sufficiently low, providing initial resistance to forgetting. Such a scheme requires further study.

Finally, it is possible that weight-dependent plasticity data reflect the statistics of *non*-weight-dependent plasticity in populations of synapses, as discussed by Debanne et al. (1999) [30]. This possibility suggests that tissue preparations and stimulation methods must be carefully considered when interpreting experimental data. Extracellular stimulation is commonly used to mimic pre-synaptic activity in STDP paradigms [133], but this method may stimulate hundreds if not thousands of synapses [94]. While intracellular stimulation of pre-synaptic neurons (used in Bi and Poo's experiments) eliminates the potential involvement of multiple pre-synaptic cells, neurons in culture (also used in Bi and Poo's experiments) typically make multiple contacts with the post-synaptic cell [29]. It is safe to say that single synapses do not mediate currents on the order of Bi and Poo's larger data ($\sim$ 2500pA). This issue overlaps with debates concerning pre- and post-synaptic contributions to plasticity [37, 92] and the degree to which plasticity is a graded phenomenon [96, 86, 94]. Varying the number of pairings in STDP experiments would directly address the latter. STDP protocols are clearly promising for the investigation of weight-dependent plasticity,

but the potentially confounding effects of populations of synapses must be carefully addressed.

**Acknowledgment**

We thank Alan Fine and Stefan Krueger for helpful discussions.

# Chapter 6

# A Population-Based Explanation of Bi and Poo's (1998) Weight-Dependent STDP Data

## 6.1 Introduction

The extrapolation of learning under the Log and Power rules by Standage, Jalil and Trappenberg (2007) [112] points to a fundamental problem with Bi and Poo's (1998) [99] weight-dependent STDP data and their interpretation. While the qualitative dependence of equilibrium weights on rate and correlation will hold for any weight-dependent STDP rule of the form discussed, the quantitative results suggest unrealistically large asymptotic weights, particularly for time-locked pre- and post-synaptic spiking at low rate (as in the STDP pairing protocol). Weights for the Power rule depend largely on the balance between potentiation and depression because this rule includes no maximum, but for the Log rule, these values correspond to the maximum weight determined by the log-linear fit in Section 2. The data show initial excitatory post-synaptic currents (EPSC's) ranging from around 25pA to 2500pA. The former is a realistic value of the strength of a single synapse. The latter is clearly not.

The effects of the larger initial weights on weight-dependent STDP rules are two-fold. Firstly, the maximum weight is too big. The second effect is more subtle. An unwanted consequence of STDP rules grounded in these data is that individual synaptic strengths may traverse the full range of initial values, despite the fact that no synaptic currents in [99] changed by more than $\sim$ 100% of control. Thus, if 60 pairings elicit weight changes on the order of 100%, many more pairings are required to elicit weight changes spanning the full range of initial weights ($\sim 10,000\%$). In effect, learning is very slow.

In this report, we provide an interpretation of the data based on a population argument. We demonstrate this interpretation with several variations of a simple model

based on two premises discussed in [30]: the range of initial EPSC measurements reflects the number of synapses recorded, and not all synapses are plastic. Evidence for non-plastic synapses in culture is found in [30]. We do not claim that plasticity is not weight-dependent, but only that Bi and Poo's weight-dependent data has been misinterpreted in weight-dependent STDP rules.

## 6.2 Simulations of Synaptic Populations

To the best of our knowledge, the only studies showing weight-dependent plasticity [99, 30, 87, 130] have used cultured preparations. Whereas axons *in vivo* typically make one or two synaptic contacts with a post-synaptic cell, axons in cultured preparations can make multiple contacts [29]. It is therefore possible that in cultured neurons, the percentage increase in strength following the application of potentiation protocols reflects the proportion of synaptic contacts activated relative to the population size. Under this hypothesis, potentiation protocols effectively identify 'functional' synaptic contacts. Conversely, the lack of substantial (percentage) weight-dependent change under depression protocols suggests that such functional synapses are equally likely to be deactivated.

We start by assuming that the number of synaptic contacts contributing to each of the data in Bi and Poo's weight-dependent experiment varies for each datum. We choose an exponential distribution in the examples below because a histogram of initial values of their data is reasonably approximated by this distribution, as shown in Figure 6.1A. In all simulations, we assign a mean of 40 synapses per population. Data on this issue are mixed, with [29, 30] citing 'several' and [95] citing 'order ten'. Because Bi and Poo's data show initial weights as high as $\sim$ 2500pA and single EPSC's are generally in the range of $10 - 30$pA, a mean number of forty bouton terminals seems fairly conservative, but changes to this mean value do not qualitatively effect our results. In the following experiments, each population has a minimum of 3 synapses and a mean EPSC of 15pA.

