Neuromorphic VLSI Designs for Spike Timing and Rate-based Synaptic Plasticity with Application in Pattern Classification

by

S. Mostafa Rahimi Azghadi

B. Eng. (Computer Hardware Engineering, First Class Rank),
Sadjad University, Iran, 2006

M. Eng. (Computer Architecture Engineering, First Class Honours),
Shahid Beheshti University, Iran, 2009

Thesis submitted for the degree of

Doctor of Philosophy

in

Electrical and Electronic Engineering,
Faculty of Engineering, Computer and Mathematical Sciences
The University of Adelaide, Australia

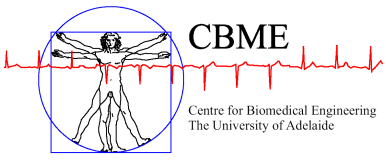
2014

Supervisors:

Dr Said Al-Sarawi, School of Electrical & Electronic Engineering

Dr Nicolangelo Iannella, School of Electrical & Electronic Engineering

Prof Derek Abbott, School of Electrical & Electronic Engineering



THE UNIVERSITY *of* ADELAIDE

To my dearest wife, Maryam
and to my Mum and Dad,
with all my love.

Contents

Abstract

This thesis presents a versatile study on the design and Very Large Scale Integration (VLSI) implementation of various synaptic plasticity rules ranging from phenomenological rules, to biophysically realistic ones. In particular, the thesis aims at developing novel spike timing-based learning circuits that advance the current neuromorphic systems, in terms of power consumption, compactness and synaptic modification (learning) abilities. Furthermore, the thesis investigates the usefulness of the developed designs and algorithms in specific engineering tasks such as pattern classification. To follow the mentioned goals, this thesis makes several original contributions to the field of neuromorphic engineering, which are briefed in the following.

First, a programmable multi-neuron neuromorphic chip is utilised to implement a number of desired rate- and timing-based synaptic plasticity rules. Specific software programs are developed to set up and program the neuromorphic chip, in a way to show the required neuronal behaviour for implementing various synaptic plasticity rules. The classical version of Spike Timing Dependent Plasticity (STDP), as well as the triplet-based STDP and the rate-based Bienenstock-Cooper-Munro (BCM) rules are implemented and successfully tested on this neuromorphic device. In addition, the implemented triplet STDP learning mechanism is utilised to train a feedforward spiking neural network to classify complex rate-based patterns, with a high classification performance.

In the next stage, VLSI designs and implementations of a variety of synaptic plasticity rules are studied and weaknesses and strengths of these implementations are highlighted. In addition, the applications of these VLSI learning networks, which build upon various synaptic plasticity rules are discussed. Furthermore, challenges in the way of implementing these rules are investigated and effective ways to address those challenges are proposed and reviewed. This review provides us with deep insight into the design and application of synaptic plasticity rules in VLSI.

Next, the first VLSI designs for the triplet STDP learning rule are developed, which significantly outperform all their pair-based STDP counterparts, in terms of learning capabilities. It is shown that a rate-based learning feature is also an emergent property

of the new proposed designs. These primary designs are further developed to generate two different VLSI circuits with various design goals. One of these circuits that has been fabricated in VLSI as a proof of principle chip, aimed at maximising the learning performance—but this results in high power consumption and silicon real estate. The second design, however, slightly sacrifices the learning performance, while remarkably improves the silicon area, as well as the power consumption of the design, in comparison to all previous triplet STDP circuits, as well as many pair-based STDP circuits. Besides, it significantly outperforms other neuromorphic learning circuits with various biophysical as well as phenomenological plasticity rules, not only in learning but also in area and power consumption. Hence, the proposed designs in this thesis can play significant roles in future VLSI implementations of both spike timing and rate based neuromorphic learning systems with increased learning abilities. These systems offer promising solutions for a wide set of tasks, ranging from autonomous robotics to brain machine interfaces.

Statement of Originality

I certify that this work contains no material, which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

	15/03/2014
Signed	Date

Acknowledgments

First and foremost, I would like to convey my deepest gratitude to my supervisors **Dr Said Al-Sarawi**, **Dr Nicolangelo Iannella** and **Prof. Derek Abbott** for their guidance and support throughout my candidature. My principal supervisor, Dr Al-Sarawi advised me with his open view and broad knowledge in the field of electronic engineering and circuit design. His critical, and thoughtful comments were always constructive and fruitful to improve the quality of my research. In addition, my co-supervisor Dr Iannella also provided me with his solid knowledge in the field of computational neuroscience and synaptic plasticity in Spiking Neural Networks. He has significantly helped me to embark in the field of neuromorphic engineering and find my direction in implementing new circuits for unexplored synaptic plasticity rules. Further, I would like to express my gratitude to my other co-supervisor, Prof. Abbott, who has been of great help, support and advice, when it counted most. With his enthusiastic supervision, he always encouraged me to conduct high quality research and publications.

Another key person whom I am strongly indebted to is **Prof. Giacomo Indiveri**. He invited me to visit his Neuromorphic Cognitive System (NCS) group in the Institute of Neuroinformatics (INI), in University/ETH Zurich, Switzerland. Prof. Indiveri's broad theoretical and experimental knowledge in the field of neuromorphic engineering was of great importance towards my PhD research. I would also like to thank his continuing support and encouragement throughout the course of last two years. In Zurich, I would like to thank my good friend and a brilliant neuromorphic researcher, Dr Saber Moradi, who introduced his IFMEM neuromorphic device to me and helped to implement my ideas on it and perform several experiments useful for my research. In addition, my sincere gratitude goes to my good friend, Dr Ning Qiao, a real VLSI chip design expert, for his significant help toward fabricating my circuit and integrating it on one of the NCS group VLSI chips. Ning was also of big help after the design was fabricated and tremendously helped me to test and measure my VLSI design. I am also indebted to my good friend, Federico Corradi, who kindly and tirelessly helped and supported me, when I was in Zurich. Other great scholars in the NCS group that were all very supportive and helpful during my long visits to Zurich, are Dr Fabio Stefanini, Dr Christian Mayr, Dora Sumislawska, Mehmet Ozdas, Mark Oswald, Hesham Mostafa Elsayed, Lorenz Muller, Jonathan Binas, Richard George and Daniel Fasnacht.

I am also thankful to David Lawrence, Kathrin Aguilar, and Simone Schumacher for their administrative assistance and support in the INI.

Zurich is really memorable for me because of my other great friends and colleagues in the INI. My great friend in the INI, Saurabh Bhargava, was always there for me to discuss both scientific and non-scientific life issues together and laugh at our problems and difficulties as PhD students. I am also thankful to other marvellous PhD and master students in the INI, Gabriela Michel, Hongjie Liu, Mitra Javadzadeh, Suraj Honnuraiah, David Bontrager, Nawal El Boghdady, Dennis Goldschmidt, Jonas Klein, Gina Paolini, Petar Ivanov, Karlis Kanders, Atanas Stankov, Asim Iqbal, Ivan Voitov, and Sofia Jativa.

I would like to express my deep gratitude to Dr Tara Hamilton from the University of New South Wales. I had the chance to first meet Dr Hamilton and discuss with her my neuromorphic research plan and ideas, in a special neuromorphic session at the IEEE ISSNIP conference in 2011, in Adelaide. Tara supported me and I was invited to the leading *Telluride Neuromorphic Engineering Workshop* that is yearly organised by the Institute of Neuromorphic Engineering (INE), in Telluride, Colorado, USA, where attendance to this workshop is by invitation only. In this three-week workshop I had great opportunity to work with other recognised researchers from around the world to discuss and verify some of my research approaches and methodology. Here, I would like to thank Prof. Ralph Etienne-Cummings from Johns Hopkins University for inviting me to this great scientific event. Also I would like to thank other workshop organisers and supporters for providing all the needed support in terms of accommodation and logistics. During the workshop I spent some time implementing STDP and TSTDP synaptic plasticity rules on the Spinnaker, a neuromorphic architecture developed at the University of Manchester. I learned about Spinnaker from two talented and enthusiastic researchers, Dr Sergio Davies, and Dr Francesco Galluppi. Besides, while in Telluride, I met Prof. Gert Cauwenberghs, one of the pioneers in the field of neuromorphic engineering, and his group members Dr Theodore Yu, and Mr Jongkil Park, from University of California, San Diego. They kindly introduced me to their developed large-scale neuromorphic chip, named HiAER IFAT. It was great to learn about leading neuromorphic projects in the world and at the same time, discuss my own research ideas with these renowned researchers. In addition, I had great pleasure to discuss my research ideas and plans with other experts in the field such as Prof. Tobi Delbruck and Dr Michael Pfeiffer from the INI, as well as Prof. Jonathan Tapson from University of

Western Sydney. I would also like to include my gratitude to Prof. Andre van Schaik from University of Western Sydney for his constructive comments and suggestions on one of my TSTDP circuit designs presented in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, in Brisbane and also during the Telluride Workshop.

Back to Australia, there are many people who helped me throughout my PhD candidature. Within the school of Electrical & Electronic Engineering, I am indebted to Dr Yingbo Zhu, who provided me with valuable help in Cadence, whenever I asked. I would also like to thank Dr Braden Philips, Dr Brian Ng and Associate Prof. Michael Liebelt for inviting me to attend and present my research at their weekly group journal club. My first days in the school was great and happy because of the support and friendliness of Dr Jega Balakrishnan, Dr Benjamin Ung and Dr Hungyen Lin. Indeed, one of the most respected colleagues I had in the school, has been Dr Muammar Kabir, who was always there for me and I could count on his help and advice. Thank you very much Muammar. I also like to thank the office & support staff of the school including Danny Di Giacomo for the logistical supply of required tools and items, IT officers, David Bowler, Mark Innes, and Ryan King, and the administrative staff, Stephen Guest, Greg Pullman, Ivana Rebellato, Rose-Marie Descalzi, Deborah Koch, Lenka Hill, and Jodie Schluter for their kindness and assistance. I would also like to thank the head of school, Associate Prof. Cheng-Chew Lim, for his support regarding my research travels and software required for my PhD thesis. I am also thankful to my other friends and colleagues within the school. First of all, I express my deep gratitude to my dear friend, Dr Ali Karami Horestani for his endless help and invaluable advice in all aspects of my life in Adelaide. In addition, I sincerely thank my other friends and colleagues Dr Pouria Yaghmaee, Amir Ebrahimi, Sam Darvishi, Mehdi Kasaee, Arash Mehdizadeh, Zahra Shaterian, Neda Shabi, Muhammad Asraful Hasan, Sarah Immanuel, Robert Moric, Tran Nguyen, and Yik Ling Lim, for making such a friendly research environment. Also my wife and I appreciate help and support of our family friends in Adelaide, Reza Hassanli, Zahra Ranjbari, Javad Farrokhi, Sahar Daghagh, Amir Mellati, Hosna Borhani, Sara Chek, Hedy Minoofar, and Mehregan Ebrahimi for their help to make our life in Adelaide so good. In addition, I am thankful to my dear Australian friend, Ben Chladek and his great family for their kindness and support.

Looking back at my *alma mater*, I am indebted to my master degree supervisor, Prof. Keivan Navi, for his great supervision and encouraging attitude toward research. Other scholars who contributed to my academic background are Dr Hamed Shah-Hosseini,

Associate Prof. Mohsen Ebrahimi-Moghadam, Associate Prof. Omid Hashemipour, Dr Ali Zakeralhosseini, Dr Namdar Saniei, and Dr Ali Jahanian. I am also thankful to my other colleagues and friends at Shahid Beheshti University, Dr Mahmoud Fazlali, Dr Hossein Torkaman, Dr Amir Kaivani, Dr Hossein Pishgar and Mr Vahab Samadi.

I recognise that this research would not have been possible without the financial assistance of Australian Government via a generous International Postgraduate Research Scholarship (IPRS) and Adelaide University Scholarship (AUS). During my candidature, I was awarded several other travel grants, scholarships and awards by several other organisations. Here, I would like to deeply appreciate these organisations support including annual travel grants from the School of Electrical & Electronic Engineering (2011-2013), OIST Japanese Nerual Network Society Travel grant (2011), the Adelaide University D. R. Stranks Postgraduate Fellowship (2012), the Adelaide University Research Abroad Scholarship (2012), the IEEE Computational Intelligence Society travel grant (2012), Brain Corporation Travel Fellowships for Spiking Neural Networks (2012), AFUW-SA Doreen McCarthy Research Bursary (2013), IEEE SA Section travel award (2013), and Australia's Defence Science and Technology Organisation (DSTO) Simon Rockliff Supplementary Scholarship (2013).

Back to my home country, my endless gratitude goes to my father and mother who always endow me with infinite support, wishes, continuous love, encouragement, and patience. I also thank my best and kindest sisters and brothers for their love and support. I also wish to express my warm and sincere thanks to my father- and mother-in-law for their kindness, guidance, and earnest wishes. Also, I would like to thank my sister-in-law for her support, inspiration, and kindness.

Last but not least, my most heartfelt thanks are due to my dearest stunning wife, *Maryam*. Words are unable to express my deepest appreciation and love to her. She stood by me in all ups and downs of my PhD and always endowed me with her endless love, and support. Darling, I love you from the bottom of my heart.

S. Mostafa Rahimi Azghadi,
March 2014,
Adelaide

Thesis Conventions

The following conventions have been adopted in this Thesis:

Typesetting

This document was compiled using LaTeX2e. Texmaker and TeXstudio were used as text editor interfaced to LaTeX2e. Inkscape and Xcircuit were used to produce schematic diagrams and other drawings.

Referencing

The Harvard style has been adopted for referencing.

System of units

The units comply with the international system of units recommended in an Australian Standard: AS ISO 1000–1998 (Standards Australia Committee ME/71, Quantities, Units and Conversions 1998).

Spelling

Australian English spelling conventions have been used, as defined in the Macquarie English Dictionary (A. Delbridge (Ed.), Macquarie Library, North Ryde, NSW, Australia, 2001).

Awards and Scholarships

2013

- Simon Rockliff Scholarship, DSTO
- Doreen McCarthy Research Bursary, AFUW-SA
- IEEE South Australia Section Student Travel Award, IEEE SA

2012

- Brain Corporation Fellowships for Spiking Neural Networks, IEEE WCCI2012
- The IEEE Computational Intelligence Society travel grant, IEEE WCCI2012
- Research Abroad Scholarship, The University of Adelaide
- D. R. Stranks fellowship, The University of Adelaide

2011

- Japanese Neural Network Society Travel Award, Okinawa Institute of Science and Technology

2010

- International Postgraduate Research Scholarships, The Australian Government
- Adelaide University Scholarships, The University of Adelaide

Publications

Journal Articles

Azghadi-M. R., Al-Sarawi-S., Iannella-N., Indiveri-G., Abbott-D. (2014b). Spike-based synaptic plasticity in silicon: Design, implementation, application, and challenges, *Proceedings of the IEEE*, **102**(5), pp. 717–737. *

Azghadi-M. R., Moradi-S., Fastnacht-D., Ozdas-M. S., Indiveri-G. (2014c). Programmable spike-timing dependent plasticity learning circuits in neuromorphic VLSI architectures, *ACM Journal on Emerging Technologies in Computing Systems*, Submitted on 15 Dec. 2013. *

Azghadi-M. R., Al-Sarawi-S., Iannella-N., and Abbott-D. (2014a). Tunable low energy, compact and high performance neuromorphic circuit for spike-based synaptic plasticity, *PLoS ONE*, **9**(2), art. no. e88326. *

Azghadi-M. R., Al-Sarawi-S., Abbott-D., and Iannella-N. (2013a). A neuromorphic VLSI design for spike timing and rate based synaptic plasticity, *Neural Networks*, **45**, pp. 70–82. *

Hashemi-S., **Azghadi-M. R.**, Zakerolhosseini-A., and Navi-K. (2014). A novel FPGA programmable switch matrix interconnection element in quantum-dot cellular automata, *International Journal of Electronics*, DOI: 10.1080/00207217.2014.936526.

Azghadi-M. R., Kavehei-O., and Navi-K. (2007b). A novel design for quantum-dot cellular automata cells and full adders, *Journal of Applied Sciences*, **7**(22), pp. 3460–3468.

Navi-K., Foroutan-V., **Azghadi-M. R.**, Maeen-M., Ebrahimpour-M., Kaveh-M., and Kavehei-O. (2009). A novel low-power full-adder cell with new technique in designing logical gates based on static CMOS inverter, *Microelectronics Journal*, **40**(10), pp. 1441–1448.

NAVI-K., SAYEDSALEHI-S., FARAZKISH-R., AND **AZGHADI-M. R.** (2010b). Five-input majority gate, a new device for quantum-dot cellular automata, *Journal of Computational and Theoretical Nanoscience*, **7**(8), pp. 1546–1553.

NAVI-K., FARAZKISH-R., SAYEDSALEHI-S., AND **AZGHADI-M. R.** (2010a). A new quantum-dot cellular automata full-adder, *Microelectronics Journal*, **41**(12), pp. 820–826.

FARAZKISH-R., **AZGHADI-M. R.**, NAVI-K., AND HAGHPARAST-M. (2008). New method for decreasing the number of quantum dot cells in QCA circuits, *World Applied Sciences Journal*, **4**(6), pp. 793–802.

Book Chapter

BONYADI-M., **AZGHADI-M. R.**, AND SHAH-HOSSEINI-H. (2008). Population-based optimization algorithms for solving the travelling salesman problem, *Traveling Salesman Problem*, Federico Greco (Ed.), ISBN: 978-953-7619-10-7, InTech, DOI: 10.5772/5586, pp. 1–34.

Conference Articles

Azghadi-M. R., Kavehei-O., Al-Sarawi-S., Iannella-N., Abbott-D. (2011c). Novel VLSI implementation for triplet-based spike-timing dependent plasticity, *Proceedings of the 7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Adelaide, Australia, pp. 158–162. *

Azghadi-M. R., Kavehei-O., Al-Sarawi-S., Iannella-N., Abbott-D. (2011d). Triplet-based spike-timing dependent plasticity in silicon, *The 21st Annual Conference of the Japanese Neural Network Society*, Okinawa, Japan, art. no. P3–35. *

Azghadi-M. R., Al-Sarawi-S., Iannella-N., and Abbott-D. (2011b). Physical implementation of pair-based spike-timing-dependent plasticity, *2011 Australian Biomedical Engineering Conference*, Darwin, Australia, Vol. 34, art. no. 141. *

Azghadi-M. R., Al-Sarawi-S., Iannella-N., and Abbott-D. (2011a). Emergent BCM via neuromorphic VLSI synapses with STDP, *5th Australian Workshop on Computational Neuroscience*, Sydney, Australia, p. 31. *

Azghadi-M. R., Al-Sarawi-S., Iannella-N., and Abbott-D. (2012b). Efficient design of triplet based spike-timing dependent plasticity, *Proc. IEEE 2012 International Joint Conference on Neural Networks (IJCNN)*, Brisbane, Australia, DOI: 10.1109/IJCNN.2012.6252820. *

Azghadi-M. R., Al-Sarawi-S., Iannella-N., and Abbott-D. (2012a). Design and implementation of BCM rule based on spike-timing dependent plasticity, *Proc. IEEE 2012 International Joint Conference on Neural Networks (IJCNN)*, Brisbane, Australia, DOI: 10.1109/IJCNN.2012.6252778. *

Azghadi-M. R., Moradi-S., and Indiveri-G. (2013b). Programmable neuromorphic circuits for spike-based neural dynamics, *11th IEEE International New Circuit and Systems (NEWCAS) Conference*, Paris, France, DOI: 10.1109/NEWCAS.2013.6573600. *

Azghadi-M. R., Al-Sarawi-S., Iannella-N., and Abbott-D. (2013c). A new compact analog VLSI model for spike timing dependent plasticity, *2013 IFIP/IEEE 21st International Conference on Very Large Scale Integration (VLSI-SoC)*, Istanbul, Turkey, pp. 7–12. *

AZGHADI-M. R., AL-SARAWI-S., ABBOTT-D., AND IANNELLA-N. (2013b). Pairing frequency experiments in visual cortex reproduced in a neuromorphic STDP circuit, *2013 20th IEEE International Conference on Electronics, Circuits, and Systems*, Abu-Dhabi, UAE, pp. 229–232. *

AZGHADI-M. R., BONYADI-M. R., AND SHAHHOSSEINI-H. (2007). Gender classification based on feedforward backpropagation neural network, *Artificial Intelligence and Innovations 2007: from Theory to Applications*, Springer US, Athens, Greece, pp. 299–304.

AZGHADI-M. R., BONYADI-M. R., HASHEMI-S., AND MOGHADAM-M. E. (2008). A hybrid multiprocessor task scheduling method based on immune genetic algorithm, *Proceedings of the 5th International ICST Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*, Hong Kong, pp. 561–564.

KAVEHEI-O., **AZGHADI-M. R.**, NAVI-K., AND MIRBAHA-A.-P. (2008). Design of robust and high-performance 1-bit CMOS full adder for nanometer design, *IEEE Computer Society Annual Symposium on VLSI, ISVLSI'08*, Paris, France, pp. 10–15.

BONYADI-M. R., **AZGHADI-M. R.**, AND HOSSEINI-H. S. (2007b). Solving traveling salesman problem using combinational evolutionary algorithm, *Artificial Intelligence and Innovations 2007: from Theory to Applications*, Springer US, Athens, Greece, pp. 37–44.

BONYADI-M., **AZGHADI-M. R.**, RAD-N., NAVI-K., AND AFJEI-E. (2007a). Logic optimization for majority gate-based nanoelectronic circuits based on genetic algorithm, *IEEE International Conference on Electrical Engineering*, Lahore, Pakistan, pp. 1–5.

HASHEMI-S., **AZGHADI-M. R.**, ZAKEROLHOSSEINI-A. (2008). A novel QCA multiplexer design, *IEEE International Symposium on Telecommunications*, Tehran, Iran, pp. 692–696.

KAZEMI-FARD-N., EBRAHIMPOUR-M., **AZGHADI-M. R.**, TEHRANI-M., AND NAVI-K. (2008). Performance evaluation of in-circuit testing on qca based circuits, *IEEE East-West Design & Test Symposium (EWDTS)*, Lviv, Ukraine, pp. 375–378.

ADINEH-VAND-A., LATIF-SHABGAHI-G., AND **AZGHADI-M. R.** (2008). Increasing testability in QCA circuits using a new test method, *IEEE International Design and Test Workshop*, Monastir, Tunisia, pp. 40–44.

ADINEH-VAND-A., PARANDIN-F., **AZGHADI-M. R.**, AND KHALILZADEH-A. (2009). Solving multi-processor task scheduling problem using a combinatorial evolutionary algorithm, *The 9th Workshop on Models and Algorithms for Planning and Scheduling Problems (MAPSP09)*, Kerkrade, the Netherlands, pp. 91–93.

Note: Articles with an asterisk (*) are directly relevant to this thesis.

List of Figures

List of Tables

Chapter 1

Introduction

THIS chapter provides the reader with introductory background on Spiking Neural Network (SNN) and discusses why neuromorphic engineering is important. Identified research gaps and the motivations behind the current study are also outlined in this chapter. Furthermore, the objectives of the thesis and the research questions and goals are discussed. Besides, the original contributions made in this thesis to reach the mentioned goals are highlighted. Finally, the structure and outline of this thesis are described.

1.1 Introduction

1.1.1 Neural Networks

The human being is one of the most mysterious and complicated creatures to our knowledge. They can understand, learn, deduce, recognise, and judge. Many researchers are inspired by various aspects of human genetic and neural mechanisms. Scientists have proposed a variety of human- and nature-inspired problem solving techniques such as the Genetic Algorithm (GA) (Goldberg 1989, Azghadi *et al.* 2008, Bonyadi *et al.* 2007) and Artificial Neural Network (ANN) (Haykin 1994, Azghadi *et al.* 2007). Artificial neural networks have attracted much attention, during the last few decades. This has resulted in three various generations of these networks (Vreeken 2003), which are based on three different types of neurons.

The first generation of artificial neural networks is composed of threshold neurons or binary units, that were firstly proposed by McCulloch and Pitts (1943). This neuron model is quite simple and just acts as a threshold detector. Put simply, when the summation of the weighted inputs of this type of neuron is more than a predefined threshold, the neuron fires, i.e. it produces a pulse, which can be used to represent a binary state. These networks are definitely suitable for performing Boolean operations, but when the problem requires analog computations and approximations, they are not practical (Vreeken 2003).

The second generation of artificial neural networks that is more realistic and closer to real neurons introduces sigmoid (i.e. logistic) neurons, which present a continuous output function rather than just a binary or quantised output (DasGupta and Schnitger 1994). In the first two generations of neural networks, the timing of input-output is not important and just their rate can carry information, but it is known that in real neurons, which can carry out many complicated learning and recognition tasks, there is a more intricate input-output relation (Dayan and Abbott 2001). This relation brings timing into action and makes a spatial-temporal information coding and transition model of neurons, so-called spiking neurons.

Spiking neurons are basic building blocks of the third generation of artificial neural networks, so-called SNN. Since synaptic plasticity in SNNs is the focus of this thesis, in the following, a brief introduction on this type of neural networks is presented.

1.1.2 Spiking Neural Networks

A SNN is composed of spiking neurons and synapses. As its name infers, a spiking neuron sends information in the form of electrical pulses, so-called spike (i.e. action potentials), to other neurons and synapses in the network. This is through the propagation of these spikes that information is transferred in the SNN (Gerstner and Kistler 2002). There is general agreement on the information coding in the form of the rate, timing, and spatial/temporal correlations of the spikes communicated among neurons. Therefore, a neuron depending on its type, its information type, its place in the network, and other factors, is able to fire spikes in various ways. For this reason, various neuron models were proposed by computational and theoretical neuroscientists to mimic the operation of real neurons. This thesis briefly discusses neurons and addresses their various models and behaviours, in the next chapter.

Besides spiking neurons, synapses are the other basic building blocks in a SNN. A synapse is the connection point of one neuron to its neighbouring neurons. It is widely believed that learning, computation and memory processes take place in synapses (Sjöström *et al.* 2008). Since the learning and computation in a SNN is a dynamic process, so the synapses that are the main blocks for learning and computation should also be dynamic and modifiable. However, the open question is how the modification takes place in the synapses within the brain, in a way that many complicated and real-time tasks such as learning, computation, and cognition are performed so smoothly and accurately. Although there is no general agreement as to the answer to this question, there are several hypotheses stating that these modifications take place in relation to the activity of pre- and post-synaptic neurons connected to the synapse (Sjöström *et al.* 2008). These activities, which are the basis of this thesis are elaborated in the following chapters.

After investigating the SNNs and their structure, scientists are able to propose models and techniques to implement these spiking networks of neurons and synapses. The implementation can be performed either in software or hardware. However, there are fundamental differences in software and hardware implementations of SNNs.

Since neuron and synapse models and behaviours are described by algorithms and mathematical terms, they can be easily implemented as software programs and run on ordinary computers. These implementations are also easily modifiable and can be altered according to the needs, in every step of the implementation process. These programs that represent neurons and synapses can then be integrated and connected

Table 1.1. Neuron and synapse quantity. Number of neurons and synapses in humans as compared to various animals (Ananthanarayanan *et al.* 2009).

	Mouse	Rat	Cat	Human
Neurons $\times 10^8$	0.160	0.550	7.63	200
Synapses $\times 10^{12}$	0.128	0.442	6.10	200

to each other, in order to from a SNN. This approach is usually noise and error-free and does not require a special technology to realise them. Despite all these benefits though, *it requires a supercomputer to simulate a mouse or cat cortex* (Frye *et al.* 2007, Ananthanarayanan *et al.* 2009).

This is because of the fact that the Von Neumann architecture, which is the foundation of all today's computers, runs programs sequentially. Therefore a large amount of processing, which in turn needs a large amount of memory, must be performed, in order to realise a portion of cortex containing billions of neurons and trillions of synapses (Frye *et al.* 2007, Ananthanarayanan *et al.* 2009). Table 1.1 shows how very large the neuron and synapse numbers are in the mouse, rat, cat, and human cortex. These numbers demonstrate why simulating a neural network requires supercomputers.

Researchers in IBM have utilised and set up a very large supercomputer with 147,456 CPUs and 144 TB of main memory, in order to perform a cat-scale cortical simulation, which is demonstrated in Ananthanarayanan *et al.* (2009). To this complexity and very large resource requirements, one may add the complexity and resources for implementing special purpose programs for neurons, synapses, and their interconnections. In addition, in terms of power consumption and area, a supercomputer, by no means is close to a biological cortex. Furthermore, a simulated cortex on the mentioned supercomputers, is an order of magnitude slower than biological neural systems (Frye *et al.* 2007, Ananthanarayanan *et al.* 2009). Therefore, this raises the question "how can researchers implement a reasonably large scale SNN, which consumes reasonable power, takes moderate area, and process in biologically plausible time scales or even faster time scales?"

Carver Mead as one of the founding fathers of silicon chips and modern electronic engineering, in his seminal neuromorphic engineering paper (Mead 1990) states,

Biological information-processing systems operate on completely different principles from those with which most engineers are familiar. For many problems, particularly those in which the input data are ill-conditioned and the computation can be specified in a relative manner, biological solutions are many orders of magnitude more effective than those we have been able to implement using digital methods. This advantage can be attributed principally to the use of elementary physical phenomena as computational primitives, and to the representation of information by the relative values of analog signals, rather than by the absolute values of digital signals. This approach requires adaptive techniques to mitigate the effect of component differences. This kind of adaptation leads naturally to systems that learn about their environment. Large-scale adaptive analog systems are more robust to component degradation and failure than are more conventional systems, and they use far less power. For this reason, adaptive analog technology can be expected to utilise the full potential of the wafer-scale silicon fabrication.

This provides us a rather clear answer to the above mentioned questions and opens horizons to a new field of engineering, as first posed by Carver Mead, *Neuromorphic Engineering*.

1.1.3 Neuromorphic Engineering

Neuromorphic engineering is the art of implementing SNN-inspired hardware (Mead 1989, Mead 1990). Neuromorphic engineers have good knowledge of both electronic engineering concepts, and biological models of neurons and synapses. They strive to use physical characteristics of silicon transistors to mimic neural systems and architectures. The silicon-based neuromorphic systems then can be used in various tasks ranging from robotics, to the Brain Machine Interface (BMI), to cognition. There are a variety of neuromorphic systems that have been implemented using Analog, Digital, and Analog/Digital Very Large Scale Integration (VLSI) technologies (Indiveri *et al.* 2011, Hamilton and Tapson 2011). In addition, with recent appearance of memristors as a new nano-electronic device (Fortuna *et al.* 2009, Eshraghian 2010, Eshraghian *et al.* 2012), which is nicely compatible to the biology of synapses, memristive neuromorphic systems are becoming quite popular (Demming *et al.* 2013, Azghadi *et al.* 2013d, Wang *et al.* 2014b, Kudithipudi *et al.* 2014, Sheri *et al.* 2014, Sheridan and Lu 2014).

In comparison to the aforementioned software implementation of neural systems, neuromorphic VLSI systems have significant benefits. These benefits include

- very high degree of integration

- very low power consumption
- real-time implementation of neural systems.

Software neural systems though, have some advantages such as reconfigurability, ease of implementation, and noise-tolerance, over the neuromorphic VLSI systems. However, when considering a large scale neural network, the advantages of VLSI neuromorphic systems are significant. That is the main reason why neuromorphic VLSI engineering has been growing and attracting attention since the pioneering ideas of Carver Mead in the late 80's (Mead 1989). Since then, neuromorphic engineers have been developing various analog, digital, and mixed-signal VLSI circuits of neurons and synapse models that have been proposed by computational and theoretical neuroscientists (Indiveri 2003, Indiveri *et al.* 2006, Hamilton *et al.* 2008, Indiveri *et al.* 2009, Indiveri *et al.* 2011, Hamilton and van Schaik 2011, Azghadi *et al.* 2013a).

Furthermore, using these silicon-based neurons and synapses, neuromorphic engineers have also been implementing biologically inspired systems that are useful for real-world applications such as pattern classification (Mitra *et al.* 2009, Giulioni *et al.* 2009), feature extraction (Vogelstein *et al.* 2007a) and orientation selectivity (Choi *et al.* 2004, Chicca *et al.* 2007). In order to perform any cognitive tasks, learning and memory are the inevitable requirements of these neuromorphic systems. However, one might ask how the required learning and memory take place in these networks to enable them to perform such challenging tasks? The motivation of this thesis is defined around this basic question.

1.2 Research Gaps and Objectives of the Thesis

The underlying mechanisms and processes responsible for learning and long-term memory in the brain has remained an important yet strongly debated subject for researchers in various fields ranging from neurophysiology through to neuromorphic engineering. It is widely believed that processes responsible for synaptic plasticity provide key mechanisms underlying learning and memory in the brain (Song *et al.* 2000, Pfister and Gerstner 2006, Sjöström *et al.* 2001, Wang *et al.* 2005). However, despite the well established agreement on the fact that learning and memory are based on synaptic efficacy (i.e. weight) plasticity (Martin *et al.* 2000, Shouval *et al.* 2002, Cooper *et al.* 2004), there are various schools of thought on answers to questions such as i) what are the

causes of these plastic behaviours? and ii) what are the connections of these behaviours to learning?

The variety of viewpoints in answering the above mentioned questions can be due to the intrinsic complexity of learning and memory phenomena. This fact causes various classes of synaptic plasticity rules, with different mechanisms of induction and expression, have been being proposed (Abbott and Nelson 2000, Mayr *et al.* 2010, Shouval 2011, Graupner and Brunel 2012). As these various rules are proposed, neuromorphic engineers implement silicon models of them. Probably the most recognised example of synaptic plasticity rules among neuromorphic engineers is Spike Timing Dependent Synaptic Plasticity (STDP) (Markram *et al.* 1997, Bi and Poo 1998). This rule has been implemented in different studies and by various groups (Bofill-I-Petit and Murray 2004, Cameron *et al.* 2005, Indiveri *et al.* 2006, Tanaka *et al.* 2009, Bamford *et al.* 2012b).

However, the research gap here is that, although the traditional form of STDP, Pair-based Spike Timing Dependent Plasticity (PSTDP), has shown success in solving some computational and learning problems both in computational neuroscience (Song *et al.* 2000, Lisman and Spruston 2010, Masquelier and Thorpe 2007, Masquelier and Thorpe 2010, Davison and Frégnac 2006) and in neuromorphic engineering (Bofill-I-Petit and Murray 2004, Cameron *et al.* 2005, Indiveri *et al.* 2006, Koickal *et al.* 2007, Tanaka *et al.* 2009, Arena *et al.* 2009, Seo *et al.* 2011), recent studies show that the simple form of PSTDP that changes the synaptic weights according to a linear summation of weight changes, is not able to account for a variety of biological experiments (Froemke and Dan 2002, Pfister and Gerstner 2006, Gjorgjieva *et al.* 2011) and hence may lack learning and computational capabilities compared to more elaborate new synaptic plasticity models. Although a variety of other synaptic plasticity models, rather than just simple PSTDP exist, the implementation of these detailed synaptic plasticity rules in VLSI and their use in various engineering applications is a rather unexplored research area, and gives rise to a gap in neuromorphic engineering. Therefore, there is a need to explore these new synaptic plasticity rules in VLSI and test them in terms of performance for generating the outcome of various biological experiments (and therefore mimicking real synapses), as well as their use in real-world applications such as in pattern classification tasks.

The main objectives of this thesis are to investigate new methods for design, analysis, and implementation of a number of novel and unexplored synaptic plasticity rules

in VLSI. The main specific rule targeted in this thesis is a derivation of the classical STDP rule that is named Triplet-based Spike Timing Dependent Plasticity (TSTDP). This rule is shown to be able to account for a variety of biological experiments (Pfister and Gerstner 2006). In addition, TSTDP exploited for pattern selection and classification (Gjorgjieva *et al.* 2011). This thesis verifies the performance of a number of proposed VLSI designs in terms of replication of the outcomes of a variety of biological experiments. It also investigates the power consumption, and silicon area of the novel proposed designs and compares them to previous VLSI designs of a variety of synaptic plasticity rules, to justify if the proposed circuits are able to help reach the long-lasting goal of having a large-scale SNN with features close to that of the brain (Poon and Zhou 2011). Furthermore, the thesis also investigates the usefulness of the TSTDP rule in pattern classification tasks and reproducing the outcomes of a number of other synaptic plasticity rules, such as rate-based Bienenstock Cooper Munro (BCM) (Bienenstock *et al.* 1982, Cooper *et al.* 2004) or timing-based suppressive model of STDP (Froemke and Dan 2002).

The newly introduced synaptic plasticity VLSI circuits tend to be used in various large-scale SNNs with increased ability of learning, and improved synaptic plasticity ability. This will lead to neuromorphic cognitive systems with higher degree of applicability in real-time cognition tasks such as pattern recognition and classification. The next Section represents a summary of original contributions made to reach the above mentioned objectives and to fill the mentioned research gaps.

1.3 Summary of Original Contributions

The original contributions presented throughout this thesis can be itemised as follows:

- A programmable hardware-software multi-neuron neuromorphic chip called Integrate and Fire with Memory (IFMEM) is set up and programmed, in a way to show the required neuronal behaviour for implementing various synaptic plasticity rules on this neuromorphic system. The silicon neuron and synapse response properties of this system are carefully controlled through specific software programs developed for this purpose. The experimental results demonstrate that the programmable IFMEM chip operates correctly over a wide range of input frequencies and therefore is suitable for use in various real-world applications, where interfacing to real-world sensors and systems is required, or even

when higher processing speed is required. The results of this study is published in the *IEEE International New Circuit and System Conference* (Azghadi *et al.* 2013d).

- For the first time, the PSTDP, TSTDP and BCM rules are successfully implemented on the IFMEM neuromorphic chip. Furthermore, a pattern classification neural network is also implemented on this multi-neuron chip. In more detail:
  (i) It is successfully shown that both PSTDP and TSTDP rules implemented using silicon neurons and programmable synapses, can demonstrate the expected behaviours, similar to those seen in biological experiments.
  (ii) It is also shown how the STDP window can be generated using the silicon neurons and synapses available on the system.
  (iii) The PSTDP rule is used and tested for generating a competitive Hebbian learning behaviour observed in computational STDP experiments.
  (iv) For the first time the TSTDP learning algorithm is implemented on the IFMEM neuromorphic hardware.
  (v) In order to test the TSTDP implementation, the rate-based BCM learning behaviour is reproduced by this implementation. This experiment shows the usefulness of this timing-based learning algorithm for generating the rate-based BCM learning behaviour.
  (vi) Finally, the implemented TSTDP learning mechanism is utilised to train a simple feedforward spiking neural network to classify some complex rate-based patterns.
  Obtained results show the high performance of the TSTDP rule in the targeted classification task. In addition, the preformed research in this part provides good view of the TSTDP rule and its properties and features, which are essential when designing VLSI TSTDP synapses in the next parts of this research. The results of the above mentioned study are presented in *The ACM Journal on Emerging Technologies in Computing Systems* (Azghadi *et al.* 2014c).

- A previous VLSI implementation of the PSTDP rule is simplified to reduce area and power consumption. The result of this study is presented in *the Engineering and Physical Sciences in Medicine and the Australian Biomedical Engineering Conference* (Azghadi *et al.* 2011b). In addition, various synaptic plasticity experiments have been performed using different PSTDP circuits. The results show that all PSTDP circuits including the modified PSTDP circuit are unable to replicate many complicated biological experiments.

- The first VLSI design for the TSTDP rule is presented in this thesis. It is shown that this voltage-based TSTDP circuit is able to mimic the outcomes of a wide range of synaptic plasticity experiments including timing-based, hybrid rate/-timing based, and rate-based synaptic plasticity experiments, under various protocols and conditions. The circuit and these primary results are presented in *The 21st Japanese Neural Network Society Annual Conference* (Azghadi *et al.* 2011d). In addition, a previous voltage-based PSTDP VLSI design proposed by Indiveri *et al.* (2006) is investigated, optimised and simulated to show the various synaptic plasticity experiments. The results show that this PSTDP circuit, similar to our previous modified PSTDP VLSI circuit (Azghadi *et al.* 2011b), fails to account for many experiments. The comparison between the performance of this PSTDP circuit and the first proposed TSTDP circuit is presented in *The IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing* (Azghadi *et al.* 2011c).

- Furthermore, the proposed voltage-based TSTDP circuit is investigated for reproducing a similar behaviour to the outcomes of a rate-based BCM experiment. The results presented in *the 2011 Australian Computational Neuroscience Workshop* (Azghadi *et al.* 2011a) demonstrate that the circuit closely mimics the sliding threshold behaviour of the BCM rule. In addition, with further investigations, it was observed that both PSTDP and TSTDP circuits are able to account for a BCM-like behaviour, while the proposed TSTDP circuit has better performance and control over this behaviour. The new results are presented in *The 2012 IEEE International Joint Conference on Neural Networks* (Azghadi *et al.* 2012a).

- The first voltage-based TSTDP design is not able to account for the exponential behaviour of the STDP learning rule. In addition, it is not able to reproduce the exponential learning window generated by the computational model of the STDP rule presented in Song *et al.* (2000). Therefore, in order to remove this deficiency, the voltage-based circuit was modified and a new synaptic plasticity circuit is proposed, which uses the current-mode design strategy based on the design proposed by Bofill-I-Petit and Murray (2004). This circuit is verified to generate the exponential learning window as well as the outcomes of all required complicated experiments. The results are presented in *The 2012 IEEE International Joint Conference on Neural Networks* (Azghadi *et al.* 2012b).

- Further investigation reveals that although the first two TSTDP designs can generate the outcomes of many experiments, none of them are able to correctly account for other set of experiments performed based on other STDP-based synaptic plasticity rules such as the suppressive STDP rule proposed by Froemke and Dan (2002). In addition, these circuits are not able to correctly regenerate the outcomes of BCM experiments under a specific condition, where the synapse is driven pre-synaptically as in the original experiments (Kirkwood *et al.* 1996). Therefore, a new high-performance VLSI design for the TSTDP rule is proposed that outperforms the other TSTDP VLSI designs in several respects. It is shown that the new proposed design has significantly lower synaptic plasticity prediction error, in comparison with previous designs for TSTDP and PSTDP rules. In addition, it is also shown that this new design can successfully account for a number of new experiments, including experiments involved with various spike triplet combinations, as well as pre-synaptic and post-synaptic driven rate-based BCM-like experiments, where the previous TSTDP and PSTDP designs do not show suitable performance and cannot mimic the experiments effectively. This new design is also tested against process variation and device mismatch. It is shown that, although the circuit is susceptible to process variation, it is possible to mitigate the effect of variations and fine-tune the circuit to its desired behaviour. In addition, the power consumption and area of the proposed design are also investigated and discussed. Obtained results are presented mainly in *Neural Networks* (Azghadi *et al.* 2013a).

- The proposed circuit has been recently fabricated as a proof of principle and the measurement results testify to the correct functionality of the fabricated circuit in performing triplet-based synaptic weight modification.

- Although the previous TSTDP design has a very high performance in reproducing the outcomes of all required experiments, compared to its PSTDP and TSTDP counterparts, it consumes significantly high power to process each spike. Furthermore, considering the number of transistors and the use of five capacitors, from which one is a very large capacitor of the size of 10 pF, this high performance TSTDP circuit occupies a large silicon area. However, in order to follow the long-lasting goal of integrating synaptic plasticity circuits in a large-scale neuromorphic system, which includes millions of these circuits, they should be of a practicable size, consume very low energy, have an acceptable performance and

tolerate the process variation to some extent. A new compact, ultra low energy, high performance, and variation tolerant design is proposed to satisfy the needs for such large-scale systems. This design utilises a 50 fF capacitor instead of the very large capacitors used in the previous TSTDP circuit, while retaining its ability to reproduce the STDP learning window, and the triplet and quadruplet experimental data. This design and its experimental results are presented in the *2013 IFIP/IEEE 21st International Conference on Very Large Scale Integration (VLSI-SoC)* (Azghadi *et al.* 2013c).

- Due to its small capacitor, the new TSTDP design presented in Azghadi *et al.* (2013c), cannot account for the frequency-dependent pairing experiments and is suitable only for experiments with high spike frequencies. Therefore, the weight capacitor in the design was increased, so that it can account for all types of experiments and is useful for processing inputs with either high or low spike rates. The presented results in the *2013 IEEE International Conference on Electronics, Circuits, and Systems* (Azghadi *et al.* 2013b), show that the new design with larger capacitor can very closely mimic the frequency-dependent pairing experiments, and is suitable for processing various input spike rates.

- Further investigations on the performance of the new design show that this design, which its synaptic plasticity performance is slightly compromised compared to the previous large high-performance circuit, significantly outperforms previous TSTDP and PSTDP designs, in all important aspects in neuromorphic engineering including power consumption, required silicon real estate, and tolerance to device mismatch. The results are presented in *PLoS ONE* (Azghadi *et al.* 2014a).

  All these features make the new proposed design an ideal device for use in large scale SNNs, which aim at implementing neuromorphic systems with an inherent capability that can adapt to a continuously changing environment, thus leading to systems with significant learning and computational abilities. This system then can be used in real-world tasks such as pattern classification.

- In addition to all these contributions, a general overview on the design and VLSI implementations of various synaptic plasticity rules, ranging from phenomenological ones (i.e. timing-based, rate-based, or hybrid rules) to biophysically realistic ones (e.g. based on calcium dependent models) is provided in this thesis.

The thesis also discusses the application domains, weaknesses and strengths of the various representative approaches proposed in the literature and provides deeper insight into the challenges that engineers face when designing and implementing synaptic plasticity rules in order to utilise them in real-world applications. Furthermore, counter approaches to tackle the challenges in designing neuromorphic engineering circuits are discussed and proposed. Besides, with a focus more on the system aspects of neuromorphic engineering, the use of various synaptic plasticity rules and circuits in real neuromorphic learning systems is discussed and these systems are analysed in terms of power consumption and silicon real estate. Also an example of an effective neuromorphic system is mentioned and it is described in detail how it learns to perform an engineering task.

The results, review and discussion mentioned above are mainly presented in *The Proceedings of the IEEE* (Azghadi *et al.* 2014b).

1.4 Thesis Outline

This thesis encompasses nine chapters as visualised in Fig. 1.1. In the current chapter, some introductory remarks on the neural networks, spiking neural networks, and neuromorphic engineering were addressed. In addition, discussions on why synaptic plasticity rules are crucial components of SNNs, were provided. Furthermore, we described the motivations of this thesis for exploring and investigating VLSI implementations of new, non-linear, and more complex synaptic plasticity rules, in order to use them in large-scale neuromorphic systems with the capability of performing real-world tasks such as pattern classification.

Chapter 2 provides a brief background on neuron and synapse models and their VLSI implementations. This will provide the required background material to build on in the thesis. It then focuses on the main topic of the thesis, synaptic plasticity rules. It reviews various plasticity rules ranging from simple phenomenological, to hybrid, to biophysically grounded rules, and compare them in terms of biological plausibility. The chapter also provides results of several Matlab simulations that were performed in the beginning of the present study to gain knowledge and better understanding of the targeted rules. These simulation were utilised as benchmark to verify the performance of the rules, when implemented in VLSI. The chapter also reviews some important

Figure 1.1. Thesis outline. The thesis is composed of 9 chapters including background and conclusion. The original contributions are distributed in chapters 3 to 8. The thesis can be divided into two parts. The first part that includes chapters 3, 4 and parts of chapter 5 is mainly concerned with the design of a neuromoprhic system. However, the second part, which includes some sections of chapter 5, and chapters 6 to 8, is dedicated to the circuit designs for synaptic plasticity rules. All chapters are virtually self-contained.

synaptic plasticity protocols that have been utilised in synaptic plasticity experiments, as well as in many experiments performed in the present thesis. In short, this chapter gives the reader an insight about the rest of the experiments and studies performed within the thesis.

Chapter 3 describes the architecture and structure of a programmable hybrid analog digital neuromorphic circuit, called IFMEM, that can be used to build compact low-power neural processing systems. In this chapter, first the architecture of the IFMEM neuromorphic system is described and then it is explained how this hybrid analog-digital Complementary Metal Oxide Semiconductor (CMOS) circuit operates correctly over a wide range of input frequencies; a feature that is essential for many applications. The chapter shows measurement results from available silicon neurons, and neuron-synapse combinations and demonstrates how required and specific behaviours are generated by programming the chip and optimising neural parameters of the silicon neurons and synapses. The provided information in this chapter helps to understand the presented results in the following chapter, which is dedicated to implementing STDP rules on the IFMEM setup, and utilising these rules to reproduce a rate-based BCM-like behaviour, and to carry out a classification task.

Chapter 4 is dedicated to the implementation of STDP rules and a pattern classification neural network on the IFMEM neuromorphic system. It is shown that both PSTDP and TSTDP rules can be implemented on this neuromorphic setup and demonstrate the expected behaviours seen in biological experiments. The chapter shows how the STDP window can be generated using the silicon neurons and synapses available on the system. Next, the STDP rule is used and tested for generating a competitive Hebbian learning behaviour observed in computational STDP experiments, which results in a bimodal synaptic weight distribution after STDP learning (Song *et al.* 2000). Then, the TSTDP learning algorithm is implemented. In order to test this implementation, the rate-based BCM learning behaviour is reproduced by this implementation. This experiment shows the usefulness of this timing-based learning algorithm for generating the rate-based BCM learning behaviour. Finally, the implemented TSTDP learning mechanism is utilised to train a simple feedforward spiking neural network to classify some complex rate-based patterns. Obtained results show the high performance of the TSTDP rule in the targeted classification task. In addition, the preformed research in this chapter provides good view of the TSTDP rule and its properties and features, which are essential when designing VLSI TSTDP synapses in the following chapters.

Chapter 5 reviews VLSI circuit implementations of various synaptic plasticity rules, ranging from phenomenological rules to biophysically realistic ones. It discusses the application domains, weaknesses and strengths of the various approaches proposed in the literature and provides deeper insight into the challenges that engineers face when designing and implementing synaptic plasticity rules in order to utilise them in real-world applications. The review performed in this chapter helps in the design process of new VLSI circuits for synaptic plasticity rules, since it highlights the challenges, applications and effective design methods and techniques, that are already known in VLSI implementations of specific rules.

Chapter 6 demonstrates that the VLSI implementations of the classical model of STDP is incapable of reproducing synaptic weight changes similar to those seen in biological experiments, which investigate the effect of either higher order spike trains (e.g. triplet and quadruplet of spikes), or simultaneous effect of the rate and timing of spike pairs on synaptic plasticity. This chapter shows that, a previously described spike triplet-based STDP rule succeeds in reproducing all of these synaptic plasticity experiments. In this chapter, synaptic weight changes using a number of widely used PSTDP circuits are investigated and it is shown how the class of PSTDP circuits fails to reproduce the mentioned complex biological experiments. In addition, a number of new STDP VLSI circuits, which act based on the timing among triplets of spikes and are able to reproduce all the mentioned experimental results, are presented. The presented circuits in this chapter are the first VLSI implementations of TSTDP rules that are capable of mimicking a wide range of synaptic plasticity experiments.

Chapter 7 introduces circuit design and implementation of a new high-performance VLSI design for the TSTDP rule that outperforms the previous TSTDP VLSI designs in several aspects. It is shown in this chapter, how different terms in the TSTDP synaptic plasticity equation, are implemented to have a very close fit to the model. This results in the proposed design to have significantly lower synaptic plasticity prediction error, in comparison with previous designs for TSTDP and PSTDP rules. In addition, it is shown that the new proposed design can successfully account for a number of experiments in addition to those mentioned for the previous TSTDP circuits, including those involved with various spike triplet combinations, as well as pre-synaptic and post-synaptic driven rate-based BCM-like experiments, where the previous TSTDP designs do not perform very efficiently. This chapter also discusses some of the main challenges in designing the proposed TSTDP circuit such as power consumption, silicon

real estate and process variations. In order to test the design against variation that leads to device mismatch, a 1000-run Monte Carlo (MC) analysis is conducted on the proposed circuit. The presented MC simulation analysis and the simulation result from fine-tuned circuits show that, it is possible to mitigate the effect of process variations in the proof of concept circuit. In addition to the performed simulations, the proposed circuit has been fabricated in VLSI as a proof of principle. The shown chip measurement results testify to the correct functionality of the fabricated circuit in performing triplet-based synaptic weight modification.

Chapter 8 introduces a new accelerated-time circuit that has several advantages over its previous neuromorphic counterparts, which were discussed in previous chapters, in terms of compactness, power consumption, and capability to mimic the outcomes of biological experiments. The proposed circuit is investigated and compared to other designs in terms of tolerance to mismatch and process variation. Monte Carlo (MC) simulation results show that the proposed design is much more stable than its previous counterparts in terms of vulnerability to transistor mismatch, which is a significant challenge in analog neuromorphic design. All these features make the proposed design an ideal circuit for use in large scale SNNs, which aim at implementing neuromorphic systems with an inherent capability that can adapt to a continuously changing environment, thus leading to systems with significant learning and computational abilities.

Chapter 9 provides concluding remarks of this thesis. It also discusses future research directions that can be followed based on the study carried out and presented in the current thesis. In addition, an outlook to the neuromorphic engineering for synaptic plasticity rules is also written in this chapter.

In a nutshell, considering the background material provided in this thesis and with regards to the versatile study performed on the design, implementation and application of spike timing-based synaptic plasticity rules, the current thesis can be of great help for readers with various backgrounds. It is useful for engineering students who want to grasp an idea around the field of neuromorphic engineering, as well as the more experienced neuromorphic engineers who need to review and learn about the VLSI implementation of synaptic plasticity rules and their applications. In addition, computational and experimental neuroscientists who would like to be familiar with the field of neuromorphic engineering and its relation with their fields of research, will find this thesis useful.

Chapter 2

Neurons, Synapses and Synaptic Plasticity

THIS chapter starts by providing a brief background on neuron and synapse models and their VLSI implementations, that is required to elucidate the material within this thesis. It then focuses on reviewing various synaptic plasticity rules ranging from simple phenomenological, to hybrid, to biophysically grounded rules, and compares them in terms of biological capabilities. The chapter provides several Matlab simulation results that were performed in the beginning of this study to gain the knowledge and better understanding of the targeted rules. This will provide the needed understanding required for implementing these rules in VLSI in the next parts of the study. The chapter also reviews some important synaptic plasticity protocols that have been utilised in synaptic plasticity experiments, as well as in many experiments performed in the present thesis. In short, this chapter gives the reader an insight of the neuron and synapse structures and provides the required information and terms that will be used in the experiments performed within this thesis.

2.1 Introduction

The basic building blocks of a SNN are neurons and synapses (Gerstner and Kistler 2002). There are various types of neurons that can be classified based on their shapes and biophysics. However, all types of cortical neurons produce electric signals—so called spikes or action potentials—and generally have a shape as demonstrated in Fig. 2.1.

In addition to the neuron, the synapse is another crucial building block of a SNN. Similar to neurons, synapses also have complex structures and behaviours. They are widely thought to be the essential components responsible for learning, memory and computational ability in the neural networks (Sjöström *et al.* 2008). A synapse, as shown in Fig. 2.1, is the contact apparatus between a pre-synaptic neuron's axon and a post-synaptic neuron soma or dendrite (see Fig. 2.1). As the figure shows, the synapse is the site for transmitting various neurotransmitter molecules to the post-synaptic neuron, through different receptors on the post-synaptic side. The underlying mechanisms of this transmission and the interactions happening in the synapse are termed synaptic plasticity rules, which are the focus of the present chapter.

This chapter is organised as follows. Section 2.2 briefly discusses the structure and behaviour of a typical neuron. Section 2.3 provides information on synapse and its structure, and also discusses various VLSI implementations of a synapse circuit. After these two background sections, Section 2.4 explains the synaptic plasticity phenomenon. Section 2.5 shows various synaptic plasticity experiments that were conducted in biology. These experiments were performed to gain an understanding of the synaptic plasticity mechanisms, in order to propose plasticity models. These synaptic plasticity models and rules are reviewed in Section 2.6. The chapter ends in Section 2.7 that includes concluding remarks on synaptic plasticity rules and sheds light on the future topics discussed in the thesis.

2.2 Spiking Neurons

Essentially, there are three types of ions in a neuron structure. An unbalanced amount of ions in and out of the neuron results in changing the input/output potential. The difference in the potential of input and output of the neuron leads to production of a spike. The spikes can be transmitted through a long tree-like structure that comes

out of the neuron body (soma) and is called an *axon*. The axon is connected to other neurons dendrites or soma. The synapse is where the axon of one neuron is connected to the dendrite or soma of another neuron—moreover, it possesses the apparatus for learning.

Currently, there is general agreement on the behaviour of biological neurons. Some conductance-based neuron models such as Hodgkin-Huxley (HH) model (Hodgkin and Huxley 1952), are able to account for almost all the behaviours, a cortical neuron exhibits. Through these models, one can investigate many biological features of real neurons and use them in various applications. Usually these conductance-based models are strongly related to the biology of neurons, dedicating one component and/or variable to implement each required component available in the neuron structure. For instance, the HH neuron model takes into account the dynamics of membrane potential, calcium (Ca), and potassium (K) currents as they occur in real neurons (Hodgkin and Huxley 1952). The main problem with this model, however, is its complex structure. This complexity is due to the intricate nature of a cortical neuron's operation. The cortical neuron can be also modelled in a simpler mathematical form, but with similar capabilities to that of the HH neuron. Examples of these models are the Izhikevich (2003) and Wilson (1999) models. In addition, previous studies show that simple neuron models such as Integrate and Fire (IF), with only a limited number of neuronal behaviours (Izhikevich 2004), are useful in simulating simplified neural systems.

In an interesting study, Izhikevich has reviewed spiking neuron models (Izhikevich 2004). He has firstly depicted 20 various behaviours that a spiking neuron can exhibit. These behaviours are shown in Fig. 2.2. Then, he reviewed 11 spiking neuron models and compared them in terms of complexity and biological plausibility.

Besides Izhikevich's neuron model, there are many other neuron models that describe the behaviour of a biological neuron by means of some equations and formulations. The most common models that can be found in the literature are HH (Hodgkin and Huxley 1952), Leaky Integrate and Fire (LIF) (Smith 2006), IF with Adaptation, Integrate-and-Fire-or-Burst (IFB) (Smith *et al.* 2000), Resonate-and-Fire model (Izhikevich 2001), Quadratic Integrate and Fire (QIF) or theta-neuron (Ermentrout 1996), Izhikevich (2003), FitzHugh (1961), Morris and Lecar (1981), Rose and Hindmarsh (1989), and Wilson (1999).

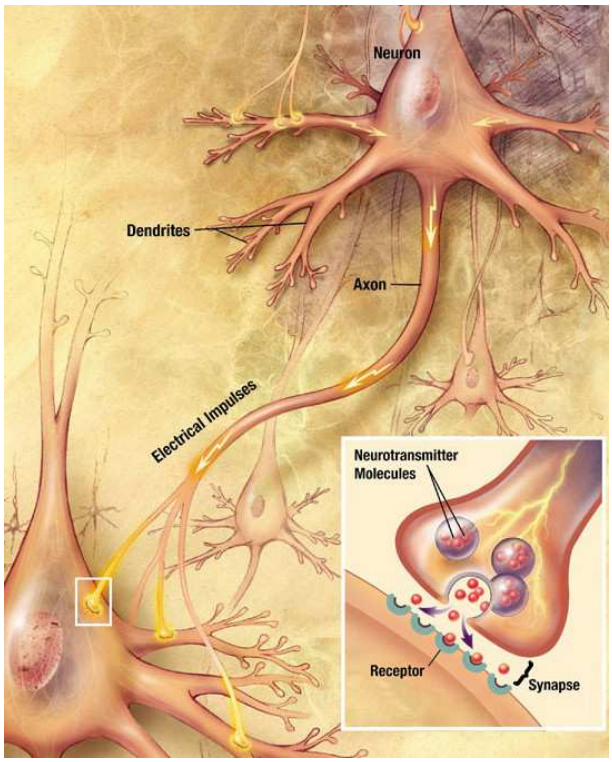


Figure 2.1. A spiking neuron and synapse construction. The figure shows various parts of a neuron and also magnifies the synapse structure, which in the case of this figure, is the connection point of pre-synaptic neuron's axon to the soma of the post-synaptic neuron. Source: US National Institutes of Health.

Since spiking neuron models can be modelled in terms of mathematical formulas, they can be fairly easily implemented in software. However, because neural systems operate in parallel, and ordinary computer systems are sequential, implementing a neuron model in software is slow. Present-day computers are very high performance, allowing straightforward simulation of a neuron in real-time. However, for a real-world application such as pattern classification, a significantly large number of these spiking neurons is needed. In addition, the software simulation will be very time consuming for a large-scale network of spiking neurons. Furthermore, utilising an expensive parallel computer to implement the required neural network system, is not economical since it requires specialised parallel software programs that are costly. Therefore, it is preferable to implement neurons in hardware, rather than via simulations. Generally, a hardware neuron, is much more practical than software neurons when considering a large-scale network of neurons for real-world applications (Smith 2006, Hamilton and van Schaik 2011, Indiveri *et al.* 2011). For a review of VLSI implementations of various neuron models, the reader is directed to Indiveri *et al.* (2011).

