

Biological learning mechanisms in spiking neuronal networks

Matthieu Gilson

Submitted in total fulfilment of the requirements of the degree of
Doctor of Philosophy

Department of Electrical and Electronic Engineering
THE UNIVERSITY OF MELBOURNE

June 2009

Copyright © 2009 Matthieu Gilson

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the author.

Abstract

IN THE present thesis biological learning mechanisms in spiking neuronal networks are investigated. In the brain, mechanisms at the molecular level modify the strengths (or weights) of the connections (synapses) between neurons, which is hypothesised to develop structure in neuronal networks depending on their activity; this learning mechanism is called 'synaptic plasticity'. This thesis focuses on a particular physiological model of synaptic plasticity: spike-timing-dependent plasticity (STDP). Using a stochastic model of the spiking neuron, the Poisson neuron, a dynamical system is derived to predict the evolution of the weights and thus of the network activity using mathematical tools from stochastic processes and dynamical systems. The main aim of the study is to gain a better understanding of the weight dynamics induced by such synaptic plasticity in recurrent neuronal networks. The emergence of weight structure in a neuronal network is determined by the interplay between the main players: neuronal mechanisms, network connectivity, stimulating input structure and learning parameters. For a broad range of parameters, STDP can generate at the same time both a stabilisation of the mean incoming weight for each neuron, synonymous to stability of the firing rates in the network, and a diverging behaviour that induces neuronal specialisation to some of its incoming connections (both for input and recurrent weights). The results presented can be linked to cortical self-organisation, for example; STDP can lead to the emergence of neuronal groups sensitive to distinct input pathways. The resulting input selectivity provides a framework for ocular dominance in the primary visual cortex. The study of the learning dynamics also contributes to obtaining insight into the neuronal information processing that occurs in the brain: it indicates a time scale at which variations of the neuronal spiking probabilities are of importance, stresses the importance of the complete

cross-correlation structure between pairs of neurons (not just coincident spiking) and supports the hypothesis of massively distributed computation through the role of neuron assemblies in driving the weight dynamics. This study can also be seen as a further step towards linking neuromodelling at the physiological level to machine learning.

Declaration

This is to certify that

1. the thesis comprises only my original work towards the PhD,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Matthieu Gilson, June 2009

Acknowledgements

I would like to express my thanks to my supervisors Anthony Burkitt, David Grayden, Doreen Thomas and Konstantin Borovkov for their patient encouragements, their support and the time they spent on my PhD project. I am greatly indebted to Prof. van Hemmen for his constructive criticisms within the context of our collaboration and the enjoyable stays at the Physik Department (T35) of the Technische Universität München. I am also grateful to the academic staff from the Department of Electrical and Electronic Engineering (EEE) and Department of Mathematics and Statistics who helped me to get started, in particular Iven Mareels, Dragan Nestic, Marty Ross, Aihua Xia, Jerry Koliha and Barry Hughes. My research activities were made possible thanks to financial support at the University of Melbourne, NICTA Victoria Research Lab, The Bionic Ear Institute and the School of Engineering from the University of Melbourne; in particular, I am grateful to Subhash Challa and Rob Shepherd for giving me the opportunity to present my work overseas. I would also like to thank Tracy Painter, Carmen Doyle, Natasha Baxter, Annette McLeod, David Strover and Jackie Brissonette for their help in organising conference travels and student daily-life issues.

Special thanks to my labmates and colleagues from The Bionic Ear Institute (BEI), EEE, S^T Vincent Hospital Melbourne and T35 for bearing with the wine-and-cheese sessions, soccer and other distractions (in random order): Michael Eager, Chris Trengove, Sean Byrnes, Daniel Taft, Hamish Meffin, Andrew Vandali, Andrew Purnama, Stas Surowiecki, Mark Harrison, Jeremy Marozeau, Andrea Varsavsky, Elma O’Sullivan-Greene, Dean Freestone, Andre Peterson (and his passion for Lipschitz), Colin Hales, Kelvin (Unit) Layton, Nick Opie, Emily O’Brien, Jennifer Che, Bahman (the Avalanche) Tahayori, Joel Villalobos, Rahil Garnavi, Tuyet Thi Anh Vu, Shafiq Massan, Byron Wicks, Levin Kuhlmann,

Thomas Hanselmann, Chien Minh Ta, Julien Ridoux, Alan Lai, Tim Nelson, Amy Halliday, Paul Friedel, Christine Vossen, Julie Goulet, Jan Diepenbrock, Andreas Sicher, Andreas Vollmayr, Moritz Bürck, Moritz Franosch and Yimin Nie.

My stay in Australia was also enjoyable thanks to my cousins Pascale et Jean-Marie Guitera, et les filles (Elma et Jade), and the friends I made there: Markus Bischofberger, Belinda van Straaten, Douggie Bell, Kelly van Burskick, Anna Henke, Zoe Riddoch, Ruth Emsden, Patrice Matrin de Barros, David, Stéphane Pettier, My Nguyen, the band of Deco; as well as a couple of more-or-less philanthropic organisations: Milawa Cheese Factory, Victoria Market, Revolver Upstairs, Cat Empire, Deco, My Corazon, Babka, Northcote Social Club, Corner Hotel, Great Britain Hotel, Batifolles.

À Aline et Jean Gilson, et Andrée et Maurice Guillaume.

À mes parents et ma famille.

À Ngà Thị Phạm.

À mes ami(e)s.

À mes professeur(e)s.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Neurons	1
1.2.1	Anatomy and functioning of a neuron	2
1.2.2	Modelling neuronal activity	3
1.2.3	Neurons as point processes	5
1.2.4	Connecting neurons	5
1.2.5	Neuronal information processing	6
1.3	Learning in neurons	6
1.3.1	Experimental evidence of synaptic plasticity	7
1.3.2	Spike-timing-dependent plasticity (STDP)	8
1.3.3	Weight dynamics induced by STDP	10
1.3.4	Functional implications of synaptic plasticity and STDP	13
1.4	Context and focus of this PhD project	14
1.5	Plan	16
2	Modelling the brain at the neuronal level	17
2.1	Introduction	17
2.2	Modelling neuronal activity in networks	17
2.2.1	Poisson neuron model	17
2.2.2	Relation to more elaborate neuron models	19
2.2.3	Model of recurrent neuronal network	20
2.3	Spike-timing-dependent plasticity (STDP)	20
2.3.1	STDP for pairwise interactions between pre- and post-synaptic spikes	20
2.3.2	Relation to rate-based models	24
3	Dynamical system to model network activity and synaptic plasticity	27
3.1	Overview	27
3.2	Description of the network	28
3.2.1	Definition of the state variables for the network	29
3.3	Slow weight evolution	31
3.4	Derivation of the network consistency equations	34
3.4.1	Short duration of the PSP kernel and of the recurrent delays	34
3.4.2	The equations describing the dynamical system	37
3.4.3	Higher-order stochastic effects of the weight dynamics	38

3.5	Generation of the input spike trains	39
3.6	Analysis of the system dynamics	41
4	Input selectivity by STDP	43
4.1	Introduction	43
4.2	General case of learning input weights with fixed recurrent weights	45
4.2.1	Homeostatic equilibrium	45
4.2.2	Emergence of a weight structure	47
4.3	Network stimulated by two homogeneous input pools	49
4.3.1	Reduction of dimensionality to study the weight drift	50
4.3.2	Firing-rate equilibrium for weak correlations	52
4.3.3	Two uncorrelated input pools with any firing rates	53
4.3.4	The two input pools have correlations and the same input firing rate	54
4.3.5	Distinct firing rates for the two correlated input pools	55
4.3.6	Extension to several homogeneous input pools	58
4.4	Symmetry breaking of the distribution of input weights with fixed recur- rent weights	58
4.4.1	Previous results concerning the weight drift	58
4.4.2	Impact of fixed recurrent connections	61
4.4.3	Non-homogeneous fixed recurrent connections	64
4.4.4	Dependence upon neuron model, initial conditions and learning parameters	65
4.5	Partial conclusion on plastic input connections	66
5	Plastic recurrent connections	71
5.1	Introduction	71
5.2	Equilibrium in a partially connected recurrent network with no external inputs	72
5.2.1	Fixed point of the firing rates	73
5.2.2	Fixed points of the weights	74
5.2.3	Stability analysis	75
5.3	Diffusion of the recurrent weights	77
5.3.1	Dispersion of the individual weights	78
5.3.2	Asymptotic pattern of recurrent weights	80
5.4	General case of learning on the recurrent weights with fixed input weights	82
5.4.1	Homeostatic equilibrium	83
5.4.2	Learning the input correlation structure	84
5.4.3	Sufficient condition for existence of fixed points	85
5.4.4	Weight dynamics for arbitrary matrix \tilde{C}_\perp	86
5.5	Network with two distinct input pathways	87
5.5.1	One input pool with spike-time correlation and one uncorrelated pool	90
5.5.2	Two input pools with balanced spike-time correlations	93
5.6	Partial conclusion on plastic recurrent connections	96

6	Self-organisation	101
6.1	Introduction	101
6.2	Effect of the weight dependence of STDP upon the learning dynamics . .	103
6.2.1	Homeostatic equilibrium	103
6.2.2	Weight specialisation	107
6.3	Representation of the input correlation in the weight structure for a single neuron	109
6.3.1	Structure of the input spike trains	109
6.3.2	Capturing the weight dynamics	111
6.3.3	Initial splitting of the input weights	112
6.3.4	Saturation of the weights for weight-dependent STDP	115
6.3.5	Generalisation to arbitrary input structure	116
7	Stability of neuronal activity in recurrent networks	119
7.1	Introduction	119
7.1.1	Non-linear Poisson neuron	119
7.1.2	Network model	121
7.1.3	Description of the network activity	122
7.2	Network dynamics	123
7.2.1	Evolution of $X(\cdot)$	123
7.2.2	Markov property	125
7.2.3	Formalism of piecewise deterministic Markov process	125
7.2.4	Description of the generator	129
7.3	Stability in the network	129
7.3.1	Ergodicity	130
7.3.2	Stationary distribution	131
7.4	Remarks on the framework presented in this chapter	136
8	Conclusion	137
8.1	Summary of original contributions and results	137
8.1.1	Theoretical framework to study learning dynamics	137
8.1.2	Weight specialisation in recurrent networks	138
8.1.3	Weight-dependence for STDP	141
8.1.4	Self-organisation in visual cortex	141
8.2	Implications for neuronal information processing	143
8.3	Future research directions	144
A	Calculations for chapter 3	147
A.1	Remarks on the input covariance structure	147
A.1.1	Definition of the external input covariance	147
A.1.2	Properties of the matrix \hat{C}^W	148
A.2	Neuron-to-input covariance consistency equation	148
A.2.1	Evaluation of the covariance using the past spiking history	149
A.2.2	Spike-triggering effect	149
A.2.3	Time-averaging	151
A.2.4	Use of the Fourier transform	152

A.2.5	Sharp distribution of delays	152
A.2.6	Impact of synaptic mechanisms on the covariance structure	153
A.2.7	Short-duration PSPs and short recurrent delays	154
A.2.8	Long recurrent delays	155
A.3	Neuron-to-neuron covariance consistency equations	155
A.3.1	Taking the autocorrelation into account	156
A.3.2	Remark on the autocorrelation structure due to the recurrent connections	159
A.3.3	Short recurrent delays	160
B	Calculations for chapter 4	163
B.1	Analysis of the drift of K due to STDP with fixed J	163
B.1.1	Symmetries of the inputs and reduction of dimensionality for K	163
B.1.2	General evolution for full input connectivity	164
B.1.3	Partial input connectivity	166
B.2	Dependence of the fixed point $K(\infty)\hat{\mathbf{h}}$ upon input correlation	167
B.3	Symmetry breaking within K for different neurons	168
B.3.1	Second moment of the stochastic evolution of K	168
B.3.2	Recurrent connections and spike-triggering effect	169
B.3.3	Arbitrary homogeneous connectivity	171
B.4	Symmetry breaking by competition between input weights	172
C	Calculations for chapter 5	173
C.1	Invertibility of $[\mathbb{1}_N - J(t)]$	173
C.2	Equilibrium induced by STDP	175
C.2.1	Fixed point of the firing rates in the presence of recurrent loops	175
C.2.2	Stability of the manifold of fixed points	175
C.2.3	Decomposition of \mathcal{L}	176
C.2.4	Homogeneous connectivity topology	177
C.3	Second order of the stochastic evolution of the weights	178
C.3.1	Analysis of the matrix $\Gamma(t, t')$	178
C.3.2	Autocorrelation effects on weight dispersion	179
C.3.3	Weight evolution for a synaptic loop $j \rightarrow i \rightarrow j$	183
C.4	Dependence of the asymptotic weight distribution on initial conditions	183
D	Simulation parameters	185

List of Figures

1.1	Drawing by Ramon Y Cajal	2
1.2	Anatomy of a neuron	3
1.3	Basic input-output functioning of a neuron	4
1.4	Increase of synaptic weight due to high-rate pre-synaptic stimulation	8
1.5	Experimental data of weight change as a function of the time difference	9
1.6	Diagrammatic representation of the interactions between pre- and post-synaptic spikes	11
2.1	Poisson neuron model	19
2.2	Example of STDP window function with weight dependence	23
3.1	Presentation of the network and notation	29
3.2	Impact of the PSP kernel ϵ on learning with the STDP window function W	36
3.3	Plots of ϵ and ζ	37
4.1	Network configurations studied in chapter 4	44
4.2	Dimension reduction for weight matrices	51
4.3	Comparison between the weight evolution of different initial conditions	55
4.4	Asymptotic weight specialisation dependence upon the difference between the input correlation strengths	56
4.5	Weight evolution for unbalanced correlations	59
4.6	Symmetry breaking of the input weights for a group of $N = 60$ neurons	63
4.7	Symmetry breaking of the input weights for two neuron groups	64
4.8	Illustration of the specialisation of two neuron groups as a function of the coupling between them	65
4.9	Schematic representation of the input weight distribution before and after learning	69
5.1	Network configurations studied in chapter 5	72
5.2	Illustration of spectra for the linear operator \mathcal{L}	77
5.3	Comparison of the evolution of the weight variance between full connectivity and partial random connectivity	79
5.4	Evolution of the distributions of incoming and outgoing weights for $N = 100$ neurons	81
5.5	Illustrative results of numerical simulations with $N = 30$ neurons	82
5.6	The recurrently connected neurons are divided into two groups.	88

5.7	Evolution of the neuron firing rates	90
5.8	Strengthening of the outgoing weights of the neuron group that receives correlated input	92
5.9	Within-group strengthening of the recurrent connections due to stimulation by correlated inputs	95
5.10	Example of a different STDP window function	96
5.11	Cross-correlogram for two neurons	96
5.12	Schematic representation of the recurrent weight specialisation before and after learning	99
6.1	Example of weight-dependent STDP window function	102
6.2	Influence of weight-dependent STDP upon the homeostatic equilibrium	106
6.3	Simultaneous evolution of the input and recurrent weights	110
6.4	Schematic representation of one neuron stimulated by m pools	111
6.5	Distribution of the roots of \mathbf{Q}	115
6.6	Evolution of the weight structure for the single neuron. Comparison between two degrees of weight dependence: (a & c) $\gamma = 0.1$ and (b & d) $\gamma = 0.02$. The neuron is stimulated by $m = 5$ pools of 30 inputs with the same firing rate $\hat{\nu}_0 = 10$ Hz and correlation levels $\hat{c}_l = 0, 0.05, 0.1, 0.15$ and 0.2 , respectively (cf. Sec. 6.3.1). (a & b) Evolution of the weights J_k (grey bundle) over 10^4 s. The thick black line represents the mean weight. (c & d) Asymptotic distribution of the weights J_k (+). The thick grey lines represent the means over each pool.	117
6.7	Evolution of the weight structure of a neuron stimulated by $m = 3$ input pools, as described in Sec. 6.3.5. (a) Evolution of the weights J_k (grey bundle) over 10^4 s. (b) Asymptotic distribution of the weights J_k (+). The plot formats and the parameters ($\gamma = 0.02$) are similar to Fig. 6.6.	118
7.1	A typical choice of neuronal activation function	120
7.2	Schematic representation of two of the network	121
7.3	Example of evolution of the process for the time point of one single source	127
7.4	Illustration of the intervals considered when evaluating the transitions between simplices.	132
8.1	Self-organisation scheme	142
8.2	Ocular dominance columns in macaque monkey	143
B.1	Example of evolution of K in two dimensions.	165
C.1	Illustration of the impact of the activation function upon the weight constraint	174

Chapter 1

Introduction

1.1 Motivation

THE brain performs many sophisticated everyday computational tasks with a speed and precision unparalleled by present-day computers. Usual examples include object recognition and target tracking by the visual pathway, sound localisation and speech recognition by the auditory pathway, and movement control in the motor cortex. The difference between the brain and computers does not lie so much in brute-force computational power, but rather in the way the processing is organised: the brain makes use of a massively parallel architecture to perform computations in a distributed fashion. How the brain achieves such a performance is of interest both in gaining a better understanding of the brain and in designing artificial applications inspired by it.

1.2 Neurons

At the end of the nineteenth century, Ramon y Cajal and Golgi uncovered the role of neurons in the information processing performed by the brain. Their observations of brain tissue by means of microscopes shed light on the structure of neuronal networks, as illustrated in Fig. 1.1, and on the way in which neurons communicate through “connections” between them, the synapses. The human brain comprises more than 10^{11} neurons organised in various structures, from microscopic circuits to macroscopic functional areas (Hubel and Wiesel 1962, Bear et al. 2007). Understanding its functioning is arguably one of the most challenging problems in the domain of complex systems.

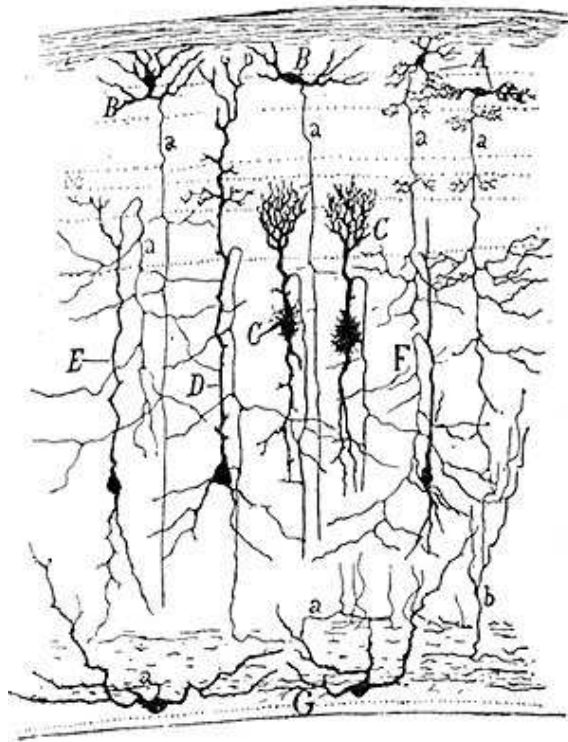


Figure 1.1: Preparation through the optic tectum (from a sparrow) impregnated with the Golgi technique. Drawing by Ramon Y Cajal (taken in May 2009 from [url nobel-prize.org/nobel.prizes/medicine/articles/cajal/images/9.gif](http://url.nobel-prize.org/nobel.prizes/medicine/articles/cajal/images/9.gif)). This drawing shows neurons of distinct types (indicated by the capital letters), and the 'a' indicates an axon.

1.2.1 Anatomy and functioning of a neuron

Neurons are believed to communicate by means of spatial and temporal variations of their membrane potential. The anatomy of a neuron involves three distinct parts with different functions with respect to the neuronal electrical activity (see Fig. 1.2): dendrites that form a tree and contain post-synaptic receptors (inputs), the cell body or soma that integrates the synaptic influx coming from the dendrites, and a long-limbed axon that terminates with pre-synaptic buttons (outputs). One typical feature of the neuronal electrical activity is the propagation of membrane depolarisation. Brief high-amplitude depolarisations that propagate from the soma along the axon are called action potentials (or spikes) and have a characteristic shape. When a spike reaches an axonal termination that “connects” to a post-synaptic neuron, neurotransmitters are released into the extra-

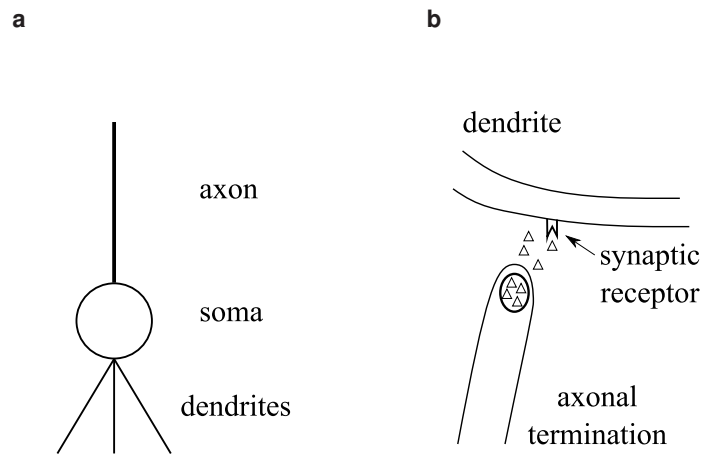


Figure 1.2: Anatomy of a neuron. (a) Schematic representation featuring the dendritic tree (inputs), the cell body (soma) and the long-limbed axon (output). (b) Detail of a synapse. The axonal termination contain vesicles (thick closed curve) of neurotransmitters (triangles) that are released when the synapse is excited by an incoming axon potential (not represented). The neurotransmitters activate synaptic receptors located on the dendrite of the post-synaptic neuron, which modifies the local membrane potential.

cellular space and excite receptors on the post-synaptic neuron (usually on dendrites). This generates a local variation of the membrane potential, which propagates towards the soma. The soma can be seen as a spatio-temporal integrator of these post-synaptic potentials (PSP) to generate an output spike, as illustrated in Fig. 1.3. The soma potential often remains set to a resting value for a few milliseconds after firing an action potential, which is referred to as the refractory period. These basic elements of the neuronal information processing actually depend upon many different mechanisms at the molecular level, such as ionic concentrations, density of ion channels, axonal myelination, and types of neurotransmitters; for a review, see Bear et al. (2007).

1.2.2 Modelling neuronal activity

The soma potential $V(t)$ for a neuron evolves over time due to leakage and the pre-synaptic influx, which is usually modelled by the following differential equation,

$$C_m \frac{\partial V(t)}{\partial t} = \frac{C_m}{\tau_m} [V_{\text{rest}} - V(t)] + I_K(t) + I_{Na}(t) + I_{Ca}(t) + I_{\text{syn}}(t), \quad (1.1)$$

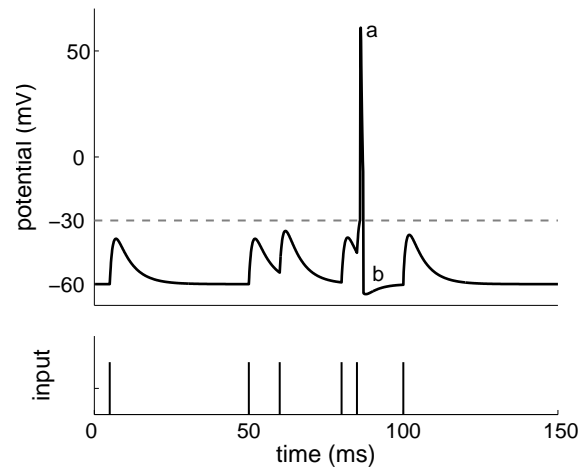


Figure 1.3: Basic input-output functioning of a neuron. Evolution of the soma potential when the neuron is stimulated by pre-synaptic spikes (input). In the absence of stimulation, the soma potential rests around -60 mV. Each pre-synaptic spike generates a partial depolarisation of the local post-synaptic membrane (e.g., on the dendrite), which propagates to the soma. When the soma potential exceeds a threshold (grey dashed line, chosen to be at -30 mV here) due to post-synaptic response following the two spikes at 80 and 85 ms, it triggers a brief high-amplitude depolarisation of the membrane (a: action potential) followed by a hyperpolarisation (b) during which the neuron is inhibited (refractory period). Then the neuron is excitable again.

where C_m is the membrane capacitance, τ_m is the passive membrane time constant due to the leakage, and V_{rest} is the resting potential. The current I_{syn} is related to incoming PSPs that propagate from the dendrites to the soma. The ionic currents I_K , I_{Na} and I_{Ca} generate an active response of the membrane to changes of its potential: in particular, they are involved in the generation and propagation of action potentials for sufficiently strong synaptic influx I_{syn} . Such details can be incorporated in a model such as that proposed by Hodgkin and Huxley (1952) to quantitatively reproduce experimental data. This model is often used with a more or less simplified compartmental geometry using numerical simulation.

However, such a modelling approach is rather untractable from a mathematical point of view and many phenomenological models of neurons have been proposed to describe the input-output functioning of neurons or groups of neurons. This includes, in particular, the integrate-and-fire (IF) neuron (Lapicque 1907, Gerstein and Mandelbrot 1964, Stein 1965, Lansky 1984, Burkitt 2006), binary neuron (McCulloch and Pitts 1943), the Poisson neuron (Kempster et al. 1998, Kempster et al. 1999) and neuronal continuous fields

(Beurle 1956, Wilson and Cowan 1972, Amari 1977, Coombes 2005). One crucial issue lies in determining at which scale to position a problem, since neuronal activity can be determined by an interplay between, for example, specific mechanisms at the molecular level (dopamine, calcium), intrinsic neuronal activation properties (bursting, irregular), and the network topology (excitatory vs. inhibitory neurons, feedback strength).

1.2.3 Neurons as point processes

One way to tackle the complexity of the mechanisms involved in the spike generation lies in separating the evolution of the soma potential during an action potential from that in the “sub-threshold” regime, i.e., between the firing of spikes. Action potentials are thus considered to be instantaneous. This idea has been used since Lapicque (1907) and is motivated by the standard shape of action potentials and their short duration (order of 1 ms) compared to other neuronal mechanisms (post-synaptic responses, leakage of the membrane potential, etc.).

In this way, a neuron can be modelled using a point process that describes the probability of firing an action potential (instantaneous event) depending upon the past activity of the neuron, its incoming synaptic stimulation or other mechanisms. This concept made it possible to study analytically the neuronal response to various input stimulations; for a review, see Gerstner and Kistler (2002) and Burkitt (2006). Spiking neurons, such as IF neurons modified with a suitable intrinsic dynamic variable, proved to be able to reproduce the variety of spiking dynamics exhibited by Hodgkin-Huxley neurons (Izhikevich 2003), which also supports this choice.

1.2.4 Connecting neurons

This project aims to study neuronal activity in networks. The cortical connectivity, on which we focus, is extremely complex and involves many types of neurons. A usual simplification consists in reducing the cortical neurons to two populations of “excitatory” (pyramidal) neurons and “inhibitory” (stellate) interneurons, with a different range for the connections between these two populations, an overall density of local connections

equal to 5-10% and roughly balanced excitation vs. inhibition for the incoming synaptic influx to each neuron. We will mostly consider only excitatory synapses, which is equivalent to assuming identical inhibition for all neurons in the network. The connection density will range from 10% to full connectivity, in order to study the effect of the connectivity density. We target general behaviours related to the spiking neuronal activity that do not depend upon the size of the network considered, but rather upon its structure.

1.2.5 Neuronal information processing

The brief duration of action potentials suggests that only their timing carries information. Using point-process modelling, it is necessary to investigate the mechanisms leading to the generation of spikes in the sub-threshold regime in order to understand how neurons process information. Within a network, neurons process information in a distributed fashion. Spiking activity at the scale of networks is still only partially understood, even though progress in the theory has recently been made in order to relate the neuronal level to the network level (Brunel and Hakim 1999, Brunel 2000, Meffin et al. 2004, Kriener et al. 2008). When observing neuronal activity, it is not clear what features of the raster plot (temporal distribution) of spikes represent information. In other words, how does one analyse spike trains: is information carried in the firing rates, spike-time correlations or other spatio-temporal features (Delorme et al. 2001, Rieke 1997)? One critical problem for neuromodelling lies in the variability of spike trains, even when evoked by the same stimulus (Shadlen and Newsome 1998). This variability discriminates between types of neurons: certain of them do exhibit reliable output spike times for some stimulation protocols (Gutkin et al. 2003, Ermentrout et al. 2008).

1.3 Learning in neurons

One particular aspect of the neuronal mechanisms that occurs at the molecular level concerns changes in the weights (or strengths) of synaptic connections, which is related to the amplitude of the PSP induced by each pre-synaptic action potential (incoming spike).

'Synaptic plasticity' describes the strengthening (potentiation) or the weakening (depression) of the synaptic weight, which depends in particular upon the neuronal spiking activity. The study of these mechanisms has become a prominent area in neuroscience (Malenka and Siegelbaum 2001, Martin et al. 2000).

1.3.1 Experimental evidence of synaptic plasticity

The usual method to investigate the change in synaptic weight *in vivo* or *in vitro* consists in finding two neurons connected by a synapse and recording the evolution of the PSP amplitude at the target neuron for a given stimulation protocol on the source neuron. Fig. 1.4(A-B) shows the change in the PSP due to pre-synaptic stimulation at 15 Hz: the experimental pathway (Exp) exhibits an increase of the depolarisation while it does not vary for the ipsilateral control pathway (Cont). Such changes can last over time and are thus denoted as long-term potentiation (LTP), as illustrated in Fig. 1.4(C) for brief stimulations separated by tens of minutes. This long-term plasticity should not be confused with short-term plasticity that occurs at time scales below seconds (Morrison et al. 2007). Conversely, long-term depression (LTD) corresponds to a decrease of the synaptic weight (Artola et al. 1990, Hirsch et al. 1992). The stimulation protocol is crucial to determine the polarity of the weight change: traditionally, LTP has been obtained for high input rates while low rates induce LTD (Bliss and Lømo 1973). However, pairing low-frequency stimulation with depolarisation of the post-synaptic membrane leads to LTP. The temporal order of a weak and a strong stimulating inputs also determines whether LTP or LTD is induced (Levy and Steward 1983).

Experiments using glutamatergic (excitatory) synapses highlighted out the role of N-methyl-D-aspartate (NMDA) receptors that are believed to act as coincidence detectors: the pre-synaptic activation triggers a release of glutamate and the depolarisation of the post-synaptic membrane lifts the blocking effect of Mg^{2+} ions (Mayer et al. 1984, Nowak et al. 1984). When these two phenomena overlap in time, calcium ions (Ca^{2+}) can flow in through the NMDA receptors and are the principal trigger for induction of LTP vs. LTD (Neveu and Zucker 1996, Yasuda and Tsumoto 1996). The concentration level in Ca^{2+} determines the activation of protein kinases and/or phosphatases, the predominance of the

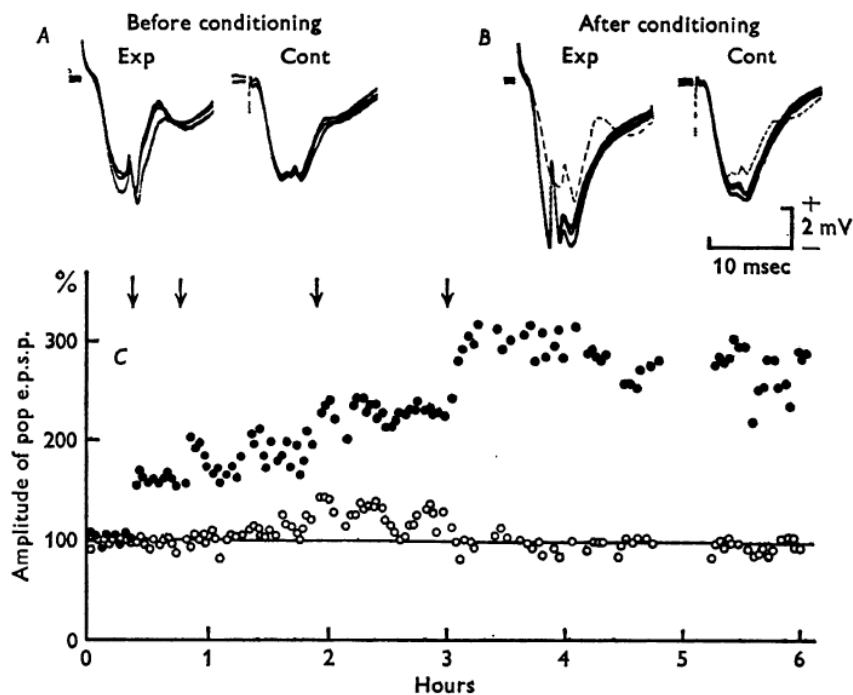


Figure 1.4: Increase of synaptic weight due to high-rate pre-synaptic stimulation; taken from Bliss and Lømo (1973). Three traces of the evoked PSP (A) before and (B) two and a half hours after conditioning (15 spikes per second for 10 s) for the experimental pathway (exp) and the ipsilateral control pathway (cont). (C) Increase of the PSP amplitude over time for the experimental pathway (filled circles) and the control pathway (open circles) in the case of consecutive conditioning stimulations (arrows).

first ones leading to LTP while LTP is induced when the second ones dominates (Lisman 1989, Colbran 2004).

1.3.2 Spike-timing-dependent plasticity (STDP)

Gerstner et al. (1996) first proposed a model for synaptic plasticity relying on the precise timing between pre- and post-synaptic spikes as a mechanism to train neurons to perform sound localisation and explain the performances recorded in the laminar nucleus in the auditory system of the barn owl. This spike-based learning rule, in contrast to rate-based rules broadly used before, allows weight dynamics to make use of spike-time information at time scale below the millisecond.

This seminal theoretical work has generated a vast interest and subsequent experi-

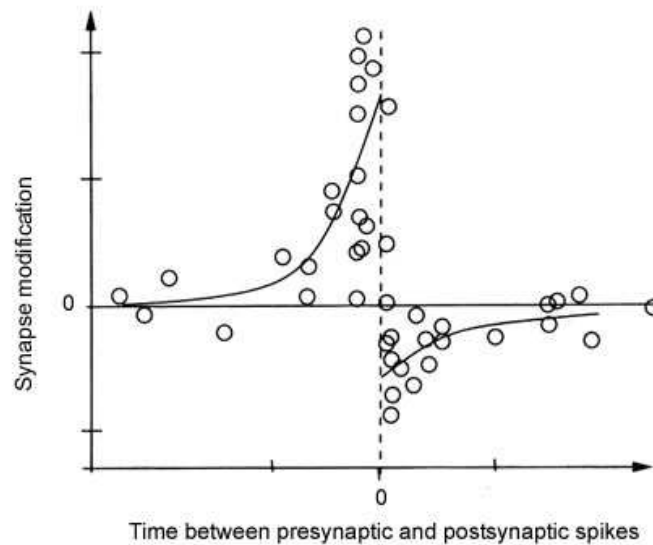


Figure 1.5: Experimental data of weight change as a function of the time difference between paired pre- and post-synaptic action potentials, fitted by two curves (solid line) for each side; data from hippocampal glutamatergic neurons in culture (Bi and Poo 1998).

mental studies have established the importance of the timing of paired pre- and post-synaptic spikes in the weight change (Markram et al. 1997, Bell et al. 1997, Magee and Johnston 1997, Bi and Poo 1998, Debanne et al. 1998, Egger et al. 1999, Feldman 2000, Bi and Poo 2001, Boettiger and Doupe 2001, Sjöström et al. 2001, Froemke and Dan 2002, Tzounopoulos et al. 2004). Most of these “early” experiment studies used short-range excitatory connections between neurons, for example, in cortical and hippocampal slices of rat, mouse, fish and tadpole. In agreement with the Hebbian postulate (Hebb 1949), pre-synaptic spikes that take part in the firing of an output spike induce an increase of the synaptic strength (potentiation); when the time order is reversed, the weight is decreased (depression). The observed weight change also fades away when the time difference between the spikes becomes larger, as illustrated in Fig. 1.5. STDP has also been observed for synapses from an excitatory to an inhibitory neurons (Bell et al. 1997, Han et al. 2000, Tzounopoulos et al. 2004, Tzounopoulos et al. 2007) and between inhibitory neurons (Holmgren and Zilberter 2001, Woodin et al. 2003). However, we will focus in this thesis on STDP between excitatory neurons.

For glutamatergic (excitatory) synapses, the main agents involved in STDP are the same as “conventional” LTP/LTD, but the question whether STDP relies on the underly-

ing same mechanisms is still controversial (Dan and Poo 2006, Caporale and Dan 2008). Indeed, STDP requires activation of NMDA receptors and elevation of the Ca^{2+} concentration at the post-synaptic site (Magee and Johnston 1997, Markram et al. 1997, Debanne et al. 1998, Bi and Poo 1998, Sjöström et al. 2001). Calcium ions are involved in several cycles with different kinetics that are used as signals to generate STDP (Magee and Johnston 1997, Sabatini et al. 2002, Rubin et al. 2005). The fact that post-synaptic NMDA receptors behave as coincidence detectors for short time scales due to fast components of calcium ions gives an explanation for the potentiation side of STDP; the depression side of STDP requires a more elaborate mechanism, such as two coincidence detectors; see Dan and Poo (2006) for a review. Some experimental studies also showed the role of pre-synaptic mechanisms such as NMDA autoreceptors in STDP (Duguid and Sjöström 2006). Integration of the synaptic inputs, for example inhibitory, in the dendritic tree affects the membrane depolarisation and thus STDP (Liu et al. 2005). The detailed shape of the STDP learning window function, cf. Fig. 1.5, was also demonstrated to depend upon the location of the synapse on the dendrite and the type of pre- and post-synaptic neurons (Froemke et al. 2005). Finally, the susceptibility of the synaptic weight to change can vary according to the current value of the weight (Bi and Poo 1998, Wang et al. 2005) or even its prior history on a longer period (Montgomery and Madison 2002). For reviews about the relationship between STDP and molecular mechanisms, see Dan and Poo (2006) and Caporale and Dan (2008).

1.3.3 Weight dynamics induced by STDP

On the modelling side, much effort has refined the STDP model initially proposed by Gerstner et al. (1996) in order to incorporate physiological mechanisms. A specific focus has been on the link between the molecular level and the phenomenological spike-timing dependence, using Ca^{2+} channels or other signaling chains for example (Senn et al. 2001, Hartley et al. 2006, Badoual et al. 2006, Benuskova and Abraham 2007, Graupner and Brunel 2007, Zou and Destexhe 2007).

Much interest has also focused on how spikes contribute to the weight change. When incorporating the effects of spike triplets or higher-order interactions, new regimes in the

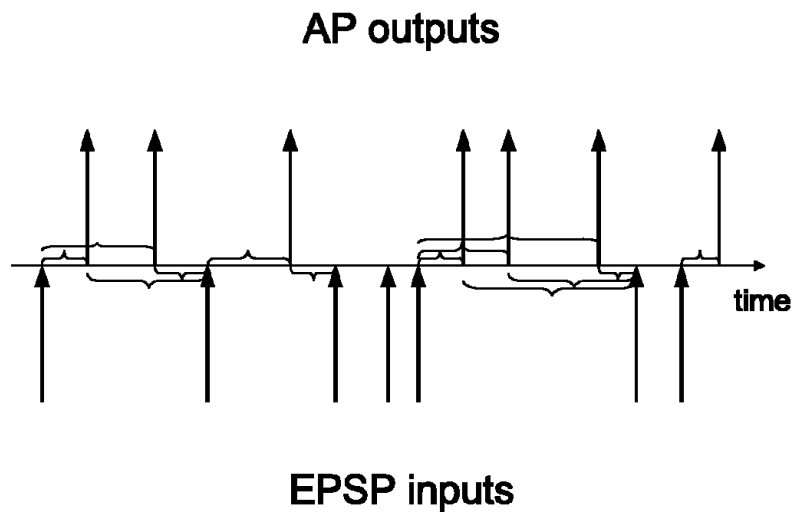


Figure 1.6: Diagrammatic representation of the interactions between pre- and post-synaptic spikes. Time goes from left to right. Each post-synaptic spike (action potential, AP) interacts only with the last and the next pre-synaptic spike (excitatory PSP). The APs are represented by vertical arrows above the time line, and PSPs are represented by vertical arrows below the time line. The brackets indicate the PSP-AP interactions included in this model: brackets above the time line represent contributions that increase the weight (potentiation), and brackets below the time line represent contributions that decrease the weight (depression). Taken from Burkitt et al. (2004).

weight dynamics appear, such as modulations of the potentiation/depression scheme for high firing rates (Sjöström et al. 2001, Kempter et al. 2001, Izhikevich and Desai 2003, Pfister and Gerstner 2006). A probabilistic interpretation of STDP also proved capable of exhibiting richer weight specialisation schemes than the original model (Appleby and Elliott 2005, Appleby and Elliott 2006, Appleby and Elliott 2007, Elliott 2008). The selection of only a portion of interactions to take into account for STDP amongst all possible pairs of pre- and post-synaptic spikes was also demonstrated to affect the weight dynamics. This implies in particular that STDP depends in a complex fashion upon the output firing rate of the neurons. Some experimental results were in favour of a restriction to the pairing of the pre- and post-synaptic spikes interaction when the input excitatory PSPs fall within more than one STDP time window of output spikes (Sjöström et al. 2001), as illustrated in Fig. 1.6. Theoretical studies investigated the influence of such spike pairing schemes upon the weight dynamics (Izhikevich and Desai 2003, Burkitt et al. 2004, Morrison et al. 2007).

The relationship between the weight dynamics induced by STDP and the input stimulation has been the particular focus of many theoretical studies (Gerstner et al. 1996, Kempster et al. 1999, Song et al. 2000, Song and Abbott 2001, Senn 2002, Gütig et al. 2003, Burkitt et al. 2004, Wensch et al. 2005, Meffin et al. 2006, Appleby and Elliott 2006, Morrison et al. 2007, Lubenov and Siapas 2008, Câteau et al. 2008, Kang et al. 2008). To obtain analytical results, a usual simplification consists in using a phenomenological model of STDP (Morrison et al. 2008) that describes the mean effect of the molecular mechanisms represented by the fitting solid curve in Fig. 1.5. Such models focus on the relationship between the weight dynamics and input spike-time correlations at a short time scale, for example, coincidentally spiking sources (Kempster et al. 1999, Song and Abbott 2001, Gütig et al. 2003, Meffin et al. 2006). In the presence of input correlations, the role of dendritic and axonal delays in sharpening or spreading synchronous activity has been pointed out (Senn 2002). However, the implications of STDP *in vivo* are not yet clear, even if some progress has been made to understand the weight dynamics corresponding to natural spike trains (Froemke and Dan 2002).

The weight dependence of the STDP rule has strong implications upon the learning dynamics. Many quantitative models have been proposed to explore the effect of this specific non-linearity in STDP (van Rossum et al. 2000, Morrison et al. 2007, Standage et al. 2007). In particular, strong weight dependence favors unimodal weight distributions whereas additive-like STDP tends to split weights into bimodal distributions (Kempster et al. 1999, Gütig et al. 2003). It is not yet clear which type of stabilisation for individual synaptic weights is more physiologically realistic.

As for network topology, the weight dynamics induced by STDP are somehow well understood for single neurons and feed-forward architectures, for which several analytical models are available (Kempster et al. 1999, Gütig et al. 2003, Meffin et al. 2006). In such cases, STDP tends to favor correlated input pathways, but elaborated stimulating inputs can generate much more complex dynamics, for example, with spike patterns where earlier spikes are picked up by STDP (Masquelier et al. 2008) or input pathways with oscillatory activity at distinct frequencies to compete between each other (Dahmen et al. 2008). In contrast, the effect of STDP in recurrent networks is still unclear even

for simple models of stimulating inputs and has only begun to be addressed, mainly using numerical simulation (Song and Abbott 2001, Wensch et al. 2005, Morrison et al. 2007, Lubenov and Siapas 2008, Câteau et al. 2008, Kang et al. 2008). In particular, there exist only a few theoretical results for recurrent architectures (Karbowski and Ermentrout 2002, Masuda and Kori 2007) due to mathematical difficulties in evaluating the effects of feedback synaptic loops. Comparison with previous results will be discussed in more details throughout the text. For a review on the dynamical implications of the various models of STDP, see Morrison et al. (2008).

1.3.4 Functional implications of synaptic plasticity and STDP

Synaptic plasticity is hypothesised to give rise to functional structure in neuronal networks and is thus central to the understanding of neuronal information processing. For example, the emergence of cortical organisation similar to that observed in the primary visual cortex (Hubel and Wiesel 1962) has been the subject of numerous studies (von der Malsburg 1973, Swindale 1996, Goodhill 2007). Such a learning scheme is believed to combine two complementary mechanisms: genesis of new connections and pathways (through growth of axons and creations of new synapses) and activity-dependent selection of synaptic connections. We constrain the present study to activity-dependent synaptic plasticity in neuronal networks.

It is not clear yet how rich a specialisation STDP can generate upon the synaptic weights in neuronal networks, especially when recurrent connections are involved. Many theoretical studies have focused on input selectivity for single neurons (Kempster et al. 1999, Gütig et al. 2003, Meffin et al. 2006) and then extended to particular structures of stimulating inputs in order to reproduce experimental results. For example, neurons were successfully trained as coincidence detectors to represent inter-aural time difference in a neuronal map (Leibold et al. 2002), tonotopic maps for oscillatory inputs with distinct frequency (Dahmen et al. 2008) and to recognise spike patterns (Masquelier et al. 2008), as well as firing-rate patterns (Fusi 2002). In recurrent networks, the relationship between STDP and spiking synchrony has been the subject of many studies (Senn 2002, Câteau et al. 2008, Lubenov and Siapas 2008). Synchronous activity is related to assembly of neu-

ronal cells and hypothesised to take part in decentralised coding of neuronal information (Hebb 1949). However, it is still controversial under which conditions STDP induces more (Izhikevich et al. 2004) or less (Iglesias et al. 2005) synchronisation. STDP has also been shown capable of enhancing the neuronal response to stimuli by reducing its variability (Bohte and Mozer 2007).

Synaptic plasticity is hypothesised to be a candidate underlying mechanism for memory (Bienenstock et al. 1982, Hopfield 1982, Levy and Steward 1983, Amit and Brunel 1997, Martin et al. 2000, Tsodyks 2002). STDP links in a natural fashion to the Hebbian postulate, as only pre-synaptic spikes that take part in the firing of an output spike lead to potentiation (Hebb 1949). STDP was shown to have interesting implications for both supervised (Pfister and Gerstner 2006, Molter et al. 2007) and unsupervised (Carnell 2009) learning. However, it is necessary to bridge the gap between the time scale of STDP, namely spike-time correlations of tens of milliseconds, and behavioural temporal associations (order of seconds), in order to link the physiological and cognitive levels (Drew and Abbott 2006).

These examples illustrate the diversity of neuronal specialisation that STDP can generate. Another striking observation concerns the versatility of generic cortical circuits that seem capable of specialising to a broad range of functions, for example in the sensory and motor areas. This suggests that a common neuronal code and plasticity mechanisms may be used to represent and learn to process very different sensory signals. Bridging the gap between local mechanisms at the scale of neurons and network dynamics is crucial to gain insight into the role of the players at the different scales (synaptic parameters, neuronal activity, connectivity and learning rule). A better understanding of the weight dynamics in relation to the input structure, in particular their spike-time correlations, should provide a unifying framework to explain the common trends behind the apparent diversity.

1.4 Context and focus of this PhD project

The present study aims to develop a mathematical framework that describes the weight dynamics in a recurrent neuronal network. The main theoretical contribution is the ex-

tension of previous work based on the Poisson neuron model (Kempster et al. 1999, Gütig et al. 2003) to the case of recurrent connectivity. This neuron model illustrates the success of stochastic point processes in modelling neuronal activity (Sec. 1.2.3). The models used in the present study will be chosen to keep the analysis as tractable as possible.

In a previous paper (Burkitt et al. 2007), we presented a framework for the analysis of STDP in recurrent networks with arbitrary topology subject to external stimulation. This framework describes how the network activity, viz., firing rates and spike-time correlations, determines the evolution of the weights that occurs on a much slower time scale than other neuronal activation mechanisms. It provides a *soluble* differential system of equations that allows us to predict the resulting development of structure within the network, in particular the asymptotic distribution of the firing rates and of the weights after a sufficiently long learning epoch (the emerged structure). However, the analysis of the weight dynamics was carried out only for a network with full recurrent connectivity and no external inputs.

This thesis presents subsequent results that analyse the more biologically realistic case in which a recurrent network with partial connectivity is stimulated by external input neurons that have a functional structure. It extends our previously developed framework to incorporate the effect of the post-synaptic response, which was simplified as a delta-function response by Burkitt et al. (2007). The study of the unsupervised learning scheme generated by STDP through the weight dynamics is a first step to link the physiological model of STDP to the domain of machine learning. We will illustrate this by addressing the following question: how can STDP generate functional self-organisation in neuronal networks? In this way, we investigate whether the behaviour of a network where STDP modifies the synaptic connections can be related to the algorithm proposed by Kohonen (1982), which leads to specialised units (areas) sensitive to some input features presented to the network, thus performing categorisation on the stimuli.

1.5 Plan

Chapter 2 introduces the models of Poisson neuron and discusses its relevance compared to other neuron models, as well as the model of weight-dependent STDP that we use. Chapter 3 presents the derivation of the theoretical framework to investigate the weight dynamics induced by additive STDP. The evolution of the network activity and synaptic weights is described by a dynamical system, which is analysed to predict the emergence of the weight structure for different network configurations:

- only the input connections are plastic (Chapter 4);
- only the recurrent connections are plastic (Chapter 5);
- both the input and recurrent connections are plastic (Chapter 6).

The framework can account for a network with an arbitrary number of neurons and external stimulating sources, but the analysis focuses on the specific case where the inputs are partitioned into homogeneous pools. The asymptotic weight structure is determined in terms of the input stimulation and the learning parameters. Chapter 6 generalises the analysis of the previous chapters to weight-dependent STDP and investigates some of its implications in terms of computation performed by STDP on the synaptic weights. Chapter 7 develops a mathematical framework based on a particular class of stochastic point processes, the piecewise deterministic Markov process, to extend our analysis to Poisson neurons with non-linear activation function. The conclusion in Chapter 8 relates the presented results to existing literature and links them to neuronal information processing.

Chapter 2

Modelling the brain at the neuronal level

This chapter introduces the models of neuron and synaptic plasticity that are used throughout the present study.

2.1 Introduction

THIS chapter describes the neuron model used in the subsequent chapters, as well as the mathematical foundation of STDP. This forms the basis for the analysis presented in later chapters: the theoretical framework derived in Chapter 3 relies on the Poisson neuron (Sec. 2.2.1) and Hebbian additive STDP (Sec. 2.3). A general version of STDP is introduced, although Chapters 4 and 5 use the simplified additive version in order to keep the analysis of the weight dynamics tractable. Chapter 6 extends the study of the weight dynamics to the more general weight-dependent STDP (non-additive).

2.2 Modelling neuronal activity in networks

2.2.1 Poisson neuron model

The Poisson neuron (Kempster et al. 1999) is a stochastic point process for which the spiking mechanism of a given neuron i is approximated by an inhomogeneous Poisson process. The corresponding intensity function $\rho_i(t)$ generates an output spike-time series $S_i(t)$, which can be represented as a sum of delta-functions or Dirac comb. One of the easiest ways to understand this model is to use discrete time, and at each time step Δt ,

the probability that the neuron fires is $\rho_i(t)\Delta t$. In addition, the probability that two or more spikes occur during Δt is $o(\Delta t)$, i.e., $o(\Delta t)/\Delta t \rightarrow 0$ when $\Delta t \rightarrow 0$, while events in disjoint intervals are independent. This homogeneous Poisson process models the effects of the currents I_K , I_{Na} and I_{Ca} , as well as the leakage related to τ_m in Eq. (1.1).

The rate function $\rho_i(t)$ of neuron i is to be related to the soma potential $V(t)$ in Eq. (1.1) and it evolves over time according to the activity of its pre-synaptic inputs (I_{syn}), as shown in Fig. 2.1:

$$\rho_i(t) = \nu_0 + \sum_k K_{ik}(t) \sum_n \epsilon \left(t - \hat{t}_{k,n} - \hat{d}_{ik} \right). \quad (2.1)$$

The constant ν_0 is the spontaneous firing rate (identical for all neurons) that accounts for other pre-synaptic connections that are not considered in detail. Each synapse indexed by k is excited by a spike-time series $\hat{S}_k(t)$, whose spikes each induce a variation of $\rho_i(t)$, namely the post-synaptic potential (PSP). The PSP is determined by the synaptic weight K_{ik} , the post-synaptic response kernel $\epsilon(t)$, and the delay \hat{d}_{ik} . The kernel function $\epsilon(t)$ models the time course of the PSP due to the current flowing into the post-synaptic neuron for one single pre-synaptic spike; $\epsilon(t)$ is normalised to one: $\int \epsilon(t) dt = 1$; and, in order to preserve causality, we have $\epsilon(t) = 0$ for $t < 0$. The overall synaptic influx is the sum of the PSPs over all past spike times $\hat{t}_{k,n}$ (related to the k^{th} synapse, and indexed by n) of the spike trains $\hat{S}_k(t)$. We often consider only positive weights, i.e., excitatory synapses, to ensure that the rate function ρ_i remains positive; otherwise, the inhibition level, as will be studied in Chapter 6, has to be chosen adequately. This model also does not usually incorporate a refractory period.

Some extensions of the Poisson neuron have been used in previous studies, such as applying an increasing and bounded activation function on the rhs of Eq. (2.1) to clip $\rho_i(t)$ to a given range. This more realistic neuron model incorporates the saturation observed for real neurons and allows us to use inhibitory synapses (ρ_i can be ensured to remain positive at all times), but the calculations are not as tractable and often require approximations. In Chapter 7, we study a network of such non-linear Poisson neurons to evaluate the effect of the activation function on the steady state of the spiking activity. This extension is a first step towards incorporating more biologically realistic features.

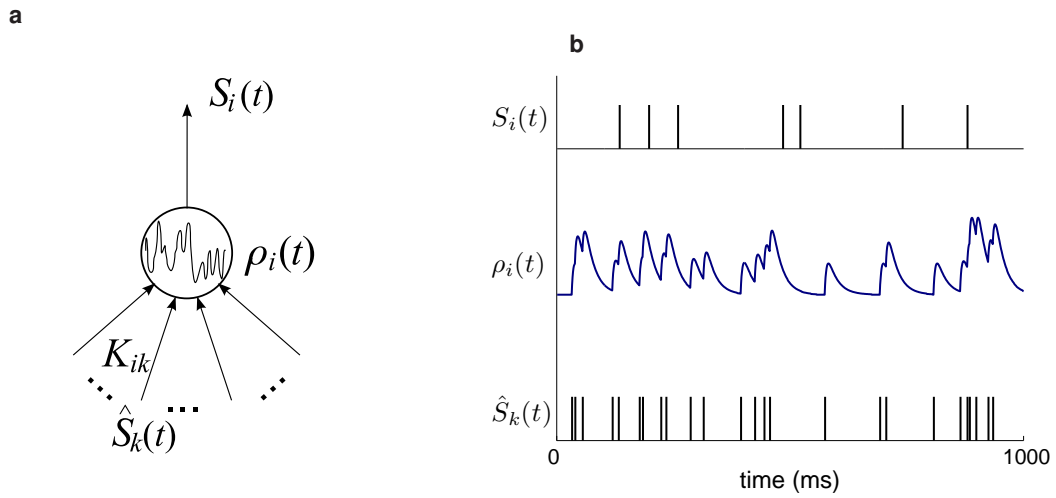


Figure 2.1: Poisson neuron model. (a) Schematic view of the neuron i . The soma (or cell body, circle) receives inputs from the synapses (*below*, indexed by k) and fires spikes that propagates on the axon (*above*), towards synaptic connections with other neurons (not represented). The flow of information is from bottom to top. (b) Illustration of the variation of $\rho_i(t)$ over 1000 ms (*middle plot*) for a given succession of pre-synaptic spikes $\hat{S}_k(t)$ (*bottom spike train*) and one of many possible randomly generated outputs, denoted by $S_i(t)$ (*top spike train*).

2.2.2 Relation to more elaborate neuron models

The Poisson neuron is a coarse approximation of the activation mechanisms that take place in real neurons. It accounts for the variability observed in some cortical neurons (Poggio and Viernstein 1964, Noda and Adey 1970), but its firing output is very noisy compared to the IF neuron, whose basic functioning is more deterministic (Burkitt 2001, Burkitt 2006, Moreno-Bote et al. 2008). Consequently, IF neurons perform better when precisely predicting the output spike-time series in response to *in vivo* currents (Rauch et al. 2003, Jolivet et al. 2008). As another example, the stability and the synchronisation of IF neurons in recurrent networks can dramatically change depending on feedback coupling and external stimulation (Brunel and Hakim 1999, Brunel 2000, Burkitt et al. 2003). Such interesting features do not emerge with Poisson neurons and a more complete study of neuronal synchronisation would thus require a more elaborate neuron model, in addition to the understanding of the weight dynamics.

However, the focus of this study is the weight dynamics that are assumed to be very slow compared to other neuronal activation mechanisms. With this separation of the

time scales (van Hemmen 2001), the activation mechanisms need not be too “realistic” to evaluate the effect of synaptic plasticity: analyses using Poisson neurons correctly predict the qualitative evolution of the weights for IF neurons (Kempster et al. 1999, Gütig et al. 2003). This suggests that the effect of STDP is mainly related to the increase of the probability of firing an output spike when receiving a post-synaptic potential, which is qualitatively captured by this neuron model.