The basic model is as follows. We assign synapses within each population the status of *on* or *off*. These states may be equated with high and low strength states respectively per the findings of O'Connor et al (2005) [95] or, perhaps more closely, with potentiated and silent states per the findings of Montgomery and Madison (2002)

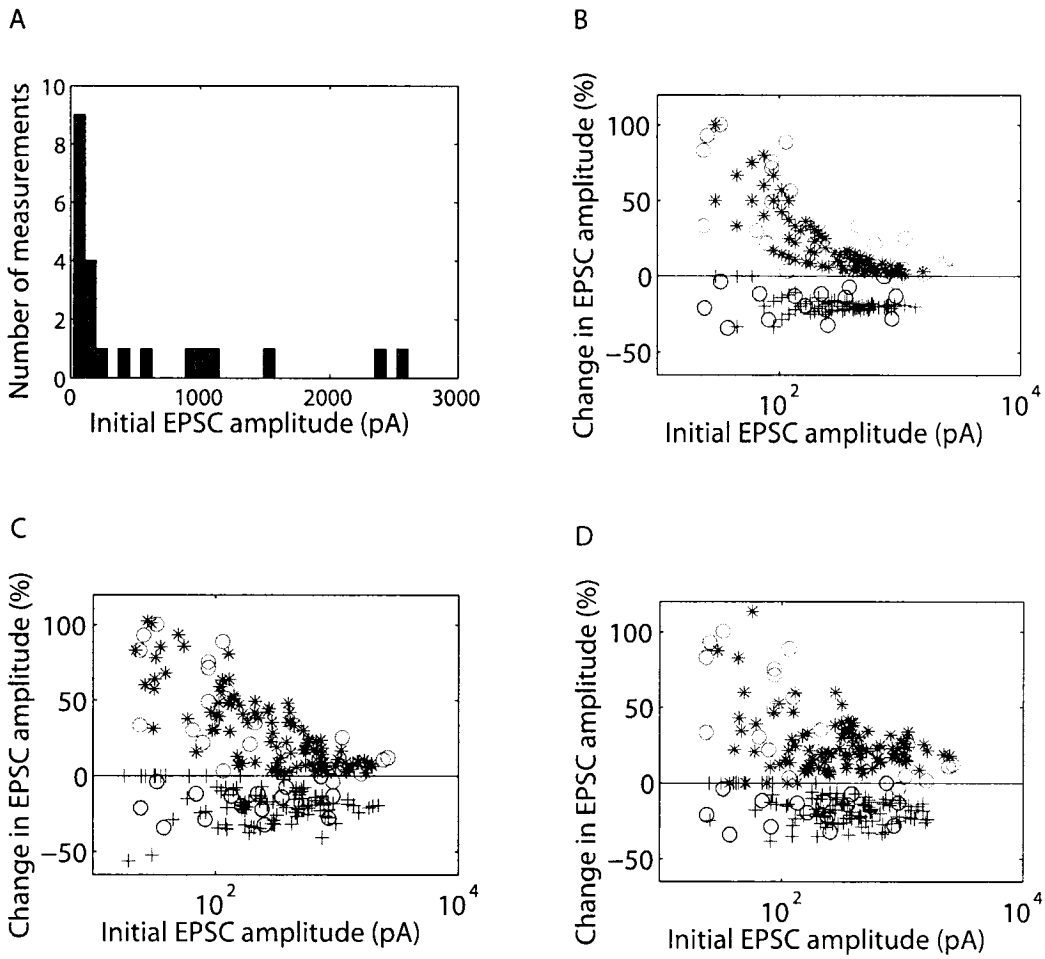


Figure 6.1: Bi and Poo's (1998) weight-dependent data may reflect the potentiation of subsets of synaptic populations in cultured cells. Synapses within each population are assumed *on* or *off* prior to STDP, where the potentiation protocol turns on (potentiates) some number $n$ of the *off* synapses. The depression protocol is assumed to turn off an average of 20% of synapses. Means of 40 synapses per group and 15pA per synapse are assumed in $B$, $C$ and $D$ where all groups have a minimum population of 3. (A) A histogram of initial weights in Bi and Poo's weight-dependent data suggests that population sizes may be approximated with an exponential distribution. (B) Results when all *on* synapses are equal in strength. Populations are initialised with 50% of synapses *on* and 50% *off*, consistent with Bi and Poo's maximum percentage change of $\sim$ 100%. Stars show results for $1 \leq n \leq 4$ where $n$ is uniformly distributed at random. Pluses show results when 20% of synapses are turned off, rounded to the nearest integer. Circles show Bi and Poo's data. (C) Weights within each population are given a Chi Squared distribution, 70% of synapses begin in the *on* state and $1 \leq n \leq 10$. For depression, synapses are turned off with probability 0.2. (D) Synapses are again given a Chi Squared distribution, but are turned on by the potentiation protocol with probability 0.5, and turned off by the depression protocol with probability 0.2. All synapses are *on* prior to STDP with probability 0.7.

[85]. While these studies show evidence for intermediate states, we assume that the 60 pairings in [99] are sufficient to fully navigate any such series of states. The computational implications of discrete state synapses are explored in [2]. We simulate the potentiation protocol by turning on some number $n$ of the *off* synapses, where $n$ is a random number. We simulate the depression protocol by randomly turning off on average 20% of the *on* synapses because Bi and Poo's weight-dependent depression data show this approximate change, regardless of initial EPSC amplitude.

### 6.2.1 Simulation 1

Figure 6.1B shows results for the simplest case where all *on* synapses are equal in strength. Because the data [99] (circles) show a maximal percentage change of $\sim 100\%$ we initialise populations with 50% of their synapses on and 50% off. Stars show results for $1 \leq n \leq 4$ where $n$ is uniformly distributed. The qualitative effects of weight-dependent relative potentiation and a constant average relative depression are reproduced by the model, but for the depression data in particular, the variance in relative change is systematically greater for small weights than for large weights.

### Simulation 2

Figure 6.1C shows results when the strengths of synapses within each population are allowed to vary. This variation may either be equated with intrinsic maxima among potentiated synapses or with synaptic strengths following graded plasticity. We give these values a Chi Squared distribution, as found by measurements of miniature EPSC's [126] and quantitative immunofluorescence [93] in cultured cells, but this distribution is not crucial to our results. With non-uniform weights, initialising populations with half their synapses on allows percentage change to exceed 100% control, but the qualitative effect of weight-dependence remains. In Figure 6.1C, 70% of synapses begin in the *on* state and we allow greater variation in $n$ with $1 \leq n \leq 10$.