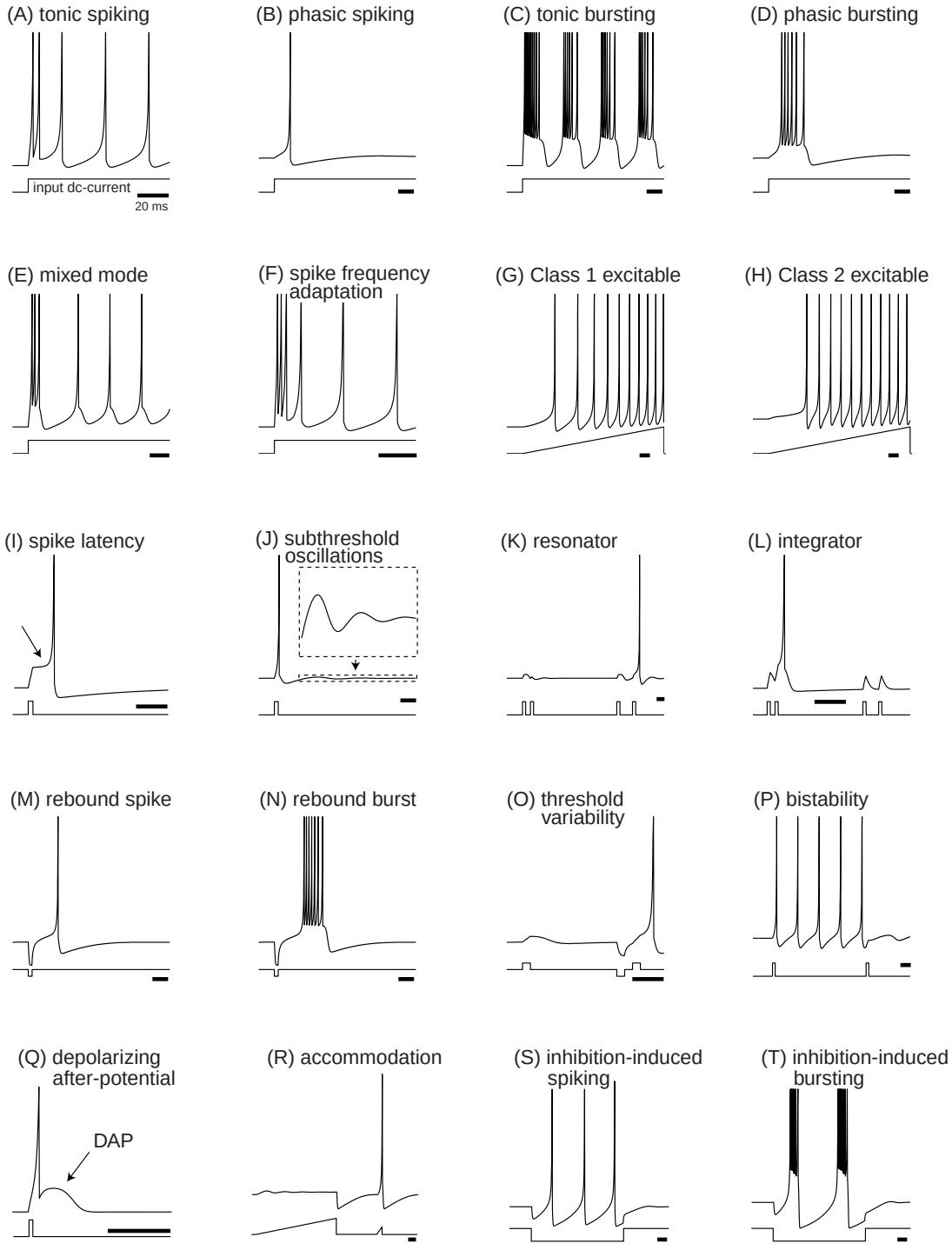


Figure 2.2. Summary of the neuro-computational properties of biological spiking neurons produced by Izhikevich model. The results shows 20 different behaviours a neuron can show under Izhikevich neuron model. The electronic version of the figure and reproduction permissions are freely available at www.izhikevich.com.

2.3 Synapses

Similar to neurons, synapses also have complex structures and behaviours. As already mentioned, they are widely thought to be the essential components responsible for learning, memory and computational ability in neural networks (Sjöström *et al.* 2008). It is believed that through some activity-dependent rules, which control some complex chemical reactions, synapses alter their efficacies (Abbott and Nelson 2000). This efficacy determines the strength and the depressing/potentiating effect, a synapse has on the spiking activity of its afferent neurons. Since neurons are associated to each other through synapses, and because they transfer data and information in the form of spikes, therefore it is absolutely essential to control the way, through which a synapse manages spiking behaviour of its post neurons.

There are different VLSI implementations for synapse. A synapse can be implemented as a simple multiplier circuit (Satyanarayana *et al.* 1992). Alternatively, a synapse can be a current source that conducts current to the post-synaptic neuron only in the duration of pre-synaptic spikes (Chicca *et al.* 2003). Here the amount of the current conveyed by the current source represents the synaptic weight. In addition to these simple implementations that do not consider the detail dynamics and behaviours of the Excitatory Post-Synaptic Current (EPSC) and Inhibitory Post-Synaptic Current (IPSC), some other VLSI implementations of synapse, take into account more detailed synaptic behaviours. A well-known example of these types of synaptic circuits is the Differential Pair Integrator (DPI) circuit proposed by Bartolozzi and Indiveri (2007). This implementation is able to account for short-term dynamics of synapse as well as reproducing the EPSC effects observed in biological synapses (Bartolozzi and Indiveri 2007).

Bartolozzi and Indiveri have reviewed various implementations of synaptic circuits in VLSI (Bartolozzi and Indiveri 2007). They have described synaptic circuits such as pulsed current-source synapse (Mead 1989), linear charge and discharge synapse (Arthur and Boahen 2004), current mirror integrator synapse (Hynna and Boahen 2001), and log-domain integrator synapse (Merolla and Boahen 2004).

In addition to these circuits, usually a silicon synapse can be implemented as a simple integrator that only injects synaptic charge to the post-synaptic neurons membrane potential—for example see Bofill-I-Petit and Murray (2004). More directly, the main duty of synaptic circuits is to control the post-synaptic current that is conveyed to the post-synaptic neuron. The magnitude of this current is determined by the synaptic plasticity circuit of the synapse that needs to be implemented separately. This synaptic

plasticity circuit functions according to a specific synaptic plasticity rule, and therefore can vary in different synaptic circuit implementations and applications. The focus of this thesis is on the design and implementation of these synaptic plasticity circuits. However, before discussing the circuit implementations, the synaptic plasticity rules are discussed and reviewed to help understanding the design and implementation of the targeted synaptic plasticity circuits.

2.4 Spiking Neurons Synaptic Plasticity

As already mentioned, there exists a significant number of hypotheses that try to approximate synaptic efficacy alterations (Mayr and Partzsch 2010). These hypotheses that govern the synaptic weight changes, are so-called synaptic plasticity models (rules). Identical to neuron models, there are a variety of synaptic plasticity models, some of which are closer to biology and have meaningful relationships to biological synapses, therefore, they are complex. On the other hand, some other models only approximate a number of biological experiments via mathematical modelling, and hence they are simpler than the former group. Generally, the main purpose of the second group of synaptic plasticity rules is to propose effective and simple rules, which are able to produce the outcomes of as many synaptic plasticity experiments as possible.

In this chapter, a number of important synaptic plasticity rules are reviewed and highlighted and their abilities in reproducing the result of various biological experiments are compared while discussing their complexities and structures. In order to have a fair comparison among various synaptic plasticity rules, these rules are compared from two aspects. The first aspect is their strength in reproducing various synaptic plasticity experiments, while the second aspect is their simplicity and suitability to be employed in large-scale neural simulations, and/or large-scale hardware realisations. Therefore, prior to investigating and reviewing various synaptic plasticity rules, a variety of biological experimental protocols that were used in the experiments performed in the neocortex have been reviewed, in order to provide the reader with an understanding of under which protocols and conditions various synaptic plasticity rules are simulated and compared. In the following sections, first some important synaptic plasticity protocols are reviewed and their structures are described. And second, some significant synaptic plasticity models are reviewed and their structures and various synaptic plasticity abilities are highlighted.

2.5 Synaptic Plasticity Experiments

In order to learn about synaptic plasticity mechanisms, experimentalists utilise specific experimental protocols, to stimulate pre- and/or post-synaptic neurons and induce and then measure alterations in the efficacy of synapses, and their ion dynamics. Understanding these alterations, with respect to the activity of the pre- and post-synaptic neurons and their dynamics, sheds light on how neurons activity affects synaptic plasticity and bring about Long Term Potentiation (LTP) and Long Term Depression (LTD). This understanding then, helps neuroscientists to approximate the behaviour of the synapse with some mathematical equations and it helps them implement a detailed model of synaptic plasticity with relation to neuronal activities.

These mathematical equations or detailed models are usually an approximation of the biological experiments. In order to have a measure, to test the efficiency of a model or a circuit in replicating the experiments, one can define an error function that represents the amount of difference among the weight changes predicted by a candidate model or circuit, and those happened in biological experiments. An instance of an error function for verifying a plasticity model, is the Normalised Mean Square Error (NMSE) function proposed and utilised in Pfister and Gerstner (2006). The NMSE is calculated using the following equation:

$$\text{NMSE} = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{\Delta w^i_{\text{exp}} - \Delta w^i_{\text{model}}}{\sigma_i}\right)^2, \tag{2.1}$$

where Δw^i_{exp}, $\Delta w^i_{\text{model}}$ and σ_i are the mean weight change obtained from biological experiments, the weight change obtained from the model or circuit under consideration, and the standard error mean of Δw^i_{exp} for a given data point i, respectively. Here, p represents the number of data points in a the data set under consideration. In order to minimise the resulting NMSEs for the model/circuit and fit their output to the experimental data, there is a need to adjust the model parameters or circuit bias parameters and time constants. This is an optimisation process of the model parameters or circuit biases to reach a minimum NMSE value and so the closest possible fit to the experimental data. A powerful synaptic plasticity model/circuit, therefore, closely mimics the outcomes of a variety of biological experiments, and reaches a minimal NMSE. Hence, the number of various synaptic plasticity experiments a single model can account for is a good measure and an indication of the model/circuit's ability to mimic biology. In the following, some of these synaptic plasticity experiments are reviewed.

These experiments have been utilised throughout this thesis, to first investigate a number of recognised biological models, and then verify the functionality and strength of the VLSI implementations of various synaptic plasticity rules.

2.5.1 Pairing Protocol

The pair-based STDP protocol has been extensively used in electrophysiological experiments and simulation studies (Bi and Poo 1998, Iannella *et al.* 2010). In this protocol, 60 pairs of pre- and post-synaptic spikes with a delay of $\Delta t = t_{\text{post}} - t_{\text{pre}}$, as shown in Fig. 2.3, are conducted with a repetition frequency of ρ Hz (in many experiments $\rho = 1$ Hz). This protocol has been utilised in experiments reported in Bi and Poo (1998), Froemke and Dan (2002), and Wang *et al.* (2005) and also have been employed in simulations and circuit designs for synaptic plasticity such as Bofill-I-Petit and Murray (2004), Indiveri *et al.* (2006), and Azghadi *et al.* (2011c).

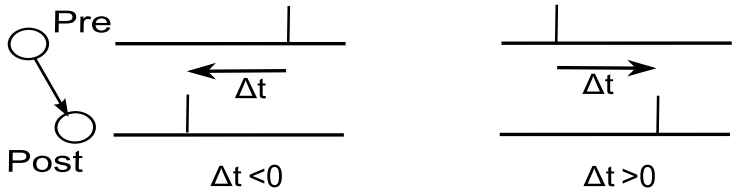


Figure 2.3. Spike pairing protocol. The figure shows how pairs of pre- and post-synaptic spikes in a pairing protocol are timed for reproducing an STDP window.

2.5.2 Frequency-dependent Pairing Protocol

In the simple pairing protocol, the repetition frequency of spike pairs kept constant and it is usually 1 Hz. However, it has been illustrated in Sjöström *et al.* (2001) that altering the pairing repetition frequency, ρ, affects the total change in weight of the synapse. The spike pairing under this protocol is shown in Fig. 2.4. It shows that in higher pairing frequencies, the order of pre-post or post-pre spike pairs does not matter and both cases will lead to LTP. However, in lower pairing frequencies, pre-post results in LTP and post-pre combination results in LTD (Sjöström *et al.* 2001, Sjöström *et al.* 2008).

2.5.3 Triplet Protocol

There are two types of triplet patterns that are used in the hippocampal experiments performed in Wang *et al.* (2005). These triplet patterns are adopted in this thesis to

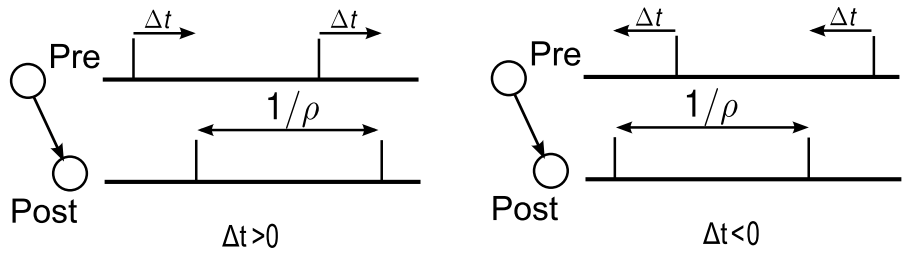


Figure 2.4. Frequency-dependent pairing protocol. The figure shows how pairs of pre- and post-synaptic spikes in a frequency-dependent pairing protocol are conducted. Here, ρ determines the repetition frequency, at which the pre-post ($\Delta t > 0$) or post-pre ($\Delta t < 0$) spike pair arrives.

compute the synaptic weight prediction error as described in Pfister and Gerstner (2006). Both of these patterns consist of 60 triplets of spikes that are repeated at a given frequency of $\rho = 1$ Hz. These triplet patterns are shown in Fig. 2.5. The first pattern is composed of two pre-synaptic spikes and one post-synaptic spike in a pre-post-pre configuration. As a result, there are two delays between the first pre and the middle post, $\Delta t_1 = t_{\text{post}} - t_{\text{pre1}}$, and between the second pre and the middle post $\Delta t_2 = t_{\text{post}} - t_{\text{pre2}}$. The second triplet pattern is analogous to the first but with two post-synaptic spikes, one before and the other one after a pre-synaptic spike (post-pre-post). Here, timing differences are defined as $\Delta t_1 = t_{\text{post1}} - t_{\text{pre}}$ and $\Delta t_2 = t_{\text{post2}} - t_{\text{pre}}$.

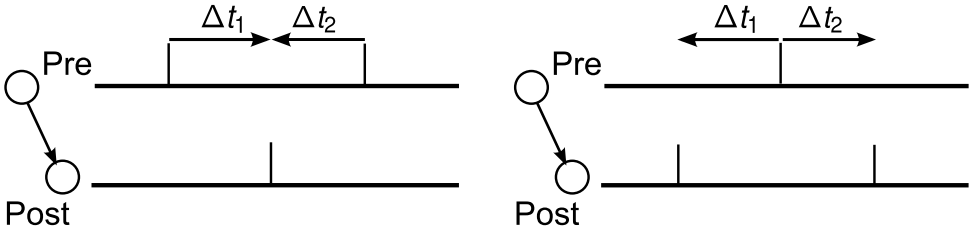


Figure 2.5. Triplet protocol. The figure shows how triplets of pre- and post-synaptic spikes in a triplet protocol are conducted.

2.5.4 Extra Triplet Protocol

In addition to the aforementioned triplet protocol employed in Pfister and Gerstner (2006), which considers only two combinations of spike triplets, there are other combinations (rather than pre-post-pre or post-pre-post) of spike triplets that have not been explored in Pfister and Gerstner (2006), but have been used in another set of multi-spike interaction experiments performed by Froemke and Dan (2002). The experimental triplet protocol as described in Froemke and Dan (2002) is as follows; a third spike

is added either pre- or post-synaptically to the pre-post spike pairs, to form a triplet. Then this triplet is repeated 60 times at 0.2 Hz to induce synaptic weight changes. In this protocol, there are two timing differences shown as $\Delta t_1 = t_{\text{post}} - t_{\text{pre}}$, which is the timing difference between the two most left pre-post or post-pre spike pairs, and $\Delta t_2 = t_{\text{post}} - t_{\text{pre}}$, which is the timing difference between the two most right pre-post or post-pre spike pairs. Fig. 2.6 demonstrates different combinations of these spike triplets.

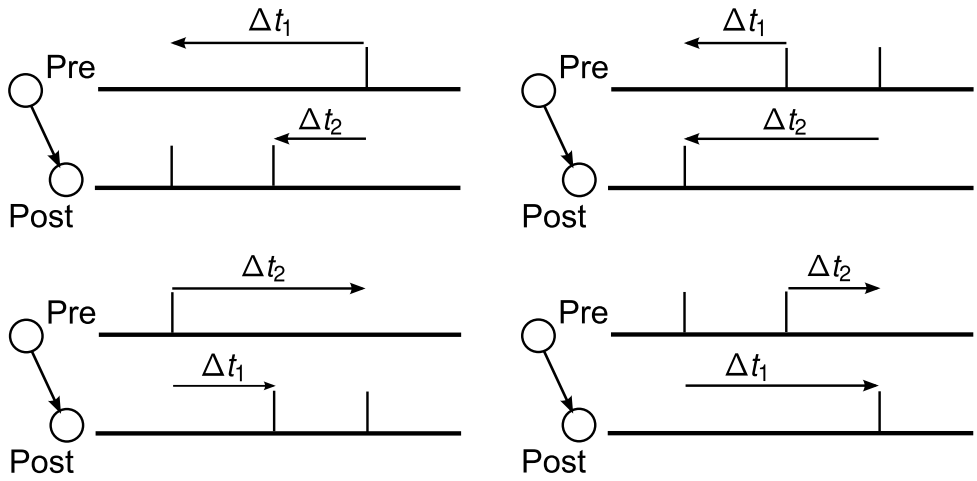


Figure 2.6. Triplet protocol for extra triplet patterns. The figure shows how extra triplets of pre- and post-synaptic spikes in the triplet protocol are timed.

2.5.5 Quadruplet Protocol

This protocol is composed of 60 quadruplets of spikes repeated at frequency of $\rho = 1$ Hz. The quadruplet is composed of either a post-pre pair with a delay of $\Delta t_1 = t_{\text{post1}} - t_{\text{pre1}} < 0$ precedes a pre-post pair with a delay of $\Delta t_2 = t_{\text{post2}} - t_{\text{pre2}} > 0$ with a time $T > 0$, or a pre-post pair with a delay of $\Delta t_2 = t_{\text{post2}} - t_{\text{pre2}} > 0$ precedes a post-pre pair with a delay of $\Delta t_1 = t_{\text{post1}} - t_{\text{pre1}} < 0$ with a time $T < 0$, where $T = (t_{\text{pre2}} + t_{\text{post2}})/2 - (t_{\text{pre1}} + t_{\text{post1}})/2$. The quadruplet patterns are shown in Fig. 2.7. Identical to Pfister and Gerstner (2006), in all quadruplet experiments in this thesis, $\Delta t = -\Delta t_1 = \Delta t_2 = 5\ \mu$s.

2.5.6 Poissonian Protocol

In order to test the ability of the targeted timing-based plasticity rules and timing-based synaptic plasticity circuits in generating a rate-based learning rule, which mimics the

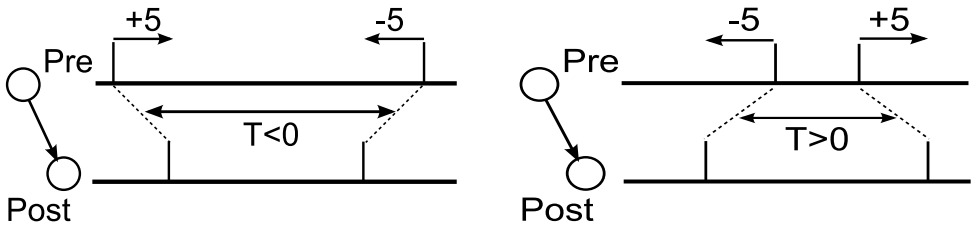


Figure 2.7. Qudruplet protocol. The figure shows how quadruplet of pre- and post-synaptic spikes in the qudruplet protocol are timed.

effects of BCM rule, a Poissonian rate-based experimental protocol is also employed. Under this protocol, the pre-synaptic and post-synaptic spike trains are generated as Poissonian spike trains with firing rate of ρ_{pre} and ρ_{post}, respectively. In these spike trains the Inter Spike Interval (ISI), are random Poissonian values. This is the same protocol that has been used in Pfister and Gerstner (2006) to show how their proposed TSTDP model can show a close mapping to the BCM model. In this thesis we utilise a similar protocol to stimulate any VLSI circuit of interest, and examine if they are capable of reproducing a BCM-like behaviour.

2.6 Synaptic Plasticity Rules

There exist a variety of synaptic plasticity models (Mayr and Partzsch 2010). Generally, these models can be classified into two fundamental groups. The first group presents abstract models, while the second consists of rules that are closer to the biophysics of the synapse. Abstract (phenomenological) models of synaptic plasticity are aimed at demonstrating how the concept of synaptic plasticity can contribute to different forms of learning and memory. On the other hand, biophysical models of synaptic plasticity are based on actual cellular and molecular mechanisms observed in synapses and are intended to demonstrate how synaptic plasticity can arise from real biological mechanisms. More specifically, the abstract rules are usually phenomenological models that aim at mimicking the outcome of plasticity but do not aim at implementing the details of the molecular mechanisms underlying synaptic plasticity in the synapse. On the other hand, biophysically grounded rules are based on implementing crucial elements underlying plastic change in a one-to-one analogy to the detailed kinetics of the synapse.

Abstract models, themselves can be classified into three various groups including: i) spike timing-based models such as STDP (Song *et al.* 2000), ii) spike rate-based models

such as Bienenstock-Cooper-Munro (BCM) rule (Bienenstock *et al.* 1982, Cooper *et al.* 2004), and iii) hybrid spike time- and rate-based models that usually consider other variables including membrane voltage in the weight update process. Examples of these hybrid models are spike driven plasticity (Brader *et al.* 2007), voltage-based STDP (Clopath and Gerstner 2010), and voltage-based BCM (Mayr and Partzsch 2010). A summary of various synaptic plasticity models can be found in Mayr and Partzsch (2010).

Although there are a variety of synaptic plasticity rules and experiments, this chapter will only investigate some of the important biophysical and phenomenological rules, which either have been already designed in VLSI, or are more suited for implementation in VLSI.

2.6.1 Phenomenological Rules

Pair-based STDP

The pair-based rule is the classical description of STDP, which has been widely used in various computational studies (Song *et al.* 2000, Iannella and Tanaka 2006, Iannella *et al.* 2010) as well as several VLSI implementations (Bofill-I-Petit and Murray 2004, Cameron *et al.* 2005, Indiveri *et al.* 2006, Tanaka *et al.* 2009, Mayr *et al.* 2010, Meng *et al.* 2011, Bamford *et al.* 2012b). The original rule expressed by Eq. 2.2 is a mathematical representation of the pair-based STDP (PSTDP) rule (Song *et al.* 2000)

$$\Delta w = \begin{cases} \Delta w^+ = A^+ \mathrm{e}^{\left(\frac{-\Delta t}{\tau_+}\right)} & \text{if } \Delta t > 0 \\ \Delta w^- = -A^- \mathrm{e}^{\left(\frac{\Delta t}{\tau_-}\right)} & \text{if } \Delta t \leq 0, \end{cases} \tag{2.2}$$

where $\Delta t = t_{\text{post}} - t_{\text{pre}}$ is the timing difference between a single pair of pre- and post-synaptic spikes. According to this model, the synaptic weight will be potentiated if a pre-synaptic spike arrives in a specified time window (τ_+) before the occurrence of a post-synaptic spike. Analogously, depression will occur if a pre-synaptic spike occurs within a time window (τ_-) after the post-synaptic spike. The amount of potentiation/depression will be determined as a function of the timing difference between pre- and post-synaptic spikes, their temporal order, and their relevant amplitude parameters (A^+ and A^-). The conventional form of STDP learning window, which is generated using Eq. 2.2 is shown in Fig. 2.8.

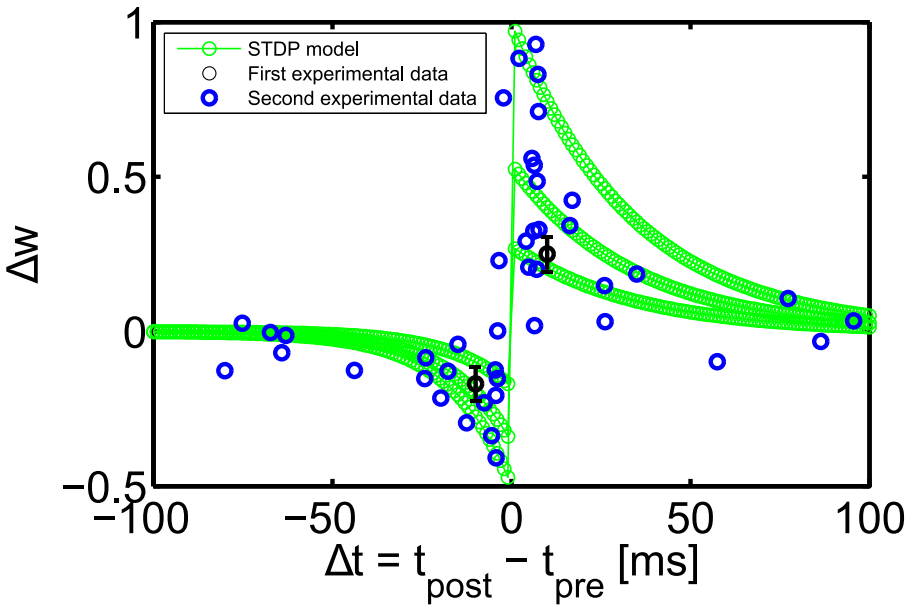


Figure 2.8. STDP learning window. Simulation results are produced under pairing protocol, and using pair-based STDP model. The synaptic parameters for generating the three shown curves (windows) using numerical simulations are as follows: $\tau_- = 16.8$ ms, and $\tau_+ = 33.7$ ms kept fixed for all three simulations, while the amplitude parameters, A^-, A^+ were altered for each simulation as $A^- = 0.5$, and $A^+ = 1$ for the first graph with the maximum $\Delta w = 1$ and the minimum $\Delta w = -0.5$; $A^- = 0.36$, and $A^+ = 0.54$ for the middle graph; and $A^- = 0.18$, and $A^+ = 0.276$ for the third graph with the maximum $\Delta w = 0.276$ and the minimum $\Delta w = -0.18$. The first experimental data shown in black are two data points with their standard deviations that are extracted from Pfister and Gerstner (2006), and the second experimental data are the normalised data extracted from Bi and Poo (1998).

Triplet-based STDP

In this model of synaptic plasticity, changes to synaptic weight are based on the timing differences among a triplet combination of spikes (Pfister and Gerstner 2006). Therefore, compared to the pair-based rule, this rule uses higher order temporal patterns of spikes to modify the weights of synapses. Triplet STDP (TSTDP) is described by

$$\Delta w = \begin{cases} \Delta w^+ = A_2^+ \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} + A_3^+ \mathrm{e}^{(\frac{-\Delta t_2}{\tau_y})} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} \\ \Delta w^- = -A_2^- \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})} - A_3^- \mathrm{e}^{(\frac{-\Delta t_3}{\tau_x})} \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})}, \end{cases} \tag{2.3}$$

where the synaptic weight is potentiated at times when a post-synaptic spike occurs and is depressed at the time when a pre-synaptic spike occurs. The potentiation and depression amplitude parameters are A_2^+, A_2^-, A_3^+ and A_3^-, while, $\Delta t_1 = t_{\mathrm{post(n)}} -$

$t_{\text{pre(n)}}$, $\Delta t_2 = t_{\text{post(n)}} - t_{\text{post(n-1)}} - \epsilon$ and $\Delta t_3 = t_{\text{pre(n)}} - t_{\text{pre(n-1)}} - \epsilon$, are the time differences between combinations of pre- and post-synaptic spikes. Here, ϵ is a small positive constant, which ensures that the weight update uses the correct values occurring just before the pre- or post-synaptic spike of interest, and finally τ_-, τ_+, τ_x and τ_y are time constants (Pfister and Gerstner 2006).

Theoretically, TSTDP rules were proposed to overcome deficiencies in the traditional pair-based STDP in being unable to reproduce the experimental outcomes of various physiological experiments such as the data generated by pairing frequency experiments performed in the visual cortex (Sjöström *et al.* 2001), or triplet, and quadruplet spike experiments performed in Wang *et al.* (2005). The main advantage of synaptic plasticity rules based upon higher order spike patterns over pair-based rules is the fact that contributions to the overall change in efficacy of traditional additive pair-based rules is essentially linear, while for higher order rules the underlying potentiation and depression contributions do not sum linearly. This fact was reported by Froemke and Dan (2002), where they show that there are non-linear interactions occurring between consecutive spikes during the presentation of higher order spike patterns. It is this underlying non-linearity that is captured in such higher order spike-based STDP rules—but is clearly lacking in pair-based STDP—and is believed to accurately model such non-linear interaction among spikes. Later, we review the suppressive STDP rule proposed by Froemke and Dan (2002), in order to discuss the mentioned non-linearity. Below some simulation results are shown that demonstrate how a minimised version of the full TSTDP rule, which is shown in Eq. 2.3, can approximate a number of biological experiments including quadruplet, triplet and STDP window experiments outcomes (Pfister and Gerstner 2006).

According to the numerical simulation results presented in Pfister and Gerstner (2006), beside the full TSTDP rule shown in Eq. 2.3, a minimised version of TSTDP rule that excludes the depression contribution of triplet of spikes, is also capable of reproducing the outcomes of several synaptic plasticity experiments. This rule that is called the first minimal TSTDP rule, is presented as

$$\Delta w = \begin{cases} \Delta w^+ = A_2^+ \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} + A_3^+ \mathrm{e}^{(\frac{-\Delta t_2}{\tau_y})} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} \\ \Delta w^- = -A_2^- \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})}. \end{cases} \tag{2.4}$$

Figure 2.9 demonstrates the results for quadruplet experimental protocol and shows how the first minimal TSTDP learning rule shown in Eq. 2.4 can generate an approximation of the quadruplet experimental results. In order to reach a good approximation of the quadruplet experiments, there is a need to optimise the six synaptic parameters shown in Eq. 2.4. These six parameters are optimised in a way that the minimal NMSE is reached for a number of experiments including pairing (window), triplet and quadruplet experiments, when compared to biological experiments. The optimised synaptic parameters that are utilised to approximate these experiments are those shown for minimal TSTDP rule in Table 2.1. These parameters are optimised in a way that approximate 13 specific synaptic weight change values shown in black data points and standard error mean bars, with minimum possible NMSE. The 13 data points include (i) three points on Fig. 2.9, (ii) eight data points in Fig. 2.10 (a) and (b), and (iii) two data points in Fig. 2.11.

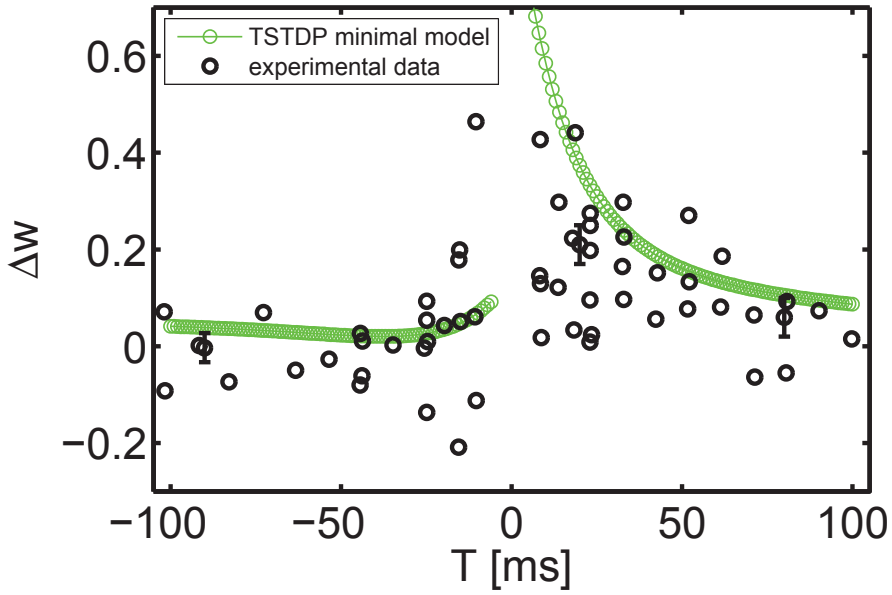


Figure 2.9. Quadruplet experiment in the hippocampus can be approximated using the first minimal TSTDP model. Simulation results are produced under quadruplet protocol. The minimal TSTDP model parameters for generating the shown synaptic weight changes are listed in Table 2.1. The experimental data shown in black are extracted from Pfister and Gerstner (2006).

Figure 2.10 shows how the first minimal TSTDP rule using the optimised synaptic parameters can approximate triplet STDP experiments. Besides, using the same optimised parameters, as those utilised for quadruplet and triplet experiments, the STDP

Table 2.1. Optimised minimal TSTDP model parameters. These parameters have been extracted from Pfister and Gerstner (2006) and used in our numerical simulations to generate the results shown in Fig. 2.9 to Fig. 2.13. Note that these parameters have been used for a minimal nearest-neighbour TSTDP model. In this table, those values that are shown as 'x', do not have any affect in the results.

Parameter name	A_2^+	A_2^-	A_3^+	A_3^-	τ_+ (ms)	τ_- (ms)	τ_y (ms)	τ_x (ms)	NMSE
First Minimal TSTDP	4.6×10^{-3}	3×10^{-3}	9.1×10^{-3}	0	16.8	33.7	48	x	2.9
2nd Minimal TSTDP	0	8×10^{-3}	5×10^{-2}	0	16.8	33.7	40	x	0.34

learning window, which is shown in Fig. 2.11 can be approximated using the first minimal TSTDP rule. The minimum NMSE that was achieved using the optimised parameters for these three experiments was equal to 2.9 as shown in Table 2.1.

In addition to the capability of simultaneously approximation of triplet, quadruplet and STDP window experiments with the same set of synaptic parameters, another minimal version of TSTDP rule (second minimal rule), is also capable of reproducing the results of the frequency-dependent pairing experiments performed in the visual cortex (Sjöström *et al.* 2001). The second minimal TSTDP model can be shown as

$$\Delta w = \begin{cases} \Delta w^+ = A_3^+ \mathrm{e}^{(\frac{-\Delta t_2}{\tau_y})} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} \\ \Delta w^- = -A_2^- \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})}, \end{cases} \tag{2.5}$$

which is simpler and takes lower number of synaptic parameters, and therefore needs a new set of parameters, in comparison with the previous minimal model for hippocampal experiments. The optimised parameters for generating a close approximation of the frequency-dependent pairing experiments are shown in the third row of Table 2.1. The experimental data and the approximated weight changes, which are computed by the second TSTDP minimal model, are shown in Fig. 2.12.

Besides the ability of reproducing timing-based experiments, which were discussed so far, the TSTDP rule has the capability to demonstrate a BCM-like behaviour. The BCM learning rule is an experimentally verified (Dudek and Bear 1992, Wang and Wagner 1999) spike rate-based synaptic plasticity rule, proposed in 1982 (Bienenstock *et al.* 1982). Unlike STDP, which is spike-timing based learning rule, synaptic modifications resulting from the BCM rule depends on the rate of the pre- and post-synaptic spikes (Bienenstock *et al.* 1982). In fact, it depends linearly on the pre-synaptic, but non-linearly on the post-synaptic neurons spiking activity. The mathematical model

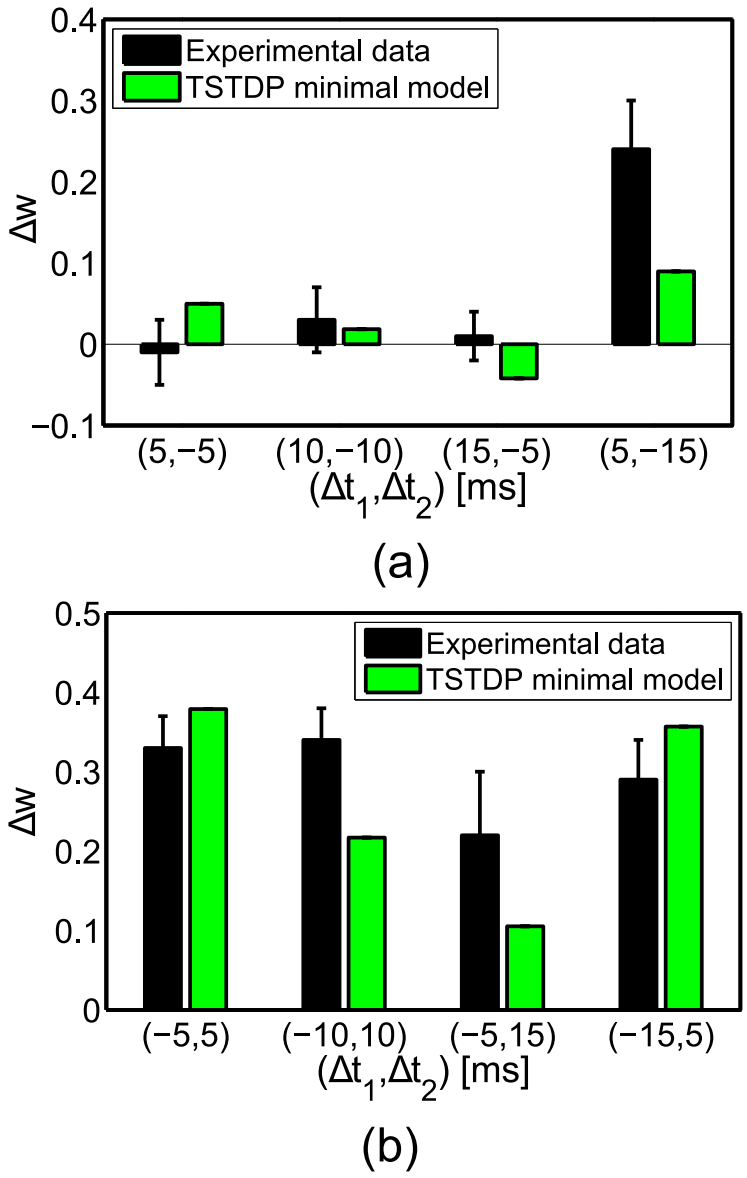


Figure 2.10. Triplet experiments in the hippocampus can be approximated using the first minimal TSTDP model. Simulation results are produced under triplet protocol presented in Pfister and Gerstner (2006). The minimal TSTDP model parameters for generating the shown synaptic weight changes are listed in Table 2.1. The experimental data, shown in black and their standard deviations are extracted from Pfister and Gerstner (2006).

of the BCM learning rule has been demonstrated in different ways, but a general, yet simple form of this model is given as,

$$\frac{\Delta w}{\Delta t} = \phi(\rho_{\text{post}}, \theta) \cdot \rho_{\text{pre}}, \tag{2.6}$$

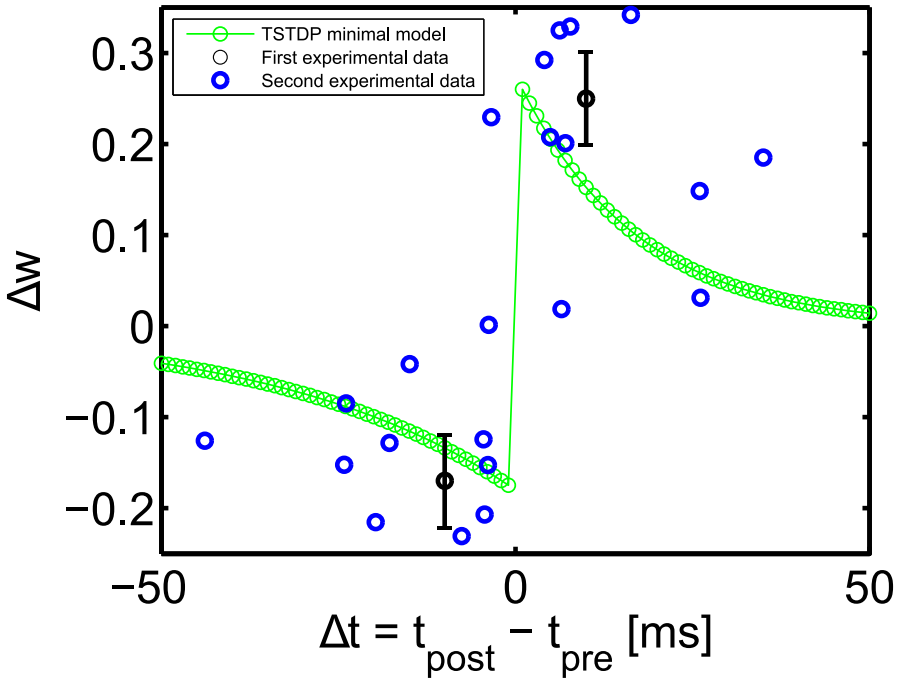


Figure 2.11. STDP learning window experiment in the hippocampus can be approximated using the first minimal TSTDP model. Simulation results are produced under pairing protocol. The minimal TSTDP model parameters for generating the shown synaptic weight changes are listed in Table 2.1. The first experimental data shown in black are two data points with their standard deviations that are extracted from Pfister and Gerstner (2006), and the second experimental data are the normalised experimental data that are extracted from Bi and Poo (1998).

where ρ_{pre} and ρ_{post} represent the pre-synaptic and post-synaptic neurons spiking activities and θ is a constant that represents some threshold (Pfister and Gerstner 2006). In addition, when $\phi(\rho_{\text{post}} < \theta, \theta) < 0$ synaptic weight will be decreased (depression), and when $\phi(\rho_{\text{post}} > \theta, \theta) > 0$, it will be increased (potentiation) and if $\phi(0, \theta) = 0$, there will be no change in synaptic weight (Pfister and Gerstner 2006).

According to the literature, the BCM rule can emerge from pair-based and triplet-based STDP rules. In 2003, Izhikevich and Desai (2003) demonstrated that, the nearest-spike interaction[1] version of PSTDP can replicate BCM-like behaviour. Furthermore, Pfister and Gerstner (2006) have reported, a triplet-based model of STDP that can also produce BCM behaviour, when long-time spike statistics are taken into account. According to

[1]Nearest-spike model considers the interaction of a spike only with its two immediate succeeding and immediate preceding nearest neighbours.

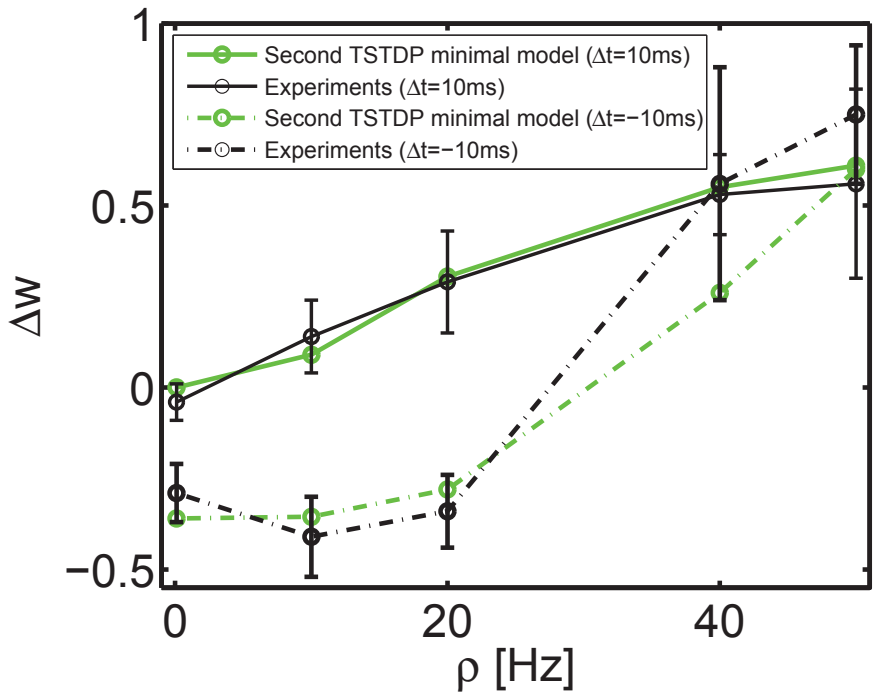


Figure 2.12. Pairing frequency experiments in the visual cortex can be approximated using the second minimal TSTDP model. Simulation results are produced under frequency-dependent pairing protocol and using the second TSTDP minimal model. The synaptic parameters for generating the shown weight changes are listed in the last row of Table 2.1. The experimental data, shown in black and their standard deviations are extracted from Pfister and Gerstner (2006).

Izhikevich and Desai (2003), under the assumption of Poissonian distribution of spike times for pre-synaptic and post-synaptic spike trains, nearest-spike pair-based STDP can give rise to the BCM rule; i.e. BCM emerges from nearest neighbour pair-based STDP; while all-to-all[2] spike interaction cannot. Furthermore, based on Pfister and Gerstner (2006), if the pre-synaptic and post-synaptic spike trains in a triplet-based STDP model are Poissonian spike trains, then BCM learning is an emergent property of the model. A mapping between BCM learning rule and the TSTDP learning rule is shown in Pfister and Gerstner (2006).

To analyse how BCM-like behaviour emerges from TSTDP, we need to go through the same analysis used by Pfister and Gerstner (2006). In this circumstance, the triplet

[2] All-to-all model considers the interaction of every single spike with all other spikes, not only with its nearest neighbours.

learning rule can be recast into a simpler form by considering the statistical properties of TSTDP weight changes which leads to the following time averaged equation,

$$\begin{aligned}\left\langle \frac{dw}{dt} \right\rangle &= -A_2^- \tau_- \rho_{\text{pre}} \rho_{\text{post}} + A_2^+ \tau_+ \rho_{\text{pre}} \rho_{\text{post}} \\ &\quad -A_3^- \tau_- \tau_x \rho_{\text{pre}}^2 \rho_{\text{post}} + A_3^+ \tau_+ \tau_y \rho_{\text{post}}^2 \rho_{\text{pre}},\end{aligned} \tag{2.7}$$

where ρ_{pre} and ρ_{post} are the pre- and post-synaptic mean firing rates, respectively. The other parameters in the above equation τ_-, and τ_+, are time constants for the pair-based contribution and τ_x, and τ_y are the corresponding time constants for the triplet-based contribution of the original triplet learning rule shown in Eq. 2.3.

Based on the rate-based BCM rule, synaptic weight change is linearly dependent on ρ_{pre} and non-linearly depends on ρ_{post} (see Eq. 2.6). In order to satisfy this condition in Eq. 2.7, A_3^- must be equal to zero and also $\rho_{\text{pre}} \ll \tau_+^{-1}$. This is a minimal case of the triplet-based STDP model—please refer to Pfister and Gerstner (2006). Also, based on the BCM learning rule definition, the synaptic weight modification threshold is a function of post-synaptic activity, i.e. $\theta = \alpha \left\langle \rho_{\text{post}}^p \right\rangle$ where $p > 1$. For triplet-based STDP, consider the case where all-to-all interactions between triplets of pre- and post-synaptic spikes; it is possible to redefine A_2^-, A_2^+ and A_3^+ in a way that the threshold be dependent on the post-synaptic firing rate, ρ_{post}^p. However, in the nearest-spike model it is not possible to change these parameters in a way to satisfy $\theta = \alpha \left\langle \rho_{\text{post}}^p \right\rangle$. Although the triplet-based nearest-spike STDP model cannot fully satisfy the second condition of a BCM learning rule (the dependency of threshold on ρ_{post}), it can elicit the properties of BCM for a limited range of frequencies. Numerical simulation results (Fig. 2.13) show how the threshold is modulated by controllable amplitude parameters (A_2^- , A_2^+ and A_3^+) for nearest spike interaction. For further details on the relation between the TSTDP and the BCM rules, refer to the text and also supplementary materials of Pfister and Gerstner (2006). Fig. 2.13 demonstrates how the second minimal TSTDP rule, with the parameters shown in Table 2.1 for visual cortex experiments, produces a BCM-fashion behaviour.

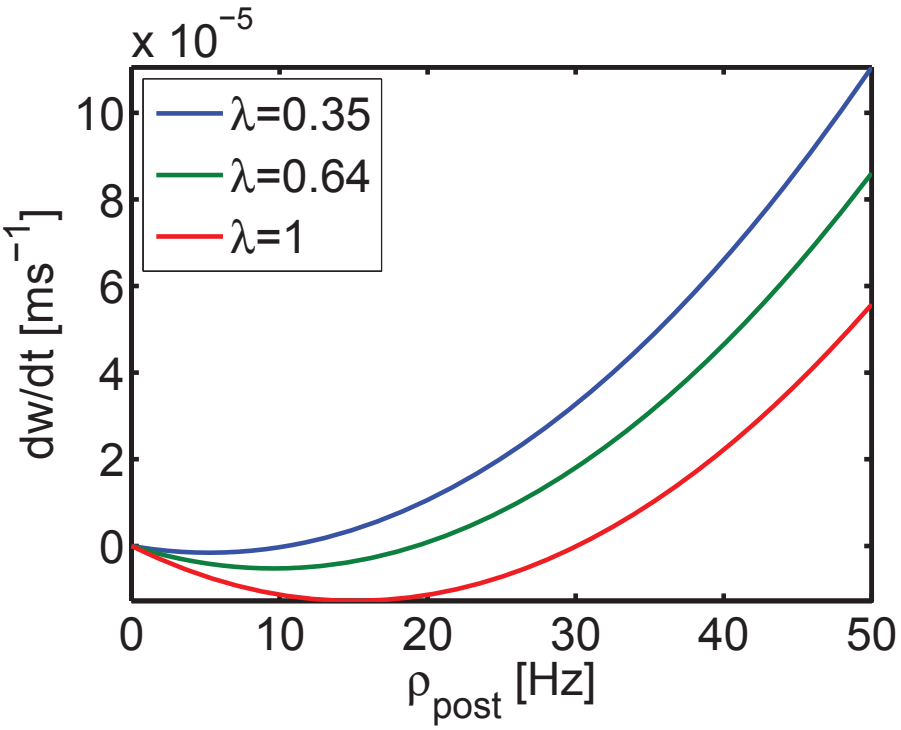


Figure 2.13. The BCM learning rule can be mapped to the TSTDP learning model. The results shown in this figure are produced under Poissonian protocol and using the second minimal TSTDP model, as the one used for generating the results in Fig. 2.12. The synaptic parameters for generating the weight changes shown in this figure are those for visual cortex and are shown in the third row (second minimal TSTDP) of Table 2.1. The three curves showcase the sliding threshold feature of the BCM rule, that in TSTDP model can be controlled by changing $\lambda = \rho^p_{\text{post}}/\rho^p_0$ (see text for details).

Suppression Model of STDP

Prior to the experiments performed by Froemke and Dan in 2002, STDP experiments were mainly directed at varying the time intervals between a pair of pre-post or post-pre of spikes. In these conventional STDP experiments, synaptic weight modification was only based on the timing difference between pairs of spikes. However, based on a new model and experiments reported in Froemke and Dan (2002), synaptic plasticity not only depends on the timing differences between pre- and post-synaptic spikes, but also depends on the spiking pattern and the inter-spike intervals of each neuron.

In light of their experimental data from slices of cat visual cortex, Froemke and Dan, proposed a synaptic plasticity rule, the so-called *suppression model*. According to this

rule, spike pairs are not independent and there is a suppressive interaction among consecutive spikes, i.e. a pre-synaptic spike before a pre-synaptic spike, or a post-synaptic before a post-synaptic spike causes suppression of the synaptic efficacy of the second spike in interaction with its neighbour spikes from the other type. This suppression subsides in time and therefore, the longer time passed from arrival of a pre/post spike, the more synaptic efficacy the incoming pre/post spike will cause. The suppression model is described by the following equation:

$$\Delta w_{ij} = \epsilon_i^{\text{pre}} \epsilon_j^{\text{post}} F(\Delta t_{ij}), \tag{2.8}$$

where Δw_{ij} is the synaptic weight change due to the i^{th} pre-synaptic spike and the j^{th} post-synaptic spike, $\epsilon_i = 1 - e^{-(t_i - t_{(i-1)})/\tau_s}$ is the efficacy of i^{th} spike and τ_s is the suppression constant. In addition $F(\Delta t_{ij})$ is defined in a similar way as defined by Eq. 2.2.

Demonstrated results in Froemke and Dan (2002) show that the suppression model is able to account for the experiments using higher order spike trains such as triplet and quadruplet of spikes, and the LTP and LTD observed in the STDP learning window. However, this rule cannot account for frequency-dependent pairing experiments reported in Sjöström *et al.* (2001). In order to remove the deficiency of the suppression model in replicating this frequency-based experiment, Froemke *et al.* proposed a revised version of their suppression model in 2006 (Froemke *et al.* 2006). Using this new model, the frequency dependent experiments also can be generated. However, this feature makes the new suppression model to require more components, which in turn makes the new model to become more complex compared to the first proposed rule.

Spike Driven Synaptic Plasticity

In addition to the timing-based synaptic plasticity rules, mentioned above, there is another synaptic plasticity rule, which acts based on the recent activities and the current state of the post-synaptic neuron, and not only according to its spike timing. This rule has been developed to resolve a shortcoming of PSTDP rule, in being unable to learn patterns of mean firing rates (Abbott and Nelson 2000). This weakness is due to the high sensitivity of the PSTDP rule to the spike timings and temporal patterns (Mitra *et al.* 2009).

In the spike-driven model of synaptic plasticity, changes in synapse efficacy occur whenever a pre-synaptic spike arrives (Brader *et al.* 2007). At this time if the post-synaptic membrane potential, V_{mem}, is higher than a threshold voltage, V_{mth}, potentiation, or if it is lower than the threshold, depression will occur requiring the fact that the amount of calcium concentration in the post-synaptic site, $C(t)$, is within a predefined boundary at the arrival of the pre-synaptic spike. In short, the synaptic efficacy, W, will change according to the following equation

$$\begin{aligned} W &= W + a; \text{ if } V_{\text{mem}}(t) > V_{\text{mth}} \text{ and } \theta^l_{\text{up}} < C(t) < \theta^h_{\text{up}} \\ W &= W - b; \text{ if } V_{\text{mem}}(t) \leq V_{\text{mth}} \text{ and } \theta^l_{\text{dn}} < C(t) < \theta^h_{\text{dn}} \end{aligned} \tag{2.9}$$

where a and b are the amount of potentiation and depression respectively. In addition, $[\theta^l_{\text{up}}, \theta^h_{\text{up}}]$ and $[\theta^l_{\text{dn}}, \theta^h_{\text{dn}}]$ are the boundaries for the calcium concentration, $C(t)$, for potentiation and depression states, respectively.

If the required conditions are not satisfied, there will be no potentiation or depression. When there is no spike coming and therefore there is no synaptic weight change, the synaptic weight, W, will drift toward either high or low synaptic weight asymptotes. The direction of the drift will depend on the values of the weights at that specific time, which can be above/below a certain threshold, θ_W (Brader *et al.* 2007, Mitra *et al.* 2009)

$$\begin{aligned} \tfrac{dW(t)}{dt} &= \alpha; \text{ if } W(t) > \theta_W \\ \tfrac{dW(t)}{dt} &= -\beta; \text{ if } W(t) \leq \theta_W. \end{aligned} \tag{2.10}$$

The internal state, $C(t)$, which represents the calcium concentration, depends on the neuron's spiking activity and changes by the following equation

$$\frac{dC(t)}{dt} = -\frac{C(t)}{\tau_C} + J_C \sum_i \delta(t - t_i), \tag{2.11}$$

where J_C determines the amount of calcium contributed by a single spike (Brader *et al.* 2007, Fusi *et al.* 2000, Sheik *et al.* 2012b).

The main difference between the SDSP rule and the STDP-type rules is that, in the SDSP rule, the timing of the spikes is replaced by the membrane potential of the post-synaptic neuron, which has a close relation with the timing of the post-synaptic neuron. One could simply assume that a neuron with the membrane potential above a certain level, a threshold, is most likely to fire a spike, and therefore a post-synaptic spike will be fired at that time. Therefore, there is a close analogy between SDSP and STDP

rules. In addition to SDSP, some other rules have also utilised this specific feature of the membrane potential available in the post-synaptic neuron site, and combined it with either pre-synaptic spike timings or with its rate, and proposed new and more powerful rules, compared to the conventional timing-based-only rule, i.e. STDP. A brief review of two membrane potential (voltage)-based rules is presented in the following sections.

Voltage-based STDP

The voltage-based STDP model proposed in Clopath *et al.* (2010) and Clopath and Gerstner (2010), is a modification of TSTDP rule proposed by Pfister and Gerstner (2006). In this rule, the combination of the post-synaptic membrane potential and the pre-synaptic spike arrival time, govern the plasticity mechanism. At the arrival of a pre-synaptic spike, the synaptic weight will be depressed, if the post-synaptic neuron has been depolarised for some time, since it shows that a post-synaptic spike happened recently. The post-synaptic spike history window, depends on the time constant, by which the post-synaptic membrane is filtered. The following equation shows the membrane potential, u, which is low-pass-filtered with the time constant τ_-,

$$\tau_- \frac{d}{dt}\overline{u}_-(t) = -\overline{u}_-(t) + u(t). \tag{2.12}$$

Depression will occur when a pre-synaptic spike arrives, if at that moment, the low-pass filtered membrane potential, $\overline{u}_-(t)$, is above a certain threshold, θ_-. Therefore, depression can be mathematically modelled as

$$\frac{d}{dt}w^- = -A_{\text{LTD}}X(t)[\overline{u}_-(t) - \theta_-]_+ \text{ if } w > 0, \tag{2.13}$$

where w shows the synaptic weight, A_{LTD} is depression amplitude, and $X(t) = \sum_i \delta(t - t_i)$, with t_i as the spike times, represents a pre-synaptic spike train. In Eq. 2.13, $[x]_+$ is equal to x, if x is positive and is 0 otherwise.

In addition, potentiation occurs if three required conditions are satisfied simultaneously. These conditions are: (i) The momentary post-synaptic voltage, u, is above a certain threshold, θ_+. (ii) The low-pass-filtered version of u with its respective time constant, τ_+, is above θ_-. (iii) A pre-synaptic spike has occurred recently and left a trace, $\overline{x}$. Therefore, depression can be mathematically modelled as

$$\frac{d}{dt}w^+ = A_{\text{LTP}}\overline{x}(t)[u(t) - \theta_+]_+[\overline{u}_+(t) - \theta_-]_+ \text{ if } w < w_{\max}, \tag{2.14}$$

where $\overline{x}$ is a low-pass-filter of the pre-synaptic spike train with time constant τ_x, and can be modelled as

$$\tau_x \frac{d}{dt}\overline{x}(t) = -\overline{x}(t) + X(t). \tag{2.15}$$

According to the mentioned equations for depression (Eq. 2.13) and potentiation (Eq. 2.14), the overall synaptic weight change can be calculated by

$$\frac{d}{dt}w = -A_{\text{LTD}}X(t)[\overline{u}_-(t) - \theta_-]_+ + A_{\text{LTP}}\overline{x}(t)[u(t) - \theta_+]_+[\overline{u}_+(t) - \theta_-]_+. \tag{2.16}$$

Clopath *et al.* have shown that their voltage-based STDP rule is capable of reproducing the outcomes of a variety of STDP experiments (Clopath and Gerstner 2010, Clopath *et al.* 2010). Their rule that is a modification of the TSTDP rule, possesses higher synaptic modification capabilities compared to the TSTDP rule. However, one should keep in mind that the complexity of this rule is higher than the TSTDP rule as well. This higher complexity along with the dependence of the rule on the membrane potential of the post-synaptic neuron, increases the complexity of VLSI implementation of this rule. This is the case for other rules with higher capabilities, and therefore higher complexity.

Voltage-based BCM

Another rule that has been recently implemented in VLSI and has a phenomenological and computational background is the Local Correlation Plasticity (LCP) rule reported in Mayr and Partzsch (2010). This rule modifies the synaptic plasticity in a BCM-like fashion. In this rule, the weight changes take place, in a relation with the current state of the post-synaptic membrane voltage, and the recent dynamic of the pre-synaptic spikes. The synaptic alteration rule that has been implemented in VLSI for this BCM-based learning circuit employs the following rule to modify the synaptic weight according to Eq. 2.17

$$dw(t)/dt = B \cdot (u(t) - \phi_u) \cdot g(t), \tag{2.17}$$

where $w(t)$ is the synaptic weight, $u(t)$ is the neuron's membrane potential, ϕ_u is a threshold between potentiation and depression, $g(t)$ is a conductance variable that

represents the post-synaptic current, I_{psc}, and therefore has its maximum value at the time of a pre-synaptic arrival and decays afterwards. The main difference between this rule and the BCM learning model, is replacing the non-linear function, ϕ in BCM model (Eq. 2.6), with a constant multiplier, B, as shown in Eq. 2.17. This results in a linear dependence of the plasticity to the membrane potential, however as shown in Mayr and Partzsch (2010), in this model, this linear dependence is translated to a non-linear dependence between plasticity and the post-synaptic rate, identical to BCM.

This rule has been shown to reproduce the outcomes of many experimental protocols including triplet, quadruplet (Wang *et al.* 2005), and pairing frequency experiments performed in the visual cortex (Sjöström *et al.* 2001). Although this rule is able to replicate many plasticity outcomes, it is prone to large errors when parameters are fitted to closely replicate experimental results. These errors are rather high compared to the TSTDP rule which is simpler and possesses fewer state variables (Mayr and Partzsch 2010, Pfister and Gerstner 2006). In addition, as already mentioned, the complexity of a synaptic plasticity rule is an essential issue when designing a rule in VLSI, and the LCP rule has higher complexity than the TSTDP rule. Besides, the LCP rule's dynamic is dependent to the neuron model, as well as synaptic conductance, comparing to the TSTDP rule that only depends on the timing of the spikes, that are easily available at the synapse pre-synaptic and post-synaptic sites. Generally, when designing a synaptic plasticity rule, there is a need to consider the complexity versus strength of the rule and set a trade-off between them. This issue will be discussed later in this thesis.