2.2.3 Model of recurrent neuronal network

When connecting neurons in a recurrent fashion, some of the pre-synaptic spike trains for neuron i in Eq. (2.1) are the outputs of other neurons. This feedback imposes a constraint upon the neuronal activity, which has non-trivial implications. Rigorously speaking, recurrently connected Poisson neurons are no longer inhomogeneous Poisson processes, but Hawkes processes (Hawkes 1971). The derivation of equations to model the learning dynamics in the presence of feedback connections is the subject of Chapter 3; the theoretical framework is general and can be applied to any network topology. The consequences for the neuronal dynamics are then analysed in Chapters 4, 5 and 6 for a particular configuration that extends previous studies (Kempster et al. 1999, Gütig et al. 2003, Meffin et al. 2006) to the case of a network: recurrently connected neurons are stimulated by homogeneous external pools of spike trains with within-pool spike-time correlations but no between-pool correlations. In this way, the spiking information conveyed by the inputs is carried by the firing rates and the pairwise correlations. More details about the generation of the input spike trains are provided in Sec. 3.5.

2.3 Spike-timing-dependent plasticity (STDP)

2.3.1 STDP for pairwise interactions between pre- and post-synaptic spikes

In order to investigate the functional effect of STDP in neuronal networks, we use a phenomenological model of STDP (Morrison et al. 2008) that describes the change in the synaptic weight induced by single spikes and pairs of pre- and post-synaptic spikes. This

choice has limitations compared to more elaborate models that include triplets of spikes in their analysis (Sjöström et al. 2001, Pfister and Gerstner 2006, Badoual et al. 2006, Appleby and Elliott 2006). However, from a functional point of view, all STDP models are sensitive to pairwise spike-time correlation. Gaining a better understanding of this relationship is the core of the present study and we ignore higher-order correlations (three-point correlation, etc.). Our version of STDP thus captures the mean effect illustrated by the interpolation curves in Fig. 1.5. This effect corresponds to the leading order of the weight change with respect to the temporal structure of the pre- and post-synaptic spike trains. It has been shown to determine the structuring of the synaptic weights for single neurons stimulated by correlated inputs (Kempster et al. 1999, Gütig et al. 2003). However, we leave to subsequent studies further non-linearities in pairwise STDP models, such as the restriction of the pairs of input and output spikes that contribute to the weight change (Burkitt et al. 2004). In other words, the present model captures the weight modification by STDP for neurons in a firing regime such that spike pairs predominate in terms of plasticity, unlike bursting; see Sec. 1.3.2 for more details.

Among all the different versions of pairwise STDP, it is still unclear which features are crucial and physiologically realistic in terms of weight dynamics, such as the degree of non-linearity related to the weight dependence (van Rossum et al. 2000, Bi and Poo 2001, Gütig et al. 2003) or the detailed shape of the learning window (Morrison et al. 2008). The present study of their functional implications for learning and neuronal dynamics may prove to be helpful in this debate.

The general model for pairwise STDP considered in this study applies to an excitatory synapse with weight J . The weight change δJ for a sole pair of pre- and post-synaptic spikes corresponding to the respective times t^{in} and t^{out} at the synaptic site is given by

$$\delta J = \eta \begin{cases} w^{\text{in}} & \text{at time } t^{\text{in}} \\ w^{\text{out}} & \text{at time } t^{\text{out}} \\ f_+(J)W_+(t^{\text{in}} - t^{\text{out}}) & \text{at time } t^{\text{out}} \text{ if } t^{\text{in}} < t^{\text{out}} \\ -f_-(J)W_-(t^{\text{in}} - t^{\text{out}}) & \text{at time } t^{\text{in}} \text{ if } t^{\text{in}} > t^{\text{out}}. \end{cases} \quad (2.2)$$

The rate-based contribution w^{in} (resp. w^{out}) accounts for the effect of each pre-synaptic (post-synaptic) spike and occurs only once per spike (Kempster et al. 1999, Burkitt et al. 2007). The STDP learning window function W_+ describes the potentiation of the weight J when the pre-synaptic spike occurs after that of the post-synaptic spike; conversely, W_- describes the depression of J when the pre-synaptic spike occurs after the post-synaptic spike. These correlation contributions are each rescaled by the function f_+ and f_- , respectively, that are illustrated in Fig. 2.2. We consider the functions f_+ , f_- , W_+ and W_- to be non-negative to correspond to Hebbian learning; for the sake of consistency, $W_+(u) = 0$ for $u > 0$ and $W_-(u) = 0$ for $u < 0$. The parameters used are in the range of values observed in the physiology (Caporale and Dan 2008). All these contributions are scaled by a learning parameter η , typically chosen to be very small, to model learning processes that occur very slowly compared to the other neuronal and synaptic mechanisms. In this study we consider learning rates such that STDP can lead to the emergence of a neuronal specialisation in tens of minutes or hours (Kempster et al. 1999, Gütig et al. 2003, Burkitt et al. 2007). In this realistic range, the weight dynamics exhibits a low level of noise and a stable asymptotic distribution (Meffin et al. 2006).

In the version of STDP proposed by Kempster et al. (1999), δJ is independent of the value of the weight J at the time of the change. In other words, $f_-(J) = f_+(J) = 1$, as represented by the grey horizontal lines in Fig. 2.2(a). The time-difference contributions in Eq. (2.2) can then be described using a single function $W = W_+ - W_-$, namely

$$W(t^{\text{in}} - t^{\text{out}}) \quad \text{at time } \max(t^{\text{in}}, t^{\text{out}}). \quad (2.3)$$

The impact on feed-forward single neurons is a homeostatic stability of the weight mean (with a fast convergence) and then the emergence of a structure among the population of weights according to the correlation between the inputs (with a slower emergence). This type of learning can be robust to noise to a certain extent. This version of STDP is referred to as “additive”. It requires the use of explicit (“hard”) bounds on the weights, because the intrinsic competition between the weights causes them to individually diverge, even when the mean weight remains constant. This phenomenon also occurs for

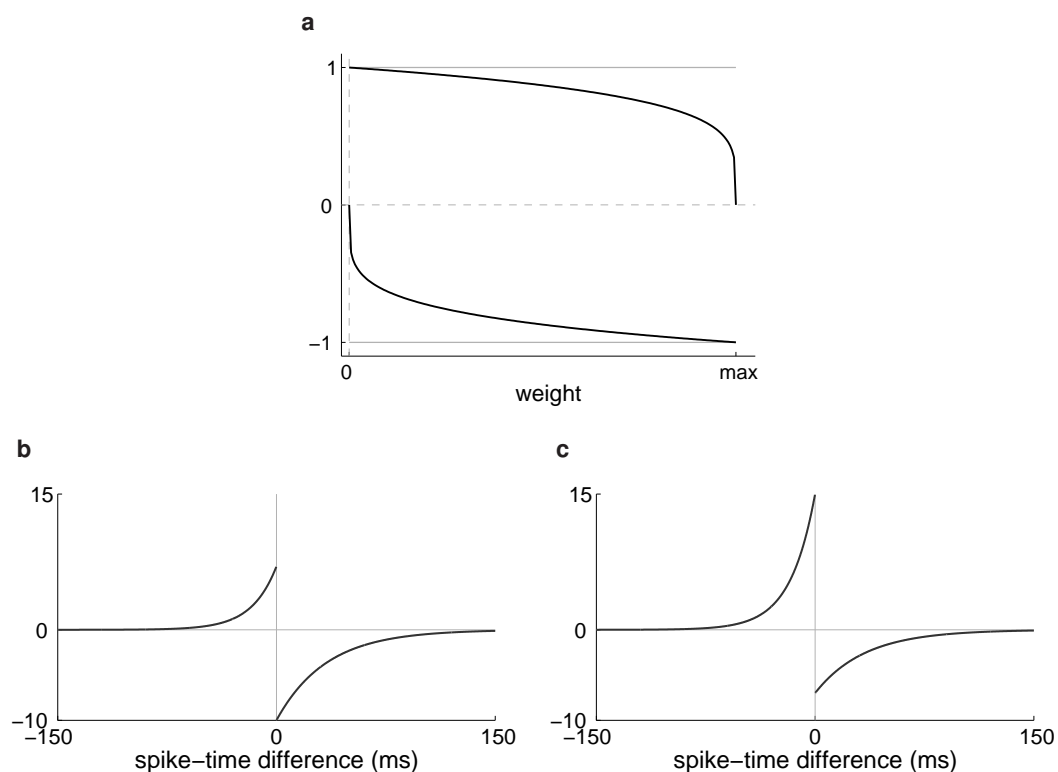


Figure 2.2: Example of STDP window function with weight dependence. (a) The weight dependence is determined by the scaling functions f_- and f_+ , which are chosen such that both $-f_-(J)$ (bottom black curve) and $f_+(J)$ (top black curve) decrease with J , which leads to more depression and less potentiation for (b) a strong synapse compared to (c) a weak synapse. The dependence in the spike-time difference is taken care of by one decaying exponential W_- for depression (right curves) with time constant 17 ms and likewise W_+ for potentiation (left curves) with 34 ms. See Appendix D for details on the parameters.

homogeneous uncorrelated inputs and the neuron then specialises in an arbitrary fashion, which is unsatisfactory.

Weight-dependence of the potentiation side of STDP has been demonstrated in experimental data (Bi and Poo 2001). Contrary to the additive version, weight-dependent STDP rules can induce “soft” bounds on the weights: a stable distribution of the weights can be obtained without explicit bounds on the synaptic weights, for example with a version of STDP where only the potentiation is weight dependent: $f_-(J) = 1$ and $f_+(J) = 1 - J$ (van Rossum et al. 2000). A drawback of this version lies in the lack of intrinsic competition between the synapses located on the same neuron; a renormalisation mechanism

was thus used to correct this unwanted feature.

Gütig et al. (2003) extended this work to the more general version of STDP: $f_-(J) = J^\alpha$ and $f_+(J) = (1 - J)^\alpha$ with the parameter $\alpha \in [0, 1]$. This non-linear version (when $\alpha > 0$) proved capable of inducing stable weight dynamics for a broad range of parameters and an interesting learning paradigm when a neuron is stimulated by two identical correlated input pools: STDP can break the symmetry between initially homogeneous weights to specialise the neuron to only one of the input pools. Such symmetry breaking is required, for example, to obtain the emergence of spatio-temporal feature maps, such as the column organisation of the primary visual cortex (Hubel and Wiesel 1962) and tonotopic maps in the auditory cortex (Schreiner et al. 2000). Input selectivity using weight-dependent SDTP for several homogeneous correlated input pools can lead to the selection of only a portion of the input pools depending on the input and learning parameters (Meffin et al. 2006).

Usually, t^{in} and t^{out} in Eq. (2.2) correspond to the times when the pre- and post-synaptic spikes affect the synaptic site, thus involving axonal and dendritic delays. Dendritic and axonal delay affects the weight dynamics when correlation is present (Senn 2002, Lubenov and Siapas 2008). Only axonal delays are considered in this work, which accounts for the axonal propagation of action potentials and the diffusion time of the neurotransmitters in the synaptic cleft. The framework can incorporate dendritic delays, but a detailed analysis is left to subsequent work, together with broad distributions of delays (as detailed in Chapter 3); some effects will also be discussed in Chapter 6.

2.3.2 Relation to rate-based models

A number of analyses (Burkitt et al. 2004, Elliott 2008) have shown the relation between STDP and rate-based learning. For a synapse k with weight J_k , rate-based plasticity can be formulated in the following general way (Sejnowski 1977, Bienenstock et al. 1982, Gerstner and Kistler 2002)

$$\Delta J_k \propto c_0(J_k) + c_1^{\text{pre}}(J_k) \phi_k + c_1^{\text{post}}(J_k) \phi_{\text{out}} + c_2^{\text{pre}}(J_k) \phi_k^2 + c_2^{\text{post}}(J_k) \phi_{\text{out}}^2 + c_{\text{corr}}(J_k) \phi_k \phi_{\text{out}}, \quad (2.4)$$

where ϕ_k and ϕ_{out} are the pre- and post-synaptic firing rates, respectively, and the coefficients c_x can be functions of the weight J_k ; the dependence in t is omitted. Contrary to the time-averaged firing rates considered by Kempter et al. (1999) and Burkitt et al. (2007) in their analysis that correspond to a moving average over a window of, say, tens of seconds (see Sec. 3.2.1 for the formal definition), the firing rates ϕ Eq. (2.4) may correspond to a medium time scale (down to tenths of a second) and often for non-spiking neurons. However, such rate-based models ignore spike-time correlations at a very short time scale (order of milliseconds), which is captured by STDP in addition to time-averaged rate-based information. This crucial limitation led to the proposition of STDP by Gerstner et al. (1996).

Rate-based models have another notorious limitation: they fail to generate both stabilisation and specialisation upon the synaptic weights at the same time (Bienenstock et al. 1982, Miller and Mackay 1994, Miller 1996, Rao and Sejnowski 2001). To correct the lack of stability, additional mechanisms such as a renormalisation for each neuron of the synaptic weights after computing the changes have been used. The biological evidence of such renormalisation is controversial, although some links with limited resources in neurotransmitters have been argued.

STDP was shown to be capable of generating both stability and competition for the input weights of a single neurons (Kempter et al. 1999, Song et al. 2000, Song and Abbott 2001, Morrison et al. 2007, Burkitt et al. 2007). One way to achieve stability for the output spiking activity when using additive STDP is to use the rate-based terms w^{in} and w^{out} in Eq. (2.2) to perform a normalisation of the incoming weights for each neuron via a dynamical equilibrium (Kempter et al. 1999, Burkitt et al. 2007). These rate-based terms are equivalent to c_1^{pre} and c_1^{post} in Eq. (2.4); they are not always incorporated in STDP rules as their physiological origin is not so clear as the learning window function W_{\pm} that was observed experimentally. From a functional point of view, using w^{in} and w^{out} or another additional renormalisation mechanism, such as an adequate weight dependence for STDP, appears to be equivalent. This will be discussed in more depth in Chapter 6.

Chapter 3

Dynamical system to model network activity and synaptic plasticity

This chapter presents the derivation of a dynamical system that describe the evolution of the synaptic weights induced by STDP within the context of a neuronal network that receives stimulation from external inputs.

3.1 Overview

IN order to incorporate the mechanisms related to the neuronal post-synaptic response in the weight dynamics, it is necessary to extend the framework presented by Burkitt et al. (2007), where the neuronal information contained in the spike trains is conveyed by firing rates and spike-time correlations (Kempster et al. 1999, Gütig et al. 2003, Meffin et al. 2006). These variables of importance to describe the neuronal activity are linked in a dynamical system together with the synaptic weights. This framework is adapted to predict the evolution of the expectation values for the firing rates, correlations and weights depending on the learning and stimulation parameters. This chapter is constrained to additive STDP.

The following mathematical assumptions (Kempster et al. 1999, van Hemmen 2001, Burkitt et al. 2007) are used in order to derive the dynamical system:

- the expectation values of the firing rates and pairwise covariances are constant in time for the external inputs, which is equivalent, here, to constant time-averaged firing rates and covariances for any realization of the spiking history;
- the separation of the time scales of the activation mechanisms and the learning

dynamics, the latter happening on a slower time scale (adiabatic hypothesis);

- the expectation values of the firing rates and pairwise covariances are quasi-constant in time for the network neurons, i.e., they only vary due to the slow learning on the weights.

3.2 Description of the network

Let us consider a network of N Poisson neurons (indexed by $1 \leq i, j \leq N$) with recurrent connections that is stimulated by M Poisson spike trains (a.k.a. external inputs or sources, indexed by $1 \leq k, l \leq M$) through input connections, as illustrated in Fig. 3.1(a). Typically both M and N are large. In addition to receiving synaptic input from the external sources, as shown for a sole neuron in Fig. 2.1, each neuron is also excited by other neurons via connections that may form feedback loops in the network, but without self-connections. The weight of the connection from input k to neuron i is denoted by $K_{ik}(t)$ and the corresponding delay is \hat{d}_{ik} (as defined in Sec. 2.2.1); likewise, we define $J_{ij}(t)$ and d_{ij} for the connection from neuron j to neuron i ; see Fig. 3.1(b). Both input and recurrent synapses share the same PSP kernel ϵ introduced in Sec. 2.2.1. We will consider both fully- and partially-connected networks, where each neuron is stimulated by some of the M inputs; n^K will denote the total number of input connections and n^J the total number of recurrent connections in the network. Partially-connected networks are generated by randomly assigning input-to-neuron and neuron-to-neuron connections. The term ‘pool’ will always refer to the external inputs, the term ‘group’ to the neurons.

Spikes are considered to be instantaneous events. We define $\hat{S}_k(t)$ as the spike-time series (Dirac comb) of the external input k ; its value is zero except at the times when a spike is fired and the spike train is described as a sum of Dirac delta-functions. If there is a spike in a given small time interval $[t, t + \delta t]$, then

$$\int_t^{t+\delta t} \hat{S}_k(t') dt' = 1, \quad (3.1)$$

where δt is “small” compared to the time scale of other neuronal mechanisms (ϵ , delays, etc.) allowing us to consider \hat{S}_k as approximate delta-functions. The spike-time series

$S_i(t)$ for network neuron i is defined similarly.

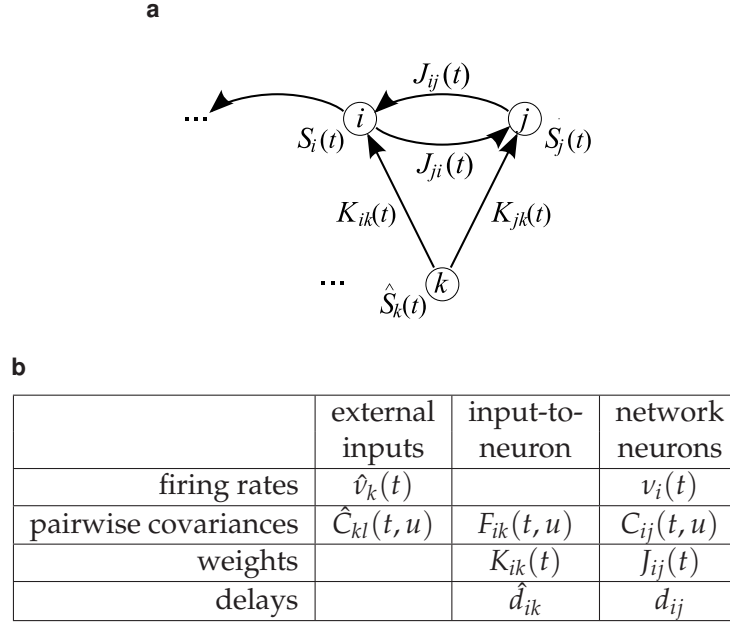


Figure 3.1: Presentation of the network and notation. (a) Schematic representation of one of the M external inputs (bottom circle, indexed by $1 \leq k \leq M$) and two of the N network neurons (top circles, $1 \leq i, j \leq N$). The input and recurrent connections have plastic weights $K_{ik}(t)$ and $J_{ij}(t)$ respectively (thick arrows). The spike trains of the input k and neuron i are denoted by $\hat{S}_k(t)$ and $S_i(t)$ respectively. (b) The table shows the variables that describe the neuronal activity: time-averaged firing rates \hat{v} and v ; time-averaged covariances \hat{C} , F and C ; and the variables related to the synaptic connections: weights K and J ; delays \hat{d} and d .

3.2.1 Definition of the state variables for the network

We now define a number of variables recapitulated in Fig. 3.1(b) to describe the activity of the network neuronal activity and contain the relevant information for STDP.

The time-averaged firing rate $v_i(t)$ for neuron i corresponds to a duration T that will be specified later in terms of the learning and neuronal dynamics (Kempster et al. 1999, Burkitt et al. 2007)

$$v_i(t) := \frac{1}{T} \int_{t-T}^t \langle S_i(t') \rangle dt', \quad (3.2)$$

where $\langle S_i(t) \rangle$ is the instantaneous firing rate averaged over the randomness. The firing rate $\hat{v}_k(t)$ for input k is defined similarly.

For fixed weights and steady inputs, the network activity as a stochastic process is

actually ergodic, which implies that $\frac{1}{T} \int_{t-T}^t S_i(t') dt' \simeq \langle S_i(t) \rangle = \text{const.}$ to a good approximation for large T . Therefore, we could simply define $v_i(t)$ as $\frac{1}{T} \int_{t-T}^t S_i(t') dt'$ in Eq. (3.2), when the weights change very slowly compared to T . We keep the notation of Eq. (3.2) to comply with the formalism developed by Kempster et al. (1999) and Burkitt et al. (2007).

Instead of the correlation coefficients used by Burkitt et al. (2007), we use the neuron-to-input time-averaged covariance F_{ik} :

$$\begin{aligned} F_{ik}(t, u) &:= \frac{1}{T} \int_{t-T}^t \langle S_i(t') \hat{S}_k(t' + u) \rangle dt' - \frac{1}{T} \int_{t-T}^t \langle S_i(t') \rangle \langle \hat{S}_k(t' + u) \rangle dt', \quad (3.3) \\ F_{ik}^\Psi(t) &:= \int_{-\infty}^{+\infty} \Psi(u) F_{ik}(t, u - \hat{d}_{ik}) du, \end{aligned}$$

where Ψ is a given kernel function and $\tilde{\Psi} = \int \Psi(u) du$ its integral value. These two formulas are usually defined for stationary second-order stochastic processes, which requires in particular constant instantaneous firing rates $\langle S_i(t) \rangle$ and $\langle \hat{S}_k(t) \rangle$ (steady inputs and fixed weights); we used $\hat{v}_k(t) \simeq \hat{v}_k(t + u)$ for u in the range of the STDP window function W .

Likewise, the time-averaged covariance C_{ij} and covariance coefficient C_{ij}^Ψ between neurons i and j are defined in the following way

$$\begin{aligned} C_{ij}(t, u) &:= \frac{1}{T} \int_{t-T}^t \langle S_i(t') S_j(t' + u) \rangle dt' - \frac{1}{T} \int_{t-T}^t \langle S_i(t') \rangle \langle S_j(t' + u) \rangle dt', \quad (3.4) \\ C_{ij}^\Psi(t) &:= \int_{-\infty}^{+\infty} \Psi(u) C_{ij}(t, u - d_{ij}) du, \end{aligned}$$

as are the time-averaged covariance \hat{C}_{kl} and covariance coefficient \hat{C}_{kl}^Ψ between inputs k and l

$$\begin{aligned} \hat{C}_{kl}(t, u) &:= \frac{1}{T} \int_{t-T}^t \langle \hat{S}_k(t') \hat{S}_l(t' + u) \rangle dt' - \frac{1}{T} \int_{t-T}^t \langle \hat{S}_k(t') \rangle \langle \hat{S}_l(t' + u) \rangle dt', \quad (3.5) \\ \hat{C}_{kl}^\Psi(t) &:= \int_{-\infty}^{+\infty} \Psi(u) \hat{C}_{kl}(t, u) du. \end{aligned}$$

Note that the input covariance coefficients \hat{C}_{kl}^Ψ do not involve delays. As explained in Appendix A.1.1, the covariances $\hat{C}_{kl}(t, u)$ by convention do not incorporate the atomic (or

point) discontinuity at $u = 0$ due to the autocorrelation of the stochastic point processes \hat{S}_k for $k = l$, namely $\langle \hat{S}_k(t) \rangle \delta(u)$, where δ is the Dirac delta-function. This means that \hat{C} represents the input correlation structure (“spiking information”) but does not contain the autocorrelation intrinsic to the neuron model.

For the sake of simplicity, we use matrix notation in the remainder of the text: vectors $v(t)$ and $\hat{v}(t)$, and matrices $F^W(t)$, $\hat{C}^W(t)$, $K(t)$ and $J(t)$.

3.3 Slow weight evolution

We now derive a learning equation to describe the evolution of the input weights due to STDP according to the activities of the pre- and post-synaptic neurons for each synapse. For a small time interval $[t, t + \delta t]$, the change in the input weight $K_{ik}(t)$ described in Eq. (2.3) can be expressed using the pre- and post-synaptic spike trains (Kempster et al. 1999)

$$\begin{aligned} \delta K_{ik}(t) &= \eta \int_t^{t+\delta t} [w^{\text{in}} \hat{S}_k(t' - \hat{d}_{ik}) + w^{\text{out}} S_i(t')] dt' \\ &+ \eta \int_{(t', u) \in \mathcal{I}(t)} W(u) S_i(t') \hat{S}_k(t' - \hat{d}_{ik} + u) du dt'. \end{aligned} \quad (3.6)$$

Recall that \hat{d}_{ik} that was defined in Sec. 2.2.1 is assimilated here to the axonal delay described in Sec. 2.3: \hat{d}_{ik} then accounts for the axonal propagation and the diffusion of neurotransmitters and we neglect the dendritic delay compared to them. Dendritic delays are not considered. Thus, $\hat{S}_k(t' - \hat{d}_{ik})$ is the delayed time series of the pre-synaptic spikes; the time difference at the synaptic site between the pre- and the post-synaptic spikes (at respective times t^{pre} and t^{post} at the somas of both neurons) is $u = t^{\text{pre}} + \hat{d}_{ik} - t^{\text{post}}$. The domain of integration $\mathcal{I}(t)$ is the subset $(t', u) \in \mathbb{R}^2$ satisfying the three conditions

$$\begin{aligned} t' &\leq t + \delta t; \\ t' - \hat{d}_{ik} + u &\leq t + \delta t; \\ t &\leq t' \text{ or } t \leq t' - \hat{d}_{ik} + u. \end{aligned} \quad (3.7)$$

The first two lines require that the spikes occur before $t + \delta t$, and the last line that at least one of them is in the time interval $[t, t + \delta t]$.

The change in weights over many independent trials (repetitions) is equivalent to averaging over a single long trial of length T . This self-averaging property of the learning requires the learning rate η to be small (van Hemmen 2001); then T can be chosen to be long compared to the time scale of the neuronal and synaptic activation mechanisms, but small compared to η^{-1} (separation of time scales). Typically, T is of the order of seconds (or tens of seconds) for synaptic mechanisms with characteristic times of tens of ms. This allows us to choose η such that δK_{ik} in Eq. (3.6) is at most a thousandth of the weight upper bound; for $\eta = 5 \times 10^{-7}$ (Appendix D), the effective learning epoch is of the order of tens of minutes. The ensemble average over the resulting random process is denoted by the angular brackets $\langle \cdot \cdot \rangle$. The rate of change for the expectation value of the external weight $\dot{K}_{ik}(t)$ is approximated by the temporal average of the summation of all $\langle \delta K_{ik}(t) \rangle$, i.e., the ensemble average taken of Eq. (3.6), over a time interval of duration T . This time-averaging allows the bounds of integration of t' in Eq. (3.6) to be slightly modified with good approximation in order to obtain (Kempster et al. 1999)

$$\begin{aligned} \dot{K}_{ik}(t) \simeq & \frac{\eta}{T} \int_{t-T}^t \left[w^{\text{in}} \langle \hat{S}_k(t' - \hat{d}_{ik}) \rangle + w^{\text{out}} \langle S_i(t') \rangle \right] dt' \\ & + \frac{\eta}{T} \int W(u) \left[\int_{t-T}^t \langle S_i(t') \hat{S}_k(t' - \hat{d}_{ik} + u) \rangle dt' \right] du . \end{aligned} \quad (3.8)$$

The terms of of Eq. (3.8) involving w^{in} and w^{out} give the time-averaged firing rates of the pre- and post-synaptic spike trains, $\hat{v}_k(t)$ and $v_i(t)$, respectively; see Fig. 3.1(b). The last term involves the time-averaged pairwise correlation (Kempster et al. 1999, Burkitt et al. 2007),

$$D_{ik}(t, u) := \frac{1}{T} \int_{t-T}^t \langle S_i(t') \hat{S}_k(t' + u) \rangle dt' ; \quad (3.9)$$

this expression is convolved with the STDP window function $W(u)$ shifted by the delay \hat{d}_{ik} , which is embodied by the coefficient

$$D_{ik}^W(t) := \int_{-\infty}^{+\infty} W(u) D_{ik}(t, u - \hat{d}_{ik}) du . \quad (3.10)$$

In order to incorporate \hat{d}_{ik} , this correlation coefficient has been modified compared to that used by Burkitt et al. (2007). They are related to the coefficient F and F^W defined in Eq. (3.3) with $\Psi = W$ in the following way:

$$\begin{aligned} F_{ik}(t, u) &= D_{ik}(t, u) - v_i(t) \hat{v}_k(t), \\ F_{ik}^W(t) &= D_{ik}^W(t) - \tilde{W} v_i(t) \hat{v}_k(t), \end{aligned} \quad (3.11)$$

where \tilde{W} is the integral value of the kernel function W . We finally obtain

$$\dot{K}_{ik}(t) \simeq \eta \left[w^{\text{in}} \hat{v}_k(t) + w^{\text{out}} v_i(t) + \tilde{W} \hat{v}_k(t) v_i(t) + F_{ik}^W(t) \right]. \quad (3.12)$$

Equation (3.12) clearly shows that STDP extends rate-based learning rules with the additional term F_{ik}^W .

In a similar way to the previous derivation for the input weights, we obtain a differential equation in the expectation value of the recurrent weight $\dot{J}_{ij}(t)$

$$\dot{J}_{ij}(t) \simeq \eta \left[w^{\text{in}} v_j(t) + w^{\text{out}} v_i(t) + \tilde{W} v_j(t) v_i(t) + C_{ij}^W(t) \right], \quad (3.13)$$

where we have used the approximation $v_j(t - d_{ij}) \simeq v_j(t)$. We have also assumed that the spike trains of neurons i and j are only weakly dependent from a probabilistic point of view; this is a reasonable assumption when each neuron receives many inputs (Burkitt et al. 2007).

The separation of time scales between the “fast” neuronal and synaptic activation mechanisms on the one hand, and the “slow” learning dynamics ($\eta \ll 1$) on the other hand allows us to capture the evolution of the network activity. Under this assumption, the neuron firing rates $v_i(t)$ can be expressed in terms of the weights $K_{ik}(t)$ and $J_{ij}(t)$ and the input firing rates $\hat{v}_k(t)$; and likewise for the covariance coefficients $D_{ik}^W(t)$ with the input covariance coefficient $\hat{C}_{kl}(t, u)$ in Fig. 3.1(b). This concept is used in the remainder

of this section to rewrite Eq. (3.8) as a dynamical system of the general form

$$[\dot{K}_{ik}(t), \dot{J}_{ij}(t)] = \mathbb{F} [K_{ik}(t), J_{ij}(t), v_0, \hat{v}_k(t), \hat{C}_{kl}(t, u)] . \quad (3.14)$$

This system of matrix equations is then used to predict the asymptotic evolution of the weights $K_{ik}(t)$ and $J_{ij}(t)$, depending on the input parameters $\hat{v}_k(t)$ and $\hat{C}_{kl}(t, u)$.

3.4 Derivation of the network consistency equations

We now derive consistency equations to express $v(t)$ and $F^W(t)$ in terms of the input parameters $\hat{v}(t)$ and $\hat{C}^W(t)$, as well as the synaptic weights $K(t)$ and $J(t)$. These equations describe the constraint of the recurrent connectivity on the neuronal activity.

3.4.1 Short duration of the PSP kernel and of the recurrent delays

Assuming that the weights $K(t)$ and $J(t)$ are quasi-stationary compared to the time scale of the PSP kernel ϵ and delays (Kempster et al. 1999), we take the ensemble average of Eq. (2.1) for neuron i to obtain

$$\langle S_i(t) \rangle = \langle \rho_i(t) \rangle = v_0 + \sum_{j \neq i} J_{ij}(t) \langle \epsilon * S_j(t - d_{ij}) \rangle + \sum_k K_{ik}(t) \langle \epsilon * \hat{S}_k(t - \hat{d}_{ik}) \rangle , \quad (3.15)$$

where $*$ denotes the convolution operation. Because T is very large compared to the neuronal time scale (ϵ and the delays), the integral of $\langle \epsilon * S_j(t - d_{ij}) \rangle$ over the time interval $[t - T, t]$ can be approximated by the integral of $\langle S_j(t) \rangle$ over the same time interval (recall that $\int \epsilon(t) dt = 1$). As a result we obtain the same matrix self-consistency equation as Burkitt et al. (2007) for the firing rates

$$\mathbf{v}(t) = [\mathbb{1}_N - J(t)]^{-1} [v_0 \mathbf{e} + K(t) \hat{\mathbf{v}}(t)] , \quad (3.16)$$

where $\mathbb{1}_N$ is the identity matrix of size N and \mathbf{e} is the column vector with N elements all equal to one (\mathbf{T} denotes the matrix transposition)

$$\mathbf{e} := [1, \dots, 1]^{\mathbf{T}}. \quad (3.17)$$

To ensure the stability of the firing rates, the matrix of the recurrent weights $J(t)$ must have all eigenvalues in the unit circle (modulus strictly smaller than one) at all times (Burkitt et al. 2007).

We now derive a consistency equation similar to Eq. (3.16) for the neuron-to-input covariance F defined in Eq. (3.3). The case of non-identical delays could be rigorously dealt with using Fourier analysis (Hawkes 1971), as shown in Appendix A.2.4. However, this method does not lead to an easily tractable solution for arbitrary PSP kernel ϵ and distribution of delays. In the present study, we consider the simplified case where all the recurrent delays are almost identical, i.e., $d_{ij} \simeq d$ for all connections $j \rightarrow i$, and likewise the input delays satisfy $\hat{d}_{ik} \simeq \hat{d}$ for all connections $k \rightarrow i$. The impact of the PSP kernel ϵ and of the recurrent delays d_{ij} can be evaluated when their two distributions are narrow in comparison to the width of the learning window W , as detailed in Appendix A.2.7, which gives

$$F^W(t) = [\mathbb{1}_N - J(t)]^{-1} K(t) \left[\hat{C}^{W*\epsilon}(t) + [W * \epsilon](0) \text{diag}(\hat{\mathbf{v}}(t)) \right]. \quad (3.18)$$

The input covariance structure \hat{C} is filtered by the PSP kernel ϵ to obtain the neuron-to-input covariance F : the effect of STDP embodied in F^W involves $\hat{C}^{W*\epsilon}$ (Kempster et al. 1999, Sprekeler et al. 2007). As a comparison, Burkitt et al. (2007) neglected this effect and used \hat{C}^W instead. The use of the covariance coefficient F instead of D (Burkitt et al. 2007) sheds a clearer light on the relationship between the neuron-to-input and the input-to-input correlation structures: the network connectivity operates on the input covariance \hat{C} through the term $(\mathbb{1}_N - J)^{-1} K$. The following approximation is made in deriving these

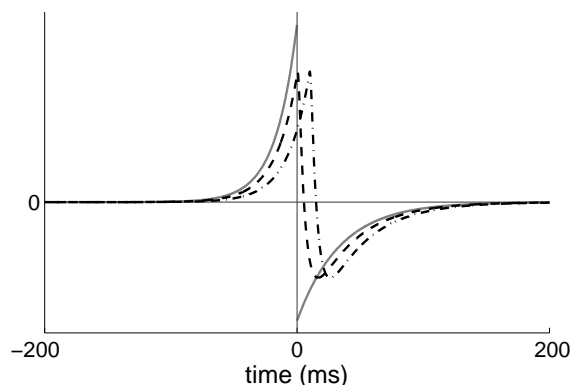


Figure 3.2: Impact of the PSP kernel ϵ on learning with the STDP window function W . The solid line represents the function W and the dashed line its convolution $W * \epsilon_d$ with the PSP kernel ϵ delayed by $d = 0.4$ ms. Globally, the shape of $W * \epsilon$ is similar to that of W , but for small $u > 0$ we have $W(u) < 0$ whereas $W * \epsilon(u) > 0$. Also plotted in dashed-dotted line is $W * \epsilon_d$ for longer delay $d = 10$ ms: increasing d shifts the curve of $W * \epsilon_d$ to the right and increases the discrepancies between the two curves. See Appendix D for details on the parameters.

equations and its accuracy is illustrated in Fig. 3.2:

$$\int W(u-r)\epsilon(r-d)dr \simeq W(u). \quad (3.19)$$

The derivation of the consistency equation for C^W is detailed in Appendix A.3. Similar to the derivation for F^W , the effect of the PSP kernel ϵ and of the recurrent delays d_{ij} can be evaluated when their two distributions in time are narrow in comparison to the width of the learning window W

$$\begin{aligned} C^W(t) & \quad (3.20) \\ &= [\mathbf{1}_N - J(t)]^{-1} K(t) \left[\hat{C}^{W*\zeta}(t) + [W * \zeta](0) \text{diag}(\hat{\mathbf{v}}(t)) \right] K^T(t) [\mathbf{1}_N - J(t)]^{-1\mathbf{T}} \\ & \quad + [\mathbf{1}_N - J(t)]^{-1} W(d) \text{diag}(\mathbf{v}(t)) [\mathbf{1}_N - J(t)]^{-1\mathbf{T}} - W(d) \text{diag}(\mathbf{v}(t)). \end{aligned}$$

Equation Eq. (3.20) describes a spatial and temporal filtering on the input covariance \hat{C} to obtain the neuron covariance C . The network connectivity operates through the term $(\mathbf{1}_N - J)^{-1} K$ that appears twice in Eq. (3.20); the same term was found in the consistency equation Eq. (3.18) for the neuron-to-input covariance F , where it appears once. The function ζ describes the temporal filtering of the PSP kernel function on \hat{C} to obtain C , as does ϵ for F ; this effect was ignored in a previous study by Burkitt et al. (2007). The

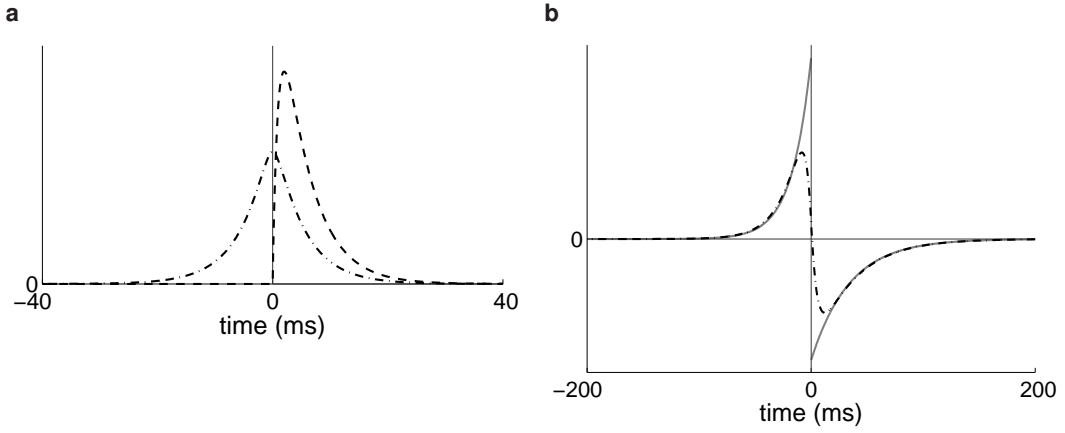


Figure 3.3: (a) Plots of ϵ (dashed line) and ζ (dashed-dotted line). (b) Plots of W (grey thick solid line) and $W * \zeta$ (black dashed-dotted line). Globally, the shapes of the two functions are similar, except for small u . We used the parameters listed in Appendix D, which correspond to $W(u) < 0$ but $[W * \zeta](u) > 0$ for very small $u > 0$. The value $[W * \zeta](0) > 0$ relates to the effect of delta-correlated inputs, cf. Eq. (3.29).

function ζ can be approximated by the self-convolution of ϵ , as illustrated in Fig. 3.3(a),

$$\zeta(r) \simeq \int \epsilon(r + r')\epsilon(r') \, dr' ; \quad (3.21)$$

see Eq. (A.39) in Appendix A.3.3 for details. The same approximation Eq. (3.19) as in the derivation for F has been made.

3.4.2 The equations describing the dynamical system

In the limit of large networks ($N \gg 1$ and $M \gg 1$) with sufficiently many synapses per neuron, we can ignore the effects due to autocorrelation, i.e., the term involving ‘diag’ in Eqs. (3.18) and (3.20). The system of equations that describe the dynamics of the firing

rates, covariance coefficients and weights reduces to

$$\boldsymbol{v} = (\mathbf{1}_N - J)^{-1} (v_0 \mathbf{e} + K \hat{\boldsymbol{v}}), \quad (3.22a)$$

$$F^W = (\mathbf{1}_N - J)^{-1} K \hat{C}^{W*\epsilon}, \quad (3.22b)$$

$$C^W = (\mathbf{1}_N - J)^{-1} K \hat{C}^{W*\zeta} K^T (\mathbf{1}_N - J)^{-1T}, \quad (3.22c)$$

$$\dot{K} = \eta \Phi_K (w^{\text{in}} \mathbf{e} \hat{\boldsymbol{v}}^T + w^{\text{out}} \boldsymbol{v} \hat{\mathbf{e}}^T + \tilde{W} \boldsymbol{v} \hat{\boldsymbol{v}}^T + F^W) \quad (3.22d)$$

$$\dot{J} = \eta \Phi_J (w^{\text{in}} \mathbf{e} \boldsymbol{v}^T + w^{\text{out}} \boldsymbol{v} \mathbf{e}^T + \tilde{W} \boldsymbol{v} \boldsymbol{v}^T + C^W). \quad (3.22e)$$

The time variable t has been omitted from all the vectors and matrices that evolve over time. Recall that hats indicate variables related to the external inputs. The constant \tilde{W} denotes the integral of the STDP window function W

$$\tilde{W} := \int_{-\infty}^{+\infty} W(u) \, du. \quad (3.23)$$

The column vectors $\hat{\mathbf{e}}$ and \mathbf{e} have all elements equal to one Eq. (3.17). The projectors Φ_K and Φ_J operate on the vector spaces of $N \times M$ and $N \times N$ matrices, respectively; they nullify the matrix components that correspond to non-existent connections in the network, in particular the diagonal terms for Φ_J .

3.4.3 Higher-order stochastic effects of the weight dynamics

The system of equations (3.22a-3.22e) describes the evolution of their expectation values, i.e., the first order of the stochastic process. In the remainder of this thesis, we refer to this leading order as the *drift* of the dynamics, in comparison to *higher orders* of the stochastic process. Phenomena such as evolution of the weight variance or symmetry breaking rely upon higher-order stochastic mechanisms and are not captured by Eqs. (3.22a-3.22e) but they can nevertheless be analyzed using this formalism (Kempster et al. 1999, Burkitt et al. 2007).

In order to evaluate the second moment of the weight dynamics, we need to consider single stochastic trajectories (realisations of the random process). The learning equation

Eq. (3.6) can be rewritten

$$\frac{dK_{ik}^{\omega}(t)}{dt} = \left[w^{\text{in}} \hat{S}_k(t - \hat{d}_{ik}) + w^{\text{out}} S_i(t) + \int W(u) S_i(t) \hat{S}_k(t + u - \hat{d}_{ik}) du \right], \quad (3.24)$$

where ω denotes a given stochastic trajectory of the process. Using the expression in Eq. (3.24), we can evaluate the multidimensional matrix coefficient

$$Y_{i,k,j,l}(t, t') := \left\langle \frac{dK_{ik}^{\omega}(t)}{dt} \frac{dK_{jl}^{\omega}(t')}{dt} \right\rangle, \quad (3.25)$$

which is related to the second moment of the weight dynamics (cf. Appendix B.3.1). The ensemble average denoted by the brackets in Eq. (3.25) is performed over all stochastic trajectories ω . In comparison, the system of equations Eq. (3.22a-3.22e) describes the expectation value of the expression Eq. (3.24) over all the trajectories, namely the drift of the weight K_{ik}

$$\dot{K}_{ik}(t) = \left\langle \frac{dK_{ik}^{\omega}(t)}{dt} \right\rangle. \quad (3.26)$$

3.5 Generation of the input spike trains

We constrain this study to a specific type of external inputs with correlations, although the present framework can be applied to arbitrary configurations. The inputs that stimulate the network are partitioned into a predetermined number of homogeneous pools, such that inputs from the same pool are correlated but independent of inputs from different pools. The firing rates of inputs within a pool are all equal to, say, $\hat{\nu}_0$. The positive within-pool correlation is generated so that, for any input, a given portion of its spikes occur at the same time as some other spikes within its pool, while the remainder occur at independent times (Gütig et al. 2003, Meffin et al. 2006).

The spikes from input k are selected from two homogeneous Poisson spike trains each of rate $\hat{\nu}_0$ such that the within-group correlation strength is $0 \leq \hat{c} \leq 1$ (Meffin et al. 2006). The first spike train is common to all inputs in the pool and generates the correlated events; distinct pools have different common reference spike trains. For a given input, the spikes are selected from this train with probability $\sqrt{\hat{c}}$, independently of other neurons in

the pool. Thus only a portion of all the neurons in the pool participate in each correlated event. The second spike train is the own independent train attached to each input and the spikes are selected from this train with probability $1 - \sqrt{\hat{c}}$. For each input k , we create a random variable $\hat{X}_k(t)$ that is one if there is an input spike at time t and zero otherwise. The correlation between the variables $\hat{X}_k(t)$ and $\hat{X}_l(t)$ corresponding to distinct sources from the same pool is given by (Gütig et al. 2003)

$$\frac{\text{Cov}[\hat{X}_k(t), \hat{X}_l(t)]}{\sqrt{\text{Var}[\hat{X}_k(t)] \text{Var}[\hat{X}_l(t)]}} = \hat{c}. \quad (3.27)$$

In this way, we obtain input spike trains $\hat{S}_k(t)$ that have ‘‘instantaneous’’ firing rates $\langle \hat{S}_k(t) \rangle = \hat{v}_0$ and pairwise covariances (for $k \neq l$)

$$\hat{C}_{kl}(t, u) \simeq \hat{c} \hat{v}_0 \delta(u) \quad (3.28)$$

that are both constant in time. The latter follows since Eq. (3.27) implies $\text{Cov}[\hat{S}_k(t), \hat{S}_l(t + u)] \simeq \hat{c} \hat{v}_0 \delta(u)$, as defined by Eq. (A.1) in Appendix A.1.1. Inputs from the same pool are only correlated for $u = 0$, which we denote as ‘delta-correlated’ inputs. We only consider positively delta-correlated inputs. It follows that, for inputs $k \neq l$,

$$\hat{C}_{kl}^{W*\epsilon} \simeq \hat{c} \hat{v}_0 [W * \epsilon](0). \quad (3.29)$$

Due to the PSP kernel, delta-correlated inputs induce a non-trivial (richer) correlation structure F in the network according to Eq. (3.18). Under the Hebbian assumption that $W(u) > 0$ for $u < 0$ (cf. Sec. 2.3),

$$[W * \epsilon](0) = \int W(u) \epsilon(-u) du > 0 \quad (3.30)$$

and the matrix $\hat{C}^{W*\epsilon}$ has non-negative elements for delta-correlation. Eq. (3.29) also implies that the spike-triggering effect for F^W , involving diag in the rhs of Eq. (3.18), is always positive, as explained in Appendix A.2.6. This is similar to the case of a feed-forward architecture (Kempster et al. 1999).

Typically, we use *small* input correlations: $0 \leq \hat{c} \leq 10^{-1}$. Within this range of *small* correlation strengths, we will discriminate between *weak* and *sufficiently strong* values when discussing their impact on the weight dynamics. Numerical simulation uses discrete time to generate the Poisson spike trains (Appendix D).

3.6 Analysis of the system dynamics

Our aim is to investigate the steady states of the firing rates and weights. For this purpose, the dynamical system is analyzed in terms of fixed point and stability in order to predict the asymptotic behaviour of the weights. This framework targets network dynamics beyond the mean-field approach in order to study the emergence of a network structure due to external stimulation. In particular, we focus on the emergence of an asymptotic weight structure in a network stimulated by two input pools, an idea inspired by Kempter et al. (1999) and Gütig et al. (2003). Minimal assumptions are made about the network connectivity (partial or full).

The term *mean* (applied to firing rates and weights) will refer to an average over the neurons, inputs, connections, etc. of the network (topological averaging), whereas *averaged* stands for time averaging, unless otherwise specified. The *homeostatic equilibrium* describes the situation where the mean firing rate and mean weight have reached an equilibrium, although individual firing rates and weights may continue to change. The expression *emergence of weight structure* will refer to the situation where the learning dynamics has imposed a specific weight structure on the network, i.e., further learning may cause individual weights to change but the qualitative character of the distribution (e.g., bimodal) will remain unchanged.

When using additive STDP (cf. Sec. 2.3 and Eq. (D.1) in Appendix D), it is necessary to introduce bounds on the input weights in numerical simulation because of their tendency to diverge due to the competition induced by STDP (Kempter et al. 1999, Burkitt et al. 2007). In this way, we focus on the splitting of the weight distribution and leave aside the stabilisation issue. The issue of stabilisation will be addressed in Chapter 6 when using weight-dependent STDP. The simulation results presented in this thesis were run using

the neuron and learning parameters listed in Appendix D.

Chapter 4

Input selectivity by STDP

In this chapter, the weight dynamics are investigated in a recurrently connected network stimulated by external inputs, where STDP modifies the plastic input connections while the recurrent weights are kept fixed.

4.1 Introduction

THIS chapter investigates the situation illustrated in Fig. 4.1 where additive STDP only modifies the input connections (thick arrows), while the recurrent weights are kept fixed (thin arrows). The dynamical system (3.22a-3.22e) reduces to

$$\mathbf{v} = (\mathbb{1}_N - J)^{-1} (\nu_0 \mathbf{e} + K \hat{\mathbf{v}}), \quad (4.1a)$$

$$F^W = (\mathbb{1}_N - J)^{-1} K \hat{C}^{W*\epsilon}, \quad (4.1b)$$

$$\dot{K} = \Phi_K(w^{\text{in}} \mathbf{e} \hat{\mathbf{v}}^T + w^{\text{out}} \mathbf{v} \hat{\mathbf{e}}^T + \tilde{W} \mathbf{v} \hat{\mathbf{v}}^T + F^W). \quad (4.1c)$$

Time has been rescaled to remove η . The study of the weight dynamics focuses on a particular network configuration where the neuronal network (top circles) are excited by two pools of input spike trains that have homogeneous within-pool firing rates and spike-time correlations (bottom circles, fill-in indicates within-pool correlations), an idea inspired by Kempster et al. (1999) and Gütig et al. (2003). Minimal assumptions are made about the network connectivity (partial or full) and input structure.

We examine whether STDP can induce a homeostatic equilibrium on the input weights, in which the mean pre-synaptic input weight and the output firing rate stabilise over time, as well as a potentiation of some input weights depending on the input correla-

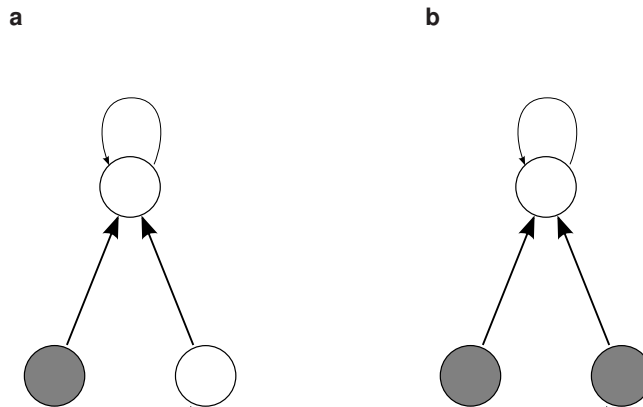


Figure 4.1: Network configurations studied in chapter 4. Top circles represent the neuronal network and bottom circles the two input pools, for which filled circles indicate non-zero within-pool correlation. Thick (resp. thin) arrows indicate plastic (fixed) weights.

tion structure. We especially investigate the differences between the present case of a recurrently connected network and previous results of a single neuron or a feed-forward network: does STDP always select the more correlated input pools? Sec. 4.4 focuses on the particular case illustrated in Fig. 4.1(b) where the network is stimulated by two input pools whose spike trains have similar characteristics, namely homogeneous initial input weights, identical firing rates and equal within-pool spike-time correlations. This relates to the concept of symmetry breaking, by which we mean the specialisation of the neurons to just *one* of the input pools. It has been previously demonstrated that STDP can implement this symmetry breaking for a single neuron stimulated by two input pools having the same firing rate and spike-time correlation: STDP causes a neuron with initially homogeneous input weights to become sensitive to *only one* of the two pools (Gütig et al. 2003). This weight specialisation can be related, for example, to the organisation of the primary visual cortex into areas sensitive to different aspects of visual stimuli, viz., ocular dominance and orientation fields (Hubel and Wiesel 1962). Possible underlying mechanisms have been the subject of many studies (von der Malsburg 1973, Swindale 1996, Choe and Miikkulainen 1998, Elliott and Shadbolt 1999, Wensich et al. 2005, Goodhill 2007) to reproduce and explain such synaptic self-organisation (Kohonen 1982). We extend the previous studies by Kempter et al. (1999) and Gütig et al. (2003) to the case of a *recurrently* connected network stimulated by two input pools as described above.

4.2 General case of learning input weights with fixed recurrent weights

4.2.1 Homeostatic equilibrium

The evolution equation of the mean input weight $K_{av} := \sum_{i,k} K_{ik} / n^K$ is given by

$$\begin{aligned} \dot{K}_{av} &= w^{\text{in}} \hat{v}_{av} + w^{\text{out}} \nu_{av} + \tilde{W} \hat{v}_{av} \nu_{av} + F_{av}^W \\ &= w^{\text{in}} \hat{v}_{av} + \frac{\nu_0 (w^{\text{out}} + \tilde{W} \hat{v}_{av})}{1 - n_{av}^J J_{av}} + n_{av}^K K_{av} \frac{\hat{v}_{av} (w^{\text{out}} + \tilde{W} \hat{v}_{av}) + \hat{C}_{av}^{W*\epsilon}}{1 - n_{av}^J J_{av}}. \end{aligned} \quad (4.2)$$

The subscript 'av' denotes the mean-averaged variable over the network, i.e., when neglecting the discrepancies among the external inputs, the neurons or the connections. We have

$$\nu_{av} = \frac{\nu_0 + n_{av}^K K_{av} \hat{v}_{av}}{1 - n_{av}^J J_{av}}. \quad (4.3)$$

The constants $n_{av}^K := n^K / N$ and $n_{av}^J := n^J / N$ denote the mean number of pre-synaptic input and recurrent connections (resp.) in the network. Eq. (4.2) is linear in K_{av} , which converges towards an equilibrium if and only if

$$\frac{\hat{v}_{av} (w^{\text{out}} + \tilde{W} \hat{v}_{av}) + \hat{C}_{av}^{W*\epsilon}}{1 - n_{av}^J J_{av}} < 0. \quad (4.4)$$

We have $1 - n_{av}^J J_{av} > 0$ since the matrix J has a spectrum in the unit circle (to prevent firing rates from diverging), so the mean recurrent feedback $n_{av}^J J_{av}$ does not change the stability of the network. For weakly correlated inputs, the covariance coefficient $\hat{C}_{av}^{W*\epsilon}$ is small compared to the mean stimulation firing rate \hat{v}_{av} (Kempster et al. 1999) and the previous stability condition reduces to

$$w^{\text{out}} + \tilde{W} \hat{v}_{av} < 0. \quad (4.5)$$

Consequently, there are four situations to consider for the homeostatic equilibrium, depending on the mean input stimulation \hat{v}_{av} :

- (i) $\tilde{W} < 0$ and $w^{\text{out}} < 0$: stable whatever the value of \hat{v}_{av} ;
- (ii) $\tilde{W} < 0$ and $w^{\text{out}} > 0$: stable for $\hat{v}_{\text{av}} > -w^{\text{out}}/\tilde{W}$;
- (iii) $\tilde{W} > 0$ and $w^{\text{out}} < 0$: stable for $\hat{v}_{\text{av}} < -w^{\text{out}}/\tilde{W}$;
- (iv) $\tilde{W} > 0$ and $w^{\text{out}} > 0$: never stable for any value of \hat{v}_{av} .

The input stimulation can thus change the stability in some cases, unlike the recurrent feedback. We recall that $\tilde{W} < 0$ in cases (i) and (ii) above leads to homeostatic stability of the learning dynamics when STDP modifies only the recurrent weights in a network with no external inputs (Burkitt et al. 2007). Case (iii) corresponds to a stability analysis already elsewhere described (Kempster et al. 1999). The simulation parameters used in numerical simulations (see Appendix D) correspond to case (i).

As a consequence of Eq. (4.2), the asymptotic value of K_{av} is given by the fixed point (if it is stable)

$$K_{\text{av}}^* = \frac{-1}{n_{\text{av}}^K} \frac{(1 - n_{\text{av}}^J J_{\text{av}}) w^{\text{in}} \hat{v}_{\text{av}} + v_0 (w^{\text{out}} + \tilde{W} \hat{v}_{\text{av}})}{\hat{v}_{\text{av}} (w^{\text{out}} + \tilde{W} \hat{v}_{\text{av}}) + \hat{C}_{\text{av}}^{W^* \epsilon}}. \quad (4.6)$$

Since we require the weights K to remain positive, the equilibrium is realisable only if the asymptotic value K_{av}^* is positive, which requires, similar to the case of a single neuron (Kempster et al. 1999),

$$w^{\text{in}} > -\frac{v_0 (w^{\text{out}} + \tilde{W} \hat{v}_{\text{av}})}{\hat{v}_{\text{av}} (1 - n_{\text{av}}^J J_{\text{av}})} > 0. \quad (4.7)$$

When the fixed point $K_{\text{av}}^* < 0$ is stable, the input weights $K(t)$ will all become quiescent. The presence of strong recurrent feedback $n_{\text{av}}^J J_{\text{av}}$ can thus cause the homeostatic equilibrium to become non-realisable. This condition is also consistent with the stability analysis in the case of learning on the recurrent weights J with no external inputs, for which $w^{\text{in}} \gg |w^{\text{out}}|$ ensures the stability of individual firing rates (Burkitt et al. 2007). Note that the particular case where $w^{\text{out}} \rightarrow 0$ does not impair the equilibrium provided $w^{\text{in}} > 0$ and $\tilde{W} < 0$.