### Simulation 3

In the previous two simulations, a greater proportion of synapses are activated in small populations than large populations because $n$ is small compared to mean population size. In the final simulation shown in Figure 6.1D, we abandon this assumption. That

is, for potentiation, all synapses in all groups are turned on with equal probability (0.5 in the figure) regardless of group size. Similarly for depression, all synapses are turned off with equal probability (0.2 as above). Population sizes again follow an exponential distribution and individual weights follow a Chi Squared distribution. Synapses begin in the *on* state with a probability of 0.7. Under this set of assumptions, the weight-dependence in the relative potentiation data is explained by the variation in intrinsic maxima, which scales away in large populations. This variation is less noticeable for depression because individual synapses are less likely to be depressed than potentiated in the simulation, consistent with experimental findings that LTP requires fewer pairings than LTD under a pairing protocol [94]. LTD has also been shown to require prolonged activation under extra-cellular stimulation and calcium uncaging, as described in [133].

## 6.3 Conclusions

Under all three sets of assumptions, the above simulations account for the weight-dependence in [99]. Our analyses rest on the assumptions that synaptic populations vary in size and that not all synapses are potentiated in experiments. The last simulation makes the assumptions that weight dependence may be explained by the variability in synaptic strength relative to population size, discussed by Debanne et al. (1999) [30] in relation to their potentiation data, and that potentiation is more readily achieved than depression.

# Chapter 7

# Discussion

In the chapters above, we have addressed two fundamental aspects of cortical memory: interactions within and between cortical regions and synaptic plasticity. In Chapter 3, we investigate the influence of bottom-up and top-down processing on the spatial distribution of selective visual attention. The model of Deco and colleagues [31, 26] provides the context for this work, where the location of visual attention is determined by hierarchical, bidirectional processing in the dorsal and ventral visual pathways. This instantiation of the biased competition model of attention [33] need not be limited to the visual system. Indeed, if a common algorithm operates throughout cortex, as originally proposed by Mountcastle (1978) [89], then biased competition may rightly be considered a *modus operandi* of cortex. In this regard, the dorsal and ventral visual streams are just two of a vast number of cortical pathways converging on higher association cortices such as prefrontal cortex (PFC). In the context of visual processing, the resolution of competition in PFC strongly influences feature selection in V1. With a uniform cortical algorithm, this competition influences processing in association cortices and across modalities.

This general framework for understanding cortex is not new. In Grossberg's Adaptive Resonance Theory (ART) [46, 47, 21], feature representations converge on a centre surround neural field (CSNF) model, where the winner-take-all (WTA) property of the model ensures the integration of these features. That is, the CSNF model ensures bottom-up compositionality. Because cortical processing is bidirectional, coactive representations at both levels of this simplified hierarchy facilitate Hebbian learning in both directions. Bottom-up learning ensures that features that regularly occur together consistently activate the same representation. Top-down activity therefor represents learned expectations, serving to predict lower level activity. Under ART, learned expectations and attention are synonymous. Top-down representations propagate their activity back to feature layers, priming the features that have consistently

activated them in the past. A mismatch between learned expectations and sensory reality is effectively a novelty signal and leads to new learning. In a more extensive hierarchy, the principles of ART fit biased competition like a glove. In Chapter 3, we equate short term memories (STM's) in PFC with foreknowledge of stimulus characteristics (features). Active maintenance of these representations serves their influence on higher sensory cortices, where representations prime the expected features for detection.

In Chapter 2, we balance the competing factors of inhibition and stabilization in a CSNF model. This work addresses two crucial aspects of cortical function. As a model of interference between STM's in PFC, it addresses the competition that biases lower-level feature detection. Additionally, if a CSNF model represents a small patch of cortex responding to a particular feature space [51], then retaining the WTA property of the model is crucial to its function as an integrator of lower level features, serving bottom-up composition in hierarchical processing. It is also crucial for matching top-down expectations to sensory features. In this regard, an important direction for future work concerns the relative strengths of bottom-up, top-down and lateral signalling in CSNF models. The WTA property of the model is not only relevant to the active maintenance of STM's, but also to the integration of convergent bottom-up and top-down activity when network parameters do not necessarily support sustained representations. For instance, in Chapter 3, lateral inhibition sharpens the input to our model of posterior parietal cortex (PP), but not enough to force a winning bubble during the life of these inputs. Under the present framework, a winning bubble may be required while inputs are converging on the centre surround network. This hypothesis requires further study.

In the context of a model such as ART, cortex cannot function without Hebbian learning. Our studies of spike time dependent plasticity (STDP) in Chapters 4, 5 and 6 address this aspect of memory. While Chapter 4 makes important contributions to our understanding of weight-dependent plasticity under the standard STDP pairing protocol, it also highlights that the common formulation of weight-dependent STDP rules [66, 62, 49] is not supported by experimental data. Chapter 5 investigates the consequences of this finding on associative learning. The focus of this work is not 'negative' in this respect. Quite the contrary, it specifically identifies problems

requiring further study, both experimentally and computationally. One such problem is addressed in Chapter 6, where we show that weight-dependent STDP data [99] may reflect *non*-weight-dependent plasticity in populations of synapses. The consequences of this possibility are far-reaching, as similar tissue preparations have been used in numerous plasticity experiments.