Other Phenomenological Synaptic Plasticity Rules

In addition to the aforementioned abstract synaptic plasticity models, there are several other rules, using which a variety of the biological experiments performed in the neocortex can be regenerated. These rules range from simple timing-based models to other calcium-based models. The dynamical two-component long-term synaptic plasticity rule that was proposed in Abarbanel *et al.* (2002), has the ability to produce some LTP and LTD experiments, and also can be mapped to the BCM rule. In addition, the STDP rule that has been combined with the BCM sliding threshold feature is another synaptic plasticity mechanism that is able to account for a range of experiments in the hippocampal (Benuskova and Abraham 2007).

For a review of phenomenological synaptic plasticity models, see Morrison *et al.* (2008). Further, for a review and discussion of biophysical and phenomenological rules refer to Mayr and Partzsch (2010).

2.6.2 Biophysical Rules

None of the rules mentioned so far truly map to the biophysics of the synapse and biochemical reactions that take place in the synapse to induce synaptic weight changes. Instead, they are all models that curve fit the outcomes of as many biological experiments as possible under a unified mathematical expression. Advances in experimental techniques, including optogenetic and molecular methods, will permit researchers to investigate intricate aspects of the biochemical network, including protein-protein interactions, which result in plastic changes at the level of synapses. This now permits the development of complicated biophysical models that take into account the observed molecular processes underlying changes in synaptic strength. Such models are expected to naturally reproduce the correct synaptic alteration for all experimental protocols. Due to the close analogy of these models with the dynamics of the synapse, these rules are usually called biophysical rules. In the following, we describe a few of these rules, in particular those, for which a VLSI implementation is also available.

Modified Ion Channel-based Plasticity

This rule not only considers calcium and its level for inducing synaptic weight changes, but also introduces the effect of other ion channels and receptors as the pathways for calcium to change in the post-synaptic neuron and therefore causes either potentiation or depression. The synaptic weight change mechanism is as follows: pre-synaptic action potentials release glutamate neurotransmitters that binds to N-methyl-D-aspartate (NMDA) receptors, and when post-synaptic activities that provide large membrane depolarisations are simultaneously present, it leads to an increase in the level of calcium (Meng *et al.* 2011). This rule is capable of reproducing both BCM (rate-based) and PSTDP (timing-based) mechanisms using a unified model. However, this model is complex and requires a large number of state variables (Meng *et al.* 2011).

Iono-Neuromorphic Intracellular Calcium-Mediated Plasticity Model

This is a synaptic plasticity rule that is focused on intracellular calcium dynamics of the synapse. It is another biophysically inspired plasticity rule that acts entirely based

on the dynamics of the ions and channels within the synapse. The rule was originally proposed by Shouval *et al.* (2002) and Shouval *et al.* (2010) and modified to be implemented in VLSI. The weight changes for the VLSI circuit are given by

$$dw = \eta([\mathrm{Ca}])(\Omega([\mathrm{Ca}]) - \lambda w), \tag{2.18}$$

where w is the current synaptic weight, $\eta([\mathrm{Ca}])$ is a superlinear function of $[\mathrm{Ca}^{2+}]_i$, $\Omega([\mathrm{Ca}])$ is the calcium dependent update rule, and λ plays the role of a learning rate (Rachmuth *et al.* 2011, Shouval *et al.* 2010).

Similar to the previous ion channel-based plasticity, this rule is shown to be capable of reproducing BCM and spike pairing synaptic plasticity experiments. However, the model is very complicated and needs several state variables to induce synaptic weight changes in a biophysical form. Another limitation is that its ability to reproduce the behaviour observed in triplet, quadruplet, and frequency-dependent pairing experiments, has not been reported.

Other Biophysical Synaptic Plasticity Rules

Since the calcium ion and its dynamics in the synaptic cleft appear to play an essential role in the synaptic plasticity, in addition to the aforementioned calcium-mediated plasticity rules, several other rules, which build upon the calcium dynamics, have been also proposed and investigated. In 2006, Shah *et al.* (2006) proposed a modified version of the Shouval's calcium-medicated rule (Shouval *et al.* 2002). This rule has been modified in order to account for the non-linear contributions of spike pairs to the synaptic plasticity when considering a natural spike trains as reported in Froemke and Dan (2002).

Besides these calcium-based models, there are some other new rules, which also utilise calcium dynamics to induce synaptic weight changes. One of these rules, that can be also counted as a phenomenological rule, is a simplified version of the Shouval's calcium-based rule. This rule utilised calcium dynamics to account for various biological experiments, and try to investigate the effects of a biophysical parameter, i.e. calcium, on the synaptic plasticity (Graupner and Brunel 2012). The dynamic of the synaptic plasticity rule is as follows:

$$\tau \frac{d\rho}{dt} = -\rho(1-\rho)(\rho_* - \rho) + \gamma_p(1-\rho)\Theta[c(t) - \theta_p] - \gamma_d \rho \Theta[c(t) - \theta_d] + \mathrm{Noise}(t), \tag{2.19}$$

where τ is the time constant of synaptic efficacy changes happening on the order of seconds to minutes. Synaptic efficacy is shown with ρ and ρ_*, which are the boundary of the basins of attraction of two stable states, one at $\rho = 0$, a DOWN state corresponding to low efficacy, and one at $\rho = 1$, an UP state corresponding to high efficacy. Here, $c(t)$ determines the momentous calcium concentration. There will be a potentiation if $c(t)$ is above a potentiation threshold, θ_d. Similarly, there will be a depression if $c(t)$ is above a depression threshold, θ_d. γ_p and γ_d are potentiation and depression rates, respectively, that will be in effect when the potentiation or depression threshold are exceeded. In addition, Θ is a Heaviside function, in which $\Theta[c - \theta] = 0$ for $c < \theta$, otherwise $\Theta[c - \theta] = 1$ and Noise(t) is an activity-dependent noise term. For further details refer to Graupner and Brunel (2012).

In addition, Badoual *et al.* (2006) proposed two synaptic plasticity mechanisms, one with a biophysical background, and the other as a phenomenological model. Both rules are based on multiple spike interactions and utilise the feature of conventional STDP rule to reproduce the outcomes of some biological experiments including the triplet experiments using the suppression STDP model presented in Froemke and Dan (2002) and the STDP learning window (Song *et al.* 2000). Although these rules are able to account for these experiments, reproducing other important experiments including the pairing frequency experiments, quadruplet and BCM, are not reported using these models.

Furthermore, another calcium-based STDP rule was recently proposed (Uramoto and Torikai 2013), which utilises three state variables in an Ordinary Differential Equation (ODE) form. This model is simpler than previous calcium-based and voltage-based STDP models as it uses lower number of state variables and does not use extra parameters such as delay in the model as used by Graupner and Brunel (2012). Although this model is simpler than some voltage- and calcium-based models and also has higher synaptic plasticity abilities than the TSTDP rule, it still is more complicated than TSTDP rule, which is a merely timing-based rule.

All the phenomenological rules mentioned above have stronger links to the biophysics of synapses than the simpler phenomenological ones. However, these rules are also more complex in structure, and therefore need more resources to be implemented in VLSI. In fact, there exist only a few implementations of biophysical rules in the literature and those implementations are against the main needs of neuromorphic engineers, i.e. low power consumption and compactness (Rachmuth *et al.* 2011, Meng *et al.*

2011). Therefore, implementing a simpler, low power, and smaller circuit with adequate synaptic capabilities, which accounts for a number of essential synaptic experiments is absolutely promising for the realisation of a large-scale biophysically plausible neuromorphic system.

2.7 Chapter Summary

In this chapter, a number of synaptic plasticity rules that are of interests to the neuromorphic community were discussed and reviewed. The rules were divided into two groups of phenomenological and biophysical plasticity rules. It was shown that these rules have different abilities and performance in reproducing the outcomes of a variety of synaptic plasticity biological experiments. As shown, some rules such as pair-based STDP are not able to account for a number of experiments, while some other rules such as TSTDP, LCP and voltage-based STDP can generate a variety of these experiments. In addition, biophysical rules were also discussed and their capabilities were reviewed. The investigations performed in this chapter provide us with a good understanding of the timing and rate-based rules and their structures. This knowledge, is useful for implementing these rules and mechanisms in VLSI and then utilising them for different applications.

In the next two chapters, three of the reviewed plasticity rules, i.e. PSTDP, TSTDP and BCM that are the rules of main interest in this thesis are implemented and tested in a programmable multi-neuron hardware neuromorphic system. Using this system the hardware implementation of these synaptic plasticity rules are verified. Then the implemented rules are utilised for classification of complex rate-based patterns.

Chapter 3

Programmable Neuromorphic Circuits for Spike-based Neural Dynamics

THIS chapter describes the architecture and structure of a programmable hybrid analog/digital neuromorphic circuit, called IFMEM, that can be used to build compact low-power neural processing systems. Here, first the architecture of the IFMEM neuromorphic system is described and then it is explained how this hybrid analog-digital CMOS circuit operates correctly over a wide range of input frequencies; a feature that is essential for many engineering applications. The chapter shows measurement results from available silicon neurons, and neuron-synapse combinations and demonstrates how specific neural behaviours can be generated by programming the chip and calibrating the silicon neurons and synapses parameters. The provided information in this chapter elucidates the presented results in the following chapter.

3.1 Introduction

As discussed in previous chapters, artificial spiking neural networks offer a promising paradigm for a new generation of brain-inspired computational models. A wide range of theoretical and computational models have already been proposed for both basic neuroscience research (Kempter *et al.* 1999, Gerstner and Kistler 2002) and practical applications (Belatreche *et al.* 2006, Rowcliffe and Feng 2008). Neuromorphic VLSI circuits represent a suitable technology for implementing these types of networks using hybrid analog/digital design techniques, and for building devices that have a very high potential in a wide range of applications such as pattern classification and recognition (Indiveri and Horiuchi 2011, Wang *et al.* 2014a, Kasabov 2014, Kasabov *et al.* 2014). In particular, the main advantage of implementing these spiking neural networks in neuromorphic VLSI technology is their compactness and low power consumption, which are critical features when implementing large scale neural architectures (Mead 1990, Poon and Zhou 2011). Note that, despite its advantages, the VLSI technology as we know it today and in the near future, cannot simply address the complexity boundary required for implementing a brain-scale neuromorphic system with billions of neurons and trillions of synapses. Therefore, there is a need to exploit new promising technologies such as memristors (Eshraghian *et al.* 2012), along with the VLSI technology to reach the required complexity boundaries (Azghadi *et al.* 2014b).

As already mentioned, the two main components when implementing a neural system are neurons and synapses. Synapses are essential components of spiking neural networks that represent the site of memory (as they store the network's synaptic weight values), and play a fundamental role in computation (as they implement crucial temporal and non-linear dynamics). In spiking neural networks, the synaptic weight is directly associated with the activity of pre-synaptic and post-synaptic neurons (Kempter *et al.* 1999). Different types of learning algorithms have been proposed, to update the synaptic weight as functions of both pre- and post-synaptic activity (Brader *et al.* 2007, Morrison *et al.* 2008). The different learning strategies have a profound effect on the post-synaptic neuron functionality and on the spiking neural network behaviour (Laughlin and Sejnowski 2003). Implementing such types of synapses and learning mechanisms in compact electronic systems is crucial, for developing efficient large-scale spiking neural networks, which learn, and for brain-inspired computing technologies that can adapt. However, as the implementation of the learning algorithm often depends on the specific application domain and on the nature of

the data to be processed, it is useful to develop compact electronic implementation of spiking neural networks, in which the weights can be adjusted by off-chip learning algorithms, e.g., implemented on a workstation, micro-controller, or Field Programmable Gate Array (FPGA).

In this chapter a programmable neuromorphic circuit that has been fabricated using standard CMOS VLSI process (Moradi and Indiveri 2011) and that can support any weight-update mechanism of interest and learning strategies, is presented. Specifically, a set of experimental results measured from the fabricated neuron and synapse circuits demonstrating how they can be calibrated to a specific targeted behaviour, is shown. Using this circuit in the following chapter, it is shown how STDP and TSTDP learning algorithms can be implemented and reproduce a number of synaptic plasticity experimental outcomes.

The results in this chapter have been mainly presented in *The 2013 New Circuits and Systems conference*, in Paris, France (Azghadi *et al.* 2013d), as well as in *The ACM Journal on Emerging Technologies in Computing Systems* (Azghadi *et al.* 2014c). Note that the IFMEM neuromorphic architecture is designed and fabricated in the Neuromorphic Cognitive Systems (NCS) group, Institute of Neuroinformatics (INI), University and ETH Zurich, Switzerland. The architecture and main characteristics of the IFMEM neuromorphic device are presented mainly in Moradi and Indiveri (2011) and Moradi and Indiveri (2014).

3.2 The IFMEM Chip

The multi-neuron chip used in this chapter is characterised by the fact that it comprises circuits that implement models of IF neurons, and a programmable memory for storing the synaptic weights. Therefore, it is referred to this device as the "IFMEM" (Integrate and Fire neurons with synaptic Memory) chip.

The photo micro-graph of the IFMEM chip comprising the programmable synapse circuits is shown in Fig. 3.1. This chip was fabricated using a standard 0.35 μm CMOS technology and is fully characterised in Moradi and Indiveri (2014). The micro-graph depicts various parts of the chip, including an on-chip 32-bit programmable bias generator (Delbrück *et al.* 2010), SRAM cells, an arbitration part, an asynchronous controller, and the "Neural Core".

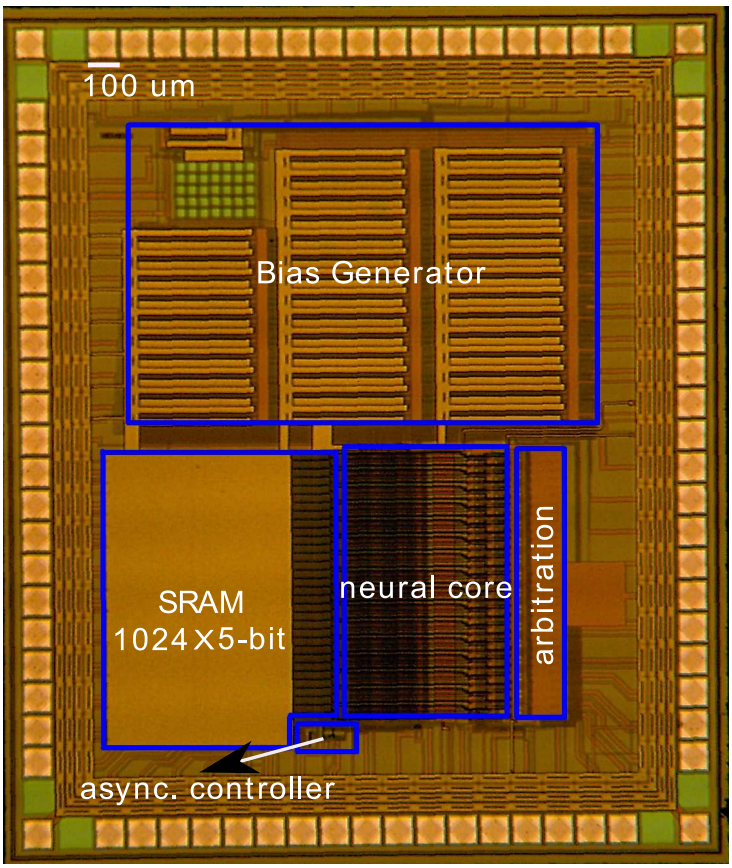


Figure 3.1. The IFMEM neuromorphic device chip micro-graph. The multi-neuron IFMEM chip was fabricated using a standard 0.35 μm CMOS technology and occupies an area of 2.1×2.5 mm^2. The programmable synapses are integrated inside the neural core block (Moradi and Indiveri 2014, Azghadi *et al.* 2013d).

All circuits on the chip that implement the neural and synapse dynamics are in the neural core block. The neuron circuits are implemented using an "adaptive exponential integrate and fire" model (Brette and Gerstner 2005, Indiveri *et al.* 2010), while the part of the synapse circuits responsible for integrating input spikes and producing temporal response properties that have biologically plausible time constants are implemented using a DPI circuit (Bartolozzi and Indiveri 2007).

The block diagram of the chip architecture is shown in Fig. 3.2(a), which demonstrates the working scheme of this programmable neural system. Depending on the input address-event, different types of synapse dynamics can be triggered: excitatory with slow time constants (e.g., to emulate NMDA-type synapses), excitatory synapses with faster time constants (e.g., to emulate AMPA-synapses), or inhibitory synapses (e.g., to emulate GABA-type synapses). Since the DPI can be used as a linear low-pass filter, it is possible to make use of a single integrator circuit for any of the synapse dynamics

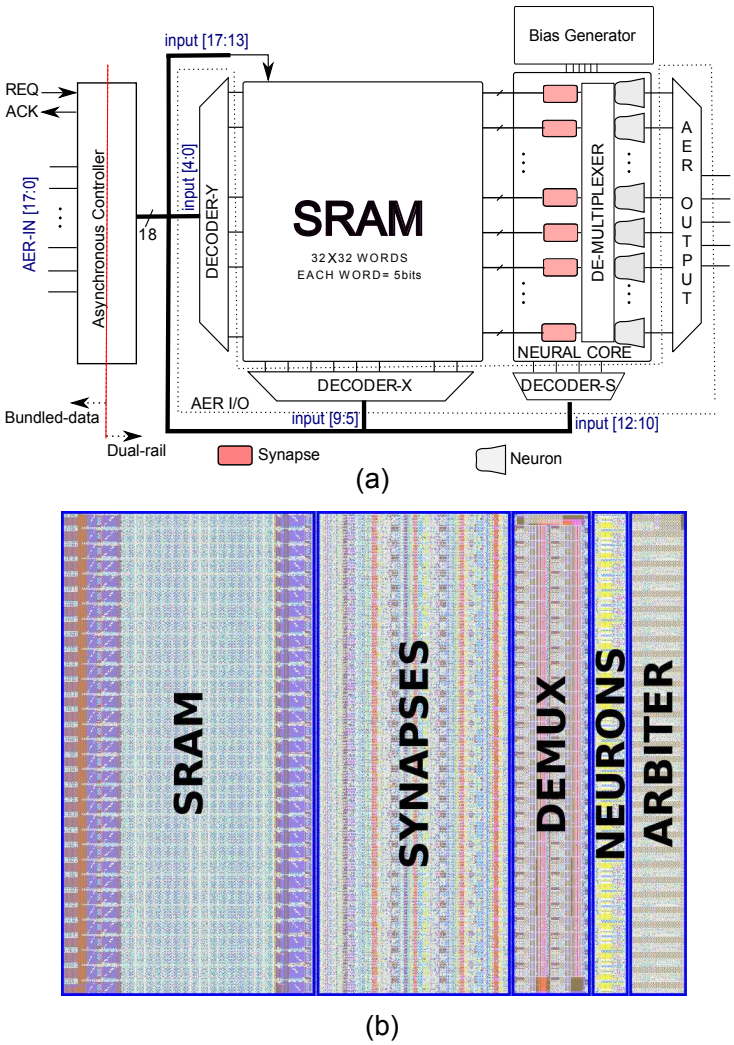


Figure 3.2. The IFMEM chip block diagram. (a) The device comprises a neural-core module with an array of synapses and integrate-and-fire neurons, an asynchronous SRAM module to store the synaptic weight values, a bias generator to set the parameters in the analog circuits, and asynchronous control and interfacing circuits to manage the AER communication. (b) Layout picture comprising the SRAM, neural core and AER output blocks. In particular, the layout of the SRAM block measures $524 \times 930\,\mu m^2$; the synapse array measures $309\,\mu$m in length, the synapse de-multiplexer measures $132\,\mu$m, the neuron array $60\,\mu$m, and the output AER arbiter $105\,\mu$m in length, while the width of all of them is equal to 930μm (Moradi and Indiveri 2014, Azghadi *et al.* 2014c).

considered (e.g., NMDA, AMPA, or GABA), and multiplex it in time to integrate the contributions from multiple spiking inputs (e.g., via multiple SRAM cells).

The analog neural components available on the chip have programmable bias parameters that can be set with an on-chip 32-bit temperature compensated programmable bias generator (Delbrück *et al.* 2010). The synaptic weights of the synapses are stored in

a 32×32 5-bit digital SRAM block, designed with asynchronous circuits for interfacing to the AER components. The digital weight values are converted into currents with an on-chip Digital to Analog Converter (DAC), so that the addressed synapse circuits produce EPSCs with amplitudes proportional to their weights.

The on-chip synapse circuits integrate incoming spikes and produce EPSCs with amplitudes proportional to their corresponding stored weights. The temporal response properties of the circuit exhibit dynamics that are biophysically realistic and have biologically plausible time constants (Bartolozzi and Indiveri 2007). The part of the synapse circuit that produces the slow temporal dynamics is the log-domain DPI filter (Bartolozzi *et al.* 2006, Mitra *et al.* 2010), shown in Fig. 3.3. By using the DPI in its linear regime, it is possible to time-multiplex the contributions from multiple spiking inputs (e.g., via multiple SRAM cells), thus requiring one single integrating element and saving precious silicon real-estate. This time multiplexing scheme, and circuits implemented on the chip of Fig. 3.1 have been fully characterised in Moradi and Indiveri (2011), while the description of the DPI synapse dynamics has been presented in Bartolozzi *et al.* (2006).

Using the synapse time-multiplexing scheme, the total number of synapses that a neuron sees is equivalent to the total number of SRAM cells present in each row. The SRAM cells can work in "feed-through" mode or in storage mode. In feed-through mode, input events contain both the address of the destination SRAM cell and the synaptic weight bits, and the synapses generate EPSCs on-line, as the data is received. In storage mode, the input events contain only the address of the destination SRAM cell, and the weight bits used by the synapses are the ones stored in the addressed SRAM cell (Moradi and Indiveri 2011). Therefore it is possible to interface the device to a workstation and use it in "feed-through" mode to train the spiking neural network on-line, with all of the hardware components in the loop, eventually storing the final synaptic weight matrix in the SRAM block at the end of the training phase. Once the training has completed, it is possible to use the device in stand-alone mode, without requiring a PC in the loop, and use the stored weights to carry out the learned task (Azghadi *et al.* 2013d, Azghadi *et al.* 2014c).

Figure 3.2(b) shows a section of the layout of the IFMEM chip comprising the main blocks mentioned above. As shown, each block is extremely compact, so it is possible in principle to scale up the network to very large sizes (e.g., a chip fabricated using an

inexpensive 0.35 μm technology, using an area of 55 mm^2 would implement a network of 512 neurons and 256k synapses each having 5 bits of precision).

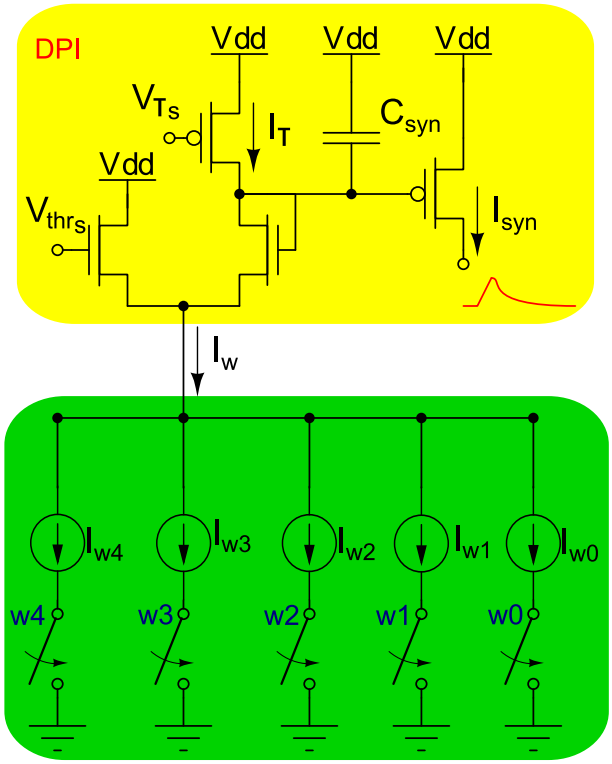


Figure 3.3. Schematic diagram of the programmable synapse circuit. The top part of the diagram represents a DPI circuit that implements the temporal dynamics. The bottom part of the diagram represents the DAC that converts the SRAM 5-bit weight into a corresponding synaptic current (Moradi and Indiveri 2014, Azghadi *et al.* 2013d).

3.3 Experimental Setup

The experimental setup, shown in Figures 3.4, consists of the three main components: a Linux PC, a generic AER interface board and, directly attached to it, a daughter-board containing the IFMEM chip. The PC is used to control and interact with the neuromorphic system. It generates spike trains that are sent to the IFMEM chip via an AER interface. The PC also monitors, records and analyses the AER output of the chip. Via a separate channel, the PC also sends bias values to IFMEM chip, which control its various circuit parameters (Azghadi *et al.* 2014c).

The *AEX* board, shown in Fig. 3.5, is a generic AER communication platform derived from the board first presented in Fasnacht *et al.* (2008). It consists of a high-speed

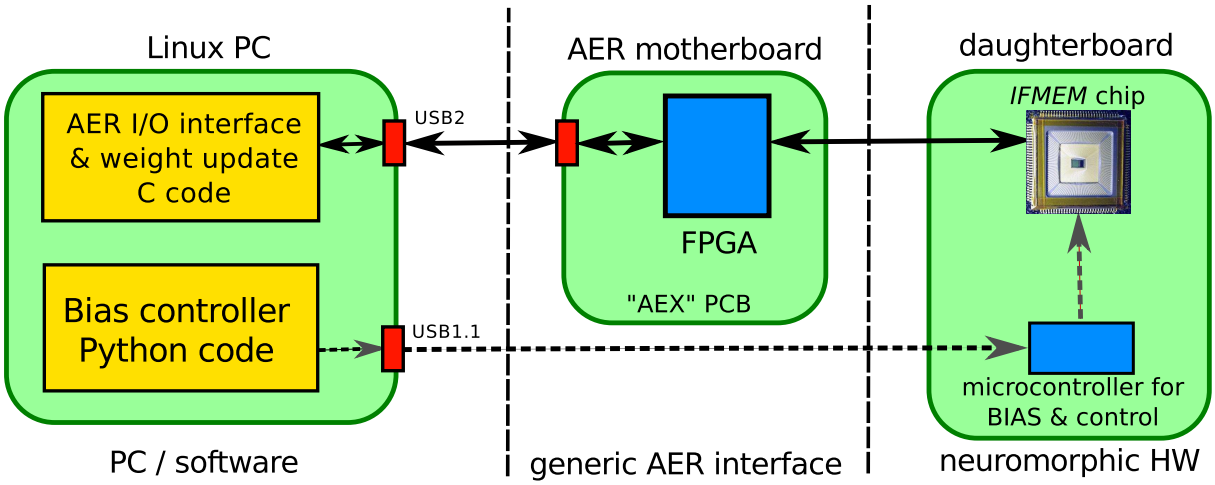


Figure 3.4. Experimental setup of the hardware-software neuromorphic system. Dashed lines represent the control path for setting analog parameters and configuring the IFMEM chip, solid lines represent the path for the address-events data flow; from and to the IFMEM chip (Azghadi *et al.* 2014c).

(480 MHz) USB2.0 interface and an FPGA device. The USB interface enables the FPGA to communicate bi-directionally with the PC attached. The FPGA receives spike trains from the PC via USB and then generates them accordingly on its Parallel AER output interface to stimulate the IFMEM chip. Vice versa, the FPGA monitors the AER output of the IFMEM chip: each address-event received by the FPGA is sent to the PC, together with a 128 ns resolution time stamp of when exactly the spike was received at the Parallel AER input of the FPGA. The AEX board also contains a high-speed Serial AER interface to communicate with other AEX boards. Since only one such board is required in the single-chip experimental setup described, the Serial AER interface was not used (Azghadi *et al.* 2014c).

Directly attached to the AEX communication board is a daughter-board. Figure 3.5 shows the two boards together. The daughterboard contains both the IFMEM chip and the circuitry needed to support the chip, such as voltage regulators and connectors to measure analog output voltages generated by the chip. It also contains a simple microcontroller that includes a full-speed (12 MHz) USB interface. Via this second USB interface the PC sends the bias values to the microcontroller. The microcontroller then programs the on-chip bias generator circuits to set the circuit bias voltages to the values specified by the user.

The IFMEM system utilises the AER protocol to transfer the required events to the chip and at the same time, record the spikes being generated by the neurons on the chip.

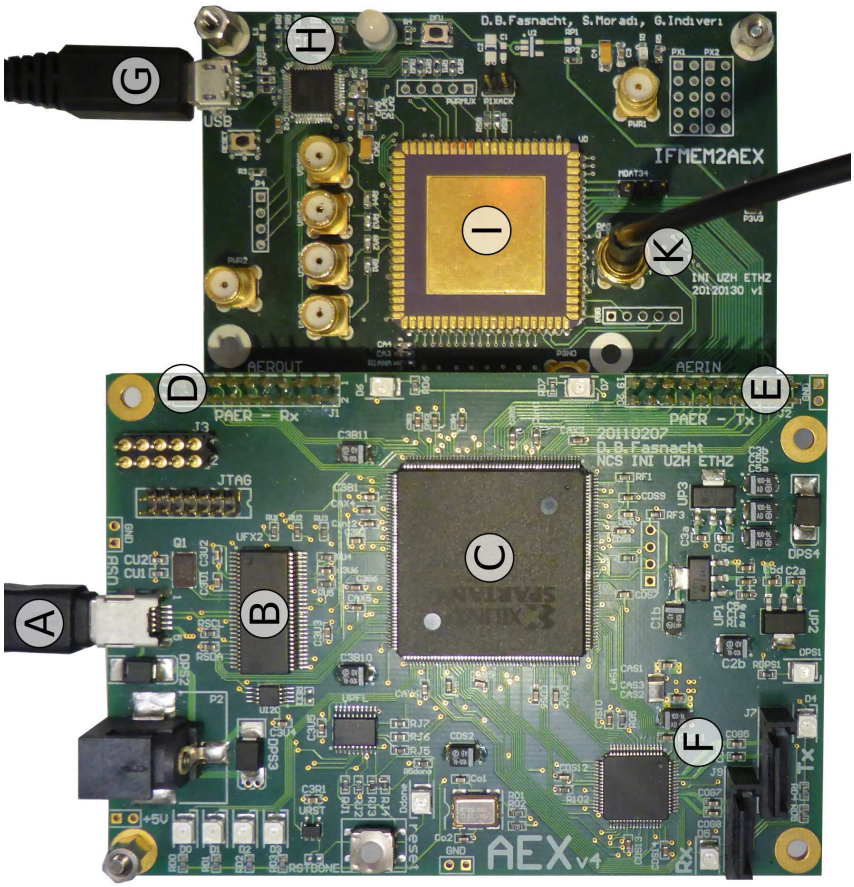


Figure 3.5. Printed circuit board of the experimental setup of the system. The AEX printed circuit board with the attached daughterboard carrying the IFMEM chip (Azghadi *et al.* 2014c):

A: High-speed USB interface for AER communication,

B: USB interface chip,

C: FPGA for AER monitoring and sequencing,

D: Parallel AER interface chip to FPGA,

E: Parallel AER interface FPGA to chip,

F: Serial AER section (unused),

G: Full-speed USB interface for IFMEM bias control,

H: Microcontroller for bias control,

I: The IFMEM chip,

K: An analog voltage output connection.

Each pre-synaptic AER address contains four slots of information including 18 bits as shown in Fig. 3.2(a). These bits describe different specifications including:

- the address of the post-synaptic neuron (5 bits),
- the address of the SRAM block containing the required synaptic weight (5 bits),
- the type (either inhibitory or excitatory) and the address of the desired synapse (3 bits), and
- the desired digital value for the synaptic weight that will be written to the addressed SRAM block (5 bits).

The weight across these virtual synapses can be modified using AER protocol and according to various learning and synaptic plasticity rules including timing-based ones e.g. STDP and triplet STDP (Azghadi *et al.* 2012a), as well as rate-based rules e.g. Bienenstock-Cooper-Munro (BCM) rule (Azghadi *et al.* 2012b). Implementing these rules will be discussed in the following chapter. Prior to implementing any synaptic plasticity algorithm on the IFMEM neuromorphic system, the response properties of the neuron and synapse circuits should be characterised.

In the following Section, first the silicon neuron response properties are shown, and then the response properties of the synapse circuits as a function of input spike frequency and of programmable weight values are characterised to evaluate their linear characteristics and dynamic range properties.

3.4 Silicon Neuron and Programmable Synapse Response Properties

The IF silicon neuron response properties can be controlled by various neural parameters that are available as bias voltages in the silicon neurons implemented on the chip. The response properties include various features of the neuron such as its refractory period, and its adaptation frequency. All of these properties can be controlled by bias voltages available for these specific purposes. As an example, Fig. 3.6 shows a spiking behaviour observed from a silicon neuron, by calibrating its bias parameters. In order to produce this behaviour, constant current was injected into the neuron circuit and the silicon neuron parameters were adjusted to obtain this biophysically realistic response properties, with biologically realistic time constants, in the order of tens of ms, as shown in the figure.

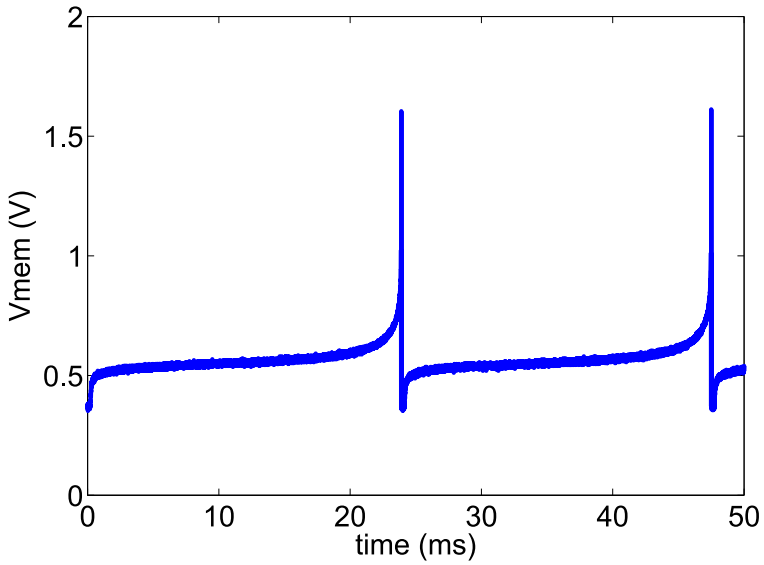


Figure 3.6. Silicon neuron response properties. Silicon neuron membrane potential (V_{mem}) in response to constant current injection is shown. Note that silicon neuorn biases are set specifically to show this biologically plausible regular spiking neural behaviour with biologically realistic time constants (Azghadi *et al.* 2014c).

Apart from the spiking behaviour of the neuron, setting a meaningful relationship between the input current injected to the neuron and its mean spiking frequency is another important characteristic of the neuron that was tested. Figure 3.7 shows an almost linear relationship between the amount of the injected current to the neuron and its output mean firing rate. Note that the spiking dynamic of the neuron in this case, that is shown in the inset of Fig. 3.7, is different from that shown in Fig. 3.6 (Azghadi *et al.* 2013d, Azghadi *et al.* 2014c).

In addition to the characteristics of the neurons, the properties of the available synapses on the chip should also be correctly characterised. Therefore, another set of measurements are carried out to calibrate the response properties of the combined synapse-neuron. For this experiment, one post-synaptic neuron and one single input synapse are employed. The response properties of the combined synapse-neuron circuit is characterised by sending input spikes to the synapse, and measuring output spikes from the neuron.

Note that the synapse circuit integrates input spikes to produce an output current (the synapse EPSC) that has a mean steady-state amplitude, which depends on both the input spike train frequency and its synaptic weight. The synapse, and if required the neuron parameters can be calibrated to achieve the desired spiking activity in relation

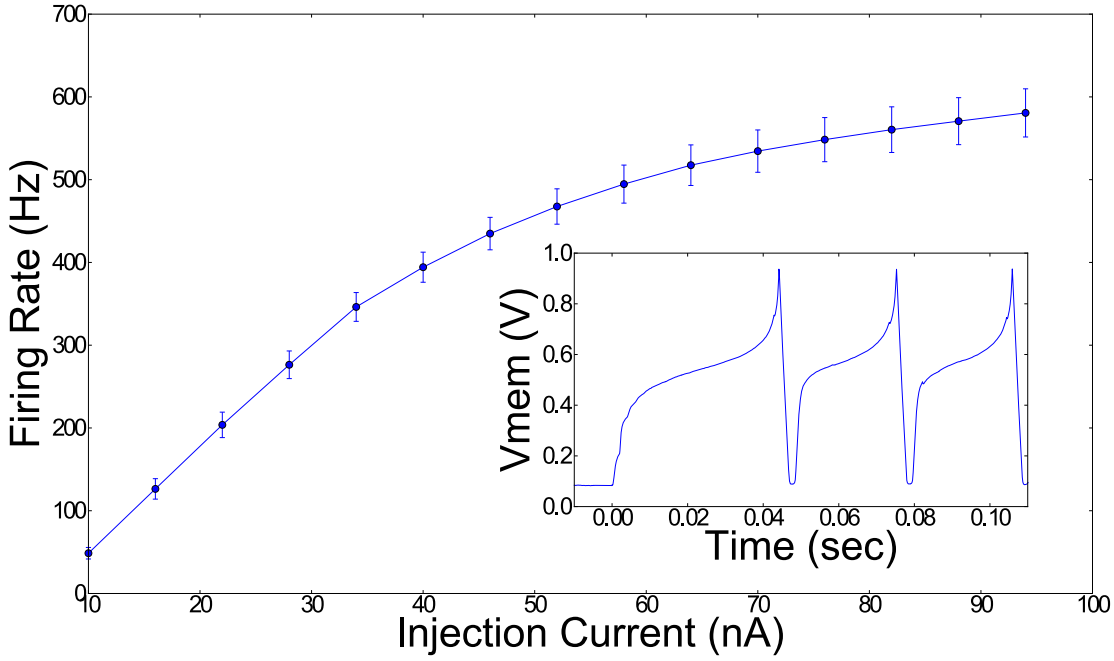


Figure 3.7. Input current versus frequency charachtristics of silicon neurons. Mean firing rates of all neurons on the chip, as a function of input current. The figure inset shows the membrane potential of a single neuron (Azghadi *et al.* 2013d).

to the synaptic weight and the input spike frequency. This spiking activity depends directly on the application, for which the Spiking Neural Network is going to be used. In order to optimise the use of the neuron and synapse circuits in various applications like computation, brain machine interface, pattern recognition etc., it is necessary to tune their analog bias parameters to specific values to result in the required features and expected behaviour of the neuromorphic chip for those applications (Azghadi *et al.* 2013a, Azghadi *et al.* 2014c).

Figure 3.8 shows how it is possible to optimise the circuit biases for a specific range of low pre-synaptic frequencies, so that the combined synapse-neuron circuits respond almost linearly to their afferent synaptic inputs, for all possible weight values that can be programmed. Under this condition, the neuron is able to show identical (gain $= 1$) or linearly higher (gain > 1) or lower (gain < 1) post-synaptic output frequency, compared to afferent pre-synaptic input frequency. Here, gain is defined as the fraction of post-synaptic to pre-synaptic firing rates in the highest synaptic weight setting ($w[4...0] = (11111)_2 = 31$). When there are synaptic inputs with various firing rates,

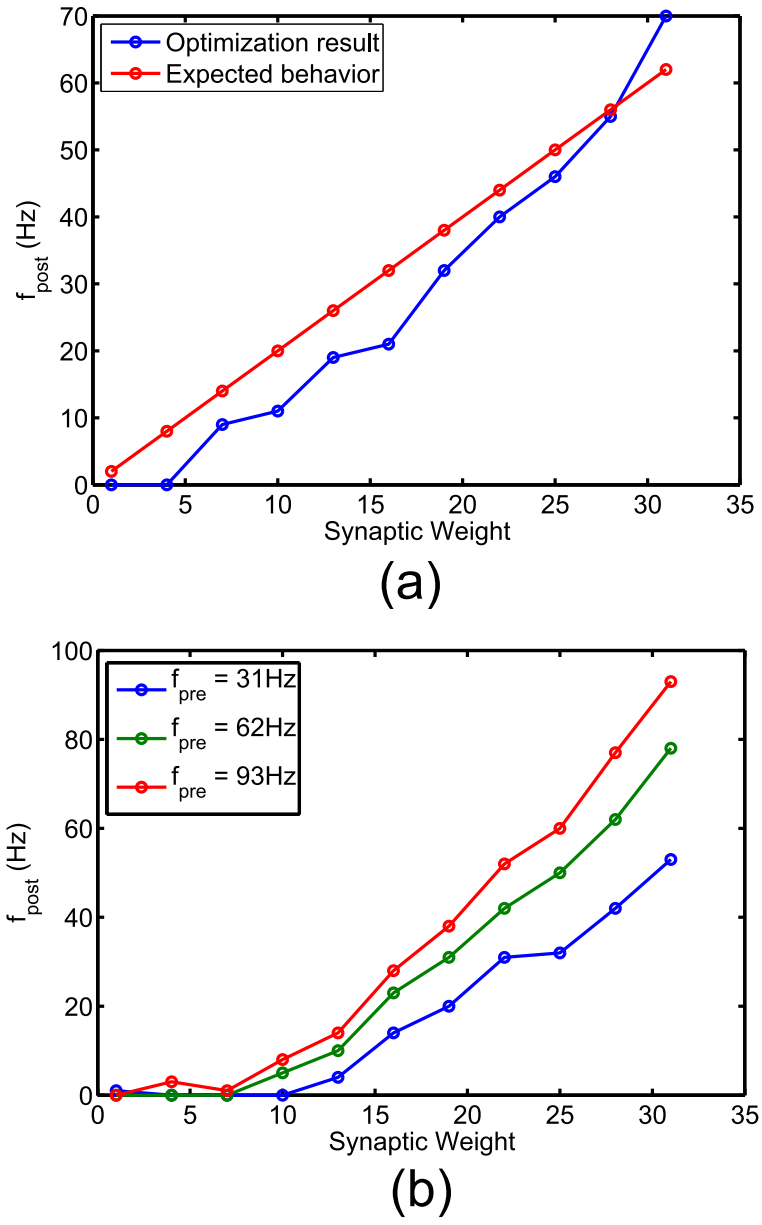


Figure 3.8. Synapse-neuron output response properties for low input frequencies. (a) The synapse-neuron bias parameters are optimised only for 31 Hz. (b) Regular spike trains with pre-synaptic input frequencies of 31, 62, and 93 Hz are applied to the synapse-neuron circuit, for different synaptic weight settings. The synapse-neuron biases are optimised to have an almost linear relationship for these three different input frequencies (Azghadi *et al.* 2013d).

the neuron and synapses should be tuned to act linearly for the whole possible input firing range of frequencies. In the calibration of the bias values in our circuit, the main sets of parameters that were tuned are those related to the five synaptic currents $I_{w0} - I_{w4}$ depicted in Fig. 3.3. Those parameters were tuned in a way to establish an

almost linear relationship between the 32-state digital synaptic weight values and the neuron post-synaptic firing rates.

Note that, figure 3.8 demonstrates the neuron-synapse response properties using a set of bias parameters optimised, in both neuron and synapse, for biologically plausible firing rates. Figure 3.8(a) shows the neuron-synapse response properties that are optimised only for 31 Hz. However, Fig. 3.8(b) demonstrates the response properties when the biases are optimised for three various pre-synaptic input frequencies, simultaneously. While Fig. 3.8(a) demonstrates a close match to the expected linear behaviour, the second figure loses some linearity.

Figure 3.9 shows similar results to Fig. 3.8, but with parameters optimised for high firing rates (e.g., for applications that need to process incoming data quickly, and for neuromorphic systems that do not need to interact with the environment). This figure shows that the biases can be optimised to achieve a very good linear relationship between synaptic weight and the neuron output mean firing rate. Fig. 3.9(a) shows the results when neuron and synapse biases are optimised for only 1 kHz input spike frequency. While this figure shows a very close match between the neuron output frequency, and an expected linear behaviour, Fig. 3.9(b) shows the results when the biases are optimised for three different input frequencies.

In addition to these response properties of the synapse-neuron combination, sometime a specific relationship between the mean firing rate of the input spike train to the synapse and that of the output of the neuron is required, with a determined synaptic weight. Therefore, the synapse should be calibrated in order to reach this behaviour. To calibrate the synapse parameters for this purpose, first the parameters of the on-chip DACs that convert the SRAM bits into programmed analog currents are adjusted. Then the synapse is stimulated with regular spike trains at different frequencies, and the neuron response is measured. Since the synapse is configured to behave as a linear filter, only a single synapse with input frequencies as high as 2 kHz is considered, in order to represent inputs from many neurons at lower frequencies—by means of the superposition principle. Figure 3.10 shows the response of a silicon neuron to these input spike trains for different synaptic weight values. As shown, we calibrated the on-chip DACs to set synaptic weights that have a low gain, even for the highest weight value ($w[4...0] = (11111)_2 = 31$).

All presented measurements in this chapter show the high degree of programmability the IFMEM chip possesses. It was shown that both neuron and synapse properties

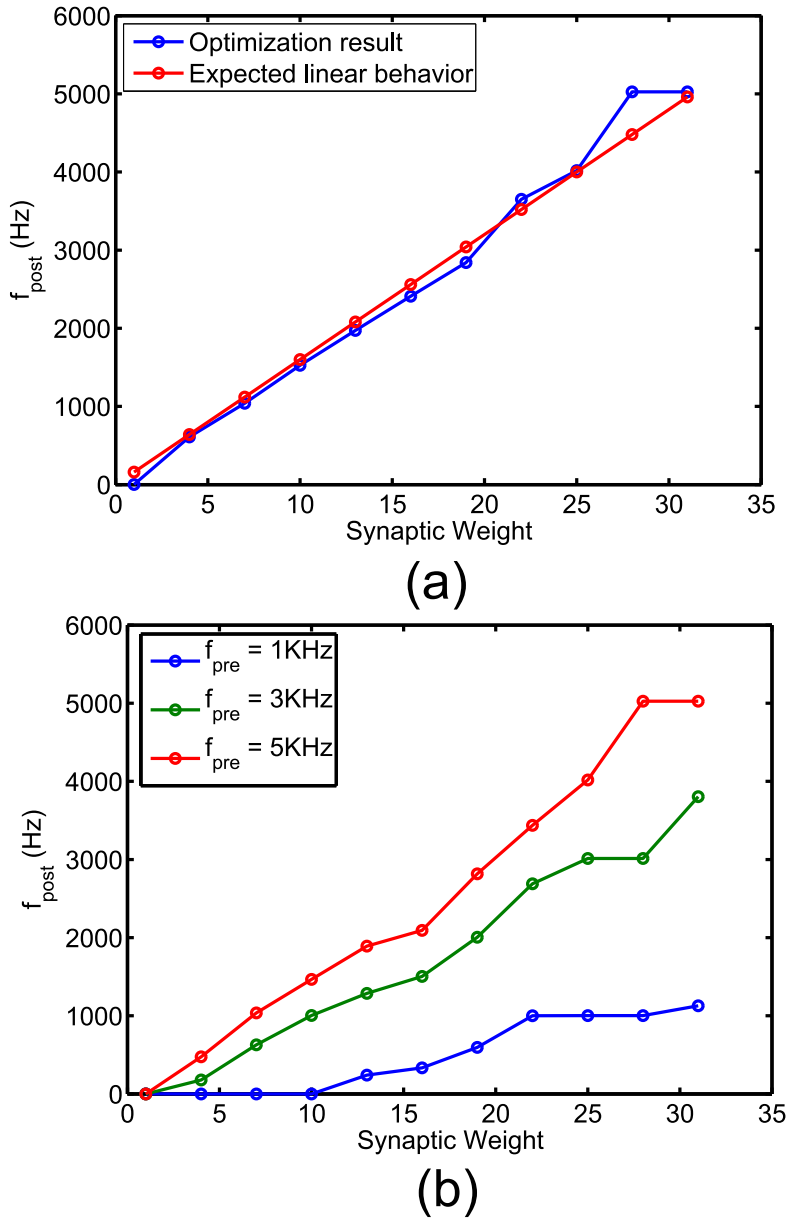


Figure 3.9. Synapse-neuron output response properties for high input frequencies. (a) The synapse-neuron bias parameters are optimised only for 1 kHz. (b) Regular spike trains with pre-synaptic input frequencies of 1, 3, and 5 kHz are applied to the synapse-neuron circuit, for different synaptic weight settings. The synapse-neuron biases are optimised to have an almost linear relationship for these three different input frequencies (Azghadi *et al.* 2013d).

can be controlled, by optimising and calibrating their bias parameters, to achieve any required behaviour from the neuron. Note that for optimising various biases that control different behaviours shown above and to reach any specific properties, the on-chip

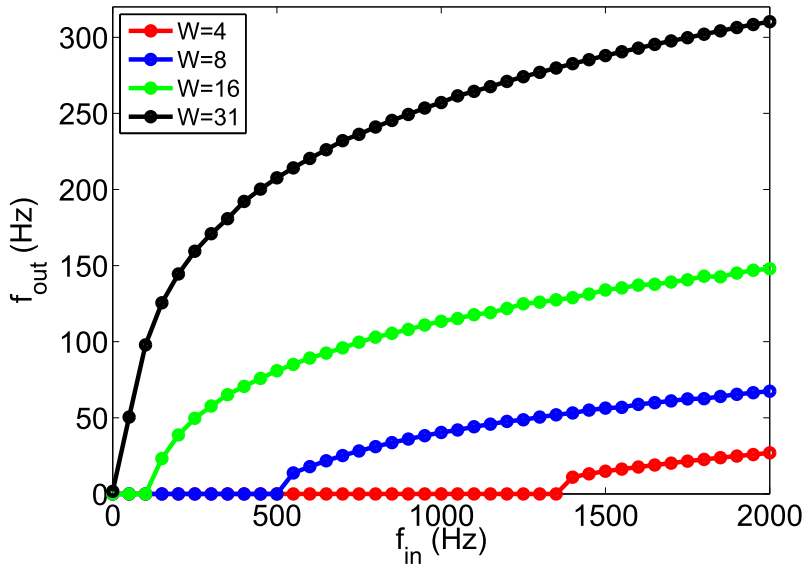


Figure 3.10. Neuron input-output response properties for various synaptic weights. Neuron output frequency versus frequency of incoming spikes, representing either a very high firing rate of a single source, or multiple sources at lower firing rates (Azghadi *et al.* 2014c).

bias generator circuit (Delbrück *et al.* 2010) was programmed using a microcontroller, which is integrated on the host PCB (see Fig. 3.5).

3.5 Chapter Summary

This chapter presented a programmable neuromorphic device, IFMEM, that can be used for implementing various spiking neural network architectures. The use of the AER representation for receiving inputs, computing with spikes, and transmitting signals in output, makes this device an ideal computational platform for building embedded neuromorphic event-based computational systems that process events generated by neuromorphic sensory systems (Liu and Delbrück 2010).

It was shown that the neurons and synapses implemented on the IFMEM chip are biophysically realistic and can provide biologically plausible time-constants if required. It was also demonstrated how the neuron-synapse circuits on the chip can be tuned to respond appropriately for different ranges of input firing rates. These features along with its programmability, make the IFMEM chip a very useful platform for implementing various synaptic plasticity rules and for different applications, such as pattern classification, and general purpose programmable neural learning systems.

Next chapter presents how the IFMEM device is used to realise various types of spike-based learning algorithms, based on either spike-timing relationships (e.g., STDP mechanisms), or spike rate-based ones (e.g. Bienenstock-Cooper-Munro (BCM) type rules). It also shows that how the IFMEM chip can be employed to classify complex rate-based patterns, using the TSTDP learning algorithm.

The implementations of various synaptic plasticity rules and also using them for a classification task, build the knowledge required for the main focus of this thesis, which is VLSI implementation of STDP rules, and using them in similar engineering applications.

Chapter 4

Implementing STDP and Pattern Classification on the IFMEM Chip

THIS chapter describes the implementation of STDP rules and a pattern classification neural network on the IFMEM neuromorphic system. It is shown that both PSTDP and TSTDP rules can be implemented on this neuromorphic setup, demonstrating the expected behaviours seen in biological experiments. This chapter shows how the STDP window can be generated using the silicon neurons and synapses available on the system. It also shows how the STDP rule is used for generating a competitive Hebbian learning behaviour observed in computational STDP experiments. Furthermore, the TSTDP learning algorithm is implemented on the chip. In order to test this implementation, it is utilised to reproduce a rate-based BCM learning behaviour. Obtained results show the usefulness of the TSTDP learning algorithm for generating the rate-based BCM learning behaviour. Finally, the implemented TSTDP learning mechanism is utilised to train a simple feedforward spiking neural network to classify some complex rate-based patterns. Obtained results show the high performance of the TSTDP rule in the targeted classification task. The experiments carried out in this chapter provide a comprehensive view of the STDP rules and their properties and features, which are essential when designing VLSI STDP synapses in the following chapters.

4.1 Introduction

As discussed in previous chapters, hardware implementations of spiking neural networks offer promising solutions for computational tasks that require compact and low power computing technologies. As these solutions depend on both the specific network architecture and the type of learning algorithm used, it is important to develop spiking neural network devices that offer the possibility to reconfigure their network topology and to implement different types of learning mechanisms. The previous chapter presented a neuromorphic multi-neuron VLSI device, IFMEM chip, with on-chip programmable event-based hybrid analog/digital circuits. The event-based nature of the input/output signals allows the use of Address-Event Representation infrastructures for configuring arbitrary network architectures, while the programmable synaptic efficacy circuits allow the implementation of different types of spike-based learning mechanisms (Moradi and Indiveri 2014, Azghadi *et al.* 2014c). This chapter focuses on the learning aspects of the IFMEM system, and shows how different Spike-Timing Dependent Plasticity learning rules can be implemented on-line, when the VLSI device is interfaced to a workstation. It will also be demonstrated how, after training, the VLSI device can act as a compact stand-alone solution for binary classification of correlated complex rate-based patterns.

Probably the most recognised synaptic plasticity rule in neuromorphic engineering is the Spike Timing Dependent Plasticity (STDP) (Bi and Poo 1998, Azghadi *et al.* 2013a). In this chapter, the focus is on implementing this important plasticity rules and reproducing its relevant behaviours observed in computational as well as biological experiments. Here, first the STDP learning algorithm is implemented to alter synaptic weights in the IFMEM system. The well-known exponential learning window associated with STDP rule (Bi and Poo 1998, Song *et al.* 2000) is also produced to verify the correct implementation of this rule using a silicon neuron biased to act linearly (as shown in Chapter 3). In addition to demonstrating the STDP learning window, it is shown how the synapses on the IFMEM chip, which their efficacies are altered by STDP, compete to control the spiking dynamic of their post-synaptic neuron, a feature that is characterised for STDP in Song *et al.* (2000).

After implementing the pair-based STDP rule and showing its associated window and replicating the bimodal behaviour in the weight distribution due to the competitive nature of STDP synapses, the chapter continues to explore the triplet-based STDP rule, which is the focus of this thesis. It is shown that this rule could reproduce rate-based

BCM-like behaviours, with a sliding threshold feature, as it is shown in the computational BCM model (Bienenstock *et al.* 1982, Cooper *et al.* 2004). It is then shown that the TSTDP learning mechanism with its parameters set to demonstrate the rate-based learning behaviour, is able to classify complex rate-based patterns with a high degree of correlation.

Reported results in this chapter are mainly presented in *The ACM Journal on Emerging Technologies in Computing Systems* (Azghadi *et al.* 2014c).

4.2 Spike Timing Dependent Plasticity (STDP)

The classical description of STDP has been widely used in various computational studies (Song *et al.* 2000, Iannella *et al.* 2010) as well as several VLSI implementations (Bofill-I-Petit and Murray 2004, Indiveri *et al.* 2006, Bamford *et al.* 2012b). The STDP rule that was discussed in Section 2.6.1, can be expressed as

$$\Delta w = \begin{cases} \Delta w^{+} = A^{+}\mathrm{e}^{\left(\frac{-\Delta t}{\tau_{+}}\right)} & \text{if } \Delta t > 0 \\ \Delta w^{-} = -A^{-}\mathrm{e}^{\left(\frac{\Delta t}{\tau_{-}}\right)} & \text{if } \Delta t \leq 0, \end{cases} \tag{4.1}$$

where $\Delta t = t_{\text{post}} - t_{\text{pre}}$ is the timing difference between a single pair of pre- and post-synaptic spikes. According to this model, the synaptic weight will be potentiated if a pre-synaptic spike arrives in a specified time window (τ_{+}) before the occurrence of a post-synaptic spike. Analogously, depression will occur if a pre-synaptic spike occurs within a time window (τ_{-}) after the post-synaptic spike. These time windows are not usually longer than 50 ms. The magnitude of potentiation/depression will be determined as a function of the timing difference between pre- and post-synaptic spikes, their temporal order, and their relevant amplitude parameters (A^{+} and A^{-}).

Fig. 4.1 demonstrates the conventional antisymmetric learning window associated with the pair-based STDP rule. For generating this window, first the neuron was set to fire spikes in response to a regular pre-synaptic spike train with the rate of 50 Hz. In this case the neuron shows a behaviour similar to the one shown in Fig. 3.8(a). Then outgoing spikes from the post-synaptic neuron were recorded. Next, the weight of the associated synapse was altered off-chip according to the PSTDP rule shown in Eq. 4.1. Figure 4.1 shows the resulting STDP weight changes that occurred due to the random time difference among pre- and post-synaptic spikes.

The figure suggests that the post-synaptic neuron spikes in a regular way similar to the pre-synaptic spikes applied to the synapse with PSTDP. It should be noted that during

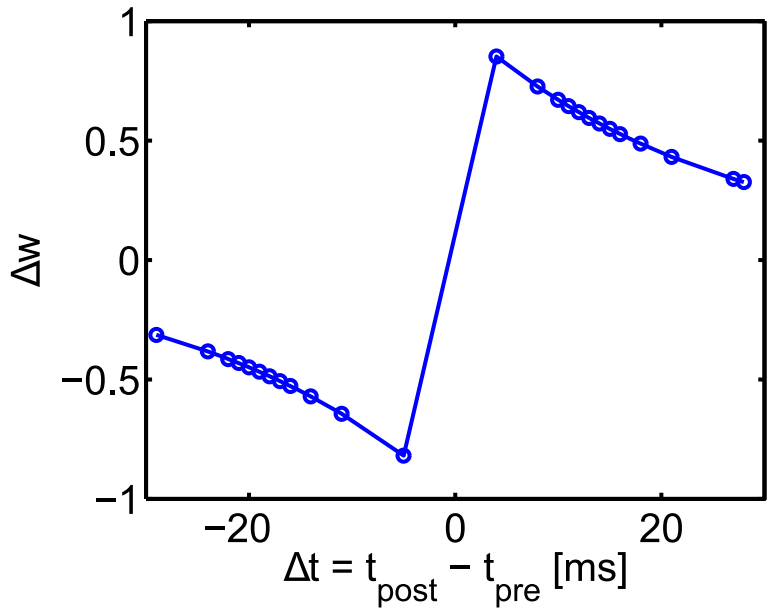


Figure 4.1. PSTDP learning window generated on the IFMEM neuromorphic device. A regular pre-synaptic spike train was applied to the neuron and the weights were modified according to the timing differences between nearest neighbour spikes in the pre- and post-synaptic spike trains. Note that, Δw in this figure determines the amount of weight change computed off-chip and according to the time differences between the spikes applied to the synapse (pre-synaptic spike) and those generated by the silicon neuron (post-synaptic spike). The STDP learning rule parameters are shown in Table 4.1.

all experiments performed in this thesis, the nearest neighbour (in contrast to all-to-all) spike interaction, in which only the immediate preceding and succeeding adjacent spikes are considered for weight modifications, is adopted. The synaptic parameters used for the STDP window experiment are shown in Table 4.1. One can change the amplitude as well as the time constants of the learning window using the parameters shown in this Table.

4.2.1 Competitive Hebbian Learning Through STDP

It is already verified that Hebbian learning has two substantial requirements to be developed. The first requirement is to control synaptic efficacy through some activity-dependent synaptic plasticity rules such as STDP, while the second requirement is a competition mechanism among synapses (Song *et al.* 2000). It is shown that, under specific circumstances, synapses, which their weights are governed by STDP, compete to control the spiking activity of their post-synaptic neuron. This competition leads to

Table 4.1. STDP parameters for producing STDP learning window on the IFMEM chip. The four required parameters of Eq. 4.1, for producing the STDP window (Song *et al.* 2000) are shown here.

Parameter name	A_2^+	A_2^-	τ_+ (ms)	τ_- (ms)
Value	1	1	25	25

divergence of synaptic weights into two distinguished groups. The first group is composed of strong synapses, which their input spikes have been more correlated so they became strong due to STDP. By contrast, the second group includes weak synapses, whose input spikes have been less correlated and therefore, they became weaker in result of STDP (Song *et al.* 2000). The competition and the resulting stable synaptic weight distribution, do not arise unless the following conditions are satisfied: (i) Imposing a hard boundary on the strength of individual synapses, and (ii) Setting synaptic parameters in a way that, synaptic weakening through STDP slightly outweighs synaptic strengthening, i.e. $A^+\tau_+ < A^-\tau_-$.

In the following, we emulate the mentioned competitive Hebbian learning experiment on our developed neuromorphic system, through a simple network composed of a single IF neuron and its 32 affiliated 5-bit digital synapses. We demonstrate how the digital synaptic weights, which act as virtual synapses governed by STDP, diverge into two distinguished groups over time. The experiments are conducted in the following manner.

First, we calibrate the silicon neuron on the chip, in a way that it is reasonably excited so that it can respond to its input spikes by firing action potentials. This is possible through increasing the injection current applied to the neuron. There are excessive control parameters for setting the dynamics of the silicon neuron including a parameter for controlling its spiking threshold, a parameter for adapting its spiking frequency, as well as a parameter for setting its refractory periods that also play fundamental roles in the spiking activity of the neuron—please refer to Moradi and Indiveri (2014) for further information.

Second, we set all 32 digital synaptic weights, which are 5-bit asynchronous SRAM cells, and therefore have 32 efficacy states, to their mid values of 16.