From Eqs. (4.6) and (4.3), the fixed point ν_{av}^* of the mean firing rate is given by

$$\nu_{\text{av}}^* = \frac{-w^{\text{in}}\hat{\nu}_{\text{av}}^2 + \nu_0 \left(1 - n_{\text{av}}^J J_{\text{av}}\right)^{-1} \hat{C}_{\text{av}}^{W^* \epsilon}}{\hat{\nu}_{\text{av}} \left(w^{\text{out}} + \tilde{W}\hat{\nu}_{\text{av}}\right) + \hat{C}_{\text{av}}^{W^* \epsilon}}. \quad (4.8)$$

For weakly correlated inputs, it reduces to

$$\nu_{\text{av}}^* \simeq -\frac{w^{\text{in}}\hat{\nu}_{\text{av}}}{w^{\text{out}} + \tilde{W}\hat{\nu}_{\text{av}}}. \quad (4.9)$$

From Eq. (4.9), we see that the fixed point ν_{av}^* is an increasing function of the mean input firing rate $\hat{\nu}_{\text{av}}$ when $w^{\text{out}} < 0$ (decreasing otherwise).

4.2.2 Emergence of a weight structure

The learning equation Eq. (4.1c) can be rewritten as a linear differential matrix equation in K ,

$$\dot{K} = \Phi_K \left[(\mathbf{1}_N - J)^{-1} KA + B \right] \quad (4.10)$$

with the two following matrices containing the input firing-rate and correlation structures

$$\begin{aligned} A &:= w^{\text{out}}\hat{\nu}\hat{\mathbf{e}}^{\mathbf{T}} + \tilde{W}\hat{\nu}\hat{\nu}^{\mathbf{T}} + \hat{C}^{W^* \epsilon}, \\ B &:= w^{\text{in}}\mathbf{e}\hat{\nu}^{\mathbf{T}} + (\mathbf{1}_N - J)^{-1} \nu_0\mathbf{e} \left(w^{\text{out}}\hat{\mathbf{e}}^{\mathbf{T}} + \tilde{W}\hat{\nu}^{\mathbf{T}} \right). \end{aligned} \quad (4.11)$$

We denote by \mathbb{M}_K the subspace of $\mathbb{R}^{N \times M}$ where the matrix K evolves, i.e., the vector subspace of matrices X such that $\Phi_K(X) = X$.

We examine the solution of Eq. (4.10), first for the case of full input connectivity (Φ_K is the identity), which depends upon the invertibility of the matrix A . We then complete the general analysis for partial connectivity (and any matrix A), which will be illustrated though a specific network example in Sec. 4.3.

Full input connectivity and invertible A

If the matrix A is invertible, the solution of Eq. (4.10) is given by

$$K(t) = K(\infty) + \sum_{n \geq 0} \frac{t^n}{n!} (\mathbb{1}_N - J)^{-n} [K(0) - K(\infty)] A^n \quad (4.12)$$

with the fixed point

$$K(\infty) = -(\mathbb{1}_N - J) B A^{-1} \quad (4.13)$$

and $K(0)$ the initial weight matrix at $t = 0$. Note that without rescaling time, t would be replaced by ηt in Eq. (4.12). The weight stability of $K(\infty)$ is determined by the eigenvalues of A since the spectrum of $\mathbb{1}_N - J$ lies in the unit circle for the sake of bounded (non-diverging) firing rates. In the same way as with the homeostatic equilibrium, the presence of recurrent connections affects the asymptotic weight matrix $K(\infty)$ as well as the rate of convergence (or divergence) of $K(t)$. Note that the spike-triggering effect, which we neglected, only adds the diagonal matrix $[W * \epsilon](0) \text{diag}(\hat{v})$ to B , which would slightly change the fixed point $K(\infty)$, but not the stability.

The weight matrix $K(t)$ converges exponentially fast towards $K(\infty)$ when the eigenvalues of A have negative real parts. The fixed point $K(\infty)$ may then be attained depending on the weight bounds. On the other hand, if A has any eigenvalue with positive real part, some components of K diverge in the direction of the principal eigenvector until hitting the bounds. Then, the relative position of the initial conditions $K(0)$ compared to that of the fixed point $K(\infty)$ in \mathbb{M}_K will determine the evolution of $K(t)$. A combination of stability and divergence gives interesting dynamics, as was shown by Kempter et al. (1999) for the single-neuron case: the first corresponds to partial equilibria (e.g., homeostatic) while the second can imply robust weight specialisation through a splitting of their distribution.

Partial input connectivity and/or non-invertible A

The matrix A is not invertible whenever there are symmetries in the input pools and in the weights K ; for example, in the case of homogeneous input pools. Details of the

general analysis are provided in Appendix B.1. In summary, we can decompose the evolution of K into three subspaces defined using the null-spaces of the matrices A and B :

- an exponential evolution (convergence in the stable case) on a subspace where “ A is invertible”, similar to the solution given in Eq. (4.12);
- a zero drift on a subspace related, for example, to symmetries of the input pools and of the input connectivity, where higher stochastic orders of the weight dynamics have a significant effect;
- a constant drift that drives weights towards their bounds in a particular direction (this case corresponds in general to very specific parameter values and we ignore it).

When the network has symmetries, the weight drift in the first subspace can be studied using a reduction of dimensionality for $K(t)$ as explained in Appendix B.1.1. In the second subspace, the weight evolution is not constrained by STDP in a way that organises the input weights and, consequently, does not correspond to learning of the input firing-rate and correlation structure. It can nevertheless be the source of organisation in the network, as will be described Sec. 4.4.

4.3 Network stimulated by two homogeneous input pools

We now illustrate the general analysis in Sec. 4.2, in particular how the input correlations determine the asymptotic weight structure, through a specific network example inspired by Kempter et al. (1999). The network is stimulated by external inputs that are divided into two homogeneous pools of the same size (indices $1 \leq k \leq M/2$ vs. $M/2 + 1 \leq k \leq M$); the two input pools may have distinct parameters, as illustrated in Fig. 4.1(a). The analysis below is carried out for full input connectivity for the sake of simplicity, but simulations show corresponding cases with partial connectivity.

4.3.1 Reduction of dimensionality to study the weight drift

The vector $\hat{\mathbf{v}}$ and the matrix $\hat{\mathbf{C}}^{W*\epsilon}$ that appear in A and B , cf. Eq. (4.11), can be expressed in terms of the M -column vector $\hat{\mathbf{e}}$, defined similarly to \mathbf{e} in Eq. (3.17),

$$\hat{\mathbf{e}} := [1, \dots, 1]^T, \quad (4.14)$$

and the M -column vector $\hat{\mathbf{h}}$, whose first $M/2$ elements are 1 and last $M/2$ elements are -1 :

$$\hat{\mathbf{h}} := [1, \dots, 1, -1, \dots, -1]^T. \quad (4.15)$$

Denoting the mean firing rates by \bar{v}_1 and \bar{v}_2 for each input pool and their correlation strengths by \hat{c}_1 and \hat{c}_2 , we have

$$\begin{aligned} \hat{\mathbf{v}} &= \frac{\bar{v}_1}{2} (\hat{\mathbf{e}} + \hat{\mathbf{h}}) + \frac{\bar{v}_2}{2} (\hat{\mathbf{e}} - \hat{\mathbf{h}}), \\ \hat{\mathbf{C}}^{W*\epsilon} &= \frac{\hat{c}_1 \bar{v}_1 [W * \epsilon](0)}{4} (\hat{\mathbf{e}} + \hat{\mathbf{h}})(\hat{\mathbf{e}} + \hat{\mathbf{h}})^T + \frac{\hat{c}_2 \bar{v}_2 [W * \epsilon](0)}{4} (\hat{\mathbf{e}} - \hat{\mathbf{h}})(\hat{\mathbf{e}} - \hat{\mathbf{h}})^T, \end{aligned} \quad (4.16)$$

where we have used Eq. (3.29). Substituting the above expressions into Eq. (4.11), we obtain the following special form for the matrices A and B

$$\begin{aligned} A &= \alpha \hat{\mathbf{e}} \hat{\mathbf{e}}^T + \beta \hat{\mathbf{h}} \hat{\mathbf{h}}^T + \gamma \hat{\mathbf{e}} \hat{\mathbf{h}}^T + \kappa \hat{\mathbf{h}} \hat{\mathbf{e}}^T, \\ B &= \alpha' \mathbf{e} \mathbf{e}^T + \beta' (\mathbf{1}_N - J)^{-1} \mathbf{e} \mathbf{e}^T + \gamma' \mathbf{e} \mathbf{h}^T + \kappa' (\mathbf{1}_N - J)^{-1} \mathbf{e} \mathbf{h}^T, \end{aligned} \quad (4.17)$$

where the constants $\alpha, \beta, \gamma, \kappa, \alpha', \beta', \gamma', \kappa'$ absorb all the input and learning parameters,

$$\begin{aligned} \alpha &= w^{\text{out}} \frac{\bar{v}_1 + \bar{v}_2}{2} + \tilde{W} \frac{(\bar{v}_1 + \bar{v}_2)^2}{4} + [W * \epsilon](0) \frac{\hat{c}_1 \bar{v}_1 + \hat{c}_2 \bar{v}_2}{4}, \\ \beta &= w^{\text{out}} \frac{\bar{v}_1 - \bar{v}_2}{2} + \tilde{W} \frac{\bar{v}_1^2 - \bar{v}_2^2}{4} + [W * \epsilon](0) \frac{\hat{c}_1 \bar{v}_1 - \hat{c}_2 \bar{v}_2}{4}, \\ \gamma &= \tilde{W} \frac{\bar{v}_1^2 - \bar{v}_2^2}{4} + [W * \epsilon](0) \frac{\hat{c}_1 \bar{v}_1 - \hat{c}_2 \bar{v}_2}{4}, \\ \kappa &= \tilde{W} \frac{(\bar{v}_1 - \bar{v}_2)^2}{4} + [W * \epsilon](0) \frac{\hat{c}_1 \bar{v}_1 + \hat{c}_2 \bar{v}_2}{4}, \end{aligned} \quad (4.18)$$

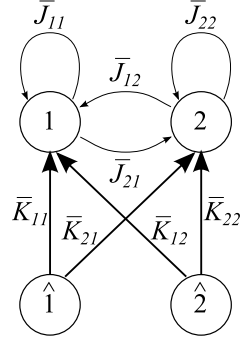


Figure 4.2: Dimension reduction for weight matrices K and J for a network of two homogeneous groups of neurons (top circles) stimulated by two homogeneous input pools (bottom circles). The variable \bar{K}_{11} , for example, corresponds to the mean input weights from pool $\hat{1}$ to group 1. The drift of the input weights can be completely studied using a reduced system of equations with \bar{K} and \bar{J} .

$$\begin{aligned}\alpha' &= w^{\text{in}} \frac{\tilde{v}_1 + \tilde{v}_2}{2}, \\ \beta' &= w^{\text{out}} v_0 + \tilde{W} v_0 \frac{\tilde{v}_1 + \tilde{v}_2}{2}, \\ \gamma' &= w^{\text{in}} \frac{\tilde{v}_1 - \tilde{v}_2}{2}, \\ \kappa' &= \tilde{W} v_0 \frac{\tilde{v}_1 - \tilde{v}_2}{2}.\end{aligned}$$

The drift $\dot{K}(t)$ is clearly zero in the subspace orthogonal to both $\hat{\mathbf{e}}$ and $\hat{\mathbf{h}}$. The interesting values to predict are the mean input weights for each neuron, embodied in the vector $K\hat{\mathbf{e}}/M$, and the difference between the mean weights from the first and the second pools, contained in $K\hat{\mathbf{h}}/M$. This reduction of dimensionality, explained in Appendix B.1.1, would still be valid for homogeneous partial input connectivity and input pools of different sizes. For the network configuration detailed in Fig. 4.2, the corresponding equivalence classes for the input weights to describe the drift $\dot{K}(t)$ are simply the mean input weights over each input pool and neuron group, such as \bar{K}_{21} from pool $\hat{1}$ to group 2.

In the following, we study the drift of the input weights using the two vectors $K\hat{\mathbf{e}}$ and

$K\hat{\mathbf{h}}$, which evolve over time according to

$$\dot{K}\hat{\mathbf{e}} = M (\mathbb{1}_N - J)^{-1} K (\alpha\hat{\mathbf{e}} + \beta\hat{\mathbf{h}}) + M\alpha' \mathbf{e} + M\beta' (\mathbb{1}_N - J)^{-1} \mathbf{e}, \quad (4.19a)$$

$$\dot{K}\hat{\mathbf{h}} = M (\mathbb{1}_N - J)^{-1} K (\gamma\hat{\mathbf{e}} + \kappa\hat{\mathbf{h}}) + M\gamma' \mathbf{e} + M\kappa' (\mathbb{1}_N - J)^{-1} \mathbf{e}. \quad (4.19b)$$

This reduced system evolves according to the eigenvalues of the matrix

$$A_r := \begin{pmatrix} \alpha & \beta \\ \gamma & \kappa \end{pmatrix}. \quad (4.20)$$

4.3.2 Firing-rate equilibrium for weak correlations

First, we constrain our study to weakly correlated inputs. In this case, we have $|\alpha| \gg |\beta|$ in Eq. (4.19a) and we can separate the evolution of $K\hat{\mathbf{e}}$ from that of $K\hat{\mathbf{h}}$, namely

$$\begin{aligned} \alpha &= \zeta + (\hat{c}_1 + \hat{c}_2)\hat{v}_{\text{av}} [W * \epsilon](0)/4 \simeq \zeta, \\ \zeta &:= w^{\text{out}}\hat{v}_{\text{av}} + \tilde{W}\hat{v}_{\text{av}}^2. \end{aligned} \quad (4.21)$$

Here ζ is much larger in absolute value than β (as well as γ and κ) in A_r , cf. Eqs. (4.18) and (4.20). It follows that $K\hat{\mathbf{e}}$ evolves much faster than $K\hat{\mathbf{h}}$, because $|\kappa| \ll |\alpha|$, in a similar way to the corresponding analysis for a single neuron (Kempster et al. 1999). We can thus consider $K\hat{\mathbf{e}}$ to be at its fixed point $K(\infty)\hat{\mathbf{e}}$ when stable, and then study the structure of the input weights through $K\hat{\mathbf{h}}$. The condition $\zeta < 0$ ensures the stability of $K(t)\hat{\mathbf{e}}$, i.e., of the mean input weight for each neuron, and is the same as the condition in Eq. (4.5). The equilibrium of all individual firing rates at v_{av}^* is equivalent to the stability of the mean input weight for each neuron at K_{av}^* .

In the remainder of Sec. 4.3, we consider small input correlations and require the stability of the firing rate for each neuron (embodied in $K\hat{\mathbf{e}}$) with an equilibrium value within the bounds, even if Eq. (4.21) is not strictly satisfied. Otherwise, input weights would end up clustered at a bound and the learning would then be null. When neglecting the inhomogeneities of J , the vector $K\hat{\mathbf{e}}$ can be approximated at the equilibrium by $K_{\text{av}}\mathbf{e}$ at

all times. However, discrepancies between the effective equilibrium value of $K\hat{\mathbf{e}}$ and the homeostatic equilibrium value $K_{\text{av}}^* \mathbf{e}$ may occur depending on the correlation strengths, weight bounds or inhomogeneities in the network and initial conditions; see Fig. 4.3 for an example.

The evolution of $K\hat{\mathbf{h}}$ described by Eq. (4.19b) corresponds to the fixed point

$$K(\infty)\hat{\mathbf{h}} = -\frac{\gamma}{\kappa} K(\infty)\hat{\mathbf{e}} - \frac{\gamma'}{\kappa} (\mathbf{1}_N - J) \mathbf{e} - \frac{\kappa'}{\kappa} \mathbf{e} \quad (4.22)$$

and it is determined by the sign of κ for the stability as well as the respective positions of $K(0)\hat{\mathbf{h}}$ and $K(\infty)\hat{\mathbf{h}}$. We now elucidate the dynamics of $K\hat{\mathbf{h}}$ depending on the input parameters.

4.3.3 Two uncorrelated input pools with any firing rates

For uncorrelated inputs with different firing rates, $\kappa = \tilde{W}(\tilde{\nu}_1 - \tilde{\nu}_2)^2/4$, so it follows from Eq. (4.19b) that $K\hat{\mathbf{h}}$ is stable when $\tilde{W} < 0$. The terms in the rhs of Eq. (4.22) are $\gamma/\kappa = 2\hat{\nu}_{\text{av}}/(\tilde{\nu}_1 - \tilde{\nu}_2)$, $\gamma'/\kappa = 2w^{\text{in}}/\tilde{W}(\tilde{\nu}_1 - \tilde{\nu}_2)$ and $\kappa'/\kappa = 2\nu_0/(\tilde{\nu}_1 - \tilde{\nu}_2)$, so the vector elements of the fixed point $K(\infty)\hat{\mathbf{h}}$ have the same sign for homogeneous recurrent connectivity, which is given by

$$-2 \frac{n_{\text{av}}^K K_{\text{av}}^* \hat{\nu}_{\text{av}} + w^{\text{in}}(1 - n_{\text{av}}^J J_{\text{av}})/\tilde{W} + \nu_0}{\tilde{\nu}_1 - \tilde{\nu}_2} = -\frac{2(1 - n_{\text{av}}^J J_{\text{av}})w^{\text{in}}}{\tilde{\nu}_1 - \tilde{\nu}_2} \frac{w^{\text{out}}/\tilde{W}}{w^{\text{out}} + \tilde{W}\hat{\nu}_{\text{av}}}. \quad (4.23)$$

We have used the expression of K_{av}^* in Eq. (4.6).

Since we required $K\hat{\mathbf{e}}$ to be stable, the conditions for homeostatic stability given in Sec. 4.2.1 must be satisfied: $w^{\text{in}} > 0$ and $w^{\text{out}} + \tilde{W}\hat{\nu}_{\text{av}} < 0$. Consequently, the sign of the vector elements of $K(\infty)\hat{\mathbf{h}}$ is the same as that of $w^{\text{out}}/\tilde{W}(\tilde{\nu}_1 - \tilde{\nu}_2)$. In both cases where the fixed point $K(\infty)\hat{\mathbf{h}}$ is stable or unstable, depending on the sign of \tilde{W} , the condition $w^{\text{out}} < 0$ corresponds to the potentiation of the input weights coming from the pool with stronger firing rate. Recall that the same condition $w^{\text{out}} < 0$ implies that the neuron firing rate ν_{av} increases with $\hat{\nu}_{\text{av}}$ at the homeostatic equilibrium (cf. Sec. 4.2.1).

4.3.4 The two input pools have correlations and the same input firing rate

We now consider the special case where both input pools have the same firing rate equal to \hat{v}_0 and the vector $\hat{v} = \hat{v}_0 \hat{\mathbf{e}}$ is homogeneous. We thus have $\gamma' = \kappa' = 0$ in Eq. (4.22), and the fixed point for $K\hat{\mathbf{h}}$ reduces to

$$K(\infty)\hat{\mathbf{h}} = -\frac{\gamma}{\kappa} K(\infty)\hat{\mathbf{e}} \simeq -\frac{\hat{c}_1 - \hat{c}_2}{\hat{c}_1 + \hat{c}_2} n_{\text{av}}^K K_{\text{av}}^* \mathbf{e}, \quad (4.24)$$

where K_{av}^* is given in Eq. (4.6). This fixed point is always unstable since $\kappa = [W * \epsilon](0)\hat{v}_0(\hat{c}_1 + \hat{c}_2)/4 > 0$, cf. Eq. (4.18), similar to the case of a single neuron (Kempton et al. 1999). This holds since $[W * \epsilon](0) > 0$ (see Sec. 3.5) and the correlation strengths \hat{c}_1 and \hat{c}_2 are positive. All the elements of the vector $K(\infty)\hat{\mathbf{h}}$ have the opposite sign to $\hat{c}_1 - \hat{c}_2$ when the equilibrium of the mean input weight for each neuron is realisable, viz., the vector $K(\infty)\hat{\mathbf{e}}$ has positive elements. It follows that the fixed point $K(\infty)\hat{\mathbf{h}}$ is determined by the balance between the input correlation strengths.

The instability will lead $K\hat{\mathbf{h}}$ to evolve in the opposite direction to the fixed point $K(\infty)\hat{\mathbf{h}}$. As a result, if the network starts with random initial input weights such that $K(0)\hat{\mathbf{h}} \simeq 0$, the weights coming from the input pool with stronger correlation will be potentiated compared to the weights from the other pool, as illustrated in Fig. 4.4(a). This is similar to that seen in the case of feed-forward architecture with $\tilde{W} > 0$ described by Kempton et al. (1999). It actually holds whatever the sign of \tilde{W} and the recurrent connections do not qualitatively change this behaviour. When the initial conditions correspond to $K(0)\hat{\mathbf{h}} \neq 0$ but not too far from 0, then regardless of their initial specialisation the input weights evolve into the “naturally” expected distribution, as shown in Fig. 4.3(a). Note that the homeostatic equilibrium did not hold asymptotically (discrepancy between the thin solid and dotted lines) in that simulation; this can happen when weights saturate and Eq. (4.2) breaks down.

However, if the input weights are initially already specialised such as $\bar{K}_{11}(0) \gg \bar{K}_{12}(0)$ then, despite $\hat{c}_2 > \hat{c}_1 = 0$, the initial “wrong” specialisation may be preserved, contrary to the expected potentiation of the weights from the more correlated input pool,

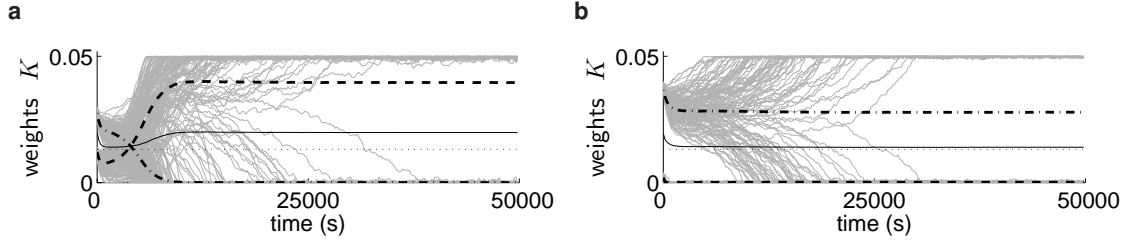


Figure 4.3: Comparison between the weight evolution of different initial conditions for the same network configuration. The network consisted of $N = 100$ neurons and two pools of $M = 100$ inputs each, for the topology described in Fig. 4.1(a). The partial input and recurrent connectivity were randomly generated with probability 30%. Input pool $\hat{2}$ had correlation ($\hat{c}_2 = 0.1$) while pool $\hat{1}$ had none; the firing rates were $\bar{v}_1 = \bar{v}_2 = 30$ Hz. The two plots show the evolution of individual weights (grey bundle); the mean weight K_{av} from the simulation (thin solid line); the analytically-predicted equilibrium value K_{av}^* (thin dotted line); the two simulated mean weights \bar{K}_{11} (thick dashed-dotted line) from the uncorrelated pool $\hat{1}$ and \bar{K}_{12} (thick dashed line) from the correlated pool $\hat{2}$. (a) For an initial distribution corresponding to the means $\bar{K}_{11}(0) = 0.028$ and $\bar{K}_{12}(0) = 0.012$, STDP inverted the weight distribution to potentiate \bar{K}_{12} (thick dashed line) eventually. The homeostatic equilibrium held momentarily and then broke down. (b) When starting with means $\bar{K}_{11}(0) = 0.038$ and $\bar{K}_{12}(0) \simeq 0.002$, the initial weight distribution was not inverted by STDP. The weights corresponding to \bar{K}_{12} (thick dashed line) are barely visible at zero in the plot. The homeostatic equilibrium held satisfactorily throughout the simulation, which determined the equilibrium value of \bar{K}_{11} (thick dashed-dotted line).

as illustrated in Fig. 4.3(b).

The specific case $K(\infty)\hat{\mathbf{h}} = 0$, which occurs for example when the two pools have the same correlation strength $\hat{c}_1 = \hat{c}_2$, will be studied in details in Sec. 4.4. When at least one of the input pools has correlation, the specific case where the matrix A_r in Eq. (4.20) is not invertible almost always leads to an unstable mean over the two input pools of the input weights for each neuron, i.e., $K(\infty)\hat{\mathbf{e}}$ will diverge to the bounds. This case is not interesting for learning and will not be considered further.

4.3.5 Distinct firing rates for the two correlated input pools

For a general choice of learning and input parameters, we have $\gamma' \neq 0$ and $\kappa' \neq 0$, and the signs of the vector elements of the fixed point $K(\infty)\hat{\mathbf{h}}$ in Eq. (4.22) depend on a complex relationship between the two mean input firing rates (\bar{v}_1 and \bar{v}_2) and mean correlation strengths (\hat{c}_1 and \hat{c}_2). Neglecting the inhomogeneities in the network, we approximate

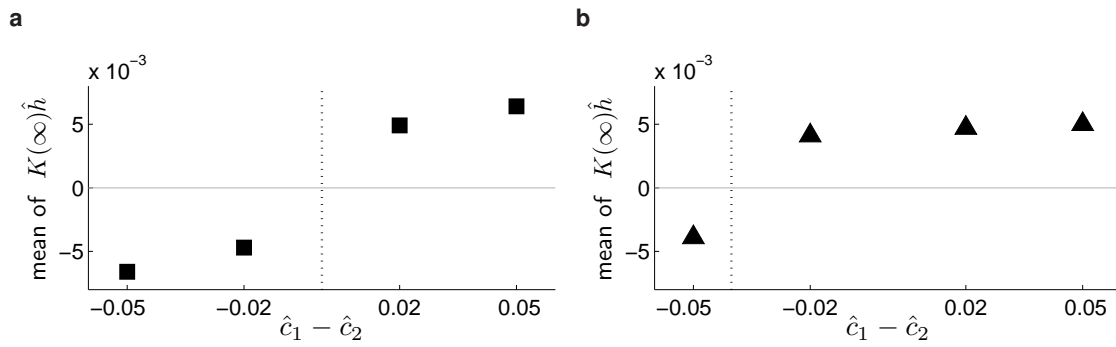


Figure 4.4: Asymptotic weight specialisation dependence upon the difference between the input correlation strengths; influence of the firing rates. The network corresponded to Fig. 4.1(a) with 30%-random partial connectivity for both input and recurrent connections; the weights were initially homogeneous with respective means $K_{av}(0) = 0.02$ and $J_{av} = 0.015$, and $\pm 10\%$ spread; both input and recurrent delays were randomly chosen with mean $7 \text{ ms} \pm 2 \text{ ms}$. In each case, the four simulations corresponded to, respectively, $\hat{c}_1 = 0.05$ and $\hat{c}_2 = 0$; $\hat{c}_1 = 0.02$ and $\hat{c}_2 = 0$; $\hat{c}_1 = 0$ and $\hat{c}_2 = 0.02$; and $\hat{c}_1 = 0$ and $\hat{c}_2 = 0.05$. The plotted points represent the mean of $K(\infty)\hat{h}$ over the neurons at the end of the simulation: it is positive if the network specialised to pool $\hat{1}$ and negative for pool $\hat{2}$. The dotted line indicates the demarcation border between the specialisations to pool $\hat{1}$ and pool $\hat{2}$, estimated using similar calculations to those for Eq. (4.25). (a) For equal input firing rates $\hat{v}_1 = \hat{v}_2 = 30 \text{ Hz}$, the difference $\hat{c}_1 - \hat{c}_2$ determines the specialisation scheme and the squares have the same sign as $\hat{c}_1 - \hat{c}_2$. (b) When the firing rates $\hat{v}_1 = 40 \text{ Hz}$ and $\hat{v}_2 = 30 \text{ Hz}$, the correlation strength \hat{c}_2 required for the network to specialise to pool $\hat{2}$ is higher: weak correlation strength still leads to the selection of pool $\hat{1}$ (triangle on the right of the demarcation dotted line with $\hat{c}_1 = 0$ and $\hat{c}_2 = 0.02$).

$K(\infty)\hat{\mathbf{e}} \simeq n_{\text{av}}^K K_{\text{av}}^* \mathbf{e}$ and $(\mathbb{1}_N - J)\mathbf{e} \simeq (1 - n_{\text{av}}^J J_{\text{av}})\mathbf{e}$ in Eq. (4.22), in order to obtain

$$K(\infty)\hat{\mathbf{h}} \simeq -\frac{n_{\text{av}}^K K_{\text{av}}^* \gamma + (1 - n_{\text{av}}^J J_{\text{av}})\gamma' + \kappa'}{\kappa} \mathbf{e}. \quad (4.25)$$

If the input firing rates are very different, the γ' and κ' terms may dominate the γ term in Eq. (4.25), and hence $K(\infty)\hat{\mathbf{h}}$ depends on the input firing rates and not the input correlation strengths. On the other hand, we can obtain an approximate condition on the mean parameters to ensure that the input correlation strengths determine the sign of the numerator of Eq. (4.25), namely

$$\left| \frac{\hat{c}_1 \bar{v}_1 - \hat{c}_2 \bar{v}_2}{\bar{v}_1 - \bar{v}_2} \right| > h(\hat{v}_{\text{av}}, \hat{C}_{\text{av}}^{W*\epsilon}), \quad (4.26)$$

where the function h is defined by Eq. (B.6) in Appendix B.2. The formula is more interesting qualitatively than quantitatively: this condition is satisfied when the difference between the correlation strengths $\hat{c}_1 - \hat{c}_2$ is sufficiently large for given input firing rates \bar{v}_1 and \bar{v}_2 , which can always be obtained when the difference $\bar{v}_1 - \bar{v}_2$ is not too large. The recurrent connectivity generally affects such a balance between the input firing rates and correlations.

Under the qualitative condition of small discrepancies between the input firing rates, all the vector elements have the same sign given by

$$\text{sgn} [K(\infty)\hat{\mathbf{h}}] = \text{sgn} \left[-\frac{\gamma}{\kappa} K(\infty)\hat{\mathbf{e}} \right] = \text{sgn} [\hat{c}_2 - \hat{c}_1] \hat{\mathbf{e}}. \quad (4.27)$$

In this situation, the input weights from the more correlated input pool will be potentiated, as illustrated in Fig. 4.5 for partial input and recurrent connectivity, $\bar{v}_1 > \bar{v}_2$ and $\hat{c}_1 < \hat{c}_2$. The corresponding simulation with uncorrelated inputs would lead to a potentiation of the input weights from the first input pool since we have used $w^{\text{out}} < 0$ (see Sec. 4.3.3). In that simulation, because of the inhomogeneities in J , the firing rates ended up stable though not clustered near that value, as shown in Fig. 4.5(a); the homeostatic equilibrium for the weights was not strictly satisfied, cf. the discrepancies between the black thin solid line compared to the thin dashed line in Fig. 4.5(b). This discrepancy

is due to the inhomogeneous recurrent connections that affect the vector of equilibrium neuronal firing rates; in other words, Eq. (4.25) is not a good approximation in this case and Eq. (4.22) should be used. Consequently, the number of potentiated weights is different for the first half and the second half of the simulated neurons in Fig. 4.5(c).

Figure 4.4(b) illustrates the dependence of the asymptotic weight specialisation upon the input firing rates and correlation strengths with the same learning parameters: stronger input correlation is necessary to select a correlated input pool when its firing rate is smaller than that of the uncorrelated pool. Compared to the case where the two input pools have the same firing rate in Fig. 4.4(a), the border (dotted line) between the potentiation of pool $\hat{1}$ and pool $\hat{2}$ is shifted to the left.

4.3.6 Extension to several homogeneous input pools

The analysis in Sec. 4.3 can be generalised to the case of an arbitrary number m of homogeneous input pools. It is possible to construct $m - 1$ vectors, $\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_{m-1}$, in a similar way to $\hat{\mathbf{h}}$ above, in order to form an orthogonal basis together with $\hat{\mathbf{e}}$ in which A and B can be expressed in a form analogous to Eq. (4.17). For m pools of the same size, the vectors $\hat{\mathbf{h}}_1, \dots, \hat{\mathbf{h}}_{m-1}$ can be constructed using the m^{th} root of unity ($\hat{\mathbf{e}}$ corresponding to one), and the decomposition of K using this basis can be obtained using the discrete Fourier transform.

4.4 Symmetry breaking of the distribution of input weights with fixed recurrent weights

4.4.1 Previous results concerning the weight drift

We now focus on the special case where we have two input pools of the same size that have the same firing rate $\hat{\nu}_0$ and same correlation strength \hat{c}_0 , as defined in Eq. (3.28). The evolution of the matrix K of the input weights can be described through two vectors $K\hat{\mathbf{e}}$ and $K\hat{\mathbf{h}}$, where $\hat{\mathbf{e}}$ is a column vector with all M elements equal to one Eq. (4.14), and $\hat{\mathbf{h}}$ is the column vector defined in Eq. (4.15) that has the first $M/2$ elements are 1 and the next

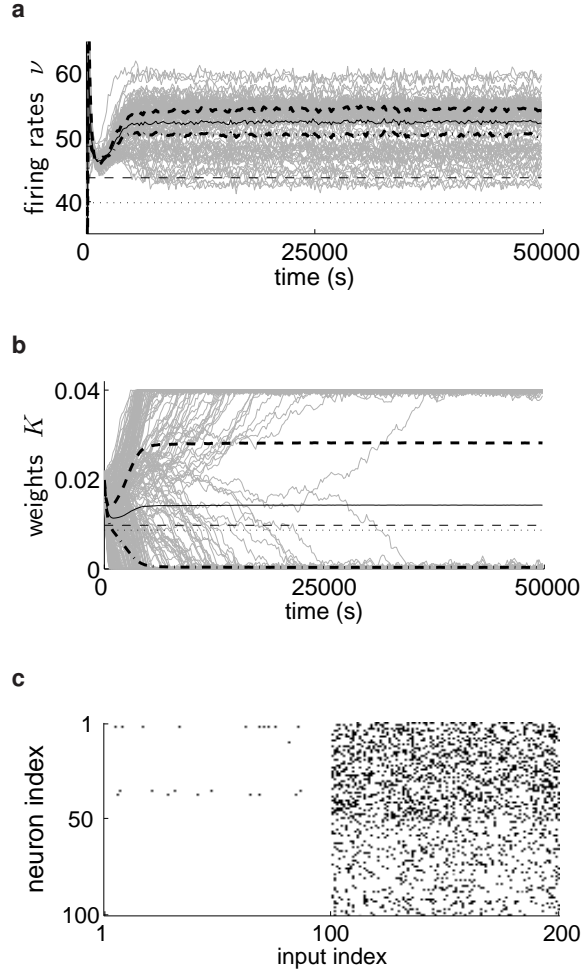


Figure 4.5: Weight evolution for unbalanced correlations. The network of $N = 100$ neurons was stimulated by two pools of $M/2 = 100$ inputs each, with partial input and recurrent connectivity (30%). The input weights were initially homogeneous around the mean value $0.02 (\pm 10\%)$ while the recurrent weights were inhomogeneous with lumped feedback $\bar{J}_{11} = 0.45$, $\bar{J}_{12} = 0$, $\bar{J}_{21} = 0.9$ and $\bar{J}_{22} = 0.45$; see Fig. 4.2. The input firing rates and correlations were set to $\bar{\nu}_1 = 35$ Hz, $\bar{\nu}_2 = 30$ Hz, $\hat{c}_1 = 0.05$ and $\hat{c}_2 = 0.1$ respectively. (a) The firing rates ν (grey bundle; mean ν_{av} in black thin solid line) eventually stabilised not far from the predicted value for the homeostatic equilibrium in Eq. (4.8) (thin dashed line; the dotted line shows the corresponding value for uncorrelated inputs); the means for each half of the network are plotted ($\bar{\nu}_1$ in thick dashed-dotted line and $\bar{\nu}_2$ in thick dashed line). (b) The input weights K individually diverged (grey bundle) while the mean weight K_{av} (thin solid line) first converged towards and then stabilised close to the homeostatic equilibrium value (thin dashed line; the thin dotted line stands for uncorrelated inputs); see Eq. (4.6). The weights from the more correlated pool $\hat{2}$ (\bar{K}_{12} in thick dashed line) became potentiated compared to those from pool $\hat{1}$ (\bar{K}_{11} in thick dashed-dotted line). (c) Emerged structure in the weight matrix K . Darker pixels represent potentiated weights. Generally only weights coming from the more correlated input pool became potentiated (input indices #101 to #200, right side).

$M/2$ elements are -1 . The elements for index i of the column vectors $K\hat{\mathbf{e}}$ and $K\hat{\mathbf{h}}$ thus are, for each neuron, the lumped sum of all input weights and the difference between the weight sums coming from each of the two pools, respectively (cf. Sec. 4.3).

We assume weak correlation and that the condition in (4.5) is satisfied, ensuring homeostatic equilibrium, that is, the stability of the mean input weights for all neurons. When the inhomogeneities of J are neglected, we can approximate $K\hat{\mathbf{e}} \simeq n_{\text{av}}^K K_{\text{av}}^* \mathbf{e}$, where n_{av}^K is the mean number of input weights per neuron and K_{av}^* is the equilibrium value of the mean input weight; the N -column vector \mathbf{e} is defined similarly to $\hat{\mathbf{e}}$ in Eq. (4.14). The mean neuron firing rate ν_{av} is then also stable and its equilibrium value can be approximated by the expression in (4.9). Details are provided in Sec. 4.3.

When the network is in homeostatic equilibrium, $K\hat{\mathbf{h}}$ describes the specialisation of each neuron to one of the two input pools: if the i^{th} vector element grows positively (negatively), then neuron i becomes sensitive to input pool $\hat{1}$ only (resp. $\hat{2}$). Since the input pools have identical firing rates and correlation strengths, the evolution of $K\hat{\mathbf{h}}$ is given by

$$\dot{K}\hat{\mathbf{h}} = F^W \hat{\mathbf{h}} = M\kappa (\mathbf{1}_N - J)^{-1} K\hat{\mathbf{h}}. \quad (4.28)$$

The present case corresponds to the analysis in Sec. 4.3 with $\gamma = \gamma' = \kappa' = 0$ in Eq. (4.19b); F^W , as defined in Eq. (3.3) with $\Psi = W$, is expressed in (4.1b). The constant κ is given by (4.18)

$$\kappa = \frac{1}{2} \hat{C}_{kl}^{W*\epsilon}(0) = [W * \epsilon](0) \frac{\hat{c}_0 \hat{\nu}_0}{2}, \quad (4.29)$$

where we have used Eqs. (3.5) and (3.28) with k and l in the same input pool, $\Psi = W * \epsilon$ and $\hat{c} = \hat{c}_0$.

The fixed point $K(\infty)\hat{\mathbf{h}} = 0$ is unstable since $\kappa > 0$. This holds because $[W * \epsilon](0) > 0$ for any ‘‘Hebbian’’ choice of STDP window function W such that $W(u) > 0$ for $u < 0$, cf. Fig. 2.2, which we assume throughout what follows. For initial conditions in which each neuron is already specialised to a given input, STDP should reinforce the initial specialisation. However, for homogeneous initial input weights (in other words the network is unorganised), $K(0)\hat{\mathbf{h}} \simeq 0$ and the state of the dynamical system lies at the unstable fixed

point so that $K\hat{\mathbf{h}}$ will grow either positively or negatively until most input weights become either saturated or quiescent. In this case, the drift for $K\hat{\mathbf{h}}$ is initially zero according to Eq. (4.28) and it is not modified by the convergence towards the homeostatic equilibrium; higher-order terms may then come into play and influence the symmetry breaking. If the neurons are not recurrently connected (in a purely feed-forward network), half of them should specialise to one input pool and the other half to the other pool (Gütig et al. 2003). In the remainder of this section, we examine the dynamics of such symmetry breaking, focusing on the impact of the recurrent connections on the specialisation pattern.

4.4.2 Impact of fixed recurrent connections

To evaluate the second moment of the stochastic evolution of the weights K , we proceed in a similar manner to Kempter et al. (1999) and Burkitt et al. (2007) for the analysis of the weight variance by evaluating $Y_{i,k,j,k}(t, t')$ as defined in Eq. (3.25) for indices i, j and $k = l$. In the remainder of this section, we assume that all the input delays are identically equal to \hat{d} , and likewise all the recurrent delays are equal to d . For each pair of input weights K_{jk} and K_{ik} from the same external input k to two neurons i and j , a connection from neuron j to neuron i induces an extra contribution to the expectation value $Y_{i,k,j,k}(t, t')$. This contribution relates to the spike-triggering effect for each pair of spikes fired by input k and neuron j , as shown in Appendix B.3, which results in a *positively* correlated evolution of the weights K_{jk} and K_{ik} for positive recurrent weights. In other words, these two weights tend to vary in the same way, either potentiated or depressed. It follows that the i^{th} and j^{th} elements of $K\hat{\mathbf{h}}$ tend to behave in the same way at the beginning of learning, when the input weights split between the two pools. This means that neurons i and j should have similar input specialisation patterns. A connection back from neuron i to neuron j reinforces this phenomenon for each pair of spikes at input k and at neuron i . The contribution is stronger when w^{in} is large and when w^{out} and \tilde{W} have the same sign; see Eq. (B.16) in Appendix B.3.

For a randomly connected network with roughly $n_{\text{av}}^K = n^K/N$ and $n_{\text{av}}^I = n^I/N$ pre-synaptic input and recurrent connections per neuron, respectively, the sum over the

whole network of the terms related to the spike-triggering effect that impact upon all the weight changes is

$$\frac{n^K n_{av}^K n_{av}^J}{MN} J_{av} v_{av} \left(w^{\text{out}} + \tilde{W} \hat{v}_{av} \right)^2 . \quad (4.30)$$

Note that this expression would be multiplied by η^2 if time had not been rescaled. The positive correlation of the evolution of all the input weights K_{ik} coming from all the inputs k in a homogeneous pool is stronger for denser input and recurrent connectivity, i.e., larger values of n_{av}^K and n_{av}^J . The expression in Eq. (4.30) is to be compared with the increase of the variance (Kempster et al. 1999, Eqs. (30) and (31)) lumped for all the n^K input weights

$$n^K \left\{ (w^{\text{in}})^2 \hat{v}_{av} + (w^{\text{out}})^2 v_{av} + \tilde{W}^2 \hat{v}_{av} v_{av} + 2\tilde{W} \hat{v}_{av} v_{av} \left[w^{\text{in}} + w^{\text{out}} + \tilde{W} (\hat{v}_{av} + v_{av}) \right] \right\} , \quad (4.31)$$

where $\tilde{W}^2 = \int [W(u)]^2 du$. The difference in order of magnitude related to the connectivity between the expressions in Eqs. (4.30) and (4.31) is given by $n_{av}^K n_{av}^J / MN$, a fraction that is almost equal to one for full input and recurrent connectivity. This means that spike-triggering effects may impact the weight dynamics in sufficiently dense networks.

As a result, neurons from a recurrently connected group tend to specialise together to the same input pool, with probability 50% for each pool, as illustrated in Fig. 4.6(a-b). This is in contrast to a network with no recurrent connections, in which each neuron specialises individually and independently to one of the two pools, as illustrated in Fig. 4.6(c-d).

A sufficiently large correlation strength \hat{c}_0 is required to ensure that the diverging behaviour of the input weights corresponds to a splitting between the two pools and therefore input selectivity (Gütig et al. 2003). This finding can be qualitatively reproduced in a calculation similar to Appendix B.3 involving K_{ik} and K_{il} for one neuron i and two inputs k and l , see Appendix B.4.

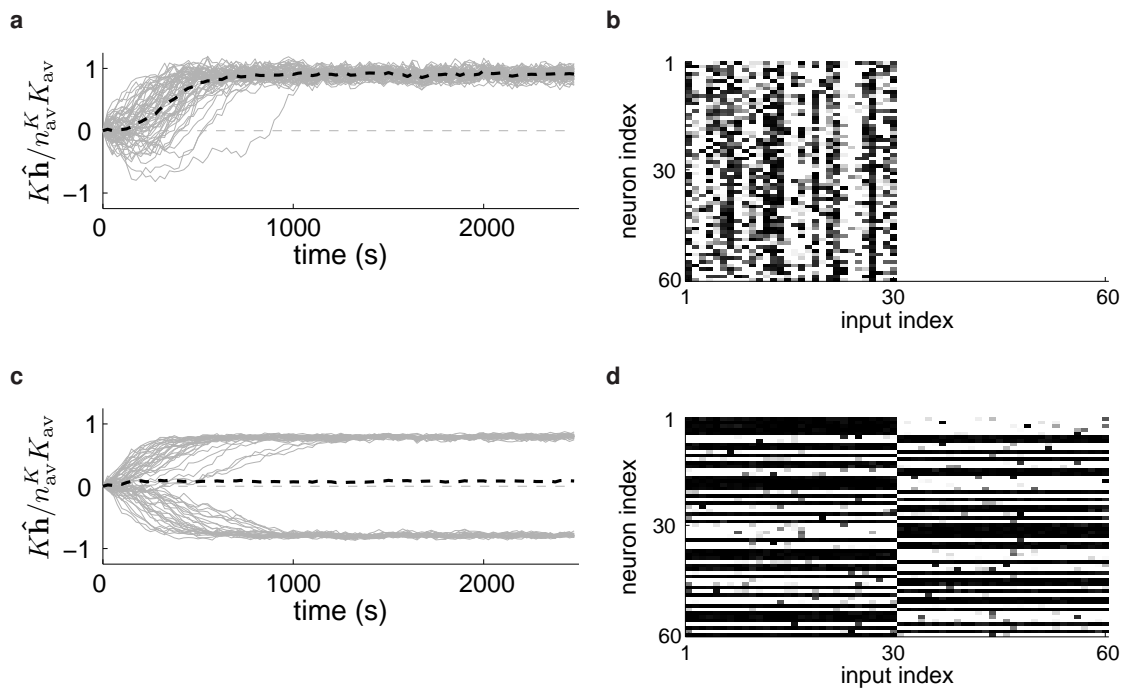


Figure 4.6: Symmetry breaking of the input weights for a group of $N = 60$ neurons and two pools of $M/2 = 30$ inputs each. (a & b) Full recurrent connectivity versus (c & d) no recurrent connections. The plots (a & c) show the traces of the elements of $K\hat{\mathbf{h}}$ for each neuron (grey thin solid lines) and the mean over all the neurons of $K\hat{\mathbf{h}}$ (black thick dashed line). The grey dashed line at zero corresponds to no specialisation. The matrix graphs (b & d) show the matrix K (neuron indexed vertically; input horizontally with first pool on the left and second pool on the right) at the end of learning; darker pixels stand for potentiated weights. For full recurrent connectivity, almost all neurons (b: 60 vs. 0) specialised to the first input pool and the mean of the $K\hat{\mathbf{h}}$ is clearly positive (a). In contrast, for no recurrent connections, the neurons specialised almost evenly between the two input pools (d: 33 vs. 27) and the mean of the $K\hat{\mathbf{h}}$ is almost zero (c).

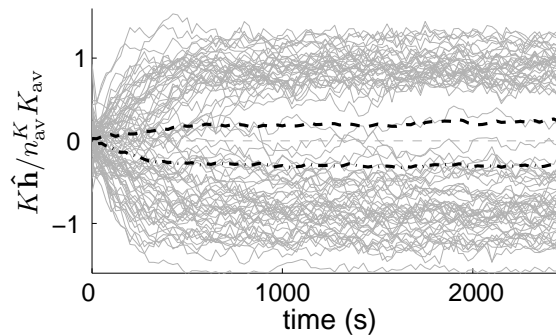


Figure 4.7: Symmetry breaking of the input weights for two pools of $M/2 = 100$ correlated inputs each, and a network made of two groups of $N = 200$ neurons each. Input weights are initially random ($\pm 10\%$ of the mean value 0.03), with partial connectivity (30%). The two groups of neurons have stronger connectivity within each (30% with mean $0.015 \pm 10\%$) than between them (10% with mean $0.008 \pm 10\%$). The plot line coding is similar to Fig. 4.6(a,c). The first group of neurons (#1-100, weight mean in thick dashed line with only a representative portion plotted) specialised to the first input pool while the second group (#101-200, mean in thick dashed-dotted line) to the second input pool.

4.4.3 Non-homogeneous fixed recurrent connections

Partial connectivity with low density and/or small recurrent weights weakens the group symmetry-breaking effect. This is illustrated in Fig. 4.7 for a network made of two groups of neurons (cf. Fig. 4.2) with partial connectivity both for the plastic input (30%) and the fixed recurrent connections (30% within-group and 10% between-group). The specialisation is weaker than that for full connectivity, cf. Fig. 4.6(a). In addition, neuron group 1 (weight mean represented by the thick dashed trace) weakly specialised to input pool $\hat{1}$, while group 2 (thick dashed-dotted trace) specialised to pool $\hat{2}$. This relates to the fact that the neuron groups have stronger feedback within themselves than between each other and thus may evolve in an independent way. The behaviour illustrated in Fig. 4.7 is interesting in that it contrasts with the expected specialisation of the network, where the two neuron groups select the same input pool because of positive coupling (recurrent weights) between them; in general, specialisation to the same input pool is more likely to occur than to different input pools.

For stronger fixed recurrent weights, the two neuron groups tend to specialise to the same input pool, whereas for smaller recurrent weights they may specialise to different input pools, as illustrated in Fig. 4.8. The y-axis indicates the degree of specialisation

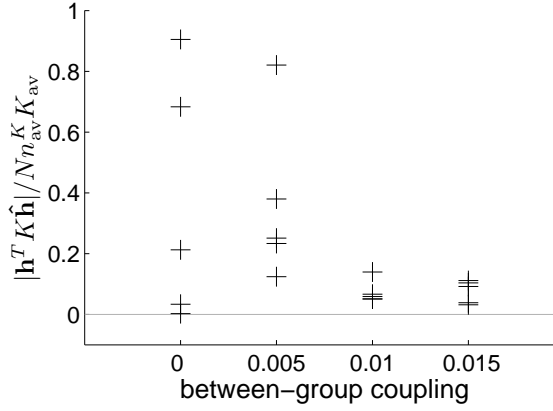


Figure 4.8: Illustration of the specialisation of two neuron groups as a function of the coupling between them. Each plotted point represents the outcome of a simulation for one network configuration. The simulated network was similar to that of Fig. 4.7, except for the strength of coupling between the two groups, that is, the recurrent weights between them that have a fixed mean weight (x-axis in the plot) with partial connectivity (15%). The y-axis is a measure of the difference in the specialisation between the two neuron groups: high values indicate specialisation to different input pools (the vector \mathbf{h} is defined in the text).

to different input pools measured by the scalar value $|\mathbf{h}^T K \hat{\mathbf{h}}| / N n_{\text{av}}^K K_{\text{av}}$, where \mathbf{h} is the N -column vector defined similar to $\hat{\mathbf{h}}$ with $N/2$ elements equal to 1 and $N/2$ equal to -1 . The probability of selecting different input pools decreases when the between-group coupling increases. The recurrent connections only have a higher-order impact on the symmetry breaking of the input weights, which induces a (probabilistic) trend to jointly specialise. For more complex network architectures, the coupling related to the recurrent connections may lead to non-trivial competition between network areas.

4.4.4 Dependence upon neuron model, initial conditions and learning parameters

The results shown here correspond to $\tilde{W} < 0$ and short recurrent delays (cf. Appendix D); similar results were obtained with $\tilde{W} > 0$ and/or larger recurrent delays (e.g. 10 ms). Note that the presence of recurrent connections changes the equilibrium value of the mean input weight K_{av}^* , which has an impact on the weight saturation through the number of potentiated vs. quiescent weights.

From a population-statistics point of view, the input firing rates, spike-time correlations and initial weights need not be exactly fine-tuned to ensure that symmetry breaking

in one way or the other is equally probable over all neurons. In other words, when $K(\infty)\hat{\mathbf{h}}$ and $K(0)\hat{\mathbf{h}}$ are not strictly zero, the uncertainty due to the initial distribution of the input and recurrent weights (when homogeneous) still leads to an equiprobable specialisation to one of the two input pools. This situation occurs in particular for partial input connectivity, where individual neurons may receive more connections from one input pool than the other. We observed that spike-triggering effects were sufficiently strong to play a role even for 30% random input connectivity and 10% spread of the initial input weights around their mean (Fig. 4.7).

In addition to the input connectivity and the initial distribution of the input weights, the stochastic nature of the Poisson neurons has an influence upon the weight dynamics: the intrinsic randomness of the output favours equiprobability of specialisation to each of the two input pools for any input spiking history. Similar results were obtained using a deterministic version of the integrate-and-fire neuron model, which required the addition of an external source of background activity in place of the spontaneous rate ν_0 of the Poisson neuron model. For this purpose, we have used an extra input pool of uncorrelated Poisson spike trains with random and fixed input connectivity.

4.5 Partial conclusion on plastic input connections

In the case of learning input connections while the recurrent weights are kept fixed, homeostatic stability of both the firing rates and weights can be obtained for a wide range of learning parameters. The stability condition Eq. (4.5) was derived for weak input correlations, but it proved to be sufficient beyond the limitation of weak correlation strengths. Stability for additive STDP requires that the effect of a single pre-synaptic spike increases the weight. In numerical simulations, we chose $w^{\text{in}} > 0$, $w^{\text{out}} < 0$ and $\tilde{W} < 0$, which leads to homeostatic stability whatever the input firing rate. This is in agreement with earlier numerical studies of integrate-and-fire neurons in feed-forward networks (Song et al. 2000) and recurrent networks (Song and Abbott 2001).

While homeostatic equilibrium is satisfied, the individual weights exhibit a diverging

behaviour, indicating strong competition between them. For two correlated input pools with firing rates in the same range (but not necessarily identical), this generally results in selecting the input connections coming from the more correlated pool (Sec. 4.3), as illustrated in Fig. 4.9(a \Rightarrow b) for the case where only one input pool has correlations. Both the convergence of the mean input weight and the specialisation (asymptotic bimodal weight distribution) are exponentially fast, the latter occurring on a slower time scale for weak input correlations. The value of the learning rate η was chosen to obtain a convergence towards the homeostatic equilibrium in hundreds of seconds, similar to Burkitt et al. (2007), and a development of a weight structure in tens of thousands of seconds (i.e., hours). Similar results were obtained with faster learning rates ($\eta = 10^{-5}$). Our results show that, even for small learning rates, the combination of equilibrium and diverging behaviour leads to the emergence of a weight structure.

When starting with an initial homogeneous distribution of input weights, the presence of the fixed recurrent connections does not qualitatively change this robust specialisation of the weights compared to a purely feed-forward architecture (Kempster et al. 1999). This behaviour is obtained whatever the shapes of the PSP kernel ϵ and the STDP window function W provided STDP is “Hebbian” (cf. Sec. 2.3). Short durations of both ϵ and the recurrent delays were required for the analysis but similar conclusions at the mesoscopic network scale held when varying these parameters (Fig. 4.4). An exception to this expected behaviour occurs for initial conditions in which the weights are already dramatically specialised in the “wrong” way or large difference between input firing rates (under certain conditions on the learning parameters); then STDP does not always select the more correlated input pathway, as shown in Fig. 4.3(b) and Fig. 4.4(b).

In the particular case of input pools with balanced firing rates and within-pool correlations, the competition induced by STDP results in symmetry breaking for a homogeneous initial distribution of the input weights (Gütig et al. 2003). For two input pools with balanced within-pool correlation but no between-pool correlation, sufficiently correlated inputs (but in the range of small correlations) are necessary in order to obtain an asymptotic bimodal distribution of the input weights that corresponds to a splitting be-

tween the two pools. This holds for a broad range of STDP parameters, provided they correspond to a stabilisation of the firing rates. The influence of non-identical but similar input firing rates and spike-time correlations has yet to be studied in more depth. This robust specialisation occurs whatever the detail of the shape of the STDP learning window function W (provided it is “Hebbian”, cf. Sec. 2.3), PSP kernel ϵ and homogeneous input delays.

During symmetry breaking, the non-learning connections can play a determining role, for example, in causing neurons with fixed excitatory recurrent connections to specialise in the same way, as illustrated in Fig. 4.9(c \Rightarrow d). This group effect takes place at the beginning of learning; when the neurons become sufficiently specialised, the drift takes over and reinforces the initial symmetry breaking because of the instability of the fixed point related to the differential equation Eq. (4.28). Inhomogeneous fixed recurrent connectivity can cause neuron groups in the network to specialise to different input pathways, as shown in Fig. 4.9(e \Rightarrow f). STDP thus provides a framework for cortical self-organisation in the textbook case where learning takes place on the excitatory connections from some external inputs whereas the remaining connections are considered fixed. Our results can be linked, for example, to the emergence of ocular-dominance areas in the primary visual cortex, when specialising to one ocular pathway (left or right eye) in the first weeks of life of new-born mammals. The two assumptions of stronger local excitatory connections than those at a longer range and of more correlation for spike trains within each ocular pathway than between the two pathways, are sufficient to qualitatively obtain the emergence of specialised recurrently connected areas sensitive to the inputs from only one eye. Other versions of STDP are expected to generate similar group specialisation so long as they generate both a homeostatic equilibrium and a splitting of the weight distribution depending upon the input correlations. Higher-order effects due to the recurrent connections may combine with non-linearities in other STDP models (Gütig et al. 2003, Burkitt et al. 2004, Appleby and Elliott 2006) or specific input structures (e.g., Leibold et al 2002) to introduce further complexity in the weight dynamics.