As a model of learning, STDP has had an immense impact on the neuroscience community. Perhaps most notably, it has lead to extensive research on the interactions between pre- and post-synaptic signals [76]. Many of these interactions have tight temporal constraints, revealing the dependence of plasticity on spike rate and spike timing on both short and long timescales [12]. STDP is consistently demonstrated in cortical preparations [133] and the recent theory of Hawkins (2004) [132] suggests that sequence learning may be crucial to cortical function, explaining how invariant representations develop across modalities and in association cortices. The temporal asymmetry of STDP would appear well suited to this learning task. Hierarchical, bidirectional, spatiotemporal learning provides an exciting direction for future research on cortical memory.

As a final word on STDP, I wish to distinguish between STDP as an experimental protocol and as a basis for computational learning rules. As an experimental protocol, STDP reveals that the order and latency of repetitive, low-frequency pre- and post-synaptic spike pairings determines the sign and magnitude of plasticity, but this spike-timing arrangement is highly artificial. Recent experiments have therefor extended STDP pairings to temporally overlapping triplets, quadruplets and quintuplets of spikes [109, 11, 130, 42], showing that STDP (as described above) breaks down under more 'natural' conditions. These extended protocols would appear to provide a model of oscillatory spiking activity, long correlated with learning in the hippocampus. Experiments have shown that rhythmic stimulation in the theta frequency range (4-12Hz) may provide optimal conditions for LTP in CA3-CA1 synapses [71, 45, 18], and recent experiments show that a minimum frequency of 5Hz is necessary for LTP in CA3-CA1 synapses under an STDP protocol, as is post-synaptic bursting [133]. However, these STDP data are qualitatively different than the temporally asymmetric data discussed above, showing that a tight temporal coincidence of pre- and post-synaptic activity yields potentiation and a more loose coincidence yields

depression. Given that CA3 and CA1 pyramidal neurons are intrinsically bursting cells, it seems unlikely that the exact timing of individual spikes should matter under these conditions. Indeed, experiments using cortical preparations also show that when post-synaptic spike quintuplets lead pre-synaptic quintuplets at burst frequencies, LTP is the result [109].

Extended STDP experiments raise other questions not yet addressed in the literature. For instance, plasticity has long been known to be a pre-synaptic phenomenon as well as a post-synaptic phenomenon. Pre-synaptically, the probability of transmitter release $[P(rel)]$ is highly plastic, where LTP and LTD are manifest as an increase and decrease in $P(rel)$ respectively [114, 37]. In Chapter 5, we conjecture that pre-synaptic plasticity may shield post-synaptic plasticity from noise, but given that $P(rel)$ is also subject to short term potentiation (STP) and depression (STD), what role is played by long term plasticity of $P(rel)$ in burst firing? With each pre-synaptic spike in a burst, $P(rel)$ increases by STP mechanisms [77]. Conversely, $P(rel)$ decreases by STD mechanisms once release has been achieved. Thus, if only one spike per burst is transmitted, as suggested by Lisman [77], then plasticity of $P(rel)$ may sharpen oscillatory activity by reducing the latency of release. This hypothesis requires further investigation. Recent experiments in Alan Fine's lab (personal correspondence) show that $P(rel)$ can be potentiated to values close to 1 with a very small number of pre- and post-synaptic pairings ($\sim 10$). If pre- and post-synaptic plasticity mechanisms do indeed follow different timecourses, the respective computational roles of these processes are largely unexplored.

With respect to computational learning rules, it is important to note that STDP rules are phenomenological, and as such, do not directly address the physiological mechanisms underlying plasticity. These rules are also grounded in a single plasticity protocol and do not easily translate to data obtained with other methods. Theoreticians have investigated so-called spike interaction models, where the contribution of individual spikes to STDP is subject to attenuation and recovery [41, 42]. Such models attempt to capture the effect of interactions between pre- and post-synaptic signals. In this regard, they provide a legitimate means of addressing plasticity under different protocols, but I believe pre- and post-synaptic spike interactions may be better captured with a dynamic variable, most often correlated with calcium flux in

the post-synaptic density [60, 108, 1]. Such generative models have the potential to address the common physiological processes believed to underly plasticity under all protocols.

## Bibliography


[1] Henry D. I. Abarbanel, Leif Gibb, R. Huerta, and M. I. Rabinovich. Biophysical model of synaptic plasticity dynamics. *Biological Cybernetics*, 89:214–226, 2003.

[2] Henry D. I. Abarbanel, Sachin S. Talathi, Leif Gibb, and M. I. Rabinovich. Synaptic plasticity with discrete state synapses. *Physical Review E*, 72:031914–1–031914–14, 2005.

[3] L. F. Abbott and Kenneth I. Blum. Functional significance of long-term potentiation for sequence learning and prediction. *Cerebral Cortex*, 6:406–416, 1996.

[4] L. F. Abbott and Sacha B. Nelson. Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–1183, 2000.

[5] Kara L. Agster, Norbert J. Fortin, and Howard Eichenbaum. The hippocampus and disambiguation of overlapping sequences. *The Journal of Neuroscience*, 2(13):5760–5768, 2002.

[6] S. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.

[7] Hisham E. Atallah, Michael J. Frank, and Randall C. O'Reilly. Hippocampus, cortex, and basal ganglia: Insights from computational models of complementary learning systems. *Neurobiology of Learning and Memory*, 82:253–267, 2004.

[8] Edward Awh and Harold Pashler. Evidence for split attentional foci. *Journal of Experimental Psychology, Human Perception and Performance*, 26(2):834–846, 2000.