Table 4.2. STDP parameters for producing competitive Hebbian learning behaviour on the IFMEM chip. The four required parameters of Eq. 4.1, for producing the competitive Hebbian learning behaviour (Song *et al.* 2000) are shown here.

Parameter name	A_2^+	A_2^-	τ_+ (ms)	τ_- (ms)
Value	0.5	0.527	20	20

Next, we apply 32 independent Poissonian spike trains with firing rates of 50 Hz to all 32 synapses. These spike trains will be time-multiplexed on a single DPI synapse, and it produces an integration of 32 synaptic currents that are proportional to each of the 32 synaptic weights. The integrated current then generates the EPSC current, which in turn is applied to the tuned post-synaptic silicon neuron. The neuron then generates action potentials in response to the synaptic currents it receives.

The timing of the post-synaptic spikes generated by the silicon neuron and the timing of pre-synaptic spikes applied to each of the 32 synapses then govern the magnitude of changes to the digital weights stored in the SRAM cells affiliated with each synapse, according to the STDP rule presented in Eq. 4.1.

Fig. 4.2 demonstrates how synaptic weights in the mentioned setup evolve over time to reach a homeostatic state, in which synaptic weights are approximately either weakened or strengthened. At time $= 0$ s, all digital synaptic weights are set to their middle value (i.e. $w[4...0] = (10000)_2 = 16$), so that a high firing rate of the silicon neuron is achieved. Then, the 32 synaptic weights are modified by STDP rule, which is implemented off-chip and updates synaptic weights stored in the SRAM cells, based on the timing difference between the current post-synaptic spike, and pre-synaptic spikes arrived immediately before or after this post-synaptic spike, in each of the 32 synapses. The synaptic parameters for this experiment are shown in Table 4.2.

After updating the weights off-chip, the modified weights are written back to their related SRAM cells through AER communication system. In response to these changes weights start to modify and distribute across the whole range of 32-state weight spectrum (see time $= 100$ s). Eventually the weight of the 32 synapses diverge into two groups, one includes stronger synapses and the other one contains weaker synapses. This divergence is mainly because those synapses whose spike trains are more correlated get stronger due to STDP, while their stronger weights acts as a positive feedback

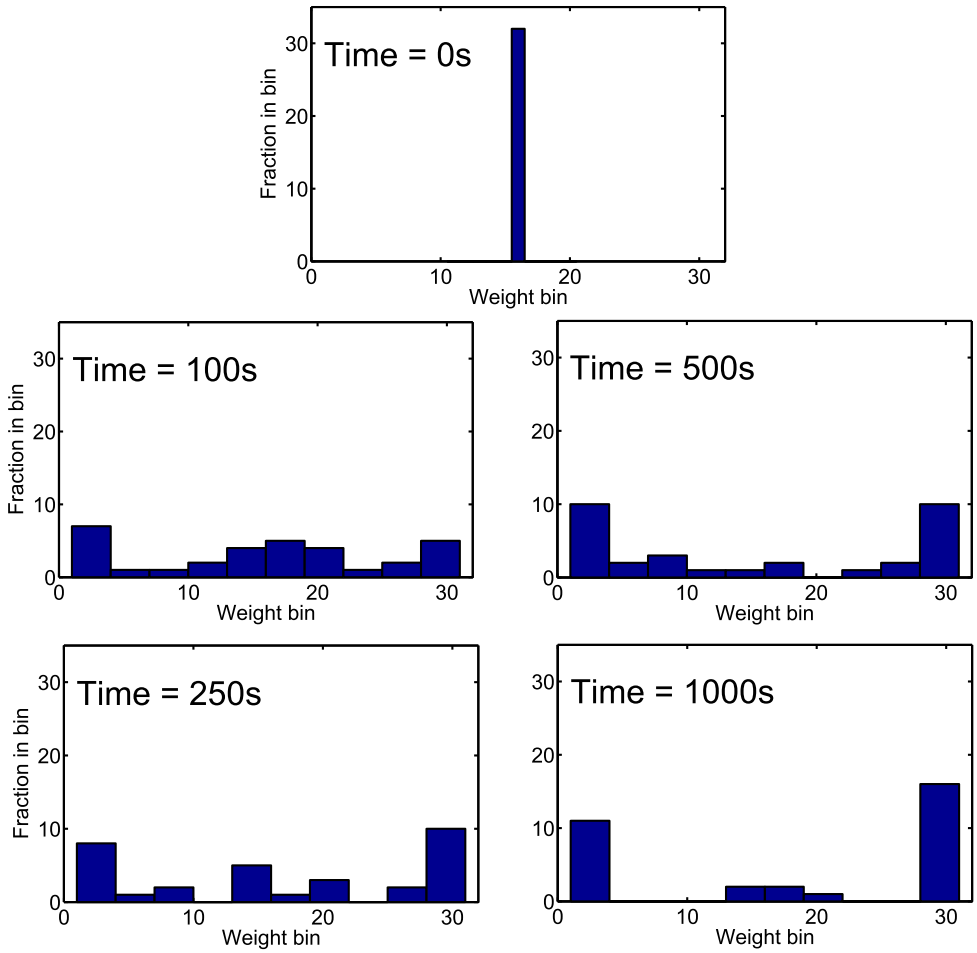


Figure 4.2. Synaptic weights evolve to reach an equilibrium state, when modified by STDP learning rule. Here, 32 synaptic weights (weight bins) each one with 32 digital states, are altered by STDP over time. The top figure shows that all 32 synaptic weights are set to 16 in the beginning at time $= 0$ s, i.e. the fraction of weights in weight bin 16 is 32. The other figures show the evolution of weights over time to reach a steady state at time $= 1000$ s. The synaptic weights stay almost fixed thereafter, and the post-synaptic neuron firing rate held in an almost direct relation to the mean firing rate of pre-synaptic spike trains (Azghadi *et al.* 2014c).

and help their weights gets even stranger. On the other hand, those synapses that receive less correlated spike trains get weaker, in this STDP competition. This feature can be used in an unsupervised form of Hebbian learning based on the correlation among input spikes.

At time = 1000 s, the weights reach an equilibrium state, in which synaptic weights do not alter anymore and the spiking activity of the post-synaptic neurons is proportionate to the mean firing rate of pre-synaptic spikes. This is an interesting feature of STDP, which let the neuron reach a steady state.

It should be noted that, in the presented experiment, the synaptic weights are bounded between 1 and 31, and each weight is rounded either upward or downward to its closest digital weight value after each synaptic weight update. In addition, further emulations suggest that the initial distribution of the 32 synaptic weights does not have a significant impact on the distribution of final weights.

4.2.2 Implementing BCM through STDP

Although BCM is an inherently rate-based rule and depends on the activities of pre- and post-synaptic neurons, recent studies have shown that timing-based triplet STDP learning rule can reproduce BCM-like functionality (Gjorgjieva *et al.* 2011, Pfister and Gerstner 2006). Here, it is demonstrated how this rate-based functionality can be realised by our software-hardware system, by using the triplet STDP learning rule (Pfister and Gerstner 2006, Gjorgjieva *et al.* 2011) to update the 5-bit synaptic weight values of the IFMEM chip.

As already mentioned, the triplet-based STDP can be formulated as

$$\Delta w = \begin{cases} \Delta w^{+} = A_2^{+} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} + A_3^{+} \mathrm{e}^{(\frac{-\Delta t_2}{\tau_y})} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} \\ \Delta w^{-} = -A_2^{-} \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})} - A_3^{-} \mathrm{e}^{(\frac{-\Delta t_3}{\tau_x})} \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})}, \end{cases} \tag{4.2}$$

where $\Delta w = \Delta w^{+}$ for $t = t_{\text{post}}$ and if $t = t_{\text{pre}}$ then the weight change is $\Delta w = \Delta w^{-}$. Here, A_2^{+}, A_2^{-}, A_3^{+} and A_3^{-} are potentiation and depression amplitude parameters, $\Delta t_1 = t_{\text{post(n)}} - t_{\text{pre(n)}}$, $\Delta t_2 = t_{\text{post(n)}} - t_{\text{post(n-1)}} - \epsilon$ and $\Delta t_3 = t_{\text{pre(n)}} - t_{\text{pre(n-1)}} - \epsilon$, are the time differences between combinations of pre- and post-synaptic spikes. Here, ϵ is a small positive constant which ensures that the weight update uses the correct values occurring just before the pre- or post-synaptic spike of interest, and finally τ_-, τ_+, τ_x and τ_y represent time constants (Pfister and Gerstner 2006).

It has been shown (Pfister and Gerstner 2006) that for Poisson distributed spike trains Eq. 4.2 can be approximated as:

$$\langle dw/dt \rangle = -A_2^{-}\tau_- \rho_{\text{pre}} \rho_{\text{post}} - A_3^{-}\tau_-\tau_x \rho_{\text{pre}}^2 \rho_{\text{post}} + A_2^{+}\tau_+ \rho_{\text{pre}} \rho_{\text{post}} + A_3^{+}\tau_+\tau_y \rho_{\text{post}}^2 \rho_{\text{post}} \tag{4.3}$$

where ρ_{pre} and ρ_{post} represent the mean firing rates of the pre- and post-synaptic spike trains, respectively.

Generally, the BCM theory suggests that the synaptic weight changes have a linear relationship with the pre-synaptic, and a non-linear relationship with the post-synaptic mean firing rates (Bienenstock *et al.* 1982). Therefore, a general description of the BCM rule can be written as:

$$\langle dw/dt \rangle = \rho_{\text{pre}} \cdot \phi(\rho_{\text{post}}, \theta) \tag{4.4}$$

where ϕ is a function that satisfies the conditions $\phi(\rho_{\text{post}} > \theta, \theta) > 0$, $\phi(\rho_{\text{post}} < \theta, \theta) < 0$ and $\phi(0, \theta) = 0$. Essentially, if the post-synaptic firing rate, ρ_{post}, is below the threshold θ, then dw/dt is negative and the synaptic weight is depressed. Conversely, the synaptic weight is potentiated if the post-synaptic firing rate is larger than the threshold θ, and it is left unchanged if $\phi = 0$, i.e., if $\rho_{\text{post}} = \theta$ (Pfister and Gerstner 2006).

The Eqs. 4.3 and 4.4 can be mapped together, if two conditions are satisfied. The first condition requires having a linear relationship between the pre-synaptic firing activity, ρ_{pre}, and the synaptic weight change, $\langle dw/dt \rangle$, as shown in Eq. 4.4. This condition is satisfied if $A_3^- = 0$, in the triplet STDP equation (Eq. 4.3). This will lead to a minimal version of the TSTDP rule presented in Pfister and Gerstner (2006), which has been shown to account for various synaptic plasticity neuroscience experiments, including those dealing with higher order spike trains (Wang *et al.* 2005). The second condition requires that the sliding threshold θ, that determines the frequency, in which depression turns to potentiation, is proportional to the expectation of the p^{th} power of the post-synaptic firing rate (ρ_{post}) (Pfister and Gerstner 2006, Bienenstock *et al.* 1982). This second condition can be satisfied if the threshold of the BCM rule is defined as

$$\theta = \langle \rho_{\text{post}}^p \rangle (A_2^- \tau_- + A_2^+ \tau_+) / \rho_0^p A_3^+ \tau_+ \tau_y. \tag{4.5}$$

Given this equation, the sliding threshold effect of the BCM rule is proportional to the post-synaptic firing rate, with the proportionality factor set by the STDP rule parameters. Previous studies have shown the possibility of mimicking the effects of BCM rule through TSTDP (Azghadi *et al.* 2013a). However, similar to the experiments performed in Gjorgjieva *et al.* (2011), they have used independent pre- and post-synaptic spike trains with mean firing rate of ρ_{pre} and ρ_{post}, respectively.

To implement BCM via the triplet STDP rule in the IFMEM chip setup, we used a single synapse, connected to a post-synaptic silicon neuron and changed its efficacy using the STDP rule of Eq. (2.3). At the beginning of the experiment, the initial weight of the synapse is set to its maximum value of 31 (i.e. $w[4...0] = (11111)_2$). This high synaptic weight makes the post-synaptic neuron fire at a high rate, proportional to the pre-synaptic firing rate (Azghadi *et al.* 2013a). The pre-synaptic spike train here is a Poisson spike train, similar to the spike trains used in previous studies (Gjorgjieva *et al.* 2011, Azghadi *et al.* 2013a). Using the AER protocol, we transmitted the software generated Poisson pre-synaptic spike train to the targeted post-synaptic silicon neuron, via a synapse with an efficacy proportional to its weight stored in the corresponding SRAM cell. The software pre-synaptic spike train, and the spike train produced by the silicon neuron, are then used to calculate the amount of weight changes in the corresponding synaptic efficacy, according to a minimal model of triplet STDP (Gjorgjieva *et al.* 2011).

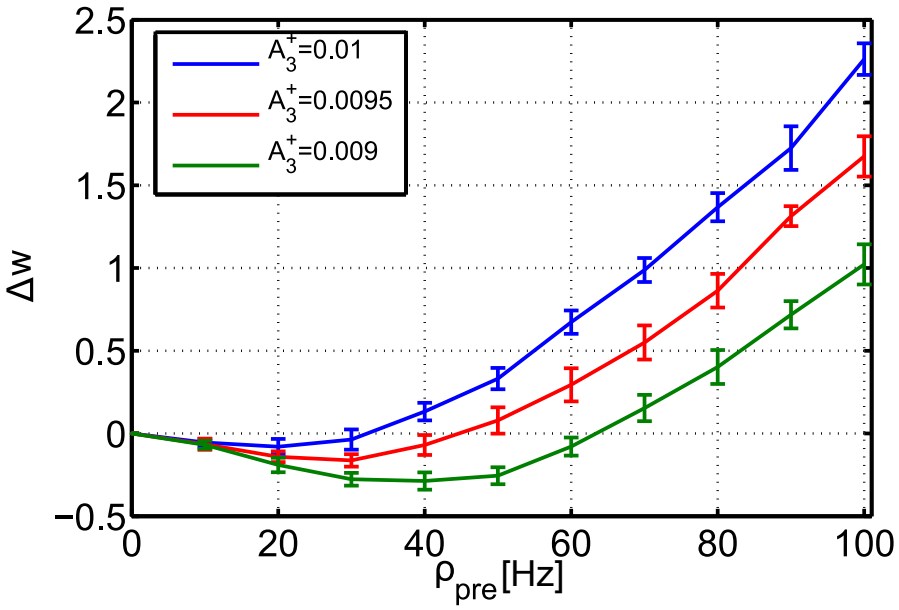


Figure 4.3. The BCM rule is implemented through TSTDP rule on the IFMEM neuromorphic chip. The sliding threshold feature of the rate-based BCM rule is replicated through Triplet STDP rule, implemented on the IFMEM chip (Azghadi *et al.* 2014c).

Figure 4.3 shows the total amount of weight change in response to Poisson spike trains of 20 s length, for a range of pre-synaptic spike rates from 0 Hz up to 100 Hz. In this figure, the sliding threshold feature of the BCM learning rule is regenerated through changing the amount of one of the parameters of the TSTDP learning rule, i.e. A_3^+. According to Eq. 4.5, with increase in A_3^+ parameter, the threshold decreases and slides

Table 4.3. Optimised TSTDP model parameters for generating BCM-like behaviour on the IFMEM chip. The eight required parameters of Eq. 4.2, for producing a BCM-like behaviour (Pfister and Gerstner 2006) are shown here. Note that x indicates 'don't care'.

Parameter name	A_2^+	A_2^-	A_3^+	A_3^-	τ_+ (ms)	τ_- (ms)	τ_y (ms)	τ_x (ms)
Value	0	0.0068	see Fig. 4.3	0	16.8	33.7	114	x

toward lower post- synaptic firing rates. Please note that, in the presented experiment, the silicon neuron parameters, as well as the synaptic weight parameters in its corresponding physical synapse, i.e. the differential pair integrator, are calibrated in a way that pre- and post-synaptic neuron are in a relatively linear relationship (Moradi and Indiveri 2014, Azghadi *et al.* 2013d). In this figure, each data point corresponds to the mean of the weight changes over 10 trials, and the error bar represents the standard deviation of the weight change over these trials. This amount of weight change can then be discretised and written back into the SRAM. The STDP parameters that have been used in this experiment are shown in Table 4.3.

4.3 Classification of Complex Correlated Patterns

In this Section, classification of complex rate-based patterns is targeted using the TSTDP learning rule. For implementing the targeted classification task, the TSTDP learning rule, with its parameters tuned for exhibiting BCM behaviour (see Fig. 4.3) are used. This section demonstrates how the TSTDP rule implemented on the IFMEM device can perform classification of binary patterns with high levels of correlations.

The neural classifier implemented on the chip is composed of one neuron and 30 synapses, which are arranged in a single layer perceptron-like architecture. The goal is to train the perceptron synaptic weights, via the TSTDP algorithm, to learn to distinguish two input patterns, UP and DOWN, in an unsupervised fashion. After training, the hardware perceptron should be able to respond with a high firing rate to pattern UP, and a low one to pattern DOWN. This is a similar experimental scenario, to the semi-supervised learning scenario utilised in an identical classification task performed using spiking neural networks (Giulioni *et al.* 2009).

The two UP and DOWN patterns can have various degrees of correlations. The correlation determines the amount of overlap in the input synapses used, and the similarity in the output response of the neuron. When there is no correlation, one pattern is applied to 15 random synapses and the other pattern is applied to the remaining 15 synapses (no overlap).

The two patterns are defined as follows. The pattern UP stimulates 15 synapses with Poisson spike trains that have a high mean firing rate of 300 Hz, while pattern DOWN comprises 15 Poisson spike trains with a low mean firing rate of 20 Hz. Therefore, in the case of zero correlation, the two patterns are likely to produce different outputs (depending on the values of the synaptic weights) even before learning. However, for the case of non-zero correlations, a random subset of N input synapses are always stimulated by high mean firing rate spike trains of 300 Hz, while the rest of the synapses are assigned to the two UP and DOWN patterns. For instance, if the number of correlated synapses is 10, 10 randomly synapses are stimulated by Poisson spike trains of 300 Hz, and the remaining 20 synapses will be reserved for the UP and DOWN patterns. In this case, pattern UP (DOWN) is presented as 10 high (low) rate spike trains that are applied to 10 random synapses from the 20 synapses, and pattern DOWN (UP) is presented to the remaining 10 synapses. In this case, because of the N common high input synapses, the two patterns will have closer mean firing rates, and therefore their classification becomes more challenging. Therefore, in the beginning of learning phase, the output frequency range of the perceptron cannot be distinguished between the two patterns and as a result, learning is required to classify the two patterns.

The training phase is composed of several trials. In each trial, one of the two patterns, UP or DOWN is randomly applied to the 30 input synapses, with a set degree of correlation, and with a new distribution of Poisson spikes. The two patterns have equal probability to be selected. For each trial, the synaptic weights are modified according to the TSTDP. In our experiment the synaptic weights reach a steady state and do not change significantly after about 20 trials, in which the input spike trains lasted 10 s each.

Figure 4.4 shows how the distribution of the neuron output firing rates, in response to the two different patterns, changes with learning, after 1, 5, 10, and 20 trials. The output neuron rates were collected over 20 classification runs, with each run comprising 20 learning trials, while 20% correlation is set among the two patterns. In each run the synaptic weights are initialised to random 5-bit values, the definition of UP

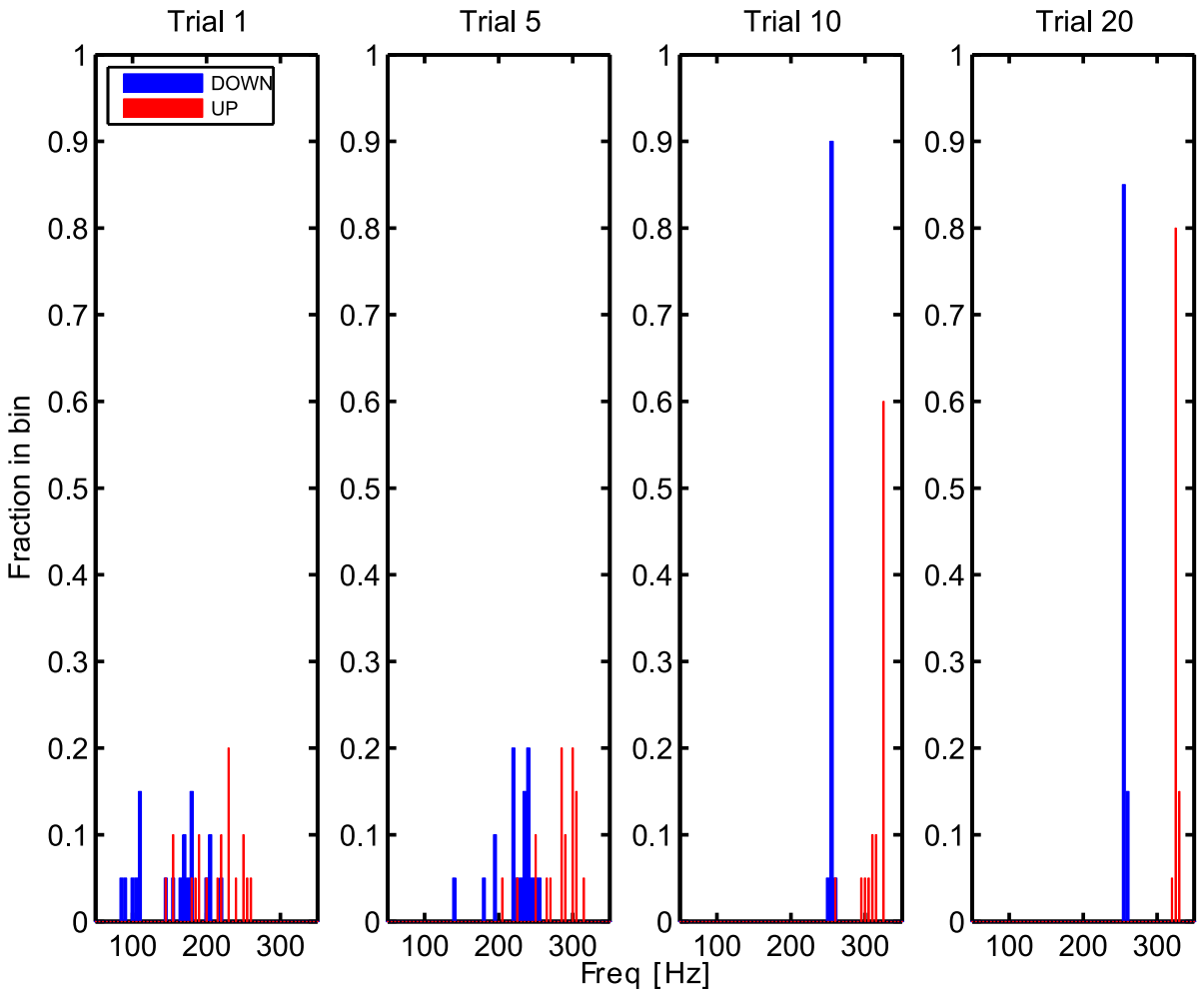


Figure 4.4. Distribution of the neuron output frequencies during different stages of learning. In the beginning of the learning phase, when initial weights are random, the neuron cannot distinguish between the two patterns. During the learning trials, the synapses are being modified and the neuron begins to effectively discriminate between the two patterns, UP and DOWN, from trial 20. In this experiment the correlation is equal to 20 %, i.e., there are 6 inputs that are common to the two patterns that always receive high firing rates (Azghadi *et al.* 2014c).

and DOWN patterns is changed, and a new random order of UP and DOWN patterns applied across trials, is defined.

As Fig. 4.4 shows after one stage of learning, trial 1, the neuron is still unable to distinguish between the two patterns, even though the weights have changed once, in result of applying either pattern UP or DOWN to the network and updating the synaptic weights according to the TSTDP learning rule. As the learning phase proceeds, the

neuron starts to fire with higher rates for pattern UP, compared to the rates for pattern DOWN. Figure 4.4 shows that after 20 trials, the neuron becomes significantly specialised to the two patterns and fire with a higher rates for pattern UP, and with a lower rate for pattern DOWN.

As expected, the utilised TSTDP learning rule tends to decrease the weights of the synapses targeted by the DOWN pattern, while it tends to increase the weights of both the UP and correlated (overlapping) synapses. After learning, the neuron will therefore fire with high firing rates when stimulated with UP patterns, and low firing rates when stimulated by DOWN patterns. While after a few trials (e.g. see second and third panels of Fig. 4.4) the neuron already performs above chance levels, many trials (20 in our experiments) are required to unambiguously classify the two patterns.

Regarding the classification performance, our results show that the implemented classification network performs robustly and holds for also large amount of correlation, i.e. more than 50 % correlation, in the input patterns. In terms of classification accuracy, we consider a DOWN pattern correctly classified if the neuron output frequency is less than a set threshold in response to that pattern; and similarly, an UP pattern is correctly classified, if the neuron response to such pattern has a firing rate higher than the threshold. In our experiments, the classifier has 100 % correct performance, even with correlation levels of 87 % (i.e., 26 overlapping synapses), if the classification threshold is adaptive (e.g., if it is set just below to the minimum frequency in response to the UP patterns). What changes however is the difference in the responses to the two patterns. Figure 4.5 shows how this difference decreases as the correlation among the input patterns increases.

Figure 4.5 depicts that with increase in the correlation level between the two patterns, the number of common active synapses between the two patterns increases and therefore, the neuron firing rate for the two patterns becomes closer. As a result, the difference in the output frequency for pattern UP and DOWN will decrease.

In contrary to an adaptive threshold, one might consider selecting a fixed threshold for the whole range of correlations. In this case, the classification accuracy of the implemented classifier will still be 100%, for patterns with up to 47% of correlations (correlation level = 14). However, the accuracy drops to 50% afterwards, because in the case of a fixed threshold, only UP patterns are correctly classified for correlation level great than 14.

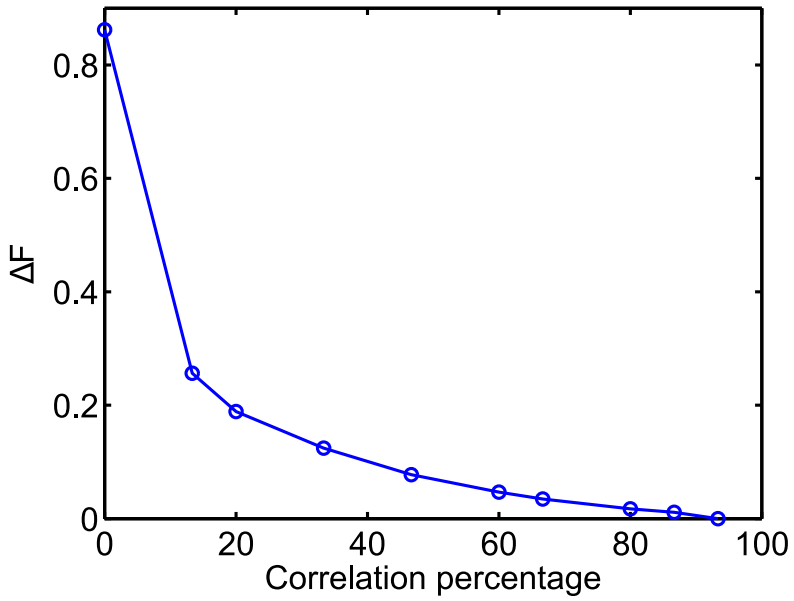


Figure 4.5. The performance of the classifier implemented on the IFMEM chip. Here $\Delta F = (F_{\mathrm{UP}}^{\mathrm{min}} - F_{\mathrm{DOWN}}^{\mathrm{max}})/F_{\mathrm{UP}}^{\mathrm{min}}$, where $F_{\mathrm{UP}}^{\mathrm{min}}$ and $F_{\mathrm{DOWN}}^{\mathrm{max}}$ are the minimum and the maximum frequencies for pattern UP and DOWN, respectively, for all 20 runs at the end of learning in trial 20. (Azghadi *et al.* 2014c).

It is worth mentioning that, in the presented implementation, contrary to previous works (Mitra *et al.* 2009, Giulioni *et al.* 2009), no *teacher signal* is used to make the neuron fire with a specific rate for pattern UP/DOWN, rather it is relied on the input patterns and random initial weights to make the neuron fire. Therefore, the implemented classification system utilises an unsupervised learning mechanism for classifying two correlated patterns.

4.4 Chapter Summary

This chapter presented how the programmable neuromorphic chip that was presented in previous chapter can be used in a hybrid software-hardware system to implement different types of spike-timing dependent plasticity learning rules. It was also demonstrated how these rules can reproduce competitive Hebbian learning and rate-based behaviours, even with the limitations of the hardware implementation (5-bit resolution for the weights, mismatch of the analog subthreshold circuits, etc.). Finally it was described how the hybrid software-hardware learning setup proposed can be used to

train a perceptron to perform binary classification in an unsupervised way, and to be robust to extremely high correlations in the input patterns.

The device and setup proposed in this chapter, therefore represents a useful real-time low-power computing platform for exploring the effectiveness of different types of spike-based learning algorithms, validating their performance at run-time on real-time custom analog/digital hardware, and implementing robust perceptron-like neural networks to carry out real-time classifications tasks. If the task can be solved after training the weights of the neural network, without requiring continuous or on-line training, then the platform proposed represents a stand-alone compact and low-power alternative to standard full-digital computing solutions (no PC is required in the loop).

This chapter along with the previous one, provided us with a good knowledge on the modelling, implementation and behaviour of both PSTDP and TSTDP rules. This knowledge and experience is quite helpful while designing these rules in VLSI, specially when these synaptic plasticity rules are employed in a synapse and integrated with a silicon neuron. The experiments performed in these two chapters are also essential if using the VLSI versions of the STDP circuits in a neuromorphic system for specific tasks such as pattern classification is targeted.

After gaining knowledge on the structure of the STDP rules and the behaviours these rules show, we can now start designing these rules in VLSI and use them to show various behaviours already observed using the IFMEM chip. However, before designing new circuits, a review of previous VLSI designs for various synaptic plasticity rules is required to build our knowledge on the state-of-the-art synaptic plasticity rules in silicon. The next chapter, is dedicated to discussion and review of various synaptic plasticity models in VLSI.

Chapter 5

Spike-based Synaptic Plasticity Rules in Silicon

THIS chapter reviews Very Large Scale Integration (VLSI) circuit implementations of various synaptic plasticity rules, ranging from phenomenological ones (i.e. timing-based, rate-based, or hybrid rules) to biophysically realistic ones (e.g. based on calcium dependent models). It discusses the application domains, weaknesses and strengths of the various representative approaches proposed in the literature and provides deeper insight into the challenges that engineers face when designing and implementing synaptic plasticity rules in order to utilise them in real-world applications. The chapter also proposes and discusses various counter approaches to tackle the challenges in neuromorphic engineering. The review performed in this chapter helps build knowledge useful for the design of new VLSI circuits for synaptic plasticity rules while considering the challenges, applications and effective design methods and techniques.

5.1 Introduction

For over a century, there has been considerable effort in attempting to find answers to the question: "how does learning and memory take place in the brain?" Although there is still no general agreement, neuroscientists agree on some general rules and hypotheses to answer this question (Hebb 2002, Cooper *et al.* 2004, Sjöström *et al.* 2008). It is agreed that learning and memory in the brain are governed mainly by complex molecular processes, which give rise to a phenomenon called synaptic plasticity. The actions of synaptic plasticity can manifest themselves through alterations in the efficacy of synapses that allow networks of cells to alter their communication. Hebbian learning, postulated by Donald Hebb in 1949 (Hebb 2002), is the foremost recognised class of synaptic plasticity rules. It has formed the foundation of a number of other plasticity rules (Gerstner and Kistler 2002). Synaptic plasticity rules can be categorised into two general groups i.e. Short Term Synaptic Plasticity (STSP) and Long-Term Synaptic Plasticity (LTSP). While STSP plays a fundamental role in decoding and processing neural signals on short time scales (Zucker and Regehr 2002, Buonomano 2000), it is LTSP that is responsible for learning and memory in the brain (Sjöström *et al.* 2008). This type of plasticity produces long lasting depression or potentiation in the synaptic weights. Specifically long-term plasticity can produce an increase in synaptic efficacy, which results in Long Term Potentiation (LTP) of the synapse, or it can produce a decrease in the synaptic weight, resulting in Long Term Depression (LTD). It is widely believed that these long term processes are the basis for learning and memory in the brain. From an engineering perspective, it is important to understand how these LTSP mechanisms can be translated into physical models that can implement adaptive learning capabilities in artificial SNNs.

There are multiple approaches that address this problem, which depend on the target application domain for the artificial SNN (e.g., for basic research, for providing tools to computational neuroscientists, or for practical real-time engineering applications). For example, some of these approaches mimic the mechanisms occurring in real synapses and model them at a *phenomenological* level, while other approaches take into account many details of the real plasticity mechanism to model biological synapses as faithfully as possible.

In order to decipher the mystery of learning through synaptic plasticity, neuroscientists typically postulate their hypotheses on how the brain learns and propose specific models of plasticity rules that can explain their theoretical and experimental observations.

Then these hypotheses can be implemented in software or hardware and tested with real-world stimuli to verify their values in addressing real-world challenges. While software implementations are ideal for exploring different hypotheses and testing different models, dedicated hardware implementations are commonly used to implement efficient neural processing systems that can be exposed to real-world stimuli from the environment and process them in real-time, using massively parallel elements that operate with time constants that are similar to those measured in biological neural systems. This approach is followed for both attempting to get a deeper understanding of how learning occurs in physical systems (including the brain), and for realising efficient hardware systems that can be used to carry out complex practical tasks, ranging from sensory processing to surveillance, robotics, or brain-machine interfaces. The synaptic plasticity models developed by neuroscientists are typically translated into electronic circuits and implemented using conventional VLSI technologies. Currently, many of these models form the foundations for developing VLSI "neuromorphic systems" (Mead 1990, Indiveri and Horiuchi 2011). This chapter reviews and discusses the most representative implementations of synaptic plasticity models presented in the literature and compare them in terms of complexity and usefulness for various applications.

The remainder of this chapter is organised as follows. Section 5.2 describes a number of fundamental building blocks useful for implementing those synaptic plasticity rules that were discussed in Chapter 2. In Section 5.3, we review a range of representative implementations of synaptic plasticity rules, designed using the aforementioned building blocks. Section 5.4 addresses the challenges that neuromorphic engineers encounter when designing these systems. Section 5.5 discusses the previously mentioned synaptic plasticity designs and approaches and points out their benefits and limitations. Section 5.6 demonstrates examples of how synaptic plasticity rules can be employed in practical applications. Finally, Section 5.7 summarises our concluding remarks for this chapter.

Presented results, review and discussion in this chapter are mainly presented in *The Proceedings of the IEEE* (Azghadi *et al.* 2014b).

5.2 Building Blocks for Implementing Synaptic Plasticity Rules in VLSI

This Section reviews the most common and useful electronic building blocks required for implementing various types of synaptic plasticity rules in VLSI.

5.2.1 Fundamental Circuit Elements

The storage of synaptic weight values, and of other state variables that need to be memorised in models of synaptic plasticity requires memory elements in VLSI. Typically, analog values are stored using capacitors. In VLSI technology, capacitors can be implemented using Metal-Oxide-Semiconductor Capacitors (MOSCAPs), or multiple layers of poly-silicon separated by an insulator (typically silicon-dioxide). These solutions usually offer the most compact and convenient way of storing variables, but they have the limitation of being leaky: as the charge stored in these devices tend to slowly leak away due to imperfect insulator used in building these devices. Alternative ways of storing analog variables involve the use of floating-gate devices (Ramakrishnan *et al.* 2011), or of dedicated analog-to-digital converters (ADCs) and digital memory circuits, such as Static Random Access Memory (SRAM) elements (Moradi and Indiveri 2011, Azghadi *et al.* 2013d). Considering the required time constants, the targeted network and its desired application, these approaches are more/less bulky and/or convenient, compared to the storage on VLSI capacitors, which is not applicable for long-time storage. Another issue that should be taken into account when selecting the storage technique for synaptic weights is the required precision needed for a given application (Pfeil *et al.* 2012). This issue is discussed in Section 5.4.8.

While capacitors are passive devices, Metal-Oxide-Semiconductor Field Effect Transistors (MOSFETs) are active and represent the main basic building block in VLSI technology (Weste and Eshraghian 1994). Depending on the voltage difference between the transistor gate and source terminals, V_{gs}, their current-voltage characteristic can dramatically change. In particular, if $V_{gs} > V_{th}$, the transistor acts in its above-threshold (i.e. strong inversion) regime. On the other hand, if $V_{gs} < V_{th}$, the transistor operates in its subthreshold (i.e. weak inversion) regime (Liu *et al.* 2002).

Neuromorphic engineers are interested in the subthreshold domain for two essential reasons. The first reason is the exponential relationship between the drain current, I_D

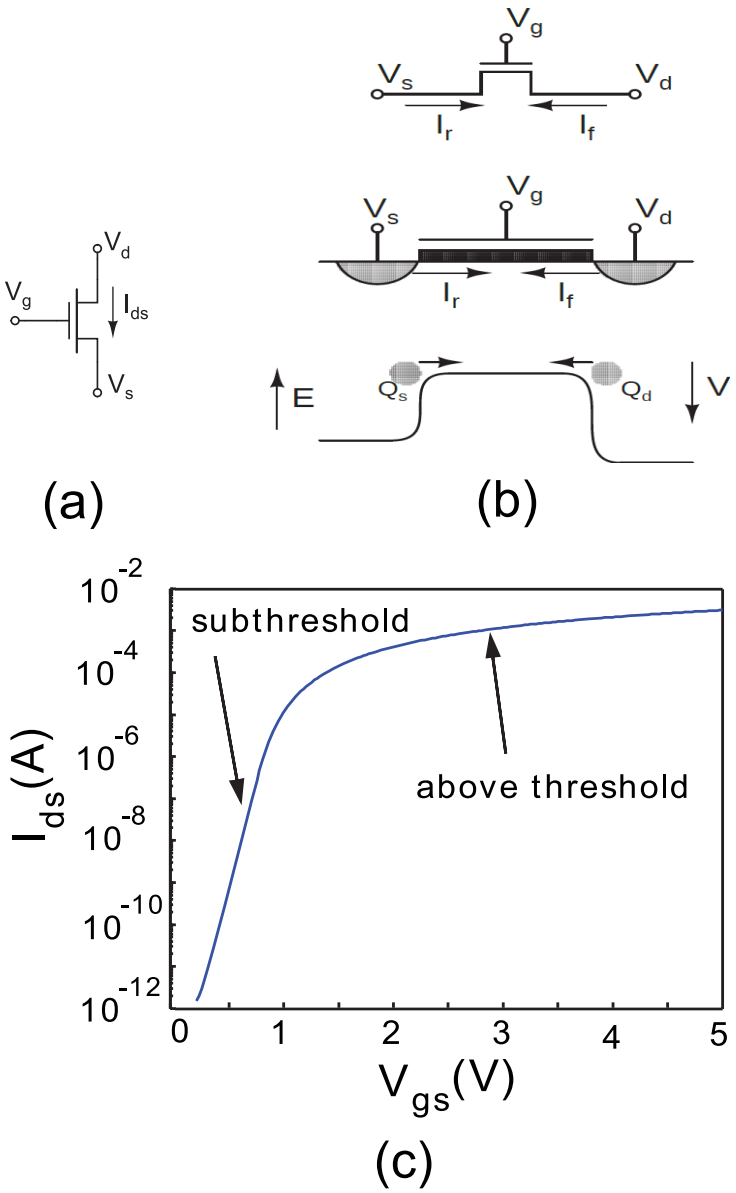


Figure 5.1. NMOS transistor in subthreshold. (a) Symbol for a NMOS transistor. (b) The drain-source current, I_{ds}, of a NMOS device in its subthreshold region of operation is a summation of two currents with opposite directions. (c) Current-voltage characteristic of the NMOS transistor, which shows significantly different behaviour for above and below threshold (Liu *et al.* 2002).

of a transistor and its gate voltage, V_{g}, as shown in Eq. 5.1,

$$I_{\text{ds}} = I_0 e^{\kappa_n V_{\text{g}}/U_T}(e^{-V_{\text{s}}/U_T} - e^{-V_{\text{d}}/U_T}), \tag{5.1}$$

where I_0 is a current-scaling parameter, κ_n denotes the n-type MOSFET subthreshold slope factor, U_T represents the thermal voltage, and V_d, V_g, and V_s are the drain, gate, and source voltages of the transistor, as it is shown in Fig. 5.1(a), relative to the bulk

potential, respectively (Liu *et al.* 2002). Fig. 5.1(b) shows that, the drain-source current shown in Eq. 5.1 is a summation of two currents in opposite directions, one is called forward current, I_f, which is a function of the gate-source voltage, and flows from the drain to the source, and the other current, I_r, the reverse current, flows from the source to the drain

$$I_{\text{ds}} = I_0 e^{\kappa_n V_{\text{g}}/U_T - V_{\text{s}}/U_T} - I_0 e^{\kappa_n V_{\text{g}}/U_T - V_{\text{d}}/U_T} = I_f - I_r. \tag{5.2}$$

If $V_{\text{ds}} > 4U_T \approx 100$ mV, as the energy band diagram in Fig. 5.1(b) shows, because of the larger barrier height (in contrast to the $V_{\text{ds}} < 4U_T$ state, where barrier heights are almost equal), the concentration of electrons at the drain end of the channel will be much lower than that at the source end, and therefore the reverse current, from source to drain, I_r becomes negligible, and the transistor will operate in the subthreshold saturation regime. Therefore, there will be a pure exponential relationship between V_{gs} and I_{ds} as

$$I_{\text{ds}} = I_0 e^{\kappa_n V_g/U_T - V_s/U_T}. \tag{5.3}$$

This exponential behaviour is analogous to the exponential relationship between the ionic conductance of a neuron and its membrane potential. Therefore, a transistor is able to directly emulate the required behaviour of an ionic conductance (Andreou *et al.* 1991). Figure 5.1(c), which is a log-linear plot, shows the drain-source current characteristic of a NMOS device, as a function of its gate-source voltage. The figure shows the exponential dependence of the current to the gate-source voltage, below the device threshold. It also shows the quadratic dependence of the current to the gate-source voltage, when the device operates in its above threshold region.

The second reason is the low-power consumption of transistors in their subthreshold regime, due to very low subthreshold currents in the order of nano to pico Ampères—see Fig. 5.1(c). Minimising power consumption is a main feature of neuromorphic circuits and it is crucial for fulfilling the ultimate goal of realising an artificial brain scale intelligent system with billions of electronic neurons and synapses. Due to these reasons many of the VLSI designs mentioned in Section 5.3, e.g. (Rachmuth *et al.* 2011, Azghadi *et al.* 2011c, Azghadi *et al.* 2012b, Azghadi *et al.* 2013a, Bofill-I-Petit and Murray 2004), exploit transistors in their subthreshold region of operation, in order to implement their desired neural dynamics and consume as little power as possible.

5.2.2 Differential Pair (DP) and Operational Transconductance Amplifier (OTA)

Differential pairs (DPs) are electronic components widely utilised in neural analog circuit design (Liu *et al.* 2002, Douglas *et al.* 1995). A DP in its basic form consists of three transistors, two of which are used for receiving the input voltages at their gates and the other one for biasing the pair by a constant current source—see Fig. 5.2(a). As shown in Fig. 5.2(c), a DP sets a sigmoidal relationship between differential input voltages and the currents flowing across each of the two differential transistors. The sigmoidal function is crucial to artificial neural networks and has been useful in describing the activities of populations of neurons (Wilson and Cowan 1972). This makes the differential pair an interesting and useful building block for neuromorphic engineers. Differential pairs can be used for various applications including spike integration for a synapse circuit (Bartolozzi and Indiveri 2007), and a rough voltage difference calculator (Mayr *et al.* 2010). They are also the heart of Operational Transconductance Amplifier (OTA)—see Fig. 5.2(b).

The OTA is another essential building block not only in neuromorphic engineering, but also in general analog integrated circuit design (Liu *et al.* 2002, Douglas *et al.* 1995, Razavi 2002). It is usually used to perform voltage mode computation and produces an output as a current. This analog component is commonly employed as a voltage-controlled linear conductor. However, in its simplest form the OTA is not really linear and usually sets a sigmoidal function between differential voltage inputs and the output current—see Fig. 5.2(c). In various VLSI implementations of neuromorphic synapses and synaptic plasticity rules, the OTA has been used in different roles (Cruz-Albrecht *et al.* 2012). In some cases, it has been used to act as an active resistor when forming a leaky integrator (Koickal *et al.* 2007, Mayr *et al.* 2010), and sometimes to act as a low-cost comparator (Mayr *et al.* 2010). In addition, a number of neuromorphic designers have carried out some changes to the basic structure of the OTA (Rachmuth and Poon 2008, Mayr *et al.* 2010, Cruz-Albrecht *et al.* 2012) to increase its symmetry, dynamic range and linearity and at the same time decrease the offset. As a result, the OTA has greater stability against noise and process variation, and gains better ability to mimic the desired neural function (Liu *et al.* 2002, Rachmuth and Poon 2008, Rachmuth *et al.* 2011, Mayr *et al.* 2010).

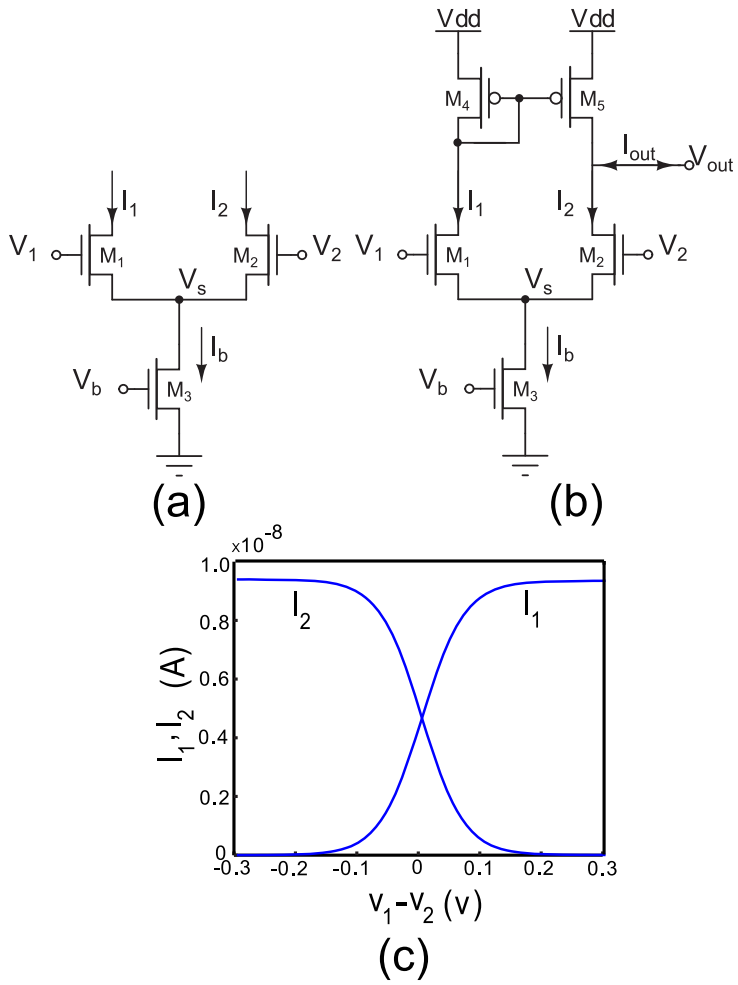


Figure 5.2. Differential Pair (DP) and Operational Transconductance Amplifier (OTA). (a) A basic Differential Pair (DP) circuit consists of three transistors. (b) The Operational Transconductance Amplifier (OTA) circuit converts the difference between its two input voltages to a corresponding current at its output. This circuit has been extensively used in the implementation of various neuromorphic devices (Liu *et al.* 2002, Rachmuth and Poon 2008, Rachmuth *et al.* 2011, Mayr *et al.* 2010). (c) The DP sets a sigmoidal relationship between differential input voltages and the currents flowing across each of the two differential transistors. This is a useful behaviour for implementing similar sigmoidal behaviour, observed in neural systems (Liu *et al.* 2002).

5.2.3 Synaptic Potential and Leaky Integrator (Decay) Circuits

When implementing a synaptic plasticity rule, there is always a need to implement some dynamics to represent potentials for potentiation and depression. These potentials start with the arrival of a spike, and can lead to potentiation/depression in the synaptic weight, if another spike arrives in the synapse before the potential vanishes. Fig. 5.3 shows a circuit that has been utilised to implement the required potentials in

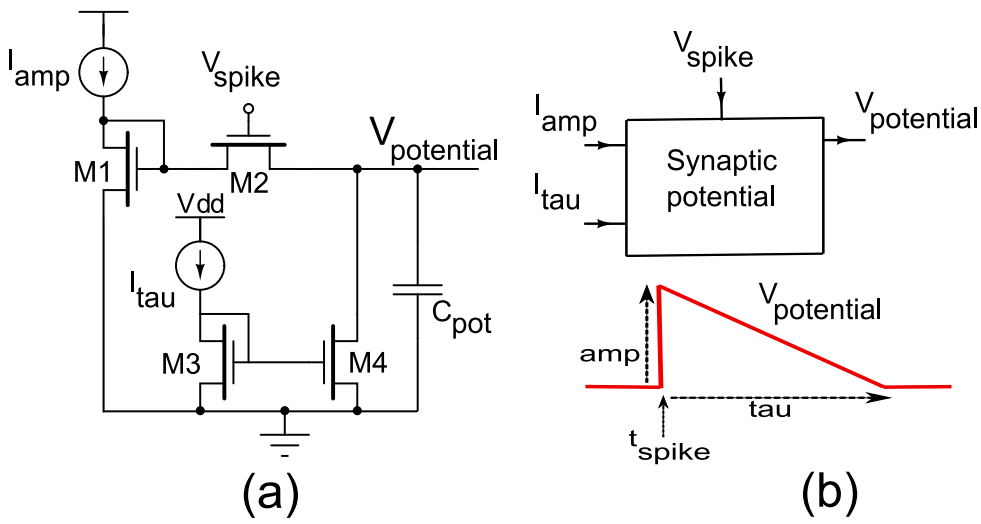


Figure 5.3. Synaptic potential (decay) circuit. (a) Synaptic potential (decay) circuit diagram. (b) Synaptic potential module. The output of this module is a decaying signal, which its time constant and amplitude are controlled by I_{tau} and I_{amp}, respectively. The decay starts once a pre/post spike arrives.

a number of PSTDP circuits including Bofill-I-Petit and Murray (2004), Azghadi *et al.* (2012b), and Azghadi *et al.* (2013a). It can also be utilised in implementing many synaptic plasticity circuit, where there is a need for controllable decay dynamics. This circuit, which acts as a leaky integrator controls both the amplitude of the generated potential signal as well as its time constant.

There is another instance of the leaky integrator, in which only the time constant is controllable and the required amplitude of the potentiation/depression may be realised with another circuit/transistor. Two different arrangements of this leaky integrator are shown in Fig. 5.4. In these circuits, the dawn of the signal determined by the arrival of a spike, and the time constant is controlled by the voltage applied (V_{tau}) to the gate of a PMOS/NMOS transistor.

In addition to the above mentioned important building blocks for analog neural designs, there are other essential circuits such as current mirrors, source-followers and current-mode Winner Take All (WTA) circuits (Liu *et al.* 2002, Razavi 2002), which are extensively used in neuromorphic analog designs including those reviewed in the next Section.

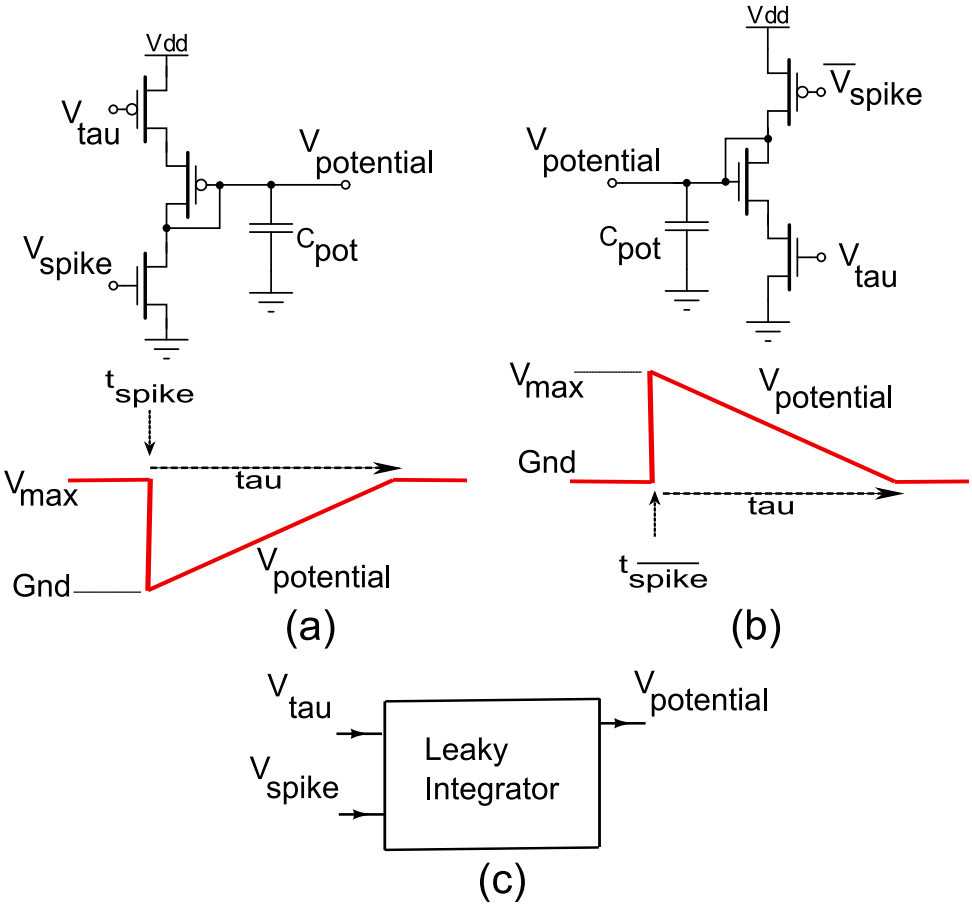


Figure 5.4. Leaky integrator circuit for producing required decay dynamics with adjustable time constants. (a) Leaky integrator for driving a PMOS transistor. (b) Leaky integrator for driving a NMOS transistor. (c) Leaky integrator module symbol.

5.3 Neuromorphic Implementation of Synaptic Plasticity Rules

The area of neuromorphic implementation of various synaptic plasticity rules has been active for over a decade and many neuromorphic research engineers have been involved in hardware realisation of various synaptic plasticity rules. Below is a review of a variety of approaches for implementing different synaptic plasticity rules.

5.3.1 Spike-based Learning Circuits

One of the first VLSI designs utilising spike-based learning mechanisms, which is closely related to many current designs for STDP, was proposed by Hafliger *et al.* (1997), even before a large body of biological evidence for STDP was published. The

learning part of the synapse circuit proposed in Hafliger *et al.* (1997) adjusts the synaptic weight stored on a capacitor, when a post-synaptic spike is received. The direction and magnitude of the adjustment i.e. charging or discharging the capacitor, which respectively corresponds to potentiation or depression of the synapse, is determined by the post-synaptic pulse width as well as pre-synaptic spiking dynamics compared to an adjustable threshold.

In 2000, another instance of a learning rule that functions based on both the occurrence of the pre-synaptic spike, and the membrane potential of the post-synaptic neuron was proposed by Fusi *et al.* (2000). This new learning rule is called Spike Driven Synaptic Plasticity (SDSP). In the SDSP rule as described in Section 2.6.1, the dynamics of the voltages produced in the neuron depends on the membrane potential, V_{mem}, of the neuron. So the SDSP rule changes the synaptic weight according to the time of pre-synaptic and membrane potential of the post-synaptic neuron. This membrane potential itself depends on the frequency of post-synaptic spikes generated by the neuron. Figure 5.5 shows a brief view of the neuron and synapse structure implemented in VLSI to realise the SDSP rule. Figure 5.5(a) shows schematic diagram of a VLSI learning neuron with an array of SDSP synapses. Multiple instances of synaptic circuits output currents into the I&F neuron's membrane capacitance (Indiveri *et al.* 2011). The I&F neuron integrates the weighted sum of the currents and produces sequences of spikes at the output.

Figure 5.5(b) shows that for implementing the SDSP synapse, a Differential Pair Integrator (DPI) (Bartolozzi and Indiveri 2007) along with a bistability circuit, are the main components, and the rest of the required components are only needed once per neuron. This figure shows the synapse with pre-synaptic weight update module. An AER asynchronous logic block receives input spikes and generates the `pre` and $\sim$ `pre` (the inverse of `pre`) pulses. An amplifier in a positive-feedback configuration, forms a bistability circuit that slowly drives the weight voltage V_{Wi} toward one of the two stable states V_{wlow} or V_{whi}. The transistors driven by the `pre` and $\sim$ `pre` pulses, together with those controlled by the V'_{UP} and V'_{DN} signals, implement the weight update. The DPI block represents a current-mode low-pass filter circuit that generates an output synaptic current I_{syn} with biologically plausible temporal dynamics. This current is then sourced into the V_{mem} node of the I&F circuit.

In addition, Fig. 5.5(c) demonstrates the neuron with post-synaptic weight control module. An I&F neuron circuit, integrates the input synaptic currents and produces a

spike train at the output. A DPI filter generates the V_{Ca} signal, encoding the neuron's mean firing rate. Voltage and current comparator circuits determine whether to update the synaptic weights of the afferent synapses, by increasing/decreasing their values.

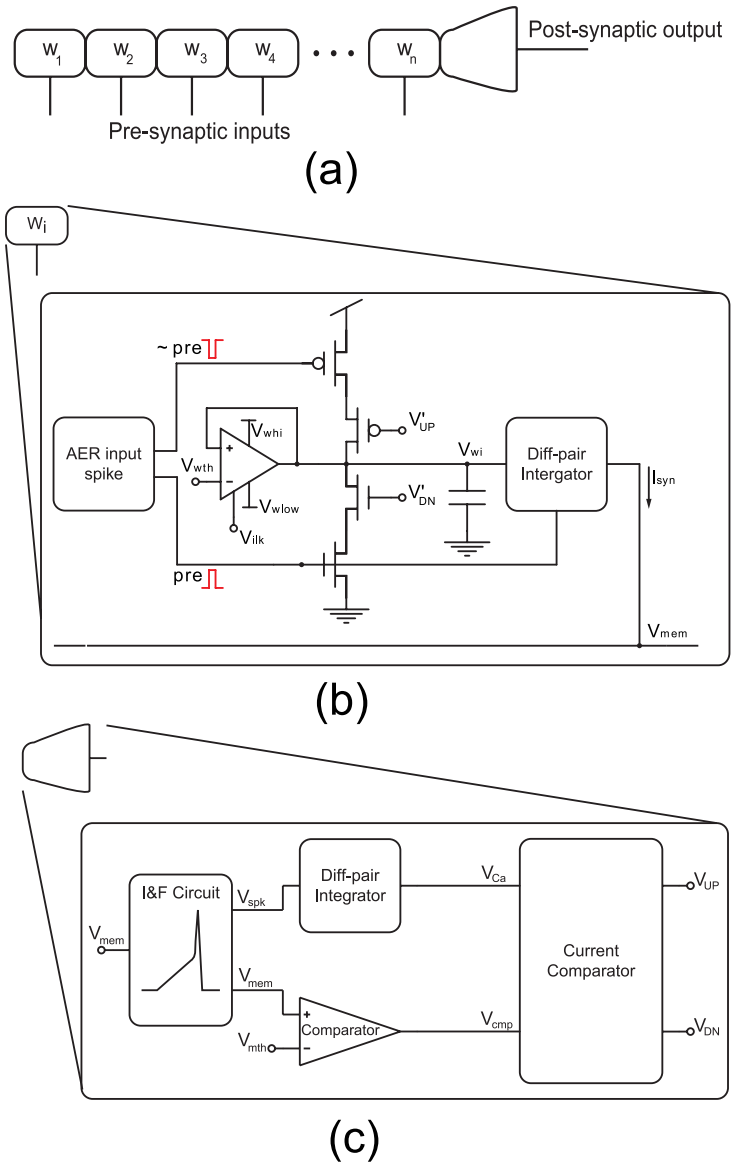


Figure 5.5. Implementation of the SDSP learning rule. (a) Schematic diagram of a VLSI learning neuron with an array of SDSP synapses. (b) Synapse with pre-synaptic weight update module. (c) Neuron with post-synaptic weight control module.

Spike-based learning circuits (Fusi *et al.* 2000, Mitra *et al.* 2009, Giulioni *et al.* 2009) alter the synaptic efficacy on pre-synaptic or post-synaptic spike arrival. They do not take into account the exact timing difference between pre-synaptic and post-synaptic spikes to induce synaptic weight changes. Therefore, spike-timing based learning circuits can be categorised as another type of synaptic plasticity learning circuits.

5.3.2 Spike Timing-Dependent Learning Circuits

Many spike-timing based circuits have been implemented by different groups and under various design strategies in VLSI (Bofill-I-Petit and Murray 2004, Cameron *et al.* 2005, Indiveri *et al.* 2006, Koickal *et al.* 2007, Tanaka *et al.* 2009, Azghadi *et al.* 2011c, Bamford *et al.* 2012b, Azghadi *et al.* 2012b, Azghadi *et al.* 2012a, Azghadi *et al.* 2013a). These circuits are classified into several design categories, and are reviewed in different subsections as follows.