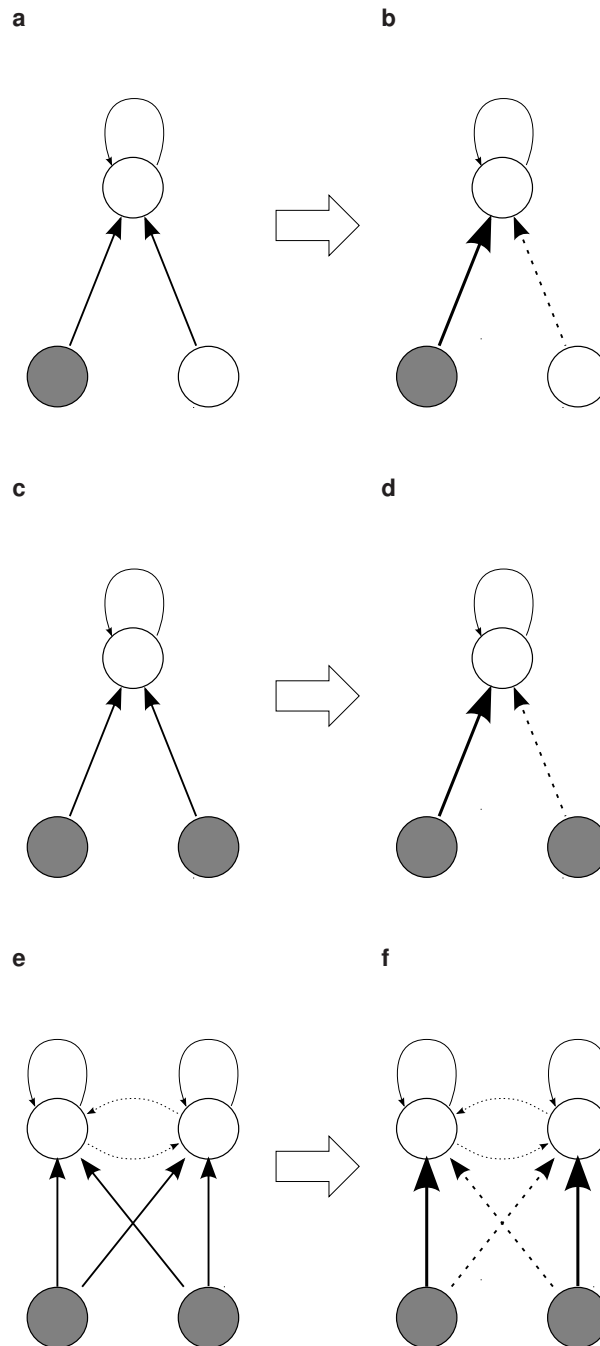


Figure 4.9: Schematic representation of the input weight distribution (a, c & e) before and (b, d & f) after learning. The neuron groups (top circles) in the network become sensitive to only one of the two correlated input pools (bottom circles) through the potentiation of some input weights (very thick arrow) at the expense of the other input weights that are depressed (dashed thick arrow).

Chapter 5

Plastic recurrent connections

This chapter investigates the weight dynamics in a network with plastic recurrent connections stimulated by fixed external inputs (in particular, fixed input weights).

5.1 Introduction

IN THIS chapter, a network where additive STDP only affects the recurrent connections with fixed input weights is considered, which is the converse situation of that studied in the previous chapter. The focus is on several specific cases illustrated in Fig. 5.1: (a) no external inputs, (b) unbalanced input pools (one has correlations while the other does not) and (c) two correlated input pools.

Previous theoretical studies have primarily investigated the weight dynamics induced by STDP for single neurons and the implications for feed-forward networks (Gerstner et al. 1996, Kempter et al. 1999, Gütig et al. 2003, Burkitt et al. 2004, Meffin et al. 2006). The cortex, however, is dominated by recurrent connections and the effect of STDP has only begun to be addressed in such recurrent networks, mainly using numerical simulation (Song and Abbott 2001, Morrison et al. 2007, Câteau et al. 2008, Lubenov and Siapas 2008). In contrast to a feed-forward architecture, the weight dynamics induced by spike-timing-dependent plasticity (STDP) in a recurrent neuronal network is not yet well understood. We investigate how the weight structure develops within the network when stimulated by input pools with homogeneous firing rates and within-pool (but no between-pool) spike-time correlations, an idea inspired by Kempter et al. (1999).

In Sec. 5.5, the study is constrained to the case of two input pools when the input weights are kept fixed (cf. Fig. 5.1). We formulate stability conditions for the mean fir-

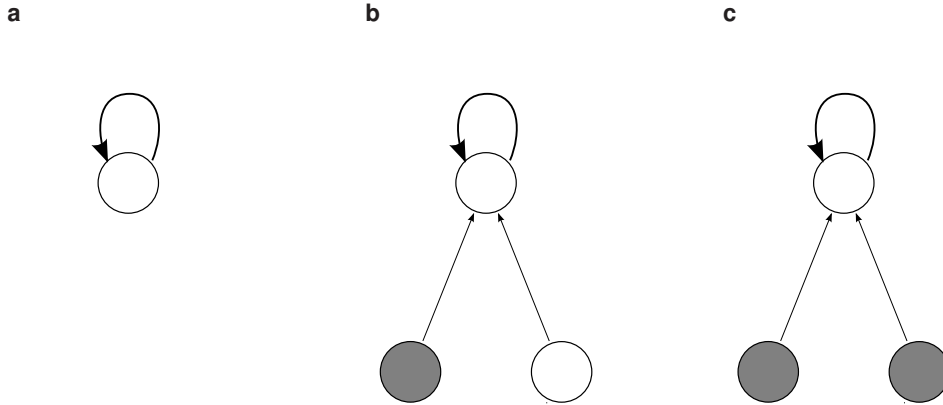


Figure 5.1: Network configurations studied in chapter 5. Top circles represent the neuronal network and bottom circles the two input pools, for which filled circles indicate non-zero within-pool correlation. Thick (resp. thin) arrows indicate plastic (fixed) weights.

ing rate and weight (homeostatic equilibrium defined in Sec. 3.6), and study the asymptotic weight distribution through a fixed-point analysis. The evolution of the recurrent weights is decomposed in order to understand how stability and specialisation can occur together.

5.2 Equilibrium in a partially connected recurrent network with no external inputs

In the case of no external inputs, the dynamical system (3.22a-3.22e) reduces to

$$\boldsymbol{\nu} = (\mathbb{1}_N - J)^{-1} \nu_0 \mathbf{e}, \quad (5.1a)$$

$$J = \Phi_J(w^{\text{in}} \mathbf{e} \boldsymbol{\nu}^T + w^{\text{out}} \boldsymbol{\nu} \mathbf{e}^T + \tilde{W} \boldsymbol{\nu} \boldsymbol{\nu}^T). \quad (5.1b)$$

Time has been rescaled to remove η . The study of the weight dynamics induced by STDP for a network with no external inputs was demonstrated for the case of full recurrent connectivity (Burkitt et al. 2007), which described the fixed points of the firing rates and of the recurrent weights, their stability, and the evolution of the weight variance. In particular, Burkitt et al. (2007) derived the stability conditions for the homeostatic equilibrium, namely the situation where the mean firing rates and mean weights stabilize although individual firing rates and weights may continue to change, and the same analysis can

be applied to the present case of partial connectivity. The stability for the mean firing rate and the mean weight over all neurons is ensured provided

$$w^{\text{in}} + w^{\text{out}} > 0 \text{ and } \tilde{W} < 0. \quad (5.2)$$

Here we study the equilibria of the dynamical system Eqs. (5.1a-5.1b) in terms of individual firing rates and mean weights (taking into account the network topology) and the corresponding stability conditions. The present analysis of the weight drift is similar to that of Burkitt et al. (2007), but here the projector Φ_J in Eq. (5.1b) is non-trivial and nullifies not only the diagonal elements, but also other elements according to the partial connectivity of the network. The matrix J belongs to the vector subspace of $\mathbb{R}^{N \times N}$ defined by $\mathbb{M}_J := \{X \in \mathbb{R}^{N \times N}, \Phi_J(X) = X\}$, whose dimension is the number of existing connections n^J .

5.2.1 Fixed point of the firing rates

We first find the equilibrium states of the network dynamics in terms of the firing rates. Setting $\dot{J} = 0$ in Eq. (5.1b) leads to the following condition on the firing rates for all existing connections $j \rightarrow i$:

$$w^{\text{in}}v_j + w^{\text{out}}v_i + \tilde{W}v_jv_i = 0, \quad (5.3)$$

that is,

$$v_i = q(v_j) := -\frac{w^{\text{in}}v_j}{w^{\text{out}} + \tilde{W}v_j}. \quad (5.4)$$

For any loop of synaptic connections from a given neuron i_0 back to itself through neurons i_1, i_2, \dots, i_{n-1} , we have

$$v_{i_{m+1}} = q(v_{i_m}) \quad \text{for } m = 0, \dots, n-1, \quad (5.5)$$

and thus $v_{i_0} = q^{\{n\}}(v_{i_0})$ where $q^{\{n\}}$ is the self-composition of q defined in Eq. (5.4) iterated n times. In other words, v_{i_0} is a fixed point of $q^{\{n\}}$ whenever there exists a loop of length n .

n in which neuron i_0 takes part. Similarly for all the v_{i_m} of the loop. Due to the special form of q , the function $x \mapsto q^{\{n\}}(x)$ has only two fixed points, namely those of q (see Appendix C.2.1). This means that at the equilibrium of the learning weight dynamics, any neuron i with non-zero firing rate v_i within a loop of arbitrary length must satisfy

$$v_i = \mu := -\frac{w^{\text{in}} + w^{\text{out}}}{\tilde{W}}. \quad (5.6)$$

We discard silent neurons at the equilibrium since the consistency equation for the firing rates Eq. (5.1a) would then imply infinitely large weights.

Consequently, the firing rates at the steady state are homogeneous provided the network connectivity is such that each neuron is part of a loop. This assumption holds when the network has sufficiently many connections (this can be related to the existence of Hamiltonian cycles in the corresponding graph). The existence of loops for every neuron is a reasonable assumption for recurrently connected networks that will be made throughout this section.

5.2.2 Fixed points of the weights

A fixed point of the network dynamics denoted by $(\boldsymbol{v}^* = \mu \mathbf{e}, J^*)$ must satisfy the following condition on the weight matrix J^* according to Eq. (5.1a)

$$J^* \mathbf{e} = \frac{\mu - v_0}{\mu} \mathbf{e}. \quad (5.7)$$

Similar to the case of full connectivity (Burkitt et al. 2007), this equation characterizes an affine space of dimension $n^J - N$ (recall that n^J is the number of connections). For example, a fully connected network without self-connections corresponds to $n^J = N(N - 1)$. A redistribution of the strengths of the incoming weights for each neuron, while keeping the sum of these weights constant at $(\mu - v_0)/\mu$, gives all the solutions J^* . In other words, the sum of the elements for each row of any J^* is equal to that constant.

In general, the dimension of the affine hyperplane of the J^* is non-zero and there is a continuum of fixed points, except for the case where a single loop links all the neurons

together (or several disjoint loops). In a single loop there is only one incoming weight per neuron, which must be equal to $(\mu - \nu_0)/\mu$ at the equilibrium; in other words, no redistribution is possible.

Recall that the matrix $\mathbb{1}_N - J^*$ must be invertible (cf. Appendix C.1); this condition can be enforced by placing bounds on the weights, which is the case in the numerical simulations. The weight matrix can then move on a manifold of fixed points denoted by \mathcal{M}^* , where the drift of the weights arising from the learning equation is zero and the weight evolution is only due to higher orders of the stochastic process (cf. Sec. 3.6), similar to the case of full connectivity (Burkitt et al. 2007).

5.2.3 Stability analysis

We derive from the learning equation Eq. (5.1b) the following linear operator that describes the evolution at the first order of the variation of the weight matrix $\Delta J := J - J^*$ around a given fixed point J^* (Burkitt et al. 2007, Sec. 5)

$$\begin{aligned} \dot{\Delta J} &\simeq \mathcal{L}(\Delta J) \\ &:= -\mu \Phi_J \left[w^{\text{in}} (\mathbb{1}_N - J^*)^{-1} \Delta J \mathbf{e} \mathbf{e}^T + w^{\text{out}} \mathbf{e} \mathbf{e}^T \Delta J^T (\mathbb{1}_N - J^*)^{-1T} \right]. \end{aligned} \quad (5.8)$$

The matrices X such that $X\mathbf{e} = 0$ form a linear subspace of \mathbb{M}_J of dimension $n^J - N$; for ΔJ in this subspace, Eq. (5.8) clearly gives $\dot{\Delta J} = 0$. This corresponds to a displacement along the fixed-point manifold \mathcal{M}^* where the learning equation does not provide any constraint to the leading order, i.e., the drift term of the stochastic weight evolution is zero (Burkitt et al. 2007).

The eigenmatrices related to the linear operator defined in the rhs of Eq. (5.8) depend on the detail of the connectivity topology in a non-trivial way, as described in Appendix C.2.2. In addition to $n^J - N$ eigenvalues equal to zero for the matrices $X \in \mathbb{M}_J$ such that $X\mathbf{e} = 0$, the spectrum of the linear operator \mathcal{L} also contains all the eigenvalues

of the following matrix L_r , as discussed in Appendix C.2.3,

$$\begin{aligned} L_r &= w^{\text{in}} L_{\text{in}} + w^{\text{out}} L_{\text{out}}, \\ L_{\text{in}} &:= -\mu R (\mathbf{1}_N - J^*)^{-1}, \\ L_{\text{out}} &:= -\mu \Phi_J [\mathbf{e} \mathbf{e}^T] (\mathbf{1}_N - J^*)^{-1}, \end{aligned} \quad (5.9)$$

where R is the diagonal matrix whose i^{th} element is the number of incoming connections for neuron i , namely

$$R = \text{diag}[\Phi_J(\mathbf{e} \mathbf{e}^T) \mathbf{e}]. \quad (5.10)$$

In the case of homogeneous recurrent connectivity, the spectrum of L_r for $|w^{\text{in}}| \gg |w^{\text{out}}|$ is almost the same as that of L_{in} , which lies in the left half-plane as illustrated in Fig. 5.2(a). It follows that the condition $w^{\text{in}} \gg |w^{\text{out}}|$ ensures eigenvalues with large negative real parts for \mathcal{L} , as illustrated in Fig. 5.2(c) and Fig. 5.2(e). On the other hand, if $|w^{\text{out}}| \gg |w^{\text{in}}|$, the eigenvalues of \mathcal{L} are almost those of L_{out} . Because of the large number of almost-zero eigenvalues in the spectrum of L_{out} as shown in Fig. 5.2(b), the eigenvalues of \mathcal{L} are then more clustered around zero, and the sign of their real parts may vary according to w^{in} . For $w^{\text{out}} \gg w^{\text{in}} > 0$, the spectrum of \mathcal{L} has eigenvalues with negative real parts (Fig. 5.2(d)). However, for $w^{\text{out}} \gg -w^{\text{in}} > 0$, the spectrum of \mathcal{L} has eigenvalues with positive real parts (Fig. 5.2(f)). These conclusions on stability are independent of the fixed point J^* used to define \mathcal{L} in Eq. (5.8), which means that the fixed-point manifold \mathcal{M}^* is actually either attractive or repulsive as a whole, i.e., either all fixed points are attractive or none of them is attractive.

It follows that in the case of random connectivity, where each neuron has the same number of incoming connections, the condition

$$w^{\text{in}} \gg |w^{\text{out}}| \quad (5.11)$$

is sufficient to ensure stability. On the other hand, the condition $w^{\text{out}} \gg |w^{\text{in}}|$ leads to weaker stability or even instability when $w^{\text{in}} < 0$. For inhomogeneous connectivity

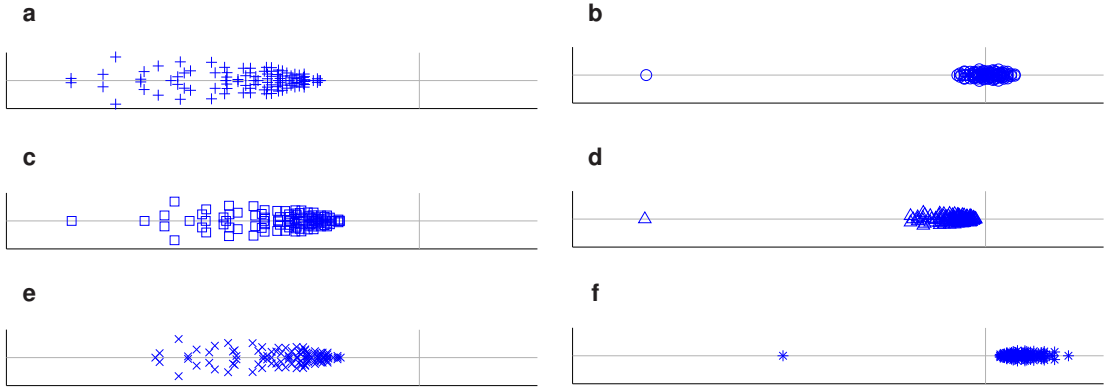


Figure 5.2: Illustration of spectra for the linear operator \mathcal{L} . The scale is the same for the five plots and the axes are the grey solid lines. (a & b) Spectrum of the matrices L_{in} (each eigenvalue is a plus) and L_{out} (circle) defined in Eq. (5.9), for a network of $N = 100$ neurons with homogeneous partial connectivity (30%) and a randomly generated fixed point J^* . These two matrices were rescaled by $w^{\text{in}} + w^{\text{out}}$. Spectrum of \mathcal{L} for: (c) $w^{\text{in}} = 4$ and $w^{\text{out}} = 1$ with squares; (d) $w^{\text{in}} = 1$ and $w^{\text{out}} = 4$ with triangles; (e) $w^{\text{in}} = 4$ and $w^{\text{out}} = -1$ with crosses; (f) $w^{\text{in}} = -1$ and $w^{\text{out}} = 4$ with stars. The cases (c) and (e) show a spectrum similar to that of L_{in} in (a), where the eigenvalues have larger negative real parts than in case (d), whose spectrum is more similar to that of L_{out} in (b). The case (f) shows many eigenvalues with positive real parts.

topology, we also expect $w^{\text{in}} \gg |w^{\text{out}}|$ to ensure stability. The condition Eq. (5.11) on the rate-based learning parameters is slightly stronger than that derived in the case of full connectivity (Burkitt et al. 2007). In order to ensure that the firing-rate equilibrium is realisable, i.e., $\mu \geq \nu_0 > 0$ in Eq. (5.6), the condition Eq. (5.11) implies that $\tilde{W} < 0$, similar to the analysis by Burkitt et al. (2007). Last, the denser the recurrent connections are, the more attractive \mathcal{M}^* is, when the stability conditions on w^{in} and w^{out} are satisfied (details are provided in Appendix C.2.4).

5.3 Diffusion of the recurrent weights

We now look in more detail at the evolution of the individual weights, when the stability conditions for the fixed-point manifold determined in Sec. 5.2, namely Eq. (5.11) and $\tilde{W} < 0$, are met. Similar to the case of full connectivity (Burkitt et al. 2007), the recurrent weights individually diverge due to the autocorrelation of the Poisson neurons, which are stochastic point-processes. In this section, we show that this weight dispersion is affected by the connectivity density. Then we investigate some properties of the asymptotic

weight distribution.

5.3.1 Dispersion of the individual weights

To study the impact of recurrent connectivity upon the evolution of the recurrent weights, we use calculations involving the higher stochastic orders of the weight dynamics similar to those in Sec. 4.4. The connectivity density affects the weight dispersion, which stems from spike-triggering effects induced by the recurrent connections. This can be studied through the coefficients

$$\Gamma_{i,j,i',j}(t,t') := \left\langle \frac{dJ_{ij}^\omega(t)}{dt} \frac{dJ_{i'j}^\omega(t')}{dt} \right\rangle, \quad (5.12)$$

which are related to the second moment of the weight dynamics, as explained in Sec. 3.4.3. The derivative $dJ_{ij}^\omega(t)/dt$ of the weight J_{ij} corresponds to one trajectory ω of the stochastic process (one realisation of the network spiking history), and it consists of weight jumps for each spike and pair of spikes; see Appendix C.3 for details. Over a homogeneously connected network with n^J recurrent connections, the sum of the contributions to $\sum \Gamma_{i,j,i',j}(t,t')$ of these spike-triggering effects is

$$(n_{av}^J)^3 J_{av} \mu w^{\text{in}} (w^{\text{in}} + w^{\text{out}}), \quad (5.13)$$

where $n_{av}^J = n^J/N$ is the mean number of incoming recurrent connections per neuron. Details are given in Appendix C.3, Eq. (C.20); note that η^2 would be present if we had not rescaled time. Under the stability conditions $w^{\text{in}} \gg |w^{\text{out}}|$ and $\tilde{W} < 0$ (cf. Sec. 5.2.3), this sum is positive, which means limiting the increase of the weight variance; this effect is stronger when w^{in} is large and \tilde{W} is small.

The expression Eq. (5.13) is to be compared with the terms due to the first-order autocorrelation of the neurons. These first-order terms are independent of the connectivity and cause the variance of the recurrent weights J to increase, hence the divergence of individual weights. For a homogeneously connected network, the sum of these terms over

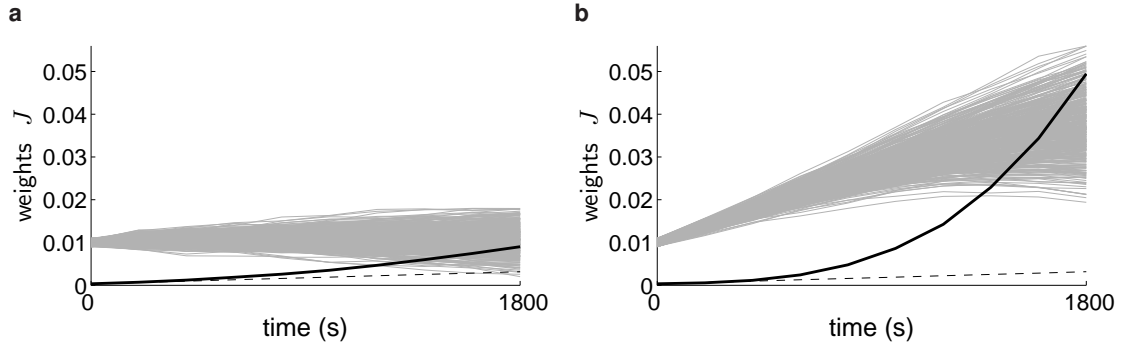


Figure 5.3: Comparison of the evolution of the weight variance between two networks of $N = 50$ neurons each, with (a) full connectivity and (b) 30% partial random connectivity. The weights were initialized to $0.01 (\pm 10\%)$ for both networks. The individual weights (grey bundles) of the fully connected network (a) tended to remain more clustered, whereas those of the partially connected network (b) became more dispersed over time; note that the equilibrium values for the mean weight in the two cases are different because of the connectivity density. After an initial period of linear growth at the predicted rate given in Eq. (5.14) (dashed line), the non-linear increase of the variance (thick solid lines - multiplied by a factor 1000 here) of the fully connected network (a) is slower than that for the partially connected network (b). No weight saturated or became quiescent during this simulation.

all connections is given by (Burkitt et al. 2007, Eq. (45))

$$n^J \left\{ \mu [(w^{\text{in}})^2 + (w^{\text{out}})^2] + \mu^2 \widetilde{W}^2 \right\} . \quad (5.14)$$

The ratio of the expressions in Eqs. (5.13) and (5.14), is given by

$$\frac{(n_{\text{av}}^J)^3 J_{\text{av}} w^{\text{in}} (w^{\text{in}} + w^{\text{out}})}{n^J \left\{ [(w^{\text{in}})^2 + (w^{\text{out}})^2] + \mu^2 \widetilde{W}^2 \right\}} \sim \frac{n^J}{N^2} n_{\text{av}}^J J_{\text{av}} . \quad (5.15)$$

This ratio ignores the details of the STDP parameters to focus on the connectivity density. We consider $n_{\text{av}}^J J_{\text{av}} = (\mu - \nu_0) / \mu < 1$ to be of order one. The denser the recurrent connections, the closer to one the ratio is. In the case of full connectivity, the ratio is approximately $n_{\text{av}}^J J_{\text{av}}$. This indicates that the weight variance increases more slowly for a network with full connectivity (Fig. 5.3(a)) than with partial connectivity (Fig. 5.3(b)).

5.3.2 Asymptotic pattern of recurrent weights

After a sufficiently long learning epoch, the recurrent weights J have evolved to either saturation or quiescence due to their increasing variance, while J remains on the fixed-point manifold \mathcal{M}^* (Burkitt et al. 2007). The matrix of the weights J exhibits a constant number of saturated weights on each row, when J remains in the attractive manifold \mathcal{M}^* , because the sum of incoming weights is then constant for each neuron as discussed in Sec. 5.2.2. However, the number of potentiated weights on each column, i.e., the sum of the outgoing weights for each neuron, may vary depending on the initial conditions, as shown in Appendix C.4 and illustrated in Fig. 5.4. The emerged weight structure in the recurrent network is thus strongly affected by the initial weight distribution. In other words, this asymptotic structure is not learned by the network in the sense of being constrained by STDP.

In the absence of external inputs, there is no weight structure to learn *per se*, but a structure may still emerge in addition to that remaining from the initial weight distribution, as described previously. For example, starting with full connectivity (except for self-connections), STDP tends to break the synaptic loops of length two between two neurons i and j , i.e., from a neuron to another one and then back to itself $j \rightarrow i \rightarrow j$. This can be explained by the second stochastic moment for two recurrent weights J_{ij} and J_{ji} in a similar manner to the calculation for the weight dispersion in Sec. 5.3.1. This second moment is related to $\Gamma_{i,j,j,i}(t, t')$, cf. Eq. (5.12), and its evaluation during the homeostatic equilibrium leads to a negative expression,

$$-2\mu [(w^{\text{in}})^2 + (w^{\text{out}})^2 + w^{\text{in}}w^{\text{out}}] < 0, \quad (5.16)$$

whatever the values for the STDP parameters, as explained in Appendix C.3.3. This means that STDP causes the weights J_{ij} and J_{ji} to diverge from each other due to the neuron autocorrelation effects in the network. Consequently, when the equilibrium value for the mean weight and the bounds are set such that half of the weights become saturated and half quiescent, the weight matrix J tends to become antisymmetric as illustrated in

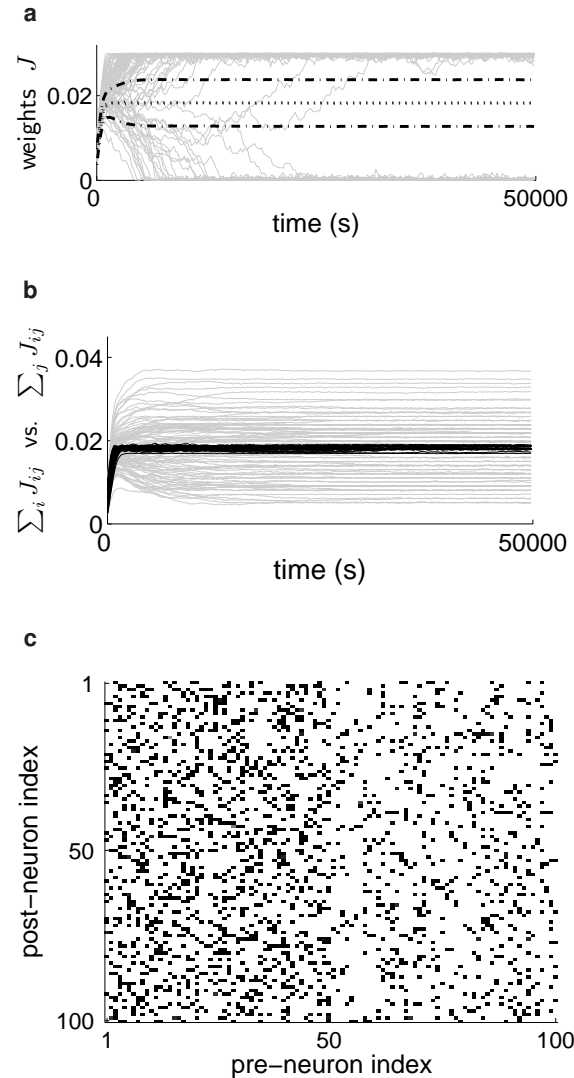


Figure 5.4: Evolution of the distributions of incoming and outgoing weights for $N = 100$ neurons. The connectivity was random with 30% probability and the initial weights were respectively set with a spread of $\pm 10\%$ around the following values: 0.1 from group 1 to 1; 0.5 from 1 to 2 and from 2 to 1; 0.25 from 2 to 2. (a) The individual weights (grey lines, with only a representative portion plotted) diverged to the bounds. The means of the incoming weights for group 1 and 2 (dotted lines, almost undistinguishable from each other) quickly converged to the same predicted equilibrium value. The means of the outgoing weights for groups 1 and 2 (dotted-dashed lines) stabilized to different values for a slightly longer period. (b) The sums for each neuron of the incoming weights (black traces) converged towards the predicted equilibrium value and remained clustered together whereas the sums of outgoing weights (grey traces) individually stabilized at different values eventually. (c) Matrix of J after 50000 s of learning. Darker elements indicate potentiated weights. The left side corresponding to weights coming from group 1 had more potentiated weights than the right side (weights from 2) at the end of the simulation, similar to the initial conditions.

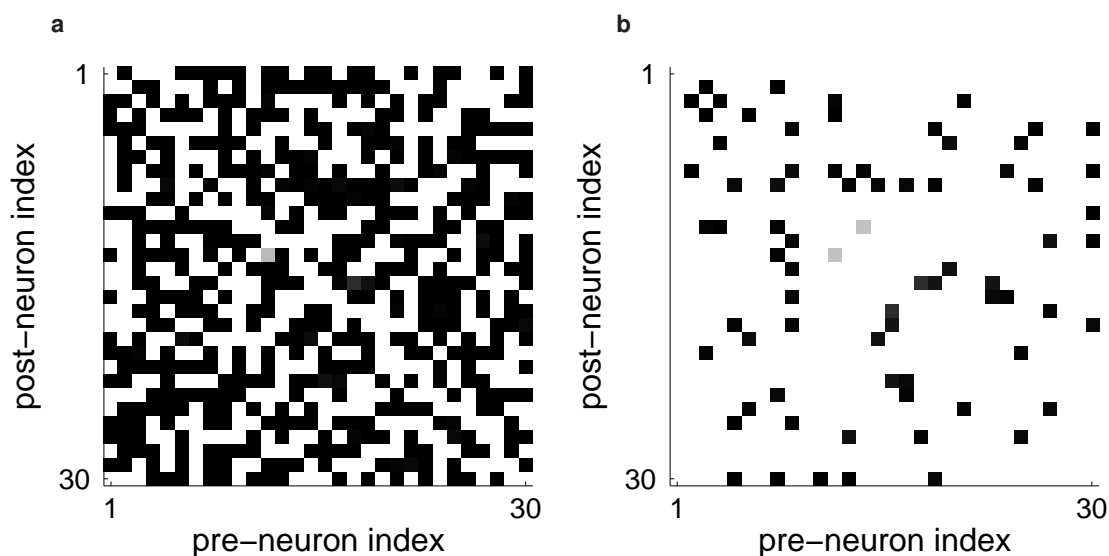


Figure 5.5: Illustrative results of numerical simulations with $N = 30$ neurons. Dark pixels indicate potentiated weights. The spontaneous rate ν_0 has been set to 22 Hz in order to obtain an equilibrium $J_{av} = 0.0177$ roughly equal to half of the weight bound $\Theta = 0.3$ such that the asymptotic matrix J has almost as many saturated as quiescent weights. (a) Antisymmetric pattern observed in the asymptotic matrix J when starting from initial full connectivity. For each pair of weights J_{ij} and J_{ji} , STDP almost always depressed one while potentiating the other, resulting in the breaking of most synaptic loops of length two in the network except for (b) 37 pairs of indices (i, j) and (j, i) (dark pixels). This corresponds to 4.25% of the total number of initial loops, which is to be compared with the expectation value of discrepancies $J_{av}/\Theta - 0.5 = 9\%$. The actual weight distribution is closer to an antisymmetric matrix than the theoretical prediction.

Fig. 5.5.

Other subtle constraints may be imposed on the recurrent weights by STDP, depending upon the specific connectivity topology and/or the distribution of delays. However, these are not the primary aim of this study and will not be pursued further here.

5.4 General case of learning on the recurrent weights with fixed input weights

In the remainder of this chapter, we consider that the network is stimulated by external input pools with fixed input connections. We analyze the general solution of Eqs. (3.22a-

3.22e), which for non-learning input weights K reduces to

$$\boldsymbol{v} = (\mathbf{1}_N - J)^{-1} \tilde{\boldsymbol{v}}, \quad (5.17a)$$

$$\dot{J} = \Phi_J \left[w^{\text{in}} \mathbf{e} \boldsymbol{v}^T + w^{\text{out}} \boldsymbol{v} \mathbf{e}^T + \tilde{W} \boldsymbol{v} \boldsymbol{v}^T + (\mathbf{1}_N - J)^{-1} \tilde{C} (\mathbf{1}_N - J)^{-1} \mathbf{1}^T \right], \quad (5.17b)$$

where we have rescaled time to remove η , and the following vector and matrix absorb the input parameters

$$\tilde{\boldsymbol{v}} := \nu_0 \mathbf{e} + K \hat{\boldsymbol{v}}, \quad (5.18)$$

$$\tilde{C} := K \hat{C}^{W*\zeta} K^T.$$

We show how STDP applied on the recurrent connections can induce both

- stable firing rates and thus stable mean incoming weights for each neuron;
- a specialisation of the recurrent weights through splitting of the outgoing weight distribution for each neuron.

5.4.1 Homeostatic equilibrium

We first examine the case of homeostatic equilibrium in the network, namely the situation when the means of the firing rates and of the weights over the whole network have reached an equilibrium. The mean value for a variable averaged over inputs, neurons or connections will be denoted using the subscript 'av'. The following differential equation for J_{av} is derived from Eq. (5.17b)

$$\dot{J}_{\text{av}} = (w^{\text{in}} + w^{\text{out}}) \nu_{\text{av}} + \left(\tilde{W} + \frac{\tilde{C}_{\text{av}}}{\tilde{\nu}_{\text{av}}^2} \right) \nu_{\text{av}}^2, \quad (5.19)$$

where we have used the following equality that comes from the averaging of Eq. (5.17a),

$$\left(1 - n_{\text{av}}^J J_{\text{av}} \right)^{-1} = \frac{\nu_{\text{av}}}{\tilde{\nu}_{\text{av}}}, \quad (5.20)$$

$n_{\text{av}}^J := n^J / N$ being the average number of presynaptic recurrent connections per neuron. This equation has the same form as that in the case with no external inputs in (5.1a-

5.1b) and can be analyzed in a similar manner, the change lying in replacement of \tilde{W} by $\tilde{W} + \tilde{C}_{av}/\tilde{v}_{av}^2$. The fixed point (ν_{av}^*, J_{av}^*) is

$$\begin{aligned}\nu_{av}^* &= -\frac{w^{\text{in}} + w^{\text{out}}}{\tilde{W} + \tilde{C}_{av}/\tilde{v}_{av}^2}, \\ J_{av}^* &= \frac{\nu_{av}^* - \tilde{\nu}_{av}}{n_{av}^J \nu_{av}^*}.\end{aligned}\quad (5.21)$$

Provided the homeostatic equilibrium is realisable, i.e., the mean firing rates and weights have positive equilibrium values, it is stable if and only if

$$\tilde{W} + \frac{\tilde{C}_{av}}{\tilde{v}_{av}^2} < 0. \quad (5.22)$$

For weak correlation, this condition reduces to $\tilde{W} < 0$. In order to ensure that the equilibrium mean firing rate is positive ($\nu_{av}^* > 0$), we require in addition that $w^{\text{in}} + w^{\text{out}} > 0$. These two conditions are the same as for the case of no external inputs (Burkitt et al. 2007). Note that the weight equilibrium is realisable only if $\nu_{av}^* > \tilde{\nu}_{av}$ (where $\tilde{\nu}_{av} \simeq \nu_0 + n_{av}^K K_{av} \hat{\nu}_{av}$), which requires $w^{\text{in}} + w^{\text{out}}$ to be sufficiently large.

5.4.2 Learning the input correlation structure

We now show that the stabilisation corresponding to the homeostatic equilibrium actually holds for all the individual neurons. In addition, STDP can also cause the weights to diverge according to the input correlation structure, thus implementing a robust specialisation in a similar way to the case of learning on the input connections (Chapter 4).

We decompose \tilde{C} into two components, one proportional to $\tilde{\mathbf{v}}\tilde{\mathbf{v}}^T$ and its complement

$$\begin{aligned}\tilde{C} &= \tilde{C}_{\parallel} + \tilde{C}_{\perp}, \\ \tilde{C}_{\parallel} &:= c_{\parallel} \tilde{\mathbf{v}}\tilde{\mathbf{v}}^T \quad \text{with } c_{\parallel} := \frac{\tilde{\mathbf{v}}^T \tilde{C} \tilde{\mathbf{v}}}{(\tilde{\mathbf{v}}^T \tilde{\mathbf{v}})^2}, \\ \tilde{\mathbf{v}}^T \tilde{C}_{\perp} \tilde{\mathbf{v}} &= 0,\end{aligned}\quad (5.23)$$

where $\tilde{\mathbf{v}}$ is defined in Eq. (5.18). Recall that the matrix J belongs to the vector subspace of $\mathbb{R}^{N \times N}$ defined by $\mathbb{M}_J := \{X \in \mathbb{R}^{N \times N}, \Phi_J(X) = X\}$, whose dimension is the number of

existing connections n^J .

5.4.3 Sufficient condition for existence of fixed points

We first analyze the special case where $\tilde{C}_\perp = 0$. Here, Eqs. (5.17a-5.17b) reduce to the same form as that obtained in the case of no external inputs (5.1a-5.1b), where $\nu_0 \mathbf{e}$ and \tilde{W} are replaced by $\tilde{\nu}$ and \tilde{W}' , respectively, with

$$\tilde{W}' := \tilde{W} + c_\parallel. \quad (5.24)$$

In particular, the fixed points (\mathbf{v}^*, J^*) of the dynamics correspond to homogeneous neuron firing rates

$$\mathbf{v}^* = \mu' \mathbf{e}, \quad (5.25a)$$

$$(\mathbf{1}_N - J^*) \mathbf{v}^* = \tilde{\nu}, \quad (5.25b)$$

where

$$\mu' := -\frac{w^{\text{in}} + w^{\text{out}}}{\tilde{W}'}. \quad (5.26)$$

For weak input correlations, μ' can be approximated by the equilibrium value μ for uncorrelated inputs

$$\mu' \simeq \mu := -\frac{w^{\text{in}} + w^{\text{out}}}{\tilde{W}}, \quad (5.27)$$

which means that the firing rates are in this case close to the equilibrium value corresponding to uncorrelated inputs. This characterizes all the fixed points provided each neuron in the network is part of a loop of synaptic connections (Sec. 5.2).

The manifold of all fixed points J^* denoted by \mathcal{M}^* is contained in an affine subspace of \mathbb{M}_J of dimension $n^J - N$, where n^J is the number of recurrent connections, according to Eq. (5.25b). Note that the matrix $\mathbf{1}_N - J^*$ must be invertible to be a valid fixed point as discussed previously in Sec. 3.4.2. When the fixed-point manifold \mathcal{M}^* is attractive, STDP causes J to converge towards \mathcal{M}^* , where it evolves due to higher orders of the stochastic process, the drift \dot{J} (cf. Sec. 3.6) being zero on \mathcal{M}^* . For homogeneous recurrent

connections, sufficient conditions such that \mathcal{M}^* is attractive are

$$w^{\text{in}} \gg |w^{\text{out}}| \quad \text{and} \quad \tilde{W} < 0; \quad (5.28)$$

refer to Sec. 5.2 for more details of the analysis and on higher stochastic orders than the drift, such as the resulting weight dispersion.

The condition $\tilde{C}_\perp = 0$ occurs in two particular situations: for uncorrelated inputs where $\tilde{C} = 0$; and in the case of homogeneous input connections since we have then $\tilde{\nu} \propto \mathbf{e}$ and $\tilde{C} \propto \mathbf{e}\mathbf{e}^T$. In the first case, the recurrent weights J can compensate for inhomogeneous input firing rates $\hat{\nu}$, causing the neuron firing rates ν to become homogeneous: the *incoming* recurrent weights of the less stimulated neurons become potentiated and those of the more stimulated neurons become depressed. In both cases, however, the asymptotic distribution of the *outgoing* recurrent weights is not constrained by STDP and strongly depends on the initial conditions, analogous to what happens in the case of no external inputs.

5.4.4 Weight dynamics for arbitrary matrix \tilde{C}_\perp

The analysis above indicates that a non-zero input correlation structure and inhomogeneous input weights are required so that the recurrent network organizes in a way that represents the input structure, i.e., at least differently from the weight dynamics for the case of no external inputs. In other words, the interesting case in terms of weight specialisation corresponds to the absence of a fixed point for the dynamical system.

When $\tilde{C}_\perp \neq 0$, the equation $\dot{J} = 0$ in Eq. (5.17b) may have no solution, but the evolution of the recurrent weights J can still be related to the manifold \mathcal{M}^* . Equation Eq. (5.17b) can be rewritten

$$\dot{J} = \Phi_J \left[-w^{\text{out}} \mathbf{e} \mathbf{G}^T - w^{\text{in}} \mathbf{G} \mathbf{e}^T + \tilde{W}' \mathbf{G} \mathbf{G}^T + (\mathbb{1}_N - J)^{-1} \tilde{C}_\perp (\mathbb{1}_N - J)^{-1T} \right], \quad (5.29)$$

where the vector $\mathbf{G} := \nu - \mu' \mathbf{e}$ evaluates for each neuron the difference between its fir-

ing rate and the common equilibrium value. For weakly correlated inputs, \tilde{C}_\perp is small and, since $(\mathbb{1}_N - J)^{-1}$ is kept invertible with bounded norm, the last term in Eq. (5.29) can be considered to be a small perturbation on the evolution of G . When the conditions Eq. (5.28) such that the manifold \mathcal{M}^* is attractive are met, there exist initial conditions on the recurrent weights such that G will converge towards zero and remain in a neighborhood of zero for bounded perturbations \tilde{C}_\perp . This means that individual neuron firing rates are then all quasi-stable and close to the equilibrium value $\mu' \simeq \mu$, and that J converges towards the vicinity of \mathcal{M}^* , provided the input correlations are sufficiently small. A rigorous analysis would consider the domain of attraction for G , namely the set of initial recurrent weights for which $J(t)$ will be driven towards \mathcal{M}^* . We assume that this domain is sufficiently large based on our study of the homeostatic equilibrium in Sec. 5.4.1, which suggests that homogeneous initial recurrent weights will converge towards \mathcal{M}^* .

While remaining in the neighborhood of \mathcal{M}^* , $\tilde{C}_\perp \neq 0$ may cause ongoing structural evolution of the weights J when it becomes the leading order while the sum of the terms involving G in Eq. (5.29) converge to a quasi-equilibrium around zero. However, the analysis is difficult in the general case and we will only consider a simple but biologically relevant special case.

5.5 Network with two distinct input pathways

We now consider a network with two homogeneous input pools that each excite half of the recurrently connected neurons, as illustrated in Fig. 5.6. The input pools have homogeneous characteristics (viz., firing rates and correlations) within each of them and no correlation between them; the neuron groups also have homogeneous characteristics. The connectivity is homogeneous from pools to groups and between groups. This network topology can be obtained after symmetry breaking by applying STDP on the input connections with balanced correlation strength between the two input pools, similar to the analysis in Sec. 4.4 (Gütig et al. 2003). We show in this section how the recurrent weights can then organize in an unsupervised way according to the input correlation structure, leading to the emergence of functional organisation.

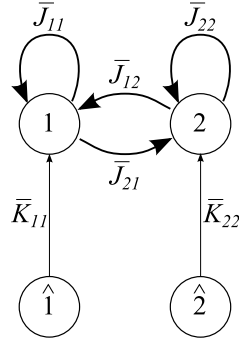


Figure 5.6: The recurrently connected neurons are divided into two groups (top circles), each being stimulated by one pool of external inputs (bottom circles). The pools and groups have homogeneous characteristics within each. The overline and the subscripts ‘1’ and ‘2’ correspond to the mean variables (\hat{v} , v , K , J , ...) over each pool of external inputs, group of neurons, etc.

Similar to the case of learning input weights K with fixed recurrent weights J analysed in Chapter 4, symmetries in the network allow us to reduce the dimensionality of the vector space \mathbb{M}_J where the matrix J evolves, in order to study the drift of J . For example, we define the variable \bar{J}_{12} as the sum of the incoming weights coming from group 2 averaged over the neurons of group 1, which gives, for full recurrent connectivity,

$$\bar{J}_{12} \simeq \frac{2}{N} \sum_{1 \leq i \leq N/2} \sum_{N/2+1 \leq j \leq N} J_{ij}. \quad (5.30)$$

Likewise, \bar{v}_1 is the mean firing rate of group 1. The variables in the reduced space correspond to equivalence classes defined modulo redistributions of incoming weights that do not modify the drift \dot{J} : in other words, two weight matrixes J and J' are in the same class \bar{J} if they induce the same drift $\dot{J} = \dot{J}'$ expressed in Eq. (5.17b). For the topology described in Fig. 5.6, these variables are

$$\begin{aligned} \bar{K} &= \begin{pmatrix} \bar{K}_{11} & 0 \\ 0 & \bar{K}_{22} \end{pmatrix}, \\ \bar{C} &= [W * \zeta](0) \begin{pmatrix} \bar{K}_{11}^2 \hat{c}_1 \bar{v}_1 & 0 \\ 0 & \bar{K}_{22}^2 \hat{c}_2 \bar{v}_2 \end{pmatrix}, \\ \bar{J} &= \begin{pmatrix} \bar{J}_{11} & \bar{J}_{12} \\ \bar{J}_{21} & \bar{J}_{22} \end{pmatrix}, \end{aligned} \quad (5.31)$$

$$\bar{\mathbf{v}} = \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix}.$$

The expression for $\bar{\mathbf{C}}$ with the correlation strengths \hat{c}_1 and \hat{c}_2 come from Eq. (3.29).

The inverse of $\mathbb{1}_2 - \bar{J}$ is

$$(\mathbb{1}_2 - \bar{J})^{-1} = \frac{1}{\Theta_{\bar{J}}} \begin{pmatrix} 1 - \bar{J}_{22} & \bar{J}_{12} \\ \bar{J}_{21} & 1 - \bar{J}_{11} \end{pmatrix}, \quad (5.32)$$

where $\Theta_{\bar{J}}$ is the determinant of $\mathbb{1}_2 - \bar{J}$

$$\Theta_{\bar{J}} := (1 - \bar{J}_{11})(1 - \bar{J}_{22}) - \bar{J}_{12}\bar{J}_{21}. \quad (5.33)$$

Substituting Eq. (5.32) in Eq. (5.17a), we obtain

$$\begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \end{pmatrix} = \frac{1}{\Theta_{\bar{J}}} \begin{pmatrix} (1 - \bar{J}_{22})\tilde{v}_1 + \bar{J}_{12}\tilde{v}_2 \\ \bar{J}_{21}\tilde{v}_1 + (1 - \bar{J}_{11})\tilde{v}_2 \end{pmatrix}. \quad (5.34)$$

The learning equation Eq. (5.17b) becomes

$$\begin{aligned} \dot{\bar{J}} &= w^{\text{in}} \begin{pmatrix} \bar{v}_1 & \bar{v}_2 \\ \bar{v}_1 & \bar{v}_2 \end{pmatrix} + w^{\text{out}} \begin{pmatrix} \bar{v}_1 & \bar{v}_1 \\ \bar{v}_2 & \bar{v}_2 \end{pmatrix} + \tilde{W} \begin{pmatrix} \bar{v}_1^2 & \bar{v}_1\bar{v}_2 \\ \bar{v}_1\bar{v}_2 & \bar{v}_2^2 \end{pmatrix} + \Omega, \\ \Omega &:= \frac{[W * \zeta](0)}{\Theta_{\bar{J}}^2} \begin{pmatrix} 1 - \bar{J}_{22} & \bar{J}_{12} \\ \bar{J}_{21} & 1 - \bar{J}_{11} \end{pmatrix} \begin{pmatrix} \bar{K}_{11}^2 \hat{c}_1 \tilde{v}_1 & 0 \\ 0 & \bar{K}_{22}^2 \hat{c}_2 \tilde{v}_2 \end{pmatrix} \begin{pmatrix} 1 - \bar{J}_{22} & \bar{J}_{21} \\ \bar{J}_{12} & 1 - \bar{J}_{11} \end{pmatrix}. \end{aligned} \quad (5.35)$$

Since the weights J are all positive and the spectrum of the matrix J is in the unit circle, as explained in Appendix C.1 (Burkitt et al. 2007), we have $0 \leq \sum_j \bar{J}_{ij} < 1$ for all i .

When the correlations strengths \hat{c}_1 and \hat{c}_2 are non-zero, the expression for $\bar{\mathbf{C}}$ in Eq. (5.31) generally corresponds to $\bar{\mathbf{C}}_{\perp} \neq 0$; hence the equation $\dot{\bar{J}} = 0$ has no solution except for specific choices of learning parameters. In the remainder of this section, we assume that the stability conditions in Eq. (5.28) are met and the correlation strengths \hat{c}_1 and \hat{c}_2 are small. As explained in Sec. 5.4.4, Eq. (5.35) causes the firing rates \bar{v}_1 and \bar{v}_2 to stabilize over time

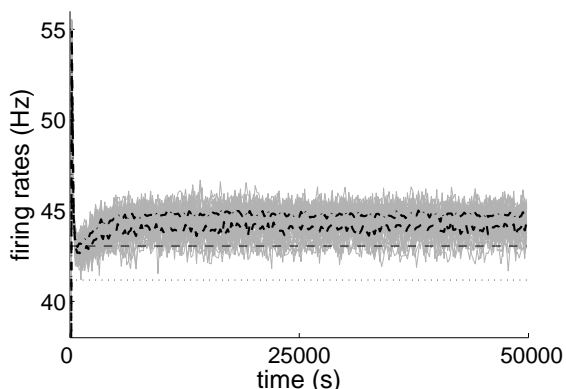


Figure 5.7: Evolution of the neuron firing rates. The network consisted of $N = 60$ neurons such that each half (#1-30 and #31-60) was stimulated by just one pool of 30 inputs, as illustrated by Fig. 5.6. The input firing rates were $\hat{v}_1 = 35$ Hz and $\hat{v}_2 = 30$ Hz; only input pool $\hat{1}$ had correlation ($\hat{c}_1 = 0.1$). After a transient from quiescence at time $t = 0$ up to 55 Hz, the individual firing rates (grey bundle) quickly converged towards the predicted equilibrium value μ' (calculated using Eq. (5.21), black thin dashed line), close to the value μ corresponding to uncorrelated inputs (black thin dotted line), and then remained quasi-homogeneous in the neighborhood of μ' . The thick dashed line (*bottom*) represents the mean firing rate for group 1 and the dashed-dotted line (*top*) the mean for group 2.

close to $\mu' \simeq \mu$, as illustrated in Fig. 5.7; the weight matrix J thus remains in the neighborhood of \mathcal{M}^* . This implies in particular that $\bar{J}_{11}\bar{v}_1 + \bar{J}_{12}\bar{v}_2 = \bar{v}_1 \simeq \text{const.}$, and consequently \bar{J}_{11} and \bar{J}_{12} will evolve in opposite directions. Likewise, $\bar{J}_{21}\bar{v}_1 + \bar{J}_{22}\bar{v}_2 = \bar{v}_2 = \text{const.}$ and \bar{J}_{21} and \bar{J}_{22} will diverge from each other. These two diverging behaviours are determined by the matrix $\bar{\bar{C}}_{\perp}$ when it is non-zero, which induces a specialisation of the recurrent connections. In the following, we analyze Ω directly in order to study the specialisation scheme, instead of $\bar{\bar{C}}_{\parallel}$ and $\bar{\bar{C}}_{\perp}$ separately.

5.5.1 One input pool with spike-time correlation and one uncorrelated pool

We first consider the case where only one pool has homogeneous non-zero spike-time correlation ($\hat{c}_1 > 0$) while the other has none ($\hat{c}_2 = 0$). The term of the learning equation Eq. (5.35) related to $\bar{\bar{C}}$ is

$$\Omega = \frac{[W * \zeta](0)\bar{K}_{11}^2\hat{c}_1\hat{v}_1}{\Theta_f^2} \begin{pmatrix} (1 - \bar{J}_{22})^2 & (1 - \bar{J}_{22})\bar{J}_{21} \\ (1 - \bar{J}_{22})\bar{J}_{21} & \bar{J}_{21}^2 \end{pmatrix}. \quad (5.36)$$

Subtracting the second term from the first term in the first row of Eq. (5.36) to evaluate the evolution of $\bar{J}_{11} - \bar{J}_{12}$, we obtain

$$\Omega_{11} - \Omega_{12} = \frac{[W * \zeta](0) \bar{K}_{11}^2 \hat{c}_1 \bar{v}_1}{\Theta_J^2} (1 - \bar{J}_{22})(1 - \bar{J}_{22} - \bar{J}_{21}). \quad (5.37)$$

The invertibility of $\mathbb{1}_N - J$ to ensure that the firing rates \bar{v} do not diverge implies that $\bar{J}_{22} + \bar{J}_{21} < 1$; it follows that $\bar{J}_{22} < 1$ also holds. Consequently, the evolution of \bar{J}_{11} and \bar{J}_{12} is determined by the sign of $[W * \zeta](0)$: there is potentiation of \bar{J}_{11} at the expense of \bar{J}_{12} for $[W * \zeta](0) > 0$, as illustrated in Fig. 5.8.

Likewise for $\bar{J}_{21} - \bar{J}_{22}$, subtracting the second term from the first term in the second row of Eq. (5.36) leads to

$$\Omega_{21} - \Omega_{22} = \frac{[W * \zeta](0) \bar{K}_{11}^2 \hat{c}_1 \bar{v}_1}{\Theta_J^2} \bar{J}_{21} (1 - \bar{J}_{22} - \bar{J}_{21}), \quad (5.38)$$

and when $[W * \zeta](0) > 0$, \bar{J}_{21} will be potentiated and \bar{J}_{22} depressed.

Putting it all together, neuron group 1, which receives correlated input, has its outgoing weights \bar{J}_{11} and \bar{J}_{21} potentiated when $[W * \zeta](0) > 0$. If this outcome can be intuitively understood for \bar{J}_{11} , it is not so straight-forward for \bar{J}_{21} . The converse situation occurs when $[W * \zeta](0) < 0$. Note that the particular value $W(0)$ does not play any role in this analysis.

In any case, the weights will diverge due to the drifts in Eqs. (5.37) and (5.38) until reaching the bounds. The distribution of the neuron firing rates may become bimodal but the discrepancies between \bar{v}_1 and \bar{v}_2 remain small for inputs with small delta-correlation, as illustrated in Fig. 5.7. Recall that the discrepancies between the individual firing rates and $\mu' \simeq \mu$ (weak correlation) relate to the absence of a fixed point for the dynamical system. The same applies for the homeostatic equilibrium of the weights in Fig. 5.8(a).

In summary, the input correlation structure determines the asymptotic distribution of the recurrent weights by strengthening (or weakening) the outgoing weights of the group that receives correlated inputs when $[W * \zeta](0)$ is positive (negative). The divergence of the weights (\bar{J}_{11} and \bar{J}_{21} vs. \bar{J}_{12} and \bar{J}_{22}) induces a robust specialisation, cf. Fig. 5.8(b).

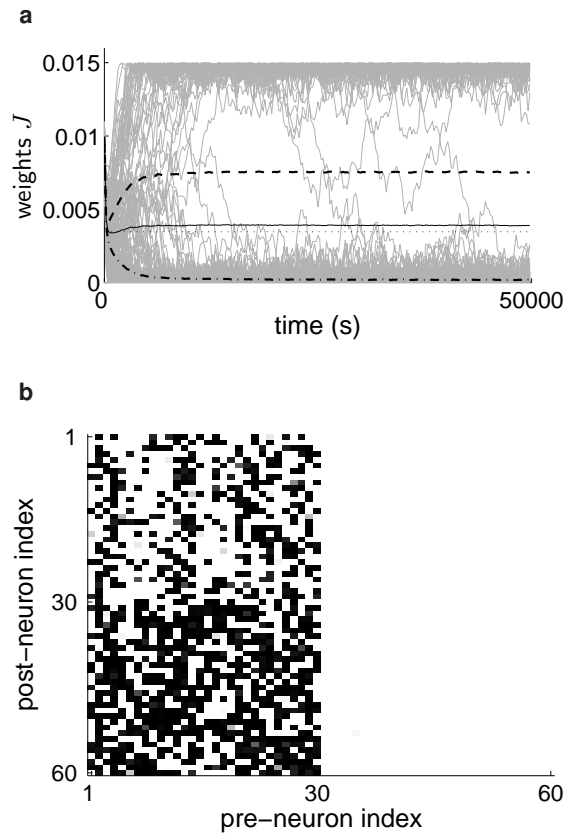


Figure 5.8: Strengthening of the outgoing weights of the neuron group that receives correlated input. The network has the same parameters as that in Fig. 5.7. (a) Evolution of the recurrent weights J . The individual weights (grey bundle, only a representative portion is plotted) diverged towards the bounds, while the homeostatic equilibrium (J_{av} in black thin solid line) was satisfied close to the predicted equilibrium value (black thin dotted line) calculated using Eq. (5.21). The recurrent weights coming from the first half ($\bar{J}_{11} + \bar{J}_{21}$ in thick dashed line) increased at the expense of those from the second half ($\bar{J}_{12} + \bar{J}_{22}$ in thick dashed-dotted line), i.e., the weights coming out of the first group that received correlated inputs were potentiated. (b) Weight matrix J after the emergence of the structure. Darker pixels indicate potentiated weights. Weights on the left side (corresponding to the mean in dashed line) were more potentiated than those on the right side.