[9] Wyeth Bair, Christof Koch, William Newsome, and Kenneth Britten. Power spectrum analysis of bursting cells in area mt in the behaving monkey. *Journal of Neuroscience*, 14(5):2870–2892, 1994.

[10] F. P. Battaglia and A. Treves. Attractor neural networks storing multiple space representations: A model for hippocampal place fields. *Physical Review*, 58(6), 1998.

[11] Guo-Qiang Bi. Spatiotemporal specificity of synaptic plasticity: cellular rules and mechanisms. *Biological Cybernetics*, 87:319–332, 2002.

[12] Guo-Qiang Bi and Jonathan Rubin. Timing in synaptic plasticity: from detection to integration. *Trends in Neurosciences*, 28(5):222–228, 2005.

[13] Elie Bienenstock. A model of neocortex. *Network: Computation in Neural Systems*, 6:179–224, 1995.

[14] Elie L. Bienenstock, Leon N. Cooper, and Paul W. Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2(1):32–48, 1982.

[15] T. V. P. Bliss and T. J. Lomo. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, 232:331–356, 1973.

[16] Christian Buchel. Cortical hierarchy turned on its head. *Nature Neuroscience*, 6(7):657–658, 2003.

[17] Anthony Burkitt, Hamish Meffin, and David B. Grayden. Spike-timing-dependent plasticity: The relationship to rate-based learning for models with weight dynamics determined by a stable fixed point. *Neural Computation*, 16:885–940, 2004.

[18] Gyorgy Buzsaki. Theta oscillations in the hippocampus. *Neuron*, 33:325–340, 2002.

[19] M. Camperi and X.-J. Wang. A model of visuospatial short-term memory in prefrontal cortex: cellular bistability and recurrent network. *J. Comput. Neurosci.*, 5:383–405, 1998.

[20] Boris Gutkin Carlo R. Laing, William C. Troy and G. Bard Ermentrout. Multiple bumps in a neuronal model of working memory. *SIAM J. Appl. Math.*, 63(1), 2002.

[21] Gail A. Carpenter and Stephen Grossberg. Adaptive pattern classificationa and universal recoding, ii: feedback, expectation, olfaction, and illusions. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*. MIT Press, 2nd edition, 2002.

[22] E.C. Cherry. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25:975–979, 1953.

[23] Paul Cisek. Integrated neural processes for defining potential actions and deciding between them: a computational model. *Journal of Neuroscience*, 65–66, 2006.

[24] Carol L. Colby and Michael E. Goldberg. Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22:319–349, 1999.

[25] A. Compte, N. Brunel, P.S. Goldman-Rakic, and X.-J. Wang. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10:910–23, 2000.

[26] S. Corchs, M. Stetter, and G. Deco. Systems-level neuronal modeling of visual attentional mechanisms. *Artificial Intelligence Review*, 20:143–160, 2003.

[27] Nelson Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24:87–114, 2001.

[28] Yang Dan and Mu ming Poo. Spike timing-dependent plasticity of neural circuits. *Neuron*, 44:23–30, 2004.

[29] Dominique Debanne, Beat H. Gahwiler, and Scott M. Thompson. Cooperative interactions in the induction of long-term potentiation and depression of synaptic excitation between hippocampal ca3-ca1 cell pairs in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 93:11225–11230, 1996.

[30] Dominique Debanne, Beat H. Gahwiler, and Scott M. Thompson. Heterogeneity of synaptic plasticity at unitary ca3-ca1 and ca3-ca3 connections in rat hippocampal slice cultures. *Journal of Neuroscience*, 19(24):10664–10671, 1999.

[31] G. Deco, O. Pollatos, and J. Zihl. The time course of selective visual attention: theory and experiments. *Vision Research*, 42:2925–2945, 2002.

[32] Sophie Deneve, Alexandre Pouget, and Peter E. Latham. Divisive normalization, line attractor networks and ideal observers. *Advances in Neural Processing Systems*, 11:104–110, 1999.

[33] Robert Desimone and John Duncan. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18:193–222, 1995.

[34] C.J. Downing and S. Pinker. The spatial structure of visual attention. In M.I. Posner and O.S.M. Marin, editors, *Attention and performance*, volume 11, pages 171–187. Erlbaum, 1985.

[35] Serena M. Dudek and Mark F. Bear. Homosynaptic long-term depression in area ca1 of hippocampus and effects of n-methyl-d-aspartate receptor blockade. *Proceedings of the National Academy of Sciences of the United States of America*, 89:4363–4367, 1992.

[36] J. Duncan and G.W. Humphreys. Visual search and stimulus similarity. *Psychological Review*, 96:433–458, 2002.

[37] Nigel J. Emptage, Christopher A. Reid, Alan Fine, and Timothy V.P. Bliss. Optical quantal analysis reveals a presynaptic component of ltp at hippocampal schaffer-associational synapses. *Neuron*, 38:797–804, 2003.

[38] Ceren Ergorul and Howard Eichenbaum. Essential role of the hippocampal formation in rapid learning of higher-order sequential associations. *The Journal of Neuroscience*, 26(15):4111–4117, 2006.

[39] C.W. Eriksen and J.D. St. James. Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, 40:225–240, 1986.

[40] Erik Fransen. Functional role of entorhinal cortex in working memory processing. *Neural Networks*, 18:1141–1149, 2005.

[41] Robert C. Froemke and Yang Dan. Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature*, 416(6879):433–438, 2002.