Analog Subthreshold Circuits

One of the first designs for PSTDP, which is the conventional form of timing-dependent plasticity, was first proposed by Bofill *et al.* (2001). In this design a transistor that operates in its subthreshold (weak inversion) region is utilised to control the amount of current flowing into/out of the synaptic weight capacitor. The direction of the changes in the voltage of the weight capacitor is determined by another circuit that generates the required signals according to the timing differences of pre- and post-synaptic spikes. This circuit utilises the same peak voltage, time constants and decays for both potentiation and depression dynamics required in the PSTDP rule (shown in Eq. 2.2), however, these values might be required to be different in various contexts. As a result, Bofill-I-Petit and Murray (2004) presented a modified version of their original design and proposed a new circuit, where potentiation and depression dynamics have their own decay constants as well as peak values. In their design, they have employed additional circuitry to form the required pre- and post-synaptic pulses for their PSTDP circuit. By adding a few transistors to the main circuit required for potentiation and depression, they also made their circuit capable of weight dependent synaptic weight modification.

Fig. 5.6(a) shows a version of the STDP circuit presented in Bofill-I-Petit and Murray (2004). In this design, two transistors (Mp and Md) that operate in their subthreshold (weak inversion) region are utilised to control the amount of current flowing into/out of the synaptic weight capacitor, C_W. The voltages that control these transistors are V_{pot} (potential for potentiation) and V_{dep} (potential for depression). These potentials produced by two instances of the synaptic potential circuit presented in Fig. 5.3. This design uses currents for controlling circuit bias parameters that correspond to the PSTDP learning rule parameters presented in Eq. 2.2. Simulation results for generating STDP learning window using this circuit are also presented in Fig. 5.6(b). This figure demonstrates the exponential decay behaviour in the learning window, which is in

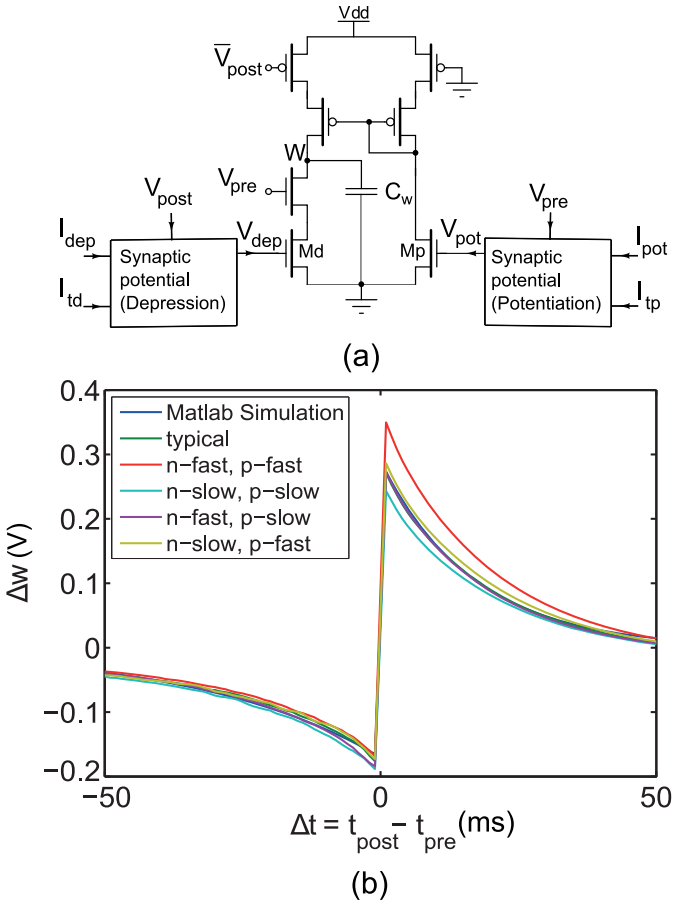


Figure 5.6. Pair-based STDP circuit with synaptic potential blocks. (a) This PSTDP rule circuit that is a modified version of the design proposed in Bofill-I-Petit and Murray (2004) is presented in Azghadi *et al.* (2012b). (b) This is the exponential learning window generated by Matlab and the PSTDP circuit under various process corners. Similar protocols and time constants to Bi and Poo (1998) are employed.

accordance with the PSTDP rule formula presented in Eq. 2.2. This exponential behaviour is reached by biasing Mp and Md, in their subthreshold regions of operation.

Another well-known PSTDP circuit is the symmetric design proposed by Indiveri *et al.* (2006). This circuit has two branches of transistors, as shown in Fig. 5.7(a). The upper branch is responsible for charging the weight capacitor, if a pre-synaptic spike precedes a post-synaptic one in a determined time, and the bottom branch is for discharging the capacitor if the reverse spike order occurs, also within a predetermined time. The potentiation and depression timings in this design are set by two leaky integrators, in which their decays are set by two bias voltages, V_{tp} and V_{td}, for potentiation and depression time constants, respectively. In addition, the amplitude of the potentiation

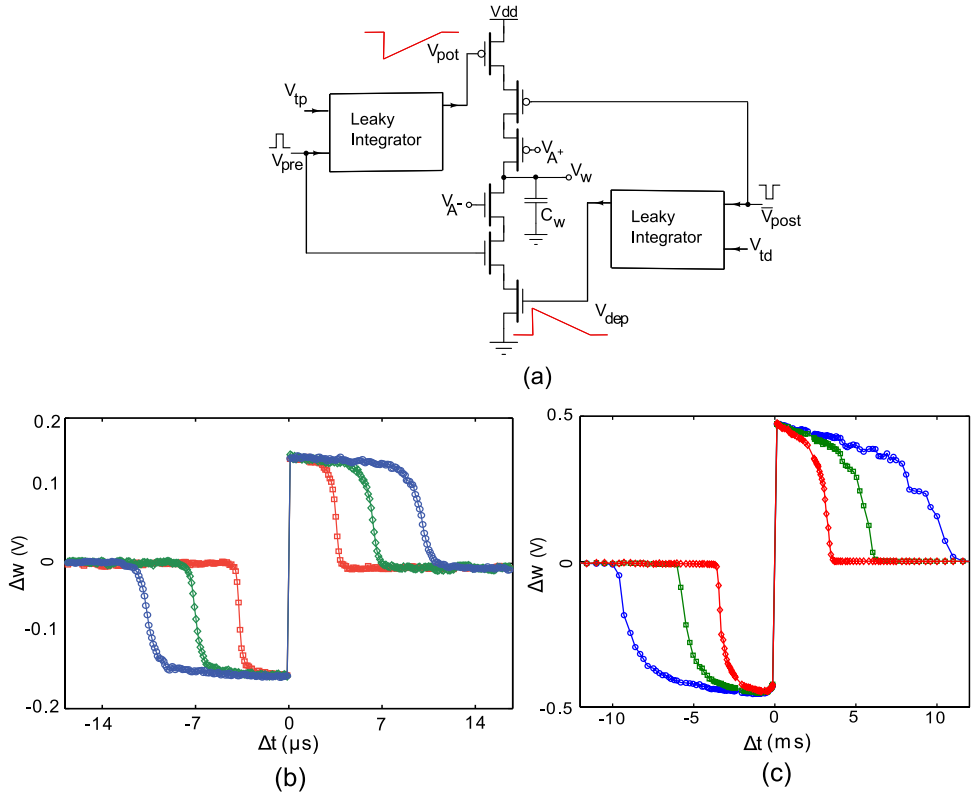


Figure 5.7. Pair-based STDP circuit with leaky integrator blocks. (a) This is a different representation of the PSTDP circuit presented in Indiveri *et al.* (2006). (b) The STDP learning window generated by the PSTDP circuit shown in (a) for various potentiation and depression time constants. A similar protocol to Bi and Poo (1998), which uses 60 pairs of pre- and post-synaptic spikes at a rate of 1 Hz, is employed. Note that simulations are carried out in accelerated time, by a factor of 1000, compared to real biological time. (c) A similar learning window was measured from the multi-neuron chip presented in Indiveri *et al.* (2006) in biologically plausible time and under the PSTDP experimental protocol utilised in Markram *et al.* (1997).

and depression are set by V_{A^+} and V_{A^-}, respectively, and the pulse width that is another important factor for the amplitude values, is kept fixed and equal to 1 μs in the shown simulation in Fig. 5.7(b), which shows the STDP learning window in an accelerated time scale. In addition, Fig. 5.7(c) shows chip measurement results for STDP learning window, in biologically plausible time (Indiveri *et al.* 2006, Azghadi *et al.* 2014c).

It is shown in a previous study that the circuit demonstrated in Fig. 5.7(a) can be minimised to decrease the number of transistors, and shrink the layout size for this circuit (Azghadi *et al.* 2011b). When considering the need for replicating this circuit in every synapse in a large-scale neural network, this little area saving can be significant.

The minimised circuit is shown in Fig. 5.8. Our simulation results show that this circuit is able to reproduce a window similar to that of figure 5.7(b), that is generated by the original PSTDP circuit shown in Fig. 5.7(a).

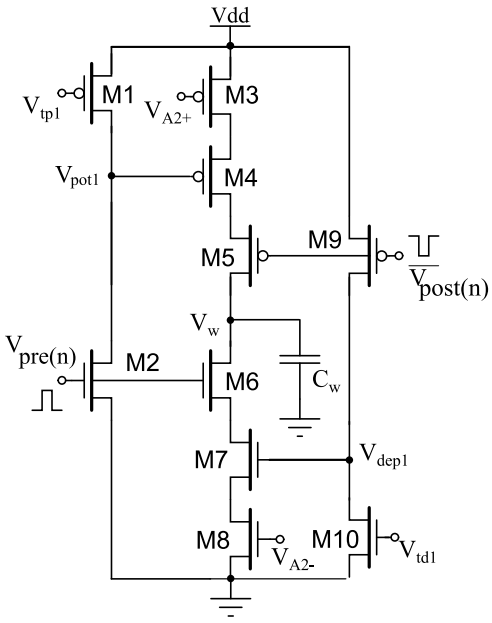


Figure 5.8. Minimised pair-based STDP circuit. This is the PSTDP circuit presented in Azghadi *et al.* (2011b).

One of the other STDP designs that has been utilised in a VLSI spiking neural network chip as part of the FACETS project is the design proposed by Schemmel *et al.* (2006). In this design the STDP circuit that is local to each synapse has a symmetric structure. The voltage potentials for potentiation or depression correspond to the quantity of charge stored on synaptic capacitors, which are discharged by a fixed rate, determined by a set of three diode connected transistors working in their subthreshold region of operation to cause an exponential decay. These capacitors later determine the amount of change in the synaptic weight corresponding to the time difference between the pre- and post-synaptic spikes.

Another PSTDP circuit that was reported in 2006, is the design proposed by Arthur and Boahen (2006). This symmetric analog design utilises a Static Random Access Memory (SRAM) cell for storing a binary state of the synapse weight, which is either high (potentiated) or low (depressed). This circuit uses leaky integrators in order to implement the required dynamic for the plasticity, similar to the designs proposed in Indiveri *et al.* (2006) and Azghadi *et al.* (2013a) and different from the design proposed

in Schemmel *et al.* (2006). Upon the arrival of a spike, the plasticity potentials are generated. They start decaying linearly thereafter, and if a complementary spike arrives in their decay intervals, its time difference with its complementary spike, determines the required level of potentiation or depression.

All aforementioned subthreshold STDP circuits, except the circuit reported in Bofill-I-Petit and Murray (2004), implement the simple conventional form of the STDP rule, known as additive STDP, in which the current synaptic weight has no effect on the synaptic plasticity mechanism (Song *et al.* 2000). Recently, a new analog VLSI design (Bamford *et al.* 2012b) for weight-dependent PSTDP (Kistler and Hemmen 2000, Guetig *et al.* 2003) was proposed. This design exploits the MOS transistor physical constraints and not any extra circuitry—as used in Bofill-I-Petit and Murray (2004)—to implement a weight-dependent PSTDP model (Bamford *et al.* 2012b). The design employs switched capacitors to implement the required leaky integrators needed for potentiation and depression.

OTA-based Circuits

In addition to the aforementioned analog STDP designs, there are other analog implementations of PSTDP in the literature that use the Operational Transconductance Amplifier (OTA) as their main building block, and have been employed in specific applications. The circuit proposed by Koickal *et al.* (2007), implements PSTDP in a symmetric fashion and by employing four instances of OTAs. Similar to the previously mentioned STDP designs, for generating the potentials for plasticity, there is a need for leaky integrators. In all STDP designs mentioned so far, these integrators are built using an RC network, where the resistor is implemented using the transistor's channel resistance, which is function of the gate to source voltage of the transistor. Here however, OTAs have been used as active resistors to generate the required long time constants for STDP, which are essential in their olfaction chip. The resistors are formed by having a feedback from the output of the OTA to its inverting input and a high resistance is reached by reducing the transconductance of the OTAs.

Another STDP circuit that utilises OTAs for implementing PSTDP, is the design proposed by Tanaka *et al.* (2009). They proposed two OTA-based circuits for implementing symmetric and asymmetric STDP. Tanaka's design uses the resistance of the transistor channels instead of OTAs for generating the required time constants needed by the STDP rule, however in a similar manner to Koickal *et al.* (2007) it still employs two

OTAs for charging and discharging the weight capacitor in case of LTP and LTD, respectively. Recently a low power design for synapses with STDP was proposed in Cruz-Albrecht *et al.* (2012). This design utilises OTAs and have transistors all biased in the deep subthreshold, while transistors have high threshold voltages, above the circuit power supply, to build low power neuron and synapses. The synapse with PSTDP weight update ability, is implemented using five OTAs from which three have enable/disable capability for disabling the synapse when no weight change is required (Cruz-Albrecht *et al.* 2012) and therefore having only leakage currents in the order of 100 fA, hence lower power consumption. The three OTAs also utilise a source degeneration technique to increase the linearity and dynamic range. The other two OTAs used in the synapse circuit are simple OTA circuits that consist of a differential pair and a set of current mirrors to generate the respective output current. The output current in both types of OTAs employed in this design are governed by the tail current of the differential pair, which itself is generated by a set of diode connected transistors stacked on top of each other to form the bias circuitry (Cruz-Albrecht *et al.* 2012).

Analog Circuits with Non-volatile Weight Storage Techniques

The Floating Gate (FG) technology is one of the popular approaches for implementing STDP-based learning mechanism that updates the charge on floating gates according to the STDP rule. By utilising FG, one can implement a spike-based learning rule such as STDP using only one transistor and some extra circuitry which can be shared for several Single Transistor Learning System (STLS) as it is named in this context (Gordon and Hasler 2002, Ramakrishnan *et al.* 2011). The other design approach with long-term weight storage capability, is the device-based design proposed by Zhang *et al.* (2010). They have proposed an ionic/Si hybrid device along with some peripheral circuitry to implement a Neural Phase Shifter (NPS), which represents the synaptic weight change as changes in the phase shift of their proposed NPS versus the time differences between pre- and post-synaptic spikes.

Digital Circuits

In addition to the above mentioned analog implementations of the STDP rule, this rule has been implemented several times using digital circuitries (Cassidy *et al.* 2007, Belhadj *et al.* 2009, Cassidy *et al.* 2011, Soleimani *et al.* 2012). Cassidy *et al.* first proposed

a FPGA implementation of a SNN that was successfully used in a few neural experiments including the demonstration of bimodal weight distribution behaviour (see Section 4.2.1 for a similar example on the IFMEM device) as a feature of synapses with STDP learning (Song *et al.* 2000, Cassidy *et al.* 2007). In 2009, neuromorphic researchers proposed three various approaches to implement STDP on FPGA and tested them in a SNN composed of analog neurons and STDP synapses that transfers their weight to the neurons through a pulse width modulation (PWM) technique (Belhadj *et al.* 2009). Cassidy and his colleagues again in 2011 proposed some other digital architectures for implementing STDP on FPGA that outperform the previous FPGA designs proposed in Belhadj *et al.* (2009), in terms of FPGA power and area consumption (Cassidy *et al.* 2011). Apart from these FPGA-based digital implementations of STDP, there are some general digital neuromorphic architectures, where a SNN can be easily configured with required learning strategy e.g. STDP. Researchers at IBM have reported implementation of a neurosynaptic core that is composed of digital Integrate and Fire (IF) neurons, addressable axons, and a crossbar arrays of SRAMs that are programmable and act as binary synapses. This design has been implemented in a 45 nm Silicon-On-Insulator (SOI) process and consumes little power (Merolla *et al.* 2011, Arthur *et al.* 2012).

In recent years, along side the IBM digital neuromorphic architectures another neuromorphic architecture inspired by the parallelism of the human brain has also been developed and tested (Furber *et al.* 2013, Painkras *et al.* 2013). This neuromorphic architecture, called Spinnaker, is a hardware platform that utilises off-the-shelf ARM processors to form a biomimetic massively parallel simulator for spiking neural networks.

Analog/Digital Circuits

Another popular approach to implement SNNs including required neural and synaptic dynamics, is the mixed signal VLSI design strategy that involves analog circuits along with digital ones that are integrated to implement the required network combination (Pfeil *et al.* 2012). For example, the neuromorphic architecture presented in Moradi and Indiveri (2014) follows mixed signal design strategies in which synapses are analog circuits (see Fig. 3.3), but with digital weights, stored on an asynchronous SRAM block (Moradi and Indiveri 2011, Moradi and Indiveri 2014), that can be updated off-chip and according to any expected plasticity rule, including timing- and

rate-based rules e.g. STDP and BCM (Azghadi *et al.* 2014b). Experimental results presented in Azghadi *et al.* (2013d) show how this programmable neuromorphic design can be tuned to exhibit biologically plausible response properties.

5.3.3 Hybrid Spike- Time and Rate Based Analog Circuit

In addition to the spike-based and spike-timing based synaptic plasticity circuits mentioned so far, there is another circuitry proposed by Mayr *et al.* (2010) that uses a hybrid learning rule composed of both timing and rate of spikes to alter the synaptic weight. This phenomenological rule was already introduced in Section 2.6.1. For generating the required exponential decay dynamics for both $u(t)$ and $g(t)$ (see Eq. 2.17), an OTA-based design approach has been utilised. Similar to the PSTDP design reported in Koickal *et al.* (2007), this design exploits a balanced OTA with negative feedback, which acts as a large resistor in the required leaky integrators. However, this design uses an active source degeneration topology to further improve the dynamic range and linearity of the integrator. These integrators are needed for both membrane potential, $u(t)$, which decays linearly back to the resting potential after a post-synaptic pulse duration is finished (hyperpolarization dynamic), as well as the post-synaptic current, $g(t)$, which decays exponentially toward zero after a pre-synaptic spike arrival. The time constants in these integrators can be tuned by changing the resistance of the OTA-based resistor that in turn can be altered by calibrating the leak bias current in the design. Beside these exponential decay dynamics, the rule needs subtraction and multiplication for changing the synaptic weight. These functions are approximated using a differential pair and its tail current (Mayr *et al.* 2010).

5.3.4 Neuromorphic Implementations of Biophysical Rules

Unlike all aforementioned neuromorphic VLSI designs of synaptic plasticity rules that are based on a rough approximation of the mechanisms that occur in the neuron and synapses, there are a few designs in the literature that go further and implement a detailed account of synaptic plasticity mechanism including the whole chemical and ionic interactions in the synaptic cleft. These designs are much more complex than the previous timing-, spike- or BCM-based designs, since they take into account detailed biophysical interactions when inducing synaptic weight changes.

The two major biophysical designs available in the literature, that are able to demonstrate both PSTDP and BCM-like behaviour are the designs proposed in Meng *et al.* (2011) and Rachmuth *et al.* (2011). These designs are based on two similar biophysical rules that were already discussed briefly in Section 2.6.2. The first design implements an elaborate biophysical synaptic plasticity model, which is based on the general biophysical processes taking place in the synapse (see Section 2.6.2). This implementation utilises current-mode design technique in order to implement the targeted biophysical rule that describes the detailed dynamic of the synaptic ion channels (Meng *et al.* 2011, Rachmuth and Poon 2008).

Recently the same group has published another iono-neuromorphic VLSI design which explores a similar approach for implementing both Spike Rate Dependent Plasticity (SRDP) and STDP using a unique biophysical synaptic plasticity model as briefly explained in Section 2.6.2. Identical to their first design, they used current-mode design technique to implement the required ion channel dynamics (Rachmuth *et al.* 2011).

5.4 Challenges in Neuromorphic Engineering

To realise artificial neural networks that can faithfully reproduce the properties of biological neural networks, and at the same time, be useful for implementing neural computation, there are different approaches and strategies, with their own advantages and limitations. Here, we focus on the challenges neuromorphic engineers face in the development of hardware artificial spiking neural networks whose foundation, architecture and structure are based on biological neural networks (Indiveri *et al.* 2009). Before investigating the obstacles and challenges in the way of implementing hardware neuromorphic systems, one might wonder why hardware is preferred over software for implementing Spiking Neural Networks? We discuss the response to this question first and then categorise the neuromorphic challenges.

In the spiking neural network software approach, the required neural architecture and its basic building blocks are implemented on conventional Von Neumann computing architectures. The main advantages of this method are, shorter design time compared to the physical implementation of neuromorphic systems, as well as the reconfigurability they offer compared to the non-reconfigurable design in most of physical design approaches. Furthermore, the input/outputs to software-based designs are technically noise-free and easy to deal with, compared to analog/digital hardware

outputs that are not so easy to use. On the other hand though, as Poon and Zhou (2011) state, the software-based approaches have significant shortcomings such as very high power consumption and very large real estate, which are major obstacles in the way of implementing a brain-scale neuromorphic system. An example of a software-based neuromorphic system implemented on digital machines, is the realisation of cat cortex on an IBM's super computer that took 147,456 CPUs and 144 TB of memory (Ananthanarayanan *et al.* 2009). Although this supercomputer is much faster than the brain in performing computation, since the computation takes place in a sequential order, compared to the massively parallel but slow computations that take place in the brain, it is still far behind the brain parallel processing capability. That is the situation, where a hardware-based approach becomes superior because of its parallel nature, similar to that of the brain.

In addition, as mentioned earlier in Section 5.3.2, another approach IBM is currently following is a software/hardware structure, in which the neural architecture is implemented in silicon as customised digital architectures and then it uses software-based off-chip synaptic plasticity rules to change the state of binary-valued synapses (Arthur *et al.* 2012). Besides, there are other neuro-computing architectures that use the software/hardware design approach and implement neurons and the neural architecture in mixed analog/digital VLSI, instead of pure digital VLSI. In these designs synaptic plasticity is performed off-chip and in software (Moradi and Indiveri 2011, Moradi and Indiveri 2014), while fast parallel silicon neurons process spikes. Furthermore, some other software/hardware designs, such as Spinnaker, simulate the entire large-scale neural system, on a special-purpose neural computer architecture (Painkras *et al.* 2013).

In contrast to software, or hybrid software/hardware systems, the neuromorphic system can be implemented mainly in hardware. The main advantage of a full-custom dedicated hardware neuromorphic system is to utilise a high degree of parallelism that allows the implemented neuromorphic system to work on biological time scales or even in accelerated time scales (Indiveri *et al.* 2011, Schemmel *et al.* 2010, Azghadi *et al.* 2013a). However, this design approach has its own limitations such as high-power consumption and large silicon area when considering a large scale neuromorphic system close to the scale of a brain. Another limitation that is inherent to analog VLSI circuits is fabrication imperfections that lead to device mismatch, which has minimal

affects on digital systems. Interestingly, the challenges and constraints faced by neuromorphic engineers for implementing synaptic learning, are similar to the ones encountered in biological learning, such as lack of long-time weight storage (Lisman 1985, O'Connor *et al.* 2005) and limited wiring. Generally, the main challenges and obstacles for implementing a large scale neuromorphic system can be summarised by the following:

5.4.1 Power Consumption

We know that the brain consists of billions of neurons and trillions of synapses, each of which consumes much less power compared to their silicon counterparts (Poon and Zhou 2011). Recently new integrated neural circuitry with a low power structure was proposed that consumes even less power per spike compared to a biological neuron (Cruz-Albrecht *et al.* 2012). Although it is a large step toward having a low power spiking neural system, it is very naive to think we are close to a neural system with a power consumption close to the brain, since this work does not consider the interconnection and communication among spikes and its required power. It also does not take into account the required complexity in the neural and synaptic structures, which is sometimes necessary for specific applications. In addition, the power consumption of a synapse or neuron heavily depends on its model parameters and their values that can change the weight modification pattern and at the same time lead to high or low power consumption. The other fact that should be considered is the spike pulse width utilised in the neuromorphic design, that can have significant affect on both functionality and power consumption of the system (Azghadi *et al.* 2014a). Therefore, an effective approach for decreasing the power consumption of a neuromorphic system is to optimise the neuron and synapse circuit bias parameters, as well as the structure of the spike pulses, in a way that while having the required functionality minimises the power consumption (Azghadi *et al.* 2014a).

One of the other effective ways for implementing low-power neuromorphic circuits is to design them using transistors operating deep in the subthreshold domain, which results in extremely low currents (e.g., in the order of pA). In addition, this approach allows the design of circuits that operate with low supply voltages, which is another strategy for reducing power consumption (Cruz-Albrecht *et al.* 2012). However, operating in the subthreshold region of operation and using lower supply voltages result in greater susceptibility to process variation and noise.

5.4.2 Process Variation and Device Mismatch

Due to inevitable variations in device parameters when fabricated in an integrated circuit technology, the resulting devices and circuits most likely will deviate in function and output when compared to simulation. Process variation shows itself more, when designing circuits using transistors biased in their subthreshold regime because of the possible variation in the transistor's threshold voltage and other parameters. Transistor mismatch is a challenge in designs including current mirrors, DPs, and those circuits that require an exact matching between several components. As stated earlier, due to the exponential behaviour and also low power consumption of transistors in their subthreshold regime, many of the spiking neural circuits including neural and synaptic weight change components, are implemented in this region. In addition, many neuromorphic VLSI designs employ current mirrors and DPs in their current- or voltage-mode structures. Therefore, these neural systems are highly susceptible to process variations and device mismatch (Azghadi *et al.* 2012b, Azghadi *et al.* 2013a, Mayr *et al.* 2010, Poon and Zhou 2011, Yu *et al.* 2013). There are various approaches to tackle the process variation and mismatch problems including the approaches mentioned in the following.

Post-Fabrication Calibration

Calibration of neural components that adjusts the circuit bias parameters, from those used in the design procedure, compensates for the variations due to the fabrication process. Although this approach can be used in small-scale neuromorphic circuits (Grassia *et al.* 2011, Azghadi *et al.* 2013a, Carlson *et al.* 2013, Carlson *et al.* 2014), and even shared biases might be useful for larger scale neuromorphic systems, for very large-scale neuromorphic systems with millions of neural circuits, this approach is impracticable (Poon and Zhou 2011). Furthermore, the calibration technique cannot account for detrimental effects of temperature and supply voltage variations, which are other sources of variations discussed later in this Section.

Utilisation of Circuit Design Techniques

The use of analog circuit design techniques that minimise the effect of process variations and device mismatch is another approach used for tackling fabrication imperfections (Poon and Zhou 2011, Rachmuth *et al.* 2011). This design approach has been successfully utilised in an iono-neuromorphic design presented in Rachmuth *et al.* (2011).

Considering large-scale neural circuity, the approach of mismatch minimisation using customised devices, seems more applicable compared to the post-fabrication calibration. This technique is also helpful for alleviating the temperature variation effects on the performance of the design. However, it results in circuits with higher complexity. This is the cost these circuits pay for robustness. Nonetheless, in the implementation of truly large-scale neural systems, the area overhead imposed by the mismatch minimisation techniques such as wide-dynamic range devices, is not significant compared to the silicon real estate occupied by the synaptic dynamic large capacitors and dense interconnects. In addition to these classical solutions for ameliorating the variations, some neuromorphic researchers have used different methods applicable to SNNs to alleviate the degradation in circuit performance due to process variation. One of these approaches that is suitable for SNNs is living with the variations (Neftci and Indiveri 2010), as discussed in the following.

System Level Approach for Mismatch Minimisation

Living with the variation and considering a mechanistic way to reduce its effect on the expected behaviour of the network is another instance of the approaches taken against variation. For example, Neftci and Indiveri have proposed an approach for compensating device mismatch by exploiting Address Event Representation (AER) in multi-neuron chips (Neftci and Indiveri 2010), which is freely available in the network and does not require adding extra circuitry or processing as mentioned for the first and second classical approaches. Another approach that exploits learning and adaptation in SNNs, is to utilise short term (Bill *et al.* 2010) or long term plasticity to alleviate process variation effects. For example, Cameron *et al.* exploit a PSTDP circuit to reduce the effect of process variation (Cameron *et al.* 2005). A similar approach that employs PSTDP in order to alleviate variations in performance of similar neurons—due to process variation and mismatch—in a SNN is also proposed in Bamford *et al.* (2012a). All previous approaches have used some techniques to reduce the effect of process variations, however, some neuromorphic engineers interestingly exploited variations in their neural circuit structure.

Exploiting the Process Variation and Device Mismatch

Sheik *et al.* have shown that the unwanted device mismatch and variations can be exploited in neuromorphic systems to implement axonal delays, which are required

and useful for neural computations (Sheik *et al.* 2012a). In another work, neuromorphic engineers exploited mismatch to perform vector-matrix multiplication in a extreme learning machine (Yao *et al.* 2013). Despite all these methods, one may keep in mind that using the process variation and mismatch such as the idea presented in Sheik *et al.* (2012a), or trying to reduce the effect of variations and mismatch is challenging due to mismatch inherent stochasticity.

All the mentioned approaches so far are only proposed to tackle the process variation and device mismatch. Although this is the most significant source of variation in analog neuromorphic models of synaptic plasticity rules, other types of variations, i.e. supply voltage and temperature variations must also be considered, especially when a large-scale neuromorphic system is targeted. The following subsection discusses the effect of variations on the design of a large-scale neuromorphic system in more details.

5.4.3 Voltage and Temperature (VT) Variation

Note that VLSI chips are susceptible to variations in many parameters and behave differently under various cases of variations. These include (i) fabrication process parameter variation such as deviation in the threshold voltage and channel length from the exact value specified for the fabrication process (this issue was discussed in details in the previous subsection); (ii) supply voltage variations from the ideal supply voltage required for the device operation; and (iii) temperature variations from the ideal temperature needed for the normal operation of the devices and interconnects. Each of these variation sources has its own corners, beyond those the device might not function properly. An instance of these corners are shown for process parameters in Fig. 5.6(b), where the STDP learning window is produced for various device corners, and fortunately for all of them it is close to the required behaviour. Since there are three simultaneous sources (i.e. PVT) of variations in an analog VLSI system, these variations should be coupled together in order to form various PVT variation corners, in which the device has its best, typical or worst characteristic. Apart from devices, variation also affects the characteristics of the interconnects, that have their own corners, which are usually different from those of the device. More directly, the device and interconnects potentially have worst performance at different corners. Considering these corners when designing the targeted application is essential, as the design might be dominated by device corners, interconnect corners, or a mixture of both (Weste and Harris 2005).

5.4.4 Silicon Real Estate

Silicon area consumption of a neuromorphic systems is related to the area used by each neuron and synapse, and to the way they are connected together. Considerations on the area required by the interconnects are listed in the next subsection. Concerning the area required by the silicon neuron and synapse designs, there are two main approaches to consider. One is the biophysically realistic approach that attempts to model in great detail the biophysics of neurons and synapses, usually producing large circuits, such as the design proposed in Meng *et al.* (2011) and Rachmuth *et al.* (2011). The other approach, that aims to implement the phenomenology of the action potential generation mechanisms but sacrificing biological fidelity, usually produces more compact circuits (Indiveri *et al.* 2011). Perhaps the most critical component however is the synapse design, as in learning architectures, most of the silicon real estate is going to be consumed by these elements. If the synapses have all the same type of (linear) temporal dynamics, it is possible to exploit the superposition principle and use one single shared (linear) temporal filter circuit to model the temporal dynamics of many synapses (Mitra *et al.* 2009). The individual synapse elements are therefore "only" required to implement the weight update mechanism and transmit a weighted pulse to the shared integrator. Naturally, the smaller the weight-update circuit, the larger the number of synapses can be integrated in the same area. There are very promising emerging technologies, e.g. based on 3D VLSI integration (Lee *et al.* 2010), 3D VLSI packaging (Al-Sarawi *et al.* 1998), and Resistive RAMS (Eshraghian 2010), which may offer ways of making extremely compact synapse elements and (consequently) extremely dense synaptic arrays (Likharev and Strukov 2005, Serrano-Gotarredona *et al.* 2013).

5.4.5 Interconnection and Routing

As in biology, wiring is a significant issue also in neuromorphic engineering. If each neuron in neuromorphic architectures were to use a dedicated wire to target its destination synapses (as real neurons use axons), then the area required by interconnects would dominate, and (given the essentially 2D nature of VLSI technology) it would be impossible to design large-scale systems. Fortunately, it is possible to exploit the very large differences in time scales between typical neuron transmission times and silicon communication circuits. In this way it is possible to time-multiplex the action

potentials generated by the silicon neurons and share wires, creating therefore "virtual axons". The most common protocol that is used in the neuromorphic community to accomplish this is based on the Address Event Representation (AER) (Deiss *et al.* 1994, Boahen 2000). In this representation, the action potentials generated by a particular neuron are transformed into a digital address that identifies the source neuron, and broadcast asynchronously on a common data bus. By using asynchronous arbiters and routing circuits (Moradi *et al.* 2013) it is therefore possible to create large-scale neural networks with reconfigurable network topologies. These networks can be distributed within the same chip e.g. among multiple neural cores (Arthur *et al.* 2012), or across multiple chips (Merolla *et al.* 2013, Chicca *et al.* 2007).

5.4.6 Electronic Design Automation for Large-Scale Neuromorphic Systems

Although there are a number of neuromorphic systems that deal with a relatively high number of analog neurons, designing large-scale neuromorphic systems is still a very complex task. One of the major obstacles on the way is the lack of an Electronic Design Automation (EDA) tool, that can facilitate the design procedure, while taking into account the targeted design requirement. There are promising recent accomplishments that exploit existing EDA tool-chains for automating the design of neuromorphic circuits. For examples see Imam *et al.* (2012) and Mostafa *et al.* (2013) for designing the asynchronous logic circuits that make up the arbiters and routers described above. However there is a need for a new generation of EDA tools that are optimised for neuromorphic architectures with hybrid analog/digital circuits, asynchronous logic circuits, and networks characterised by very large fan-in and fan-out topologies.

5.4.7 Bias Generation for Neuromorphic Circuits

The complex behaviour of neural circuitries including neurons and synapses are controlled by many parameters including synapse potentiation and depression time constants and amplitudes, neuron spiking thresholds, spiking frequency adaptation, and refractory period parameters. For controlling silicon neurons and synapses, these parameters should be presented as small-scale and high accuracy voltages and currents to silicon neurons and synapses. Generating these bias voltages and currents, which usually span over a wide range, usually needs a specific dedicated VLSI circuit that

generates these values in a programmable and reconfigurable manner. Fortunately, there are a number of high resolution, wide-dynamic range, temperature compensated analog programmable bias generator circuitries already available in the literature, which can be used for synaptic plasticity circuits and systems (Delbrück and Van Schaik 2005, Delbrück *et al.* 2010). Considering large-scale neuromorphic systems with large number of neurons and synapses, a bias sharing technique for neurons and synapses that are laid out closely seems a practicable approach as it has been used in Stanford University Neurogrid chips (Gao *et al.* 2012).

The challenges mentioned in the previous subsections are engaged with typical neuromorphic systems and are not specific to synaptic plasticity circuits. However, a specific challenge on the way of implementing required synaptic plasticity rules and integrating them into network of neurons, is the synaptic weight storage method, which is discussed in more details in the following subsection.

5.4.8 Synaptic Weight Storage and Stabilisation

Synaptic weight storage is another big challenge neuromorphic engineers encounter in the process of synaptic plasticity circuit implementation. When implementing the circuit in VLSI, the synaptic weight is usually represented as the amount of charge stored across a weight capacitor (Azghadi *et al.* 2013a). However, this weight is not stable as the charge on the capacitor will leak away and therefore the learnt synaptic weight will be lost. This instability is due to the leakage (i.e. off-current) of the transistor that is in the order of fA. Therefore, the synaptic weight cannot be preserved longer than hundreds of milliseconds to a few seconds, depending on the capacity of the weight capacitor.

As there is a direct relationship between the stability of the voltage stored on a capacitor and its capacity, some previous STDP designs have used bulky capacitors (Bofill-I-Petit and Murray 2004, Bamford *et al.* 2012b, Cruz-Albrecht *et al.* 2012, Azghadi *et al.* 2013a), which takes up a large portion of the precious silicon real estate, in order to preserve the weight value at least for the period of time required for their desired experiments. This does not appear compatible with the goal of neuromorphic engineering, which ultimately aims to integrate a large scale neural system with billions of synaptic circuits. Therefore a number of other approaches have been sought, in order

to address this obstacle on the way of realising long-term plasticity in silicon. These approaches are briefly reviewed as follows.

Accelerated-time Synaptic Plasticity Circuits

A number of neuromorphic designers have used a time-scaling approach, in order to tackle the mentioned problem of weight storage (Schemmel *et al.* 2006, Mayr *et al.* 2010, Azghadi *et al.* 2013a). This approach utilises circuits that operate in accelerated time scales compared to the timing of real neurons, therefore, these circuits operate orders of magnitude (10^3 to 10^5) faster than biological neurons and synapses. An instance of an accelerated neuromorphic system is the BrainScaleS wafer-scale system (Schemmel *et al.* 2010). The main advantages of this approach are, (i) increased speed of emulating large-scale neuromorphic systems that is useful for long experiments, and (ii) a higher degree of density and integration, due to smaller capacitors. On the other hand, main disadvantages of this approach are (i) the inability of the accelerated-time system to be directly interfaced to biological sensors and systems (Hamilton *et al.* 2008, Lichtsteiner *et al.* 2008, Hamilton *et al.* 2009) with biological time-scales, and (ii) the high degree of communications in the network, which requires high-performance control systems.

Utilising Reverse-biased Transistors to Decrease Leakage

Since the main reason for instability of the voltage stored on a weight capacitor is the leakage current, one may increase the stability by reducing the leakage current. Using reverse-biased transistors in the path of charging/discharging weight capacitors, reduces leakage currents and therefore increases the weight stability on that capacitor. This approach was first proposed by Linares-Barranco and Serrano-Gotarredona (2003). Recently, it has been successfully exploited in Bamford *et al.* (2012b) for storing the weight for a longer period of time in the order of hundreds of milliseconds. In order to reverse bias the transistors in the design, as it is proposed in Bamford *et al.* (2012b), the Gnd and Vdd are shifted few hundreds of millivolts toward Vdd and Gnd, respectively. By reducing the supply voltage slightly or increasing the ground voltage level, the transistor back gate will be in both cases at higher voltages, resulting in an increase in the threshold voltage that leads to reduced leakage current.

Bistability Mechanism

Another approach for synaptic weight stabilisation, which has been used in a number of synaptic plasticity circuits and for various learning rules, is a bistability mechanism that is based on the idea of having the long-term state of a synapse either potentiated or depressed. As shown in Fig. 5.5(b), in this approach, an amplifier with positive feedback is utilised to drive the synaptic weight stored on the weight capacitor and updated by the desired synaptic plasticity rule in the short-term, slowly either upward or downward depending on the current value of the synaptic weight that is above or below a predetermined threshold (Chicca *et al.* 2003, Indiveri *et al.* 2006, Mitra *et al.* 2009). Furthermore, some other neuromorphic circuits use the same approach but with storing the weight on a SRAM i.e. only two stable states (Arthur and Boahen 2006), or mapping and storing the modified weight on a multi-stage analog memory (Hafliger and Kolle Riis 2003, Hafliger 2007). Using this method, the synaptic weight stored on the capacitor is updated whenever there are spikes, but as soon as there is no spike, the weight is driven toward a high or low analog value, depending on the current synaptic weight on the capacitor, which can be either potentiated, above a certain threshold, or depressed, below that threshold, respectively.

Bistability mechanism has experimental support, as well as benefits over the use of large weight capacitors, in large neuromorphic systems. Considering a large network of neurons and synapses, on long time scales, synaptic efficacy can be assumed to have only two high (potentiated) or low (depressed) values. This assumption is compatible with experimental data (Bliss and Collingridge 1993, Petersen *et al.* 1998). In addition, from a theoretical perspective, it has been argued that the performance of associative networks is not necessarily degraded if the dynamic range of the synaptic efficacy is restricted even into two stable states (Amit and Fusi 1994). Furthermore, bistable synapses can be implemented in a small area compared to having large-scale capacitors for preserving the synaptic weights for longer periods of time (Indiveri *et al.* 2006). Due to these benefits, this technique is a suitable approach to be used in all of our reviewed synaptic plasticity circuits including the STDP and TSTDP circuits.

Despite the usefulness of the bistability mechanism for short term learning and weight changes, for permanent storage of synaptic weights, which are quantised to two high-/low states using the bistable technique, there is a need for a non-volatile storage technique. A number of these storage techniques have been discussed in the following.

Digitising the Synaptic Weight and Storing it in Memory

This approach has been followed in a few ways. In one of the pioneering works on neural networks presented in 1989, the approach was to serially and periodically refresh the analog weights stored on the capacitor with the weight stored in the memory (Eberhardt *et al.* 1989). This approach however needs bulky Digital-to-Analog Converters (DACs) and Analog-to-Digital Converters (ADC). In addition, designing these devices in the subthreshold low-current regime is a crucial task, as there is a high demand for low power consumption and small silicon area in neuromorphic applications (Poon and Zhou 2011). Moradi and Indiveri have used a single current-mode DAC, available beside each synapse integrated circuit, in order to convert 5-bit digitally stored synaptic weights in asynchronous SRAMs, to a current that drives the synapse integrator (Moradi and Indiveri 2011, Moradi and Indiveri 2014). Therefore, the synaptic weights here are considered as virtual synapses and their weights come into effect whenever they receive a spike from the AER system (Moradi and Indiveri 2011, Moradi and Indiveri 2014). This approach utilises a time multiplexing technique and therefore only uses one DAC per several synapse memory cell. Whilst this saves silicon area, it causes extra communication. A similar approach of using virtual synapses with digitised synaptic weights, has been employed by other neuromorphic engineers, in order to tackle both synaptic weight storage and also reduce area usage (Vogelstein *et al.* 2007b). In Pfeil *et al.* (2012), the authors discuss the issue of digitising weight on the PSTDP rule and show that considering other constraints of neuromorphic designs, increasing the weight storage resolution is not necessarily useful for PSTDP.

Floating Gate (FG)

Floating Gate technology (FG) is another possible approach for nonvolatile storage of synaptic weights. It has been exploited extensively in neuromorphic systems to implement Hebbian-based and STDP learning rules (Ramakrishnan *et al.* 2011, Gordon and Hasler 2002, Holler *et al.* 1989). As already mentioned in Section 5.3.2, this storage technique leads to a compact single transistor implementation of STDP, which saves significant silicon area. However, the drawback for this approach is the severe mismatch that occurs in the tunnelling and/or injection processes. Also the requirement for a special purpose CMOS process and extra control circuitry are other drawbacks of this approach.

Memristor

The memristor as the fourth circuit element (Chua 2011) possesses invaluable characteristics including non-volatility, low power, and high density (Strukov *et al.* 2008, Fortuna *et al.* 2009, Eshraghian 2010, Eshraghian *et al.* 2012) which are the features have always being sought for implementing large scale neuromorphic systems. Therefore, memristors may be a possible solution for solving the problem of synaptic weight storage (Jo *et al.* 2010, Zamarreño-Ramos *et al.* 2011, Pershin and Di Ventra 2012, Wang *et al.* 2014b, Sheri *et al.* 2014). It can also be integrated with CMOS (Eshraghian *et al.* 2011) in order to form a non-volatile synapse circuit (Jo *et al.* 2010, Indiveri *et al.* 2013). These hybrid CMOS/memristor synapse circuits then can be utilised to implement both computational and detailed biophysical synaptic plasticity learning rules that are quite useful for neural computation (Afifi *et al.* 2009, Ebong and Mazumder 2012, Azghadi *et al.* 2013d). Although the memristor has significant strengths, its weaknesses such as need for read operation, the accuracy of device programming, the device yield, and its variation and non-linearity should be considered as well.

5.5 Discussion

In this chapter, after a brief review of some basic circuit building blocks useful for synaptic plasticity models, the design and implementation of those models in VLSI were discussed under various design strategies. Then, the main challenges neuromorphic engineers encounter when designing neuromorphic systems were presented. Furthermore, several methods to address those challenges were discussed. In this Section, first the goals for VLSI implementation of various synaptic plasticity rules are discussed, and then the performance of the reviewed VLSI designs in reaching these goals as well as an efficient implementation are reviewed. Understanding these goals and challenges, as well as the effective and useful strategies for designing various synaptic plasticity rules is essential while new designs for other synaptic plasticity models are proposed and discussed in the remainder of this thesis.

Shouval has argued in a recent article *What is the appropriate description level for synaptic plasticity?* (Shouval 2011). He mentioned that the suitable and required level of description when considering synaptic plasticity is not yet known, but it essentially depends on what we attempt to reach. He stated that in some sense a simple plasticity rule such

as TSTDP theory of synaptic plasticity might provide a sufficient level of plasticity description, if only reproducing a set of experiments, including higher order phenomena, is needed. However, he also indicated that further research is needed to understand what is the required level of synaptic plasticity description and how detailed the biophysical synaptic plasticity models ought to be for various applications.

The same argument holds for the application domain of synaptic plasticity rules, specially those that are implemented in silicon. Before designing any synaptic plasticity rule in VLSI, one should first consider, what are the goals, application area, and needs for learning. Then, they should find the required level of synaptic plasticity description, ideal for their applications and proportionate to their other design requirements, such as area and power consumption. At this step the designers must set a trade-off between various design aspects, such as the required complexity of the network and synaptic plasticity mechanism, the needed level of weight storage precision, and the limitations in power consumption and silicon area. It is at this stage that the designer selects an appropriate synaptic plasticity rule from the continuum of rules, depending on the design specification and required synaptic details. Some therefore might choose simple phenomenological rules such as STDP (Azghadi *et al.* 2013a), while other opt for more detailed biophysical rules (Rachmuth *et al.* 2011).

Many of the mentioned designs in Section 5.3, have the simple research goal to implement, verify and understand desired synaptic plasticity rules such as pair-based STDP (Indiveri *et al.* 2006), triplet-based STDP (Azghadi *et al.* 2013a), BCM (Azghadi *et al.* 2012a, Azghadi *et al.* 2013a), BCM-like LCP (Mayr *et al.* 2010), and biophysical ion channel-based plasticity models (Rachmuth *et al.* 2011, Meng *et al.* 2011). In order to understand and verify a synaptic plasticity rule, reproducing biological experiments and replicating the experimental data using various models (VLSI circuits) is quite helpful. For instance, many of the VLSI implementation of STDP rules (Bofill-I-Petit and Murray 2004, Indiveri *et al.* 2006, Schemmel *et al.* 2006, Tanaka *et al.* 2009, Koickal *et al.* 2007, Bamford *et al.* 2012b, Azghadi *et al.* 2013a) have shown to be able to reproduce the STDP learning window, which represents both potentiation and depression as they occur in biology (Bi and Poo 1998, Markram *et al.* 1997). However, the way these circuits reproduce the window and other experimental results and the performance of the circuits are different.

If a plasticity circuit can show a close match to experimental data, therefore it can be of great help for neuroscience studies. Typically, the VLSI designs for conventional

PSTDP rule are able to reproduce the STDP learning window (Bofill-I-Petit and Murray 2004, Indiveri *et al.* 2006, Schemmel *et al.* 2006, Tanaka *et al.* 2009, Koickal *et al.* 2007, Bamford *et al.* 2012b). They are also able to show BCM-like characteristics (Azghadi *et al.* 2012a), similar to the BCM behaviours that were reproduced using computational STDP models (Pfister and Gerstner 2006, Izhikevich 2003). However, the PSTDP rule circuits are not able to reproduce more complicated experimental data. Therefore, modified and extended version of these circuits, which are based on updated STDP computational models such as Triplet-based STDP (TSTDP) learning model (Pfister and Gerstner 2006), will be needed.

Besides these synaptic plasticity models and VLSI designs, sometimes it is desirable to replicate the dynamics of synapses in detail, and study the effect of neuromodulators and intracellular signalling. In these cases, all the ion channels and synapse dynamics should be taken into account in the targeted VLSI implementation. These detailed implementations are potentially useful also for various neuroprosthesis, neural-based control, and machine learning tasks (Rachmuth *et al.* 2011).

Apart from the level of complexity and the required level of synaptic details, there are some other factors that affect the complexity, hence the area and power consumption of neuromorphic synaptic plasticity rules in VLSI. One of these factors is the weight storage techniques (see Section 5.4.8) and the required level of precision when storing the weight. This approach is also related to the targeted application and the goal of the implemented VLSI synaptic plasticity rule (Pfeil *et al.* 2012). In general, designers should always consider the trade-off between the level of synaptic weight precision they require for their target applications and the resources they can afford. For instance, neuromorphic engineers have shown that for synchrony detection application in spiking neural networks, a 4-bit synaptic weight precision is viable (Pfeil *et al.* 2012). On the other hand, a number of other studies show that if a suitable learning mechanism is used, then even binary weight precision controlled by the bistability mechanism is sufficient for the task of classifying complex rate-based patterns (Mitra *et al.* 2009).

Table 5.1 summarises the key properties of some neuromorphic systems for learning and synaptic plasticity applications. Note that the estimated area and power consumption data in this Table only reflect the reported data in the related papers. These numbers are dependent on many parameters including the synaptic plasticity rule implemented, the synaptic plasticity parameters, the weight storage techniques, and the network stimulation patterns and protocols. Since in some papers, the exact power

consumption and area requirement of the synapse is not available, the total power and area of the chip are divided by the number of synapses and neurons on the chip, to calculate a rough value of the size and power requirement of the synapse. Also, note that the calculated estimated area for each synapse, encompasses both the synapse circuit as well as its synaptic plasticity circuit, which may be reported or implemented separately in the related papers.

The feedforward network with STDP learning presented in Bofill-I-Petit and Murray (2004) successfully implements the targeted synchrony detection, but it consumes significant power and occupies a large area. The high power consumption is due to power hungry biasing current distribution network designed to minimise mismatch between synapses. In addition, the area of the designed STDP circuit is significantly large due to huge capacitors of the order of several pFs.

The implementation of an array of neurons with bistable STDP synapses (Indiveri *et al.* 2006), is the next design that has better power and size performance compared to the first mentioned design (Bofill-I-Petit and Murray 2004). Furthermore, this neuromorphic system utilises the AER communication protocol and therefore is reconfigurable, in contrary to the hard-wired network structure presented in Bofill-I-Petit and Murray (2004). The next two neural networks with STDP synapses, mentioned in Table 5.1, are also configurable. This feature helps to customise the neural network topology where there is a need for various studies and applications, such as the designs Indiveri *et al.* (2006) and Schemmel *et al.* (2006), which have been used to show STDP learning window, LTP and LTD characteristics. In terms of silicon real estate required for the STDP circuit, the design in Schemmel *et al.* (2006) has a compact structure that occupies an area of 50 μm^2 for the STDP circuit and 100 μm^2 for the synapse including STDP, DAC and memory cell for storing the synaptic weight. Power consumption information for this FACETS accelerated-time neuromorphic architecture is not listed. The neuromorphic learning chip presented in Arthur and Boahen (2006) also uses STDP and on-chip SRAM cells to store a binary state of the synaptic weight updated by the STDP circuit. Considering the number of neurons and synapses in this architecture and the overall area of the chip presented in Arthur and Boahen (2006), which is 10 mm^2, this design that has been used for learning patterns, also has a compact synapse size, on par with the FACETS project chip (Schemmel *et al.* 2006).

The next design reviewed in Table 5.1 is an adaptive olfactory neural network with on-chip STDP learning (Koickal *et al.* 2007). There is no power consumption information

Table 5.1. Synaptic plasticity circuit comparison.

Learning network (Neuron, Synapse)	Estimated power (Supply)	Estimated area (Design Tech.)	Weight storage (Precision)	Plasticity	Application
Feedforward (Bofill-I-Petit and Murray 2004) (5, 40)	0.34 mW (5 V)	19902 μm^2 (0.6 μm)	Capacitor (Analog)†	STDP	Synchrony detection
(Re)configurable (Indiveri *et al.* 2006) (32, 256)	Not available	4495 μm^2 (0.8 μm)	Capacitor (Bistable)‡	STDP	General purpose
(Re)configurable (Schemmel *et al.* 2006)** (384, 98304)	Not available	100 μm^2 (0.18 μm)	SRAM (4 bits)	STDP	The FACETS project
(Re)configurable (Arthur and Boahen 2006) (1024, 21504)	Not available	440 μm^2 (0.25 μm)	SRAM (1 bit)	STDP	Learning patterns
Adaptive olfactory (Koickal *et al.* 2007) (9, 81)	Not available	72000 μm^2 (0.6 μm)	Capacitor (Analog)†	STDP	Odor classification
Hopfield Feedback (Tanaka *et al.* 2009) (5, 10)	250 μW (3.3 V)	5000 μm^2 (0.25 μm)	Capacitor (Analog)†	STDP	Associative memory
Single Neuron-Synapse (Cruz-Albrecht *et al.* 2012) (1, 1)	(@100 Hz) 37 pW (0.6 V)	4823 μm^2 (90 nm)	Capacitor (Analog)†	STDP	Test design
Recurrent/Hopfield (Seo *et al.* 2011) (256, 64k)	(UVT*) 8 nW (0.53 V)	(UVT*) 13 μm^2 (45 nm)	SRAM (1 bit)	STDP	Various cognitive tasks
Weight-dependent STDP (Ramakrishnan *et al.* 2011) (-, 20k)	Not available	100 μm^2 (0.35 μm)	Floating gate (Analog)	Weight-dependent STDP	STDP test design
Weight-dependent STDP (Bamford *et al.* 2012b) (32, 2k)	(@1 KHz) 60 pW (3.3 V)	400 μm^2 (0.35 μm)	Capacitor (Analog)†	Weight-dependent STDP	STDP test design
(Re)configurable (Mitra *et al.* 2009) (16, 2048)	Not available	3000 μm^2 (0.35 μm)	Capacitor (Bistable)‡	SDSP (Brader *et al.* 2007)	Pattern classification
(Re)configurable (Mayr *et al.* 2010, Mayr *et al.* 2013)** (16, 512)	(@1 MHz) 11 μW (3.3 V)	700 μm^2 (0.18 μm)	Memory (4 bits)	LCP (Mayr and Partzsch 2010)	Beyond STDP experiments
Single Synapse (Meng *et al.* 2011) (-, 1)	500 nW (1.2 V)	4 mm^2 (0.15 μm)	Not available	Iono-Neuromorphic (Meng *et al.* 2011)	STDP and BCM experiments
Single Synapse (Rachmuth *et al.* 2011) (2, 2)	50 nW (5 V)	8 mm^2 (1.5 μm)	Digital bistable	Iono-Neuromorphic (Rachmuth *et al.* 2011)	STDP and BCM experiments

* UVT = Ultra high Voltage Threshold

** The chip operates in an accelerated time of 10^4 compared to the real biological time.

† The synaptic weight value will decay if not refreshed. Nonvolatile weight storage techniques should be considered to retain the updated weights for later use (see Section 5.4.8).

‡ See Section 5.4.8.

available in the paper. In addition, the exact area occupied by neurons and synapses on the chip has not been reported. However, considering the die area of the fabricated olfactory chip, the OTA-based synapse circuit with STDP occupies an area larger than the area required for the design mentioned in Schemmel *et al.* (2006).

Tanaka *et al.* developed an accelerated-time neuromorphic chip with STDP learning in a Hopfield network for associative memory. Although they used a similar VLSI technology to the design presented in Schemmel *et al.* (2006), their implemented synapse takes up significantly larger silicon area. The power consumption of the synapse presented in this work is also 250 μW, which is high for a synapse circuit compared to other designs presented in Table 5.1. In another attempt for implementing STDP, Cruz-Albrecht *et al.* designed a test low-energy STDP circuit and have verified their design in terms of producing STDP learning window and its power consumption (Cruz-Albrecht *et al.* 2012). The STDP synapse presented in this work consumes only 37 pW of power at 100 Hz. On the other hand, this design that utilises different OTAs for realising a STDP learning window, considering its 90 nm design technology, occupies a large silicon area of 64,823 μm^2.

Comparing to all previously mentioned STDP-based learning circuits and systems, the neuromorphic learning network presented in Seo *et al.* (2011), with 256 neurons and 64K synapses, that only consumes 8 nW of power and occupies roughly 13 μm^2 per synapse in the UVT chip, is the most efficient neuromorphic design. It is shown in Seo *et al.* (2011) that this design can be configured for various cognitive tasks such as pattern recognition and classification as well as associative memory.

Further to these designs, Bamford *et al.* (2012b) developed a weight-dependent STDP (W-STDP) circuit, which is different from designs mentioned so far that implemented conventional form of STDP. They showed that the W-STDP design can be implemented using the physical constrains of CMOS transistors and therefore their design has an acceptable area and a low power consumption compared to previous STDP designs. Another W-STDP design is the single transistor synapse device proposed in Ramakrishnan *et al.* (2011). This device utilises a floating gate transistor to implement W-STDP, while the synaptic weight changes are stored in a non-volatile manner in the floating gate. It is shown that this device is able to demonstrate LTP, LTD and STDP behaviors, and is highly scalable.

All neuromorphic systems mentioned so far have used STDP as the learning mechanism in their networks. However, as already mentioned, other synaptic plasticity rules

have also been implemented and tested in neuromorphic systems for applications and synaptic plasticity experiment replications. One of the first designs that used a different rule than STDP for a classification task, was the design presented in Mitra *et al.* (2009) that employs SDSP learning algorithm for synaptic plasticity. The area of this design is comparable to the area required for the STDP learning rule, implemented in previous designs. The authors have also shown the significant performance of the implemented neural network with SDSP learning in classifying complex rate-based patterns (Mitra *et al.* 2009).

Another neuromorphic system that implements a different synaptic plasticity rule rather than STDP is the design presented in Mayr *et al.* (2010) and Mayr *et al.* (2013). This design implements a BCM-like voltage-dependent rule called LCP (See 2.6.1) to replicate synaptic plasticity experiments beyond STDP such as triplet (Froemke and Dan 2002) and frequency-dependent STDP (Sjöström *et al.* 2001). Considering the higher ability in replicating synaptic plasticity experiments compared to STDP, this circuit has higher complexity. However, the presented design in Mayr *et al.* (2013) is in par with most of the STDP designs presented so far in both power and area requirements.

There are also a few biophysical VLSI neuromorphic designs available in the literature that take into account details of synaptic plasticity phenomena and implement its underlying mechanism with a high degree of similarity to biological synapses, in silicon (Meng *et al.* 2011, Rachmuth *et al.* 2011). This similarity results in the specific ability of these synapses to account for both SRDP and STDP experiments and replicate intracellular dynamics of the synapse, where simple previous synapses with STDP fail. It also leads to large silicon area requirement for these circuits, while their reported power consumption is reasonable comparing to most of the other VLSI synaptic plasticity designs presented in Table 5.1.

In addition to the custom made hardware systems that opt to implement a specific type of learning (synaptic plasticity) rule and use it in a specifically designed and structured spiking neural network for an application or neuromorphic research, general neural architectures, such as the Spinnaker (Furber *et al.* 2013) can be instructed, using software, to implement any desired spiking neural network (whether simple or complex) with any learning rule of choice. In Spinnaker system, the targeted neural network is numerically simulated in multiple core processors and the synaptic weights are stored in shared Dynamic Random Access Memory (DRAM). This neural architecture utilises

asynchronous design strategy for global routing in its design, so that the power consumption of the design can potentially be improved. It also uses low-power microprocessors and DRAMs to reduce the power consumption of the system. However, implementing a specific synaptic plasticity rule in this general neural architecture consumes more power than a typical custom VLSI design of that rule, due to its software based approach.

5.6 Applications of Neuromorphic Circuits with Synaptic Plasticity

In order to implement a system with the capabilities close to the brain, many neuromorphic engineers have been following a bottom-up design strategy. Therefore, they start with building basic blocks of the brain in silicon. One of the main building blocks is the synapse that itself includes the synaptic weight plasticity mechanism. This is the main block that brings about learning, memory and computational properties of the neural system (Gerstner and Kistler 2002). In this section, we briefly discuss and review how VLSI implementation of various synaptic plasticity rules can be useful in learning and real-world applications.

Since working with live creatures and measuring experimental data from biological sources is time-consuming and challenging, maybe one of the first applications for a neuromorphic system that contains both neurons and synapses with any desired synaptic plasticity rule, is its use for experimental neuroscientists. These scientists can use a neuromorphic system, which acts according to a desired synaptic plasticity rule and neural combination, and therefore experiment with various features and characteristics in that system. For example, the biophysically inspired iono-neuromorphic circuits proposed in Meng *et al.* (2011) and Rachmuth *et al.* (2011), provide useful insight into how the calcium level alters in a real synapse.

Furthermore, since it is widely believed that synaptic plasticity underlies learning and computational power in the brain (Gerstner and Kistler 2002, Shouval 2011), various mechanisms that have direct or hypothetical relation to the synaptic plasticity experiments, are being used as the learning part of a spiking neural network, to perform various cognitive and machine learning tasks (Mitra *et al.* 2009, Oliveri *et al.* 2007, Masquelier and Thorpe 2010, Ebong and Mazumder 2012, Neftci *et al.* 2013, Schmuker *et al.* 2014, Khosla *et al.* 2014).

It is known that, the spiking behaviour and the activity of the pre- and post-synaptic neurons cause the synapses in the network to adapt themselves to these activities (i.e. learn). These activities that are coded in the form of spikes, represent the input to the network. It is therefore, absolutely essential to first have the correct spike coding structure to effectively represent data to the neural network, and then it is critical to adapt the synapses in a proper way, which is efficient for learning the current type of inputs to the network. This means that the learning mechanism, i.e. the synaptic plasticity rule, can heavily depend on the structure of input to the network, which in turn depends on the application. Sometimes neuroscientists modify a rule or even combine a number of rules to use them for their intended applications. It means that after a careful study of the nature of the input and the required process on that to reach the desired output, they decide on the structure of the learning method.