These results are similar to those described in Chapter 4 when STDP is applied on the input connections with fixed recurrent weights: stability of some components of the weight matrix (in that case, the incoming weights for each neuron) in parallel with a diverging behaviour corresponding to a splitting between the different weight sets according to the input correlations.

This specialisation described above contrasts with the “redistribution” of the incoming weights in the case of uncorrelated inputs, which results in homogeneous equilibrium firing rates for the neurons, as explained in Sec. 5.4.3. In particular, when $\hat{v}_1 > \hat{v}_2$, we have

$$(1 - \bar{J}_{22} - \bar{J}_{21})\bar{v}_1 = (1 - \bar{J}_{11} - \bar{J}_{12})\bar{v}_2, \quad (5.39)$$

which means that the incoming weights to group 2, namely $\bar{J}_{22} + \bar{J}_{21}$, are potentiated at the expense of the weights to group 1, i.e., $\bar{J}_{22} + \bar{J}_{21}$. This weight compensation is a consequence of an equilibrium, hence it is weaker than the potentiation resulting from the diverging behaviour due to input correlation.

5.5.2 Two input pools with balanced spike-time correlations

For $\hat{c}_1 > 0$ and $\hat{c}_2 > 0$, we use an equivalent equation to Eq. (5.36) for \hat{c}_2 by permuting the indices in order to obtain an equation similar to Eq. (5.37) that gives the direction of the evolution of $\bar{J}_{11} - \bar{J}_{12}$,

$$\begin{aligned} \Omega_{11} - \Omega_{12} = & \frac{[W * \zeta](0)\bar{K}_{11}^2\hat{c}_1\bar{v}_1}{\Theta_J^2} (1 - \bar{J}_{22})(1 - \bar{J}_{22} - \bar{J}_{21}) \\ & - \frac{[W * \zeta](0)\bar{K}_{22}^2\hat{c}_2\bar{v}_2}{\Theta_J^2} \bar{J}_{12}(1 - \bar{J}_{11} - \bar{J}_{12}). \end{aligned} \quad (5.40)$$

Consider balanced input firing rates equal to \hat{v}_{av} , balanced correlation strengths equal to \hat{c}_{av} and balanced input weights $\bar{K}_{11} = \bar{K}_{22}$. For homogeneous initial recurrent weights equal to J_{av} , Eq. (5.40) reduces to

$$\Omega_{11} - \Omega_{12} = \frac{[W * \zeta](0)\bar{K}_{11}^2\hat{c}_{av}\hat{v}_{av}}{\Theta_J^2} (1 - n_{av}^I J_{av})^2. \quad (5.41)$$

Consequently, for $[W * \zeta](0) > 0$, \bar{J}_{11} will be initially potentiated at the expense of \bar{J}_{21} . Likewise, \bar{J}_{22} will be initially potentiated at the expense of \bar{J}_{12} . When starting from homogeneous recurrent weights, the weight evolution satisfies $\bar{J}_{11} \simeq \bar{J}_{22}$ and $\bar{J}_{12} \simeq \bar{J}_{21}$ at the beginning of learning. This leads to

$$\Omega_{11} - \Omega_{12} = \frac{[W * \zeta](0) \bar{K}_{11}^2 \hat{c}_{av} \hat{v}_{av}}{\Theta_{\bar{J}}^2} (1 - \bar{J}_{11} - \bar{J}_{12})^2. \quad (5.42)$$

As a result, \bar{J}_{11} becomes increasingly potentiated and \bar{J}_{12} depressed. The condition $[W * \zeta](0) > 0$ implies the strengthening of the within-group connections due to the input correlation when starting from homogeneous initial weights J , as illustrated in Fig. 5.9. This conclusion still holds when there are small inhomogeneities between the input firing rates, correlations or input weights. In the converse situation, when $[W * \zeta](0) < 0$, the within-group connections were weakened and the between-group connections were strengthened.

Recall that the analysis in Chapter 3 assumed very short recurrent delays d . Simulations run using $d = 0.4 \pm 0.2$ ms showed the expected outcomes described above for both $[W * \zeta](0)$ positive and negative. However, for longer delays $d \simeq 3 - 50$ ms, results similar to that in Sec. 5.5.1, or even the opposite of the expected behaviour for $[W * \zeta](0) > 0$ were observed, i.e., depression instead of potentiation of the within-group weights. The specialisation observed in numerical simulation was weaker in this case than that described in Sec. 5.5.1. The desired strengthening of within-group connections was obtained for larger recurrent delays ($d = 3 \pm 1$ ms) using a different learning window function W shifted such that $W(u) = 0$ for $u = t^{\text{in}} - t^{\text{out}} = 1$ ms, which corresponds to potentiation around the origin, as illustrated in Fig. 5.10. The potentiation observed in simulations is weaker for larger recurrent delays; shifting the curve of W more to the right allows the use of longer delays. This indicates the importance of the shape of W around the origin when interacting with the narrowly correlated neuronal activity within the network, due to the narrow within-pool input correlations. This is illustrated in Fig. 5.11, which shows that the correlation extends over the domain from -20 ms and $+20$ ms. The discrepancies between the theoretical prediction and the simulation result

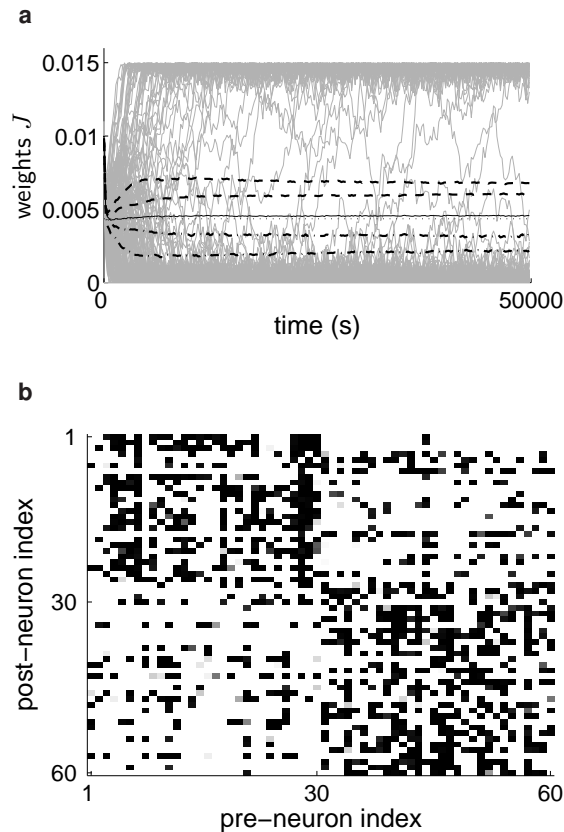


Figure 5.9: Within-group strengthening of the recurrent connections due to stimulation by correlated inputs. The network and figure are similar to Fig. 5.8, except that the two input pools have the same firing rate $\hat{\nu}_1 = \hat{\nu}_2 = 30$ Hz and balanced within-pool correlation ($\hat{c}_1 = \hat{c}_2 = 0.1$). (a) Evolution of the recurrent weights J . The individual weights (grey bundle, only a representative portion is plotted) diverged towards the bounds, while the homeostatic equilibrium was satisfied (mean J_{av} in black thin solid line, prediction in black thin dotted line). The two means of the within-group connections (\bar{J}_{11} and \bar{J}_{22} in dashed lines) were potentiated while those of the between-group connections (\bar{J}_{12} and \bar{J}_{21} in dashed-dotted lines) were depressed. (b) Matrix J after the emergence of the weight structure. Darker pixels indicate potentiated weights. The weights in the top-left and bottom-right quarters (corresponding to the two means in dashed line) were more potentiated than those in the top-right and bottom-left quarters.

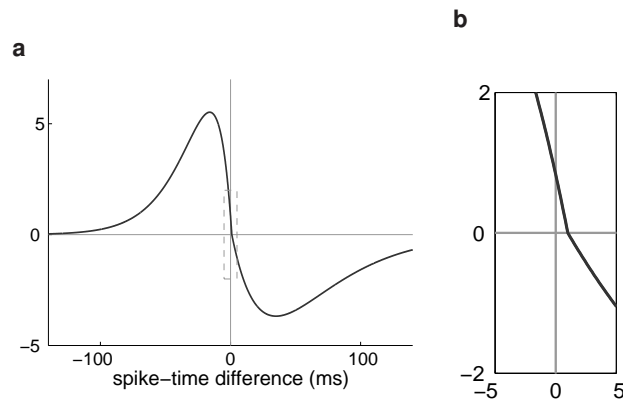


Figure 5.10: (a) Example of a different STDP window function. Each branch is an alpha function with the same time constants 17 ms (potentiation) and 34 ms (depression) as for Fig. 2.2. (b) Enlarged portion of (a) around the origin (a: dashed box) illustrating that the curve has been shifted to the right such that its zero corresponds to $u = 1$ ms (see insert on the right).

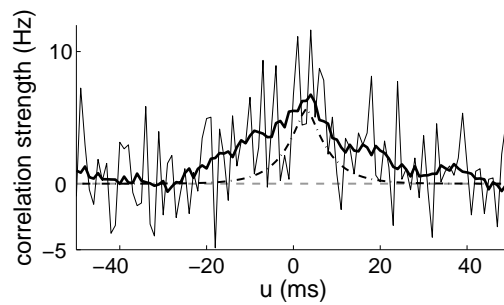


Figure 5.11: Cross-correlogram (black thin line) for two neurons from the same group in the network in Fig. 5.9 (without learning) simulated for 100 s. The black thick line represents the curve smoothed over 10 ms. The dashed-dotted line indicates the theoretical prediction using Eq. (3.18).

relates to the approximation made to derive Eq. (3.18), which only considered the first order of recurrence for the feedback connections, cf. Eq. (A.21). This explains why the actual distribution is broader.

5.6 Partial conclusion on plastic recurrent connections

Stability of the neuron firing rates, and thus of the incoming weight means, can be obtained for a wide range of learning parameters. The conditions $w^{\text{in}} > |w^{\text{out}}| > 0$ and $\tilde{W} < 0$ are sufficient in the case of weak input correlation. They correspond, respectively, to an increase of the weight due to each pre-synaptic spike, by a greater amount than

the effect of each single post-synaptic spike (either potentiation or depression), and to more depression than potentiation induced by the STDP window function W for uncorrelated inputs (negative integral value). This is in agreement with the theoretical analysis of the learning dynamics in a recurrently connected network with no external inputs (Burkitt et al. 2007) and earlier studies of numerical simulations of recurrently connected integrate-and-fire neurons (Song and Abbott 2001, Morrison et al. 2007).

The rate-based learning constants w^{in} and w^{out} were used in order to obtain homeostatic equilibrium for the weights, so that a structure could emerge depending on the input correlation. Unlike the STDP learning window W , they are not experimentally motivated. We expect the stability conclusions to hold in most cases for similar stabilizing mechanisms, such as weight normalisation (van Rossum et al. 2000) or a suitable weight-dependency for W (van Rossum et al. 2000, Gütig et al. 2003), so long as the resulting weight dynamics leads to an effective homeostatic equilibrium.

In order to obtain a non-trivial specialisation of the recurrent weights during the firing rate equilibrium, a network topology is necessary where different neuron groups receive distinct inputs with correlation. Otherwise, the weight dynamics is equivalent to that in a network with no external inputs. When conditions are met such as those described in Sec. 5.5.1 and 5.5.2, the individual weights exhibit strong competition that can result in the emergence of a feed-forward synaptic pathway or the strengthening of within-group connections for learning on J , as illustrated in Fig. 5.12 for two input pools. Very short recurrent delays were required to obtain within-group strengthening of recurrent connections (Sec. 5.5.2), which corresponds to the assumption in Sec. 3.4.1 that was made in order to derive the dynamical system Eqs. (3.22a-3.22c).

This robust specialisation scheme is determined by the covariance coefficient matrix $\hat{C}^{W*\zeta}$ in Eq. (3.22b). This matrix embodies the interplay between the correlation structure of the external inputs, the STDP window function W , and the PSP response kernel ϵ . The asymptotic weight distribution will be determined by the input correlations, when they are sufficiently large, rather than the input firing rates. For the network configurations in Fig. 5.12, the durations of ϵ and of the delays played a crucial role when STDP

modifies the recurrent connections. This is in contrast to the case of plastic input connections (Chapter 4) where the weights specialise in the same fashion irrespective of the delays and shapes of ϵ and W for Hebbian STDP. The different schemes of potentiation vs. depression that were observed depending upon $\hat{C}^{W*\zeta}$ may explain the contradictory behaviours observed in numerical simulations by Izhikevich et al. (2004) and Iglesias et al. (2005), which generated debate about whether STDP induces more or less synchronisation in recurrent networks. Indeed, stronger synchrony can be obtained by potentiating within-group connections, whereas desynchronisation would correspond to their depression.

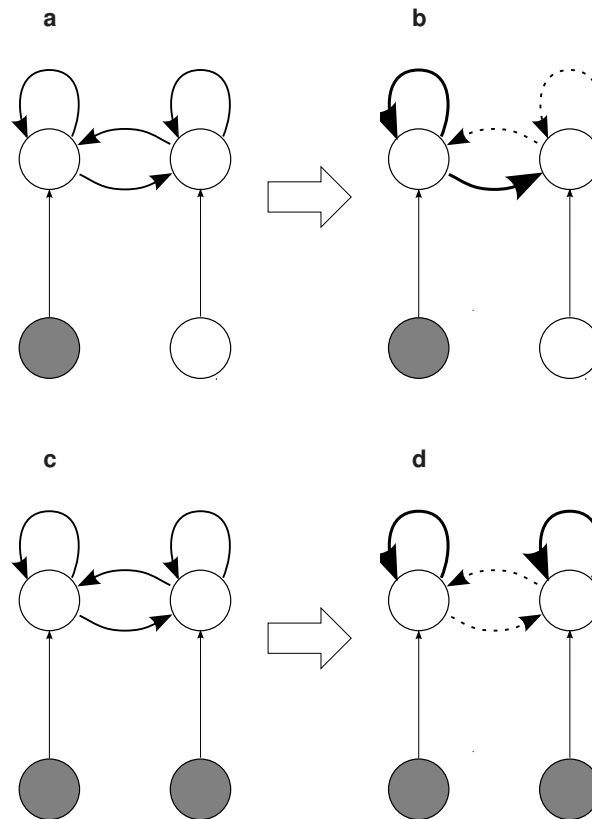


Figure 5.12: Schematic representation of the recurrent weight specialisation (a & c) before and (b & d) after the learning epoch for two input configurations (top vs. bottom). In each case, among the two sets of incoming connections to each neuron group (top circle) in the network, one becomes potentiated at the expense of the other. (a \Rightarrow b) When one input pool dominates in terms of spike-time correlations (filled bottom circle) compared to the other (open bottom circle), the neuron group that receives more correlations takes over in the recurrent network through the potentiation of its outgoing recurrent weights (thick arrow), while those of the other group are depressed (dashed arrow). (c \Rightarrow d) For two input pools with balanced spike-time correlations (filled bottom circles), the within-group recurrent connections are potentiated (thick arrow) while the between-group connections are depressed (dashed arrow).

Chapter 6

Self-organisation

This chapter relates the results presented in Chapter 4 and 5 to the context of network self-organisation such as that observed in the visual cortex. The framework developed in Chapter 3 is extended to incorporate a weight-dependent STDP. In Sec. 6.3, preliminary results of the implications of STDP in term of signal processing are presented, as a further step in the theory towards the domain of machine learning.

6.1 Introduction

IN THIS chapter, a weight-dependent STDP rule is considered as described in Eq. (2.2). The influence of the corresponding non-linearity will be illustrated through the scaling functions f_+ and f_- inspired by Gütig et al. (2003):

$$\begin{aligned} f_+(J) &= \left(1 - \frac{J}{J_{\max}}\right)^\gamma, \\ f_-(J) &= \left(\frac{J}{J_{\max}}\right)^\gamma. \end{aligned} \tag{6.1}$$

The parameter γ (in this chapter) relates to the degree of the weight dependence in the model: $\gamma = 0$ corresponds to additive STDP and $\gamma = 1$ to a multiplicative version similar to that used by van Rossum et al. (2000) for the depression side. The influence of γ is illustrated in Fig. 6.1.

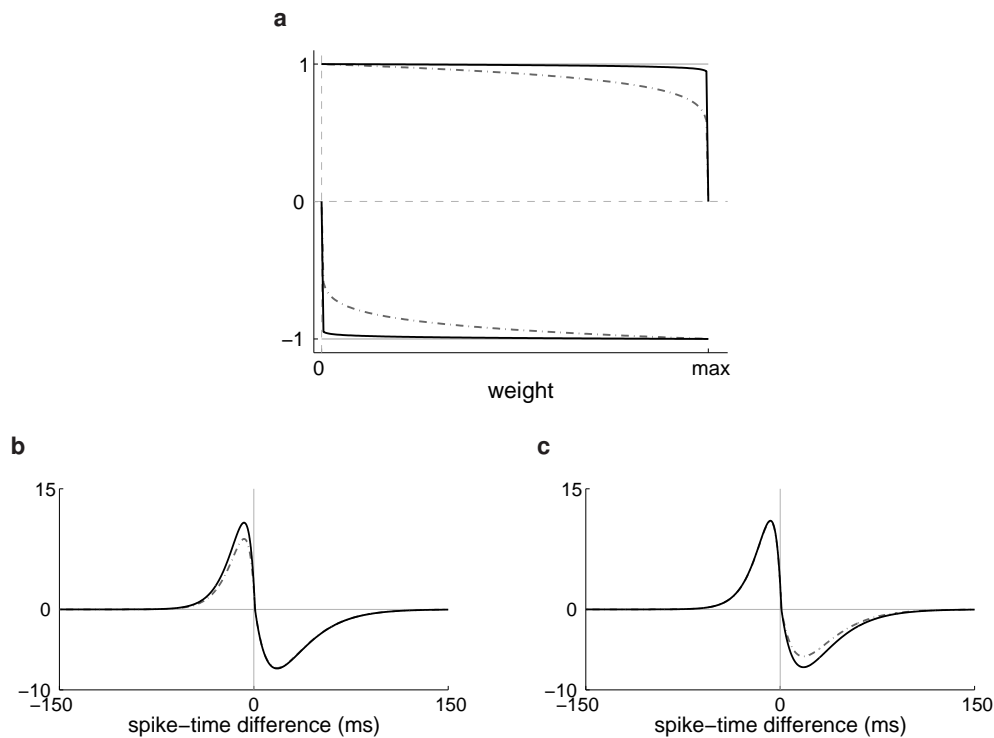


Figure 6.1: Example of weight-dependent STDP window function. (a) Representation of the functions $-f_-$ (bottom curve) and f_+ (top curve) that determine the weight dependence in the STDP model. The thick black solid curves correspond to $\gamma = 0.01$, the thick dashed-dotted grey curves to $\gamma = 0.1$ for the expression in Eq. (6.1); additive STDP ($\gamma = 0$) corresponds to constant functions at $+1$ and -1 , respectively. The use of f_+ and f_- leads to less potentiation and more depression for (b) a strong synapse $J = 0.9J_{\max}$ compared to (c) a weak synapse $J = 0.1J_{\max}$. The dependence upon the spike-time difference is taken care of by one alpha function W_+ for depression (right curves) with time constant 8.5 ms and likewise W_- for potentiation (left curves) with time constant 17 ms.

6.2 Effect of the weight dependence of STDP upon the learning dynamics

We obtain the following differential equations to describe the evolution of the drift of the input weight K_{ik} and recurrent weight J_{ij} :

$$\begin{aligned} \dot{K}_{ik} \simeq & \eta \left\{ w^{\text{in}} \hat{v}_k + w^{\text{out}} v_i + \left[f_+(K_{ik}) \tilde{W}_+ - f_-(K_{ik}) \tilde{W}_- \right] \hat{v}_k v_i \right. \\ & \left. + f_+(K_{ik}) F_{ik}^{W_+} - f_-(K_{ik}) F_{ik}^{W_-} \right\} \end{aligned} \quad (6.2a)$$

$$\begin{aligned} \dot{J}_{ij} \simeq & \eta \left\{ w^{\text{in}} v_j + w^{\text{out}} v_i + \left[f_+(J_{ij}) \tilde{W}_+ - f_-(J_{ij}) \tilde{W}_- \right] v_j v_i \right. \\ & \left. + f_+(J_{ij}) C_{ij}^{W_+} - f_-(J_{ij}) C_{ij}^{W_-} \right\}. \end{aligned} \quad (6.2b)$$

6.2.1 Homeostatic equilibrium

Neglecting the inhomogeneities of the network and the covariance terms in the learning equations Eqs. (6.2a) and (6.2b), we obtain the following equations for the mean input and recurrent weights:

$$\begin{aligned} \dot{K}_{\text{av}} \simeq & \eta \left\{ w^{\text{in}} \hat{v}_{\text{av}} + w^{\text{out}} v_{\text{av}} + \left[f_+(K_{\text{av}}) \tilde{W}_+ - f_-(K_{\text{av}}) \tilde{W}_- \right] \hat{v}_{\text{av}} v_{\text{av}} \right\} \\ \equiv & \eta G(v_{\text{av}}, K_{\text{av}}) \end{aligned} \quad (6.3a)$$

$$\begin{aligned} \dot{J}_{\text{av}} \simeq & \eta \left\{ (w^{\text{in}} + w^{\text{out}}) v_{\text{av}} + \left[f_+(J_{\text{av}}) \tilde{W}_+ - f_-(J_{\text{av}}) \tilde{W}_- \right] v_{\text{av}}^2 \right\} \\ \equiv & \eta H(v_{\text{av}}, J_{\text{av}}). \end{aligned} \quad (6.3b)$$

For the sake of simplicity, the functions G and H have been defined, as well as

$$g(x) := f_+(x) \tilde{W}_+ - f_-(x) \tilde{W}_-. \quad (6.4)$$

A fixed-point $(v_{\text{av}}^*, K_{\text{av}}^*, J_{\text{av}}^*)$ of this dynamical system must nullify the above expressions for \dot{K}_{av} and \dot{J}_{av} . We require a non-zero equilibrium value for the mean firing rate $v_{\text{av}}^* \neq 0$ in Eqs. (6.3a) and (6.3b). Note that weight-dependent STDP is necessary here, as additive

STDP leads to the two following equalities

$$\begin{aligned} \nu_{\text{av}} &= -\frac{w^{\text{in}} \hat{\nu}_{\text{av}}}{w^{\text{out}} + (\tilde{W}_+ - \tilde{W}_-) \hat{\nu}_{\text{av}}} \\ \nu_{\text{av}} &= -\frac{w^{\text{in}} + w^{\text{out}}}{\tilde{W}_+ - \tilde{W}_-}, \end{aligned} \quad (6.5)$$

which cannot be satisfied simultaneously for general choices of input and learning parameters.

Influence of rate-based learning terms

The particular case where $w^{\text{in}} = w^{\text{out}} = 0$ has been previously studied for a single neuron (van Rossum et al. 2000, Burkitt et al. 2004). From Eqs. (6.3a) and (6.3b), we have

$$g(K_{\text{av}}^*) = g(J_{\text{av}}^*) = 0. \quad (6.6)$$

In this case, the equilibrium values of the mean weights, K_{av}^* and J_{av}^* , only depend upon the function g ; they are independent of the input firing rate $\hat{\nu}_{\text{av}}$. The equilibrium value of the mean firing rate for the neurons is then determined by Eq. (3.22a).

In the case where $w^{\text{in}} \neq 0$ and $w^{\text{out}} \neq 0$, the following necessary conditions must be satisfied in order to obtain a non-zero equilibrium mean firing rate ν_{av}^*

$$w^{\text{in}} \hat{\nu}_{\text{av}} + [w^{\text{out}} + g(K_{\text{av}}^*) \hat{\nu}_{\text{av}}] \nu_{\text{av}}^* = 0 \quad (6.7a)$$

$$w^{\text{in}} + w^{\text{out}} + g(J_{\text{av}}^*) \nu_{\text{av}}^* = 0. \quad (6.7b)$$

Except for particular values of the input and learning parameters, this implies

$$w^{\text{in}} g(J_{\text{av}}^*) \hat{\nu}_{\text{av}} = (w^{\text{in}} + w^{\text{out}}) [w^{\text{out}} + g(K_{\text{av}}^*) \hat{\nu}_{\text{av}}]. \quad (6.8)$$

The consistency equation Eq. (3.22a) for the firing rates gives the additional constraint

$$-\frac{(w^{\text{in}} + w^{\text{out}}) (1 - n_{\text{av}}^I J_{\text{av}}^*)}{g(J_{\text{av}}^*)} = v_0 + n_{\text{av}}^K K_{\text{av}}^* \hat{v}_{\text{av}}, \quad (6.9)$$

after using the equality Eq. (6.7b). Combining Eqs. (6.8) and (6.9), the mean recurrent weight J_{av} must be a zero of the following function h at the equilibrium:

$$h(J_{\text{av}}^*) = 0 \quad (6.10)$$

with

$$h(J_{\text{av}}) := w^{\text{in}} g(J_{\text{av}}) \hat{v}_{\text{av}} - (w^{\text{in}} + w^{\text{out}}) \left[w^{\text{out}} + g \left(-\frac{(w^{\text{in}} + w^{\text{out}}) (1 - n_{\text{av}}^I J_{\text{av}})}{n_{\text{av}}^K g(J_{\text{av}}) \hat{v}_{\text{av}}} - \frac{v_0}{n_{\text{av}}^K \hat{v}_{\text{av}}} \right) \hat{v}_{\text{av}} \right]. \quad (6.11)$$

Fig. 6.2 illustrates the change of h defined in Eq. (6.11) and the corresponding equilibrium values of K_{av}^* and J_{av}^* for different input firing rates \hat{v}_{av} . In the presence of w^{in} and w^{out} , the weight equilibrium values are still uniquely determined, but depend upon the mean input firing rate \hat{v}_{av} . It is possible to choose parameters such that this dependency is weak: see Fig. 6.2(c) for J_{av}^* .

Stability

In the presence of w^{in} and w^{out} , the stability of the fixed point $(v_{\text{av}}^*, K_{\text{av}}^*, J_{\text{av}}^*)$ is given by the Jacobian matrix expressed using the partial derivatives of G and H in Eqs. (6.3a) and (6.3b):

$$\begin{pmatrix} \frac{\partial G}{\partial K_{\text{av}}} (v_{\text{av}}^*, K_{\text{av}}^*, J_{\text{av}}^*) & \frac{\partial G}{\partial J_{\text{av}}} (v_{\text{av}}^*, K_{\text{av}}^*, J_{\text{av}}^*) \\ \frac{\partial H}{\partial K_{\text{av}}} (v_{\text{av}}^*, K_{\text{av}}^*, J_{\text{av}}^*) & \frac{\partial H}{\partial J_{\text{av}}} (v_{\text{av}}^*, K_{\text{av}}^*, J_{\text{av}}^*) \end{pmatrix} =: \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{pmatrix}, \quad (6.12)$$

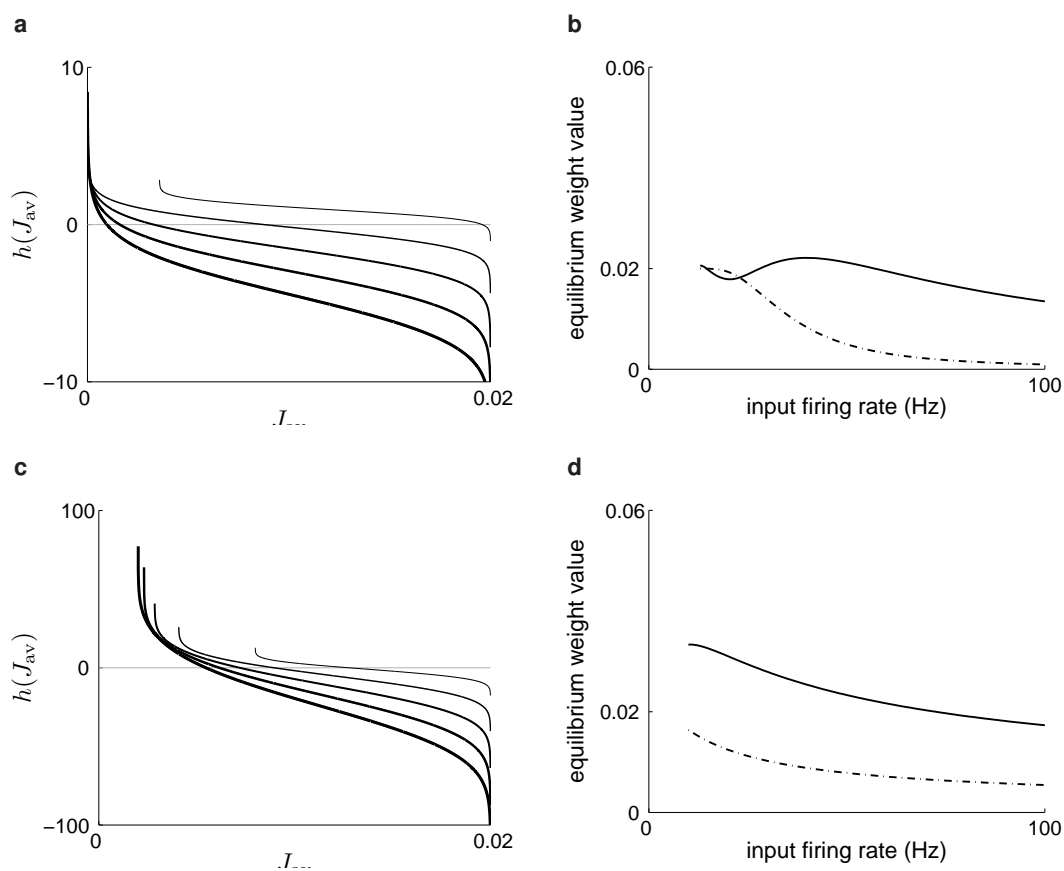


Figure 6.2: Influence of weight-dependent STDP upon the homeostatic equilibrium. Comparison between (a & b) almost additive STDP corresponding to $\gamma = 0.01$ and (c & d) a stronger weight dependence with $\gamma = 0.1$. (a & c) Plots of the function h in Eq. (6.11) for $\hat{v}_{av} = 20, 40, 60, 80$ and 100 Hz (thin to thick line). The zero of h when crossing the origin (grey dashed line) corresponds to the equilibrium value J_{av}^* . The plot is restricted to the domain of J_{av} for which the value K_{av} given by Eq. (6.9) is within the bounds. (b & d) Equilibrium mean weights K_{av}^* (solid line) and J_{av}^* (dashed-dotted line) as functions of the input firing rate \hat{v}_{av} . Parameters were $n_{av}^K = 100$, $n_{av}^J = 100$, $K_{max} = 0.06$, $J_{max} = 0.02$, 30% partial connectivity and those detailed in Appendix D.

where

$$\begin{aligned}
 \alpha_1 &= -w^{in} \frac{n_{av}^K \hat{v}_{av}}{(1 - n_{av}^J J_{av}) v_{av}^*} + g'(K_{av}^*) \hat{v}_{av} v_{av}^*, \\
 \alpha_2 &= -w^{in} \frac{n_{av}^J \hat{v}_{av}}{1 - n_{av}^J J_{av}}, \\
 \alpha_3 &= -(w^{in} + w^{out}) \frac{n_{av}^K \hat{v}_{av}}{1 - n_{av}^J J_{av}}, \\
 \alpha_4 &= -(w^{in} + w^{out}) \frac{n_{av}^J v_{av}^*}{1 - n_{av}^J J_{av}} + g'(J_{av}^*) (v_{av}^*)^2.
 \end{aligned} \tag{6.13}$$

We have used the equalities in Eq. (6.7a) and

$$\begin{aligned}\frac{\partial v_{av}}{\partial K_{av}} &= \frac{n_{av}^K \hat{v}_{av}}{1 - n_{av}^J J_{av}} \\ \frac{\partial v_{av}}{\partial J_{av}} &= \frac{n_{av}^J v_{av}}{1 - n_{av}^J J_{av}}.\end{aligned}\quad (6.14)$$

The stability of the mean weights K_{av}^* and J_{av}^* implies that of v_{av}^* ; it requires that the Jacobian matrix has eigenvalues with negative real parts. In other words, the trace of the Jacobian matrix must be negative and its determinant must be positive:

$$\begin{aligned}\alpha_1 + \alpha_4 &< 0 \\ \alpha_1 \alpha_4 - \alpha_2 \alpha_3 &> 0.\end{aligned}\quad (6.15)$$

We actually require the stronger conditions $\alpha_1 < 0$ and $\alpha_4 < 0$ to ensure stability for the weight dynamics when only the input or the recurrent weights are plastic, while the other set remains fixed. The following conditions are sufficient to ensure stability for any input firing rate

$$\begin{aligned}w^{\text{in}} + w^{\text{out}} &> 0 \\ w^{\text{in}} &> 0 \\ g' &< 0,\end{aligned}\quad (6.16)$$

where g is defined in Eq. (6.4). Note that, in the absence of the rate-based terms w^{in} and w^{out} , the condition $g' < 0$ leads to stability since $g'(K_{av}^*)$ and $g'(J_{av}^*)$ are negative.

6.2.2 Weight specialisation

Now we consider the homeostatic equilibrium to be satisfied, which means that the term related to the spike-time covariance in (6.2a) becomes the leading order. For all these terms to generate effective weight specialisation amongst the synapses, they must cause the weights to exhibit a diverging behaviour in a similar manner to additive STDP when starting from a homogeneous weight distribution (see Sec. 4.5). This means that, for each

k , the term $f_+(K_{\text{av}}^*) F_k^{W_+} - f_-(K_{\text{av}}^*) F_k^{W_-}$ must have the same sign as $F_k^{W_+} - F_k^{W_-}$. This is the case irrespective of the input correlation structure whenever the fixed point K_{av}^* corresponds to similar values:

$$f_-(K_{\text{av}}^*) \simeq f_+(K_{\text{av}}^*) . \quad (6.17)$$

It follows that the splitting of the weights is not affected by the weight dependence of STDP provided the mean weight equilibrium value is far from the bounds (cf. Fig. 6.1), which means that Eq. (6.17) is satisfied. This is supported by previous results, which showed that our choice of monotonic functions f_- and f_+ ensures the splitting of weights between homogeneous pools with the same correlation level, not within pools (Meffin et al. 2006). This effective neuronal specialisation requires sufficiently strong input correlations, in agreement with the studies by Gütig et al. (2003) and results presented in Chapter 4.

In the special case of two input pools that have within-pool correlation with respective levels \hat{c}_1 and \hat{c}_2 and firing rate \hat{v}_0 as described in Sec. 3.5, we have, for two inputs k and l from the first pool,

$$\begin{aligned} C_{kl}^{W_+ * \epsilon} &= \hat{c}_1 \hat{v}_0 [W_+ * \epsilon](0) > 0 \\ C_{kl}^{W_- * \epsilon} &= 0 , \end{aligned} \quad (6.18)$$

and likewise for \hat{c}_2 for the second pool. Since f_+ is positive, the signs of the vector $F^{W_+} \hat{\mathbf{h}}$ are determined. Consequently, the scheme of potentiation vs. depression for the input weight starting from an initially homogeneous distribution is given by the relative correlation levels, in the same fashion as for additive STDP. For the recurrent connections and $C^{W_{\pm}}$, the coefficients $C_{kl}^{W_+ * \zeta}$ and $C_{kl}^{W_- * \zeta}$ are strong and weak, respectively, for a suitable choice of STDP window functions W_{\pm} which results in potentiation at the origin for small u (STDP function shifted to the right, cf. Fig. 5.10). It then follows that the specialisation scheme for the recurrent weights is similar to that for additive STDP.

Figure 6.3 corroborates these predictions for an initially homogeneous network stimulated by two input pools: STDP induces both stabilization of the mean weights (rep-

resented in Fig. 6.3(a-b) by thick dashed lines) and a specialisation of the individual weights. Neurons in the network become selective to one of the two input pools (Fig. 6.3(c)). After labelling each neuron according to its specialisation, thus defining two groups, the within-group connections (solid lines in Fig. 6.3(d)) in the recurrent network are observed to be strengthened at the expense of the between-group connections (dashed lines). At the end of the learning epoch, we obtain the emergence of two groups of neurons (40 and 60 neurons, respectively, in this simulation), each being selective to a different input pool, see Fig. 6.3(e-f).

6.3 Representation of the input correlation in the weight structure for a single neuron

The previous section showed that weight-dependent STDP can perform effective neuronal specialisation in a recurrent network stimulated by two external pools. In order to further investigate the functional implications, we now focus on some aspects of the computation performed by STDP and show how the input spike-time correlation structure can be encoded in the input weight structure. We constrain this section to a single neuron and consider a more elaborate input structure than previously. The purpose of these preliminary results is to obtain more insight in the general unsupervised learning scheme induced by STDP and establish links with the domain of machine learning.

6.3.1 Structure of the input spike trains

Previous work (Kempster et al. 1999, Gütig et al. 2003, Meffin et al. 2006) showed the major role played by the input correlations in determining the weight dynamics induced by pairwise STDP. We consider in this section an extension of the configurations previously studied: M inputs are partitioned into m homogeneous pools of the same size with distinct within-pool correlation levels \hat{c}_l , $1 \leq l \leq m$; inputs from different pools are not correlated with each other. We assume the correlations to be sorted in increasing order: $0 \leq \hat{c}_1 < \dots < \hat{c}_l < \dots < \hat{c}_m < 1$. We constrain our study to the same input firing rate $\hat{\nu}_0$ for all input pools, assuming that not-too-large inhomogeneities in the firing rates

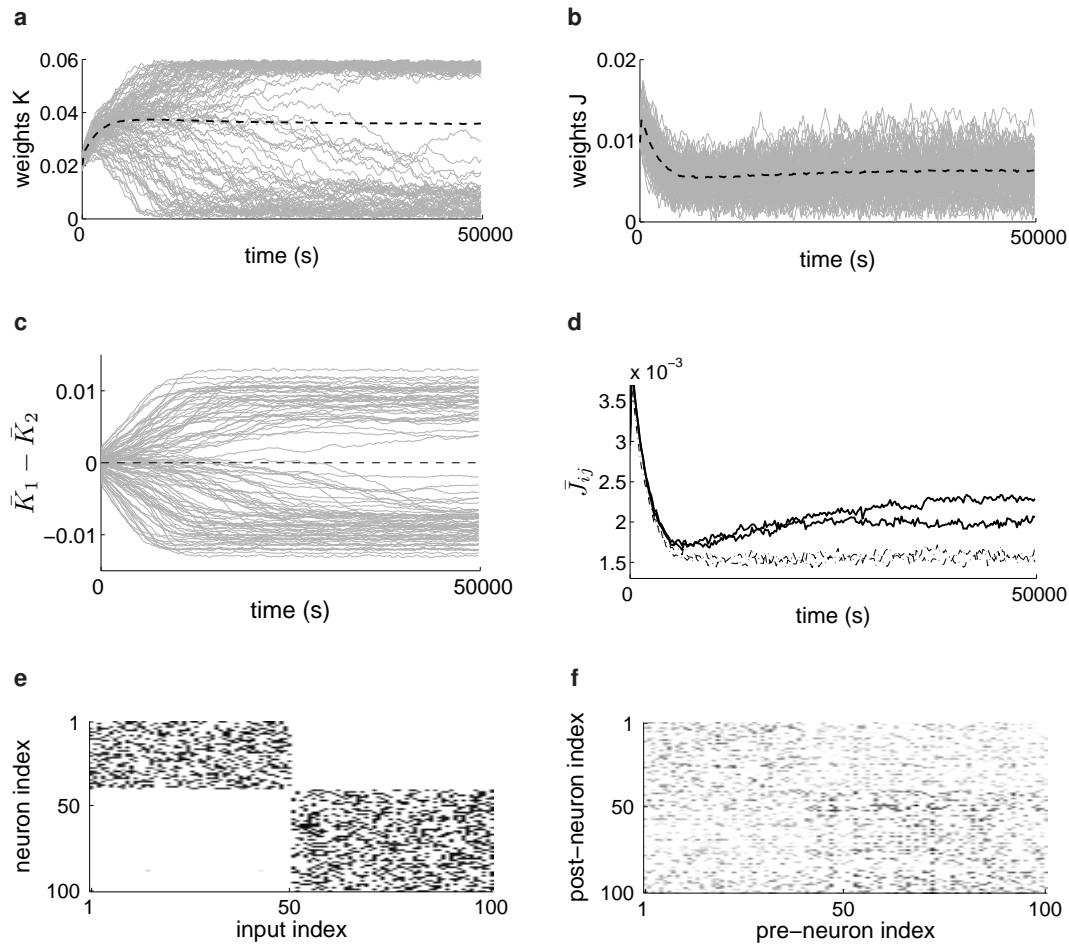


Figure 6.3: Simultaneous evolution of the input and recurrent weights. The network consists of $N = 100$ neurons stimulated by two pools of $M/2 = 50$ inputs each. The two pools have the same firing rate 30 Hz and correlation level $\hat{c} = 0.2$. The initial weights were homogeneous. (a-b) Traces of the individual weights (grey bundles) and stabilisation of their means (thick dashed lines). The input weight distribution (a) clearly became bimodal whereas the recurrent weight (b) remained unimodal. (c) Neuronal specialisation. Traces of the difference between the input weights from the first and second pools. A first group of 40 neurons become selective to the first input pool (increasing curves) while a second group of 60 did so to the second pool (decreasing curves). (d) Structuring of the recurrent connections. The connections within the two neuronal groups defined above became potentiated (solid lines) while those between the two groups were depressed (dashed lines). (e-f) Asymptotic weight matrices for the input and recurrent weights. Darker pixels indicate potentiated weights. The indices of the neurons have been arranged according to which group they belong to.

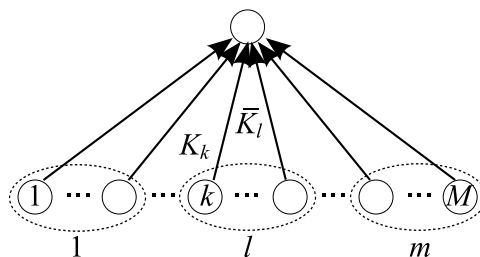


Figure 6.4: Schematic representation of one neuron (top circle) stimulated by M inputs (bottom circles) partitioned into m pools (dashed ellipsoids). The input connection from the external source k has a plastic weight $K_k(t)$; the mean weight over pool l is denoted by $\bar{K}_l(t)$.

between the pools do not impair the conclusions, as was shown for the case of two input pools.

We study the weight dynamics for the input connections of a single neuron that is stimulated by these m pools, as illustrated in Fig. 6.4. We denote by $S(t)$ and $\hat{S}_k(t)$ the spike trains of the neuron and each external input k , respectively. The connection from input k to the neuron has plastic weight $K_k(t)$.

6.3.2 Capturing the weight dynamics

We use the framework developed in Chapter 3 to analyse the evolution of the input weights K_k . We obtain a differential equation to describe the evolution of the drift of the input weights K_k

$$\dot{K}_k \simeq w^{\text{in}} \hat{v}_k + w^{\text{out}} \nu + \left[f_+(K_k) \tilde{W}_+ - f_-(K_k) \tilde{W}_- \right] \hat{v}_k \nu + f_+(K_k) F_k^{W_+} - f_-(K_k) F_k^{W_-}, \quad (6.19)$$

where

$$\tilde{W}_\pm := \int W_\pm(u) du. \quad (6.20)$$

Time has been rescaled to remove the learning rate η . In this section, we use the following simplified notation: $\nu(t)$ and $\hat{v}_k(t)$ for the time-averaged firing rates of the neuron and each input k , respectively, and the covariance coefficient $F_k^{W_\pm}$ between the neuron and input k ; cf. Eq. (3.3).

We consider the weight-dependent STDP to lead to an effective homeostatic equilibrium, which means that the term related to the spike-time covariance in Eq. (6.19) becomes the leading order. We also assume that the spike-time covariances generate proper weight specialisation over the pool of synapses, i.e., Eq. (6.17) is satisfied. Now we examine the weight specialisation for the case of several input pools with within-pool correlation, cf. Sec. 6.3.1. We first use additive STDP and similar calculations to the analysis in Chapter 4, then we take the weight dependence into account.

6.3.3 Initial splitting of the input weights

For input pools that have within-pool correlation but no between-pool correlation, as described in Sec. 6.3.1, we have for two inputs k and l from a pool with correlation level \hat{c}_l and firing rate \hat{v}_0

$$\begin{aligned} C_{kl}^{W_+ * \epsilon} &= \hat{c}_l \hat{v}_0 [W_+ * \epsilon](0) > 0, \\ C_{kl}^{W_- * \epsilon} &= 0. \end{aligned} \quad (6.21)$$

We assume pools of the same size and reduce the dimension of the problem and examine the mean weights \bar{K}_l over each pool l to investigate the emergence of the structure. In this reduced space, the input covariance matrix \bar{C} is diagonal for the input structure detailed in Sec. 6.3.1. To simplify the notation, we write $\bar{C}^{W_+ * \epsilon} = \hat{v}_0 [W_+ * \epsilon](0) \Lambda$ with

$$\Lambda = \begin{pmatrix} \hat{c}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \hat{c}_m \end{pmatrix}. \quad (6.22)$$

Using the expression for the activation dynamics of the Poisson neuron Eq. (2.1), the learning equation (6.19) for additive STDP can be rewritten in the matrix equation,

$$\begin{aligned} \dot{\bar{K}} &= w^{\text{in}} \hat{\mathbf{v}}^{\text{T}} + w^{\text{out}} [v_0 + (M/m)\bar{K}\hat{\mathbf{v}}] \hat{\mathbf{e}}^{\text{T}} + \tilde{W} [v_0 + (M/m)\bar{K}\hat{\mathbf{v}}] \hat{\mathbf{v}}^{\text{T}} \\ &\quad + (M/m)\bar{K}\bar{C}^{W_+ * \epsilon} \\ &= \beta_1 \hat{\mathbf{e}}^{\text{T}} + \bar{K} (\beta_2 \hat{\mathbf{e}} \hat{\mathbf{e}}^{\text{T}} + \beta_3 \Lambda), \end{aligned} \quad (6.23)$$

6.3 Representation of the input correlation in the weight structure for a single neuron 13

where $\dot{\bar{K}}$ is a row vector, $\bar{v} = \hat{v}_0 \bar{\mathbf{e}}$ is the column vector of the mean firing rates over each pool, and the superscript \mathbf{T} is the matrix transposition. The coefficients β_1 , β_2 and β_3 absorb the input and learning parameters:

$$\begin{aligned}\beta_1 &= w^{\text{in}} \hat{v}_0 + w^{\text{out}} v_0 + \tilde{W} v_0 \hat{v}_0, \\ \beta_2 &= w^{\text{out}} \frac{M}{m} \hat{v}_0 + \tilde{W} \frac{M}{m} \hat{v}_0^2, \\ \beta_3 &= [W_+ * \epsilon](0) \frac{M}{m} \hat{v}_0.\end{aligned}\tag{6.24}$$

When starting from an initially homogeneous distribution, i.e., $\bar{K}(0) \propto \bar{\mathbf{e}}^{\mathbf{T}}$, it follows from (6.23) that \bar{K} can be decomposed into components in the basis $\bar{\mathbf{e}}^{\mathbf{T}} \Lambda^r$ with $0 \leq r \leq m-1$:

$$\bar{K} = \sum_{0 \leq r \leq m-1} \zeta_r \bar{\mathbf{e}}^{\mathbf{T}} \Lambda^r \tag{6.25}$$

and the evolution of the coefficients ζ_r is given by

$$\dot{\zeta}_0 = \beta_1 + \beta_2 \sum_r \zeta_r \left(\sum_l \hat{c}_l^r \right) - p_0 \beta_3 \zeta_{m-1}, \tag{6.26a}$$

$$\dot{\zeta}_r = \beta_3 (\zeta_{r-1} - p_r \zeta_{m-1}), \quad \text{for } 1 \leq r \leq m-1. \tag{6.26b}$$

We have used the equality

$$\bar{\mathbf{e}}^{\mathbf{T}} \Lambda^r \bar{\mathbf{e}} = \sum_l \hat{c}_l^r \tag{6.27}$$

and the following expression for the polynomial \mathbf{P} that nullifies the matrix Λ ,

$$\mathbf{P}(X) = \prod_l (X - \hat{c}_l) = X^m + \sum_{0 \leq r \leq m-1} p_r X^r, \tag{6.28}$$

which implies $\Lambda^m = -\sum_r p_r \Lambda^r$.

When assuming small correlation levels $\hat{c} < 1$, the stability conditions for the homeostatic equilibrium, namely $w^{\text{out}} < 0$ and $\tilde{W} < 0$, imply $\beta_2 < 0$ and thus lead to the stability of the component ζ_0 . In this case, the terms $\sum_l \hat{c}_l^r$ for $r \geq 1$ and $p_0 = (-1)^m \prod_l \hat{c}_l$ in Eq. (6.26a) are indeed then much smaller than the leading-order factor β_2 for ζ_0 in the

rhs and the stability is given by the homogeneous differential equation in ζ_0 without the perturbation:

$$\dot{\zeta}_0 = \beta_1 + \beta_2 \zeta_0. \quad (6.29)$$

This equation is equivalent to Eq. (6.3a) in the study of the homeostatic equilibrium. Consequently, the stable asymptotic value ζ_0^* for small correlations is close to the solution of the homogeneous equation Eq. (6.29), which we assume to be positive (realisable equilibrium), namely

$$\zeta_0^* \simeq -\frac{\beta_1}{\beta_2} \simeq K_{\text{av}}^* > 0. \quad (6.30)$$

Since $\beta_3 > 0$ Eq. (6.21), the stability of the other coefficients ζ_r for $r \geq 1$ is given by the following matrix according to Eq. (6.26b)

$$R = \begin{pmatrix} 0 & 0 & 0 & -p_1 \\ 1 & 0 & 0 & -p_2 \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & 1 & -p_{m-1} \end{pmatrix}. \quad (6.31)$$

The spectrum of the matrix R comprises the roots of the polynomial

$$Q(X) = X^{m-1} + \sum_{1 \leq r \leq m-1} p_r X^{r-1} = [P(X) - P(0)] / X. \quad (6.32)$$

Since the roots of P are non-negative reals, all roots of Q have positive real parts. This is illustrated in Fig. 6.5 for a specific example with $m = 10$ pools.

It follows that the behaviour of the ζ_r ($r \geq 1$) is unstable and these coefficients each diverge from their respective fixed point given by Eq. (6.26b), namely

$$\begin{aligned} \zeta_{m-1}^* &= \frac{\zeta_0^*}{p_1}, \\ \zeta_r^* &= p_{r+1} \zeta_{m-1}^* = \frac{p_{r+1} \zeta_0^*}{p_1}, \quad \text{for } 1 \leq r \leq m-2. \end{aligned} \quad (6.33)$$

The leading order for the structure of the input weights is given by ζ_1 since Λ has much

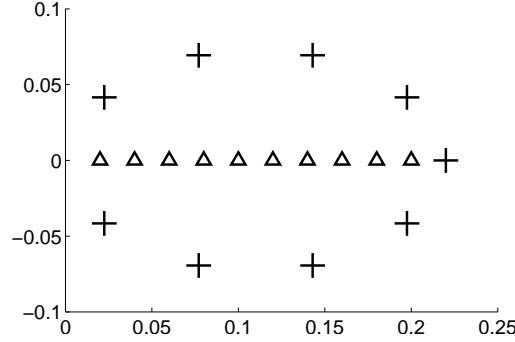


Figure 6.5: Distribution of the roots (+) of Q for ten input pools with correlations $\hat{c}_l = 0.02, 0.04, \dots, 0.2$ (\triangle). The horizontal axis stands for the real part, the vertical axis for the imaginary part. All the roots have positive real parts.

larger elements than Λ^r for $r \geq 2$ under the assumption of small correlations. We have

$$\zeta_1^* = \frac{p_2}{p_1} \zeta_0^* = \frac{\sum_{l' \neq l''} \frac{p_0}{\hat{c}_l \hat{c}_{l''}}}{\sum_l \frac{p_0}{-\hat{c}_l}} K_{\text{av}}^* = - \frac{\left(\sum_l \frac{1}{\hat{c}_l} \right)^2 - \sum_l \frac{1}{\hat{c}_l^2}}{\sum_l \frac{1}{\hat{c}_l}} K_{\text{av}}^*. \quad (6.34)$$

Neglecting the inhomogeneities in the correlation levels \hat{c}_l , we obtain the following approximation of ζ_1^* to give an idea of its order of magnitude:

$$\zeta_1^* \simeq - \frac{m-1}{\hat{c}_{\text{av}}} K_{\text{av}}^* \ll 0. \quad (6.35)$$

As a result, ζ_1 will always increase when starting from roughly homogeneous weights, which corresponds to $\zeta_1(0) \simeq 0$. The weights will then become structured according to $\hat{\mathbf{e}}^T \Lambda$, synonymous to stronger potentiation for larger \hat{C}_l . In addition, as ζ_0 remains roughly constant at ζ_0^* , the weights from pools with weaker \hat{c}_l will actually be depressed. This means that the degree of potentiation or depression of the input weights depends upon the correlation level of the corresponding pool.

6.3.4 Saturation of the weights for weight-dependent STDP

Now we consider the effect of scaling functions f_+ and f_- similar to those in Fig. 2.2: they gradually attenuates the potentiation for the individual weights above the homeostatic equilibrium value J_{av}^* and likewise with the depression for the weights below J_{av}^* .

The weight dependence used here acts as a spring: moving towards the bounds becomes harder when getting closer. This means that the weights with stronger potentiation due to the magnitude of the spike-time correlation coefficients F^{W+} , in relation to $\hat{C}^{W+*\epsilon}$ as described in Se. 6.3.3, will grow towards a higher stabilization value than those with weaker potentiation. The weights that are depressed experience a similar graduated stabilization towards quiescence.

When varying γ , strong weight dependence ($\gamma = 0.1$) leads to a weak specialization of the weights and their distribution may actually remain unimodal, as illustrated in Fig. 6.6(a) for a single neuron stimulated by $m = 5$ input pools. On the contrary, almost-additive STDP ($\gamma = 0.02$) induces stronger weight specialization: the asymptotic weight distribution in Fig. 6.6(b) is multimodal. In both cases, stronger correlation results in more potentiation, but the spreading of the asymptotic distribution is broader for almost-additive STDP, as illustrated in Fig. 6.6(d) to be compared with Fig. 6.6(c).

6.3.5 Generalisation to arbitrary input structure

These results can be extended to the case of an arbitrary input correlation structure. Since the set of diagonalisable matrices is dense, \hat{C} can be transformed to be roughly diagonal after a change of basis. For that new basis, the evolution of the input weights will proceed according to the principal components of the correlation structure in a similar way to the case of increasing input correlation levels \hat{c}_l studied above. Following (6.26b), the evolution of ζ_1 is given by

$$\dot{\zeta}_1 = \beta_3 (\zeta_0 - p_1 \zeta_{m-1}) . \quad (6.36)$$

Making the further assumption $p_1 \simeq mp_0/\hat{c}_{av} \simeq m\hat{c}_{av}^{m-1} \ll 1$, the evolution of ζ_1 will always be increasing, which means that the weights will become structured according to Λ . In other words, the evolution of the input weights will occur in a similar way to Sec.6.3.4: the weights will evolve in increasing order with respect to stronger covariance components of the input correlation structure.