[42] Robert C. Froemke, Ishan A. Tsay, Mohamad Raad, John D. Long, and Yang Dan. Contribution of individual spikes in burst-induced long-term synaptic modification. *Journal of Neurophysiology*, 95:1620–1629, 2006.

[43] S. Funahashi, C.J. Bruce, and P.S. Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61:331–349, 1989.

[44] Dileep George and Jeff Hawkins. A hierarchical bayesian model of invariant pattern recognition in the visual cortex. In *The International Joint Conference on Neural Networks 2005*, 2005.

[45] Yoram J. Greensteinnext, Constantine Pavlides, and Jonathan Winson. Long-term potentiation in the dentate gyrus is preferentially induced at theta rhythm periodicity. *Brain Research*, 438(1–2):331–334, 1988.

[46] Stephen Grossberg. Adaptive pattern classification and universal recoding: I. parallel development and coding of neural feature detectors. *Biologial Cybernetics*, 23:187–202, 1976.

[47] Stephen Grossberg. Adaptive pattern classificationa and universal recoding, ii: feedback, expectation, olfaction, and illusions. *Biologial Cybernetics*, 23:121–134, 1976.

[48] Stephen Grossberg. How does the cerebral cortex work? learning, attention and grouping by the laminar circuits of visual cortex. *Spatial Vision*, 12:163–186, 1999.

[49] R. Gutig, R. Aharonov, S. Rotter, and Haim Sompolinsky. Learning input correlations through nonlinear temporally asymmetric hebbian plasticity. *Journal of Neuroscience*, 23:3697–3714, May 2003.

[50] S. Hahn and A. Kramer. Further evidence for the division of attention between non-contiguous locations. *Visual Cognition*, 5:217–256, 1998.

[51] David Hansel and Haim Sompolinsky. Modeling feature selectivity in local cortical circuits. In Christof Koch and Idan Segev, editors, *Methods in neuronal modeling, from ions to networks*, chapter 13, pages 499 – 467. MIT Press, second edition, 1998.

[52] Michael E. Hasselmo. Runaway synaptic modification in models of cortex: Implications for alzheimer's disease. *Neural Networks*, 7(1):13–40, 1994.

[53] Michael E. Hasselmo. Neuromodulation and cortical function: modeling the physiological basis of behavior. *Behavioural Brain Research*, 67:1–27, 1995.

[54] Donald O. Hebb. *The Organisation of Behaviour*. John Wiley, New York, 1949.

[55] John Hertz, Anders Krogh, and Richard G. Palmer. *Introduction to the theory of neural computation*. Addison-Wesley Publishing Company, 1991.

[56] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117:500–544, 1952.

[57] L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews: Neuroscience*, 2:194–203, 2001.

[58] Eugene M. Izhikevich and Niraj S. Desai. Relating stdp to bcm. *Neural Computation*, 15(7):1511–1523, 2003.

[59] Ole Jensen and John E. Lisman. Theta/gamma networks with slow nmda channels learn sequences and encode episodic memory: role of nmda channels in recall. *Learning and Memory*, 3:264–278, 1996.

[60] Uma R. Karmarkar and Dean V. Buonomano. A model of spike-timing dependent plasticity: one or two coincidence detectors? *Journal of Neurophysiology*, 88:507–513, 2002.

[61] Richard Kempter, Wulfram Gerstner, and J. Leo van Hemmen. Hebbian learning and spiking neurons. *Physical Review E*, 59(4):4498–4514, 1999.

[62] Adam Kepecs, Mark C.W. van Rossum, Sen Song, and Jesper Tegner. Spike-timing-dependent plasticity: common themes and divergent vistas. *Biological Cybernetics*, 87:446–458, 2002.

[63] Jeansok J. Kim and Mark G. Baxter. Multiple brain-memory systems: the whole does not equal the sum of its parts. *Trends in Neurosciences*, 24(6):324–330, 2001.

[64] Alfredo Kirkwood, Marc G. Rioult, and Mark F. Bear. Experience-dependent modification of synaptic plasticity in visual cortex. *Nature*, 381:526–528, 1996.

[65] M. V. Kiselev. Statistical approach to unsupervised recognition of spatio-temporal patterns by spiking neurons. In *The International Joint Conference on Neural Networks 2003*, pages 2843–2847, 2003.

[66] Werner M. Kistler and J. Leo van Hemmen. Modeling synaptic plasticity in conjunction with the timing of pre- and postsynaptic action potentials. *Neural Computation*, 12:385–405, 2000.

[67] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4):219–227, 1985.

[68] Alexandre Kuhn, Ad Aertsen, and Stefan Rotter. Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Computation*, 15:67–101, 2003.

[69] D. LaBerge and V. Brown. Theory of attentional operations in shape identification. *Psychological Review*, 96:101–124, 1989.

[70] Carlo R. Laing and William C. Troy. Two-bump solutions of amari-type models of neuronal pattern formation. *Physica D*, 178, 2003.

[71] John Larson and Gary Lynch. Induction of synaptic potentiation in hippocampus by patterned stimulation involves two events. *Science*, 232(4753):985–988, 1986.

[72] W. B. Levy and O. Steward. Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, 8:791–797, 1983.

[73] William B. Levy. A computational approach to hippocampal function. In R.D. Hawkins and G.H. Bower, editors, *Computational Models of Learning in Simple Systems*, volume 23 of *The Psychology of Learning and Motivation*, pages 243–305. Academic Press, 1989.

[74] William B. Levy. A sequence predicting ca3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus*, 6:579–590, 1996.