The study presented in D'Souza *et al.* (2010) shows an example of a learning method, that couples STDP and Spike Frequency Adaptation (SFA) technique for updating synaptic weights, to enable learning in a perceptron like structure. This work proposes an effective platform for sensory guided processing, where two sources of auditory and visual sensory inputs, result in changes in the perceptron neuron spiking activity. It is shown that visual inputs can act as teacher in their used perceptron learning mechanism, while auditory inputs are used for updating the synaptic weights and learning the input auditory patterns (D'Souza *et al.* 2010). Another example is the neuromorphic architecture developed for object recognition and motion anticipation using a modified version of STDP (Nere *et al.* 2012).

In another study, TSTDP is used to generate receptive field development, which is a well-known feature of the rate-based BCM rule (Gjorgjieva *et al.* 2011). Gjorgjieva *et al.* showed that TSTDP can learn up to third-order spatio-temporal correlations that is of importance in neural coding applications (Pillow *et al.* 2008) where the PSTDP rule lacks this capability, even though it is also able to account for the BCM rate-based rule under specific assumptions (Izhikevich 2003, Azghadi *et al.* 2012a). This is useful for developing direction and speed selectivity in the visual cortex (Gjorgjieva *et al.* 2011). Therefore, this rule appears to be useful for pattern classification applications.

The previous three examples show that depending on the needs for the application, and with mathematical and computational analysis, modification to synaptic plasticity rules can be useful for performing tasks, which cannot be carried out with the simple form of the plasticity rules such as STDP, SFA, and BCM. Therefore, the nature and

needs of an application and its inputs has a direct impact on the synaptic plasticity mechanism and hence on its VLSI implementation.

Perhaps the conventional form of STDP, which is according to the formulation shown in Song *et al.* (2000) and is in agreement with PSTDP experiments in Bi and Poo (1998), is the most examined type of synaptic plasticity rule that has been exploited for learning, in various applications ranging from pattern recognition (Masquelier *et al.* 2008), to dataset classification (Oliveri *et al.* 2007), and to topographic mapping formation (Bamford *et al.* 2010). The pair-based STDP has been also utilised for many learning tasks including receptive field development through cortical reorganisation (Young *et al.* 2007), motion anticipation (Nere *et al.* 2012), unsupervised learning of visual features (Masquelier and Thorpe 2007), learning cross-modal spatial transformations (Davison and Frégnac 2006), object recognition (Masquelier and Thorpe 2010), odour data classification (Hsieh and Tang 2012), associative memory type of learning using STDP (Tanaka *et al.* 2009), temporal synchrony detection (Bofill-I-Petit and Murray 2004), robot navigation control (Arena *et al.* 2007, Arena *et al.* 2009) and associative memory, as well as variability and noise compensation tasks (Arthur and Boahen 2006). Although some of these learning applications such as the last five mentioned works, have been successfully implemented as part of a neuromorphic system, many of the other synaptic plasticity rules that have been modelled based on biological experiments performed *in vivo* and *in vitro*, are yet to be explored by neuromorphic engineers for other applications. Examples of these plasticity rules that have not been explored for any application are the hybrid rules proposed in Mayr and Partzsch (2010), Clopath and Gerstner (2010), and Graupner and Brunel (2012) as well as biophysical-based rule proposed in Rachmuth *et al.* (2011), Shouval *et al.* (2002) and Meng *et al.* (2011). In addition to the spike timing-based rules, other spike-based rules such as the SDSP (Brader *et al.* 2007) rule is shown to be useful in other applications such as supervised learning for real-time pattern classification (Mitra *et al.* 2009).

In general, when considering implementing learning (synaptic plasticity) circuits for specific applications such as robotics, neuroprostheses, brain machine interfaces, neural computation, and control, a number of design aspects should be taken into account including (i) the nature of inputs to the system that should be learned; (ii) the level of complexity the implemented system and application can afford; (iii) the use of most appropriate synaptic plasticity rule, in terms of VLSI implementation complexity and performance in processing input neuronal data, which can account for the required

level of performance for the targeted application; (iv) the possible need for modifying the structure of available synaptic plasticity rules for better performance, lower implementation complexity, or input data processing. As an example, here we review a neuromorphic learning network and answer the above mentioned questions about it.

As already discussed, one of the most effective implementations of a VLSI SNN capable of learning to preform a real-world task is the design presented in Seo *et al.* (2011). This neuromorphic system is composed of 256 neurons and 256×256 synapses, in a crossbar array structure to be used for various applications including an associative memory task. The above mentioned questions are answered regarding this system.

(i) The input to this system can be set as 256 spike trains, each one corresponding to a neuron in the network. These 256 spike trains encode the information embedded in the input pattern and present it to the network of neurons. The network changes its weights according to a PSTDP algorithm, in the training phase, when patterns are presented to the network for learning. In the test phase, the neurons are presented with a partial version of the original pattern, and the network through its weights reflects the learned complete pattern, as output spikes.

(ii) The complexity of the targeted task and the number of patterns that can be learned using this neuromorphic system is directly related to the complexity of the network, i.e. its reconfigurability and neuron and synapses count. Since in the present network only 256 neurons with binary synapses are used, as the results in Seo *et al.* (2011) show, the network can only learn 0.047 patterns per neuron in an associative memory task. It is also shown that, if synapses with 4-bit precision are used instead of binary synapses, the learning capacity of the hardware network increases up to 0.109 patterns per neuron.

(iii) The spiking network implemented in this work has a highly reconfigurable structure with on-chip probabilistic PSTDP learning, thanks to its crossbar architecture and transposable synapse SRAM cells, which make PSTDP possible. Therefore, it can realise various network topologies and perform different cognitive tasks. Obviously, for implementing more complex tasks and learning a high number of patterns, a large-scale network with high-precision synapses is needed. Since this design is a basic building block for a scalable neuromorphic system, this extension can be carried out easily. The performance of the associative memory task presented for this network—see (Seo *et al.* 2011)—shows that for this application, simple binary PSTDP synapses integrated with digital integrate and fire neurons are enough. However, one can use

other synaptic plasticity rules with higher complexity, such as those reviewed in this thesis including TSTDP, SDSP, LCP and even biophysically-based rules, to reach better performance or perform more complex tasks. However, the higher cost of these more complex learning circuits should be considered.

(iv) In addition to the main chip that contains 64K probabilistic binary PSTDP synapses and 256 neurons, three different variants of this chip were investigated to follow different targets such as area, power consumption, and learning capability. It is shown that the system with higher learning capability consumes the highest amount of power and occupies the largest silicon real estate among all designs.

In addition to this digital event-driven synchronous neuromorphic learning network that can be scaled up for various real-time learning tasks, in another work IBM scientists have proposed a similar digital event-driven neuromorphic synaptic core, but this time utilised asynchronous operation to decrease the active power of the system, and also implemented learning off-chip. This system has been successfully used in various applications including pattern recognition and auto-associative memory. It also shows a one-to-one correspondence with a neural programming model that makes it possible to realise any type of learning task that can be modelled in software (Arthur *et al.* 2012). The questions mentioned above can be answered for this other neuromorphic learning circuit along the same line as the first discussed design (Seo *et al.* 2011).

It is worth mentioning that the IBM neuromorphic learning network, presented in Seo *et al.* (2011), utilised digital silicon neurons and binary silicon synapses. Therefore, this neuromorphic learning system is not technically subject to device mismatch. However, as already mentioned in Section 5.4.2, when designing a network of analog learning circuits, the device mismatch can lead to inhomogeneity in synaptic plasticity circuits across the network, which may result in an imbalance in potentiation and depression rates, which can affect the learning performance of the system in any targeted application. Hence, a careful assessment of the effect of mismatch while designing neuromorphic learning systems is essential (Poon and Zhou 2011).

5.7 Chapter Summary

This chapter reviewed VLSI implementations of various synaptic plasticity rules ranging from simple timing-based rules such as pair-based STDP rules, to hybrid rules that consider both timing and rate of spikes, to complicated detailed biophysical rules. In

addition, various challenges that neuromorphic engineers encounter when designing neuromorphic systems and synaptic plasticity rules were discussed. Furthermore, the use of these various approaches and the performance of various synaptic plasticity circuits that are implemented as part of different neuromorphic systems were reviewed and compared. Besides, applications of neuromorphic systems with different plasticity rules were highlighted and discussed.

The presented review and discussion in this chapter provide us with a deep insight to the field of designing VLSI circuits for synaptic plasticity rules as part of a neuromorphic system that tends to be used in an engineering application. This insight is quite helpful while designing new circuits for unexplored synaptic plasticity rules such as the TSTDP rule, which is the focus of this thesis. The following chapters present new circuit designs for the TSTDP rule, and show how these designs are able to reproduce the outcomes of a variety of synaptic plasticity experiments ranging from timing-based experiments to rate-based experiments, while they follow various design strategies and targets.

Chapter 6

First VLSI Designs for Triplet-based Spike Timing Dependent Plasticity

AS discussed in Chapter 2, classical model of STDP is incapable of reproducing synaptic weight changes similar to those seen in specific biological experiments that investigate the effects of either higher order spike trains, or the simultaneous effects of the rate and timing of spike pairs on synaptic plasticity. Significantly, a previously described TSTDP rule succeeds in reproducing all of these synaptic plasticity experiments. In this chapter, first, synaptic weight changes using a number of widely used PSTDP circuits are investigated. The investigations show that the class of PSTDP circuits fails to mimic the outcomes of the mentioned complex biological experiments. Then, two new TSTDP VLSI circuits, which are able to reproduce the outcomes of many complex experiments, are presented. To the best of our knowledge, the presented circuits in this chapter are the first VLSI implementations of the TSTDP rule. The new STDP VLSI circuits significantly improve upon previous circuits. These new circuit capabilities in closely mimicking the outcomes of various biological experiments, may play a significant role in future neuromorphic VLSI systems with increased learning ability.

6.1 Introduction

Spike Timing-Dependent Plasticity (STDP) is an unsupervised synaptic plasticity rule that induces changes in individual synaptic weights, based on the timing difference between pre- and post-synaptic spikes (Song *et al.* 2000, Bi and Poo 1998). The classical STDP model employs a pair of spikes (pre-post or post-pre) as the trigger for changes in synaptic plasticity (Song *et al.* 2000, Bi and Poo 1998). However, if the repetition frequency of spike pairs is increased, this model fails to correctly reproduce synaptic weight changes as observed in physiological experiments (Sjöström *et al.* 2001). Furthermore, it is not able to account for experiments using triplet or quadruplet of spikes (Wang *et al.* 2005, Pfister and Gerstner 2006). An explanation for these shortcomings is that traditional pair-based STDP does not account for known nonlinear interactions between successive spikes when more complex spike patterns are used (Froemke and Dan 2002). In order to resolve the short-comings of the classical pair-based model, a simple yet elegant STDP model was proposed by Pfister and Gerstner (2006), where synaptic weight changes based on triplets of spikes was developed.

As already mentioned in Chapter 5, research focusing on translating computational synaptic plasticity models into neuromorphic devices including VLSI implementations of both spiking neurons and synaptic plasticity, in particular STDP, has increased in popularity over the last decade. When considering a VLSI implementation of STDP, several issues need to be addressed such as power consumption, circuit area, noise, output dynamic range, and the accuracy of the design in implementing the targeted synaptic plasticity rule. Currently there are several VLSI implementations of pair-based STDP rule such as the designs presented in Bofill-I-Petit and Murray (2004), Cameron *et al.* (2005), Indiveri *et al.* (2006), Schemmel *et al.* (2006), Tanaka *et al.* (2009), Ramakrishnan *et al.* (2011), and Bamford *et al.* (2012b). However, to the best of our knowledge there is no VLSI implementation available for the TSTDP rule other than those presented in this thesis.

This chapter proposes two new VLSI implementations for TSTDP. These implementations are based on two previous PSTDP circuits already presented in Bofill-I-Petit and Murray (2004) and Indiveri *et al.* (2006). It is shown that both previous PSTDP designs lack the ability to account for the experiments presented in previous experimental papers (Sjöström *et al.* 2001, Froemke and Dan 2002, Wang *et al.* 2005). Furthermore, It is demonstrated that both proposed TSTDP circuits succeed in closely reproducing all the experimentally observed biological effects.

The remainder of this chapter is organised as follows. In Section 6.2, two sample PSTDP circuits are presented and described. Section 6.3 introduces the new TSTDP circuits and describes how they are related to the TSTDP learning rule. Section 6.4 provides information about experimental setup including, experimental protocols, data sets and error functions. Experimental results are provided in Section 6.5, where it is shown that the TSTDP circuit successfully mimic many experiments, while the class of PSTDP circuits, including the two example circuits presented here fails. Section 6.6 compares the two proposed designs and also compares them to previous works that implemented other synaptic plasticity rules with similar or less ability. The chapter finishes by conclusions in Section 6.7.

The results shown in this chapter are presented in *The 21st Japanese Neural Network Society Annual Conference* (Azghadi *et al.* 2011d), as well as in *The 7th IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing* (Azghadi *et al.* 2011c), and *The 2012 IEEE International Joint Conference on Neural Networks* (Azghadi *et al.* 2012b, Azghadi *et al.* 2012a).

6.2 VLSI Implementation of Pair-based STDP

Before discussing the structure of the circuits implementing the PSTDP rule, looking at the PSTDP model that was already described in Section 2.6.1 is useful. In the following, first the PSTDP model is presented and then the circuit models are shown. Note that different parts of both of these circuits are mapped to different terms in the PSTDP model to facilitate understanding the circuits.

6.2.1 Pair-based STDP Model

In the PSTDP plasticity model, potentiation occurs when a pre-synaptic spike precedes a post-synaptic spike; otherwise depression occurs, where weight changes can be governed by a temporal learning window. The classical STDP temporal learning window can be expressed as (Song *et al.* 2000)

$$\Delta w = \begin{cases} \Delta w^{+} = A^{+}\mathrm{e}^{\left(\frac{-\Delta t}{\tau_{+}}\right)} & \text{if } \Delta t > 0 \\ \Delta w^{-} = -A^{-}\mathrm{e}^{\left(\frac{\Delta t}{\tau_{-}}\right)} & \text{if } \Delta t \leq 0, \end{cases} \tag{6.1}$$

where $\Delta t = t_{post} - t_{pre}$ is the time difference between a single pair of post- and pre-synaptic spikes, τ_+ and τ_- are time constants of the learning window, and finally A^+ and A^- represent the maximal weight changes for potentiation and depression, respectively.

A VLSI implementation of the PSTDP rule must account for various parts of the rule shown in Eq. 6.1. As already discussed in Section 5.3.2, there are various VLSI implementations of PSTDP rule available in the literature. In this chapter, two simple PSTDP circuits as representatives of the class of PSTDP circuit models are selected and tested for generating various synaptic plasticity experiments to verify their abilities. The two chosen circuits are the PSTDP circuit proposed by Indiveri *et al.* (2006), and the current mode circuit presented by Bofill-I-Petit and Murray (2004). These circuits are chosen due to their different voltage- and current-mode structures, so that two different techniques are verified in our investigations. In addition, both of these designs are simple and have been proven as silicon devices, which are used for some learning applications. Besides, they use two different types of leaky integrators (see Section 5.2.3) and produce different STDP learning windows (Bofill-I-Petit and Murray 2004, Indiveri *et al.* 2006). In the following section, it is shown how these circuits implement an approximation of the PSTDP model using CMOS transistors.

6.2.2 Indiveri's PSTDP Circuit Model

The implementation by Indiveri *et al.* (2006) was adopted, due to its low power and small area. Fig. 6.1(a) depicts the Indiveri's pair-based STDP circuit schematic and Fig. 6.1(b) demonstrates its resulting temporal learning window for various τ_+ and τ_- (V_{tp}, V_{td}). The timing of pre- and post-synaptic spikes are used to induce weight changes across C_w. This circuit results in a learning window which captures the essential features of STDP, where there are two distinct regions, one for potentiation where $\Delta t \geq 0$ and depression for $\Delta t < 0$.

When a pre-synaptic pulse, V_{pre}, or a post-synaptic pulse ($\overline{V}_{post}$) occurs, V_{pot}(V_{dep}) will be set to zero (V_{dd}). Note that V_{pot} (V_{dep}) then changes linearly over time to reach V_{dd} (zero), and represents the required time constants τ_+ (τ_-). These time constants can be set by changing the gate voltage of the corresponding transistor, i.e. V_{tp} (V_{td}). Fig. 6.1(b) demonstrates the variation of the learning window for different values of V_{tp} (V_{td}), i.e. τ_+ (τ_-). So, if a V_{pre} ($\overline{V}_{post}$) pulse occurs during time determined by its corresponding

time constant, τ_- (τ_+), the output capacitor will be discharged (charged) by a current that is proportional to the value of V_{dep} (V_{pot}) and V_{A-} (V_{A+}).

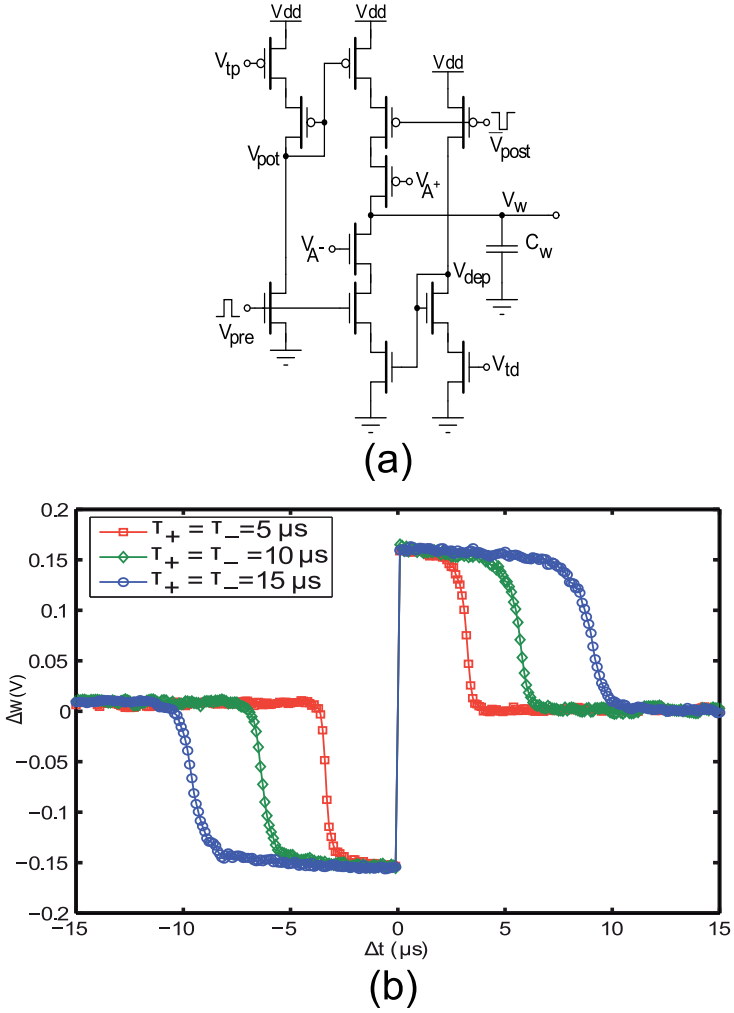


Figure 6.1. Indiveri's PSTDP circuit model. (a) Schematic circuit diagram of Indiveri *et al.* (2006). (b) The learning window of the circuit based on our simulations in an accelerated time scale.

6.2.3 Bofill and Murray's PSTDP Circuit Model

The PSTDP circuit presented by Indiveri *et al.* (2006) cannot reproduce the required exponential behaviour seen in Eq. 6.1, while the circuit proposed by Bofill-I-Petit and Murray (2004) shown in Fig. 6.2 can.

The PSTDP circuit presented in Bofill-I-Petit and Murray (2004) implements a weight-dependent STDP rule, in which the current synaptic weight has an impact on the amount of potentiation/depression. In order to make this PSTDP circuit compatible

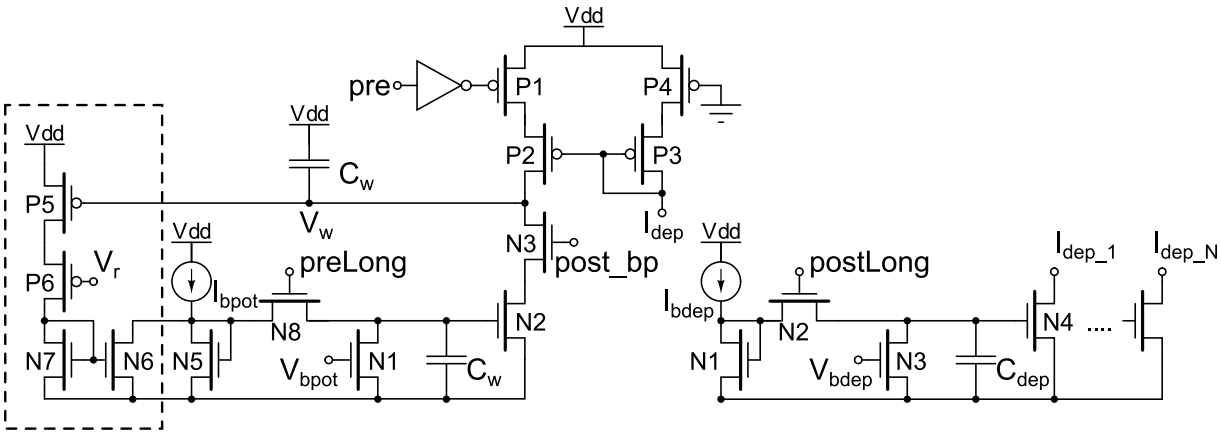


Figure 6.2. Bofill and Murray's PSTDP circuit model. This circuit that is presented in Bofill-I-Petit and Murray (2004) has a weight-dependency part, shown in the dashed-box, which depending on the current weight controls the amount of weight change.

with the normal PSTDP rule, i.e. weight independent, which is also the case for the TSTDP rule, some modifications are needed. These modifications lead to the circuit shown in Fig. 6.3(a). The modifications are as follows: (i) Since the classical PSTDP model (Eq. 6.1) is weight independent, in the modified circuit, the weight dependency part (shown in the dashed box in Fig. 6.2) is omitted. (ii) Potentiation and depression in the modified circuit are represented through an increase or a decrease in the amount of charges stored on the weight capacitor, respectively, which is in contrast to the circuit presented in Bofill-I-Petit and Murray (2004). (iii) Also, in order to simplify the circuit, preLong and postLong pulses which should be generated by an additional circuitry, were replaced with Vpre and Vpost. These signals represent the input pre- and post-synaptic pulses in the modified circuit. (iv) Furthermore, using bias voltages for time constants control results in significant variation in time constants under various 3σ process corners. So, in order to make the circuit more robust against this condition, the bias voltages were represented as the gate-to-source voltage of a number of diode connected transistors that are biased by current sources (M3 and M14). Fig. 6.3(b) demonstrates that using this approach results in a slight change in time constants at all different process corners when compared with MATLAB simulations. Beside time constants, the amplitude constants are also implemented as current sources (I_{pot} and I_{dep}) to be less prone to process variations compared to the original circuit shown in Fig. 6.2, and can later be used to fine tune the time constants when needed. This approach suggests that the time constants, as well as the amplitude parameters are more robust against device mismatch.

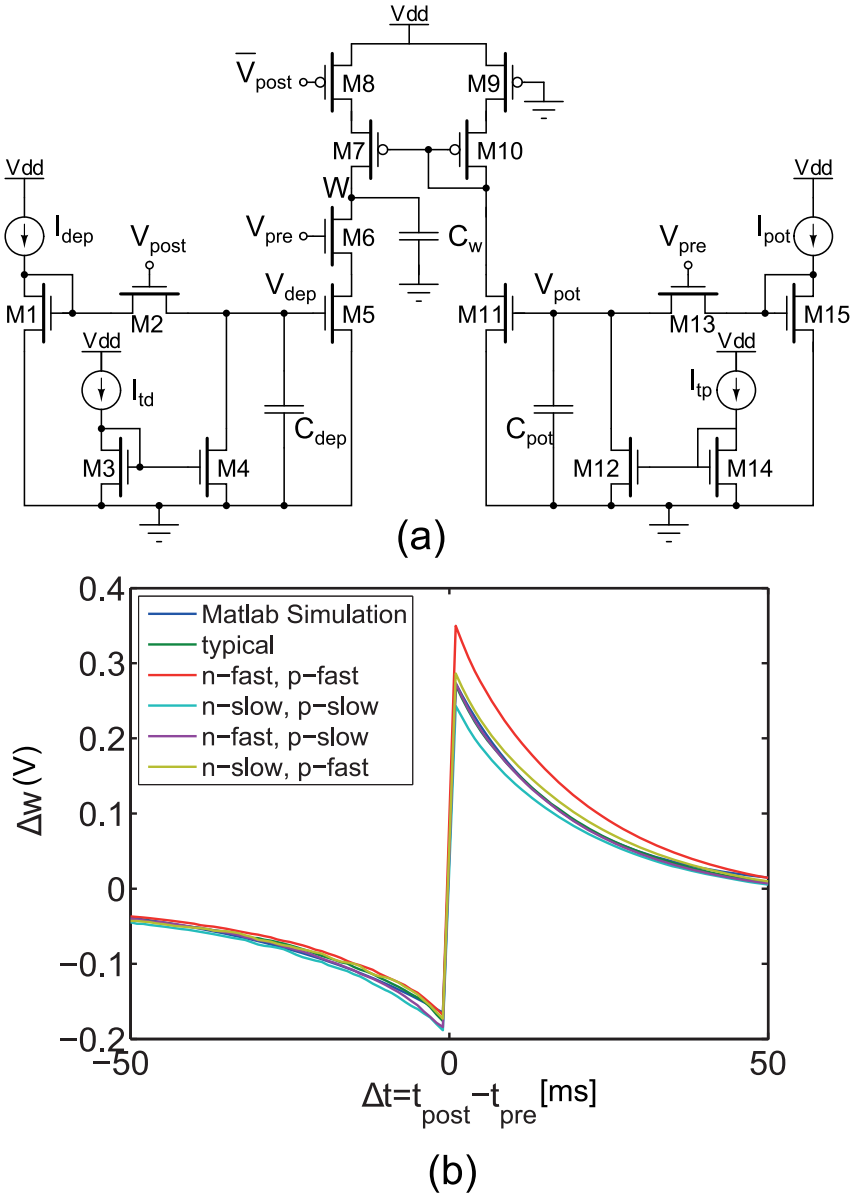


Figure 6.3. Modified PSTDP circuit. (a) The Bofill and Murray's circuit shown in Fig. 6.2 was modified (see the text for more detail). (b) STDP learning window generated by Matlab simulation and also using different transistor process corners for the modified PSTDP circuit shown in part (a). Note that similar protocols and time constants to Bi and Poo (1998) are employed.

6.3 VLSI Implementation of Triplet-based STDP

Before discussing the structure of the circuits that implement the TSTDP rule, looking at the TSTDP model that was already described in Section 2.6.1 is useful. In the following, first the TSTDP model is presented and then the proposed circuit models are shown.

6.3.1 Triplet-based STDP

As already mentioned, previous studies illustrated that classical STDP model fails to reproduce the experimental outcomes involving higher order spike patterns such as triplets and quadruplets of spikes (Wang *et al.* 2005, Pfister and Gerstner 2006). Furthermore, it fails to account for the observed dependence of synaptic plasticity on the repetition frequency of pairs of spikes (Sjöström *et al.* 2001). To resolve these issues, pair-based STDP was extended to include spike triplets that resulted in a spike-triplet-based STDP learning rule which could sufficiently reproduce previously reported physiological experiments. Based on the triplet synaptic learning rule presented in Pfister and Gerstner (2006), the triplet synaptic modification rule can be written as

$$\Delta w = \begin{cases} \Delta w^{+} = A_2^{+} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} + A_3^{+} \mathrm{e}^{(\frac{-\Delta t_2}{\tau_y})} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} \\ \Delta w^{-} = -A_2^{-} \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})} - A_3^{-} \mathrm{e}^{(\frac{-\Delta t_3}{\tau_x})} \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})}, \end{cases} \tag{6.2}$$

where $\Delta w = \Delta w^{+}$ if $t = t_{\text{post}}$ and $\Delta w = \Delta w^{-}$ if $t = t_{\text{pre}}$. Here, A_2^{+}, A_2^{-}, A_3^{+} and A_3^{-} are constants, $\Delta t_1 = t_{\text{post}} - t_{\text{pre}}$, $\Delta t_2 = t_{\text{post(n)}} - t_{\text{post(n-1)}} - \epsilon$ and $\Delta t_3 = t_{\text{pre(n)}} - t_{\text{pre(n-1)}} - \epsilon$, are time difference between combinations of pre- and post-synaptic spikes, τ_-, τ_+, τ_x and τ_y are time constants, and finally ϵ is a small positive value which selects the contribution to the weight change just before the final spike of the presented triplet (Pfister and Gerstner 2006). Hence, triplet-based model induces weight change in proportion to eight parameters (in comparison to four parameters for classical pair-based model); four potentiation parameters (A_2^{+}, τ_+, A_3^{+}, and τ_y) and four depression parameters (A_2^{-}, τ_-, A_3^{-}, and τ_x).

The TSTDP model is not just a simple change in the degree of freedom of the PSTDP model, but it tries to overcome some deficiencies of the PSTDP model. In Pfister and Gerstner (2006), it is addressed that the TSTDP model, removed two main problems of the PSTDP formula. These problems and how the TSTDP solves them are as follows: (i) As PSTDP considers just pairs of spikes, for any value of $A^{+} > 0$, if a pre-synaptic spike precedes a post-synaptic one, it brings about potentiation, while according to Sjöström *et al.* (2001), at low repetition frequencies, there is no potentiation. In the TSTDP model, this deficiency can be solved by setting ${A_2}^{+}$ to a small value or in the case of minimal TSTDP rule, to zero. So it makes the potentiation very small, which can be neutralised by a bit of depression, or it can be zero. (ii) Considering biological experiments in Sjöström *et al.* (2001), for $\Delta t > 0$, potentiation will increase with the increase in frequency. However, this behaviour cannot be generated by PSTDP, as when

the frequency of pairs of spikes increases, it causes the pairs to interact with each other, so it causes no significant potentiation. This problem can be solved again by correctly tuning the TSTDP parameters. In this case, ${A_3}^+$ should be strong enough to make the potentiation win over depression and so to have depression in high frequencies (Pfister and Gerstner 2006).

A VLSI implementation of the TSTDP rule must account for all the mentioned details as well as the various parts of the rule shown in Eq. 6.2. In this chapter, two new designs for TSTDP are proposed, which account for the mentioned details and various parts of the TSTDP rule. These circuits are tested for generating various synaptic plasticity experiments to verify their abilities. The two proposed circuits are built upon the two example PSTDP circuits shown in Section 6.2.

6.3.2 Voltage-mode Triplet-based STDP Circuit Model

Unlike the pair-based model, in the triplet model, a pre-synaptic (post-synaptic) spike further to having an affect on its successive post-synaptic (pre-synaptic) spike can also have an affect on its consecutive pre-synaptic (post-synaptic) spike(s). In the proposed triplet circuit, two more pulses, $V_{\text{post(n}-1)}$ and $\overline{V}_{\text{pre(n}-1)}$, are used in addition to $\overline{V}_{\text{post(n)}}$ and $V_{\text{pre(n)}}$, as shown in Fig. 6.4.

These extra pulses result in the required nonlinearity in the triplet-based model (Pfister and Gerstner 2006). The circuit works as follows: upon the arrival of a post-synaptic pulse, $\overline{V}_{\text{post(n)}}$, the M5, M10 and M18 transistor switches turn on. Then M10 sets a de-potentiating voltage V_{dep1} to V_{dd}. This voltage then starts decaying linearly with time which can result in depression, if a pre-synaptic pulse, $V_{\text{pre(n)}}$ arrives during the time V_{dep1} is decaying to zero (τ_- time constant). In this situation, C_w will be discharged through M7-M9 by a current that is limited by the M7 bias voltage ($V_{A_2^-}$).

In contrast to M10, which can result in depression after receiving a post-synaptic pulse, M5 and M18 can lead to two different potentiations. The first one can occur if M5 turns on during time constant of V_{pot1} (τ_+). This potentiation will be through M4-M6 and is proportional to the bias voltage at M6 ($V_{A_2^+}$). The second potentiation term can charge C_w through M16-M19 and is proportional to $V_{A_3^+}$ if M18 is on at the required time, i.e. when V_{pot1} and V_{pot2} have still kept M16 and M17 on, respectively. This is the term that distinguishes triplet from pair-based STDP, as there is no such term in pair-based STDP. Similarly, upon the arrival of a pre-synaptic pulse, $V_{\text{pre(n)}}$, a potentiating voltage

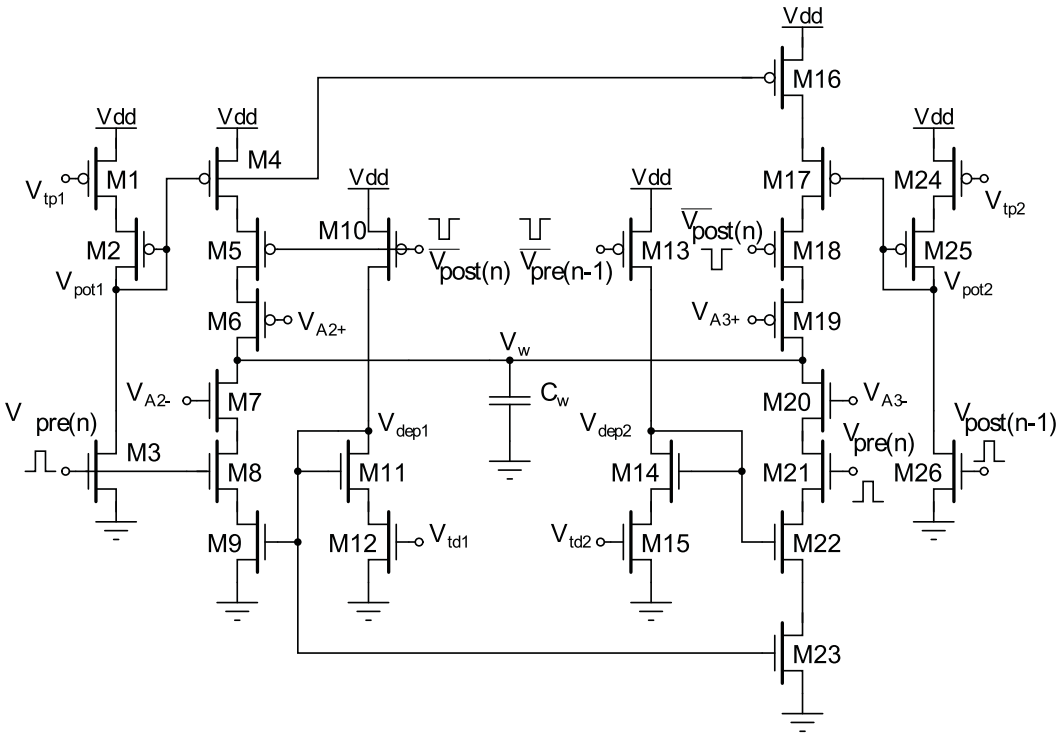


Figure 6.4. Proposed voltage-mode full TSTDP circuit. This figure shows a symmetric voltage-mode design for triplet-based STDP rule (Azghadi *et al.* 2011d).

V_{pot} is set to zero and starts increasing linearly in time which can result in potentiation when a $\overline{V}_{\text{post(n)}}$ pulse arrives within the τ_+ time constant. In addition, two possible depressions proportional to A_2^- and A_3^- can take place, if this pre-synaptic pulse is in the interval area of effect of V_{dep1} and V_{dep2}, i.e. in τ_- and τ_x time constants. It is worth mentioning that the required time constants in the proposed circuit, τ_-, τ_+, τ_x and τ_y, are adjusted by altering their corresponding bias voltages, V_{td1}, V_{tp1}, V_{td2} and V_{tp2}.

6.3.3 Current-mode Triplet-based STDP Circuit Model

Fig. 6.5 presents the proposed current-mode circuit for the TSTDP model. In this circuit, there are eight parameters that can be tuned by controlling eight bias currents as follows: the first four currents including I_{dep1}, I_{pot1}, I_{dep2} and I_{pot2} represent the amplitude of synaptic weight changes for post-pre, pre-post, pre-post-pre and post-pre-post combinations of spike triplets, respectively. Another control parameter for these amplitude values in the circuit is the pulse width of the spikes which was kept fixed during all experiments in this chapter (1 μs). In addition to these amplitude parameters, four more currents control the required time constants in the model for post-pre, pre-post,

pre-post-pre and post-pre-post combinations of spike triplets, and can be adjusted using I_{td1}, I_{tp1}, I_{td2} and I_{tp2} respectively.

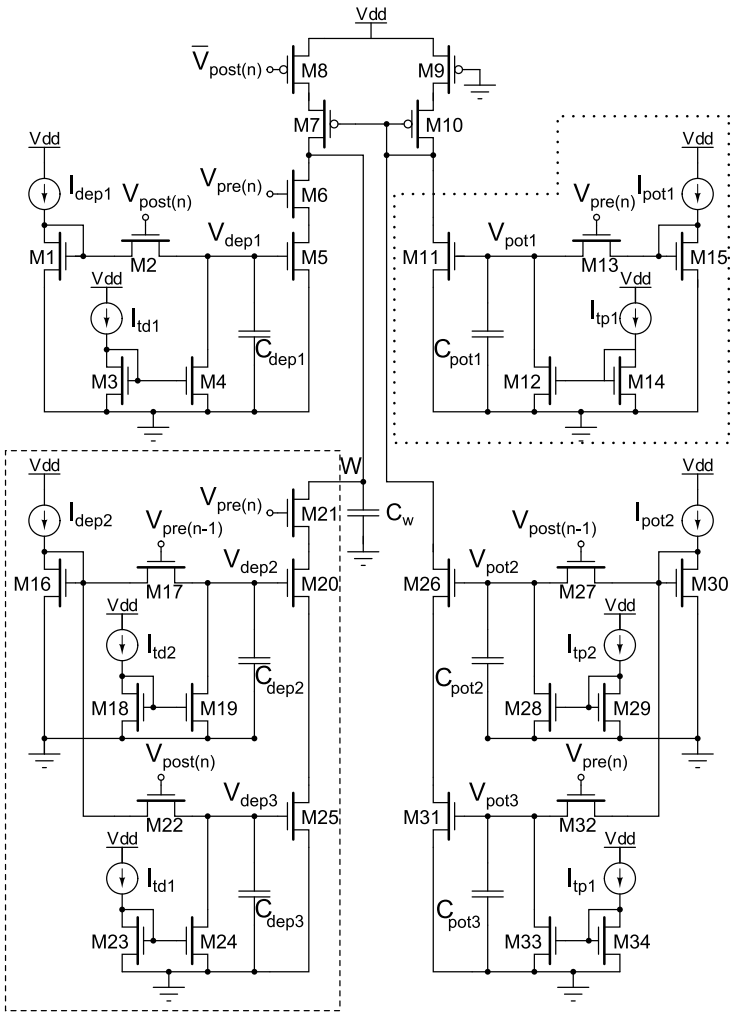


Figure 6.5. Proposed current-mode full TSTDP circuit. This figure shows a current-mode circuit for the TSTDP rule (Azghadi *et al.* 2012b).

The proposed circuit works as follows: upon the arrival of a post-synaptic pulse, $V_{post(n)}$, M2, M8 and M22 switch on. At this time, I_{dep1} can charge the first depression capacitor, C_{dep1}, through M2 to the voltage of V_{dep1}. After finishing $V_{post(n)}$, V_{dep1} starts decaying linearly through M4 and with a rate proportional to I_{td1}. Now, if a pre-synaptic pulse, $V_{pre(n)}$ arrives at M6 in the decaying period of V_{dep1}, namely when M5 is still active, the weight capacitor, C_W, will be discharged through M5-M6 transistors and a depression occurs due to the occurrence of a pre-synaptic pulse in the interval of affect of a post-synaptic spike (post-pre combination of spikes). Additionally, if a pre-synaptic spike arrives at M13, soon before the present post-synaptic spike

at M8, the weight capacitor can be charged through M7-M8 transistors and a potentiation happens. This potentiation happens because the current post-synaptic spike is in the time of affect of a pre-synaptic spike (pre-post combination of spikes). The amount of potentiation depends on V_{pot1}, which itself can be tuned by the relevant amplitude parameter I_{pot1}. Also, the activation interval of M11 can be modified by changing the related time constant parameter I_{tp1}. Furthermore, another contribution to potentiation can occur if a previous post-synaptic pulse, $V_{\text{post(n-1)}}$, arrives at M27 soon enough before the current post-synaptic happens at M8 and also before a pre-synaptic pulse happens at M32 (this is the same pulse as for M13). In this situation, the weight capacitor can be charged again through M7-M8 and by an amount proportional to V_{pot2} and V_{pot3}. This is a triplet interaction in the proposed circuit that leads to the required non-linearity mentioned in the triplet learning rule, appears. A similar description holds for the situation when a pre-synaptic pulse occurs at M6, M13 and M21 transistors. But this time one potentiation and two depression events can happen if the appropriate situation is provided.

The first two parts of this current-mode TSTDP circuit (on the top left and the top right) are identical to the PSTDP circuit presented in Fig. 6.3(a). Also, the two bottom parts of this circuit carry out the triplet terms interactions. This circuit is in correspondence to the full triplet learning rule presented in Pfister and Gerstner (2006) which takes into account all four possible potentiation and depression interactions. However, as it is shown in following sections, only some of the terms are really necessary to reproduce the expected biological experiments. This is referred to as minimal triplet learning rule in Pfister and Gerstner (2006), which makes the required circuit simpler and smaller.

In the next section, the two proposed circuits for the TSTDP model and the mentioned PSTDP circuits are verified for reproducing the outcomes of several biological experiments. These circuits are used in a specific experimental setup, which is explained below.

6.4 Experimental Setup

6.4.1 VLSI Circuit Simulation Setup

The presented simulations of the two proposed TSTDP circuits as well as the two mentioned PSTDP circuits were carried out using parameters for a 0.35 μm standard CMOS

technology in HSpice. The widths and lengths used for all transistors in presented designs are 0.7 μm and 0.35 μm, respectively. The capacitor values for these circuits are as follows. Indiveri's PSTDP and the proposed voltage-mode TSTDP circuits weight capacitors $C_w = 1$ pF. Bofill and Murray's PSTDP circuit and the proposed current-mode TSTDP circuits weight capacitors $C_w = 10$ pF, and other capacitors in these designs are set to 10 fF.

The synaptic weight capacitors in the proposed designs occupy a large portion of the silicon area, as it is the case for almost all synaptic circuits. Therefore, one of the concerns here is to reduce this capacitor value in order to make the area of the design smaller. There are a number of approaches to reduce the size of these capacitors as discussed in Section 5.4.8. In addition, our simulation results also show that, it is possible to scale the capacitor value (and hence, its size) to a significantly smaller value and optimise the circuit biases for the amount of weight changes stored on this smaller capacitor. However, the minimum fitting error (discussed in the next section) might significantly increase.

It should be noted that, during all experiments in this thesis, the nearest spike interaction, which considers the interaction of a spike only with its two immediate succeeding and immediate preceding nearest neighbours, was used (see Section 2.6.1). In addition, the simulations presented in this chapter are all carried out in an accelerated time scale of 1000 compared to real biological time scale. It means that in the performed simulations, each ms corresponds to one second of actual biological time. For the sake of simplicity while comparing the results from the proposed circuits, to the one from biological experiments, all the simulations are scaled back to the real time. The approach of time-scaling has been used extensively in many previous neuromorphic studies such as Schemmel *et al.* (2006), Tanaka *et al.* (2009) and Mayr *et al.* (2010).

6.4.2 Data Sets

In order to compare pair-based and triplet-based VLSI implementations to experimental data, the same experimental protocols and experimental data sets that were used in Pfister and Gerstner (2006) and explained in Chapter 2, are adopted. The simulations were conducted using two types of data sets: The first data set originates from experiments on the visual cortex (Sjöström *et al.* 2001), which investigated how altering the repetition frequency of spike pairings affects the overall synaptic weight change.

This data set is composed of 10 data points that are obtained from Table 1 of Pfister and Gerstner (2006) that represents experimental weight change, Δw, for two different Δt's, and as a function of the frequency of spike pairs under a pairing protocol in the visual cortex. The second experimental data set that was utilised, originates from hippocampal cultures experiments from Wang *et al.* (2005), which examined pairing, triplet and quadruplet protocols effects on synaptic weight change. This data set consists of 13 data points obtained from Table 2 of Pfister and Gerstner (2006). This data set shows the experimental weight change, Δw, as a function of the relative spike timing Δt, Δt_1, Δt_2 and T under pairing, triplet and quadruplet protocols in hippocampal cultures.

6.4.3 Error Function

Identical to Pfister and Gerstner (2006) that tests its proposed triplet model, as well as a PSTDP model simulation results against the experimental data and reports their differences as Normalised Mean Square Error (NMSE) for each data set, here the mentioned PSTDP and TSTDP circuit simulation results are verified under same conditions. The mentioned NMSE (Pfister and Gerstner 2006) is calculated using the following equation (presented already in Section 2.5):

$$E = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{\Delta w^i_{\text{exp}} - \Delta w^i_{\text{cir}}}{\sigma_i}\right)^2, \tag{6.3}$$

where Δw^i_{exp}, Δw^i_{cir} and σ_i are the mean weight change obtained from biological experiments, the weight change obtained from the circuit under consideration, and the standard error mean of Δw^i_{exp} for a given data point i, respectively; p represents the number of data points in a specified data set (can be 10 or 13).

In order to minimise the resulting NMSE for a circuit, there was a need to adjust the parameters and time constants to minimise the resulting NMSE. In the following subsections, the circuit simulation results and applied bias currents, and voltages for setting the required parameters, in order to have the minimum achieved NMSEs for each circuit under test, are reported.

The following section shows the experimental circuit results after optimising circuit biases for the four different PSTDP and TSTDP circuits presented in the previous sections.

6.5 Experimental Circuit Results

6.5.1 Indiveri's PSTDP Circuit Results

In order to test the accuracy of Indiveri's PSTDP circuit, Fig. 6.1(a), first the circuit was checked for producing the frequency-dependent pairing experimental data, i.e. first data set. The circuit parameters were optimised to minimise the NMSE for reproducing the 10 data points available in the first data set. Obtained results shown in Fig. 6.6 clearly demonstrates that this PSTDP circuit fails to mimic the experimental data. The final set of optimised parameters for the shown results are those reported in Table 6.1 and the minimal resulting NMSE was 10.03. This error is close to the error reported in the best case of parameters obtained from classical pair-based STDP model simulations reported in Pfister and Gerstner (2006) where NMSE $\cong$ 7.5—data obtained from Fig. 6 of Pfister and Gerstner (2006).

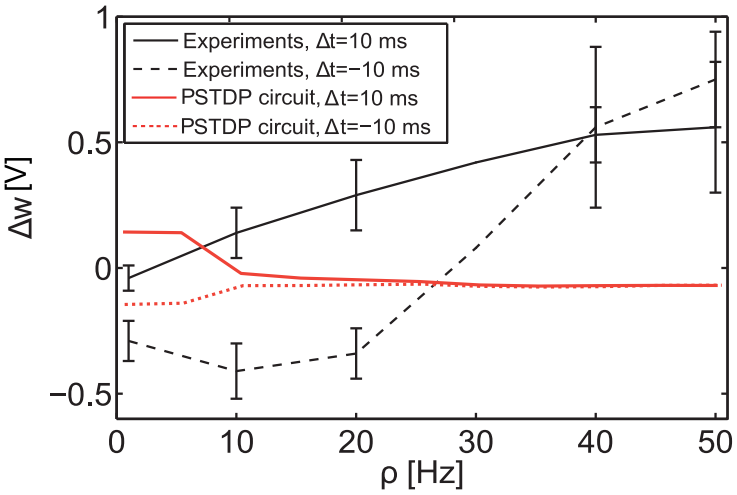


Figure 6.6. Indiveri's PSTDP circuit fails to reproduce the outcomes of frequency-dependent pairing experiments. Frequency-dependent pairing protocol (see Section 2.5.2) is applied to the circuit. Here, ρ is the repetition rate of pre- and post-synaptic spike pairs. Note that there is no experimental data available at $\rho = 30$ Hz.

In addition, further simulations on pairing, quadruplet and triplet protocols were conducted. Again, we optimised the parameters of the VLSI implementation of pair-based STDP so that the NMSE was minimal across the entire data set, i.e. for all three protocols, we employed similar V_{A^+}, V_{A^-}, τ_+ and τ_- values. Obtained results show that the classical VLSI implementation for pair-based STDP, like its mathematical model, fails to reproduce the experimental data obtained using quadruplet, Fig. 6.7(a), and triplet protocols, Fig. 6.7(b)-(c). The NMSE in this case was 11.3, which is close to the

Table 6.1. Indiveri's PSTDP circuit bias voltages for mimicking two different data sets and their resulting NMSEs. The first data set includes 10 data points from pairing frequency experiments presented in Sjöström *et al.* (2001) and the second data set consists of 13 data points from pairing experiments (2 points), quadruplet experiments (3 points) and two different triplet experiments (each one with 4 points) presented in Wang *et al.* (2005).

Data set	V_{A^+} (V)	V_{A^-} (V)	V_{tp} (V)	V_{td} (V)	NMSE
First	2.36	0.69	2.87	0.29	10.03
Second	2.28	0.68	2.8	0.25	11.36

Table 6.2. Bofill and Murray's PSTDP circuit bias currents for mimicking two different data sets and their resulting NMSEs. Similar data sets to Table 6.1 are used.

Data set	I_{pot} (nA)	I_{dep} (nA)	I_{tp} (pA)	I_{td} (pA)	NMSE
First	150	150	24	18	7.26
Second	410	190	20	5	10.76

optimal value obtained from the pair-based model in Pfister and Gerstner (2006)—NMSE $\cong$ 10.5, data obtained from Fig. 6 in Pfister and Gerstner (2006).

6.5.2 Bofill and Murray's PSTDP Circuit Results

Simulation results for the frequency-dependent pairing experiments, using the circuit presented in Fig. 6.3(a) is demonstrated in Fig. 6.8. This figure shows how this PSTDP circuit, similar to the previous PSTDP circuit, fails to reproduce the observed experimental results in visual cortex reported in Sjöström *et al.* (2001). The minimal NMSE obtained in this situation was 7.26, which is consistent with the reported minimal achieved error using computer simulation of the PSTDP rule in Fig. 6A of Pfister and Gerstner (2006). The four required bias currents for controlling the model parameters are reported in Table 6.2.

Simulation results of the second data set also suggest that this PSTDP circuit fails to reproduce experimental results observed in hippocampal cultures. Fig. 6.9(a) shows

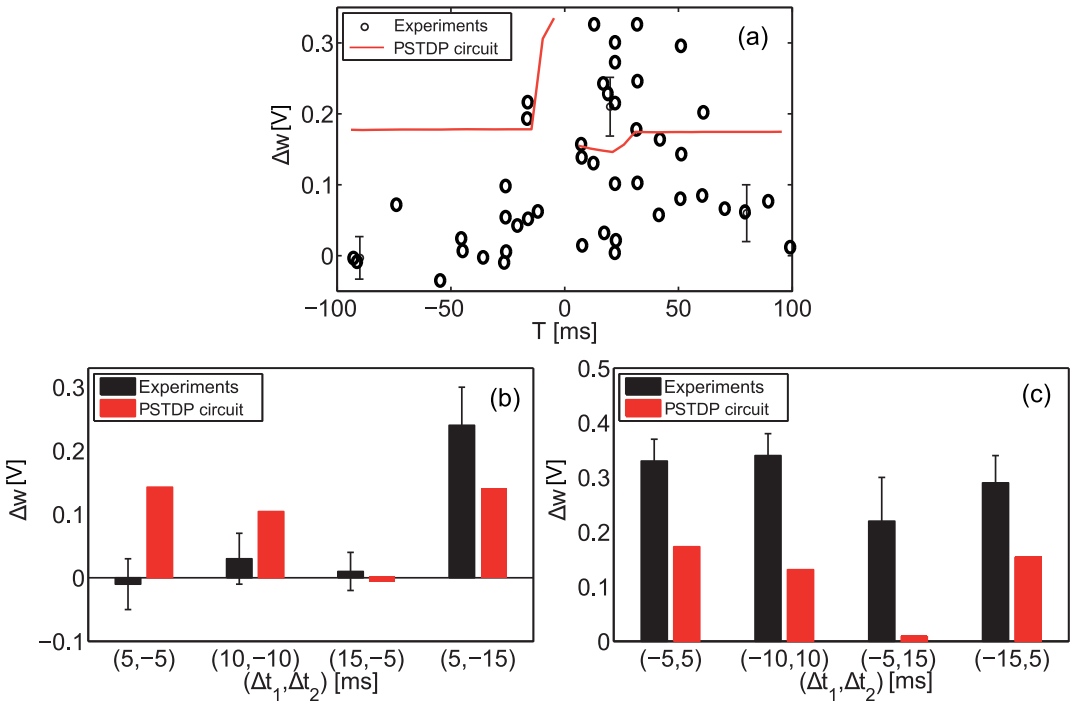


Figure 6.7. Indiveri's PSTDP circuit fails to reproduce the outcomes of triplet and quadruplet experiments. (a) Quadruplet spikes were applied to the circuit, under the quadruplet protocol, and the biases were optimised to minimise the NMSE to fit the 3 quadruplet experimental data points (with error bars) shown in (a), as well as the 8 triplet data bars shown in (b) and (c). (b) Pre-post-pre triplet of spikes were applied to the circuit according to the triplet protocols. (c) Same as (b), but for the post-pre-post spike triplets.

the circuit simulation results along with experimental data for the quadruplet protocol, while Fig. 6.9(b)-(c) represent the results under triplet protocols for pre-post-pre and post-pre-post combinations of spike triplets, respectively. The minimal NMSE obtained in this situation was 10.76, again consistent with the reported results in Fig. 6B of Pfister and Gerstner (2006). The four required bias currents for controlling the model parameters are reported in Table 6.2.

6.5.3 Proposed Voltage-mode TSTDP Circuit Simulation Results

The simulation results shown in Figs. 6.10 and 6.11 demonstrate that the proposed VLSI triplet-based circuit has a significantly improved weight change prediction capability in comparison to its pair-based counterparts. Like pair-based circuit experiments, Fig. 6.10 shows the total weight change induced by a pairing protocol for various pulse

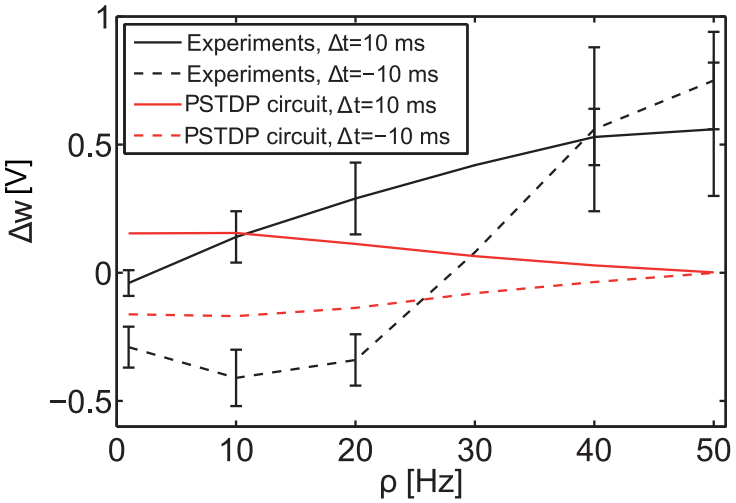


Figure 6.8. Bofill and Murray's PSTDP circuit fails to reproduce the outcomes of frequency-dependent pairing experiments. Frequency-dependent pairing protocol (see Section 2.5.2) is applied to the circuit. Here, ρ is the repetition rate of pre- and post-synaptic spike pairs. Note that there is no experimental data available at $\rho = 30$ Hz.

Table 6.3. Proposed voltage-mode TSTDP circuit bias voltages for mimicking two different data sets and their resulting NMSEs. Similar data sets to Table 6.1 are used.

Data set	V_{A2+} (V)	V_{A2-} (V)	V_{tp1} (V)	V_{td1} (V)	V_{A3+} (V)	V_{A3-} (V)	V_{tp2} (V)	V_{td2} (V)	NMSE
First	2.49	0.59	2.49	0.59	2.36	0.25	2.44	2.6	0.82
Second	2.49	0.66	2.4	0.59	2.3	0.25	2.45	2.7	3.46

repetition rates. As can be seen from the figure, a better match between the experimental data and simulations was observed. The NMSE achieved was 0.82, which is far better than the NMSE for the pair-based case and much closer to the NMSE $= 0.22$, obtained through analytical calculation of the triplet-based model, given in Table 3 of Pfister and Gerstner (2006). The optimised bias voltages to reach this NMSE are shown in Table 6.3.

In addition to the bias optimisation performed on the circuit, to approximate the first data set, the circuit bias parameters were also optimised to approximate the outcome of triplet, and quadruplet experiments. Achieved results for these experiments are shown in Fig. 6.11. The minimal obtained NMSE for this case was 3.46 that is close to the one achieved using the optimised parameters for the TSTDP model (NMSE $= 2.9$) presented in Pfister and Gerstner (2006). The optimised bias voltages to reach this NMSE are shown in Table 6.3.

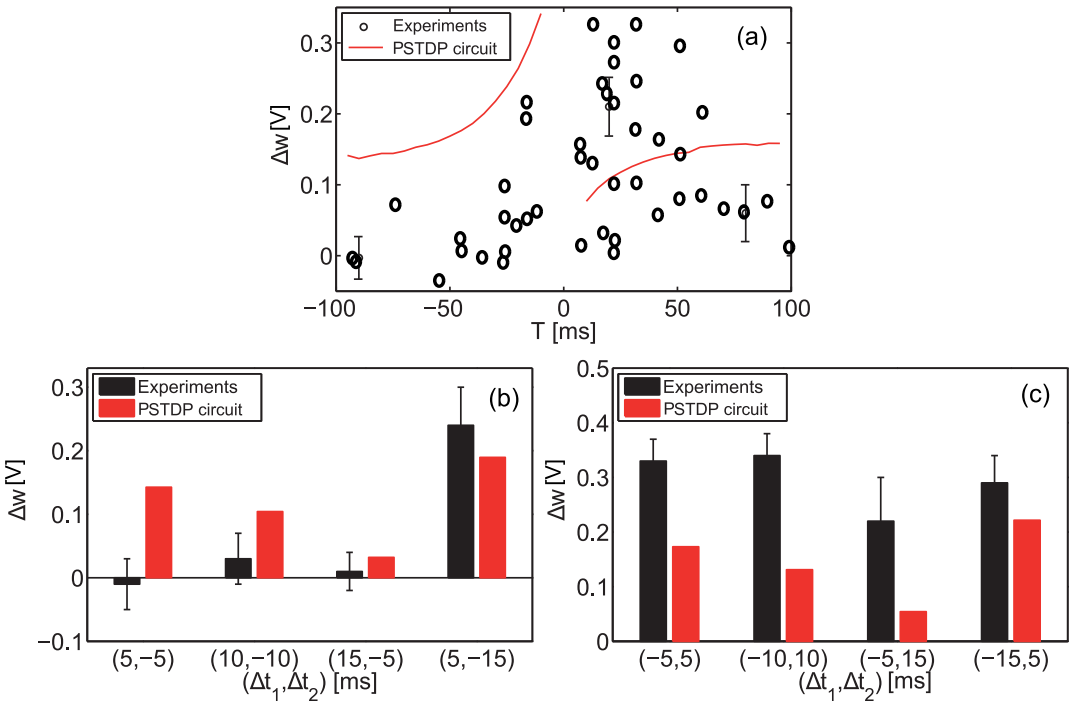


Figure 6.9. Bofill and Murray's PSTDP circuit fails to reproduce the outcomes of triplet, and quadruplet experiments. (a) Quadruplet spikes were applied to the circuit, under the quadruplet protocol, and the biases were optimised to minimise the NMSE to fit the 3 quadruplet experimental data points (with error bars) shown in (a), as well as the 8 triplet data bars shown in (b) and (c). (b) Pre-post-pre triplet of spikes were applied to the circuit according to the triplet protocols. (c) Same as (b), but for the post-pre-post spike triplets.

Note that in the presented simulations for the proposed voltage-mode design, the full TSTDP circuit, which is in accordance to the full TSTDP model is used. However, as discussed in Pfister and Gerstner (2006), the rule can be minimised without having significant effect on the performance of the rule in generating the required experiments. Similarly, in the case of the proposed circuits, simulation results for replicating the shown experiments using minimal versions of the circuit (and minimal circuit), show similar performance in reproducing the outcomes of the targeted experiments. In the case of the minimal model, the corresponding circuit that contains 26 transistors, is minimised to a circuit with 16 transistors for reproducing the visual cortex experiments (first date set). In this case M4-M6, M13-M15 and M20-M23 are removed from the full TSTDP circuit shown in Fig. 6.1. In addition, the circuit is minimised to 19 transistors for the hippocampal experiments (second data set), since M13-M15 and M20-M23 transistors are removed from the full TSTDP circuit.