To illustrate this, we examine another example of a neuron stimulated by three input pools: one pool with no spike-time correlations, two pools with within-pool correlations

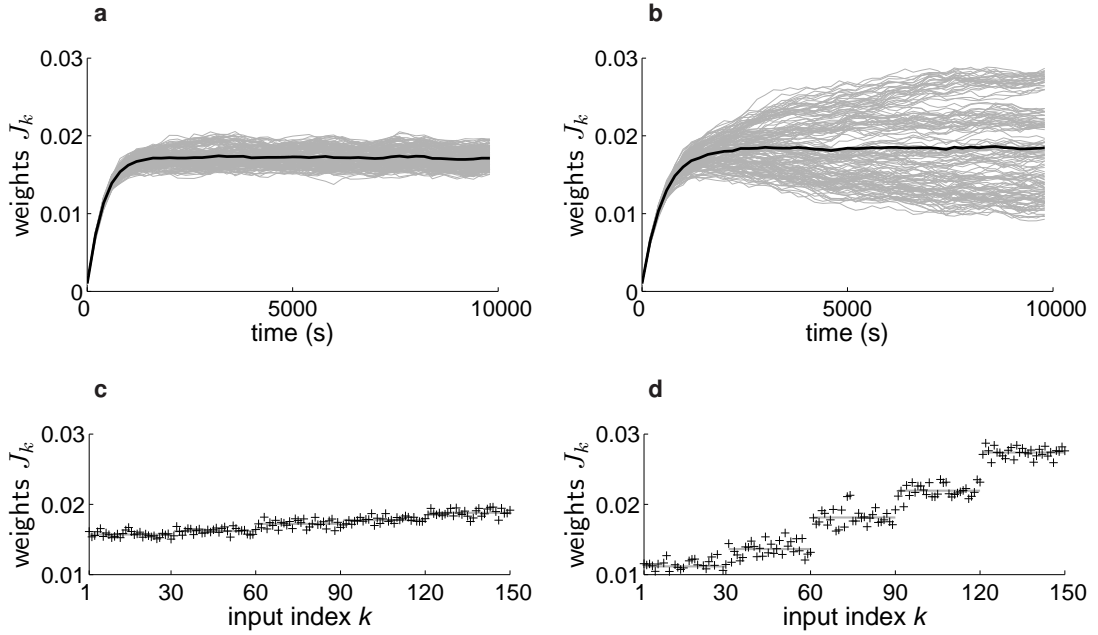


Figure 6.6: Evolution of the weight structure for the single neuron. Comparison between two degrees of weight dependence: (a & c) $\gamma = 0.1$ and (b & d) $\gamma = 0.02$. The neuron is stimulated by $m = 5$ pools of 30 inputs with the same firing rate $\hat{\nu}_0 = 10$ Hz and correlation levels $\hat{c}_l = 0, 0.05, 0.1, 0.15$ and 0.2 , respectively (cf. Sec. 6.3.1). (a & b) Evolution of the weights J_k (grey bundle) over 10^4 s. The thick black line represents the mean weight. (c & d) Asymptotic distribution of the weights J_k (+). The thick grey lines represent the means over each pool.

as described in Sec 6.3.1, but also with correlations between them such that the inputs from the second pool tend to fire 5 ms before those of the third pool. In this case, STDP select only the second pool, as illustrated in Fig. 6.7. The repression of the third pool can be explained by the matrix Λ , whose elements have signs according to the following diagram:

$$\Lambda \sim \begin{pmatrix} 0 & 0 & 0 \\ 0 & + & + \\ 0 & - & + \end{pmatrix}. \quad (6.37)$$

The form of Λ in this case follows from our choice of Hebbian STDP, which favors causality. The repeated firing of the third pool, which has within-pool spike-time correlation, after the second one causes STDP to discard the third pool, as it does for the first uncorrelated pool.

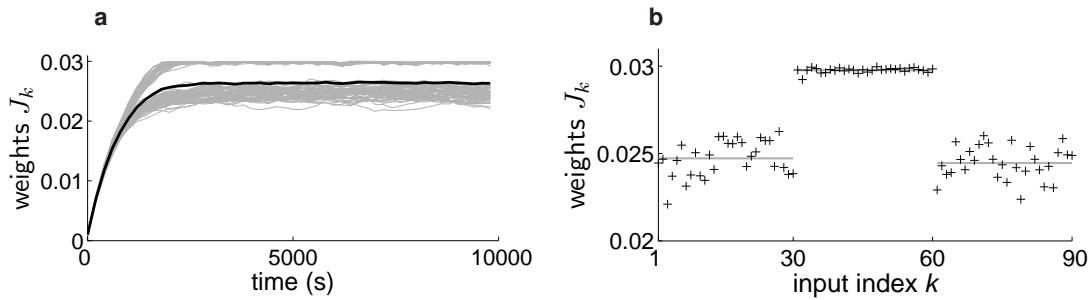


Figure 6.7: Evolution of the weight structure of a neuron stimulated by $m = 3$ input pools, as described in Sec. 6.3.5. (a) Evolution of the weights J_k (grey bundle) over 10^4 s. (b) Asymptotic distribution of the weights J_k (+). The plot formats and the parameters ($\gamma = 0.02$) are similar to Fig. 6.6.

In summary, STDP performs a kernel PCA, with the kernels determined by an interplay between the learning and neuronal parameters, namely W and ϵ . Another particularity is the normalization of the weights due to the homeostatic equilibrium, which may depend upon the input firing rate, as illustrated in Fig. 6.2. Previous work showed that a normalization constraint on the weights can scale between different flavors of PCA, such as k-means and graph-cut algorithms (Xu et al. 2009). STDP for a single neuron is thus capable to perform a generalized version of kernel PCA in an elaborate manner that depends upon the external inputs, as was suggested previously by van Rossum and Turrigiano (2001). In this way, STDP extends Oja's rule that extracts the strongest component of the rate-based correlation (Oja 1982). Partial connectivity and inhomogeneities are expected to bring more computational power since different areas in the network will then deal with various aspects of the input correlation structure. These results link the level of physiological modeling to machine learning and sheds light to the functional property of STDP through the induced learning dynamics.

Chapter 7

Stability of neuronal activity in recurrent networks

This chapter studies the ergodicity of the stochastic process representing the spiking activity of recurrently connected neurons. The neuron model used here extends the Poisson neuron introduced in Chapter 2. The focus is on the stationary properties of the spiking activity. Learning does not occur in the network, that is, all weights are kept fixed here.

7.1 Introduction

IN this chapter, an extension of the Poisson neuron model presented in Sec. 2.2.1 is considered, introducing a non-linear activation function to model the neuronal firing saturation. A framework to investigate the spiking dynamics in a network with (possibly) both excitatory and inhibitory synapses with fixed weights is developed, leaving aside the learning mechanisms. This chapter can be seen as a first step to the study of the neuronal correlation structure in a recurrently connected network of neurons that are more elaborate than the model used in the previous chapters of this thesis. Further results arising from this new framework will be important to understand learning in more elaborated neuronal networks, beyond the Poisson neuron model.

7.1.1 Non-linear Poisson neuron

Extending the original model corresponding to Eq. (2.1), the soma potential $\rho_i(t)$ is determined by an activation function σ (assumed to be continuous) that operates on the total

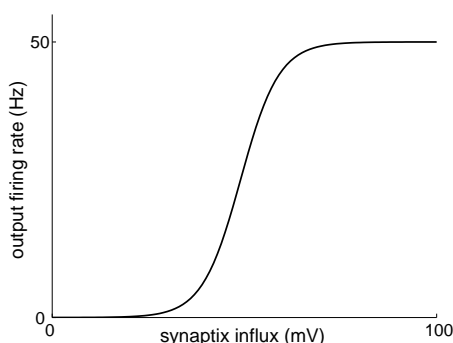


Figure 7.1: A typical choice of neuronal activation function. The boundedness of σ relates to the saturation of the neuron for strong stimulation observed in the biology. The plot represents a function of the form $\sigma(x) = a/(1 + e^{b-x})$ for two positive constants a and b .

synaptic influx (sum of PSPs):

$$\rho_i(t) = \sigma \left[\nu_0 + \sum_{k,n} K_{ik} \epsilon_{ik}(t - \hat{t}_{k,n}) \right]. \quad (7.1)$$

At rest (absence of pre-synaptic activity), the synaptic influx (within the square brackets) is equal to ν_0 , which models background activity that is not considered in detail; $\rho_i(t)$ is then equal to the spontaneous firing rate $\sigma(\nu_0)$. The total synaptic influx (within the square brackets) is thus the sum of ν_0 and the PSPs determined by the PSP kernel function $\epsilon_{ik}(t) \geq 0$ and the synaptic weight K_{ik} . Note that the use of σ allows the weights K_{ik} to be negative, which corresponds to inhibitory synapses.

The original Poisson model (Kempster et al. 1999) used in the previous chapters corresponds to the case where σ is the identity function. In this section, we assume σ to be positive and bounded,

$$0 < \sigma(x) \leq \Lambda < \infty \quad \text{for } x \in \mathbb{R}. \quad (7.2)$$

The positivity of σ ensures that neurons spontaneously fire spikes at rest. A typical choice of σ is illustrated in Fig. 7.1. All neurons have the same background activity ν_0 and the same activation function σ . This extended model accounts for the firing saturation of real neurons.

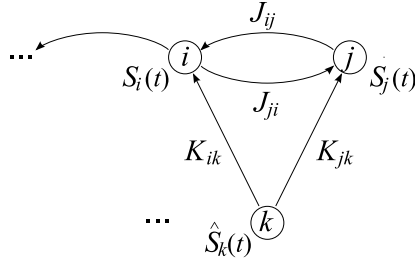


Figure 7.2: Schematic representation of two of the N network neurons (top circles) that are stimulated by one of the M external sources (bottom circle). The fixed weights of the input connections are denoted by K and the fixed recurrent weights by J . The network topology can be arbitrary.

7.1.2 Network model

Similar to Chapter 3, we consider a network of N Poisson neurons with firing intensities $\rho_i(t)$, $1 \leq i \leq N$, as defined in Eq. (7.1). The network neurons are excited by M external sources (or external inputs), with respective constant intensities $\hat{\rho}_k$,

$$\hat{\rho}_k \leq \Lambda < \infty, \quad k = 1, \dots, M. \quad (7.3)$$

The connection from external input k to neuron i is specified by a fixed weight K_{ik} and a PSP kernel function $\hat{\epsilon}_{ik}$, cf. Eq. (7.1). Likewise, the connection from neuron j to neuron i is specified by a fixed weight J_{ij} and a PSP kernel function ϵ_{ij} . We assume that all PSP kernels $\hat{\epsilon}_{ik}(t)$ and $\epsilon_{ij}(t)$ fade out when t is larger than a certain value (compact support), in addition to being non-negative and continuous. Note that $\hat{\epsilon}_{ik}(t)$ and $\epsilon_{ij}(t)$ absorb the synaptic delays in this chapter.

The framework developed in Chapter 3 corresponds to the situation where the PSP kernels are identical and equal to ϵ for all synapses (both input and recurrent connections), but incorporate individual synaptic delays \hat{d}_{ik} and d_{ij} , namely

$$\hat{\epsilon}_{ik}(t) = \epsilon(t - \hat{d}_{ik}) \quad \text{and} \quad \epsilon_{ij}(t) = \epsilon(t - d_{ij}). \quad (7.4)$$

We keep the general formulation in the present study.

Physiologically speaking, there is no “self-connection” from a neuron to itself ($J_{ii} = 0$

for all i). However, we keep the general formulation where such J_{ii} can be non zero for the sake of generality, since this framework could be applied to other problems where this condition may not be relevant.

7.1.3 Description of the network activity

We set the time origin at $t = 0$, the network being silent before that. Following Eq. (7.1), only the recent history of the network activity affects the potential of the network neurons and thus their probability of firing. We denote by Θ the “short-term memory depth” of the network:

$$\Theta := \max \left\{ \sup \{t : \max_{i,k} \hat{\epsilon}_{ik}(t) > 0\}, \sup \{t : \max_{i,j} \epsilon_{ij}(t) > 0\} \right\}. \quad (7.5)$$

This means that the “immediate” effect of a spike at time t_0 in the network will have faded out after the time Θ has passed (for $t \geq t_0 + \Theta$).

The variable $\hat{v}_k(t)$ counts the number of spikes fired by the k^{th} external source in the time interval $]t - \Theta, t]$. When $\hat{v}_k(t) \geq 1$, we denote by $\hat{\tau}_{k,1}(t)$ the time till the disappearance of the last spike from the memory window $]t - \Theta, t]$:

$$\hat{\tau}_{k,1}(t) := \hat{t}_{k,n} - t + \Theta \in]0, \Theta], \quad (7.6)$$

where $n = n(k, t)$ is the number of spikes fired by external source k in the time interval $[0, t]$. Likewise,

$$\hat{\tau}_{k,2}(t) := \hat{t}_{k,n-1} - t + \Theta > \hat{\tau}_{k,3}(t) > \dots > \hat{\tau}_{k,\hat{v}_k(t)}(t) > 0 \quad (7.7)$$

are the respective times elapsed since the second last, third last, etc., spikes coming from the external source k . The state of source k is determined by the collection of all these *time points* complemented by infinitely many zeros:

$$\hat{\zeta}_k(t) := (\hat{\tau}_{k,1}(t), \dots, \hat{\tau}_{k,\hat{v}_k(t)}(t), 0, 0, \dots) \in E_0, \quad (7.8)$$

where

$$E_0 := \left\{ (s_1, s_2, \dots) \in [0, \Theta]^{\mathbb{N}} : s_1 > \dots > s_n > 0, s_{n+1} = \dots = 0 \text{ for some } n \geq 0 \right\}. \quad (7.9)$$

Similarly, for the network neuron i , we define the spike-count variable $v_i(t)$ and the vector of the time points $\zeta_i(t) := (\tau_{i,1}, \dots, \tau_{i,v_i(t)}, 0, 0, \dots)$.

The state of the whole network,

$$X(t) := (\hat{\zeta}(t), \zeta(t)) = (\hat{\zeta}_1(t), \dots, \hat{\zeta}_M(t), \zeta_1(t), \dots, \zeta_N(t)), \quad (7.10)$$

is composed from the two vectors $\hat{\zeta}(t)$ and $\zeta(t)$, whose components are the infinite-dimensional vectors of time points defined similarly to Eq. (7.8). It is also convenient to use vector notation for the event-counting variables: $\hat{v}(t) = (\hat{v}_1(t), \dots, \hat{v}_M(t))$ for the external sources and $v(t) = (v_1(t), \dots, v_N(t))$ for the network neurons.

7.2 Network dynamics

7.2.1 Evolution of $X(\cdot)$

First note that all the trajectories of the process $X(\cdot)$ are right-continuous by construction and $X(\cdot)$ is a cadlag process (Doob 1953). The dynamics of the process $X(\cdot)$ can be described in the following way. All the time points $\hat{\tau}_{k,\cdot}$ and $\tau_{i,\cdot}$ decrease at a unit rate over time, until they “disappear” from our description when reaching 0. Indeed, they do not have any direct influence on the network state after that time. If t_0 is the time when $\hat{\tau}_{k,\hat{v}_k(t)}$ reaches 0, the spike-counting variable is decreased by one unit at time t_0

$$\hat{v}_k(t_0) = \hat{v}_k(t_0-) - 1. \quad (7.11)$$

So long as source k does not fire a spike, all components of the vector $\hat{\zeta}_k$ satisfy for $h > 0$

$$\hat{\tau}_{k,m}(t+h) = (\hat{\tau}_{k,m}(t) - h)_+, \quad m = 1, \dots, \hat{v}_k(t), \quad (7.12)$$

where $s_+ = \max\{s, 0\}$ denotes the positive part of $s \in \mathbb{R}$. The same applies to $\zeta_i(t) = (\tau_{i,1}, \dots, \tau_{i,v_i(t)}, 0, 0, \dots)$ for the network neurons.

Sources and neurons can fire at any time. The probability of firing one spike for the external source k during the time interval $]t, t + h]$ is

$$\hat{\rho}_k h (1 + o(1)) \quad \text{as } h \rightarrow 0 +. \quad (7.13)$$

The probability of firing more than one spike (even for different sources or neurons) during that time interval is $o(h)$. This implies that the probability for different neurons to have a spike at the same time is zero. When the external source k fires (say, at time t_1), its spike-counting variable is increased by one:

$$\hat{v}_k(t_1) = \hat{v}_k(t_1-) + 1, \quad (7.14)$$

and a new time point (time till disappearance of the immediate influence of the new spike) $\hat{\tau}_{k,1}(t_1) = \Theta$ is inserted in the first position of the “spike history” of the source k , while all the already listed ones are relabelled by shifting their subscripts by one. That is, the state variable immediately prior to t_1 ,

$$\hat{\zeta}_k(t_1-) = (\hat{\tau}_{k,1}(t_1-), \dots, \hat{\tau}_{k,2}(t_1-), \dots, \hat{\tau}_{k,m}(t_1-), 0, 0, \dots), \quad (7.15)$$

is updated to become

$$\hat{\zeta}_k(t_1) = (\hat{\tau}_{k,1}(t_1) = \Theta, \hat{\tau}_{k,2}(t_1) = \hat{\tau}_{k,1}(t_1-), \dots, \hat{\tau}_{k,m+1}(t_1) = \hat{\tau}_{k,m}(t_1-), 0, \dots), \quad (7.16)$$

where $m = \hat{v}_k(t_1-) = \hat{v}_k(t_1) - 1$.

The state variable $\zeta_i(t)$ related to the network neuron i evolves in a similar way. The only difference lies in the probability of firing in the time interval $]t, t + h]$, which is $\rho_i(t)h(1 + o(1))$ where $\rho_i(t)$ is determined by the network state $X(t)$ in a similar fash-

ion to Eq. (7.1):

$$\rho_i(t) = \sigma \left[v_0 + \sum_{j,n} J_{ij} \epsilon_{ij} (\Theta - \tau_{j,n}) + \sum_{k,m} K_{ik} \hat{\epsilon}_{ik} (\Theta - \hat{\tau}_{k,m}) \right]. \quad (7.17)$$

In summary, we see that the value of $X(\cdot)$ can change by a jump (firing of a spike), while between the jumps it varies in a continuous deterministic way.

7.2.2 Markov property

Due to the definition of Θ , the network state $X(t)$ entirely determines the firing rate for each network neuron at time t using Eq. (7.17), and the firing rate for the external input k is constant at $\hat{\rho}_k$. Thus, the probability of transition from the current state $X(t) = \underline{\mathbf{x}}$ (all state variables will be underlined) to another state $X(t+h) = \underline{\mathbf{x}'}$ after an arbitrary time increment $h > 0$ is completely specified by the information contained in $X(t)$. This means that, for a given event

$$A = \{X(t+h_1) \in B_1, \dots, X(t+h_m) \in B_m\}, \quad (7.18)$$

where $0 < h_1 < \dots < h_m$ and B_1, \dots, B_m are Borel sets, we have

$$\Pr[A | \mathcal{F}(t)] = \Pr[A | X(t)], \quad (7.19)$$

where $\mathcal{F}(t)$ denotes the natural filtration (past history up to time t) of the process $X(\cdot)$. In other words, $X(\cdot)$ is a continuous-time homogeneous piecewise-deterministic Markov process.

7.2.3 Formalism of piecewise deterministic Markov process

We now adapt notation from Davis (1984) and Jacobsen (2006) to the present study.

State space and filtration

All sources and neurons have the same “individual state space” E_0 defined in Eq. (7.9), in which each of the respective variables $\hat{\zeta}_k(t)$ or $\zeta_i(t)$ evolves. The set E_0 is contained in the infinite-dimensional cube $[0, \Theta]^{\mathbb{N}}$. The silent state of a source or neuron, when no spike has been fired in the recent history (determined by Θ), i.e. all time points are zeros, is denoted by $\mathbf{0}$; the silent state for the whole network is $\underline{\mathbf{0}} := (\mathbf{0}, \dots, \mathbf{0})$.

The process $X(\cdot)$ takes values in the measurable product space (E, \mathcal{E}) , where $E = E_0^{M+N}$ and \mathcal{E} is the product σ -algebra generated by cylindrical sets in E_0 with Borel bases. In the following sections, when considering a given state $X(t) = \underline{\mathbf{x}} \in E$, the variables $\hat{\nu}$, ν , $\hat{\zeta}$, ζ will refer to the corresponding values associated with $\hat{\nu}$, ν , $\hat{\zeta}$, ζ , respectively.

Deterministic vector field

Without a stochastic jumps (firing of a spike), the deterministic evolution of the state is the uniform decrease of the positive components (time points) of $\underline{\mathbf{x}}$ at a unit rate. When the smallest of them turns into zero, that component stops changing. Consider the state $\underline{\mathbf{x}} = (\hat{\mathbf{z}}, \mathbf{z}) \in E$. The deterministic trajectory of the state variable $X(t)$ starting from $\underline{\mathbf{x}}$ at time $t_0 > 0$ is denoted by $\Phi(t - t_0, \underline{\mathbf{x}})$:

$$\Phi(h, \underline{\mathbf{x}}) = \left((\hat{\mathbf{z}}_1 - h)_+, \dots, (\hat{\mathbf{z}}_M - h)_+, (\mathbf{z}_1 - h)_+, \dots, (\mathbf{z}_N - h)_+ \right), \quad (7.20)$$

where the positive part $(\cdot)_+$ applies to all components of each infinite-dimensional vector $\mathbf{s} = (s_1, s_2, \dots)$ and $\mathbf{s} - h = (s_1 - h, s_2 - h, \dots)$ for $h \in \mathbb{R}$. Note that the specific value of t_0 does not matter here since the process is time homogeneous. For the particular case of the silent state $\underline{\mathbf{0}}$, we have for $h \geq 0$

$$\Phi(h, \underline{\mathbf{0}}) = \underline{\mathbf{0}}. \quad (7.21)$$

In Fig. 7.3, the straight arrows (1), (2) and (5), as well as staying at the origin (3), illustrate the deterministic evolution of the state.

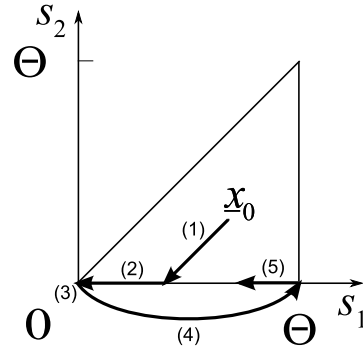


Figure 7.3: Example of evolution of the process for the time point of one single source. Starting from $\underline{x}_0 = (s_1, s_2, 0, \dots)$ corresponding to two time points, both s_1 and s_2 decrease at unit rate until s_2 reaches zero (1); then only s_1 varies until reaching zero (2). After some time in the silent state (3), the source fires (4) and a new time point s_1 set to Θ starts decreasing (5).

We define the time-invariant one-sided vector field \mathcal{X} at state \underline{x} such that (Davis 1984)

$$\mathcal{X}f(\underline{x}) = \frac{\partial f[\Phi(h, \underline{x})]}{\partial h} \quad (7.22)$$

using the right-sided derivative for any given smooth enough real-valued function $f : E \rightarrow \mathbb{R}$. Following Sec. 7.2.1, \mathcal{X} simply operates on a differentiable function f as

$$\mathcal{X}f(\underline{x}) = \frac{\partial f(\underline{x})}{\partial \mathbf{u}(\underline{x})} \quad (7.23)$$

for $\underline{x} \in E$, where the vector $\mathbf{u}(\underline{x}) = (\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_M, \mathbf{u}_1, \dots, \mathbf{u}_N)$ has components

$$\hat{u}_{k,m}(\underline{x}) := -\mathbb{1}(\hat{\mathbf{z}}_{k,m} > 0), \quad u_{i,n}(\underline{x}) := -\mathbb{1}(\mathbf{z}_{i,n} > 0). \quad (7.24)$$

We require f in Eq. (7.23) to be path-continuous and path-differentiable (Jacobsen 2006).

Jump times and survivor function

We denote by T_n ($n \geq 0$) the jump times of our process, namely each time a spike is fired in the network (including the external sources). Due to the boundedness of the stochastic intensities for the neurons and sources, only a finite number of jumps occur

during a finite time interval, with probability 1. As in Davis (1984), we define the function $\lambda : E \rightarrow \mathbb{R}^+$ that determines the probability of any event occurring in the whole network when the system is in a given state \underline{x} . It lumps the probabilities of all neurons or sources to fire a spike (cf. Sec. 7.2.1) by summing up the intensities $\hat{\rho}_k$ and ρ_i in Eq. (7.17) determined by the state \underline{x} :

$$\lambda(\underline{x}) = \sum_k \hat{\rho}_k + \sum_i \tilde{\rho}_i(\underline{x}), \quad (7.25)$$

where we have defined $\tilde{\rho}_i(\underline{x})$ as the potential of neuron i evaluated according to Eq. (7.17) when the network is in state \underline{x} :

$$\tilde{\rho}_i(\underline{x}) := \sigma \left[v_0 + \sum_{j,n} J_{ij} \epsilon_{ij}(\Theta - z_{j,n}) + \sum_{k,m} K_{ik} \epsilon_{ik}(\Theta - \hat{z}_{k,m}) \right]. \quad (7.26)$$

Note that the function λ is clearly bounded according to Eq. (7.2).

Consider the network in a state $X(t_0) = \underline{x}$ at time t_0 . According to the definition of Φ in Sec. 7.2.3, the “survival function” at time for $h \geq 0$ is given by

$$F(\underline{x}, t_0 + h) := \Pr \{ \text{no jump in } [t_0, t_0 + h] | X(t_0) = \underline{x} \} = \exp \left(- \int_0^h \lambda(\Phi(h', \underline{x})) dh' \right). \quad (7.27)$$

Note that the survival function only depends upon the network state \underline{x} at time t_0 and the time h elapsed since t_0 , but not upon the specific time t_0 .

Stochastic transitions

We now introduce some notation to handle more easily the modifications of the lists of time points at the jump epochs. We define the two following operations on any given state $\underline{x} = (\hat{\mathbf{z}}, \mathbf{z})$. First, for a given state $\underline{x} \in E$ and a source index k , $\underline{x}^{\circ,k}$ is the state of E corresponding to the source k firing: the value $\hat{\mathbf{z}}_k = (\hat{z}_{k,1}, \hat{z}_{k,2}, \dots, \hat{z}_{k,m}, 0, 0, \dots)$ is replaced by $(\Theta, \hat{z}_{k,1}, \dots, \hat{z}_{k,m-1}, \hat{z}_{k,m}, 0, \dots)$, cf. Eqs. (7.15) and (7.16). This is illustrated by the curved arrow (4) in Fig. 7.3. Likewise, $\underline{x}^{\bullet,i}$ modifies the vector \mathbf{z}_i for neuron i .

The function $Q(\underline{y}, \underline{x}) : E \times E \rightarrow [0, 1]$ determines the probability of transition from the

state \underline{x} to \underline{y} , given that a jump from \underline{x} occurs. For $\underline{x} \in E$, we have

$$Q(\underline{y}, \underline{x}) = \begin{cases} \hat{\rho}_k \lambda(\underline{x})^{-1} & \text{if } \underline{y} = \underline{x}^{\circ,k}, \\ \tilde{\rho}_i(\underline{x}) \lambda(\underline{x})^{-1} & \text{if } \underline{y} = \underline{x}^{\bullet,i}, \\ 0 & \text{otherwise.} \end{cases} \quad (7.28)$$

7.2.4 Description of the generator

From Davis (1984, Theorem 5.5) and Jacobsen (2006, Eq. (7.76)), the (extended) generator \mathcal{A} of our piecewise deterministic process $X(\cdot)$ is defined by

$$\mathcal{A}f(\underline{x}) = \mathcal{X}f(\underline{x}) + \lambda(\underline{x}) \int (f(\underline{y}) - f(\underline{x})) Q(d\underline{y}, \underline{x}), \quad (7.29)$$

for $f \in \mathfrak{F}$, where the domain \mathfrak{F} of the generator consists of all bounded functions $f : E \rightarrow \mathbb{R}$ that are path-continuous and path-differentiable for the deterministic vector field \mathcal{X} (Jacobsen 2006). Using Eqs. (7.23) and (7.28), the expression for the generator applied to $f \in \mathfrak{F}$ at $\underline{x} \in E$ has the form

$$\mathcal{A}f(\underline{x}) = \frac{\partial f(\underline{x})}{\partial \mathbf{u}(\underline{x})} + \sum_k \hat{\rho}_k [f(\underline{x}^{\circ,k}) - f(\underline{x})] + \sum_i \tilde{\rho}_i(\underline{x}) [f(\underline{x}^{\bullet,i}) - f(\underline{x})]. \quad (7.30)$$

7.3 Stability in the network

We want to know if the piecewise deterministic Markov process $X(\cdot)$ from Sec. 7.2.3 is ergodic. A positive answer to this question was given in Bremaud and Massoulié (1996, Massoulié (1998) for a larger class of kernels ϵ (no requirement for a compact support) and both linear and non-linear activation functions σ . In particular, the convergence of the process towards a stationary regime (stable spiking intensities) is ensured by either the boundedness of the activation function σ or a condition on the strength of the recurrent connections otherwise. However, our framework gives a simpler proof of the ergodicity for the bounded case and is more general than the process studied in Bremaud and Massoulié (1996, Massoulié (1998) in the sense that it can be extended to more complex

models than Hawkes processes (Poisson neurons). We show the positive recurrence of the silent state, which is visited by the system in finite mean time.

7.3.1 Ergodicity

Consider the silent state $\underline{\mathbf{0}} = (0, \dots, 0)$. Starting from an arbitrary state $X(t_0) = \underline{\mathbf{x}}_0$ at time $t = t_0$, there exists a lower bound for the probability to enter $\underline{\mathbf{0}}$ at time $t = t_0 + \Theta$. Because the $\hat{\rho}_k$ and the function σ are all bounded by Λ , we have according to (7.27)

$$F(\underline{\mathbf{x}}, t_0 + \Theta) \geq \left(\exp \left(- \int_{t_0}^{t_0 + \Theta} \Lambda \, dt \right) \right)^{M+N} = e^{-\Lambda(M+N)\Theta} > 0. \quad (7.31)$$

Now consider the network starting at the state $\underline{\mathbf{x}}_0$ at time $t = 0$. Denote by s^* the first time point on the time lattice with span Θ when the network is in the state $\underline{\mathbf{0}}$:

$$s^* = \inf \{ m\Theta : X(m\Theta) = \underline{\mathbf{0}}, \quad m \geq 1 \}, \quad (7.32)$$

and by t^* the first time the network is in the state $\underline{\mathbf{0}}$:

$$t^* = \inf \{ t > 0 : X(t) = \underline{\mathbf{0}} \}. \quad (7.33)$$

Since $t^* \leq s^*$, we have from the uniform lower bound of Eq. (7.31) that, for any state $\underline{\mathbf{x}}_0$ and a given integer $m \geq 1$

$$\begin{aligned} \Pr \{ t^* \geq m\Theta \mid X(0) = \underline{\mathbf{x}}_0 \} &\leq \Pr \{ s^* \geq m\Theta \mid X(0) = \underline{\mathbf{x}}_0 \} \\ &\leq \left(1 - e^{-\Lambda(M+N)\Theta} \right)^m. \end{aligned} \quad (7.34)$$

This clearly implies that the expected first hitting time of the state $\underline{\mathbf{0}}$ is finite:

$$\begin{aligned}
\mathbb{E}(t^* \mid X(0) = \underline{\mathbf{x}}_0) &= \int_0^\infty \Pr\{t^* > t\Theta \mid X(0) = \underline{\mathbf{x}}_0\} dt \\
&\leq \sum_{m \geq 0} \Theta \Pr\{t^* \geq m\Theta \mid X(0) = \underline{\mathbf{x}}_0\} \\
&\leq \sum_{m \geq 0} \Theta \left(1 - e^{-\Lambda(M+N)\Theta}\right)^m \\
&= \Theta e^{\Lambda(M+N)\Theta}.
\end{aligned} \tag{7.35}$$

Thus $\underline{\mathbf{0}}$ is a positive recurrent state for our network and clearly $\Pr\{t^* < \infty \mid X(0) = \underline{\mathbf{x}}_0\} = 1$ for any initial state $\underline{\mathbf{x}}_0$. As it is obvious that the Markov process $X(\cdot)$ is aperiodic and stochastically continuous, it follows that it is strongly ergodic (Borovkov 1998, Th.1 in §18). This implies that there exists a stationary distribution Π on E for the process $X(\cdot)$.

7.3.2 Stationary distribution

In the special case of a linear activation function σ , the equilibrium firing rate is the same for any PSP kernels, $\hat{\epsilon}_{ik}$ and ϵ_{ij} (Hawkes 1971). However, the stationary distribution Π in the state space E may depend on these parameters. For a function f in the domain \mathfrak{F} of the generator, we have (Jacobsen 2006, pp. 184)

$$\int_E \mathcal{A}f(\underline{\mathbf{x}}) \Pi(d\underline{\mathbf{x}}) = 0. \tag{7.36}$$

This functional equation allows us in principle to determine Π in the general case, but we will now focus on a particular illustrative case.

Single neuron with one feedback self-connection

Consider a single neuron with a feedback self-connection with weight J and for the sake of simplicity without input connection. The intensity function of the neuron is simply denoted by $\tilde{\rho}(\underline{\mathbf{x}})$. In other words, the neuron is driven by the spontaneous activity related to ν_0 and its own past activity. The state $\underline{\mathbf{x}}$ of the single neuron evolves in E_0 , which will

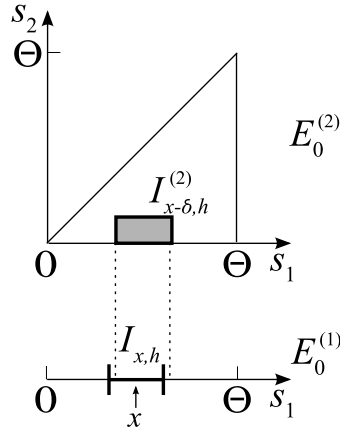


Figure 7.4: Illustration of the intervals considered when evaluating the transitions between simplices.

be considered for convenience as a union of finite-dimensional simplices

$$E_0^{(n)} := \{(s_1, \dots, s_n) \in [0, \Theta]^n : s_1 > \dots > s_n > 0\} \quad (7.37)$$

so that $E_0 = \bigcup_{n \geq 0} E_0^{(n)}$. The case $n = 0$ corresponds to the silent state and we use the convention $E_0^{(0)} = \mathbf{0}$. We will assume that the stationary distribution has a density ψ_n on each simplex $E_0^{(n)}$, $n \geq 1$, with respect to the corresponding volume (Lebesgue) measure, and an atom at the silent state $\mathbf{0}$. We further assume that each ψ_n has well-defined finite limits on the boundaries of the corresponding simplex. This implies in particular that the densities are bounded.

Fix an arbitrary $x \in]0, \Theta[\subseteq E_0^{(1)}$ and consider two scalars $h \rightarrow 0+$ and $\delta \rightarrow 0+$ such that $I_{x, h} :=]x, x + h[\subset E_0^{(1)}$ and assuming w.l.o.g. that also $I_{x+\delta, h} \subset E_0^{(1)}$, as illustrated in Fig. 7.4. We have

$$\begin{aligned} \int_{I_{x, h}} \psi_1(x') dx' &= \Pr\{X(t) \in I_{x, h}\} \\ &= \Pr\{X(t + \delta) \in I_{x, h}\} \\ &= \int_{E_0} \Pr\{X(t + \delta) \in I_{x, h} \mid X(t) = \mathbf{x}'\} \Pr\{X(t) \in d\mathbf{x}'\}. \end{aligned} \quad (7.38)$$

For the integrand in the last integral to be non-zero, we need to have either

$$X(t) \in I_{x+\delta,h} \quad (7.39a)$$

$$\text{or } X(t) \in (I_{x,h} \times]0, \delta[) \subset E_0^{(2)} \quad (7.39b)$$

$$\text{or } \dots \quad (7.39c)$$

and no jump happened during the time interval $]t, t + \delta[$; note that we have considered δ sufficiently small such that $]x, x + h[\cap]\ominus - \delta, \ominus[= \emptyset$. In the first case of Eq. (7.39a), the corresponding contribution to the integral is given by

$$\begin{aligned} & \int_{I_{x+\delta,h}} \Pr \{ \text{no jump during }]t, t + \delta[\mid X(t) = \underline{x}' \} \Pr \{ X(t) \in d\underline{x}' \} \quad (7.40) \\ &= \int_{I_{x+\delta,h}} \exp \left[- \int_0^\delta \tilde{\rho}(\underline{x}' - t') dt' \right] \psi_1(x') dx', \end{aligned}$$

where $\underline{x}' = (x', 0, 0, \dots)$. Using continuity of $\tilde{\rho}$, we see that, for $x' \in I_{x+\delta,h}$, $t' \in]0, \delta[$, one has $\tilde{\rho}(\underline{x}' - t') = \tilde{\rho}(\underline{x}) + o(1)$. It follows that the integral in the rhs of Eq. (7.39a) is equal to

$$[1 - (\tilde{\rho}(\underline{x}) + o(1))\delta] \int_{I_{x+\delta,h}} \psi_1(x') dx' = \int_{I_{x+\delta,h}} \psi_1(x') dx' - \tilde{\rho}(\underline{x}) \psi_1(x) h\delta + o(h\delta), \quad (7.41)$$

after constraining the evaluation up to the order $h\delta$. Furthermore, using our assumption of continuity of ψ_2 in the closure of $E_0^{(2)}$, we see that the corresponding contribution to the rhs of Eq. (7.38) when Eq. (7.39b) holds is

$$\begin{aligned} & \int_{I_{x+\delta,h} \times]0, \delta[} \Pr \{ \text{no jump during }]t, t + \delta[\mid X(t) = \underline{x}' \} \Pr \{ X(t) \in d\underline{x}' \} \quad (7.42) \\ &= \int_{I_{x+\delta,h} \times]0, \delta[} \exp \left[- \int_0^\delta \tilde{\rho}(\underline{x}' - t') dt' \right] \psi_2(z'_1, z'_2) dz'_1 dz'_2 \\ &= (1 + o(1)) \psi_2(x, 0+) h\delta \end{aligned}$$

with $\underline{x}' = (z'_1, z'_2, 0, \dots)$. Other contributions involving $E_0^{(3)}, \dots$ are $o(h\delta)$. Putting it all together, we obtain the following relation by rewriting Eq. (7.38):

$$\int_{I_{x,h}} \psi_1(x') dx' = \int_{I_{x+\delta,h}} \psi_1(x') dx' - \tilde{\rho}(\underline{x}) \psi_1(x) h\delta + \psi_2(x, 0+) h\delta + o(h\delta). \quad (7.43)$$

Subtracting the first term on the rhs from the lhs, we have

$$\int_x^{x+h} \psi_1(x') dx' - \int_{x+\delta}^{x+\delta+h} \psi_1(x') dx' = \int_x^{x+h} [\psi_1(x') - \psi_1(x' + \delta)] dx', \quad (7.44)$$

which gives after reorganising Eq. (7.43) and dividing by $h\delta$

$$\frac{1}{h} \int_x^{x+h} \frac{\psi_1(x') - \psi_1(x' + \delta)}{\delta} dx' = -\tilde{\rho}(\underline{\mathbf{x}}) \psi_1(x) + \psi_2(x, 0+) + o(1). \quad (7.45)$$

Finally, assuming that ψ_1 is continuously differentiable on $E_0^{(1)}$, we obtain the relation

$$\psi_1'(x) = \tilde{\rho}(\underline{\mathbf{x}}) \psi_1(x) - \psi_2(x, 0+), \quad \underline{\mathbf{x}} = (x, 0, 0, \dots). \quad (7.46)$$

Similar calculations for $\underline{\mathbf{x}} = (z_1, \dots, z_n, 0, \dots)$ and ψ_n , $n \geq 1$, lead to

$$\begin{aligned} \frac{\partial \psi_n(z_1, \dots, z_n)}{\partial \mathbf{u}_n} &= \lim_{h \rightarrow 0+} \frac{\psi_n(z_1 + h, \dots, z_n + h)}{h} \\ &= \tilde{\rho}(\underline{\mathbf{x}}) \psi_n(z_1, \dots, z_n) - \psi_{n+1}(z_1, \dots, z_n, 0+), \end{aligned} \quad (7.47)$$

where \mathbf{u}_n is the n -dimensional vector with all elements equal to one: $\mathbf{u}_n = (1, \dots, 1)$.

Now considering the atom $\mathbf{0}$ for the stationary distribution Π , we proceed similarly:

$$\begin{aligned} \Pi(\{\mathbf{0}\}) &= \Pr\{X(t) = \mathbf{0}\} \\ &= \Pr\{X(t + \delta) = \mathbf{0}\} \\ &= \int_{E_0} \Pr\{X(t + \delta) = \mathbf{0} \mid X(t) = \underline{\mathbf{x}}'\} \Pr\{X(t) \in d\underline{\mathbf{x}}'\}. \end{aligned} \quad (7.48)$$

Non-zero contributions to the integrand of the rhs correspond to

$$\begin{aligned} X(t) &= \mathbf{0} \\ \text{or } X(t) &\in]0, \delta[\subset E_0^{(1)} \\ \text{or } X(t) &\in (]0, \delta[^2) \cap E_0^{(2)} \\ \text{or } &\dots \end{aligned} \quad (7.49)$$

and no jump happened during the time interval $]t, t + \delta[$. The contribution in the first case is given by

$$\exp \left[- \int_0^\delta \tilde{\rho}(\mathbf{0}) dt' \right] \Pi(\{\mathbf{0}\}) = [1 - (\tilde{\rho}(\mathbf{0}) + o(1)) \delta] \Pi(\{\mathbf{0}\}) . \quad (7.50)$$

The contribution corresponding to $X(t) \in E_0^{(n)}$ in Eq. (7.49) is

$$\begin{aligned} & \int_{(]0, \delta^{[n]} \cap E_0^{(n)})} \Pr \{ \text{no jump during }]t, t + \delta[\mid X(t) = \underline{\mathbf{x}}' \} \Pr \{ X(t) \in d\underline{\mathbf{x}}' \} \quad (7.51) \\ &= \int_{(]0, \delta^{[n]} \cap E_0^{(n)})} \exp \left[- \int_0^\delta \tilde{\rho}((\underline{\mathbf{x}}' - t')_+) dt' \right] \psi_n(z'_1, \dots, z'_n) dz'_1 \dots dz'_n \\ &= [1 - (\tilde{\rho}(\mathbf{0}) + o(1)) \delta] (1 + o(1)) \frac{\delta^n}{n!} \psi_n(0, \dots, 0) , \end{aligned}$$

using the assumption of continuity for $\tilde{\rho}$ and ψ_n . Taking the leading order in δ for the expression in Eq. (7.48), we obtain

$$\Pi(\{\mathbf{0}\}) = [1 - (1 + o(1)) \delta \tilde{\rho}(\mathbf{0})] \Pi(\{\mathbf{0}\}) + \delta \psi_1(0+) + o(\delta) . \quad (7.52)$$

Reorganising, dividing by δ and taking the limit when $\delta \rightarrow 0+$ finally leads to

$$\tilde{\rho}(\mathbf{0}) \Pi(\{\mathbf{0}\}) = \psi_1(0+) . \quad (7.53)$$

We finally give an example about how to use Eqs. (7.46), (7.47) and (7.53) to evaluate the stationary densities ψ_n of the process on each simplex $E_0^{(n)}$, $n \geq 1$. We can construct a truncated process $X_n(\cdot)$ that behaves similarly to $X(\cdot)$ when the number of spikes $\nu(t) < n$ for a given $n > 1$ and that does not fire a spike when $\nu(t) = n$. In this way, . Using the fact that $X(\cdot)$ has a fast-decaying probability of reaching $E_0^{(n)}$ for large n , it can be shown that the stationary densities corresponding to the truncated process $X_n(\cdot)$ converges more than exponentially fast when $n \rightarrow \infty$ towards those for $X(\cdot)$. This allows us to use the “boundary” condition $\psi_{n+1} = 0$ to evaluate $\psi_{n'}$ for $n' \leq n$.

7.4 Remarks on the framework presented in this chapter

These preliminary results showed that this framework is fitted to studying the stationary properties of the network. Using the generator in Eq. (7.30), we hope to obtain more insight into the evolution of the stochastic process. This framework is a tentative to examine the pairwise correlation structure and its implications for elaborate neuron models than the Poisson neuron (Sec. 2.2.1). The present framework could be generalised to any neuron model for which the effect of a pre-synaptic spike vanishes after a given period (short-term memory): the key consists in expressing the probability of firing for each neuron depends upon the past spiking history in the network. For integrate-and-fire neurons, other directions of work to achieve similar goals are currently explored (Moreno-Bote et al. 2008). The derivation of consistency equations for the firing rates and spike-time correlations similar to Eqs. (3.22a-3.22c) for more elaborated neuron models is crucial to analyse the effect of STDP.

Chapter 8

Conclusion

8.1 Summary of original contributions and results

IN THIS thesis a theoretical framework is presented to investigate the effect of STDP in recurrently connected neuronal networks. The analysis has been carried out for particular network configurations in order to understand how the weight dynamics results from an interplay between the neuronal properties, the network connectivity, the input structure and the learning parameters. This led to determining conditions on the parameters for which STDP generates neuronal specialisation in the network in a fashion that corresponds to self-organisation.

8.1.1 Theoretical framework to study learning dynamics

The mathematical framework for analysing the weight dynamics induced by STDP presented in Chapter 3 relies on the Poisson neuron model (Kempster et al. 1999) and can account for any arbitrary (excitatory) connectivity topology and input structure. The STDP rule describes the change in synaptic weight resulting from each spike and pair of pre- and post-synaptic spikes. By averaging over the spike statistics, we obtain differential equations to describe the evolution of the weights (first stochastic moment). In addition to rate-based learning, STDP involve an additional term related to the spike-time covariance at a short time scale between the pre- and post-synaptic spike trains. In this sense, STDP extends the rate-based description of synaptic plasticity.

The analysis presented in this thesis incorporates the effect of the post-synaptic response, in this way extending previous work (Burkitt et al. 2007); however, dendritic

delays (Senn 2002) are ignored. For richer external input spike trains than the delta-correlated pools considered in Sec. 3.5, the post-synaptic response may play an important role (Sprekeler et al. 2007). The derivation of the covariance self-consistency equations Eqs. (3.18) and (3.20) is a cornerstone of this analysis, which was made tractable by using the Poisson neuron model (Kempster et al. 1999). These equations are crucial for the evaluation of spike-driven effects of STDP in recurrent networks, which cannot be captured by rate-based learning. The evolution of the input weights for slow learning is described by a dynamical system, which is analyzed in terms of fixed point and stability in order to predict the asymptotic behaviour of the weights. This framework targets network dynamics beyond the mean-field approach in order to study the emergence of a network structure due to external stimulation. Most of the analysis (Chapters 3-5) focuses on additive STDP in order to keep the analysis of the weight specialisation as tractable as possible, but weight-dependent STDP is addressed in Chapter 6.

8.1.2 Weight specialisation in recurrent networks

Both stability and competition for the weights were obtained for a broad range of learning parameters, in all cases studied in Chapters 4, 5 and 6. This interesting combination of behaviours arises from a homeostatic equilibrium, in which the mean incoming weight is constrained to a stable value for each neuron, and a splitting of the weight distribution occurs on a smaller time scale depending on the input correlations (for small correlation levels).

The conditions on the learning parameters that ensure the homeostatic equilibrium for the weights, irrespective of the input stimulation level, correspond to STDP inducing more depression than potentiation for uncorrelated inputs, i.e. the condition $\tilde{W} < 0$. This is in agreement with earlier numerical studies using integrate-and-fire neurons where the rate-based learning terms w^{in} and w^{out} are absent (Song et al. 2000, Song and Abbott 2001, Morrison et al. 2007). For additive STDP, w^{in} and w^{out} are necessary to obtain stability; when using weight-dependent STDP that induces alone stability, these rate-based terms modify but do not suppress the equilibrium of the mean weight. The weight equilibrium enforces the stabilisation of the firing rate for each neuron. The con-

clusions for the fixed point of the firing rates are valid for any neuron model, provided the correlation structure between the neurons is sufficiently weak. The firing rates are then all constrained to the same equilibrium value due to the learning equation Eq. (3.22b). The stability conditions derived using the Poisson neuron model are expected to hold equally well for other neuron models with excitatory synaptic weights, although the actual equilibrium values would then depend on the neuronal activation parameters. Inhomogeneities in the neuronal properties and/or learning parameters would induce inhomogeneities in the equilibrium values, but not impair the equilibria.

When the inputs have spike-time correlations, the initial weight distribution is, in general, modified to comply with the predicted specialisation scheme, which depends upon the input correlation structure embodied in \hat{C} , cf. Eq. (3.5). An exception to this expected behaviour only occurs for initial conditions in which the weights are already dramatically specialised in the “wrong” way or there are large differences between input firing rates, i.e., that would contradict and over-ride the specialisation trend induced by the spike-time correlations. This was shown for input selectivity, for which STDP potentiates, in general, synaptic connections coming from more correlated inputs (Sec. 4.3.4).

When the weight drift is small, such as during symmetry breaking for the input weights, the recurrent connections may play a determining role, even when they are non-plastic. For example, excitatory recurrent connections may cause the neurons to specialise in the same way, as illustrated in Sec. 4.4. This group effect takes place at the beginning of learning; when the neurons become sufficiently specialised, the drift takes over and reinforces the initial symmetry breaking because of the corresponding instability of the weight dynamics.

In order to obtain a non-trivial specialisation for the recurrent weights, a network topology is necessary where different neuron groups receive distinct inputs with correlation. Otherwise, the weight dynamics is equivalent to that in a network with no external inputs. When conditions are met, such as those described in Sec. 5.5.1 and 5.5.2, the individual weights exhibit strong competition that can result in the emergence of a feed-forward synaptic pathway or the strengthening of within-group connections for plastic

recurrent connections. The weight specialisation is determined by the interplay between the correlation structure of the external inputs, the STDP window function W and the PSP response kernel ϵ , in contrast to the case of input selectivity, where the details of W , ϵ and the delays are not important. The different schemes of potentiation vs. depression that were observed depending upon the sign and magnitude of $\hat{C}^{W*\zeta}$ may explain the contradictory behaviours observed in numerical simulations (Izhikevich et al. 2004, Iglesias et al. 2005), which generated debate about whether STDP induces more or less synchronisation in recurrent networks.

In the generalised case where the network receives external inputs from more than two pools (with small within-pool spike-time correlations), the following behaviours are expected:

- For sufficiently large input spike-time correlations, the splitting of the weight distribution depends on the input correlation structure, irrespective of (not too large) inhomogeneities in the input firing rates.
- Weights coming from input pools with stronger spike-time correlations are potentiated.
- The specialisation of input weights corresponds to splitting between the input pools.
- Neuron groups that receive strong (positive) spike-time correlation will experience a potentiation of their outgoing weights, provided the conditions on W , ϵ and the recurrent delays in Sec. 5.5 are met.

A learning rate $\eta = 5 \times 10^{-7}$ corresponds to a convergence towards the homeostatic equilibrium in hundreds of seconds, similar to Burkitt et al. (2007), and a development of a weight structure in tens of thousands of seconds (i.e., hours). Similar results were obtained with faster learning rates ($\eta = 10^{-5}$), provided that the consequent noise does not destroy the homeostatic equilibrium. More noise also reduces the dependence upon the

initial conditions. Our results show that, even for very small learning rates, the combination of equilibrium and diverging behaviour leads to the emergence of a weight structure.

8.1.3 Weight-dependence for STDP

Weight-dependent versions of STDP modify the weight dynamics compared to additive STDP, such as the homeostatic equilibrium and the weight distribution for uncorrelated inputs (van Rossum et al. 2000, Gütig et al. 2003, Morrison et al. 2007, Morrison et al. 2008). However, when starting from an initial homogeneous distribution of recurrent weights and for correlated inputs, the weights split in a similar manner to the additive model so long as the weight competition is strong enough. This was observed for STDP with a weak non-linearity related to the weight dependence, i.e., almost additive STDP.

The rate-based learning constants w^{in} and w^{out} were necessary in the analysis using additive STDP (Chapters 4 and 5) to obtain homeostatic equilibrium for the weights, which allowed a weight structure to emerge that resulted from the input spike-time correlation. They do not impair the local character of the learning rule. We expect the stability conclusions to hold in most cases for similar stabilising mechanisms, such as weight scaling (van Rossum et al. 2000), provided the combination with STDP leads to effective homeostatic equilibrium for the weights.

8.1.4 Self-organisation in visual cortex

The results presented in this thesis can be linked, for example, to the emergence of ocular-dominance areas in the primary visual cortex, when neuronal circuits specialise to one ocular pathway (left or right eye) in the first weeks of life of new-born mammals (Swindale 1996). It was shown that the assumption of more correlation for spike trains within each ocular pathway than between the two pathways is sufficient for STDP to cause the emergence of specialised recurrently connected areas sensitive to the inputs from only one eye, as illustrated in Fig. 8.1. STDP thus provides a framework to explain the emergence of ocular dominance (Fig. 8.2). Higher-order effects due to the recurrent connections may combine with non-linearities in other STDP models (Gütig et al. 2003, Burkitt

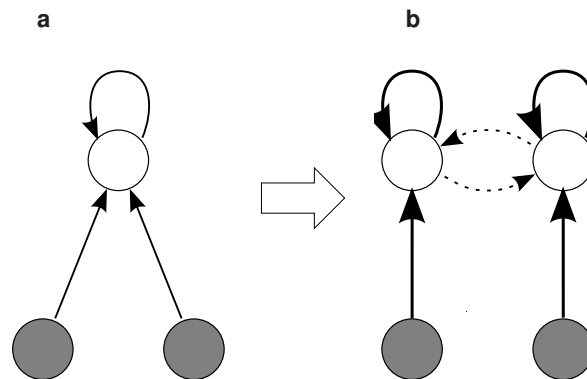


Figure 8.1: Self-organisation scheme. The initially homogeneous weight distribution is modified by STDP to become asymptotically bimodal depending on the input stimulation. Neuron groups emerge in response to the input correlation structure and specialise to only one of the correlated input pathways.

et al. 2004, Appleby and Elliott 2006) or specific input structures (e.g., Leibold et al 2002) to introduce further complexity in the weight dynamics.

Our results are intended to shed analytical light on previous work that used numerical simulations to show the emergence of a cortical-like organisation due to STDP (Choe and Miikkulainen 1998, Wensich et al. 2005). The present study has made minimal assumptions about the network topology and the input firing rate and correlation structures in order to explore the input specialisation behaviour in a recurrent network. Further study of the weight dynamics is required in a more detailed network topology that takes into account spatial structure in agreement to that in the cortex, such as short-range excitatory and medium-range inhibitory connections in the visual cortex (von der Malsburg 1973). The results presented here have some bearing on previous work on ocular dominance (von der Malsburg 1973, Swindale 1996, Elliott and Shadbolt 1999, Goodhill 2007); most of the models proposed or cited by von der Malsburg (1973) and Swindale (1996) interestingly combine the same dynamical ingredients as those shown here to be generated by STDP, namely a combination of stabilisation and divergence.

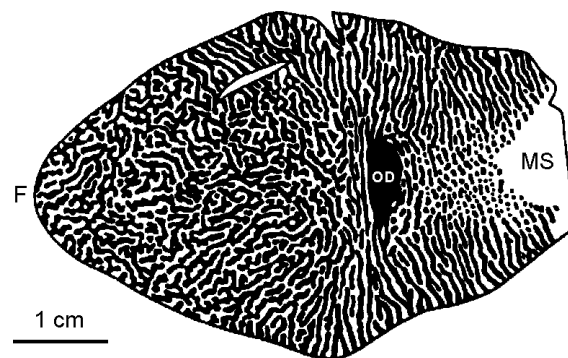


Figure 8.2: Ocular dominance columns in macaque monkey. The picture shows the pattern over nearly the complete visual hemifield in a macaque monkey. The outer boundaries of the pattern correspond to the vertical midline of the visual field; F indicates the fovea; OD the optic disc, and MS the monocular segment. The pattern is a drawing made from a montage of sections stained for cytochrome oxidase in a monkey which had lost one eye over a year prior to sacrifice. Taken from Florence and Kaas (1992).

8.2 Implications for neuronal information processing

Our results show the importance of spike-time correlations in generating a structure amongst synaptic weights, in agreement with previous studies (Kempster et al. 1999, Gütig et al. 2003, Song et al. 2000, Song and Abbott 2001). The time scale associated with these correlations are of the order of milliseconds. Experimental studies involving time bins of several tens of milliseconds (Tang et al. 2008, Carrillo-Reid et al. 2009) may thus only capture a portion of the relevant spike-timing information. The role played by these correlations in the encoding of neuronal information is still under debate and active investigation. It was shown that the spiking dynamics of integrate-and-fire neurons, either isolated or within networks, are sensitive to the correlation structure of their inputs (Salinas and Sejnowski 2002, Burkitt 2006, Moreno-Bote et al. 2008, Kriener et al. 2008). A better understanding of the interplay between the learning and spiking dynamics is a promising way of providing insight into the encoding of neuronal information.

The learning dynamics are determined by an interplay between STDP, the spike-time correlation structure and the network topology. The correlations themselves depend upon the input correlation structure and the neuronal mechanisms (especially the PSP response in the framework presented here). This elaborate self-organisation scheme is capable of rich behaviour. We have shown in Sec. 6.3 how STDP can encode into the

weight structure the spike-time correlation structure of inputs stimulating a single neuron. Namely, for homogeneous pools with distinct levels of within-pool correlation, weights can be sorted in increasing order with respect to the correlation of their corresponding pool. This can be extended to an arbitrary correlation structure and the algorithm performed by STDP then has the flavour of principal component analysis (PCA), as was suggested previously by van Rossum and Turrigiano (2001). In this sense, STDP extends Oja's rule that relies on rate-based learning (Oja 1982). These preliminary results nicely link the modelling at the physiological level to machine learning and sheds light on the functional properties of STDP in neuronal networks. Further studies incorporating networks with recurrent connections should provide interesting developments to investigate how neurons process spiking information in a self-organizing distributed fashion (Kohonen 1982). The ability of STDP to preprocess temporal inputs has already received support (Carnell 2009), when applied to the lateral connections of a network that acts as a reservoir of functions to extract information on the inputs, a.k.a. the liquid state machine (Maass 1997, Maass et al. 2002). To close the loop, it is also necessary to investigate the influence of the weight structure resulting from learning upon the spiking dynamics in the network (Amit and Brunel 1997). In particular, the spike-time correlation described in this study corresponds to fast variations of the probability of firing of the neurons, such as that required, for example, for image recognition (Thorpe et al. 2002).