[75] J.E. Lisman, J.M. Fellous, and X.-J. Wang. A role for nmda-receptor channels in working memory. *Nature Neurosci.*, 1(4):273–5, 1998.

[76] John Lisman and Nelson Spruston. Postsynaptic depolarization requirements for ltp and ltd: a critique of spike timing-dependent plasticity. *Nature Neuroscience*, 8(7):839–841, 2005.

[77] John E. Lisman. Bursts as a unit of neural information: making unreliable synapses reliable. *Trends in Neurosciences*, 20(1):38–43, 1997.

[78] John E. Lisman. Relating hippocampal circuitry viewpoint to function: Recall of memory sequences by reciprocal dentate-ca3 interactions. *Neuron*, 22:233–242, 1999.

[79] G.S. Lynch, T. Dunwiddie, and V. Gribkoff. Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature*, 266:737–739, 1977.

[80] Robert C. Malenka and Roger A. Nicoll. Long term potentiation: a decade of progress? *Science*, 285(5435):1870–1874, 1999.

[81] Henry Markram, Joachim Lbke, Michael Frotscher, and Bert Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215, 1997.

[82] D. Marr. Simple memory: A theory of archicortex. *Philosophical Transactions of the Royal Society of London*, 262(841):23–81, 1971.

[83] S. J. Martin, P.D. Grimwood, and R. G. M. Morris. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annual Review of Neuroscience*, 23:649–711, 2000.

[84] P.A. McCormick, R.M. Klein, and S. Johnston. Splitting vs. sharing focal attention: comment on castiello and umilta (1992). *Journal of Experimental Psychology*, 24(1):350–357, 1998.

[85] Johanna M. Montgomery and Daniel V. Madison. State-dependent heterogeneity in synaptic depression between pyramidal cell pairs. *Neuron*, 33:765–777, 2002.

[86] Johanna M. Montgomery and Daniel V. Madison. Discrete synaptic states define a major mechanism of synapse plasticity. *Trends in Neurosciences*, 27(12):744–750, 2004.

[87] Johanna M. Montgomery, Paul Pavlidis, and Daniel V. Madison. Pair recordings reveal all-silent synaptic connections and the postsynaptic expression of long-term potentiation. *Neuron*, 29:691–701, 2001.

[88] V.B. Mountcastle and I. Darian-Smith. *Neural mechanisms of somethesia*, chapter 2, pages 1372–1423. St. Louis: Mosby, 12 edition, 1968.

[89] Vernon B. Mountcastle. An organizing principle for cerebral function: the unit module and the distributed system. In F. O. Schmitt, editor, *The Mindful Brain*. MIT Press, 1978.

[90] M.M. Muller, P. Malinowski, T. Gruber, and S.A. Hillyard. Sustained division of the attentional spotlight. *Nature*, 424, 2003.

[91] Ulric Neisser. *Cognitive Psychology*. New York: Appleton, Century, Crofts, 1967.

[92] Roger A. Nicoll, Susumu Tomita, and David S. Bredt. Auxiliary subunits assist ampa-type glutamate receptors. *Science*, 311:1253–1256, 2006.

[93] Richard J. O'Brien, Sunjeev Kamboj, Michael D. Ehlers, Kenneth R. Rosen, Gerald D. Fischbach, and Richard L. Huganir. Activity-dependent modulation of synaptic ampa receptor accumulation. *Neuron*, 21(5):1067–1078, 1998.

[94] Daniel H. O'Connor, Gayle M. Wittenberg, and Samuel S.-H. Wang. Dissection of bidirectional synaptic plasticity into saturable unidirectional processes. *Journal of Neurophysiology*, 94:1565–1573, 2005.

[95] Daniel H. O'Connor, Gayle M. Wittenberg, and Samuel S.-H. Wang. Graded bidirectional synaptic plasticity is composed of switch-like unitary events. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9679–9684, 2005.

[96] Carl C.H. Petersen, Robert C. Malenka, Roger A. Nicoll, and John J. Hopfield. All-or-none potentiation at ca3-ca1 synapses. *Proceedings of the National Academy of Sciences of the United States of America*, 95:4732–4737, 1998.

[97] Russell A. Poldrack and Mark G. Packard. Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*, 41:245–251, 2003.

[98] M.I. Posner, C.R.R. Snyder, and B.J. Davidson. Attention and detection of signals. *Journal of Experimental Psychology, General*, 109:160–174, 1980.

[99] Guo qiang Bi and Mu ming Poo. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18:10464–10472, 1998.

[100] A. Redish, A. Elga, and D. Touretzky. A coupled attractor model of the rodent head direction system. *Network: Computation in Neural Systems*, 7:671–685, 1996.

[101] Jonathan Rubin, Daniel D. Lee, and H. Sompolinsky. Equilibrium properties of temporally asymmetric hebbian plasticity. *Physical Review Letters*, 86:364, 2001.

[102] D. E. Rumelhart and J. L. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure.* MIT Press, Cambridge, MA, 1986.

[103] Alexei Samsonovich. *Attractor map theory of the hippocampal representation of space.* PhD thesis, University of Arizona, 1997.

[104] Werner X. Schneider. An introduction to 'mechanisms of visual attention: a cognitive neuroscience perspective'. *Visual Cognition*, 5:1–8, 1998.

[105] Walter Senn. Beyond spike timing: the role of nonlinear plasticity and unreliable synapses. *Biological Cybernetics*, 87:344–355, 2002.

[106] Neel T. Shah, Luk Chong Yeung, Leon N. Cooper, Yidao Cai, and Harel Z. Shouval. A biophysical basis for the inter-spike interaction of spike-timing-dependent plasticity. *Biological Cybernetics*, 95:113–121, 2006.