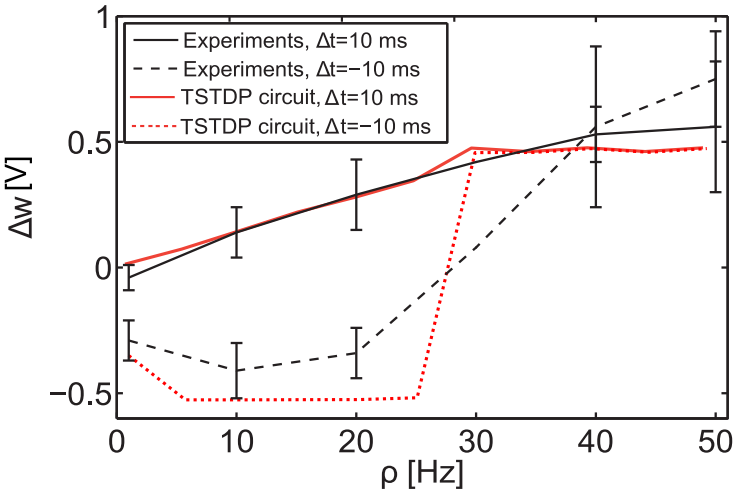


Figure 6.10. Proposed voltage-mode TSTDP circuit mimics the outcomes of frequency-dependent pairing experiments. Frequency-dependent pairing protocol (see Section 2.5.2) is applied to the circuit. Here, ρ is the repetition rate of pre- and post-synaptic spike pairs. Note that there is no experimental data available at $\rho = 30$ Hz.

6.5.4 Proposed Current-mode TSTDP Circuit Simulation Results

In order to test the proposed current-mode TSTDP circuit under the mentioned protocols and using the two data sets, firstly the full TSTDP circuit was employed. This circuit is shown in Fig. 6.5 and consists of four distinct parts each of them related to one of the pair or triplet combination of spikes. However, as stated in Pfister and Gerstner (2006), only some of these combinations are really necessary and play significant roles in synaptic weight change under different protocols. Therefore, the full TSTDP circuit was changed to two minimal TSTDP circuits in correspondence to two minimal TSTDP rules in Pfister and Gerstner (2006). In these circuits, the inconsequential parts of the proposed full-triplet circuit are removed to have the minimal circuits. This is in the contrary to the voltage-mode design simulations that utilised the full TSTDP circuit and optimised all circuit biases to reach a minimal NMSE.

As it can be extracted from the last line of Table 3 in Pfister and Gerstner (2006), the minimal TSTDP rule that is capable of reproducing the expected visual cortex weight change experiments (the first data set), sets pre-post and pre-post-pre spike combination amplitude parameters to zero. This means that this rule neither require the pre-post interactions of spikes, nor the pre-post-pre interactions to take part in synaptic weight modification. Therefore, these parts are not needed also in the corresponding minimal TSTDP circuit. This circuit composed of 19 transistors (exclude the parts in

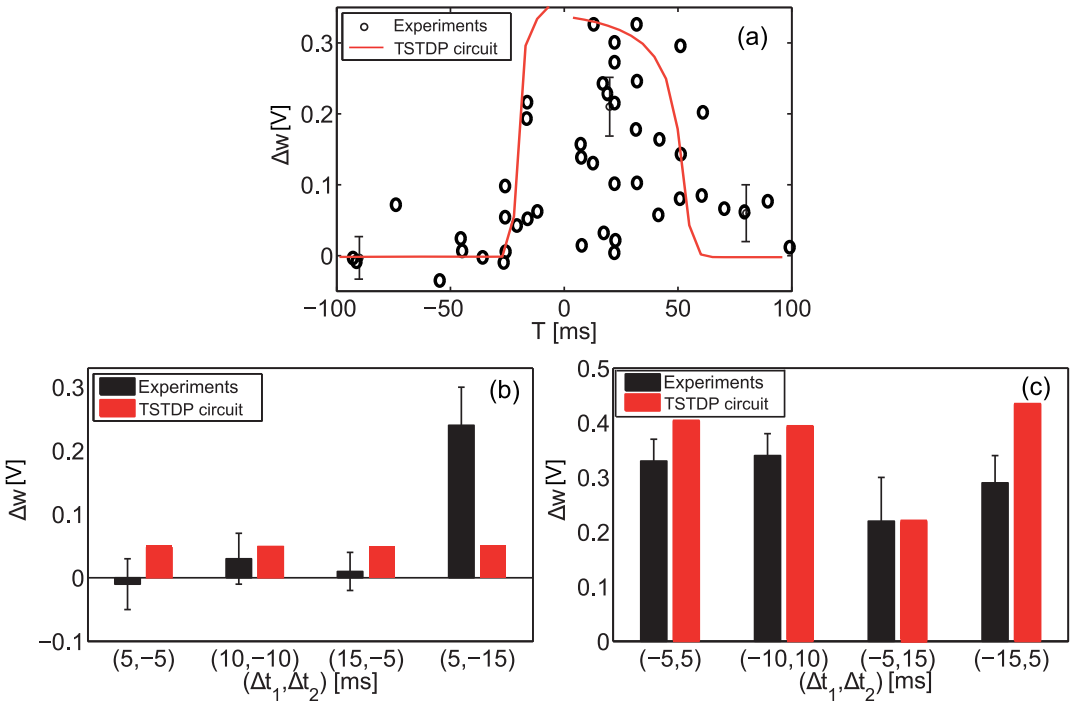


Figure 6.11. Proposed voltage-mode TSTDP circuit mimics the outcomes of triplet, and quadruplet experiments. (a) Quadruplet spikes were applied to the circuit, under the quadruplet protocol, and the biases were optimised to minimise the NMSE to fit the 3 quadruplet experimental data points (with error bars) shown in (a), as well as the 8 triplet data bars shown in (b) and (c). (b) Pre-post-pre triplet of spikes were applied to the circuit according to the triplet protocols. (c) Same as (b), but for the post-pre-post spike triplets.

the dashed and dotted boxes in the circuit presented in Fig. 6.5) and it can reproduce very similar results to the full TSTDP circuit which contains 34 transistors (Fig. 6.5). The minimum NMSE obtained for the first data set and using this first minimal TSTDP circuit was 0.64, which is near the minimum NMSE obtained by means of computer simulations of minimal TSTDP model, NMSE = 0.34, obtained from Table 3 of Pfister and Gerstner (2006). The five required bias currents for controlling the model parameters are reported in Table 6.4.

Furthermore, the minimum obtained NMSE for the second data set using the second minimal TSTDP circuit is 2.25. The achieved results are shown in Fig. 6.13(a)-(c). The second minimal TSTDP circuit is composed of the whole top parts and the right bottom part of the full TSTDP circuit presented in Fig. 6.5—see the last line of Table 4 in Pfister and Gerstner (2006). The obtained NMSE using this circuit is slightly better than the NMSE obtained using minimal TSTDP model and by means of computer simulations,

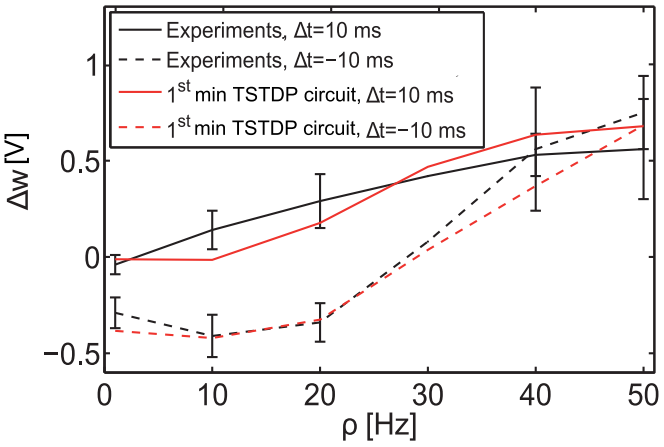


Figure 6.12. Proposed current-mode TSTDP circuit mimics the outcomes of frequency-dependent pairing experiments. Frequency-dependent pairing protocol (see Section 2.5.2) is applied to the circuit. Here, ρ is the repetition rate of pre- and post-synaptic spike pairs. Note that there is no experimental data available at $\rho = 30$ Hz.

Table 6.4. First TSTDP circuit bias currents and its resulting NMSE. The first data set includes 10 data points from pairing frequency experiments presented in Sjöström *et al.* (2001).

Data set	I_{dep1} (nA)	I_{tp1} (nA)	I_{td1} (pA)	I_{pot2} (μA)	I_{tp2} (pA)	NMSE
First	300	40	40	1.5	50	0.64

NMSE $= 2.9$, extracted from Table 4 of Pfister and Gerstner (2006). The six required bias currents for controlling the model parameters are recorded in Table 6.5.

6.6 Discussion

In this chapter, a current-mode (Azghadi *et al.* 2012b) and a voltage-mode (Azghadi *et al.* 2011c) VLSI design, were proposed to implement the TSTDP learning rule. Although it is shown that both circuits are able to account for various synaptic plasticity experiments, the presented results in this chapter suggest that these two circuits have different performance in reproducing the outcomes of various experiments. From the NMSEs obtained for each of the circuits and for two different data sets, it is clear that the current-mode design has a better performance in reproducing the experiments outcomes. This is mainly because of the fact that the current-mode design produces a better exponential behaviour, which is closer to the TSTDP model. On the other hand,

Table 6.5. Second TSTDP circuit bias currents and its resulting NMSE. The second data set consists of 13 data points from pairing experiments (2 points), quadruplet experiments (3 points) and two different triplet experiments (each one with 4 points) presented in Wang *et al.* (2005).

Data set	I_{pot1} (nA)	I_{dep1} (nA)	I_{tp1} (pA)	I_{td1} (pA)	I_{pot2} (nA)	I_{tp2} (pA)	NMSE
Second	160	130	28	20	400	10	2.25

the voltage-mode design has a simpler structure and uses a smaller number of transistors and capacitors, compared to the current-mode design. Therefore, each of the proposed circuit models for TSTDP rule has its own pros and cons in terms of area, power consumption, or synaptic accuracy. The voltage-mode design presents a better area and power performance, but lacks in terms of accuracy, while the current-mode design presents a high accuracy at the cost of more area and power.

Previous computational studies show that PSTDP (Izhikevich and Desai 2003) and TSTDP (Pfister and Gerstner 2006) rules under specific circumstances can reproduce BCM-like learning behaviour with a sliding threshold feature (Bienenstock *et al.* 1982, Cooper *et al.* 2004). As part of the investigations in this thesis, the proposed voltage-mode TSTDP circuit was examined to generate the BCM-like learning behaviour. Obtained results presented in Azghadi *et al.* (2011a) demonstrate that the proposed circuit can successfully generate this behaviour. In another study (Azghadi *et al.* 2012a), in order to compare the performance of both PSTDP and TSTDP circuits for producing the BCM-like behaviour, the voltage-mode PSTDP circuit proposed by Indiveri *et al.* (2006) as well as the proposed voltage-mode TSTDP circuit were stimulated according to a Poissonian protocol (see Section 2.5.6) to reproduce the required behaviour. Simulation results demonstrate that the TSTDP circuit significantly produces the threshold-based behaviour of the BCM. Also, the results suggest that the PSTDP circuit is able to account for the BCM-like behaviour (Azghadi *et al.* 2012a).

In addition to the two proposed TSTDP designs, which have capabilities beyond the PSTDP circuits, there is a previous VLSI implementation proposed by Mayr *et al.* (2010) that is capable of reproducing the mentioned biological experiments (except the quadruplet protocol which has not been shown) similar to the proposed triplet circuits. In terms of functionality, these implementations are different, since the design of Mayr *et al.* (2010) implements the BCM-like rule, requiring the voltage of the neuron to take part

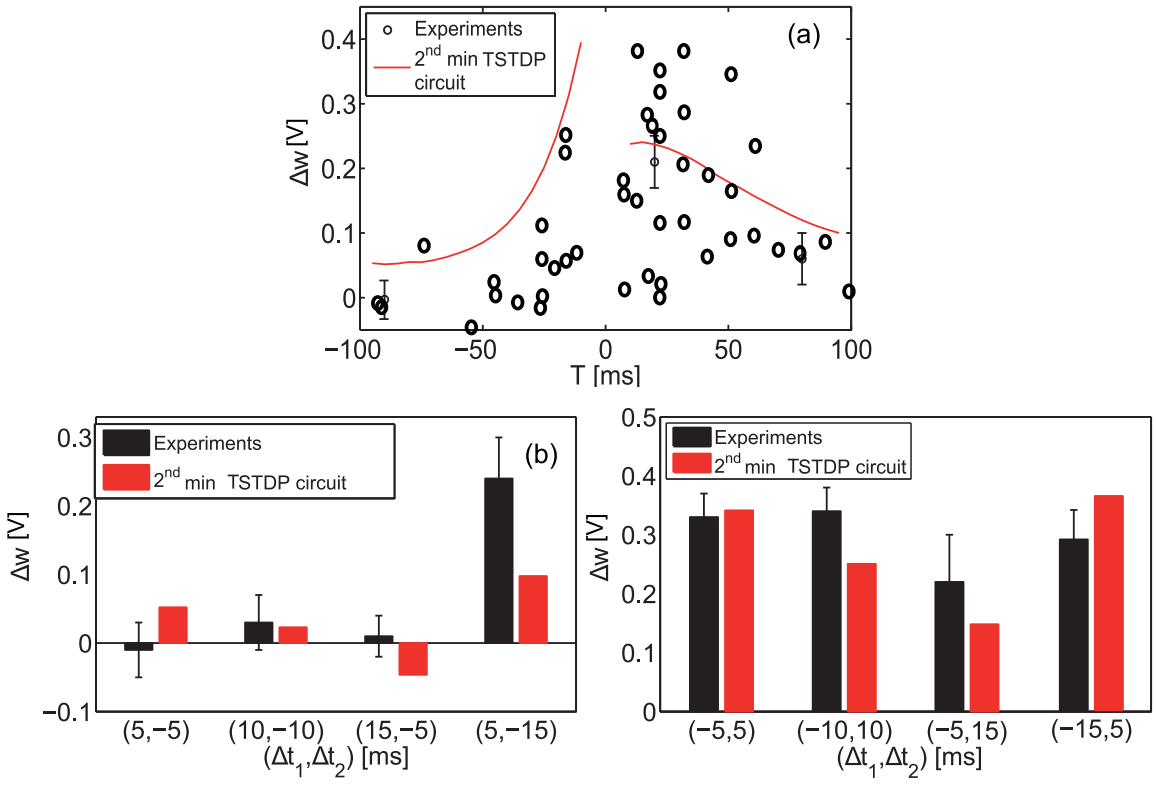


Figure 6.13. Proposed current-mode TSTDP circuit mimics the outcomes of triplet, and quadruplet experiments. (a) Quadruplet spikes were applied to the circuit, under the quadruplet protocol, and the biases were optimised to minimise the NMSE to fit the 3 quadruplet experimental data points (with error bars) shown in (a), as well as the 8 triplet data bars shown in (b) and (c). (b) Pre-post-pre triplet of spikes were applied to the circuit according to the triplet protocols. (c) Same as (b), but for the post-pre-post spike triplets.

in learning, and eventually leads to compatible changes in the neuron architecture. By contrast, the proposed designs in this chapter are based upon triplets of spikes from which the required triplet, quadruplet, and pairing frequency experiments are extracted. In addition, unlike the circuit presented in Mayr *et al.* (2010), the proposed VLSI circuits act as STDP circuits, which can be simply used to connect to other sets of (neuromorphic) neurons of choice. Furthermore, it is shown that the proposed circuits not only can reproduce the required behaviours seen in the mentioned experiments,

but also they can be tuned to mimic those experimentally observed behaviour with a small error, while Mayr *et al.* (2010) just depict the behaviour and not the required values observed in biological experiments. Also, the proposed circuits would require smaller silicon real estates and presents lower power consumption when compared to the circuit presented in Mayr *et al.* (2010).

6.7 Chapter Summary

This chapter reviewed two different VLSI circuits that implement the PSTDP rule and showed how these circuits as the representative of the class of PSTDP circuits, fail to account for a variety of synaptic plasticity experiments. The chapter then introduced two new VLSI circuits of another STDP rule, which in contrast to the normal PSTDP rule that alters the synaptic weight according to the timing of pairs of spikes, works based on the timing of triplets (and not pairs) of spikes. It is shown how the newly proposed VLSI designs generate the outcomes of frequency-dependent pairing experiments, as well as pairing, triplet, and quadruplet biological experiments, with a small error, while the PSTDP circuits have significantly large errors. Although the proposed VLSI circuits presents a novel design in the area of VLSI design for synaptic plasticity rules, the minimal NMSE obtained using these circuits is still significant and should be improved. In addition, the complexity and power consumption of the proposed designs can also be improved using different design techniques and a better approximation of the TSTDP rule. Furthermore, the proposed VLSI designs in this chapter are not able to correctly account for some other synaptic plasticity experiments involving other spike triplets (Froemke and Dan 2002) than those used in the current chapter.

Next chapter proposes a new TSTDP circuit that improves the synaptic ability compared to the designs presented in this chapter and results in significantly lower error, while reproducing the outcomes of many experiments, some of which cannot be replicated by the circuits presented in this chapter.

Chapter 7

High-performance TSTDP VLSI Design

THIS chapter introduces circuit design and implementation of a new high-performance VLSI design for the TSTDP rule that outperforms the other TSTDP VLSI designs in several aspects. It is shown in this chapter, how different terms in the TSTDP synaptic plasticity equation, are implemented to have a very close fit to the model. This results in the proposed design to have significantly lower synaptic plasticity prediction error, in comparison with previous designs for TSTDP and PSTDP rules. In addition, it is shown that the new proposed design can successfully account for a number of new experiments, including experiments involved with various spike triplet combinations, where the previous TSTDP designs do not show acceptable performance and cannot mimic the experiments effectively. This chapter also discusses some of the main challenges in designing the proposed TSTDP circuit such as power consumption, silicon real estate and process variations. We show that it is possible to mitigate the effect of process variations in the proposed circuit. In addition, the power consumption and area of the proposed design are also investigated and discussed in this chapter. The proposed circuit has been fabricated as a proof of principle. Performed chip measurement results testify the correct functionality of the fabricated circuit in performing triplet-based synaptic weight modification.

7.1 Introduction

As already discussed in Chapter 2, neuro-physiological experiments have illustrated that plastic changes to synapses can occur via spike-timing, varying the frequency of inputs to the neuron, or changes to internal concentration of calcium in the neuron's spine apparatus (Bi and Poo 1998, Sjöström *et al.* 2001, Wang *et al.* 2005). Many theoretical and experimental studies have focused on studying changes to synaptic strength caused by STDP (Gerstner *et al.* 1996, Bi and Poo 1998, Song *et al.* 2000, Froemke and Dan 2002, Wang *et al.* 2005, Pfister and Gerstner 2006, Iannella and Tanaka 2006, Iannella *et al.* 2010). At the same time, there have been attempts at translating such rules to VLSI circuit implementations (Bofill-I-Petit and Murray 2004, Indiveri *et al.* 2006, Tanaka *et al.* 2009, Rachmuth *et al.* 2011, Azghadi *et al.* 2011a, Azghadi *et al.* 2011b, Azghadi *et al.* 2011c, Azghadi *et al.* 2011d, Azghadi *et al.* 2012a, Azghadi *et al.* 2012b). These attempts represent the crucial technological steps in developing smart VLSI chips with adaptive capabilities similar to that of the mammalian brain. The long term aim is to have VLSI circuits that can learn to adapt to changes and result in modifying their functionality to improve their performance. The realisation of such adaptive VLSI circuits will have widely varying applications ranging from artificial bionic prostheses through to improved autonomous navigation systems.

The main contribution of this chapter is to significantly advance previous VLSI implementations of triplet-based STDP and introduce a new synaptic analog circuit that possesses some critical capabilities that have not been demonstrated in previous VLSI implementations. The proposed circuit not only can replicate known outcomes of STDP, including the effects of input frequency, but also it is capable of mimicking BCM-like behaviour (Bienenstock *et al.* 1982). It improves the synaptic weight change modification ability and results in less error while curve fitting the experimental data. In addition, the proposed circuit captures important aspects of both timing- and rate-based synaptic plasticity that is of great interest for researchers in the field of neuromorphic engineering, specifically to those who are involved in experiments dealing with learning and memory *in-silico*.

This chapter is organised as follows. Section 7.2 represents a different arrangement of the previously introduced TSTDP rule, that is helpful in understanding the design procedure of the new proposed TSTDP circuit. In Section 7.3, a description of the proposed circuit operation is given. Section 7.4 is dedicated to simulation results where the various capabilities of the proposed circuit are illustrated. Section 7.5 discusses

and describes the effects of process variation and transistor mismatch on the proposed design, and suggests a tuning mechanism to overcome the performance degradation in the presence of physical variations. Section 7.6 describes the VLSI implementation of the proposed high-performance design and presents chip measurement results. Section 7.7, provides a discussion of advantages as well as limitations of the proposed design and suggests ways and their costs, in order to reduce the limitations of the novel presented circuit. Section 7.8 gives concluding remarks of the chapter.

The results shown in this chapter are presented mainly in *Neural Networks* (Azghadi *et al.* 2013a).

7.2 A Different Arrangement of Triplet-based STDP

In order to have a better understanding of the structure of the proposed circuit in this chapter, the TSTDP synaptic plasticity rule is again mentioned here. However, a different arrangement of the rule presented in Eq. 2.3 has been utilised, which facilitates understanding the structure of the new proposed circuit. This new arrangement of the TSTDP rule (Pfister and Gerstner 2006) is given by

$$\Delta w = \begin{cases} \Delta w^{+} = \mathrm{e}^{\left(\frac{-\Delta t_1}{\tau_+}\right)}\left(A_2^{+} + A_3^{+}\mathrm{e}^{\left(\frac{-\Delta t_2}{\tau_y}\right)}\right) \\ \Delta w^{-} = -\mathrm{e}^{\left(\frac{\Delta t_1}{\tau_-}\right)}\left(A_2^{-} + A_3^{-}\mathrm{e}^{\left(\frac{-\Delta t_3}{\tau_x}\right)}\right), \end{cases} \tag{7.1}$$

where $\Delta w = \Delta w^{+}$ for $t = t_{\text{post}}$ and if $t = t_{\text{pre}}$ then the weight change is $\Delta w = \Delta w^{-}$. A_2^{+}, A_2^{-}, A_3^{+} and A_3^{-} are potentiation and depression amplitude parameters, $\Delta t_1 = t_{\text{post(n)}} - t_{\text{pre(n)}}$, $\Delta t_2 = t_{\text{post(n)}} - t_{\text{post(n-1)}} - \epsilon$ and $\Delta t_3 = t_{\text{pre(n)}} - t_{\text{pre(n-1)}} - \epsilon$, are the time differences between combinations of pre- and post-synaptic spikes. Here, ϵ is a small positive constant which ensures that the weight update uses the correct values occurring just before the pre or post-synaptic spike of interest, and finally τ_-, τ_+, τ_x and τ_y are time constants (Pfister and Gerstner 2006). Prior to this TSTDP model, there was another rule proposed by Froemke and Dan (2002) which considers higher order temporal patterns (quadruplets) of spikes to induce synaptic modification. Both of these rules tend to explore the impact of higher order spike patterns on synaptic plasticity. In this study, the proposed analog circuit aims to mimic the model presented in Eq. 7.1.

7.3 High-performance Circuit for TSTDP Rule

The new high-performance circuit produces a close fit to the outcomes of the TSTDP rule. Fig. 7.1 presents the proposed circuit implementation of the full TSTDP model. In the full TSTDP model, there are eight parameters that can be tuned in the proposed circuit, by controlling eight bias currents as follows: I_{dep1}, I_{pot1}, I_{dep2} and I_{pot2} represent the amplitude of synaptic weight changes for post-pre (A_2^-) and pre-post (A_2^+) spike pairs, and pre-post-pre (A_3^-) and post-pre-post (A_3^+) combinations of spike triplets, respectively. Another control parameter for these amplitude values in the circuit is the pulse width of the spikes, which was kept fixed during all experiments in this chapter (1 μs). In addition to these amplitude parameters, the required time constants in the model for post-pre (τ_-), pre-post (τ_+), pre-post-pre (τ_x) and post-pre-post (τ_y) spike patterns, can be adjusted using I_{td1}, I_{tp1}, I_{td2} and I_{tp2} respectively (see Eq. 7.1 and Fig. 7.1).

The proposed circuit works as follows: upon the arrival of a pre-synaptic pulse, $V_{\text{pre(n)}}$, M9 and M15 are switched on. At this time, I_{pot1} can charge the first potentiation capacitor, C_{pot1}, through M9 to the voltage of V_{pot1}. After finishing $V_{\text{pre(n)}}$, V_{pot1} starts decaying linearly through M11 and with a rate proportional to I_{tp1}. Now, if a post-synaptic pulse, $V_{\text{post(n)}}$ arrives at M13 in the decaying period of V_{pot1}, namely when M12 is still active, the weight capacitor, C_{W}, will be discharged through M12-M13 transistors and a potentiation occurs because of the arrival of a post-synaptic pulse in the interval of effect of a pre-synaptic spike (pre-post combination of spikes). Additionally, if a post-synaptic spike has arrived at M19, soon before the current pre-synaptic spike at M15, the weight capacitor can be charged through M14-M15 transistors and a depression happens. This depression happens because the present pre-synaptic spike is in the time of effect of a post-synaptic spike (post-pre combination of spikes). The amount of depression depends on V_{dep1}, which itself can be tuned by the relevant amplitude parameter I_{dep1}. Also, the activation interval of M18 can be modified by changing the related time constant parameter I_{td1}. Furthermore, another contribution to depression can occur if a previous pre-synaptic pulse, $V_{\text{pre(n}-1)}$, has arrived at M26 soon enough before the current pre-synaptic happens at M15 and also before a post-synaptic pulse happens at M19. In this situation, the weight capacitor can be charged again through M14-M15 by an amount proportional to an effect of both V_{dep2} and V_{dep1}, simultaneously. This triplet interaction leads to the required non-linearity mentioned in the triplet learning rule.

A similar description holds for the situation when a post-synaptic pulse occurs at M13 and M19 transistors. But this time one depression will take place as the result of charging the weight capacitor up through M14-M15 and because of an arriving post-synaptic spike at M19 before a pre-synaptic spike at M15. Besides, two potentiation events can happen if an appropriate situation is provided to discharge the weight capacitor because of a pre-post or a post-pre-post combination of spikes. Note that, in this implementation, the synaptic strength is inversely proportional to the voltage stored on the weight capacitor, C_W. However, for the sake of simplicity when comparing the achieved results to experimental data, the weights are shown in a consistent way to biological data, i.e. potentiation with positive strength and depression with negative strength.

Upon examination of the TSTDP expression (Eq. 7.1), there are four different parts that need to be implemented, in order to satisfy the equation as accurately as possible. The proposed circuit (Fig. 7.1) is composed of four leaky integrators which are arranged in a way that form the required addition and multiplications in the formula in a simple manner. Furthermore, in order to have the exponential behaviour required for the TSTDP rule, M5, M12, M18 and M25 are biased in the subthreshold region of operation. The most left part of the circuit implements the potentiation triplet component of the rule using a simple leaky integrator and the resulting current produced by this part ($I_{\text{pot}-\text{trip}}$) is given by

$$I_{\text{pot}-\text{trip}} = A_3^+ \text{e}^{(\frac{-\Delta t_2}{\tau_y})}, \tag{7.2}$$

where I_{pot2} represents A_3^+, I_{tp2} can control τ_y and finally $\Delta t_2 = t_{\text{post(n)}} - t_{\text{post(n}-1)} - \epsilon$ controlled by M2 and M13. Next, $I_{\text{pot}-\text{trip}}$ is added up to I_{pot1} current which represents A_2^+ in the TSTDP formula (Eq. 7.1). Hence, the amount of current going to M8 transistor is given by

$$I_{\text{M8}} = A_2^+ + A_3^+ \text{e}^{(\frac{-\Delta t_2}{\tau_y})}. \tag{7.3}$$

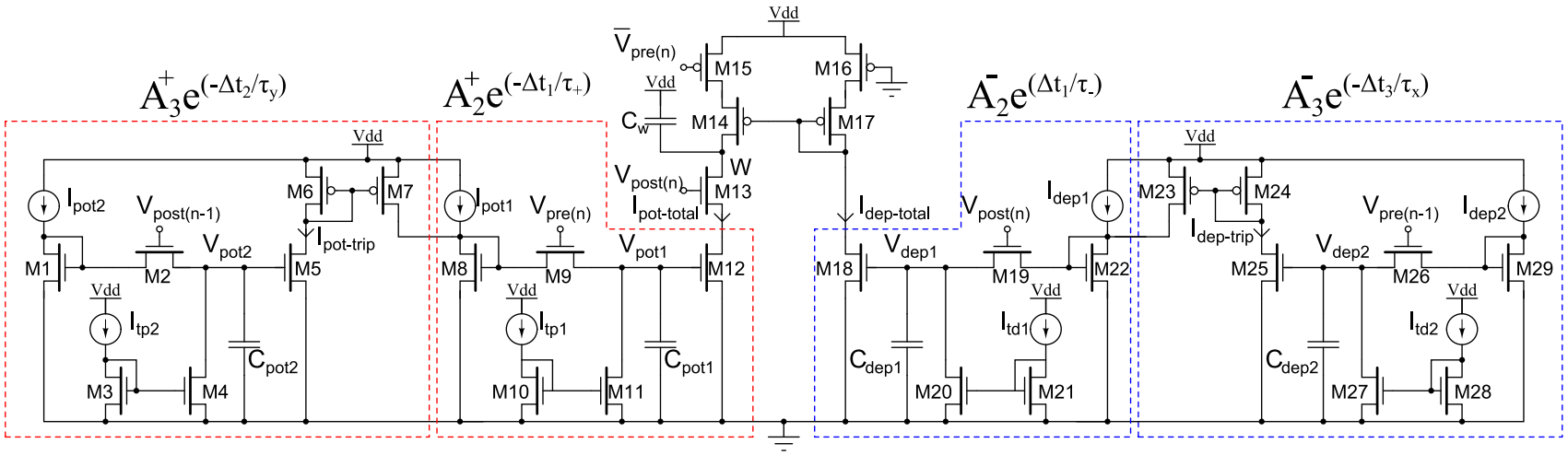


Figure 7.1. Proposed circuit for the full triplet-based STDP rule. Each section of the circuit aims to implement one part of the rule mentioned in Eq. 7.1. There are two potentiation parts which are shown in red dashed boxes and two depression parts that are presented in blue dashed boxes. The analytical term corresponding to each part of the circuit is depicted on each box also. Note that, the synaptic weight is inversely proportional to the value over weight capacitor, C_W.

This current then goes to the second leaky integrator on the second left box in Fig. 7.1 and will result in $I_{\text{pot}-\text{total}}$ passing through M12 and M13 and discharging the weight capacitor, C_W, hence causes a potentiation equal to Δw^+. The amount of this current which is in result of the contribution of both triplet and pair-based spike patterns, can be written as

$$I_{\text{pot}-\text{total}} = e^{\left(\frac{-\Delta t_1}{\tau_+}\right)}\left(A_2^+ + A_3^+ e^{\left(\frac{-\Delta t_2}{\tau_y}\right)}\right), \tag{7.4}$$

where I_{tp1} can control τ_+ and finally $\Delta t_1 = t_{\text{post(n)}} - t_{\text{pre(n)}}$ is controlled by M9 and M13.

The same approach applies for the depression part of Eq. 7.1. There are two leaky integrators (the blue boxes in Fig. 7.1), each one is responsible for building an exponential current and the final current ($I_{\text{dep}-\text{total}}$) which will be mirrored through M14 and M17 into the weight capacitor and result in charging the weight capacitor and hence depression. This is the full TSTDP circuit which realises the full-TSTDP rule (Eq. 7.1). However, according to the analytical calculations and numerical simulations presented in Pfister and Gerstner (2006), some parts of the full TSTDP rule may be omitted without a significant effect on the efficiency of the rule when replicating biological experiments. Pfister and Gerstner called these new modified rules, minimal triplet rules.

According to the first minimal TSTDP rule, when generating the biological experiment outcomes for the visual cortex data set presented in Sjöström *et al.* (2001), the triplet contribution for depression, as well as the pairing contribution of the potentiation parts of Eq. 7.1 can be dismissed (i.e. $A_3^- = 0$ and $A_2^+ = 0$) and the outcome will be quite similar to using the full TSTDP rule—Table 3 in Pfister and Gerstner (2006). Furthermore, the second minimal TSTDP rule which considers a zero value for A_3^- (Eq. 7.1) has quite similar consequences to the full TSTDP rule and allows reproducing the hippocampal culture data set experimental data presented in Wang *et al.* (2005).

As the rules are simplified, the full TSTDP circuit also can be minimised. This minimisation can be performed by removing those parts of the circuit that correspond to the omitted parts from the full TSTDP model. These parts are M23-M29 transistors which can be removed when $I_{\text{dep2}} = 0$ (i.e. $A_3^- = 0$). Also I_{pot1} can be set to zero, as it represents A_2^+ that is not necessary for the first minimal triplet rule. The resulting minimal circuit based on these assumptions is shown in Fig. 7.2 with two separate parts for potentiation and depression.

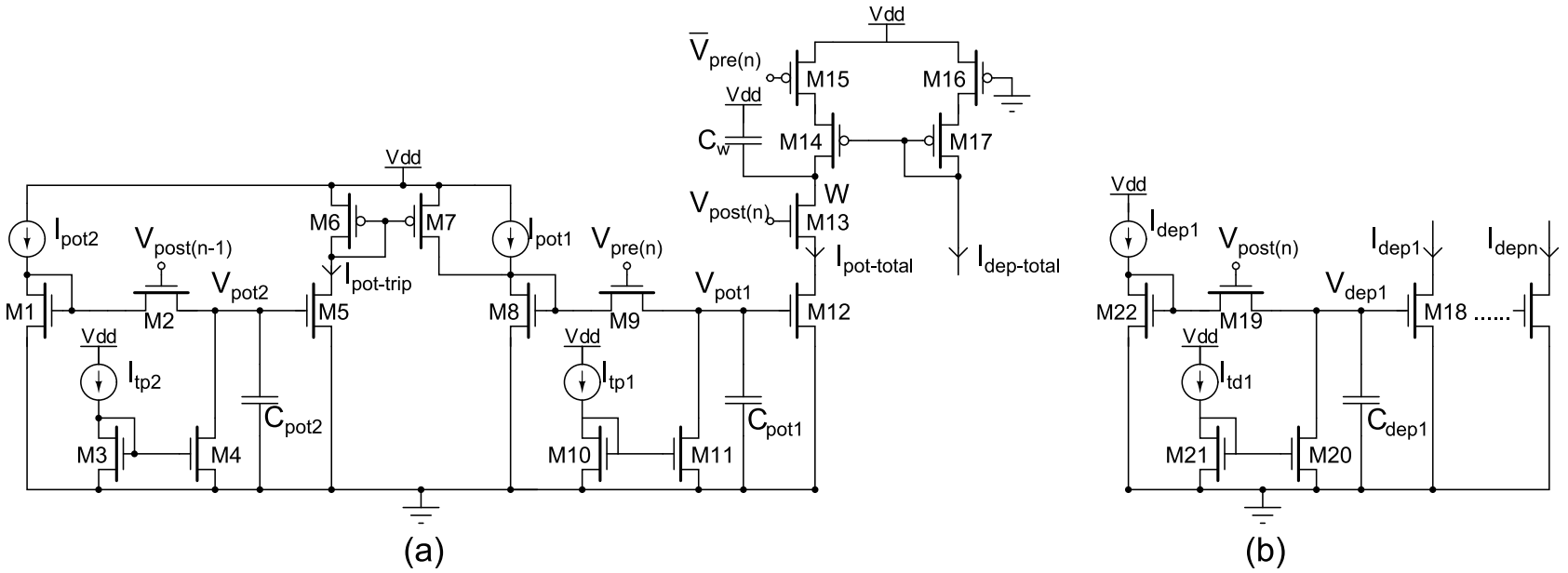


Figure 7.2. Proposed minimal triplet-based STDP circuit. The synaptic weight is inversely proportional to the value over weight capacitor, C_W. (a) This part of the circuit brings about potentiation due to pre-post and post-pre-post combinations of spikes. Potentiation means discharging the weight capacitor through M12-M13. This part of the circuit should be replicated for all synapses (on all dendrite branches come from various pre-synaptic neurons). (b) This section of the circuit is responsible for depression through charging the weight capacitor. This part needs to be implemented only once per neuron and it can result in area saving as it does not need to be replicated for all synapses.

The potentiation part (a) which is composed of two leaky integrators is responsible for voltage decrements across the weight capacitor (potentiation), in case of pre-post or post-pre-post of spike patterns in the required timing periods. This part receives two inputs backpropagated from the post-synaptic neuron ($V_{\text{post(n}-1)}$, and $V_{\text{post(n)}}$), and another input forwarded from a pre-synaptic neuron ($V_{\text{pre(n)}}$). As there can be several synapses on each post-synaptic neuron, this part of the minimal circuit which receives inputs from different pre-synaptic neurons, needs to be replicated for every synapse. However, the depression part of the minimal circuit, part (b), just receives an input from the post-synaptic neuron and hence can be replicated once per neuron. That is why we represent the potentiation and depression inversely to the charge stored on the weight capacitor. As the number of neurons is significantly lower than the number of synapses, this area saving can result in a significantly smaller area for a large neuromorphic system with TSTDP synapses. A similar approach was also utilised by Bofill-I-Petit and Murray (2004).

7.4 Simulation Results

The proposed circuit shown in Fig. 7.2 was simulated using parameters for the 0.35 μm C35 CMOS process by AMS. All transistors in the design are set to 1.05 μm wide and 0.7 μm long. The capacitor values are 10 pF for the weight capacitor and 100 fF for all the capacitors in the leaky integrators. The circuit was simulated in Spectre within Cadence and circuit bias optimisation was performed using HSpice and Matlab, as described later in this chapter. All reported experiments in this chapter assume the nearest spike interaction, which considers the interaction of a spike only with its two immediate succeeding and preceding nearest neighbours (see Section 2.6.1). Furthermore, in order to facilitate the simulation of the circuits, a scaling approach, which has been used in similar VLSI implementations of synaptic plasticity e.g. (Schemmel *et al.* 2006, Tanaka *et al.* 2009, Mayr *et al.* 2010), was adopted, which uses a time scale of microseconds to represent milliseconds, i.e a scaling factor of 1000. However, in all simulation results presented in this chapter, the results are scaled back to biological time in order to facilitate comparisons with published data from biological experiments.

In order to validate the functionality of the proposed TSTDP circuit, 12 different patterns of spikes including spike pairs (four patterns), spike triplets (six patterns) and spike quadruplets (two patterns) were utilised. These patterns were applied to the circuit and recorded weight changes were compared to their corresponding experimental

data. All simulation results show a good match to their related experimental data. The first and second simulations were performed using two different data sets and for different experimental protocols. The optimisation scheme and the data fitting method used here were that of Pfister and Gerstner (2006). The required experimental protocols, different sets of data, the data fitting method as well as the achieved simulation results, are explained and presented in the following subsections.

Additionally, for the third set of simulations, the proposed circuit was examined for generating weight changes using all six possible spike triplet patterns presented in Froemke and Dan (2002). Furthermore, the circuit was also used to reproduce the weight changes produced by the rate-based BCM rule under a Poissonian protocol. The achieved results for these two simulations, the triplet and Poissonian protocols are also explained in the following subsections.

7.4.1 The Proposed Circuit Response to Various Experimental Protocols

The experimental protocols used to stimulate the proposed triplet circuit are those explained in Section 2.5. The results presented in this section are due to the different protocols to the proposed circuit.

Pairing Protocol

Fig. 7.3 shows that the proposed minimal triplet circuit can reproduce the exponential learning window produced by both PSTDP and TSTDP models, under the conventional pairing protocol described in Section 2.5.1 and adopted in many experiments (Bi and Poo 1998, Wang *et al.* 2005). This exponential learning window can also be reproduced using a number of the previously described PSTDP circuits e.g. Bofill-I-Petit and Murray (2004).

However, it has been illustrated in Sjöström *et al.* (2001) that altering the pairing repetition frequency affects the total change in weight of the synapse. As it is shown in Azghadi *et al.* (2011c) and Azghadi *et al.* (2012b) and was discussed in Chapter 6, PSTDP circuits are not capable of reproducing such biological experiments that investigators examine the effect of changes in pairing frequency on synaptic weight. However, Fig. 7.4 illustrates how the proposed TSTDP circuit can readily reproduce these experiments.

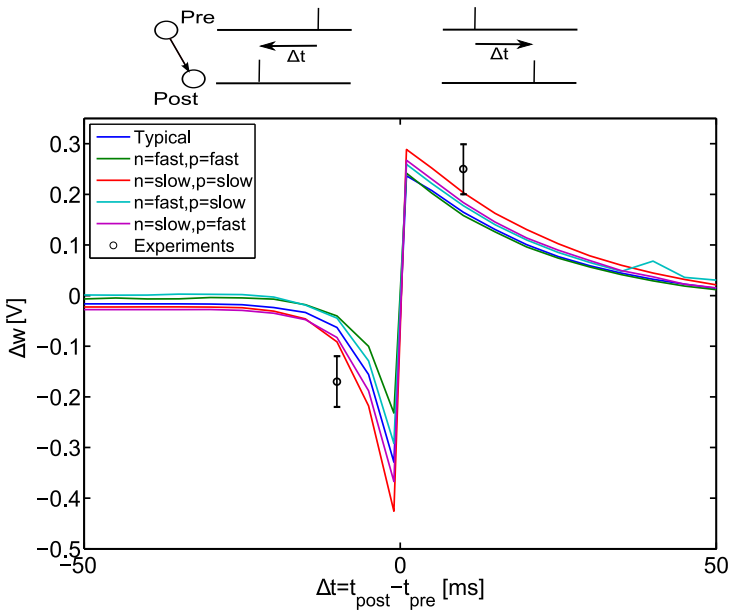


Figure 7.3. Exponential learning window produced by the proposed minimal TSTDP circuit. The learning window was produced based on the pairing protocol for different transistor process corners. The required bias currents taken for the triplet circuit correspond to the hippocampal culture data set (Table 7.1). Experimental data and error bars are extracted from Wang *et al.* (2005).

Triplet Protocol

There are two types of triplet patterns that are used in the hippocampal experiments, which are also adopted in this chapter to compute the prediction error as described in Section 2.5.3. Figures 7.5(a)-(b) show how the proposed minimal triplet circuit produces a close fit to the triplet experiments reported in Wang *et al.* (2005).

Quadruplet Protocol

Fig. 7.6 shows the weight changes produced by the proposed minimal TSTDP circuit under quadruplet protocol conditions. This protocol is explained in detail in Section 2.5.5. Identical to Pfister and Gerstner (2006), in all quadruplet experiments in this chapter, $\Delta t = -\Delta t_1 = \Delta t_2 = 5\ \mu s$. Note that none of the previously proposed PSTDP circuits are capable of showing such a close fit, neither to triplet, nor to quadruplet experiments (Azghadi *et al.* 2011c, Azghadi *et al.* 2012b).

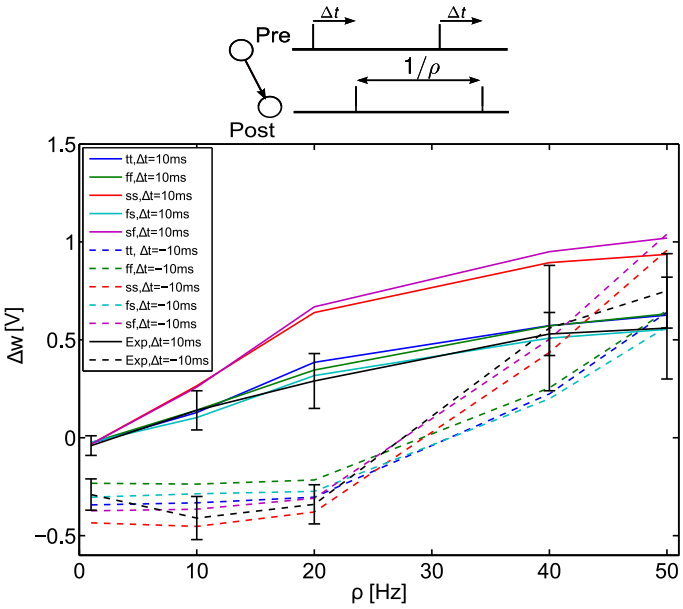


Figure 7.4. Weight changes produced by the proposed minimal TSTDP circuit under a frequency-dependent pairing protocol. This figure demonstrates weight changes in a pairing protocol as a function of the pairing frequency, ρ. The synaptic weight changes are reproduced by the proposed minimal TSTDP circuit for different transistor process corners. Experimental data points and error bars are extracted from Sjöström *et al.* (2001)—no data point at $\rho = 30$ Hz. The required bias currents taken for the triplet circuit correspond to the visual cortex data set (Table 7.1).

7.4.2 Data Sets

The proposed circuit is expected to be capable of reproducing experimental weight changes induced by pairing, triplet and quadruplet protocols in hippocampal cultures reported in Wang *et al.* (2005). It should also be able to reproduce experimental weight changes induced by a pairing protocol and in the presence of spike pairing frequency changes, in the visual cortex presented in Sjöström *et al.* (2001). In order to check if the proposed circuit is capable of doing so, simulations were conducted using two types of data sets similar to those used in Chapter 6. The first data set originates from experiments on the visual cortex which investigated how altering the repetition frequency of spike pairings affects the overall synaptic weight change. This data set is composed of 10 data points—obtained from Table 1 of Pfister and Gerstner (2006)—which represent experimental weight change, Δw, for two different Δt's, and as a function of the frequency of spike pairs under a pairing protocol in the visual cortex (10 black data points and error bars shown in Fig. 7.4). The second experimental data set that

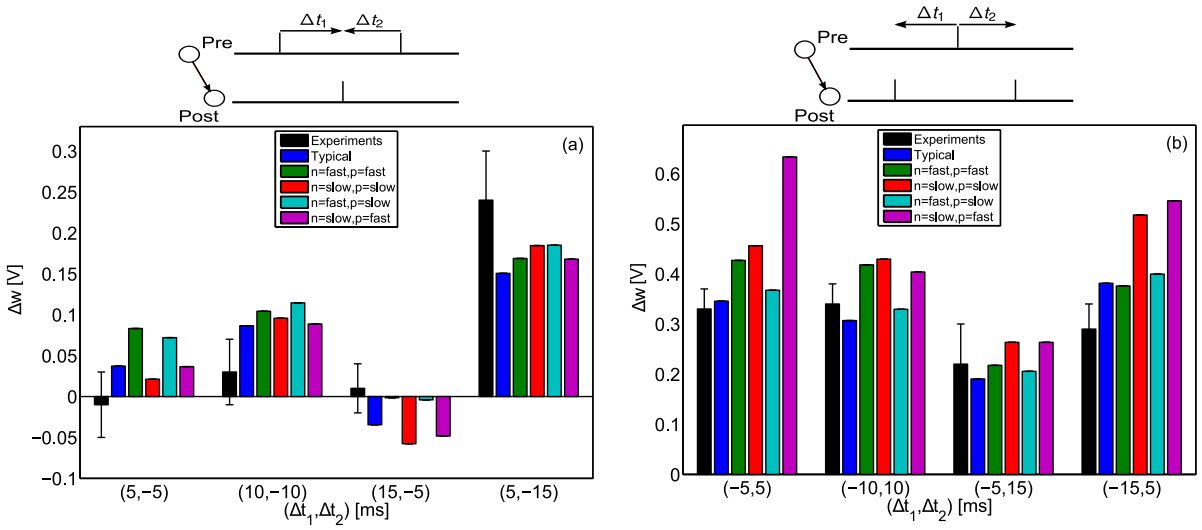


Figure 7.5. Weight changes produced by the proposed minimal TSTDP circuit under triplet protocol for two different spike triplet combinations. (a) This shows weight changes induced by the triplet protocol for the pre-post-pre combination of spikes. The synaptic weight changes are reproduced by the proposed minimal TSTDP circuit for different transistor process corners. (b) This shows the same case as (a), but for the post-pre-post combination of spikes. The required bias currents taken for the triplet circuit correspond to the hippocampal culture data set (Table 7.1).

was utilised originates from hippocampal culture experiments, which examine pairing, triplet and quadruplet protocols effects on synaptic weight. This data set consists of 13 data points obtained from Table 2 of Pfister and Gerstner (2006) including (i) two data points and error bars for pairing protocol in Fig. 7.3, (ii) three data points and error bars for quadruplet protocol in Fig. 7.6, and (iii) eight data points and error bars for triplet protocol in Figures 7.5(a) and (b). This data set shows the experimental weight change, Δw, as a function of the relative spike timing Δt, Δt_1, Δt_2 and T under pairing, triplet and quadruplet protocols in hippocampal culture.

7.4.3 Data Fitting Approach

Identical to Pfister and Gerstner (2006) that test their proposed triplet model simulation results against the experimental data using a Normalised Mean Square Error (NMSE) for each of the data sets, and similar to the error measurements in experiments performed in Chapter 6, the proposed circuit is verified by comparing its simulation results with the experimental data and ensuring a small NMSE value. The NMSE is calculated using the following equation:

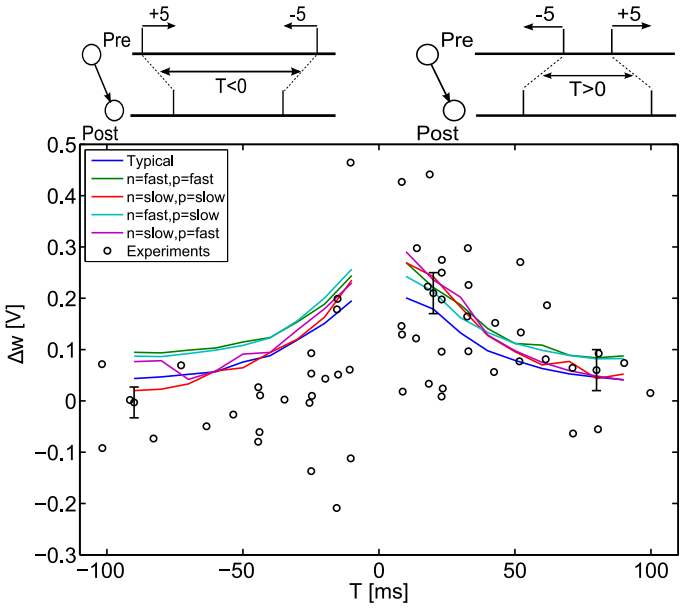


Figure 7.6. Weight changes produced by the proposed minimal TSTDP circuit under quadruplet protocol. The synaptic weight changes are reproduced by the proposed minimal TSTDP circuit for different transistor process corners. The required bias currents taken for the triplet circuit correspond to the hippocampal culture data set (Table 7.1). Experimental data points and error bars are after Wang *et al.* (2005).

$$\mathrm{NMSE} = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{\Delta w^{i}_{\mathrm{exp}} - \Delta w^{i}_{\mathrm{cir}}}{\sigma_i}\right)^2, \tag{7.5}$$

where $\Delta w^{i}_{\mathrm{exp}}$, $\Delta w^{i}_{\mathrm{cir}}$ and σ_i are the mean weight change obtained from biological experiments, the weight change obtained from the circuit under consideration, and the standard error mean of $\Delta w^{i}_{\mathrm{exp}}$ for a given data point i, respectively; p represents the number of data points in a specified data set (can be 10 or 13).

In order to minimise the resulting NMSEs for the circuit and fit the circuit output to the experimental data, there is a need to adjust the circuit bias parameters and time constants. This is an optimisation process of the circuit bias currents which results in reaching a minimum NMSE value and so the closest possible fit to the experimental data. In the following subsection, the optimisation method used to tune the circuit bias currents is introduced.

Table 7.1. Minimal TSTDP circuit bias currents and the resulted NMSEs for the two data sets. The visual cortex data set includes 10 data points from pairing frequency experiments presented in Sjöström *et al.* (2001) and the hippocampal data set consists of 13 data points from pairing experiments (2 points), quadruplet experiments (3 points) and two different triplet experiments (each one with 4 points) presented in Wang *et al.* (2005).

Data set	I_{pot1}	I_{dep1}	I_{tp1}	I_{td1}	I_{pot2}	I_{tp2}	NMSE
Visual cortex	0	220 nA	500 pA	140 pA	1.15 μA	80 pA	0.33
Hippocampal	130 nA	190 nA	900 pA	170 pA	280 nA	140 pA	1.74

7.4.4 Optimisation Method

In order to minimise the NMSE function mentioned above and achieve the highest analogy to the experimental data, the circuit bias currents which tunes the required parameters from the model should be optimised as it is the case for TSTDP model parameters (Eq. 7.1). For this purpose, Matlab and HSpice were integrated in a way to minimise the NMSE resulted from circuit simulations using the Matlab built-in function `fminsearch`. This function finds the minimum of an unconstrained multi-variable function using a derivative-free simplex search method. Table 7.1 demonstrates bias currents achieved from the mentioned optimisation method in order to reach the minimum NMSE for the two sets of data: the visual cortex data set and the hippocampal culture data set. The minimum obtained NMSEs for the visual cortex and hippocampal data sets are also presented in Table 7.1. These results are consistent with the obtained NMSEs using TSTDP model reported in Pfister and Gerstner (2006).

In addition to the above mentioned experiments that have been carried out in Pfister and Gerstner (2006), the proposed design has been additionally tested for all possible combination of spike triplets. Applied protocol and more explanation on these experiments are provided in the following subsection.

7.4.5 Extra Triplet Patterns

Apart from reproducing the behaviour of the TSTDP model proposed by Pfister and Gerstner (2006), the proposed circuit is also able to reproduce the observed weight modifications for other combinations (rather than pre-post-pre or post-pre-post) of

spikes triplets which have not been explored in Pfister and Gerstner (2006), but have been used in another set of multi-spike interaction experiments performed by Froemke and Dan (2002). In these experiments, six different combinations of spike triplets induce synaptic weight changes. These changes in Froemke and Dan (2002) have been calculated according to a suppressive model described in Section 2.6.1.

The simulation protocol (for suppressive STDP model) as described in Froemke and Dan (2002) is as follows; a third spike is added either pre- or post-synaptically to the pre-post spike pairs, to form a triplet. Then this triplet is repeated 60 times at 0.2 Hz to induce synaptic weight changes. Here, the same protocol has been used to stimulate the proposed minimal TSTDP circuit. In this protocol, there are two timing differences shown as $\Delta t_1 = t_{\text{post}} - t_{\text{pre}}$ which is the timing difference between the two most left pre-post or post-pre spike pairs, and $\Delta t_2 = t_{\text{post}} - t_{\text{pre}}$ which is the timing difference between the two most right pre-post or post-pre spike pairs.

Although the proposed circuit implements the triplet model presented in Pfister and Gerstner (2006), and not the suppressive model in Froemke and Dan (2002), obtained results shown in Figs 7.7(a)-(b) demonstrate qualitative regional agreement with the reported results in Froemke and Dan (2002). Nonetheless, there is a direct contrast between our circuit results and their results in the post-pre-post case of spike patterns.

Indeed, the weight changes induced by the pre-post-post, post-post-pre, pre-pre-post, and pre-post-post spike triplets are significantly matched to the weight changes resulted from the similar spike patterns obtained from the Froemke-Dan model. However, there is a slight difference in the results for pre-post-pre and a significant difference in the results for post-pre-post spike combinations when using these two different models. Right bottom square in Fig. 7.7(a) which represents the post-pre-post case shows potentiation as it is the case for the post-pre-post spike pattern case in Fig. 7.5(b) also. However Froemke-Dan model results show a depression for this spike combination—Fig. 3b in Froemke and Dan (2002). According to the discussion provided in Pfister and Gerstner (2006), the difference in the result is due to the nature of the original suppressive rule where post-pre-post contributions gave rise to a depression, in contrast to TSTDP where this specific combination leads to potentiation. Note that the Froemke-Dan revised model presented in 2006 addressed this issue, since in this model there are two different potentiation and depression saturation values (Froemke *et al.* 2006). This revised model now reproduces the expected experimental outcomes from Sjöström *et al.* (2001).

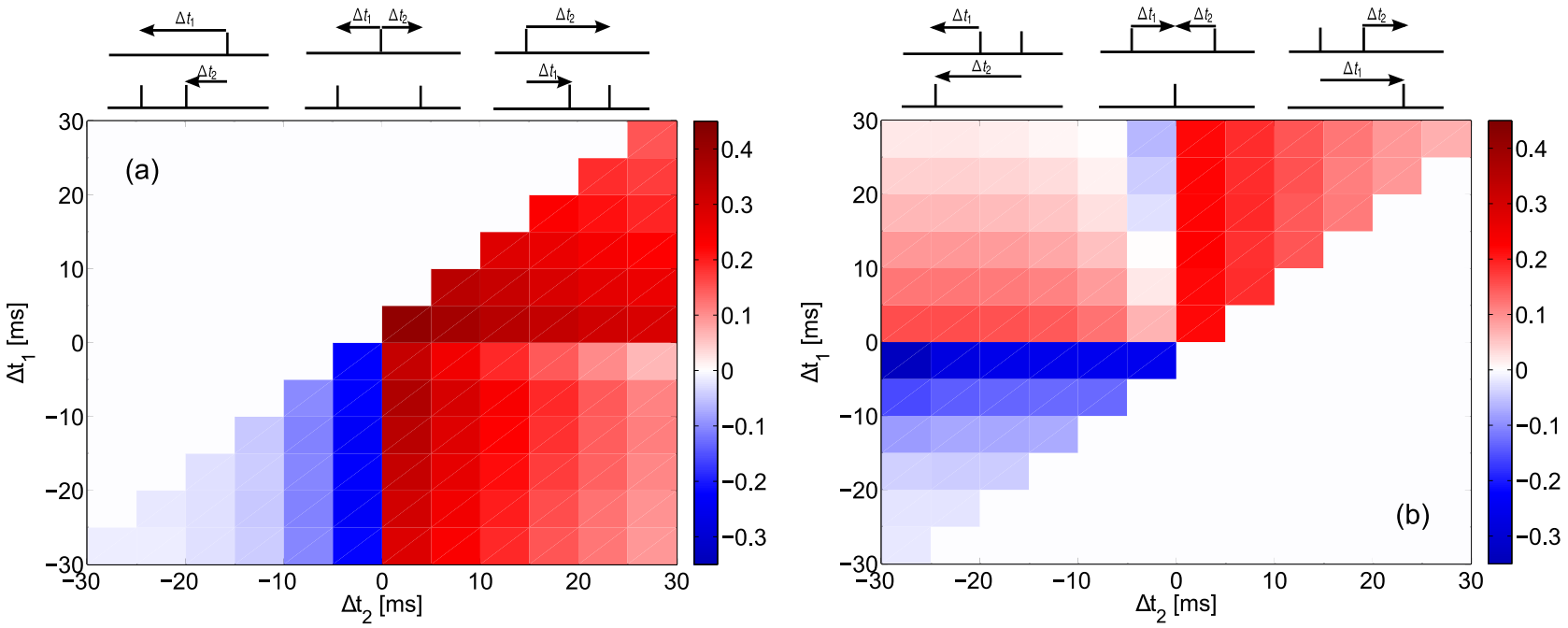


Figure 7.7. Synaptic weight changes in result of the extra triplet protocol and using the proposed minimal TSTDP circuit. (a) Synaptic weight changes in result of the extra triplet protocol for pre-post-post (top right triangle), post-post-pre (bottom left triangle) and post-pre-post (right bottom square) combination of spikes. (b) Synaptic weight changes in result of the extra triplet protocol for pre-post-pre (top left square), pre-pre-post (top right triangle) and post-pre-pre (left bottom triangle) combination of spikes. The required bias currents taken for the triplet circuit correspond to the hippocampal culture data set (Table 7.1).

7.4.6 Poissonian Protocol for the BCM Rate-based Learning

As already mentioned, in addition to the ability of reproducing the synaptic weight changes resulting from the pairing protocol (both window and change in pairing frequency), triplet protocol and quadruplet protocol (which all demonstrate the influence of timing-based variations of inputs on the synaptic weights), the proposed circuit also has the ability to give rise to a rate-based learning rule which mimics the effects of BCM. In order to demonstrate how the proposed circuit can reproduce a BCM-like behaviour, a Poissonian protocol has been used as follows. Under this protocol, the pre-synaptic and post-synaptic spike trains are generated as Poissonian spike trains with firing rate of ρ_{pre} and ρ_{post}, respectively. This is the same protocol that has been used in Pfister and Gerstner (2006) to show how their proposed TSTDP model can show a close mapping to the BCM model. This chapter utilises a similar protocol to stimulate the minimal TSTDP circuit and examines if it is capable of reproducing a similar BCM-like behaviour as in Pfister and Gerstner (2006).

In order to extract BCM-like characteristics, as described by the general BCM model (shown in Eq. 2.6), out of the TSTDP rule, Pfister and Gerstner (2006) used a minimal TSTDP rule by setting $A_3^- = 0$. They specifically observed the statistical nature of the weight changes associated with this rule including the time averaged learning dynamics of the weight changes. Consequently, in order to show that the circuit is capable of reproducing similar BCM-like behaviour, the same protocol as used by Pfister and Gerstner (2006), has been implemented here. Therefore, either I_{dep2} must be set to zero in the full-triplet circuit (Fig. 7.1), or the circuit can be changed to the minimal TSTDP circuit presented in Fig. 7.2. The simulation results for the Poissonian protocol and using the proposed minimal TSTDP circuit are shown in Fig. 7.8.