8.3 Future research directions

Non-linear neuronal activation mechanisms may play a significant role in determining the neuron covariance structure. The framework developed in Chapter 7 aims to investigate the effect of non-linearities in the neuronal activation mechanisms upon the network spiking dynamics. Another challenge is to apply this framework to more complex neuron models, such as a Poisson neuron with non-linear activation function and the integrate-and-fire neuron (Burkitt 2006, Moreno-Bote et al. 2008). Preliminary simulation results (not presented in this thesis) satisfactorily showed that networks of integrate-and-fire neurons behave for some parameters according to the prediction made using the Poisson

neuron model.

The present study was constrained to using only narrow distributions of axonal delays and did not investigate the evolution of synchronisation between neurons. Previous work showed that STDP can induce non-trivial synchrony structure between neurons (Câteau et al. 2008) and that dendritic delays play an important role (Senn 2002, Lubenov and Siapas 2008). The present framework can be adapted to incorporate these aspects with a view to studying neuronal synchrony in networks (Izhikevich et al. 2004, Iglesias et al. 2005, Câteau et al. 2008). Note that the shift of STDP in Sec. 5.5.2 is equivalent to the introduction of dendritic delays, while keeping the sum of the axonal and dendritic delay constant.

The introduction of a population of inhibitory neurons as well as a more detailed network connectivity (e.g., short-range excitatory and medium-range inhibitory connections) would provide a further step towards a more realistic model of the visual cortex (Swindale 1996). Richer input correlation structures, such as oscillatory firing rates or spike patterns, will be of interest in applying our framework to self-organisation in the auditory pathways.

Appendix A

Calculations for chapter 3

A.1 Remarks on the input covariance structure

A.1.1 Definition of the external input covariance

In Eq. (3.3) the following definition for the covariance between two external inputs k and l is used

$$\text{Cov}[\hat{S}_k(t), \hat{S}_l(t+u)] := \langle \hat{S}_k(t) \hat{S}_l(t+u) \rangle - \langle \hat{S}_k(t) \rangle \langle \hat{S}_l(t+u) \rangle. \quad (\text{A.1})$$

The inputs \hat{S}_k are second-order stationary processes, which means that these functions $\text{Cov}[\hat{S}_k(t), \hat{S}_l(t+u)]$ are constant in t . Similar to Hawkes (1971), we take the convention that all the $\text{Cov}[\hat{S}_k(t), \hat{S}_l(t+u)]$ are continuous at $u = 0$, which means that they do not include the atomic discontinuity for $u = 0$ and $k = l$ due to the autocorrelation of the stochastic point-processes \hat{S}_k . We refer to ‘complete covariance’ for the second moment that includes the extra contribution $\langle \hat{S}_k(t) \rangle \delta(u)$ for each pair $k = l$, where δ is the Dirac delta function and $\langle \hat{S}_k(t) \rangle$ the constant firing rate.

This convention aims to discriminate between the intrinsic covariance resulting from autocorrelation (always present even for uncorrelated inputs) and the correlation structure that encodes spike synchronization. For uncorrelated inputs, the matrix $\hat{C}(t, u)$ defined in Eq. (3.3) satisfies $\hat{C}(t, u) = 0$ for all $u \in \mathbb{R}$. Therefore, it can be related to the spike-timing information conveyed by the external inputs and encoded in their covariance. However, in the derivation of the self-consistency covariance equations Eq. (3.18), we incorporate terms related to the autocorrelation of the external inputs in order to as-

sess their impact on the learning dynamics.

A.1.2 Properties of the matrix \hat{C}^W

Since the stochastic processes $\hat{S}_k(t)$ have time-invariant first and second stochastic moments, we have $\langle \hat{S}_k(t) \rangle = \text{const.}$ and

$$\langle \hat{S}_k(t) \hat{S}_l(t+u) \rangle = \langle \hat{S}_l(t) \hat{S}_k(t-u) \rangle \quad (\text{A.2})$$

for all indices k and l , which implies $\hat{C}(t, u) = \hat{C}^T(t, -u)$ because the input firing rates $\langle \hat{S}_k(t) \rangle$ are constant for all k . When convoluting with a given kernel $\Psi(t)$, we obtain

$$\begin{aligned} \int_{-\infty}^{+\infty} \Psi(-u) \hat{C}(t, u) du &= - \int_{+\infty}^{-\infty} \Psi(u) \hat{C}(t, -u) du \\ &= \int_{-\infty}^{+\infty} \Psi(u) \hat{C}^T(t, u) du, \end{aligned} \quad (\text{A.3})$$

where we have used a change of variable $u \rightarrow -u$. In particular, this implies that $\hat{C}^V = (\hat{C}^W)^T$ for the time-reverse of the STDP window function $V(u) = W(-u)$.

Moreover, for homogeneous pairwise correlation of inputs, we have

$$\langle \hat{S}_k(t) \hat{S}_l(t+u) \rangle = \langle \hat{S}_l(t) \hat{S}_k(t+u) \rangle, \quad (\text{A.4})$$

which implies that the function $u \mapsto \langle \hat{S}_k(t) \hat{S}_l(t+u) \rangle$ is symmetric in u and thus the matrix \hat{C}^W is symmetric in this case.

A.2 Neuron-to-input covariance consistency equation

In this appendix we derive the self-consistency equations for the covariance coefficients presented in Sec. 3.4.1, which leads to Eq. (3.18). This analysis includes the spike-triggering effects induced by the autocorrelation of the external inputs and of the neurons (Kempster et al. 1999), which were sometimes neglected (Burkitt et al. 2007). In addition, we incorporate the fine-timing effects such as delays and the time course of the PSP response.

The expression for the time-average covariance coefficient F_{ik} in Eq. (3.3) arises from the following definition equivalent to (A.1)

$$\text{Cov}[S_i(t), \hat{S}_k(t+u)] := \langle S_i(t) \hat{S}_k(t+u) \rangle - \langle S_i(t) \rangle \langle \hat{S}_k(t+u) \rangle. \quad (\text{A.5})$$

We consider $\langle \hat{S}_k(t) \rangle$ to be constant in time, which implies that the instantaneous firing rate $\langle S_i(t) \rangle$ for neuron i is quasi-constant due to the slow variation of the weights. The last term in the above expression reduces to $v_i(t) \hat{v}_k(t)$ in Eq. (3.3).

A.2.1 Evaluation of the covariance using the past spiking history

The analysis presented here is based upon that of Hawkes processes (Hawkes 1971). Hawkes processes are stationary second-order processes, and the stationary property strictly holds here for fixed weights $K_{ik}(t)$ (J_{ij} being constant), which we assume in the remainder of this appendix (their dependence upon t will be suppressed here).

The pairwise neuron-to-input correlation $\langle S_i(t) \hat{S}_k(t+u) \rangle$ can be evaluated using the same “stochastic expansion” as Kempter et al. (1999) and Burkitt et al. (2007, Sec. 3.4). It consists of using the definition of the intensity function $\rho_i(t)$ (cf. Eq. (2.1)) and depends on the past activity of the external inputs and of the neurons,

$$\langle S_i(t) \hat{S}_k(t+u) \rangle = \langle \rho_i(t) \hat{S}_k(t+u) \rangle. \quad (\text{A.6})$$

However, this equality hides effects induced by autocorrelation, which arise since $\rho_i(t)$ has an implicit dependence upon \hat{S}_k .

A.2.2 Spike-triggering effect

The expression of $\rho_i(t)$ in Eq. (2.1) can be written as

$$\rho_i(t) = v_0 + \sum_j J_{ij} (\epsilon * S_j)(t - d_{ij}) + \sum_l K_{il} (\epsilon * \hat{S}_l)(t - \hat{d}_{il}). \quad (\text{A.7})$$

When $l = k$, an extra contribution in Eq. (A.6) due to the autocorrelation of the external input k needs to be taken into account, since this term is defined not to be included in $\text{Cov}[\hat{S}_k(t), \hat{S}_l(t')]$ for $k = l$ and $t = t'$ (cf. Eq. (A.1)), as discussed in Appendix A.1.1. When substituting Eq. (A.7) into Eq. (A.6), the term corresponding to $k = l$ is

$$\int \epsilon(r) \hat{S}_k(t - \hat{d}_{ik} - r) \hat{S}_k(t + u) dr \quad (\text{A.8})$$

and taking the ensemble average $\langle \dots \rangle$ leads to

$$\int \epsilon(r) \langle \hat{S}_k(t - \hat{d}_{ik} - r) \hat{S}_k(t + u) \rangle dr + \int \epsilon(r) \langle \hat{S}_k(t + u) \rangle \delta(u + r + \hat{d}_{ik}) dr, \quad (\text{A.9})$$

where δ denotes the Dirac delta function. The second term of Eq. (A.9) is the spike-triggering effect: each pre-synaptic spike from input k induces an extra contribution due to the autocorrelation of input k . The integral in r and the ensemble average brackets $\langle \dots \rangle$ were swapped (Fubini theorem) in order to obtain Eq. (A.9), and ϵ can be taken out of the angular brackets since it is a deterministic function. The spike-triggering effect occurs for $t - r - \hat{d}_{ik} = t + u$, i.e., $r + u + \hat{d}_{ik} = 0$, and it reduces to

$$\epsilon(-u - \hat{d}_{ik}) \langle \hat{S}_k(t + u) \rangle. \quad (\text{A.10})$$

Taking this spike-triggering effect into account, Eq. (A.6) becomes

$$\begin{aligned} \langle S_i(t) \hat{S}_k(t + u) \rangle &= v_0 \langle \hat{S}_k(t + u) \rangle \\ &+ \sum_j J_{ij} \int_{-\infty}^{+\infty} \epsilon(r) \langle S_j(t - r - d_{ij}) \hat{S}_k(t + u) \rangle dr \\ &+ \sum_l K_{il} \int_{-\infty}^{+\infty} \epsilon(r) \langle \hat{S}_l(t - r - \hat{d}_{il}) \hat{S}_k(t + u) \rangle dr \\ &+ K_{ik} \epsilon(-u - \hat{d}_{ik}) \langle \hat{S}_k(t + u) \rangle. \end{aligned} \quad (\text{A.11})$$

A.2.3 Time-averaging

We substitute the equalities (A.11) and (3.15) in equation (A.5) to express $\text{Cov}[S_i(t'), \hat{S}_k(t' + u)]$

$$\begin{aligned}
& \text{Cov}[S_i(t'), \hat{S}_k(t' + u)] & (A.12) \\
&= \sum_j J_{ij} \int \epsilon(r) \text{Cov}[S_i(t' - r - d_{ij}), \hat{S}_k(t' + u)] \, dr \\
&\quad + \sum_l K_{il} \int \epsilon(r) \text{Cov}[\hat{S}_l(t' - r - \hat{d}_{il}), \hat{S}_k(t' + u)] \, dr \\
&\quad + K_{ik} \epsilon(-u - \hat{d}_{ik}) \langle \hat{S}_k(t' + u) \rangle .
\end{aligned}$$

Then we integrate in t' over the time interval $[t - T, t]$ with T much larger than the time scale of the activation mechanisms, so that we can neglect the impact of the two changes of variables $t' \rightarrow t' + r + d_{ij}$, $t' \rightarrow t' + r + \hat{d}_{il}$ and $t' \rightarrow t' - u$, for the terms in the rhs of Eq. (A.12) resp., as Kempter et al. (1999).

$$\begin{aligned}
& \int \text{Cov}[S_i(t'), \hat{S}_k(t' + u)] \, dt' & (A.13) \\
&= \sum_j J_{ij} \int \int \epsilon(r) \text{Cov}[S_i(t'), \hat{S}_k(t' + u + r + d_{ij})] \, dr \, dt' \\
&\quad + \sum_l K_{il} \int \int \epsilon(r) \text{Cov}[\hat{S}_l(t'), \hat{S}_k(t' + u + r + \hat{d}_{il})] \, dr \, dt' \\
&\quad + K_{ik} \int \epsilon(-u - \hat{d}_{ik}) \langle \hat{S}_k(t') \rangle \, dt' .
\end{aligned}$$

Ignoring the slight modifications caused by the changes of variables in t' , the integration bounds are $t' \in [t - T, t]$ and $r \in \mathbb{R}$. After the further changes of variables $r \rightarrow r - d_{ij}$ and $r \rightarrow r - \hat{d}_{il}$, we obtain

$$\begin{aligned}
F_{ik}(t, u) &= \sum_j J_{ij} \int \epsilon(r - d_{ij}) F_{jk}(t, u + r) \, dr & (A.14) \\
&\quad + \sum_l K_{il} \int \epsilon(r - \hat{d}_{il}) \hat{C}_{lk}(t, u + r) \, dr \\
&\quad + K_{ik} \epsilon(-u - \hat{d}_{ik}) \hat{v}_k(t) ,
\end{aligned}$$

where we have used the time-averaged firing rates and covariances defined in equations (3.2), (3.3) and (3.5).

A.2.4 Use of the Fourier transform

The Fourier operator \mathcal{F} between the domains of u and ω for a given function f (\mathbf{i} is the complex root of -1) is given by

$$\mathcal{F}f(\omega) := \int_{-\infty}^{+\infty} f(u) \exp(-\mathbf{i}\omega u) \, du. \quad (\text{A.15})$$

We evaluate the Fourier transform $\mathcal{F}F(\omega)$ of $F(t, u)$ using matrix notation and Eq. (A.14), for fixed t ,

$$\mathcal{F}F(\omega) = \underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \mathcal{F}F(\omega) + \underline{K}(\omega) \mathcal{F}\epsilon(-\omega) \mathcal{F}\hat{C}(\omega) + \underline{K}(\omega) \mathcal{F}\epsilon(-\omega) \text{diag}(\hat{\mathbf{v}}), \quad (\text{A.16})$$

where we defined $\underline{K}_{ik}(\omega) := K_{ik} \exp(\mathbf{i}\hat{d}_{ik}\omega)$ and $\underline{J}_{ij}(\omega) := J_{ij} \exp(\mathbf{i}d_{ij}\omega)$; $\text{diag}(X)$ is the diagonal matrix whose diagonal elements are the vector X .

A.2.5 Sharp distribution of delays

In order to simplify the expressions for \underline{K} and \underline{J} in Eq. (A.16), we now assume that all the recurrent delays are identical ($d_{ij} = d$) and all the input delays are identical ($\hat{d}_{ik} = \hat{d}$). This is a good approximation for sharp distributions of each type of delay. To obtain $\mathcal{F}F^W(\omega)$ ($F^W(t)$ is given in Eq. (3.3)), we multiply Eq. (A.16) by $\exp(-\mathbf{i}\hat{d}\omega) \mathcal{F}W(-\omega)$,

$$\begin{aligned} \mathcal{F}F^W(\omega) &= \underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \mathcal{F}F^W(\omega) \\ &\quad + K \mathcal{F}(W * \epsilon)(-\omega) \mathcal{F}\hat{C}(\omega) \\ &\quad + K \mathcal{F}(W * \epsilon)(-\omega) \text{diag}(\hat{\mathbf{v}}), \end{aligned} \quad (\text{A.17})$$

where $\underline{K}(0) = K$. The expression for $\mathcal{F}F^W(\omega)$ is

$$\mathcal{F}F^W(\omega) = [\mathbb{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)]^{-1} \left[K \mathcal{F}\hat{C}^{W*\epsilon}(\omega) + \mathcal{F}(W * \epsilon)(-\omega) K \text{diag}(\hat{\nu}) \right], \quad (\text{A.18})$$

similar to that given in Hawkes (1971, Eq. (21)).

Expanding the inverse $[\mathbb{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)]^{-1}$ in a power series and taking the inverse Fourier transform of Eq. (A.18), $F^W(t)$ can be rigorously expressed

$$F^W(t) = \sum_{n \geq 0} J^n K \hat{C}^{W*\epsilon*\epsilon_d^{\{n\}}}(t) + \sum_{n \geq 0} \left[W * \epsilon * \epsilon_d^{\{n\}} \right](0) J^n K \text{diag}(\hat{\nu}(t)), \quad (\text{A.19})$$

where $\epsilon_d(t) := \epsilon(t - d)$ and $\epsilon_d^{\{n\}}$ is the n^{th} iterated self-convolution of ϵ_d . We use the convention $W * \epsilon * \epsilon_d^{\{0\}} = W * \epsilon$. The two series are well defined for any PSP kernel ϵ provided all eigenvalues of the weight matrix J have a modulus strictly less than one. Note that the spike-triggering effect is of order M^{-1} compared to the remainder of the synaptic influx for each neuron (in the case of full input connectivity), embodied by the presence of $\text{diag}(\hat{\nu})$ in the last term of Eq. (A.19).

A.2.6 Impact of synaptic mechanisms on the covariance structure

When incorporating the effect of the PSP kernel ϵ and of the recurrent delay d , the input covariance $\hat{C}(t, u)$ in Eq. (A.19) is convolved with $W * \epsilon * \epsilon_d^{\{n\}}$ and not W alone. This implies that the separation between depression and potentiation for $W * \epsilon$ is slightly shifted to the right compared to that of W , as illustrated in Fig. 3.2. Consequently, an input spike that arrives almost immediately after the neuron fires does not cause depression but rather potentiation.

For homogeneous delta-correlated inputs with correlation strength \hat{c}_{av} and firing rate

\hat{v}_{av} (cf. Sec. 3.5), we have

$$\hat{C}^{W*\epsilon*\epsilon_d^{\{n\}}} = \hat{c}_{\text{av}} \hat{v}_{\text{av}} \left[W * \epsilon * \epsilon_d^{\{n\}} \right] (0). \quad (\text{A.20})$$

Then, $[W * \epsilon * \epsilon_d^{\{n\}}](0) > 0$ for all $n \geq 0$ provided $W(u) \geq 0$ for $u < 0$, as described in Sec. 2.3. This means that delta-correlated inputs always induce non-zero correlation coefficients $\hat{C}^{W*\epsilon}$ and thus a non-zero correlation structure F^W in the network. This provides a finer approximation than in Burkitt et al. (2007) and contrasts with the predictions stated in that previous paper, i.e., uncorrelated inputs will induce no correlation structure within the network. Note that the spike-triggering effect is always positive.

The terms of the series for $n \geq 1$ in Eq. (A.19) arise due to the recurrent connections. Only a finite number of terms are non-zero, since $\epsilon_d^{\{n\}}$ vanishes uniformly when $n \rightarrow \infty$ provided ϵ is not a Dirac delta function (i.e., has a finite time course), and the series reduces to a polynomial in J .

A.2.7 Short-duration PSPs and short recurrent delays

In general, the expression in Eq. (A.19) is not tractable. However, we can approximate the solution $\mathcal{F}F^W$ in Eq. (A.18) by making the further assumptions that d is small compared to the time scale of W and that ϵ has a short time course compared to that of W . This implies that

$$\underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \simeq \underline{J}(0) \mathcal{F}\epsilon(0) = J, \quad (\text{A.21})$$

which leads to the expression for $F^W(t)$ in Eq. (3.18) and corresponds to the analysis and the simulations in Chapter 4.

Under these assumptions, the approximation of Eq. (A.21) used to derive Eq. (3.18) is equivalent to the following approximation in Eq. (A.17)

$$\mathcal{F}F^{W*\epsilon_d}(\omega) = \exp(\mathbf{id}\omega) \mathcal{F}\epsilon(-\omega) \mathcal{F}F^W(\omega) \simeq \mathcal{F}F^W(\omega). \quad (\text{A.22})$$

In the time domain, this corresponds to

$$\int W(u-r)\epsilon(r-d)dr \simeq W(u). \quad (\text{A.23})$$

The discrepancies are illustrated in Fig. 3.2, where W is represented by a solid line and $W * \epsilon_d$ by a dashed line. This approximation may be the source of the small discrepancies observed when comparing analytical solutions with numerical simulations.

A.2.8 Long recurrent delays

When d is large compared to the time scale of W , such that $[W * \epsilon * \epsilon_d^{\{n\}}](0) = 0$ for $n \geq 1$, only the first term of the series for $n = 0$ remains in Eq. (A.19), which becomes

$$F^W(t) = K \hat{C}^{W*\epsilon}(t) + [W * \epsilon](0) K \text{diag}(\hat{v}(t)). \quad (\text{A.24})$$

In this case, F^W does not depend on the recurrent weights J and has the same expression as in the case for a feed-forward architecture, i.e., $J = 0$ (Kempster et al. 1999, Sprekeler et al. 2007).

A.3 Neuron-to-neuron covariance consistency equations

In this appendix, we derive the self-consistency equations for the covariance coefficient C^W presented in Sec. 3.4.1, which leads to Eq. (3.18). This analysis includes the spike-triggering effects induced by the autocorrelation of the external inputs and of the neurons (Kempster et al. 1999), which were neglected in Burkitt et al. (2007). In addition, we incorporate the fine-timing effects such as delays and the time course of the PSP response.

The instantaneous neuron covariance is given the following definition equivalent to (A.1)

$$\text{Cov}[S_i(t), S_j(t+u)] := \langle S_i(t')S_j(t'+u) \rangle - \langle S_i(t') \rangle \langle S_j(t'+u) \rangle. \quad (\text{A.25})$$

As in Appendix A.1 for the input covariances, we consider that the function of u in

Eq. (A.25) and thus $C_{ij}(t, u)$ in Eq. (3.3) are continuous in $u = 0$.

A.3.1 Taking the autocorrelation into account

Similar to Appendix A.2, we use the definition of $\rho_i(t)$,

$$\langle S_i(t) S_j(t+u) \rangle = \langle \rho_i(t) S_j(t+u) \rangle, \quad (\text{A.26})$$

where the rhs incorporates terms due to autocorrelation. However, the derivation of the consistency equation for the neuron-to-neuron covariance is a bit more complex than for the input-to-neuron covariance, due to the recurrent connections and the probabilistic interdependence with the past spiking history of the network that they imply. Here, we adapt the original derivation in Hawkes (1971, Eqs. (12) and (24)).

We freeze t and consider the Fourier transform of $C(t, u)$ by integrating u , using the adiabatic assumption that the weights J are quasi-constant. The key-point of this derivation is that C satisfies

$$C(t, -u) = C^T(t, u), \quad (\text{A.27})$$

under the assumption of slow learning for the weights (quasi time-invariant first and second stochastic moments), as explained in Appendix A.1.2. In order to calculate $C(t, u)$, we use Eq. (A.26)

$$\begin{aligned} \langle S_i(t) S_j(t+u) \rangle &= v_0 \langle S_j(t+u) \rangle + \sum_{i'} J_{i'i'} \langle (\epsilon * S_{i'})(t - d_{i'i'}) S_j(t+u) \rangle \\ &+ \sum_k K_{ik} \langle (\epsilon * \hat{S}_k)(t - \hat{d}_{ik}) S_j(t+u) \rangle + J_{ij} \epsilon(-u - d_{ij}) \langle S_j(t+u) \rangle. \end{aligned} \quad (\text{A.28})$$

The last term is a spike-triggering effect due to a recurrent connection, corresponding to $t - r - d_{ij} = t + u$. It is important to note that this equality is only valid for the half-plane $u < 0$.

We then use Eq. (3.15) and proceed to the time-averaging over $[t - T, t]$ in order to

obtain

$$\begin{aligned} C_{ij}(t, u) &= \sum_{i'} J_{i'j} \int \epsilon(r - d_{i'j}) C_{i'j}(t, u + r) dr \\ &+ \sum_k K_{ik} \int \epsilon(r - \hat{d}_{ik}) F_{jk}(t, -u - r) dr + J_{ij} \epsilon(-u - d_{ij}) v_j(t). \end{aligned} \quad (\text{A.29})$$

The following changes of variable were made: $t \rightarrow t + r + d_{i'j}$ and $r \rightarrow r - d_{i'j}$ for terms in the first sum in i' on the rhs, and $t \rightarrow t - u + d_{ij}$ and $r \rightarrow r - \hat{d}_{ik}$ for the second sum in k . Note the negative sign for u in F_{jk} . Recall that this is only valid for $u < 0$, so information from the past is used to evaluate the impact of the autocorrelation on the covariance structure and we evaluate the total effect using the symmetry in u of C , cf. Eq. (A.27).

We now assume, as in Appendix A.2, that the delays $d_{ij} = d$ and $\hat{d}_{ik} = \hat{d}$ are sharply distributed. We would like to take the Fourier transform of Eq. (A.29) in order to obtain an equivalent of the consistency equation for the input-to-neuron covariance in Appendix A.2. However, Eq. (A.29) is only valid for $u < 0$. As in Hawkes (1971, Eq. (22)), we introduce the matrix \check{C} defined by the Fourier transform on the rhs of Eq. (A.29) less the Fourier transform of C , namely $\mathcal{F}C(\omega)$

$$\begin{aligned} \check{C}(\omega) &:= -\mathcal{F}C(\omega) + \underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \mathcal{F}C(\omega) + \underline{K}(-\omega) \mathcal{F}\epsilon(\omega) \mathcal{F}F^T(-\omega) \\ &+ \underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \text{diag}(\mathbf{v}). \end{aligned}$$

The matrix $\check{C}(\omega)$ thus defined incorporates the effects induced by autocorrelation for the “future” ($u > 0$ in Eq. (A.28)). An argument on the regularity of $\omega \mapsto \check{C}(\omega)$ (i.e., \check{C} is holomorphic) is used in Hawkes (1971) to evaluate it.

Expressing $\mathcal{F}C$ in terms of \check{C} from Eq. (A.30)

$$\begin{aligned}
\mathcal{F}C(\omega) &= [\mathbb{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)]^{-1} & (A.30) \\
&\quad \left[-\check{C}(\omega) + \underline{K}(-\omega) \mathcal{F}\epsilon(\omega) \mathcal{F}F^T(-\omega) + \underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \text{diag}(\mathbf{v}) \right] \\
&= [\mathbb{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)]^{-1} \left[\underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \text{diag}(\mathbf{v}) - \check{C}(\omega) \right] \\
&\quad + [\mathbb{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)]^{-1} \underline{K}(\omega) \mathcal{F}\epsilon(-\omega) [\mathcal{F}\hat{C}(\omega) + \text{diag}(\hat{\mathbf{v}})] \\
&\quad \underline{K}^T(-\omega) \mathcal{F}\epsilon(\omega) [\mathbb{1}_N - \underline{J}(-\omega) \mathcal{F}\epsilon(\omega)]^{-1T} ,
\end{aligned}$$

using the expression of $\mathcal{F}F$ in (A.16),

$$\mathcal{F}F(\omega) = [\mathbb{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)]^{-1} \underline{K}(\omega) \mathcal{F}\epsilon(-\omega) [\mathcal{F}\hat{C}(\omega) + \text{diag}(\hat{\mathbf{v}})] , \quad (A.31)$$

and also $\mathcal{F}\hat{C}(-\omega) = \mathcal{F}\hat{C}^T(\omega)$, cf. Appendix A.1.2.

Now using Eq. (A.27) with the expression of $\mathcal{F}C$ in Eq. (A.30), we obtain

$$\begin{aligned}
&[\mathbb{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)]^{-1} \left[\underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \text{diag}(\mathbf{v}) - \check{C}(\omega) \right] & (A.32) \\
&= \left[\underline{J}(-\omega) \mathcal{F}\epsilon(\omega) \text{diag}(\mathbf{v}) - \check{C}(-\omega) \right]^T [\mathbb{1}_N - \underline{J}(-\omega) \mathcal{F}\epsilon(\omega)]^{-1T} ,
\end{aligned}$$

since the last term involving \hat{C} and $\text{diag}(\hat{\mathbf{v}})$ in the rhs of Eq. (A.30) satisfies an equality similar to Eq. (A.27).

Equation (A.32) can be reorganized

$$\begin{aligned}
&\underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \text{diag}(\mathbf{v}) + [\mathbb{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)] \check{C}^T(-\omega) & (A.33) \\
&= \text{diag}(\mathbf{v}) \mathcal{F}\epsilon(\omega) \underline{J}^T(-\omega) + \check{C}(\omega) [\mathbb{1}_N - \underline{J}(-\omega) \mathcal{F}\epsilon(\omega)]^T .
\end{aligned}$$

The equality Eq. (A.33) allows us to define a function that is regular on the whole plane ω , each side being regular for the half of the plane related to the sign of the imaginary part of ω . This requires assumptions on the exponentially-fast decay of $\epsilon(u)$ for $u \rightarrow +\infty$ and of elements of \check{C} . The so-defined holomorphic function vanishes when $|\omega| \rightarrow \infty$, which implies that it is actually zero on the whole plane. Consequently, we have the expression of \check{C} in terms of the weight matrices K and J (modified to incorporate the effect of the PSP

kernel ϵ and delays), of the covariance between the neurons and the inputs (through F) and of the autocorrelation of the processes ($\text{diag}(\boldsymbol{v})$),

$$\check{C}(\omega) = -\text{diag}(\boldsymbol{v}) \mathcal{F}\epsilon(\omega) \underline{J}^{\mathbf{T}}(-\omega) [\mathbf{1}_N - \underline{J}(-\omega) \mathcal{F}\epsilon(\omega)]^{-1\mathbf{T}}. \quad (\text{A.34})$$

Finally, we obtain the expression for $\mathcal{F}C$ by using Eq. (A.34) in Eq. (A.30), similar to the expression for $\mathcal{F}F$ in Appendix A.2:

$$\begin{aligned} \mathcal{F}C(\omega) &= [\mathbf{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)]^{-1} \quad (\text{A.35}) \\ &\quad \left\{ \underline{K}(\omega) \mathcal{F}\epsilon(-\omega) [\mathcal{F}\hat{C}(\omega) + \text{diag}(\hat{\boldsymbol{v}})] \underline{K}^{\mathbf{T}}(-\omega) \mathcal{F}\epsilon(\omega) \right. \\ &\quad \left. + \underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \text{diag}(\boldsymbol{v}) + \text{diag}(\boldsymbol{v}) \mathcal{F}\epsilon(\omega) \underline{J}^{\mathbf{T}}(-\omega) \right. \\ &\quad \left. - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \text{diag}(\boldsymbol{v}) \mathcal{F}\epsilon(\omega) \underline{J}^{\mathbf{T}}(-\omega) \right\} \\ &\quad [\mathbf{1}_N - \underline{J}(-\omega) \mathcal{F}\epsilon(\omega)]^{-1\mathbf{T}} \\ &= [\mathbf{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)]^{-1} \underline{K}(\omega) \mathcal{F}\epsilon(-\omega) [\mathcal{F}\hat{C}(\omega) + \text{diag}(\hat{\boldsymbol{v}})] \\ &\quad \underline{K}^{\mathbf{T}}(-\omega) \mathcal{F}\epsilon(\omega) [\mathbf{1}_N - \underline{J}(-\omega) \mathcal{F}\epsilon(\omega)]^{-1\mathbf{T}} \\ &\quad + [\mathbf{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)]^{-1} \text{diag}(\boldsymbol{v}) [\mathbf{1}_N - \underline{J}(-\omega) \mathcal{F}\epsilon(\omega)]^{-1\mathbf{T}} - \text{diag}(\boldsymbol{v}). \end{aligned}$$

A.3.2 Remark on the autocorrelation structure due to the recurrent connections

The autocorrelation effects are the three terms in the first expression of $\mathcal{F}C(\omega)$ in Eq. (A.35) involving ‘diag’. Note that the complete covariance impacts upon STDP, i.e., including the first-order autocorrelation that corresponds to $u = 0$, which is added to $C(t, u)$ in the convolution with W in the learning equation. This actually corresponds to the last term $\text{diag}(\boldsymbol{v})$ in the second expression of $\mathcal{F}C(\omega)$ in Eq. (A.35). Refer to Appendix A.1.1 for the distinction between covariance and complete covariance (Hawkes 1971) and its relationship to the encoding of neuronal information.

By naively taking only the spike-triggering effect as an extra contribution in Eq. (A.26),

one obtains $\underline{J}(\omega) \mathcal{F}\epsilon(-\omega) \text{diag}(\boldsymbol{\nu})$. The consideration of the double expansion

$$\langle S_i(t) S_j(t+u) \rangle = \langle \rho_i(t) \rho_j(t+u) \rangle \quad (\text{A.36})$$

using the expression for $\rho_i(t)$ in Eq. (2.1) may lead to the expression of Eq. (A.35), since it preserves the symmetry between the neurons i and j . However, care must be taken to ensure that the terms involving $\text{diag}(\boldsymbol{\nu})$ are correctly considered.

A.3.3 Short recurrent delays

The function $\mathcal{F}C^W$ can be obtained by multiplying the lhs of Eq. (A.35) by the term $\exp(-\mathbf{id}\omega) \mathcal{F}W(-\omega)$, to incorporate the impact of W and the delays d_{ij} . The inverse of $\mathbb{1}_N - \underline{J}(\omega) \mathcal{F}\epsilon(-\omega)$ could be developed in a power series in order to obtain a rigorous expression of $C^W(t)$. This actually leads to a double series because of the two occurrences of the inverse matrix. For delta-correlated inputs and an STDP window function W with compact support, only a finite number of terms remain in the double series when ϵ is different from the Dirac delta function (cf. Appendix A.2). Unlike the input-to-neuron covariance, a larger value for the delay d does not lead to a single term and the expression is more difficult to handle.

To simplify the expression for $\mathcal{F}C$ in Eq. (A.35), we use the approximation in (A.21) assuming the short durations of ϵ and d . This allows us to deal with the inverses and express C in the time domain u using the inverse Fourier transform. We obtain the following expression of the term $C^W(t) + W(d) \text{diag}(\boldsymbol{\nu}(t))$ due to STDP in the rhs of the learning matrix equation of J (cf. Sec. A.3.2)

$$\begin{aligned} & C^W(t) + W(d) \text{diag}(\boldsymbol{\nu}(t)) \\ &= [\mathbb{1}_N - J]^{-1} K \left[\mathcal{F}\hat{C}^{W*\zeta}(t) + [W * \zeta](0) \text{diag}(\hat{\boldsymbol{\nu}}) \right] K^T [\mathbb{1}_N - J]^{-1T} \\ & \quad + W(d) [\mathbb{1}_N - J]^{-1} \text{diag}(\boldsymbol{\nu}) [\mathbb{1}_N - J]^{-1T}. \end{aligned} \quad (\text{A.37})$$

The function ζ describes the impact of the PSP kernel on the input covariance structure.

It corresponds to the inverse Fourier transform of (cf. Eq. (A.35))

$$\exp(\mathbf{i} \hat{d} \omega) \mathcal{F}\epsilon(-\omega) \exp(-\mathbf{i} \hat{d} \omega) \mathcal{F}\epsilon(\omega) \exp(-\mathbf{i} d \omega), \quad (\text{A.38})$$

but reversed in time since the convolution with W corresponds to $\mathcal{F}W(-\omega)$, i.e.,

$$\zeta(r) := \int \epsilon(r + r' + d) \epsilon(r') \, dr' \simeq \int \epsilon(r + r') \epsilon(r') \, dr'. \quad (\text{A.39})$$

The expression in Eq. (A.37) differs from its equivalent in Burkitt et al. (2007) by the autocorrelation terms involving ‘diag’ and the convolution $W * \zeta$. The recurrent connections induce intrinsic correlation structure within the network, which are at the first order of the recurrence described by ζ . This may partly explain the small discrepancies between theoretical predictions and simulation results in Burkitt et al. (2007).

Appendix B

Calculations for chapter 4

B.1 Analysis of the drift of K due to STDP with fixed J

We present in this appendix the main arguments about the general solution of the differential equation Eq. (4.10) that describes the evolution of input weights $K(t)$. The results are summarized in Sec. 4.2.2.

B.1.1 Symmetries of the inputs and reduction of dimensionality for K

Here we decompose the space \mathbb{M}_K , in which $K(t)$ evolves according to Eq. (4.10), depending on the symmetries of the input pools and input connectivity. We want to reduce the dimensionality of the matrix $K(t)$ in order to eliminate the subspaces within which the drift $\dot{K}(t) = 0$ always, in order to focus on the complementary subspace where the drift is meaningful and leads to the development of a structure in $K(t)$. We constrain this section to full input connectivity (Φ_K is the identity and $\mathbb{M}_K = \mathbb{R}^{N \times M}$) but the results can also be applied to the case of partial connectivity, after the transform detailed in Appendix B.1.3.

For each symmetry of the input pools and input connectivity, say inputs $\hat{1}$ and $\hat{2}$ belong to the same input pool and are interchangeable, we can construct a M -column vector $\hat{\mathbf{u}}_D := [1, -1, 0, \dots, 0]^T$ such that $A\hat{\mathbf{u}}_D = 0$ and $B\hat{\mathbf{u}}_D = 0$, which leads to $\dot{K}(t)\hat{\mathbf{u}}_D = 0$ always whatever the value of $K(t)\hat{\mathbf{u}}_D$. This implies that the value of $K\hat{\mathbf{u}}_D$ is not constrained by the drift of the dynamics determined by Eq. (4.10). Furthermore, higher stochastic orders of the weight dynamics may affect this value without any effect on the drift of K . The i^{th} element of the column vector $K\hat{\mathbf{u}}_D$ corresponds to the difference $K_{i1} - K_{i2}$ between the weights from these two inputs. A displacement of K along this sole direction in \mathbb{M}_K ,

i.e., modifying $K_{i1} - K_{i2}$ and preserving all other matrix components in a suitable basis, consists in a redistribution between the weights K_{i1} and K_{i2} .

In order to study the drift $\dot{K}(t)$, we can thus define equivalence classes \bar{K} of the matrices K in \mathbb{M}_K modulo such redistributions of weights that do not impact the drift. In other words, matrices K belonging to the same class \bar{K}_0 have the same drift \dot{K}_0 . The drift of $K(t)$ is completely captured by the evolution of $\bar{K}(t)$ in the “reduced” vector space of the equivalence classes determined by the symmetries. The evolution of K within classes \bar{K} is only due to higher orders of the stochastic processes. A similar reduction can be performed when the neurons and recurrent weights also have symmetries, as described in Fig. 4.2.

In the reduced space, equivalence classes \bar{K} lump the weights that correspond to symmetries. Taking the example above with $\#1$ and $\#2$, \bar{K} is only concerned about the sum $K_{i1} + K_{i2}$ for all indices i , not the difference $K_{i1} - K_{i2}$; we thus reduce \mathbb{M}_K by M components. Using matrix notation, we focus on $K\hat{\mathbf{u}}_S$ with $\hat{\mathbf{u}}_S := [1, 1, 0, \dots, 0]^T$ and not $K\hat{\mathbf{u}}_D$, as defined above. Generalizing to the case of many symmetries, the elements of \bar{K} can be taken equal to the mean input weights (instead of the sums of weights) averaged over the considered inputs and neurons, when several neurons are involved.

Such a reduction of dimensionality can also be applied in the case where the parameters within an input pool or a neuron group are not strictly identical, but where they are sharply distributed and the connectivity can be considered homogeneous up to some “noise” in the parameters. Then the equivalence classes correspond to the respective mean weights.

B.1.2 General evolution for full input connectivity

Let us consider the evolution of the drift $\dot{K}(t)$ described in Eq. (4.10) when the matrix A defined in Eq. (4.11) is non-invertible. We also assume full input connectivity: the matrix K evolves in the space $\mathbb{M}_K = \mathbb{R}^{N \times M}$. We show how the structure of A and B determines the evolution of the matrix $K(t)$. We first assume that A is diagonalizable and the space

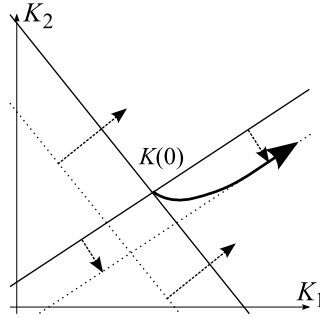


Figure B.1: Example of evolution of K in two dimensions. The direction of each thick solid line is determined by \hat{u}_1 and \hat{u}_2 resp.; their intersection corresponds to $K(0)$. The thick dotted lines correspond to one stable fixed point $K(\infty)\hat{u}_1$ (dashed arrows pointing towards the dotted line) and one unstable fixed point $K(\infty)\hat{u}_2$ (dashed arrows pointing away from the dotted line). Learning causes the value of $K(t)$ to reach the line corresponding to $K\hat{u}_1 = K(\infty)\hat{u}_1$ while pushing it away from the intersection of the dashed lines (thick arrow) until reaching the upper bound of K_1 .

\mathbb{R}^M can be decomposed into the direct sum of three subspaces of eigenvectors \hat{u} .

By restricting A to the quotient space obtained by factoring out the null-space of A , we can use the same formula as Eq. (4.12), since the restriction of A is invertible on the quotient space ($A\hat{u} \neq 0$ in this subspace; take \hat{u}_5 in Appendix B.1.1 for example). The restriction of the weight matrix K thus converges or diverges according to the (non-zero) eigenvalues of the restriction of A . In this case, K will evolve subject to constraints determined by A and B and the network will learn the input firing-rate and correlation structures. For example, the case of one stable fixed point and one unstable fixed point for two components of K is illustrated in Fig. B.1.

We now examine the behavior of the weights in the intersection of the two null-spaces of A and B , i.e., eigenvectors \hat{u} such that $A\hat{u} = 0$ and $B\hat{u} = 0$, which implies $\dot{K}\hat{u} = 0$. Such vectors \hat{u} exist, for example, when the network has symmetries, cf. \hat{u}_D in Appendix B.1.1. In this subspace, only higher orders of the stochastic dynamics (cf. Sec. 3.6) drive the evolution of the weights K and the value of $K\hat{u}$ can be arbitrary; it depends in particular on the initial conditions. Changing the value of $K\hat{u}$ corresponds to a redistribution of the strengths of the weights that has no impact on the weight structure.

Finally, for any eigenvector \hat{u} such that $A\hat{u} = 0$ and $B\hat{u} \neq 0$, we have $\dot{K}\hat{u} = B\hat{u} = \text{const}$. Thus, $K\hat{u}$ will grow linearly in time until the weights hit the bounds. This situation,

however, corresponds to very specific values of the input and learning parameters. For example, from Eq. (4.11) the choice $\hat{\mathbf{v}} = -w^{\text{out}}\hat{\mathbf{e}}/\tilde{W}$ with uncorrelated inputs ($\hat{C}^W = 0$) gives $A = 0$ and $B = w^{\text{in}}\hat{\mathbf{e}}\hat{\mathbf{v}}^T \neq 0$, when $w^{\text{in}} \neq 0$. We do not investigate this case any further.

In general, A is not diagonalizable and the decomposition above is not as simple; for example, the image of A can intersect with its null-space. But the set of diagonalizable matrices is dense in \mathbb{M}_K and thus the previous analysis can be extended to all matrices in \mathbb{M}_K , because the behavior of K qualitatively depends on the eigenvalues of A only.

B.1.3 Partial input connectivity

We now look at the dynamics of $K(t)$ for partial input connectivity, when Φ_K nullifies some terms related to non-existing connections; the space \mathbb{M}_K is then a strict subspace of $\mathbb{R}^{N \times N}$. Instead of considering K a matrix, we take it as a column vector \check{K} indexed by the duplet (i, k) such that the connection $k \rightarrow i$ exists and we omit the elements nullified by Φ_K . Eq. (4.10) becomes

$$\dot{\check{K}} = L\check{K} + \check{B}, \quad (\text{B.1})$$

where \check{B} is the column vector constructed from B in the same way as \check{K} from K , and L is a square matrix of dimension $n^K \times n^K$ defined by its elements indexed by $\{(i, k)(i', k')\}$

$$L_{\{(i, k)(i', k')\}} := \check{J}_{ii'} A_{k'k}, \quad (\text{B.2})$$

where $\check{J} := (\mathbb{1}_N - J)^{-1}$. This equation can be analyzed in the same way as in Appendix B.1.2, with a basis of column vectors $\check{\mathbf{u}}$ instead of $\hat{\mathbf{u}}$. The cases where $\check{\mathbf{u}}^T L$ and $\check{\mathbf{u}}^T \check{B}$ are zero or non-zero are to be considered as above (note the transposition 'T'). It follows that the evolution of \check{K} can be decomposed into evolution within three subspaces as in Appendix B.1.2.

When A is invertible, a generalization of Eq. (4.12) can be written as

$$K(t) = K(\infty) + \sum_{n \geq 0} \frac{t^n}{n!} \tilde{K}_n \quad (\text{B.3})$$

with

$$\begin{aligned} \tilde{K}_{n+1} &:= \Phi_K \left[(\mathbb{1}_N - J)^{-1} \tilde{K}_n A \right], \\ \tilde{K}_0 &:= K(0) - K(\infty), \\ K(\infty) &= -\Phi_K \left[(\mathbb{1}_N - J) \Phi_K(B) A^{-1} \right]. \end{aligned} \quad (\text{B.4})$$

B.2 Dependence of the fixed point $K(\infty)\hat{\mathbf{h}}$ upon input correlation

Here we derive a condition on the input correlation strengths \hat{c}_1 and \hat{c}_2 such that the sign of the elements of the fixed point $K(\infty)\hat{\mathbf{h}}$ in Eq. (4.25) is determined by the balance between the correlations \hat{c}_1 and \hat{c}_2 and not by that of the firing rates \tilde{v}_1 and \tilde{v}_2 . Recall that this sign determines the evolution of K , i.e., which input pathway is potentiated by learning, for homogeneous initial input weights.

We focus on the role of the correlation strengths \hat{c}_1 and \hat{c}_2 in the numerator $n_{\text{av}}^K K_{\text{av}}^* \gamma + (1 - n_{\text{av}}^J J_{\text{av}}) \gamma' + \kappa'$. Multiplying the numerator by the denominator of K_{av}^* in Eq. (4.6) gives

$$\begin{aligned} & \left[(1 - n_{\text{av}}^J J_{\text{av}}) w^{\text{in}} \hat{v}_{\text{av}} + \nu_0 (w^{\text{out}} + \tilde{W} \hat{v}_{\text{av}}) \right] \left[\tilde{W} \hat{v}_{\text{av}} \frac{\tilde{v}_1 - \tilde{v}_2}{2} + [W * \epsilon](0) \frac{\hat{c}_1 \tilde{v}_1 - \hat{c}_2 \tilde{v}_2}{4} \right] \\ & - \left[(1 - n_{\text{av}}^J J_{\text{av}}) w^{\text{in}} \frac{\tilde{v}_1 - \tilde{v}_2}{2} + \tilde{W} \nu_0 \frac{\tilde{v}_1 - \tilde{v}_2}{2} \right] \left[\hat{v}_{\text{av}} (w^{\text{out}} + \tilde{W} \hat{v}_{\text{av}}) + \hat{C}_{\text{av}}^{W * \epsilon} \right] \quad (\text{B.5}) \\ & = \frac{\tilde{v}_1 - \tilde{v}_2}{2} \hat{C}_{\text{av}}^{W * \epsilon} \left[- (1 - n_{\text{av}}^J J_{\text{av}}) w^{\text{in}} - \tilde{W} \hat{v}_0 \right] \\ & \quad + \frac{\hat{c}_1 \tilde{v}_1 - \hat{c}_2 \tilde{v}_2}{4} [W * \epsilon](0) \left[(1 - n_{\text{av}}^J J_{\text{av}}) w^{\text{in}} \hat{v}_{\text{av}} + \nu_0 (w^{\text{out}} + \tilde{W} \hat{v}_{\text{av}}) \right], \end{aligned}$$

where we have used the means $\hat{v}_{\text{av}} = (\tilde{v}_1 + \tilde{v}_2)/2$ and $\hat{C}_{\text{av}}^{W * \epsilon} = [W * \epsilon](0) (\hat{c}_1 \tilde{v}_1 + \hat{c}_2 \tilde{v}_2)/4$; the expressions for γ , γ' and κ' are given in Eq. (4.18).

If the difference $\hat{c}_1 \tilde{v}_1 - \hat{c}_2 \tilde{v}_2$ dominate the rhs of Eq. (B.5), then the correlation strengths

\hat{c}_1 and \hat{c}_2 determine the sign of the fixed point $K(\infty)\hat{\mathbf{h}}$. This condition can be rewritten

$$\left| \frac{\hat{c}_1 \bar{v}_1 - \hat{c}_2 \bar{v}_2}{\bar{v}_1 - \bar{v}_2} \right| > \frac{2\hat{C}_{\text{av}}^{W*\epsilon}}{[W * \epsilon](0)} \left| \hat{v}_{\text{av}} + \frac{w^{\text{out}} v_0}{(1 - n_{\text{av}}^I J_{\text{av}}) w^{\text{in}} + \tilde{W} v_0} \right|^{-1}. \quad (\text{B.6})$$

B.3 Symmetry breaking within K for different neurons

This appendix details some calculations related to the study of the impact of recurrent connections on the symmetry breaking performed by STDP on input connections through the second stochastic moment of their weight dynamics.

B.3.1 Second moment of the stochastic evolution of K

Here we consider $Y_{i,k,j,k}(t, t')$ defined in Eq. (3.25) for indices i, j and $k = l$. This coefficient relates to the relative evolution of the weights K_{ik} and K_{jk} : the sign of $Y_{i,k,j,k}(t, t')$ indicates whether K_{ik} and K_{jk} tend to evolve in the same direction or not (potentiation or depression). We only consider the simplified case of identical input firing rates $\hat{v}_k = \hat{v}_0$ and spike-time correlation (\hat{c}_0). From Eq. (3.24) we have

$$\begin{aligned} & \frac{dK_{ik}^\omega(t)}{dt} \frac{dK_{jk}^\omega(t')}{dt} \\ &= \left[(w^{\text{in}})^2 \hat{S}_k(t - \hat{d}) \hat{S}_k(t' - \hat{d}) + (w^{\text{out}})^2 S_i(t) S_j(t') \right. \\ & \quad + w^{\text{in}} w^{\text{out}} \hat{S}_k(t - \hat{d}) S_j(t') + w^{\text{in}} w^{\text{out}} S_i(t) \hat{S}_k(t' - \hat{d}) \\ & \quad + w^{\text{in}} \int W(u) S_i(t) \hat{S}_k(t + u - \hat{d}) \hat{S}_k(t' - \hat{d}) du \\ & \quad + w^{\text{in}} \int W(u) \hat{S}_k(t - \hat{d}) \hat{S}_k(t' + u' - \hat{d}) S_j(t') du' \\ & \quad + w^{\text{out}} \int W(u) S_i(t) \hat{S}_k(t + u - \hat{d}) S_j(t') du \\ & \quad \left. + w^{\text{out}} \int W(u') S_i(t) \hat{S}_k(t' + u' - \hat{d}) S_j(t') du' \right. \\ & \quad \left. + \int \int W(u) W(u') S_i(t) \hat{S}_k(t + u - \hat{d}) \hat{S}_k(t' + u' - \hat{d}) S_j(t') du du' \right]. \end{aligned} \quad (\text{B.7})$$

The leading-order drift obtained when taking the expectation value of the sum of these nine terms is $\langle \frac{dK_{ik}^\omega(t)}{dt} \rangle \langle \frac{dK_{jk}^\omega(t')}{dt} \rangle$, while neglecting the autocorrelation effects and some probabilistic interdependence of the spike trains \hat{S}_k , S_i and S_j . This leading-order term is almost zero when the input mean weights are stable around their equilibrium value for all neurons, which follows from

$$\left\langle \frac{dK_{ik}^\omega(t)}{dt} \right\rangle = \dot{K}_{ik}(t) = 0 \quad (\text{B.8})$$

for all indices k and i . Consequently, higher orders involving autocorrelation effects of inputs and neurons may have an impact on the evolution of K_{ik} and K_{jk} when they have reached the homeostatic equilibrium. We identify different kinds of contributions: the first-order autocorrelation terms that are independent of the network connectivity, which will not be discussed here, see Kempter et al. (1999) for details; spike-triggering effects (second-order in terms of autocorrelation) that depend on the connectivity; and further orders that will not be considered, i.e., terms that arise from recurrent synaptic paths of length two or more.

B.3.2 Recurrent connections and spike-triggering effect

We focus on the spike-triggering effects related to recurrent connections when taking the ensemble average of Eq. (B.7) for two given neurons $i \neq j$ and a given input k . First, we consider a single recurrent connection $j \rightarrow i$ with weight $J_{ij} > 0$, ignoring all other recurrent connections. Spike-triggering effects due to the autocorrelation of input k arise in the second, seventh, eighth and ninth terms of the rhs of Eq. (B.7).

In the second term of Eq. (B.7), taking the ensemble average of $S_i(t)S_j(t')$ induces an additional term $J_{ij} \epsilon(t - t' - d) \langle S_j(t') \rangle$ due to the autocorrelation of neuron j , which arises from the relationship

$$\langle S_i(t)S_j(t') \rangle = \langle \rho_i(t)S_j(t') \rangle, \quad (\text{B.9})$$

where $\rho_i(t)$ involves $J_{ij} [\epsilon * S_j](t - d)$. This gives the following contribution to $Y_{i,k,j,k}(t, t')$ induced by J_{ij}

$$(w^{\text{out}})^2 J_{ij} \epsilon(t - t' - d) \langle S_j(t') \rangle . \quad (\text{B.10})$$

For each spike fired by neuron j at time t' , there is a non-zero contribution given by Eq. (B.10) for all times $t \geq t' + d$ such that $\epsilon(t - t' - d) \neq 0$.

The seventh term of Eq. (B.7) gives

$$\begin{aligned} w^{\text{out}} J_{ij} \int W(u) \epsilon(t - t' - d) \langle S_j(t') \hat{S}_k(t + u - \hat{d}) \rangle du \\ \simeq w^{\text{out}} J_{ij} \epsilon(t - t' - d) \langle S_j(t') \rangle \int W(u) \langle \hat{S}_k(t + u - \hat{d}) \rangle du , \end{aligned} \quad (\text{B.11})$$

where $S_j(t')$ and $\hat{S}_k(t + u - \hat{d})$ are taken to be independent, which is equivalent to considering only the leading-order in terms of the autocorrelation of neuron j . Likewise, the eighth term of Eq. (B.7) gives

$$\begin{aligned} w^{\text{out}} J_{ij} \int W(u') \epsilon(t - t' - d) \langle S_j(t') \hat{S}_k(t' + u' - \hat{d}) \rangle du' \\ \simeq w^{\text{out}} J_{ij} \epsilon(t - t' - d) \langle S_j(t') \rangle \int W(u') \langle \hat{S}_k(t' + u' - \hat{d}) \rangle du' . \end{aligned} \quad (\text{B.12})$$

Finally, the ninth term in Eq. (B.7) gives

$$\begin{aligned} J_{ij} \int \int W(u) W(u') \epsilon(t - t' - d) \langle S_j(t') \hat{S}_k(t + u - \hat{d}) \hat{S}_k(t' + u' - \hat{d}) \rangle du du' \\ \simeq J_{ij} \epsilon(t - t' - d) \langle S_j(t') \rangle \int \int W(u) W(u') \langle \hat{S}_k(t + u - \hat{d}) \rangle \langle \hat{S}_k(t' + u' - \hat{d}) \rangle du du' . \end{aligned} \quad (\text{B.13})$$

Summing the four terms in Eqs. (B.10), (B.11), (B.12) and (B.13), we obtain the total contribution to $Y_{i,k,j,k}(t, t')$ induced by the single weight J_{ij}

$$J_{ij} \epsilon(t - t' - d) \langle S_j(t') \rangle \left[w^{\text{out}} + \int W(u) \langle \hat{S}_k(t + u - \hat{d}) \rangle du \right]^2 , \quad (\text{B.14})$$

which is positive since the instantaneous firing rate $\langle S_j(t') \rangle$, ϵ and the recurrent weight J_{ij} are positive.

This additional contribution implies that $Y_{i,k,j,k}(t, t')$ is more positive in the presence

of the recurrent connection $j \rightarrow i$. This induces a more positively correlated evolution of K_{ik} and K_{jk} , which means that they tend to evolve in the same direction: either they both increase or both decrease.

Since weights vary slowly compared to the time scale of the neuronal activation mechanisms related to ϵ , d and \hat{d} , we integrated Eq. (B.14) over time to obtain the time-averaged effect. In the case of homogeneous inputs, this leads to

$$J_{ij} \nu_{\text{av}} \left(w^{\text{out}} + \tilde{W} \hat{\nu}_{\text{av}} \right)^2, \quad (\text{B.15})$$

since the integral of ϵ is normalized to one. Using the approximation of the equilibrium value of ν_{av} in Eq. (4.9), the expression Eq. (B.15) becomes

$$-J_{ij} \hat{\nu}_{\text{av}} w^{\text{in}} \left(w^{\text{out}} + \tilde{W} \hat{\nu}_{\text{av}} \right) > 0. \quad (\text{B.16})$$

Recall that $(w^{\text{out}} + \tilde{W} \hat{\nu}_{\text{av}}) < 0$ is required for homeostatic stability.

Note that, because J has no self-connections, the diagonal terms J_{ii} do not contribute to the variance of the input weights, which is related to $Y_{i,k,i,k}(t, t')$.

B.3.3 Arbitrary homogeneous connectivity

We now consider the situation when each input and recurrent connection have the probability n^K/NM and $n^J/N(N-1)$ resp. of existing (recall that n^K and n^J are the number of input and recurrent connections, resp.). We average Eq. (B.15) over the whole network for all triplets (k, i, j) to obtain the time-averaged contribution to $\sum_{k \rightarrow i} \sum_{k \rightarrow j} Y_{i,k,j,k}(t, t')$ due to all recurrent connections

$$\begin{aligned} & MN(N-1) \left(\frac{n^K}{NM} \right)^2 \frac{n^J}{N(N-1)} J_{\text{av}} \nu_{\text{av}} \left(w^{\text{out}} + \tilde{W} \hat{\nu}_{\text{av}} \right)^2 \\ & \simeq \frac{n^K n_{\text{av}}^K n_{\text{av}}^J}{MN} J_{\text{av}} \nu_{\text{av}} \left(w^{\text{out}} + \tilde{W} \hat{\nu}_{\text{av}} \right)^2, \end{aligned} \quad (\text{B.17})$$

where $n_{\text{av}}^K = n^K/N$ and $n_{\text{av}}^J = n^J/N$ are the mean numbers per neuron of incoming external input connections and incoming recurrent connections, respectively.