[107] Stewart Shipp. The brain circuitry of attention. *Trends in Cognitive Sciences*, 8(5):223–230, 2004.

[108] Harel Z. Shouval, Mark F. Bear, and Leon N Cooper. A unified model of nmda receptor-dependent bidirectional synaptic plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10831–10836, 2002.

[109] Per Jesper Sjostrom, Gina G. Turrigiano, and Sacha B. Nelson. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32:1149–1164, 2001.

[110] Sen Song, Kenneth D. Miller, and L. F. Abbott. Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3(9):919–926, 2000.

[111] D.I. Standage, T.P. Trappenberg, and R.M. Klein. A continuous attractor neural network model of divided visual attention. *Proceedings of the International Joint Conference on Neural Networks, July 31 - August 4, 2005, Montreal, Canada*, 2005.

[112] Dominic Standage, Sajiya Jalil, and Thomas Trappenberg. Computational consequences of experimentally derived spike-time and weight dependent plasticity rules. *Biological Cybernetics*, 96:615–623, 2007.

[113] Dominic I. Standage, Thomas P. Trappenberg, and Raymond M. Klein. A continuous attractor neural network model of divided visual attention. In *Proceedings of the International Joint Conference on Neural Networks 2005*, pages 2897–2902, 2005.

[114] Charles F. Stevens and Yanyan Wang. Changes in reliability of synaptic function as a mechanism for plasticity. *Nature*, 371:704–707, 1994.

[115] S.M. Stringer, E.T. Rolls, and T.P. Trappenberg. Self-organizing continuous attractor networks with multiple activity packets and the representation of space. *Neural Networks*, 17:5–27, 2004.

[116] S.M. Stringer, T.P. Trappenberg, E.T. Rolls, and I.E.T. Araujo. Self-organising continuous attractor networks and path integration: One-dimensional models of head direction cells. *Network: Computation in Neural Systems*, 13:217–242, 2002.

[117] Jesper Tegner and Adam Kepecs. An adaptive spike-timing-dependent plasticity rule. *Neurocomputing*, 44-46:189–194, 2002.

[118] Richard F. Thompson and Jeansok J. Kim. Memory systems in the brain and localization of a memory. *Proceedings of the National Academy of Sciences of the United States of America*, 93:13438–13444, 1996.

[119] Thomas P. Trappenberg. *Fundamentals of computational neuroscience.* Oxford University Press, 2002.

[120] Thomas P. Trappenberg and Dominic I. Standage. Multi-packet regions in stabilized continuous attractor networks. *Neurocomputing, in press*, 2005.

[121] T.P. Trappenberg. Why is our capacity of working memory so large? *Neural Information Processing: Letters & Reviews*, 1(3):97–101, 2003.

[122] T.P Trappenberg, M. Dorris, R.M. Klein, and D.P. Munoz. A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience*, 13:256–271, 2001.

[123] Anne Treisman. Feature binding, attention and object perception. *Proceedings of the Royal Society*, 1998.

[124] M. Tsodyks and T. Sejnowski. Associative memory and hippocampal place cells. *Int. Journal of Neural Systems*, 6(supp. 1995):81–86, (1995).

[125] John K. Tsotsos. On the relative complexity of active vs. passive visual search. *International Journal of Computer Vision*, 7:127–141, 1992.

[126] Gina G. Turrigiano, Kenneth R. Leslie, Niraj S. Desai, Lana C. Rutherford, and Sacha B. Nelson. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391(6670):892–896, 1998.

[127] M. Usher, M. Stemmler, C. Koch, and Z. Olami. Network amplification of local fluctuations causes high spike rate variability, fractal firing patterns and oscillatory local-field potentials. *Neural Computation*, 6:795–836, 1994.

[128] M. C. W. van Rossum, G. Q. Bi, and G. G. Turrigiano. Stable hebbian learning from spike timing-dependent plasticity. *Journal of Neuroscience*, 20(23):8812–8821, 2000.

[129] Gene V. Wallenstein and Michael E. Hasselmo. Gabaergic modulation of hippocampal population activity: sequence learning, place field development and the phase precession effect. *Learning and Memory*, 3:264–278, 1996.

[130] Huai-Xing Wang, Richard C. Gerkin, David W. Nauen, and Guo-Qiang Bi. Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nature Neuroscience*, 8(2):187–193, 2005.

[131] H.R. Wilson and J.D. Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13:55–80, 1973.

[132] Jeff Hawkins with Sandra Blakeslee. *On Intelligence*, chapter 6, pages 106–176. Owl Books, 2004.

[133] Gayle M. Wittenberg and Samuel S.-H. Wang. Malleability of spike-timing-dependent plasticity at the ca3ca1 synapse. *The Journal of Neuroscience*, 26(24):6610–6617, 2006.

[134] J.M. Wolfe. Guided search 2.0: a revised model of visual search. *Psychonomic Bulletin and Review*, 1(2):202–238, 1994.

[135] Christina Zelano, Moustafa Bensafi, Jess Porter, Joel Mainland, Brad Johnson, Elizabeth Bremner, Christina Telles, Rehan Khan, and Noam Sobel. Attentional modulation in human primary olfactory cortex. *Nature*, 8:114–120, 2004.

[136] Kechen Zhang. Representation of spatial orientation by the intrinsic dynamics of head-direction cell ensembles: a theory. *Journal of Neuroscience*, 16(4):2112–2126, 1996.