B.4 Symmetry breaking by competition between input weights

We consider the equivalent of Eq. (B.7) for $\frac{dK_{ik}^\omega(t)}{dt} \frac{dK_{il}^\omega(t')}{dt}$ with two external inputs $k \neq l$ and recurrently connected neuron i . When k and l come from the same correlated input pool with correlation strength \hat{c} , an additional contribution for $t = t'$ to $Y_{i,k,i,l}(t, t')$ defined in Eq. (3.25) arises from the autocorrelation of inputs $k \neq l$, namely

$$\hat{c}\hat{v}_{\text{av}} \left(w^{\text{in}} + \tilde{W}v_{\text{av}} \right)^2. \quad (\text{B.18})$$

This contribution multiplied by the number of external input connections n^K is to be compared with the evaluation of the increase of the external input weight variance in Eq. (4.31), which “generates” the symmetry breaking. In order for the symmetry breaking to occur between different external input pools and not within the pools, it is necessary that the correlation strength \hat{c} be sufficiently large in order that the expression in Eq. (B.18) is comparable with that in Eq. (4.31), as shown by Gütig et al. (2003).

Appendix C

Calculations for chapter 5

C.1 Invertibility of $[\mathbf{1}_N - J(t)]$

In Chapter 3 and in Burkitt et al. (2007) we require that the matrix $[\mathbf{1}_N - J(t)]$ is invertible for all times t since the contrary would imply a divergence of the firing rates, cf. Eq. (3.22a). Actually, the possibility of diverging firing rates is related to the properties of our Poisson neuron model. This can be illustrated with a single neuron with spontaneous rate $v_0 > 0$ connected to itself by a scalar weight J . In this case, the synaptic input is $v_0 + Jv$ and the resulting firing rate v is determined by

$$v = \frac{v_0}{1 - J}. \quad (\text{C.1})$$

Provided $0 \leq J < 1$, the firing rate is finite and positive.

This constraint on the upper bound of J is relaxed if, instead of our version of the Poisson neuron model, we introduce an upper bound on the firing rate v . For example, we may use a sigmoidal-like function σ such that the firing rate is defined by the self-consistency relation

$$v = \sigma(v_0 + Jv). \quad (\text{C.2})$$

This gives a solution of v for any value of $J \geq 0$, as illustrated in Fig. C.1.

This can be extended to the case of several recurrently connected neurons, where J is a matrix. When σ is the identity function, the spectrum of J must then be within the unit circle. This condition on the spectrum relates to the expansion of $(\mathbf{1}_N - J)^{-1}$ in a power series, which is well defined for eigenvalues whose absolute values are strictly less than

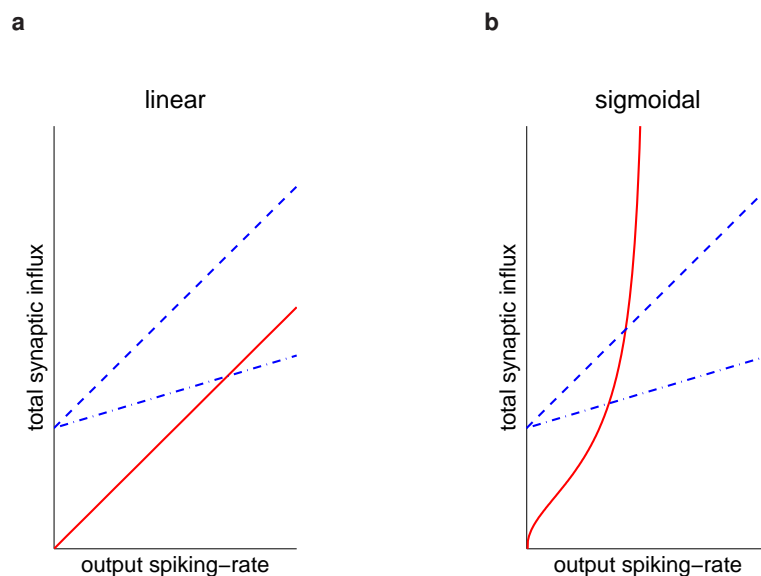


Figure C.1: Illustration of the impact of the activation function upon the weight constraint. This figure compares two Poisson neurons with (a) a linear and (b) a sigmoidal activation function σ . Each neuron is connected to itself with a scalar weight J . The dashed and the dashed-dotted lines correspond to different values of the weight J (resp. 1 and 0.3) in Eq. (C.2), while the solid lines correspond to the activation function. The intersection point determines the firing rate self-consistently constrained by the recurrent loop. For $J = 1$, the left plot has no solution, whereas the right plot does have a solution.

one.

A bounded activation function σ allows us to remove the upper bounds on the weights. However, the framework of this thesis exploits the linearity of the Poisson neuron model to make the analysis tractable. The qualitative behavior is expected to be the same for the Poisson neuron model with non-linear activation function in terms of equilibria and stability, provided the activation function σ is continuous, increasing and bounded. In this case, Eq. (C.2) always has a unique bounded solution ν for any given value of J .

In simulations, an explicit bound on the weights J was introduced (generally around 0.9 for the sum of incoming recurrent weights). This ensured that $[\mathbb{1}_N - J(t)]$ remained invertible at all times. Consequently, for certain parameter values, the simulations showed some discrepancies from the analytical predictions.

C.2 Equilibrium induced by STDP

This appendix contains a number of derivations whose results are discussed in Sec. 5.2.

C.2.1 Fixed point of the firing rates in the presence of recurrent loops

We define the function q as in Eq. (5.4)

$$q(x) = -\frac{w^{\text{in}}x}{w^{\text{out}} + \tilde{W}x}. \quad (\text{C.3})$$

For a synaptic loop of length n , the firing rate v_i of any neuron i within the loop satisfies $q^{\{n\}}(v_i) = v_i$ where $q^{\{n\}}$ denotes the n^{th} iteration of the self-composition of q , i.e., $q^{\{n\}} := q \circ \dots \circ q$.

For each $n \geq 0$, the function $q^{\{n\}}$ has a fractional form $ax/(b + cx)$ (proof by recurrence; a , b and c depend on n). Thus it has two fixed points at most, determined by the quadratic equation $ax - x(b + cx) = 0$. Since q has two fixed points, $q^{\{n\}}$ has the same two fixed points: 0 and $\mu := -(w^{\text{in}} + w^{\text{out}})/\tilde{W}$.

C.2.2 Stability of the manifold of fixed points

We study the spectrum of the endomorphism related to the first-order derivative of the learning equation around a given fixed point J^* , defined in Eq. (5.8). In the following analysis, we fix J^* and denote by \mathcal{L} the endomorphism that operates on matrices $X \in \mathbb{M}_J$

$$\mathcal{L}(X) = -\mu \Phi_J \left[w^{\text{in}} (\mathbb{1}_N - J^*)^{-1} X \mathbf{e} \mathbf{e}^T + w^{\text{out}} \mathbf{e} \mathbf{e}^T X^T (\mathbb{1}_N - J^*)^{-1T} \right]. \quad (\text{C.4})$$

Recall that \mathbb{M}_J is the space of $N \times N$ real matrices X such that $\Phi_J(X) = X$, i.e., matrices with non-zero elements only for indices (i, j) corresponding to an existing connection $j \rightarrow i$ in the network. The dimension of \mathbb{M}_J is equal to the number of recurrent connections n^J . \mathcal{L} has at least $n^J - N$ eigenmatrices related to the eigenvalue 0, since any matrix X such that $X\mathbf{e} = 0$ implies $\mathcal{L}(X) = 0$.

If the real parts of all eigenvalues in the spectrum of \mathcal{L} are negative (i.e., in left half of

the complex plane), then the fixed point J^* is stable. When all the J^* have negative real-part eigenvalues, the fixed-point manifold \mathcal{M}^* is attractive. Any J^* with one eigenvalue or more in the right half-plane will be unstable.

C.2.3 Decomposition of \mathcal{L}

We now study the N remaining eigenmatrices that do not correspond to the subspace $X\mathbf{e} = 0$. The columns of the matrix $(\mathbb{1}_N - J^*)^{-1}$ are denoted by the N -column vectors \mathbf{g}_i for $1 \leq i \leq N$, namely $\mathbf{g}_i = (\mathbb{1}_N - J^*)^{-1} \mathbf{x}_i$ with \mathbf{x}_i the i^{th} N -column vector of the canonical basis of \mathbb{R}^N (with all elements equal to zero except that on the i^{th} row, which is equal to one). We denote by A_{ij} the matrices of the canonical basis of \mathbb{M}_J with all elements equal to zero except the element on the i^{th} row and j^{th} column. For each index i , all the matrices $\mathcal{L}(A_{ij})$ are identical, since $A_{ij}\mathbf{e} = \mathbf{x}_i$; thus we fix an index $j_i = j(i)$ and one matrix $\check{A}_i = A_{ij_i} \in \mathbb{M}_J$. For a given i , the identical images $\mathcal{L}(A_{ij})$ can be expressed in terms of the $\check{A}_{i'}$ with $1 \leq i' \leq N$ and a matrix $Z(i) \in \mathbb{M}_J$ such that $Z(i)\mathbf{e} = 0$

$$\begin{aligned} \mathcal{L}(A_{ij}) &= \mathcal{L}(\check{A}_i) \\ &= -\mu \Phi_J \left[w^{\text{in}} \mathbf{g}_i \mathbf{e}^{\text{T}} + w^{\text{out}} \mathbf{e} \mathbf{g}_i^{\text{T}} \right] \\ &= -\mu \sum_{i'} \left\{ w^{\text{in}} \mathbf{x}_{i'}^{\text{T}} \Phi_J [\mathbf{g}_i \mathbf{e}^{\text{T}}] \mathbf{e} + w^{\text{out}} \mathbf{x}_{i'}^{\text{T}} \Phi_J [\mathbf{e} \mathbf{g}_i^{\text{T}}] \mathbf{e} \right\} \check{A}_{i'} + Z(i). \end{aligned} \quad (\text{C.5})$$

The matrix $Z(i)$ corresponds to the specific redistribution of the coefficients of $\mathcal{L}(A_{ij})$, where all the elements on each row i' are summed to form the coefficient of the element $\check{A}_{i'} = A_{i'j_{i'}}$ (for a matrix X , the corresponding sum is $\mathbf{x}_{i'}^{\text{T}} X \mathbf{e}$). This redistribution for each $\mathcal{L}(A_{ij}) = \mathcal{L}(\check{A}_i)$ only depends on i and not on j . In other words, we reduce the dimensionality of \mathbb{M}_J and work with classes of equivalent matrices $\Delta J \in \mathbb{M}_J$ that induce the first-order drift $\dot{\Delta} J \simeq \mathcal{L}(\Delta J)$, defined modulo the subspace $\{X \in \mathbb{M}_J, X\mathbf{e} = 0\}$.

Therefore, we can express the endomorphism \mathcal{L} in the basis of \mathbb{M}_J consisting of the N matrices \check{A}_i , and a linearly-independent family of $n_J - N$ matrices $X \in \mathbb{M}_J$ such that $X\mathbf{e} = 0$ to complete the basis

$$\mathcal{L} \sim \begin{pmatrix} L_r & 0 \\ L_Z & 0 \end{pmatrix}, \quad (\text{C.6})$$

where we assimilate \mathcal{L} with its matrix in the basis defined above. The $(n^J - N) \times N$ matrix L_Z is the expression of the $Z(i)$ of Eq. (C.5) in the subbase of $\{X \in \mathbb{M}_J, X\mathbf{e} = 0\}$. The $N \times N$ matrix L_r is given by

$$(L_r)_{ij} = -\mu \left(w^{\text{in}} n_i^J (\mathbf{g}_j)_i + w^{\text{out}} \sum_{i' \rightarrow i} (\mathbf{g}_j)_{i'} \right). \quad (\text{C.7})$$

The matrix element $(L_r)_{ij}$ corresponds to the expression of $L_r(\check{A}_i)$ in terms of the $\check{A}_{i'}$ in Eq. (C.5); note that i and j in Eq. (C.7) correspond to the indices i' and i resp. in Eq. (C.5). Note that $(\mathbf{g}_j)_i$ is the i^{th} element of the vector \mathbf{g}_j defined above, i.e., the element of $(\mathbb{1}_N - J^*)^{-1}$ for indices (i, j) . The sum $\sum_{i' \rightarrow i}$ is a sum over all i' such that there exists a connection from i' to i ; n_i^J is the number of incoming connections of neuron i . This decomposition allows us to study the non-zero spectrum of \mathcal{L} , which coincides with that of L_r according to Eq. (C.6), excluding the $n^J - N$ eigenvalues equal to zero. Using Eq. (C.7), we obtain Eq. (5.9), where R defined in Eq. (5.10) is the diagonal matrix with i^{th} element equal to n_i^J . Note that for the case of full connectivity except for self-connections, this links to the analysis by Burkitt et al. (2007).

C.2.4 Homogeneous connectivity topology

The matrix R in Eq. (5.10) can be approximated by $n_{\text{av}}^J \mathbb{1}_N$ in the case of random connectivity with roughly the same number $n_{\text{av}}^J = n^J / N$ of incoming connections per neuron. It follows that $L_{\text{in}} \simeq -\mu n_{\text{av}}^J (\mathbb{1}_N - J^*)^{-1}$. The spectrum of J^* is assumed to be in the unit circle at all times (cf. Appendix C.1), which means that the spectrum of $(\mathbb{1}_N - J^*)^{-1}$ lies in the right half of the complex plane. Since $\mu > 0$, the spectrum of L_{in} (crosses in Fig. 5.2(a)) is in the left half-plane, i.e., its eigenvalues have negative real parts. The spectrum of L_{out} (circles in Fig. 5.2(b)) contains $N - 1$ eigenvalues roughly equal to zero due to the presence of $\Phi_J[\mathbf{e} \mathbf{e}^T]$ (it is strictly zero for full connectivity except for self-connections) and one non-zero eigenvalue related to the eigenvector \mathbf{e} given by

$$-\frac{\mathbf{e}^T L_{\text{out}} \mathbf{e}}{N} \simeq -\frac{n_{\text{av}}^J \mu^2}{v_0} < 0, \quad (\text{C.8})$$

which also lies in left half-plane. We have used the approximation $\Phi_J[\mathbf{e}\mathbf{e}^T]\mathbf{e} \simeq n_{\text{av}}^J \mathbf{e}$.

The discussion about the spectrum L_r depending on the values of w^{in} and w^{out} is detailed in Sec. 5.2.3. For $w^{\text{out}} > 0$ and $w^{\text{in}} > 0$, we expect the spectrum to remain in the left half-plane, contained within the convex hull of the spectra of L_{in} and L_{out} expanded by the scale factor $w^{\text{in}} + w^{\text{out}}$. The conclusions on the stability are the same for all fixed points J^* , and hence they determine whether the whole fixed-point manifold \mathcal{M}^* is attractive or not. Denser recurrent connectivity also gives larger positive values of n_{av}^J in $R \simeq n_{\text{av}}^J \mathbf{1}_N$ and in Eq. (C.8). This implies stronger stability of the fixed points J^* when the conditions on w^{in} and w^{out} are met.

C.3 Second order of the stochastic evolution of the weights

In this appendix, we provide details of calculations useful to evaluate the structural evolution of the recurrent weights due to STDP, which occurs after the fast convergence towards the homeostatic equilibrium described in Sec. 5.2. The weight dispersion can be related to the second moment of the stochastic evolution of the weight matrix J , through the multidimensional matrix $\Gamma(t, t')$ whose elements are defined in Eq. (5.12). We show how the connectivity is involved in the evaluation of this matrix, due to the autocorrelation of the neuron activity.

C.3.1 Analysis of the matrix $\Gamma(t, t')$

The trace of this matrix was used in order to evaluate the linear increase of the weight variance due to STDP near the beginning of the learning epoch for $t = t'$ and zero recurrent delays $d_{ij} = 0$ (Burkitt et al. 2007). The variance is the expectation value of the trace of the matrix product involving the derivative of J ,

$$\begin{aligned} \text{Var}(J)(t) &= \left\langle \frac{1}{nJ-1} \sum_{j \rightarrow i} [J_{ij}(t) - J_{\text{av}}(t)]^2 \right\rangle \\ &= \frac{1}{nJ-1} \left\langle \text{trace} \left\{ [J(t) - J_{\text{av}}(t)\Phi_J(\mathbf{e}\mathbf{e}^T)] [J(t) - J_{\text{av}}(t)\Phi_J(\mathbf{e}\mathbf{e}^T)]^T \right\} \right\rangle, \end{aligned} \quad (\text{C.9})$$

where $\sum_{j \rightarrow i}$ is the sum over the existing connections. When the network is at the homeostatic equilibrium, the mean weight over the network (considered “deterministic”) satisfies $J_{av}(t) = \text{const}$. It follows that the growth rate of the weight variance is given by

$$\frac{d\text{Var}(J)}{dt}(t) = \frac{2}{(n^J - 1)t} \int_0^t \text{trace} \left\langle \frac{dJ^\omega(t)}{dt} \left[\frac{dJ^\omega(t')}{dt} \right]^T \right\rangle dt', \quad (\text{C.10})$$

where $\frac{dJ^\omega(t)}{dt}$ denotes here the derivative of the weight matrix J for one stochastic trajectory; it is different from the drift $\dot{J}(t)$ (expectation value). Note that before the homeostatic equilibrium is reached, the variance will evolve both due to deterministic and stochastic contributions depending on the initial value of the variance if the weights are not homogeneous at the beginning of the learning. The stochastic part can then be evaluated using $\Gamma(t, t') - \dot{J}(t) \dot{J}^T(t')$ instead of $\Gamma(t, t')$ alone in Eq. (C.10).

The non-diagonal elements of $\Gamma(t, t')$ can also be related to the stochastic dispersion of the weights J . The sign of $\Gamma_{i,j,i',j}(t, t')$ indicates whether the two incoming weights J_{ij} and $J_{i'j}$ of neurons i and i' evolve in the same direction (potentiation or depression): when positive, they tend to both either increase together or decrease together. Sets of weights for which $\sum \Gamma_{i,j,i',j}(t, t')$ (synaptic connections involving indices $i \neq i'$ and j) are more positive will exhibit a smaller dispersion. However, these terms do not directly relate to the generation of the increasing variance described by Eq. (C.9) (Burkitt et al. 2007).

C.3.2 Autocorrelation effects on weight dispersion

We consider now the situation of network evolution at the equilibrium, i.e., the weight matrix $J(t)$ is on the manifold of fixed points \mathcal{M}^* at all times without reaching the bounds and its drift $\dot{J}(t) = \left\langle \frac{dJ^\omega(t)}{dt} \right\rangle = 0$ with $\nu(t) = \mu \mathbf{e}$. We want to evaluate the impact of the recurrent connectivity on the evolution of $\Gamma_{i,j,i',j}(t, t')$.

Impact of $J_{i'i'}$ on $\Gamma_{i,j,i',j}(t, t')$

Here we evaluate the effect of the presence of a single recurrent connection $i' \rightarrow i$ at the first order of the recurrence, by naively deriving the spike-triggering effects related

to $J_{ii'}$ in $\Gamma_{i,j,i',j}(t, t')$. We use similar calculations to those in Appendix B.3 to evaluate the common evolution of input weights. Using Eq. (3.6) to express the variation of the weights J_{ij} and $J_{i'j}$ for one stochastic trajectory, we obtain

$$\frac{dJ_{ij}^{\circ}(t)}{dt} \frac{dJ_{i'j}^{\circ}(t')}{dt'} = \left[w^{\text{in}} S_j(t-d) + w^{\text{out}} S_i(t) + \int W(u) S_i(t) S_j(t+u-d) du \right] \quad (\text{C.11})$$

$$\left[w^{\text{in}} S_j(t'-d) + w^{\text{out}} S_{i'}(t') + \int W(u') S_{i'}(t') S_j(t'+u'-d) du' \right],$$

where we assumed that all the recurrent delays are equal to d . Four terms induced by spike-triggering effects related to $J_{ii'}$ arise from Eq. (C.11) and contribute to $\Gamma_{i,j,i',j}(t, t')$ when taking the ensemble average on Eq. (C.11).

First, $(w^{\text{out}})^2 S_i(t) S_{i'}(t')$ involves $J_{ii'}$ through the dependence of $S_i(t)$ on the past synaptic input history of $S_{i'}(t)$, according to Eq. (2.1), namely for index $j' = i'$,

$$\rho_i(t) = v_0 + \sum_{j'} J_{ij'} \int \epsilon(r) S_{j'}(t-r-d) dr. \quad (\text{C.12})$$

This leads to an extra contribution due to the autocorrelation of $S_{i'}$ for $j' = i'$ and $t-r-d = t'$,

$$J_{ii'} \epsilon(t-t'-d) \langle S_{i'}(t') \rangle. \quad (\text{C.13})$$

Note that this expression is *a priori* valid only for $t' < t$, but it actually holds in general since $\epsilon(t-t'-d) = 0$ for $t' \geq t$. Second, $[w^{\text{out}} S_i(t)] [\int W(u') S_{i'}(t') S_j(t'+u'-d) du']$ gives

$$w^{\text{out}} J_{ii'} \epsilon(t-t'-d) \int W(u') \langle S_{i'}(t') S_j(t'+u'-d) \rangle du' \quad (\text{C.14})$$

$$\simeq w^{\text{out}} J_{ii'} \epsilon(t-t'-d) \langle S_{i'}(t') \rangle \int W(u') \langle S_j(t'+u'-d) \rangle du',$$

where the spike trains $S_{i'}$ and S_j are taken to be independent (we only evaluate the leading order here). Third, the term $[\int W(u) S_i(t) S_j(t+u-d) du] [w^{\text{out}} S_{i'}(t')]$ gives

$$w^{\text{out}} J_{ii'} \epsilon(t-t'-d) \langle S_{i'}(t') \rangle \int W(u) \langle S_j(t+u-d) \rangle du. \quad (\text{C.15})$$

Fourth and last, $[\int W(u)S_i(t)S_j(t+u-d)du][\int W(u')S_{i'}(t')S_j(t'+u'-d)du']$, where the function W is involved twice, gives

$$J_{i'i'}\epsilon(t-t'-d)\langle S_{i'}(t') \rangle \int \int W(u)W(u')\langle S_j(t+u-d) \rangle \langle S_j(t'+u'-d) \rangle du du'. \quad (\text{C.16})$$

Summing the terms in Eqs. (C.13), (C.14), (C.15) and (C.16), we obtain the total contribution to $\Gamma_{i,j,i',j}(t,t')$ due to the single weight $J_{i'i'}$, at the leading order:

$$J_{i'i'}\epsilon(t-t'-d)\langle S_{i'}(t') \rangle \left[w^{\text{out}} + \int W(u)\langle S_j(t+u-d) \rangle du \right]^2. \quad (\text{C.17})$$

The coefficient of $J_{i'i'}$ in Eq. (C.17) is positive, which tends to cause the weights J_{ij} and $J_{i'j}$ to evolve in the same direction, either potentiation or depression, cf. Appendix C.3.1.

When the network is at the equilibrium, $v_i(t) = \mu$ for each neuron i and the time-averaged contribution to the weight coupling $\Gamma_{i,j,i',j}(t,t')$ due to $J_{i'i'}$ given in Eq. (C.17) becomes

$$J_{i'i'}\mu \left(w^{\text{out}} + \tilde{W}\mu \right)^2 = J_{i'i'}\mu (w^{\text{in}})^2, \quad (\text{C.18})$$

where we have used the normalization of the PSP kernel function ($\int \epsilon = 1$) and the definition of μ in Eq. (5.6).

Impact of J_{ji} on $\Gamma_{i,j,i',j}(t,t')$

Similar to the calculation above, we now evaluate the effect of J_{ji} on $\Gamma_{i,j,i',j}(t,t')$ using Eq. (C.12) and examine the spike-triggering effect due to the autocorrelation of S_i . We find the equivalent to Eq. (C.18) for the time-averaged contribution at the equilibrium,

$$J_{ji}\mu \left(w^{\text{out}} + \tilde{W}\mu \right) \left(w^{\text{in}} + \tilde{W}\mu \right) = J_{ji}\mu w^{\text{in}}w^{\text{out}}, \quad (\text{C.19})$$

where we have used Eq. (5.6). The sign of the coefficient of J_{ji} can either be positive or negative here. For example, our choice of parameters corresponds to $w^{\text{in}} > 0$ and $w^{\text{out}} < 0$ (cf. Appendix D), so the contribution in Eq. (C.19) is negative in this case.

Link to the density of recurrent connections

It follows from this analysis that the stronger the recurrent connections are in a neuron group, the more its weights tend to evolve together. Weakly connected sets of weights are more likely to exhibit individual weights that evolve in different directions (potentiation vs. depression). For homogeneous recurrent connectivity, where each connection has the probability $n^J/N(N-1)$ of existing (n^J is the number incoming recurrent connections), the lumped effect for the whole network corresponds to the sum of the terms in Eqs. (C.18) and (C.19) for the two possible connections $i' \rightarrow i$ and $i \rightarrow j$, and for all triplets (i, i', j) when the connections $j \rightarrow i$ and $j \rightarrow i'$ exist:

$$\begin{aligned} \frac{1}{t} \int \sum_{j \rightarrow i} \sum_{j \rightarrow i'} \Gamma_{i,j,i',j}(t, t') dt' &\simeq \frac{N(N-1)(N-2)(n^J)^3}{[N(N-1)]^3} J_{\text{av}} \mu w^{\text{in}} (w^{\text{in}} + w^{\text{out}}) \quad (\text{C.20}) \\ &\simeq -(n_{\text{av}}^J)^3 J_{\text{av}} \frac{w^{\text{in}} (w^{\text{in}} + w^{\text{out}})^2}{\tilde{W}}, \end{aligned}$$

where we have used the definition of μ in Eq. (5.6) and taken the limit of a large network ($N \gg 1$ neurons). Recall that $n_{\text{av}}^J = n^J/N$ is the mean number of incoming recurrent connections per neuron. Note that the triplets are ordered so that the triplet (i, i', j) accounts for the connections $j \rightarrow i$ and $j \rightarrow i'$; the connections $i \rightarrow i'$ and $i' \rightarrow j$ are taken into account by the triplet (i', i, j) . The overall effect is positive provided w^{in} and \tilde{W} have opposite signs, which is the case where the fixed-point manifold \mathcal{M}^* of the weights J is attractive (cf. Sec. 5.2.3 and 5.2.2): $w^{\text{in}} > 0$ and $\tilde{W} < 0$.

The contributions due to the recurrent connections at the first order of the recurrence are captured by the spike-triggering effects in Eq. (C.20). Similar to the expansion $(\mathbb{1}_N - J)^{-1} = \sum_n J^n$, it is possible to rigorously incorporate higher orders of autocorrelation induced by these spike-triggering effects. Since the network contains only positive weights, all of these effects are positive and accumulate. The higher order terms decay exponentially and consequently do not substantially change the result obtained in Eq. (C.20). This can be illustrated for a scalar J such that $1 - J < 1$ with a “safety” margin (J is not too close to 1), where $J/(1 - J)$ and J are of the same order.

C.3.3 Weight evolution for a synaptic loop $j \rightarrow i \rightarrow j$

Now we evaluate the effect of STDP on a given synaptic loop of length two between neurons i and j via the evolution of $\Gamma_{i,j,j,i}(t, t')$ defined in a similar manner to Eq. (5.12) with different indices. Similar to Eq. (C.11), we use Eq. (3.6) in order to express the relative evolution of the weights J_{ij} and J_{ji} for one stochastic trajectory, which relates to

$$\frac{dJ_{ij}^\omega(t)}{dt} \frac{dJ_{ji}^\omega(t')}{dt'} = \left[w^{\text{in}} S_j(t-d) + w^{\text{out}} S_i(t) + \int W(u) S_i(t) S_j(t+u-d) du \right] \quad (\text{C.21})$$

$$\left[w^{\text{in}} S_i(t'-d) + w^{\text{out}} S_j(t') + \int W(u') S_j(t') S_i(t'+u'-d) du' \right].$$

We consider the network to be at the homeostatic equilibrium in order to evaluate the effects due to the autocorrelation of the neurons, $v_i = \mu = -(w^{\text{in}} + w^{\text{out}})/\tilde{W}$ for all i , cf. Eq. (5.6). In this case, the leading order of the terms that arise is negative, independent of the learning parameters,

$$2 \left[w^{\text{in}} w^{\text{out}} \mu + (w^{\text{in}} + w^{\text{out}}) \tilde{W} \mu^2 \right] = -2\mu \left[(w^{\text{in}})^2 + (w^{\text{out}})^2 + w^{\text{in}} w^{\text{out}} \right] < 0, \quad (\text{C.22})$$

since the polynomial in x of the second order $x^2 + ax + a^2$ is always positive for any value of the coefficient a . Note that we did not use the Poisson neuron model here.

C.4 Dependence of the asymptotic weight distribution on initial conditions

We consider a specific example of evolution of the weights J with full connectivity except for self-connections, so that in this case Φ_J only nullifies the diagonal terms of its matrix argument. The sums of the outgoing weights for each neuron are given by the elements of the row vector $\mathbf{e}^T J$, which according to Eq. (3.22b) is

$$\mathbf{e}^T J = w^{\text{in}} \mathbf{e}^T \mathbf{e} \mathbf{v}^T + w^{\text{out}} \mathbf{e}^T \mathbf{v} \mathbf{e}^T + \tilde{W} \mathbf{e}^T \mathbf{v} \mathbf{v}^T - (w^{\text{in}} + w^{\text{out}}) \mathbf{v}^T - \tilde{W} \mathbf{v}^T \text{diag}(\mathbf{v}). \quad (\text{C.23})$$

We consider initial conditions for which the sums on the incoming weights are identical for each neuron, but the sums of the outgoing weights are inhomogeneous: $J\mathbf{e} \propto \mathbf{e}$ but $\mathbf{e}^T J$ is not proportional to \mathbf{e}^T . This implies homogeneous firing rates, i.e., $\boldsymbol{\nu} \propto \mathbf{e}$, since $\dot{J}\mathbf{e} \propto \mathbf{e}$ at all times using a similar equation to that above. Then $\mathbf{e}^T \dot{J}$ reduces to

$$\mathbf{e}^T \dot{J} = (N-1)v_{\text{av}} \left(w^{\text{in}} + w^{\text{out}} + \tilde{W}v_{\text{av}} \right) \mathbf{e}^T.$$

Consequently, the sums of the outgoing weights (i.e., on each column of the matrix J) will evolve identically; hence, the initial discrepancies will remain after the learning stabilizes, when $v_{\text{av}} = \mu$.

This example illustrates that STDP does not reorganize the sums of the outgoing weights for each neuron, as it does for the sums of incoming weights in order to obtain homogeneous neuron firing rates. Likewise, for an initial inhomogeneous vector of firing rates $\boldsymbol{\nu}$, $\mathbf{e}^T J$ will be modified until $\boldsymbol{\nu}$ converges to $\mu\mathbf{e}$, which may cause inhomogeneities to develop even if initially $\mathbf{e}^T J$ is homogeneous. As a result, the asymptotic value of $\mathbf{e}^T J$ is not constrained by STDP and this evolution does not relate to learning *per se*. A similar conclusion can be drawn for the case of partial connectivity.

Appendix D

Simulation parameters

The results in Chapters 4 and 5 were obtained using discrete-time numerical simulation and the parameters listed in Table D.1, unless stated otherwise. At each time step, the probability of firing for each source and neurons is computed depending on the past spiking history and the new spikes are determined by random draws; then the weights are modified accordingly.

The additive STDP window function W is given by

$$W(u) = \begin{cases} c_P \exp\left(\frac{u}{\tau_P}\right) & \text{for } u < 0 \\ -c_D \exp\left(-\frac{u}{\tau_D}\right) & \text{for } u > 0. \end{cases} \quad (\text{D.1})$$

In Chapter 6, weight-dependent STDP with alpha functions was used:

$$\begin{aligned} W_+(u) &= 2c_P \frac{u}{\tau_P/2} \exp\left(\frac{u}{\tau_P/2}\right) \quad \text{for } u < 0, \\ W_-(u) &= -2c_D \exp\left(-\frac{u}{\tau_D}\right) \quad \text{for } u > 0. \end{aligned} \quad (\text{D.2})$$

The constant 2 is used to obtain a decaying profile similar to that of additive STDP.

The PSP kernel ϵ is defined by

$$\epsilon(t) = \begin{cases} \frac{\exp(t/\tau_B) - \exp(t/\tau_A)}{\tau_B - \tau_A} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0. \end{cases} \quad (\text{D.3})$$

The synaptic weights are not normalized, but defined such that the sum of the incoming weights for each neuron is of the order of one. This implies that the effective rate of

Table D.1: Table of simulation parameters

time step	10^{-4} s
simulation duration	10^5 s
Input Poisson spike trains	
firing rates	$\hat{\nu}_{av} = 30 - 35$ Hz
correlation strength	$\hat{c}_{av} = 0 - 0.2$
Poisson neurons	
spontaneous firing rate	$\nu_0 = 5$ Hz
Synapses	
rise time constant	$\tau_A = 1$ ms
decay time constant	$\tau_B = 5$ ms
mean of recurrent delays	$d = 0.4 \pm 0.2$ ms
mean of input delays	$\hat{d} = 7 \pm 1$ ms
STDP	
learning parameter	$\eta = 10^{-5} - 5 \times 10^{-7}$
pre-synaptic rate-based coeff.	$w^{in} = 4$
post-synaptic rate-based coeff.	$w^{out} = -0.5$
potentiation time constant	$\tau_P = 17$ ms
potentiation scaling coefficient	$c_P = 15$
depression time constant	$\tau_D = 34$ ms
depression scaling coefficient	$c_D = 10$

change per second for the weights is at least two orders of magnitude (10^{-2}) below their upper bound. These parameters are in the same range as those used in previous studies Kempter et al. (1999) and Burkitt et al. (2007).

Bibliography

- Amari, S. I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics* 27(2), 77–87.
- Amit, D. J. and N. Brunel (1997). Dynamics of a recurrent network of spiking neurons before and following learning. *Network-Computation in Neural Systems* 8(4), 373–404.
- Appleby, P. A. and T. Elliott (2005). Synaptic and temporal ensemble interpretation of spike-timing-dependent plasticity. *Neural Computation* 17(11), 2316–2336.
- Appleby, P. A. and T. Elliott (2006). Stable competitive dynamics emerge from multi-spike interactions in a stochastic model of spike-timing-dependent plasticity. *Neural Computation* 18(10), 2414–2464.
- Appleby, P. A. and T. Elliott (2007). Multispikes interactions in a stochastic model of spike-timing-dependent plasticity. *Neural Computation* 19(5), 1362–1399.
- Artola, A., S. Brocher, and W. Singer (1990). Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual-cortex. *Nature* 347(6288), 69–72.
- Badoual, M., Q. Zou, A. P. Davison, M. Rudolph, T. Bal, Y. Fregnac, and A. Destexhe (2006). Biophysical and phenomenological models of multiple spike interactions in spike-timing dependent plasticity. *International Journal of Neural Systems* 16(2), 79–97.
- Bear, M. F., B. W. Connors, and M. A. Paradiso (2007). *Neuroscience: exploring the brain* (3rd ed. ed.). Philadelphia: Lippincott Williams & Wilkins.
- Bell, C. C., V. Z. Han, Y. Sugawara, and K. Grant (1997). Synaptic plasticity in a

- cerebellum-like structure depends on temporal order. *Nature* 387(6630), 278–281.
- Benuskova, L. and W. C. Abraham (2007). STDP rule endowed with the BCM sliding threshold accounts for hippocampal heterosynaptic plasticity. *Journal of Computational Neuroscience* 22(2), 129–133.
- Beurle, R. L. (1956). Properties of a mass of cells capable of regenerating pulses. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences* 240(669), 55–87.
- Bi, G. Q. and M. M. Poo (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience* 18(24), 10464–10472.
- Bi, G. Q. and M. M. Poo (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience* 24, 139–166.
- Bienenstock, E. L., L. N. Cooper, and P. W. Munro (1982). Theory for the development of neuron selectivity - orientation specificity and binocular interaction in visual-cortex. *Journal of Neuroscience* 2(1), 32–48.
- Bliss, T. V. P. and T. Lømo (1973). Long-lasting potentiation of synaptic transmission in dentate area of anesthetized rabbit following stimulation of perforant path. *Journal of Physiology - London* 232(2), 331–356.
- Boettiger, C. A. and A. J. Doupe (2001). Developmentally restricted synaptic plasticity in a songbird nucleus required for song learning. *Neuron* 31(5), 809–818.
- Bohte, S. M. and M. C. Mozer (2007). Reducing the variability of neural responses: A computational theory of spike-timing-dependent plasticity. *Neural Computation* 19(2), 371–403.
- Borovkov, A. A. (1998). *Ergodicity and stability of stochastic processes*. New York: J. Wiley.
- Bremaud, P. and L. Massoulié (1996). Stability of nonlinear Hawkes processes. *Annals of Probability* 24(3), 1563–1588.
- Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience* 8(3), 183–208.

- Brunel, N. and V. Hakim (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation* 11(7), 1621–1671.
- Burkitt, A. N. (2001). Balanced neurons: analysis of leaky integrate-and-fire neurons with reversal potentials. *Biological Cybernetics* 85(4), 247–255.
- Burkitt, A. N. (2006). A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological Cybernetics* 95(1), 1–19.
- Burkitt, A. N., M. Gilson, and J. L. van Hemmen (2007). Spike-timing-dependent plasticity for neurons with recurrent connections. *Biological Cybernetics* 96(5), 533–546.
- Burkitt, A. N., H. Meffin, and D. B. Grayden (2003). Gain modulation and balanced synaptic input in a conductance-based neural model. *Neurocomputing* 52-4, 221–226.
- Burkitt, A. N., H. Meffin, and D. B. Grayden (2004). Spike-timing-dependent plasticity: The relationship to rate-based learning for models with weight dynamics determined by a stable fixed point. *Neural Computation* 16(5), 885–940.
- Caporale, N. and Y. Dan (2008). Spike timing-dependent plasticity: A Hebbian learning rule. *Annual Review of Neuroscience* 31, 25–46.
- Carnell, A. (2009). An analysis of the use of Hebbian and anti-Hebbian spike time dependent plasticity learning functions within the context of recurrent spiking neural networks. *Neurocomputing* 72(4-6), 685–692.
- Carrillo-Reid, L., F. Tecuapetla, O. Ibanez-Sandoval, A. Hernandez-Cruz, E. Galarraga, and J. Vargas (2009). Activation of the cholinergic system endows compositional properties to striatal cell assemblies. *Journal of Neurophysiology* 101(2), 737–749.
- Câteau, H., K. Kitano, and T. Fukai (2008). Interplay between a phase response curve and spike-timing-dependent plasticity leading to wireless clustering. *Physical Review E* 77(5), 051909.
- Choe, Y. and R. Miikkulainen (1998). Self-organization and segmentation in a laterally connected orientation map of spiking neurons. *Neurocomputing* 21(1-3), 139–157.
- Colbran, R. J. (2004). Protein phosphatases and calcium/calmodulin-dependent pro-

- tein kinase ii-dependent synaptic plasticity. *Journal of Neuroscience* 24(39), 8404–8409.
- Coombes, S. (2005). Waves, bumps, and patterns in neural field theories. *Biological Cybernetics* 93(2), 91–108.
- Dahmen, J. C., D. E. Hartley, and A. J. King (2008). Stimulus-timing-dependent plasticity of cortical frequency representation. *Journal of Neuroscience* 28(50), 13629–39.
- Dan, Y. and M. M. Poo (2006). Spike timing-dependent plasticity: From synapse to perception. *Physiological Reviews* 86(3), 1033–1048.
- Davis, M. H. A. (1984). Piecewise-deterministic Markov-processes – a general-class of non-diffusion stochastic-models. *Journal of the Royal Statistical Society Series B - Methodological* 46(3), 353–388.
- Debanne, D., B. H. Gähwiler, and S. M. Thompson (1998). Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures. *Journal of Physiology - London* 507(1), 237–247.
- Delorme, A., L. Perrinet, and S. J. Thorpe (2001). Networks of integrate-and-fire neurons using rank order coding B: Spike timing dependent plasticity and emergence of orientation selectivity. *Neurocomputing* 38, 539–545.
- Doob, J. L. (1953). *Stochastic processes* New York, Wiley.
- Drew, P. J. and L. F. Abbott (2006). Extending the effects of spike-timing-dependent plasticity to behavioral timescales. *Proceedings of the National Academy of Sciences of the United States of America* 103(23), 8876–8881.
- Duguid, I. and P. J. Sjöström (2006). Novel presynaptic mechanisms for coincidence detection in synaptic plasticity. *Current Opinion in Neurobiology* 16(3), 312–322.
- Egger, V., D. Feldmeyer, and B. Sakmann (1999). Coincidence detection and changes of synaptic efficacy in spiny stellate neurons in rat barrel cortex. *Nature Neuroscience* 2(12), 1098–1105.
- Elliott, T. (2008). Temporal dynamics of rate-based synaptic plasticity rules in a stochastic model of spike-timing-dependent plasticity. *Neural Computation* 20(9),

2253–2307.

- Elliott, T. and N. R. Shadbolt (1999). A neurotrophic model of the development of the retinogeniculocortical pathway induced by spontaneous retinal waves. *Journal of Neuroscience* 19(18), 7951–7970.
- Ermentrout, G. B., R. F. Galan, and N. N. Urban (2008). Reliability, synchrony and noise. *Trends in Neurosciences* 31(8), 428–434.
- Feldman, D. E. (2000). Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron* 27(1), 45–56.
- Froemke, R. C. and Y. Dan (2002). Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature* 416(6879), 433–438.
- Froemke, R. C., M. M. Poo, and Y. Dan (2005). Spike-timing-dependent synaptic plasticity depends on dendritic location. *Nature* 434(7030), 221–225.
- Fusi, S. (2002). Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. *Biological Cybernetics* 87(5-6), 459–470.
- Gerstein, G. L. and B. Mandelbrot (1964). Random walk models for spike activity of single neuron. *Biophysical Journal* 4(1P1), 41–68.
- Gerstner, W., R. Kempter, J. L. van Hemmen, and H. Wagner (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature* 383(6595), 76–78.
- Gerstner, W. and W. M. Kistler (2002). *Spiking neuron models: single neurons, populations, plasticity*. New York: Cambridge University Press.
- Goodhill, G. J. (2007). Contributions of theoretical modeling to the understanding of neural map development. *Neuron* 56(2), 301–311.
- Graupner, M. and N. Brunel (2007). STDP in a bistable synapse model based on CaMKII and associated signaling pathways. *PLoS Computational Biology* 3(11), 2299–2323.
- Gütig, R., R. Aharonov, S. Rotter, and H. Sompolinsky (2003). Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *Journal of Neuroscience* 23(9), 3697–3714.

- Gutkin, B., G. B. Ermentrout, and M. Rudolph (2003). Spike generating dynamics and the conditions for spike-time precision in cortical neurons. *Journal of Computational Neuroscience* 15(1), 91–103.
- Han, V. Z., K. Grant, and C. C. Bell (2000). Reversible associative depression and nonassociative potentiation at a parallel fiber synapse. *Neuron* 27(3), 611–622.
- Hartley, M., N. Taylor, and J. Taylor (2006). Understanding spike-time-dependent plasticity: A biologically motivated computational model. *Neurocomputing* 69(16-18), 2005–2016.
- Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society Series B - Statistical Methodology* 33(3), 438–443.
- Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*. New York: Wiley.
- van Hemmen, J. L. (2001). *Handbook of biological physics, Vol. 4: Neuro-informatics and neural modelling, Theory of synaptic plasticity*, pp. 771–823. Amsterdam: Elsevier.
- Hirsch, J. C., G. Barrionuevo, and F. Crepel (1992). Homosynaptic and heterosynaptic changes in efficacy are expressed in prefrontal neurons - an in-vitro study in the rat. *Synapse* 12(1), 82–85.
- Hodgkin, A. L. and A. F. Huxley (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology - London* 117(4), 500–544.
- Holmgren, C. D. and Y. Zilberter (2001). Coincident spiking activity induces long-term changes in inhibition of neocortical pyramidal cells. *Journal of Neuroscience* 21(20), 8270–8277.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America - Biological Sciences* 79(8), 2554–2558.
- Hubel, D. H. and T. N. Wiesel (1962). Receptive fields, binocular interaction and functional architecture in cats visual cortex. *Journal of Physiology - London* 160(1), 106–

164.

- Iglesias, J., J. Eriksson, F. Grize, M. Tomassini, and A. E. P. Villa (2005). Dynamics of pruning in simulated large-scale spiking neural networks. *Biosystems* 79(1-3), 11–20.
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks* 14(6), 1569–1572.
- Izhikevich, E. M. and N. S. Desai (2003). Relating STDP to BCM. *Neural Computation* 15(7), 1511–1523.
- Izhikevich, E. M., J. A. Gally, and G. M. Edelman (2004). Spike-timing dynamics of neuronal groups. *Cerebral Cortex* 14(8), 933–944.
- Jacobsen, M. (2006). *Point process theory and applications: marked point and piecewise deterministic processes*. Boston: Birkhäuser.
- Jolivet, R., R. Kobayashi, A. Rauch, R. Naud, S. Shinomoto, and W. Gerstner (2008). A benchmark test for a quantitative assessment of simple neuron models. *Journal of Neuroscience Methods* 169(2), 417–424.
- Kang, S., K. Kitano, and T. Fukai (2008). Structure of spontaneous UP and DOWN transitions self-organizing in a cortical network model. *PLoS Computational Biology* 4(3), e1000022.
- Karbowski, J. and G. B. Ermentrout (2002). Synchrony arising from a balanced synaptic plasticity in a network of heterogeneous neural oscillators. *Physical Review E* 65(3), 031902.
- Kempter, R., W. Gerstner, and J. L. van Hemmen (1999). Hebbian learning and spiking neurons. *Physical Review E* 59(4), 4498–4514.
- Kempter, R., W. Gerstner, and J. L. van Hemmen (2001). Intrinsic stabilization of output rates by spike-based Hebbian learning. *Neural Computation* 13(12), 2709–2741.
- Kempter, R., W. Gerstner, J. L. van Hemmen, and H. Wagner (1998). Extracting oscillations: Neuronal coincidence detection with noisy periodic spike input. *Neural Computation* 10(8), 1987–2017.

- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43(1), 59–69.
- Kriener, B., T. Tetzlaff, A. Aertsen, M. Diesmann, and S. Rotter (2008). Correlations and population dynamics in cortical networks. *Neural Computation* 20(9), 2185–2226.
- Lansky, P. (1984). On approximations of steins neuronal model. *Journal of Theoretical Biology* 107(4), 631–647.
- Lapicque, L. (1907). Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *Journal de Physiologie et de Pathologie Générale* 9, 620–635.
- Leibold, C., R. Kempter, and J. L. van Hemmen (2002). How spiking neurons give rise to a temporal-feature map: From synaptic plasticity to axonal selection. *Physical Review E* 65(5), 051915.
- Levy, W. B. and O. Steward (1983). Temporal contiguity requirements for long-term associative potentiation depression in the hippocampus. *Neuroscience* 8(4), 791–797.
- Lisman, J. (1989). A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. *Proceedings of the National Academy of Sciences of the United States of America* 86(23), 9574–9578.
- Liu, Q. S., L. Pu, and M. M. Poo (2005). Repeated cocaine exposure in vivo facilitates LTP induction in midbrain dopamine neurons. *Nature* 437(7061), 1027–1031.
- Lubenov, E. V. and A. G. Siapas (2008). Decoupling through synchrony in neuronal circuits with propagation delays. *Neuron* 58(1), 118–131.
- Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models. *Neural Networks* 10(9), 1659–1671.
- Maass, W., T. Natschläger, and H. Markram (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation* 14(11), 2531–2560.
- Magee, J. C. and D. Johnston (1997). A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science* 275(5297), 209–213.

- Malenka, R. C. and S. A. Siegelbaum (2001). *Synapses*, Synaptic plasticity: diverse targets and mechanisms for regulating synaptic efficacy, pp. 393–454. Baltimore, The John Hopkins University.
- von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in striate cortex. *Kybernetik* 14(2), 85–100.
- Markram, H., J. Lubke, M. Frotscher, A. Roth, and B. Sakmann (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *Journal of Physiology - London* 500(2), 409–440.
- Martin, S. J., P. D. Grimwood, and R. G. M. Morris (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annual Review of Neuroscience* 23, 649–711.
- Masquelier, T., R. Guyonneau, and S. J. Thorpe (2008). Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PLoS ONE* 3(1), e1377.
- Massoulié, L. (1998). Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic Processes and Their Applications* 75(1), 1–30.
- Masuda, N. and H. Kori (2007). Formation of feedforward networks and frequency synchrony by spike-timing-dependent plasticity. *Journal of Computational Neuroscience* 22(3), 327–345.
- Mayer, M. L., G. L. Westbrook, and P. B. Guthrie (1984). Voltage-dependent block by Mg^{2+} of NMDA responses in spinal-cord neurons. *Nature* 309(5965), 261–263.
- McCulloch, W. S. and W. Pitts (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 115–133.
- Meffin, H., J. Besson, A. N. Burkitt, and D. B. Grayden (2006). Learning the structure of correlated synaptic subgroups using stable and competitive spike-timing-dependent plasticity. *Physical Review E* 73(4), 041911.
- Meffin, H., A. N. Burkitt, and D. B. Grayden (2004). An analytical model for the 'large, fluctuating synaptic conductance state' typical of neocortical neurons in vivo. *Journal of Computational Neuroscience* 16(2), 159–175.

- Miller, K. D. (1996). Synaptic economics: Competition and cooperation in synaptic plasticity. *Neuron* 17(3), 371–374.
- Miller, K. D. and D. J. C. Mackay (1994). The role of constraints in Hebbian learning. *Neural Computation* 6(1), 100–126.
- Molter, C., U. Salihoglu, and H. Bersini (2007). The road to chaos by time-asymmetric Hebbian learning in recurrent neural networks. *Neural Computation* 19(1), 80–110.
- Montgomery, J. M. and D. V. Madison (2002). State-dependent heterogeneity in synaptic depression between pyramidal cell pairs. *Neuron* 33(5), 765–777.
- Moreno-Bote, R., A. Renart, and N. Parga (2008). Theory of input spike auto- and cross-correlations and their effect on the response of spiking neurons. *Neural Computation* 20(7), 1651–1705.
- Morrison, A., A. Aertsen, and M. Diesmann (2007). Spike-timing-dependent plasticity in balanced random networks. *Neural Computation* 19(6), 1437–1467.
- Morrison, A., M. Diesmann, and W. Gerstner (2008). Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics* 98(6), 459–478.
- Neveu, D. and R. S. Zucker (1996). Postsynaptic levels of Ca²⁺ (i) needed to trigger LTD and LTP. *Neuron* 16(3), 619–629.
- Noda, H. and W. R. Adey (1970). Firing variability in cat association cortex during sleep and wakefulness. *Brain Research* 18(3), 513–526.
- Nowak, L., P. Bregestovski, P. Ascher, A. Herbet, and A. Prochiantz (1984). Magnesium gates glutamate-activated channels in mouse central neurons. *Nature* 307(5950), 462–465.
- Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology* 15(3), 267–273.
- Pfister, J. P. and W. Gerstner (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *Journal of Neuroscience* 26(38), 9673–9682.
- Poggio, G. F. and L. J. Viernstein (1964). Time series analysis of impulse sequences of thalamic somatic sensory neurons. *Journal of Neurophysiology* 27(4), 517–535.

- Rao, R. P. N. and T. J. Sejnowski (2001). Spike-timing-dependent Hebbian plasticity as temporal difference learning. *Neural Computation* 13(10), 2221–2237.
- Rauch, A., G. La Camera, H. R. Luscher, W. Senn, and S. Fusi (2003). Neocortical pyramidal cells respond as integrate-and-fire neurons to in-vivo-like input currents. *Journal of Neurophysiology* 90(3), 1598–1612.
- Rieke, F. (1997). *Spikes: exploring the neural code*. Cambridge, Mass.: MIT Press.
- van Rossum, M. C. W., G. Q. Bi, and G. G. Turrigiano (2000). Stable Hebbian learning from spike timing-dependent plasticity. *Journal of Neuroscience* 20(23), 8812–8821.
- van Rossum, M. C. W. and G. G. Turrigiano (2001). Correlation based learning from spike timing dependent plasticity. *Neurocomputing* 38, 409–415.
- Rubin, J. E., R. C. Gerkin, G. Q. Bi, and C. C. Chow (2005). Calcium time course as a signal for spike-timing-dependent plasticity. *Journal of Neurophysiology* 93(5), 2600–2613.
- Sabatini, B. L., T. G. Oertner, and K. Svoboda (2002). The life cycle of Ca²⁺ ions in dendritic spines. *Neuron* 33(3), 439–452.
- Salinas, E. and T. J. Sejnowski (2002). Integrate-and-fire neurons driven by correlated stochastic input. *Neural Computation* 14(9), 2111–2155.
- Schreiner, C. E., H. L. Read, and M. L. Sutter (2000). Modular organization of frequency integration in primary auditory cortex. *Annual Review of Neuroscience* 23, 501–529.
- Sejnowski, T. J. (1977). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology* 4(4), 303–321.
- Senn, W. (2002). Beyond spike timing: the role of nonlinear plasticity and unreliable synapses. *Biological Cybernetics* 87(5-6), 344–355.
- Senn, W., H. Markram, and M. Tsodyks (2001). An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. *Neural Computation* 13(1), 35–67.
- Shadlen, M. N. and W. T. Newsome (1998). The variable discharge of cortical neurons:

- Implications for connectivity, computation, and information coding. *Journal of Neuroscience* 18(10), 3870–3896.
- Sjöström, P. J., G. G. Turrigiano, and S. B. Nelson (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* 32(6), 1149–1164.
- Song, S. and L. F. Abbott (2001). Cortical development and remapping through spike timing-dependent plasticity. *Neuron* 32(2), 339–350.
- Song, S., K. D. Miller, and L. F. Abbott (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience* 3(9), 919–926.
- Sprekeler, H., C. Michaelis, and L. Wiskott (2007). Slowness: An objective for spike-timing-dependent plasticity? *PLoS Computational Biology* 3(6), 1136–1148.
- Standage, D., S. Jalil, and T. Trappenberg (2007). Computational consequences of experimentally derived spike-time and weight dependent plasticity rules. *Biological Cybernetics* 96(6), 615–623.
- Stein, R. B. (1965). A theoretical analysis of neuronal variability. *Biophysical Journal* 5(2), 173–194.
- Swindale, N. V. (1996). The development of topography in the visual cortex: A review of models. *Network-Computation in Neural Systems* 7(2), 161–247.
- Tang, A., D. Jackson, J. Hobbs, W. Chen, J. L. Smith, H. Patel, A. Prieto, D. Petrusca, M. I. Grivich, A. Sher, P. Hottowy, W. Dabrowski, A. M. Litke, and J. M. Beggs (2008). A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *Journal of Neuroscience* 28(2), 505–518.
- Thorpe, S. J., N. Bacon, G. A. Rousset, M. J. M. Mace, and M. Fabre-Thorpe (2002). Rapid categorisation of natural scenes: feedforward vs feedback contribution evaluated by backward masking. *Perception* 31, 150–150.
- Tsodyks, M. (2002). Spike-timing-dependent synaptic plasticity - the long road towards understanding neuronal mechanisms of learning and memory. *Trends in Neurosciences* 25(12), 599–600.
- Tzounopoulos, T., Y. Kim, D. Oertel, and L. O. Trussell (2004). Cell-specific,

- spike timing-dependent plasticities in the dorsal cochlear nucleus. *Nature Neuroscience* 7(7), 719–725.
- Tzounopoulos, T., M. E. Rubio, J. E. Keen, and L. O. Trussell (2007). Coactivation of pre- and postsynaptic signaling mechanisms determines cell-specific spike-timing-dependent plasticity. *Neuron* 54(2), 291–301.
- Wang, H. X., R. C. Gerkin, D. W. Nauen, and G. Q. Bi (2005). Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nature Neuroscience* 8(2), 187–193.
- Wenisch, O. G., J. Noll, and J. L. van Hemmen (2005). Spontaneously emerging direction selectivity maps in visual cortex through STDP. *Biological Cybernetics* 93(4), 239–247.
- Wilson, H. R. and J. D. Cowan (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal* 12(1), 1–24.
- Woodin, M. A., K. Ganguly, and M. M. Poo (2003). Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl⁻ transporter activity. *Neuron* 39(5), 807–820.
- Xu, L., M. White, and D. Schuurmans (2009). Optimal reverse prediction: A unified perspective on supervised, unsupervised and semi-supervised learning. In International Conference on Machine Learning (ICML-09); url = <http://www.cs.ualberta.ca/~dale/papers.html>.
- Yasuda, H. and T. Tsumoto (1996). Long-term depression in rat visual cortex is associated with a lower rise of postsynaptic calcium than long-term potentiation. *Neuroscience Research* 24(3), 265–274.
- Zou, Q. and A. Destexhe (2007). Kinetic models of spike-timing dependent plasticity and their functional consequences in detecting correlations. *Biological Cybernetics* 97(1), 81–97.