In this figure, each data point at each post-synaptic frequency, ρ_{post}, demonstrates the average value of weight changes for ten different realisations of post-synaptic and pre-synaptic Poissonian spike trains. In addition, each error bar shows the standard deviation of the weight changes over these ten trials. The demonstrated results were produced using the bias currents which correspond to the visual cortex data set (Table 7.1). In the circuit, $V_{\text{post(n-1)}}$, $V_{\text{post(n)}}$, $\overline{V}_{\text{pre(n)}}$ and $V_{\text{pre(n)}}$ are Poissonian spike trains where ρ_{post}, ρ_{post}, ρ_{pre} and ρ_{pre} denote their average firing rates, respectively. The three different curves presented in Fig. 7.8 display three different weight modification thresholds. In the original BCM rule, these thresholds are related to the post-synaptic

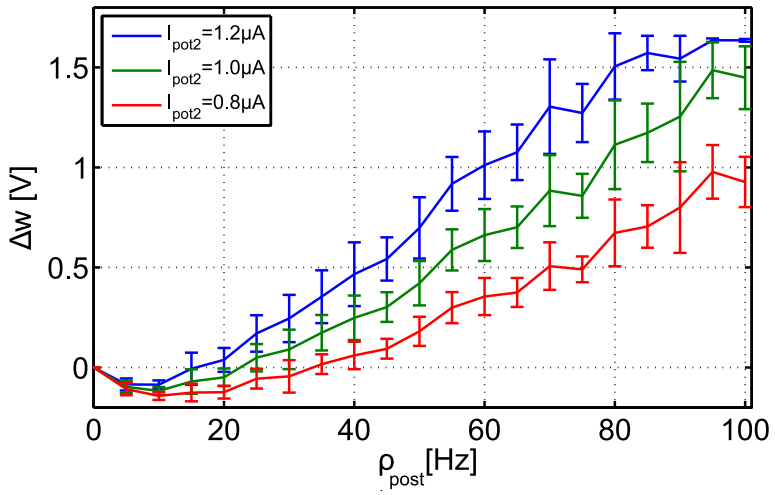


Figure 7.8. The proposed TSTDP circuit can generate BCM-like behaviour. The required bias currents for the circuit correspond to those used for the visual cortex data set (Table 7.1). The three different curves show the synaptic weight changes according to three different synaptic modification thresholds demonstrating the points where LTD changes to LTP, which is controlled by the current I_{pot2}. The threshold is adjustable using the TSTDP rule parameters. In order to move the sliding threshold toward left or right, the I_{pot2} parameter can be altered as depicted in this figure. The rate of pre-synaptic spike trains, ρ_{pre}, is 10 Hz in all experiments. Each data point shows the mean value of the weight changes for 10 different trials and the error bars depict the standard deviations of the weight changes for each value of ρ_{post} for these trials.

firing rate, ρ_{post}. Based on Pfister and Gerstner (2006), the modification threshold for the all-to-all spike interactions can be expressed as

$$\theta = \left\langle \rho_{\text{post}}^{p} \right\rangle \frac{(A_2^- \tau_- A_2^+ \tau_+)}{(\rho_0^p A_3^+ \tau_+ \tau_y)}, \tag{7.6}$$

where $\left\langle \rho_{\text{post}}^{p} \right\rangle$ is the expectation over the statistics of the p^{th} power of the post-synaptic firing rate and $\rho_0^p = \left\langle \rho_{\text{post}}^{p} \right\rangle$ for large time constants (10 min or more). However, for the nearest-spike model which is the case for the proposed TSTDP circuit, it is not possible to derive a closed form expression for the modification threshold based on ρ_{post}^{p}, however for post-synaptic firing rate up to 100 Hz, a similar behaviour to what Eq. 7.6 presents is inferable from the simulation results (supplementary materials of Pfister and Gerstner (2006)). The three different curves in Fig. 7.8 are the results of three different values for I_{pot2} currents which correspond to three different values of

A_3^+. This simulation suggests that the proposed circuit not only can reproduce timing-based experimental outcomes, but also can reproduce some rate-based synaptic weight modifications.

Other examples of post-synaptically driven BCM-like behaviour can be found in Appendix A; for these the circuit simulations were conducted by fixing the pre-synaptic rates to 5 Hz and 15 Hz, respectively and post-synaptic rates varied from 0 to 50 Hz. For both these cases a BCM-like behaviour was observed.

To analyse how BCM-like behaviour emerges from TSTDP, we need to go through the same analysis used by Pfister and Gerstner (2006). In this circumstance, the triplet learning rule can be recast into a simpler form by considering the statistical properties of TSTDP weight changes which leads to the following time averaged equation,

$$\begin{aligned}\left\langle \frac{dw}{dt} \right\rangle &= -A_2^- \tau_- \rho_{\text{pre}} \rho_{\text{post}} + A_2^+ \tau_+ \rho_{\text{pre}} \rho_{\text{post}} \\ &\quad -A_3^- \tau_- \tau_x \rho_{\text{pre}}^2 \rho_{\text{post}} + A_3^+ \tau_+ \tau_y \rho_{\text{post}}^2 \rho_{\text{pre}},\end{aligned} \tag{7.7}$$

where ρ_{pre} and ρ_{post} are the pre- and post-synaptic firing rates, respectively. The other parameters in the above equation τ_-, and τ_+, are time constants for the pair-based contribution and τ_x, and τ_y are the corresponding time constants for the triplet-based contribution of the original triplet learning rule by Pfister and Gerstner (2006).

By considering the mapping of Eq. 7.7 into a mathematically similar functional form of the BCM rule, shown in Eq.2.6—following the method as described in Pfister and Gerstner (2006)—one can simply set $A_3^- = 0$ and for simplicity, keep A_2^- and A_2^+ constant in Eq. 7.7. This gives rise to the following expression

$$\begin{aligned}\left\langle \frac{dw}{dt} \right\rangle &= -A_2^- \tau_- \rho_{\text{pre}} \rho_{\text{post}} + A_2^+ \tau_+ \rho_{\text{pre}} \rho_{\text{post}} \\ &\quad + A_3^+ \tau_+ \tau_y \rho_{\text{post}}^2 \rho_{\text{pre}}.\end{aligned} \tag{7.8}$$

The above equation, given an appropriate choice of parameter values, can mimic BCM-like nonlinear weight change dynamics by keeping ρ_{pre} fixed and altering the value of the ρ_{post}; under these conditions, one can numerically illustrate that the weight changes as a function of increasing post-synaptic frequency, has a similar profile to the weight changes of the original BCM rule as described by Eq. 2.6.

However, one must keep in mind an important aspect of the original BCM experiments (Kirkwood *et al.* 1996, Cooper *et al.* 2004) in order not to introduce any misconceptions about the original BCM rule. This aspect (excluding neuromodulatory

effects) is that the original experiments were conducted using increasing pre-synaptic frequency of inputs (Kirkwood *et al.* 1996). It is a well-known and undisputed fact that neurophysiological experiments have shown that pre-synaptic activity typically drives post-synaptic responses, and changes in post-synaptic firing rate only occurs as a result of changes to input activity. Put simply, changes in post-synaptic firing cannot be considered independent from changes in pre-synaptic activity, they are functionally related. Hence, in a more precise physiological terms, the firing rate of the post-synaptic neuron really needs to be considered as a function of its pre-synaptic inputs. A more informative analysis of the weight dynamics of the triplet rule should take this fact about pre- and post-synaptic firing rate, i.e. $\rho_{\text{post}} = F(\rho_{\text{pre}})$, into account. Hence changing the post-synaptic firing rates should really be driven by changes in pre-synaptic firing rates, as they do in any neurophysiological setting; in this manner one can deduce a more informative link between the plasticity model and the original BCM rule. Changing ρ_{post} while keeping the pre-synaptic firing rate ρ_{pre} fixed, needs to be viewed with caution as it represents a misinterpretation in the application of the original stimulus protocol used in LTD/LTP experiment, despite leading to BCM-like weight changes.

As a check that our circuit can reproduce BCM-like behaviour which is driven by pre-synaptic (rather than post-synaptic) activity, we have repeated our circuit simulations but made the naive assumption that post-synaptic firing rate is a linear function of the pre-synaptic firing rate, i.e. $\rho_{\text{post}} = A\rho_{\text{pre}}$ and for the sake of simplicity we let $A = 1$, i.e $\rho_{\text{post}} = \rho_{\text{pre}}$. Despite such a crude approximation, the circuit successfully was able to mimic BCM-like behaviour where weight changes were pre-synaptically driven, as illustrated in Fig. 7.9. In this figure, each data point shows the mean value of the weight changes for 10 different trials and the error bars depict the standard deviations of the associated weight changes.

Additionally, Matlab simulations were conducted using both the linear Poisson neuron model and the Izhikevich model, in order to assess whether such models can reproduce pre-synaptically driven BCM-like changes to synaptic strength. We found that in the case of increasing the pre-synaptic activity, the resulting synaptic weight changes followed a BCM-like profile where for low pre-synaptic activity, there was no alteration to synaptic weight; for moderate levels of pre-synaptic activity, gave rise to depression (LTD) and for further increases in (pre-synaptic) activity led to potentiation (LTP). Such a pre-synaptically driven BCM-like profile of synaptic change occurs for each

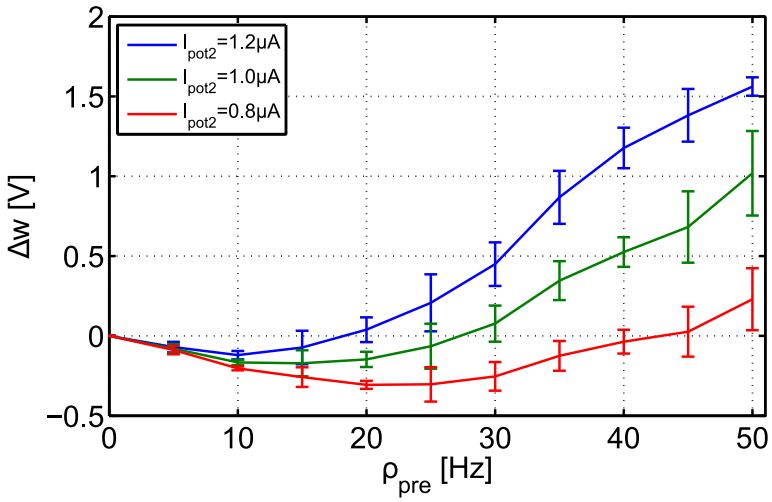


Figure 7.9. The proposed TSTDP circuit can generate pre-synaptically driven BCM-like weight changes. In this simulation, post-synaptic firing rate is a linear function of pre-synaptic firing rate.

above stated neuron model and the results of these simulations are presented in Appendix A. These preliminary Matlab simulations were pursued in order to inform us whether combining a circuit based model of a neuron with our TSTDP circuit will lead to a circuit implementation capable of both timing and rate-based synaptic plasticity changes.

7.5 Mismatch and Variation

Neuromorphic models are an approximation to biological experiments and as we can see from these experiments there is a significant variation associated with them (Bi and Poo 1998, Wang *et al.* 2005, Pfister and Gerstner 2006). Nonetheless, it is of interest to produce circuits that mimic these models, the most important usually being the trend of the circuit behavior. Having said that, as mentioned in Section 5.4.2, the variation and mismatch inherent in the fabrication process of transistors in submicron scales and subthreshold design regime are major concern when designing analog CMOS neuromorphic circuits especially in large-scale. The majority of neuromorphic models are designed in the subthreshold regime to gain the required neuronal behavior and at the same time enjoy less power consumption compared to above threshold operational region. However, the subthreshold regime usually brings about severe transistor threshold voltage variations as well as inevitable transistor mismatch (Poon and Zhou 2011).

In order to minimise the effect of variations and mismatch, the analog VLSI signal processing guidelines proposed by Vittoz (1985) can be adopted. It should be acknowledged that a complete elimination of mismatch and variations is not possible.

The proposed circuit uses a number of transistors operating in the subthreshold region and also includes current mirrors. Therefore, it is expected that this circuit will be susceptible to process variations. In order to show that the proposed design is process tolerant, two types of analysis were performed here. First, the proposed design was simulated using the worst case process corners of the AMS 0.35 μm CMOS model. The simulation results shown in Figs. 7.3 to 7.6 demonstrate that under the process corners, the proposed circuit functions within expectation (reasonable bounds) and can show the expected behaviour in all cases. These figures show that there are slight variations in the amplitudes, which can be adjusted by retuning the circuit's bias currents. This robustness suggests that the physical implementation of the proposed design would be also robust and work within the expected design boundaries—Chapter 4 of Weste and Harris (2005).

Furthermore, since the proposed design utilises current mirrors to copy currents and set the required amplitudes and time constants of the TSTDP model, the effect of transistors mismatch on the circuit performance must be considered. Therefore as the second variation analysis, the proposed circuit was examined against process variation and transistor mismatch. For this purpose, a 1000 Monte Carlo (MC) runs were performed on the proposed circuit in order to test its operational robustness.

The process variation scenario is as follows. All the circuit transistors go under a local variation which changes their absolute parameter values in the typical model. The process parameter that was chosen to go under these variations is the transistor threshold voltage, which is one of the most important process parameters especially in the proposed design consisting of transistors operating in the subthreshold region of operation (Seebacher 2005). The parameters vary according to the AMS C35 MC process statistical parameters. The threshold voltages of PMOS and NMOS transistors varied according to a one sigma Gaussian distribution, which can change the threshold voltages by up to 30 mV. This variation can be much less if larger transistor aspect ratios be used for the design. According to Pelgrom's law (Pelgrom *et al.* 1989), the variation has a relation with transistors aspect ratio. According to the transistor sizing in the proposed circuit, a $\sigma(\Delta V_{\text{th}})$ equal to 10 mV might be faced. Under such a circumstance,

a 1000 MC runs were performed on the proposed circuit for three different cases, as described below.

As the first case of MC analysis, the circuit was stimulated by a pairing protocol to reproduce the exponential STDP learning window in the presence of the mentioned local variation. The circuit bias currents correspond to those used for the typical model and hippocampal data set reported in Table 7.1. Note that Fig. 7.10 shows a 1000 MC analysis performed on the proposed circuit. This figure also shows that the proposed circuit is less susceptible to process variation and the overall LTP/LTD exponential behaviour can be preserved. However, the strength of the proposed circuit is in its controllability using the bias currents. The observed variations in the design can be alleviated by means of readjusting the circuit bias currents. This tuning can be conducted even after the circuit is fabricated. If the fabricated circuit performance changes from the expected characteristics, the circuit bias currents, which serve as inputs to the fabricated chip, can be retuned to reach the required behaviour.

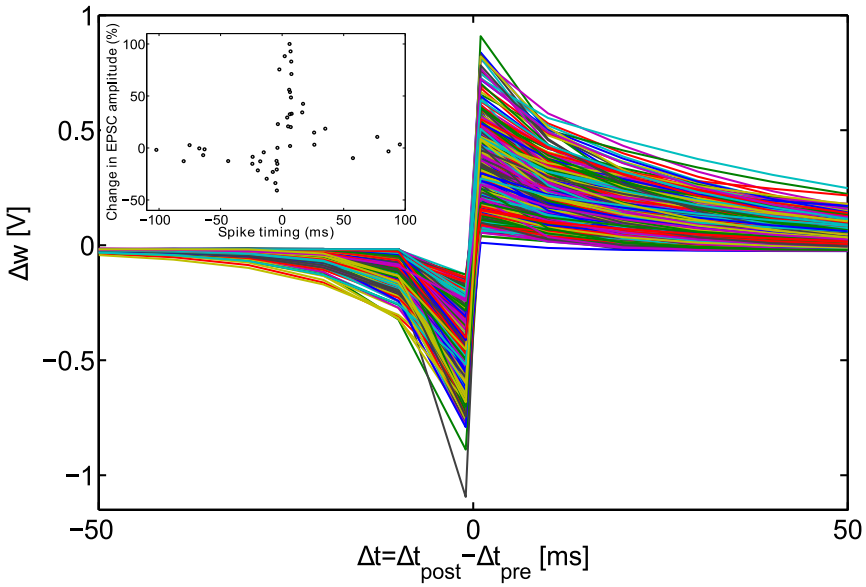


Figure 7.10. STDP learning windows produced in 1000 MC runs using the optimised bias parameters for the hippocampal data set. Each curve represents the weight change obtained from the minimal TSTDP circuit that is stimulated according to the pairing protocol to reproduce the STDP window. In each run, a random variation of transistors threshold voltages occurs. The circuit bias parameters are those used for the typical transistor model for the hippocampal data set, which are reported in Table 7.1. The inset figure shows the STDP experimental data extracted from Bi and Poo (1998). Note the similarity between the simulation results and the experimental data.

The second analysis was performed under similar process variation conditions to the first case, but this time the circuit was stimulated by the frequency-dependent pairing protocol to reproduce the visual cortex data set and the resulting NMSEs were computed for 1000 MC runs. The circuit bias currents correspond to those used for the typical model, which are reported in Table 7.1. The obtained results are shown in Fig. 7.11. Furthermore, as the third case, the same analysis was carried out for the hippocampal data set and bias parameters presented in Table 7.1. Achieved results are demonstrated in Fig. 7.12. Figures 7.11 and 7.12 show significant variations in the value of NMSE compared to the typical transistor parameters that the circuit bias currents (see Table 7.1) were optimised for. Despite these deviations in the NMSE values under process variations, they are easily treatable by retuning the bias currents.

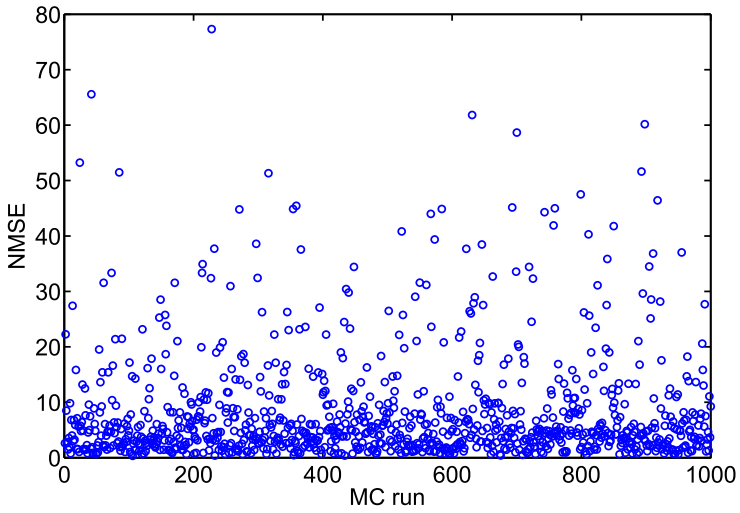


Figure 7.11. NMSEs obtained to reproduce the visual cortex data set in 1000 MC runs, using the optimised bias parameters for this data set. Each run represents a NMSE value obtained from the minimal TSTDP circuit that is stimulated according to the freqneucy-dependent pairing protocol to mimic the visual cortex data set. The circuit bias parameters are those used for the typical transistor model for the visual cortex data set, which are reported in Table 7.1. In each run, a random variation of transistors threshold voltages occurs.

In the case of the visual cortex data set, the worst NMSE was almost 78 that is much larger than a minimal NMSE obtained using the typical model, NMSE $= 0.33$. Also, in the case of hippocampal data set, the worst NMSE is about 306, which is again significantly bigger than the NMSE obtained using the typical model, NMSE $= 1.74$. These major deviations can be significantly reduced by retuning circuit bias currents

and optimising them to get a minimal NMSE, in the presence of process variation. It means that, some bias tuning should be performed on the circuit to reach a minimal NMSE comparable to the design target.

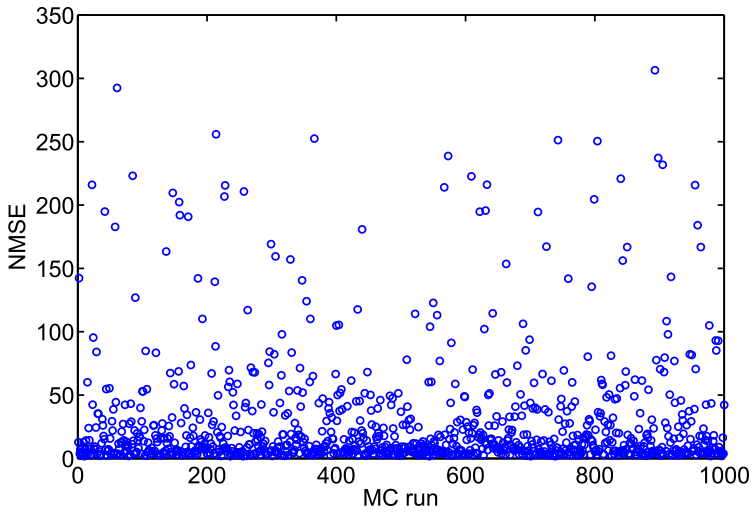


Figure 7.12. NMSEs obtained to reproduce the hippocampal data set in 1000 MC runs, using the optimised bias parameters for this data set. Each run represents a NMSE value obtained from the minimal TSTDP circuit that is stimulated according to the triplet, qudruplet, and pairing protocols to mimic the hippocampal data set. The circuit bias parameters are those used for the typical transistor model for the hippocampal data set, which are reported in Table 7.1. In each run, a random variation of transistors threshold voltages occurs.

As an example, the worst case of NMSE for the hippocampal data set (NMSE $= 306.4$) is in the case of some big changes in the threshold voltages around 30 mV. In the presence of these parameter variations in the design, all circuit bias currents were adjusted again and a new minimum NMSE was obtained. The achieved NMSE, which is equal to 1.92, is consistent with the design expectations. The retuned circuit bias currents in this case are given in Table 7.2. Using these new bias currents, the required behaviour for the pairing experiment (Fig. 7.3), the quadruplet experiment (Fig. 7.6), as well as the triplet experiment (Figures 7.5(a) and (b)) were well observed. The same approach was considered for the visual cortex data set, and the worst NMSE($= 78$) changed to an acceptable NMSE $= 0.47$ that can faithfully represent the required frequency-dependent behaviour in the pairing visual cortex experiment shown in Fig. 7.4.

Both worst case and MC analysis performed on the circuit show the robustness and the controllability of the design in the presence of physical variations. Hence, despite

Table 7.2. Retuned TSTDP circuit bias currents and the resulted NMSEs in the presence of the worst case variation in 1000 MC runs. The table shows the NMSEs in the presence of the worst case variation in 1000 MC runs shown in Figures 7.11 and 7.12. NMSEs were equal to 78 and 306.4 for the visual cortex and the hippocampal data sets, respectively, but they were brought back to the shown NMSEs by readjusting the circuit bias currents from the values shown in Table 7.1.

Data set	I_{pot1}	I_{dep1}	I_{tp1}	I_{td1}	I_{pot2}	I_{tp2}	NMSE
Visual cortex	0	260 nA	1 nA	120 pA	590 nA	150 pA	0.47
Hippocampal	510 nA	240 nA	270 pA	860 pA	110 nA	180 pA	1.92

the fact that the proposed design has some susceptibility to process variations, a post-fabrication calibration is possible through retuning the bias currents of the design to achieve a minimal NMSE to faithfully reproduce the needed learning behaviour.

7.6 VLSI Implementation of the Proposed TSTDP Circuit and Chip Measurement Results

The proposed TSTDP circuit was designed and integrated on a multi-neuron chip, named MN256R1, fabricated in collaboration with the Neuromorphic Cognitive Systems (NCS) group in the Institute of Neuroinformatics (INI), Uni/ETH Zurich, Switzerland. Figure 7.13(a) shows the chip and its bounding wires under the microscope. The test block of the TSTDP circuit is shown in the bottom of the chip. In addition, Fig. 7.13(b) demonstrate the TSTDP circuit block under the microscope. Note to the five large capacitors sitting besides each other to form the required biological time constants as well as to restore the synaptic weight for longer times. Unlike the circuit simulations shown earlier in this chapter that utilised accelerated time to reproduce the TSTDP behaviour, the proof circuit is fabricated to realise biologically plausible time constants in the order of tens of ms. In addition, the updated synaptic weight value must be retained to perform various synaptic plasticity experiments in the system. Therefore, large capacitors were employed in the chip to ensure the required long time constants, as well as keeping the updated synaptic weight changes for required time. The issue of large capacitors in neuromorphic design to obtain large time constants was already discussed in Section 5.4.8.

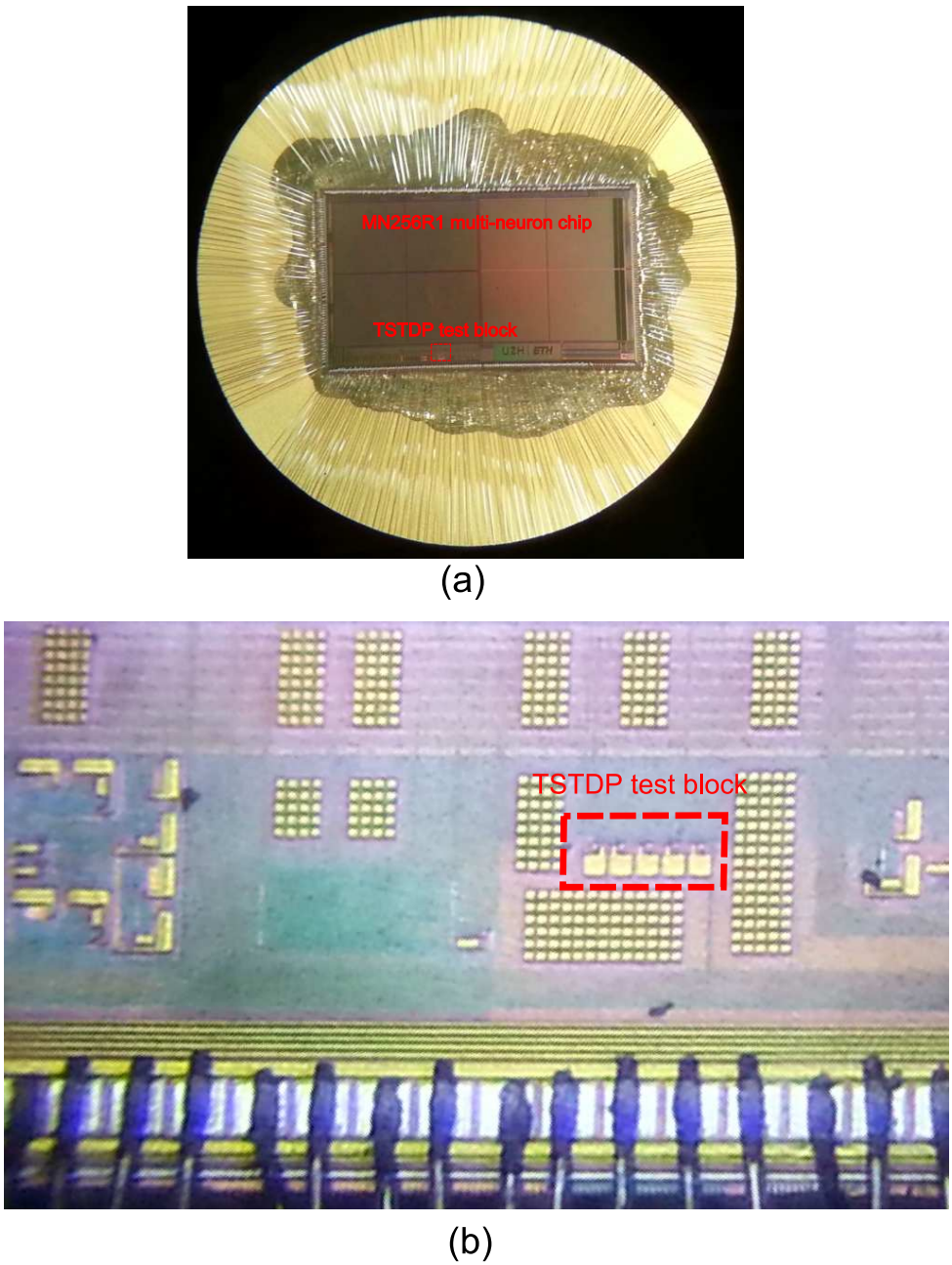


Figure 7.13. The MN256R1 multi-neuron chip under the microscope. (a) This figure shows the complete chip including neurons and plastic synapses. The TSTDP circuit block is shown inside a red dashed box. (b) This photo micrograph shows further detail of the TSTDP circuit block and its five large capacitors. Note that the relevant circuit is inside the area surrounded by the red dashed box.

The chip was designed and fabricated using a 1-poly 6-metal 0.18 μm AMS CMOS process. The utilised design kit was Hit-Kit 4.01, and the supply voltage is 1.8 V. The layout of the implemented TSTDP circuit is shown in Fig. 7.14. The required silicon area for the implemented circuit is 165 μm ×60 μm, from which almost 75 percent is

occupied by the five capacitors. In addition to the four time constants capacitors i.e. C_{pot1}, C_{dep1}, C_{pot2} and C_{dep2}, as shown in Fig. 7.1, another capacitor is the C_w capacitor, which has the same size as the time constant capacitors.

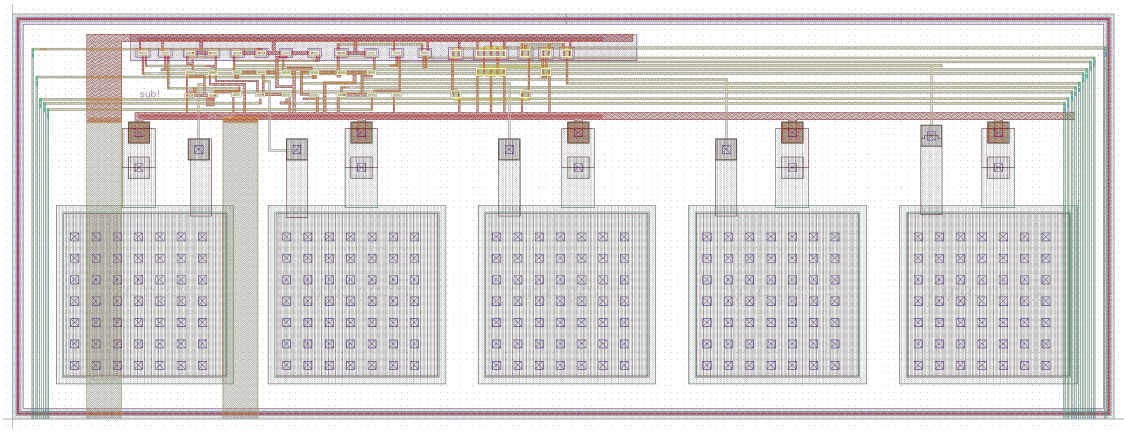


Figure 7.14. The layout of the TSTDP circuit implemented on the MN256R1 chip. The layout shows various parts of the TSTDP circuit, including five large capacitors with the size of 500 fF in order to reach biologically plausible time constants in the order of ms, in contrary to the circuit simulations that were performed in an accelerated time scale and therefore four of the capacitors related to the TSTDP time constants were smaller and only the weight capacitor was large for retaining the synaptic weight for longer time. The occupied area by the implemented TSTDP circuit is $165\ \mu m \times 60\ \mu m$.

Performed chip measurement results show that the implemented TSTDP device functions as expected and correctly follows the weight update pattern of the TSTDP rule. As an example, Fig. 7.15 depicts measurement results from the TSTDP circuit, while it is stimulated with four required signals, namely pre, post, pre(n-1) and post(n-1). These signals are generated as pulses with 1 ms width. Each signal is repeated with a delay of 20 ms for a period of time. The time difference among pre and post-synaptic spikes is set to be 10 ms. This time difference can be both increased and decreased, which therefore results in different synaptic weight changes amplitudes, which is shown as the fourth signal in Fig. 7.15. The figure shows that the synaptic weight, W, is first depressed when a pre-synaptic spike arrives. This is due to a pre-post-pre (the first pre in this pattern is not shown in the figure) spike triplet, which results in depression, as expected. Next the synaptic weight is increased (potentiated), which is in result of a post-pre-post (the first pre in this pattern is not shown in the figure) spike combination, as expected. Finally, the synaptic weight is shown to be depressed again at the arrival of the second pre signal shown in the figure. This depression is in result of the pre-post-pre triplet, which is shown in the figure. Note that,

the circuit also needs a delayed version of pre- and post-synaptic spikes, pre(n-1) and post(n-1), to function correctly. The figure only shows the post(n-1) signal, which is a duplicate of post spike train, with 1 ms delay. Note that the circuit also needs to receive pre(n-1), which is a delayed version of pre spike train with 1 ms delay, to function according to the TSTDP equation (see Eq. 7.1).

Figure 7.15 shows stronger potentiation than depression. As already discussed in Section 7.2, in the TSTDP rule, the magnitude of potentiation and depression are determined by eight synaptic parameters including pair- and triplet-based potentiation and depression time constants and magnitude parameters. The same set of parameters are available in the implemented circuit as eight input analog biases that are set by a programmable bias generator device. In the shown measurement, the circuit is set in a way to show stronger potentiation than depression.

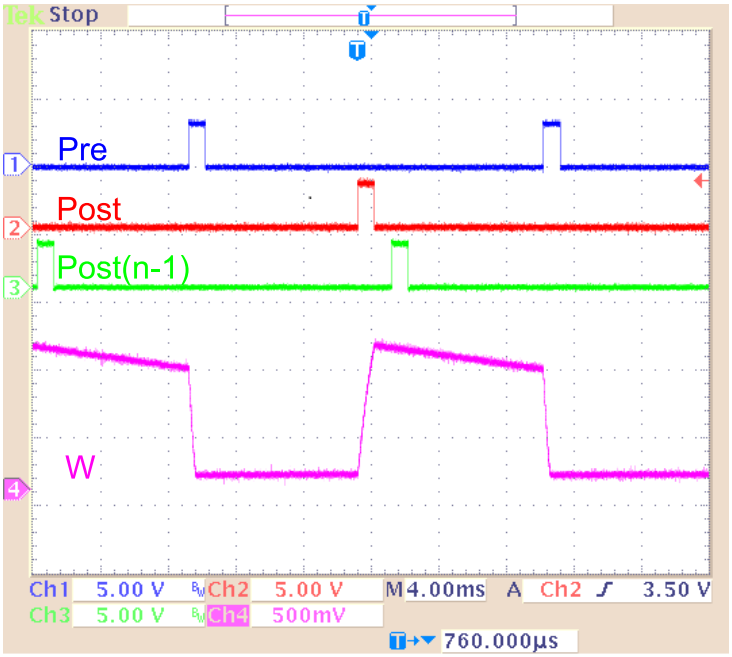


Figure 7.15. Measurement results of the fabricated TSTDP circuit. This figure demonstrtaes synaptic weight changes due to pre-post-pre, post-pre-post and pre-post-pre spike combinations from left to right. The synaptic weight signal, W, shows the voltage changes across the weight capacitor. As expected, a pre-post-pre combination of spikes results in depression, while a post-pre-post spike triplet results in potentiation. The magnitude of weight change at the time of each spike is controlled by pair and triplet potentiation and depression time constants and amplitude parameters (see circuit description in Section 7.3 and equation 7.1) that are applied to the circuit as 8 different analog biases.

The MN256R1 is composed of 256 low power IF neurons. The implemented TSTDP learning circuit can be interfaced to 256 low power IF neurons on the MN256R1 chip, using the available AER system already discussed in Sections 3.2 and 3.3. The AER system is also useful to generate the required pre(n-1) and post(n-1) spike trains which are delayed versions of pre- and post-synaptic spike trains.

7.7 Discussion

As it is shown in this chapter, the proposed VLSI design for the TSTDP learning rule produces a very close fit to the experimental data. Having a minimal fitting error has been the primary goal of the proposed design. Although this goal has been successfully satisfied, considering the challenges mentioned in Section 5.4, this design faces some challenges when integrated in a large-scale SNN, to be used in a learning or computation application. One of the first challenges that is generic to almost all VLSI implementations of synaptic plasticity models is the silicon real estate used by the design and its weight storage technique. In the proposed design, a large capacitor has been utilised to store the synaptic weights for long period of times in the orders of hundreds of ms. However, in a large-scale neuromorphic system having such large capacitors for storing the synaptic weight is not practicable. A number of methods to tackle the weight storage technique problem was discussed in Section 5.4.8.

A counter approach that is appropriate in the proposed design is to use a smaller capacitor along with a bistability circuit (Indiveri *et al.* 2006). However, this technique has its own problem, as it confines the synaptic weights to two final steady states and therefore, the analog nature of the weight is compromised. In addition, integrating the TSTDP circuit with a memristive memory element is also another option to save silicon area (Azghadi *et al.* 2013d). This approach has its own problems such as high variability though. Besides, the use of compact CMOS memory elements to store the synaptic weight (Seo *et al.* 2011) is another promising solution for storing the synaptic weight. Either of the mentioned methods are useful to decrease the silicon are required for weight storage, however none of them decrease the large area required for the TSTDP circuit itself that contains 37 transistors, and four capacitors. In the next chapter, a new design will be proposed that minimise the number of transistors used to realise the TSTDP rule, and decrease the size of the weight capacitor by a factor of 10, while loosing just a little of synaptic weight change prediction ability.

In addition to the weight storage, another challenge the proposed design faces to be used in large-scale systems is its susceptibility to process variation, similar to all analog VLSI designs in micro/nano meter design regime. Although the presented MC simulation analysis and the result from fine-tuned circuits (presented in Section 7.5) show that it is possible to minimise the effect of process variations, fine tuning of the circuit bias parameters in a large spiking neural network of neurons and proposed triplet synapses sounds to be not practical. Hence, a variation aware design technique is required that take into account the process variation while designing the desired neuronal circuit. The technique utilised to alleviate variations and mismatch in our design is using the rules of transistor matching proposed by Vittoz (1985). However, in the worst case variation condition, this approach is not adequate.

An interesting approach available in the literature that is useful for the proposed TSTDP circuit when it is integrated in a large-scale neuromorphic system is the use of already available AER framework in the system to fine-tune the circuit. This approach has been successfully utilised in previous works for both small-scale (Neftci and Indiveri 2010) and large-scale SNNs (Choudhary *et al.* 2012).

Another approach available in the literature, which is not useful for our proposed circuit design, is the design technique proposed and well utilised by Rachmuth *et al.* (2011) and Meng *et al.* (2011) in neuromorphic modelling of ion channel and ionic dynamics. This variation aware design technique exploits source degeneration and negative feedback methods to increase the dynamic range of input voltages of transistors and make them robust against mismatch errors that happen mainly because of the low input voltage dynamic range in traditional subthreshold current mode circuits (Poon and Zhou 2011). The exploitation of the source degeneration and negative feedback design techniques can be another alternative to build a network of neurons with TSTDP synapses, which are not susceptible to process variations.

Apart from the above mentioned challenges, power consumption of the proposed design is another essential feature of a neuromorphic learning circuit, which should be considered. The simulation results show that the proposed design consumes almost 60 pJ energy per spike, which is rather high comparing to some other synaptic plasticity circuit implementations as presented in Table 5.1. Note that in a large-scale neuromorphic system that contains billions of synaptic plasticity circuit, saving a little of energy per circuit is critical. Next chapter proposes a new TSTDP design that not only

is more compact, but also is more energy efficient than the current design. This becomes possible with the cost of losing a bit of weight modification accuracy.

7.8 Chapter Summary

A high performance synaptic plasticity circuit for the TSTDP rule was proposed in this chapter. The presented results show that this new circuit is able to very closely fit many biological experiments including rate-based (Bienenstock *et al.* 1982, Pfister and Gerstner 2006), timing-based (Bi and Poo 1998, Wang *et al.* 2005), and hybrid rate and timing-based experiments (Sjöström *et al.* 2008). It was also shown that the circuit is able to closely match the experiments for the extra triplet patterns of a suppressive STDP model presented in Froemke and Dan (2002). Furthermore, the rate-based behaviour of the proposed timing-based circuit was examined under various experimental protocols to stimulate the circuit both pre- and post-synaptically to mimic the results presented in computational model experiments—post-synaptically driven—(Izhikevich 2003, Pfister and Gerstner 2006) as well as the pre-synaptically driven biological experiments (Kirkwood *et al.* 1996).

In addition to its biological accuracy and plausibility, the circuit was investigated in terms of process variation. As reported in the chapter, although the inherent analog VLSI process variation may result in incorrect circuit operation, it is possible to retune the circuit to perform its expected function. However, designing a circuit with the same capabilities, but less prone to variation will be of high interest, since in this case the functionality is not severely affected by the variations, so the circuit may work as expected without the need for retuning.

Besides, the investigation on the power consumption of the circuit shows that, despite having a high performance when reproducing the biological experiments, it consumes relatively high energy, compared to some of the low energy designs available in the literature (Cruz-Albrecht *et al.* 2012). Besides, it was discussed that this design that utilises a large capacitor for storing the synaptic weight occupies large silicon area, a problem that is generic to most of the synaptic plasticity learning circuits (Bofill-I-Petit and Murray 2004, Indiveri *et al.* 2006, Koickal *et al.* 2007). Therefore, there is a need for a new design that not only reproduces all the mentioned biological experiments with an acceptable accuracy, but also is very low energy and compact.

In the next chapter, a novel ultra low-energy, and compact design is presented that improves the previous designs in all essential aspects, but in result its synaptic weight modification ability is slightly compromised.

Chapter 8

Low Energy and Compact Neuromorphic Circuit for Spike-based Synaptic Plasticity

THIS chapter introduces a new accelerated-time circuit that has several advantages over its previous neuromorphic counterparts, which were discussed in previous chapters, in terms of compactness, power consumption, and capability to mimic the outcomes of biological experiments. The proposed circuit is investigated and compared to other designs in terms of tolerance to mismatch and process variation. Monte Carlo (MC) simulation results show that the proposed design is much more stable than its previous counterparts in terms of vulnerability to transistor mismatch, which is a significant challenge in analog neuromorphic design. All these features make the proposed circuit an ideal device for use in large scale spiking neural networks, which aim at implementing neuromorphic systems with an inherent capability that can adapt to a continuously changing environment, thus leading to systems with significant learning and computational abilities.

8.1 Introduction

The brain processes large amounts of data in real-time in the presence of noise, while consuming little power. It also takes little space and has extraordinary processing features. The ultimate goal for neuromorphic engineers is to develop a cybernetic system, which closely mimics the capabilities of the brain. To reach this goal, as already mentioned, understanding and implementing *in silico* the main components of cortical networks, i.e. neurons and synapses, is a crucial first step.

Currently, the dynamical behaviour of biological neurons is best understood through biophysically detailed models, such as the Hodgkin-Huxley (HH) model (Hodgkin and Huxley 1952), which given the correct parameters, can replicate various experimentally observed response properties. Using such models one can develop hypotheses about cortical circuit behaviour and any underlying computations taking place. The complexity of such biophysical models can be a prohibitive bottleneck when translation into silicon is desired. For this reason simpler models, such as the Integrate-and-Fire (IF) (Izhikevich 2004, Indiveri *et al.* 2011), have been adopted in simulating networks, even though they lack the dynamic realism of real cortical circuits.

Identical to neuron models, there are a variety of synaptic plasticity models. Some of these models embrace certain features of real biological synapses, however they tend to be complex in their (mathematical) formulation. On the other hand, other models have been mathematically formulated to replicate the outcomes of a subset of known experiments. Their representation is typically simpler in form allowing, in some cases, reduced problematic translation into silicon. Generally, the main purpose of such simplified rules is to mimic, as accurately as possible, the outcomes of various experimental synaptic plasticity protocols (Mayr and Partzsch 2010, Morrison *et al.* 2008).

In this chapter, a new VLSI implementation of a malleable synaptic circuit that is capable of mimicking the outcomes of various synaptic plasticity experiments, is proposed. It is demonstrated that the new design has a compact structure and possesses low power features, which are required for VLSI implementations of large-scale spiking neural networks. In addition, the robustness of the proposed circuit is verified against transistor mismatch and process variations. The results show that the new circuit is a reliable design in terms of transistor mismatch. These features make this new design an ideal learning component that may benefit various VLSI synaptic plasticity systems. The proposed circuit is of potential interest for future large scale neuromorphic circuits

with significantly high numbers of neurons and synapses, where low power consumption, compactness, accuracy and mismatch tolerance are absolutely essential.

The results shown in this chapter are presented mainly in *PLOS ONE* (Azghadi *et al.* 2014a), as well as in *The 2013 IFIP/IEEE International VLSI/SOC conference* (Azghadi *et al.* 2013c), and in *The 20th IEEE International Conference on Electronics, Circuits and Systems* (Azghadi *et al.* 2013b).

8.2 Minimal Representation of Triplet STDP Model

As discussed and shown in Section 2.6.1, the weight changes in the triplet-based STDP (TSTDP) model of synaptic plasticity occur according to the timing differences among triplet of spikes in contrary to the pair-based STDP, which alters the synaptic weight based on the timing differences between pairs of spikes. The TSTDP rule is described by

$$\Delta w = \begin{cases} \Delta w^{+} = A_2^{+} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} + A_3^{+} \mathrm{e}^{(\frac{-\Delta t_2}{\tau_y})} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} \\ \Delta w^{-} = -A_2^{-} \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})} - A_3^{-} \mathrm{e}^{(\frac{-\Delta t_3}{\tau_x})} \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})}, \end{cases} \tag{8.1}$$

where the synaptic weight can be decreased (depressed) if a pre-synaptic spike occurs, or can be increased (potentiated) at the time when a post-synaptic spike arrives. Here, A_2^{+}, A_3^{+} and A_2^{-}, A_3^{-} are the potentiation and depression amplitude parameters, respectively. In addition, $\Delta t_1 = t_{\mathrm{post(n)}} - t_{\mathrm{pre(n)}}$, $\Delta t_2 = t_{\mathrm{post(n)}} - t_{\mathrm{post(n-1)}} - \epsilon$ and $\Delta t_3 = t_{\mathrm{pre(n)}} - t_{\mathrm{pre(n-1)}} - \epsilon$, are the time differences between combinations of pre- and post-synaptic spikes, while ϵ is a small positive constant, which ensures that the weight update uses the correct values occurring just before the pre or post-synaptic spike of interest. In Eq. 8.1, τ_- and τ_x are depression time constants, while τ_+ and τ_y are potentiation time constants (Pfister and Gerstner 2006).

In addition, it was pointed out that since the TSTDP rule utilises higher order temporal patterns of spikes, it is shown to be able to account for the outcomes of several experimental protocols including the frequency-dependent pairing experiments performed in the visual cortex (Sjöström *et al.* 2001), or triplet, and quadruplet spike experiments performed in the hippocampus (Wang *et al.* 2005). Note that, the PSTDP rule fails to reproduce the outcomes of these experiments (see Chapter 6). This is due to a linear summation of the effect of potentiation and depression in the PSTDP rule, while the underlying potentiation and depression contributions in the TSTDP rule, do not sum linearly (Froemke and Dan 2002).

Numerical simulation results presented in Section 2.6.1 demonstrate how a minimised version of the full TSTDP rule, which is shown in Eq. 8.1, can approximate a number of biological experiments performed in hippocampal including quadruplet, triplet and STDP window experiments outcomes. This minimised TSTDP rule is presented as

$$\Delta w = \begin{cases} \Delta w^{+} = A_2^{+} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} + A_3^{+} \mathrm{e}^{(\frac{-\Delta t_2}{\tau_y})} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} \\ \Delta w^{-} = -A_2^{-} \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})}. \end{cases} \tag{8.2}$$

This model is able to account for quadruplet, triplet, and pairing (window) experiments as shown in Pfister and Gerstner (2006) and Azghadi *et al.* (2013a). In addition to the capability of simultaneously approximation of triplet, quadruplet and STDP window experiments with the same set of synaptic parameters, another minimal version of TSTDP rule, is also capable of reproducing the results of the frequency-dependent pairing experiments performed in the visual cortex (Sjöström *et al.* 2001). The minimal model for this experiment can be shown as

$$\Delta w = \begin{cases} \Delta w^{+} = A_3^{+} \mathrm{e}^{(\frac{-\Delta t_2}{\tau_y})} \mathrm{e}^{(\frac{-\Delta t_1}{\tau_+})} \\ \Delta w^{-} = -A_2^{-} \mathrm{e}^{(\frac{\Delta t_1}{\tau_-})}, \end{cases} \tag{8.3}$$

which is simpler and utilises a lower number of synaptic parameters, and therefore needs a new set of parameters, in comparison with the previous minimal model for hippocampal experiments.

Besides the ability of reproducing timing-based experiments, the TSTDP rule has the capability to demonstrate BCM-like behaviour. The BCM learning rule is an experimentally verified (Dudek and Bear 1992, Wang and Wagner 1999) spike rate-based synaptic plasticity rule, proposed in Bienenstock *et al.* (1982). Unlike STDP, which is a spike-timing based learning rule, synaptic modifications resulting from the BCM rule depends on the rate (activity) of the pre- and post-synaptic spikes (Bienenstock *et al.* 1982).

In the following section, a novel VLSI design for TSTDP rule is proposed that have fewer number of transistors, smaller area, and lower power consumption, than all previously published circuits, yet with all their synaptic capabilities. These features make this design an ideal learning component for large-scale neuromorphic circuits. It is shown that the proposed circuit is able to faithfully reproduce the outcomes of many biological experiments, when examined under experimental protocols mentioned in Section 2.5.

8.3 Proposed Low Energy and Compact STDP Circuit

The proposed design in this chapter is implemented based on a different arrangement of the TSTDP rule presented in Eq. 2.3. This new arrangement is given by

$$\Delta w = \begin{cases} \Delta w^{+} = \mathrm{e}^{\left(\frac{-\Delta t_1}{\tau_+}\right)}\left(A_2^{+} + A_3^{+}\mathrm{e}^{\left(\frac{-\Delta t_2}{\tau_y}\right)}\right) \\ \Delta w^{-} = -\mathrm{e}^{\left(\frac{\Delta t_1}{\tau_-}\right)}\left(A_2^{-} + A_3^{-}\mathrm{e}^{\left(\frac{-\Delta t_3}{\tau_x}\right)}\right). \end{cases} \tag{8.4}$$

The new TSTDP circuit is demonstrated in Fig. 8.1. This symmetric circuit operates as follows: When a pre-synaptic spike, $V_{\mathrm{pre(n)}}$, is received at the gate of M6 at $t_{\mathrm{pre(n)}}$, V_{pot1} reaches ground resulting in switching on M8, and then starts to increase linearly toward V_{dd}. The rate of this increase is determined by V_{tp1} that is applied to the gate of M5, and corresponds to the pairing potentiation time constants, τ_+, which is present in both pairing and triplet potentiation terms as shown in the first line of Eq. 8.1.

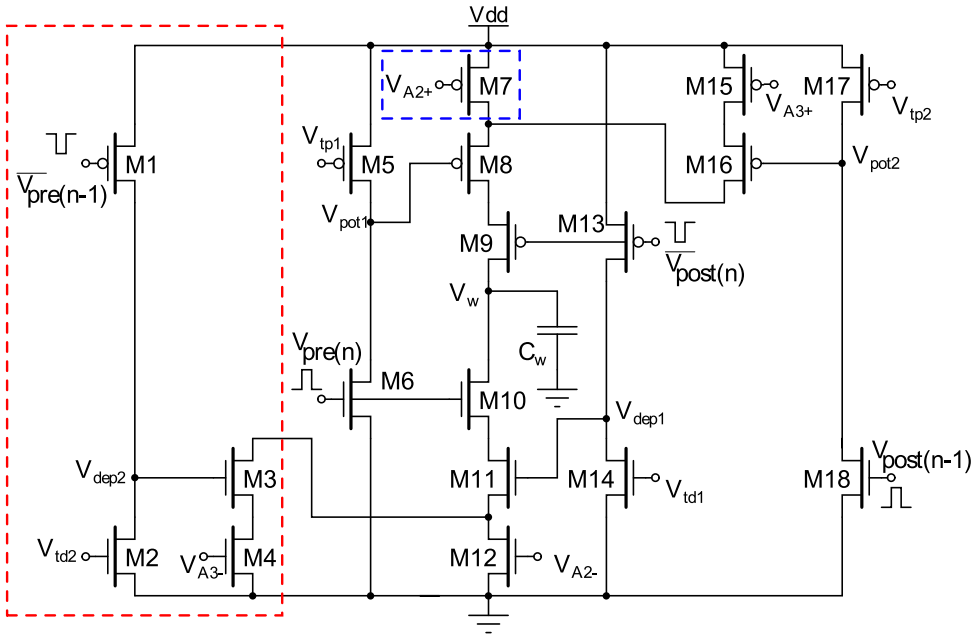


Figure 8.1. Proposed circuit for the full TSTDP rule. This circuit corresponds to the modified full TSTDP rule shown in Eq. 8.4. The minimal circuit that corresponds to the first minimal TSTDP model shown in Eq. 8.2 does not include transistors M1-M4 shown in the red dashed-box. Furthermore, the minimal TSTDP circuit that corresponds to the second minimal TSTDP model shown in Eq. 8.3, does not include the M1-M4 transistors, nor the M7 transistor, shown in the blue dashed-box.

In fact, V_{pot1} is a triangular voltage, which is controlled by the leaky integrator composed of the output conductance of M5 and the gate capacitor of M8, to control the

existence of the potentiation in the first place and allows a current, I_{pot}, to flow through the potentiation branches (M7-M9 and/or M15-M16-M8-M9) at the time of arrival of a post-synaptic spike at M9, $t_{post(n)}$.

The linear increase of V_{pot1}, which starts at $t_{pre(n)}$, and leads to charging the weight capacitor through M8 once $t_{post(n)}$ arrives, is approximately proportional to

$$e^{(\frac{-\Delta t_1}{\tau_+})},$$

where $\Delta t_1 = t_{post(n)} - t_{pre(n)}$ and τ_+ approximates by V_{tp1}. This term is repeated twice in the first line of Eq. 8.1, and can be factorised as it is shown in the first line of Eq. 8.4.

Furthermore, the addition term shown in the second term of first line of Eq. 8.4 that determines the amount of potentiation as a result of both pair and triplet interactions, is approximated through a sum of two currents that charge the weight capacitor, C_w, and represent synaptic weight potentiation. The first current is controlled by the controllable voltage $V_{A_2^+}$, while the second one is determined by both the second potentiation dynamic V_{pot2}, as well as the controllable voltage $V_{A_3^+}$. When a post-synaptic spike arrives at M18, V_{pot2} reaches ground and after the post-synaptic pulse duration is finished, it starts to increase linearly toward V_{dd}. The rate of this increase is determined by V_{tp2} that is applied to the gate of M17, and corresponds to the triplet potentiation time constants, τ_y. Therefore, the current flowing through M15-M16 can be an approximation of

$$A_3^+ e^{(\frac{-\Delta t_2}{\tau_y})},$$

where $\Delta t_2 = t_{post(n)} - t_{post(n-1)}$. The current flowing through M15-M16 transistors accumulates with the current flowing through M7 transistor (which is controlled by gate voltage $V_{A_2^+}$) and forms the total current that is approximately proportional to

$$A_2^+ + A_3^+ e^{(\frac{-\Delta t_2}{\tau_y})},$$

and it represents an approximation of the second term of the first line of Eq. 8.4.

The same dynamic operates in the depression half of the proposed circuit, in which currents flow away from the weight capacitor, C_w, and represent synaptic weight depression. In this part, current sinks away from the weight capacitor through M10-M12, if there has been a pre-synaptic action potential that arrives at M10, in a specified time window defined by V_{td1} (which corresponds to τ_-), after a post-synaptic

spike arrives at M13. The amount of this current is first determined by the time difference between the pre- and post-synaptic spikes (Δt_1) and then by the controllable voltage, $V_{A_2^-}$. Therefore, this current approximates

$$A_2^- e^{(\frac{\Delta t_1}{\tau_-})},$$

where $\Delta t_1 = t_{post(n)} - t_{pre(n)}$. This is the pairing depression current that flows away from the weight capacitor and results in depression due to post-pre spike pairs.

In addition, another current that can discharge the capacitor and results in depression, will flow through M10-M11-M3-M4, if two conditions are satisfied. First, if there has been a previous pre-synaptic spike, $V_{pre(n-1)}$, in a specified time window, set by V_{td2} (which corresponds to τ_x), before the current pre-synaptic spike, $V_{pre(n)}$, arrives at M10 gate. And second, if a post-synaptic spike arrived at M13 gate in a specified time window set by V_{td1} before the current and after the previous pre-synaptic spikes. The magnitude of this current is first controlled by the time difference between the pre- and post-synaptic spikes (Δt_1), second with the time difference between the $V_{pre(n)}$ and $V_{pre(n-1)}$ spikes, (Δt_3), and then by controllable voltage, $V_{A_3^-}$. Therefore, this current approximates

$$A_3^- e^{\frac{\Delta t_1}{\tau_-}} e^{\frac{-\Delta t_3}{\tau_x}},$$

where $\Delta t_1 = t_{post(n)} - t_{pre(n)}$ and $\Delta t_3 = t_{pre(n)} - t_{pre(n-1)}$. This is the triplet depression current that flows away from the weight capacitor and results in depression due to pre-post-pre spike triplet.

If the above two currents accumulate together, they form the depression term of both Eqs. 8.1 and 8.4 that are equal as follows,

$$-A_2^- e^{(\frac{\Delta t_1}{\tau_-})} - A_3^- e^{\frac{\Delta t_1}{\tau_-}} e^{\frac{-\Delta t_3}{\tau_x}} = -e^{(\frac{\Delta t_1}{\tau_-})}(A_2^- + A_3^- e^{\frac{-\Delta t_3}{\tau_x}}), \tag{8.5}$$

where the negative sign indicates that the current is depressive and that it flows away from the weight capacitor.

Note that the above explanations contain assumptions that approximate the TSTDP rule using our proposed circuit. However, from a circuit analysis point of view, if M3-M4, M7-M12, and M15-M16 operate in the subthreshold regime (Liu *et al.* 2002), the

analytical expressions for I_{pot} and I_{dep}, which are potentiation and depression currents, respectively, can written as follows

$$I_{\text{pot}}(t) = \frac{I_0}{e^{-(\kappa/U_T)V_{\text{pot1}}(t-t_{\text{pre(n)}})}+e^{-(\kappa/U_T)V_{A_{2+}}}} + \frac{I_0}{e^{-(\kappa/U_T)V_{\text{pot1}}(t-t_{\text{pre(n)}})}+e^{-(\kappa/U_T)V_{\text{pot2}}(t-t_{\text{post(n-1)}})}+e^{-(\kappa/U_T)V_{A_{3+}}}}, \tag{8.6}$$

$$I_{\text{dep}}(t) = \frac{I_0}{e^{-(\kappa/U_T)V_{\text{dep1}}(t-t_{\text{post(n)}})}+e^{-(\kappa/U_T)V_{A_{2-}}}} + \frac{I_0}{e^{-(\kappa/U_T)V_{\text{dep1}}(t-t_{\text{post(n)}})}+e^{-(\kappa/U_T)V_{\text{dep2}}(t-t_{\text{pre(n-1)}})}+e^{-(\kappa/U_T)V_{A_{3-}}}} \tag{8.7}$$

where $t_{\text{pre(n)}}$ and $t_{\text{post(n)}}$ are current pre- and post-synaptic spike times respectively, while $t_{\text{pre(n-1)}}$ and $t_{\text{post(n-1)}}$ are the times at which the previous pre- and post-synaptic spikes have arrived. Therefore, the voltage change in synaptic weight, shown as V_w in Fig. 8.1, is approximated as:

$$\Delta V_w = \begin{cases} \Delta V_w^+ = \left(\frac{I_{\text{pot}}(t_{\text{post(n)}})}{C_{\text{pot1}}}\right)\Delta t_{\text{spk}} \\ \Delta V_w^- = \left(\frac{I_{\text{dep}}(t_{\text{pre(n)}})}{C_{\text{dep1}}}\right)\Delta t_{\text{spk}} \end{cases} \tag{8.8}$$

where Δt_{spk} are the width of pre- and post-synaptic spike pulses, and C_{pot1} and C_{dep1} are the parasitic capacitance at the gate of M8, and M11, respectively. Please note that, in the proposed circuit, similar to the TSTDP model, whenever a pre-synaptic spike arrives at $t_{\text{pre(n)}}$, a depression can happen, while a potentiation can happen whenever a post-synaptic spike arrives.

Below, experimental results of the proposed circuit are presented. In addition, the circuit is compared with previous synaptic plasticity circuits in terms of power consumption, area and ability in reproducing the outcomes of various biological experiments.

8.4 Experimental Results

8.4.1 Experimental Setup

This section provides information about the experimental setup, under which simulations are performed. These simulations are carried out in order to verify the performance of the proposed circuit and compare its performance with published synaptic plasticity circuits in the literature.

Minimal TSTDP Circuits

As already discussed, in order to regenerate the outcomes of several biological experiments, minimal models of the TSTDP rule, shown in Eqs. 8.2 and 8.3 are sufficient. Matlab simulation results of the first minimal model, presented in Pfister and Gerstner (2006) demonstrate that the first minimal TSTDP model, shown in Eq. 8.2, can efficiently generate STDP window, triplet, and quadruplet experiments, using the synaptic parameters optimised for these experiments. In addition, according to another set of numerical simulations, the frequency-dependent pairing experiments and also the BCM-like rate-based experiments, can be regenerated through the second minimal model, shown as Eq. 8.3, and by employing the synaptic parameters optimised for the frequency-dependent pairing experiments. As the full TSTDP rule is minimised, the proposed circuit that approximates the full TSTDP rule, can also be further modified and hence the number of transistors is reduced from the 18 transistors required for the full TSTDP circuit shown in Fig. 8.1.

This chapter presents experimental results of two minimal TSTDP circuits that correspond to the two minimal TSTDP models mentioned in previous paragraph. According to the minimal rules shown in both Eqs. 8.2 and 8.3, the depression contribution of the spikes triplet interactions can be neglected without having a significant effect on the circuit performance in reproducing the targeted biological experiments. The triplet depression part in the full TSTDP circuit shown in Fig. 8.1, is the four transistors surrounded in the red-dashed box. Therefore, the minimal TSTDP circuit, is the one shown in Fig. 8.1 minus the part enclosed in the red-dashed box, i.e only 14 transistors are needed to regenerate all desired biological experiments, as it is shown in Fig. 8.2. This is the first minimal TSTDP circuit.

In addition, the numerical simulation results suggest that, for generating the frequency-dependent pairing experiments, as well as the BCM experiment, further to the triplet depression part of the circuit, the pairing potentiation part is not also necessary and can be removed (see Section 2.6.1). Therefore, in the case of second minimal TSTDP rule, shown in Eq. 8.3, A_2^+ can be set to zero. As a result, one more transistor that is shown in the blue dashed-box can be also removed from the proposed full TSTDP circuit (Fig. 8.1) and therefore only 13 transistors are required for generating the mentioned pairing and BCM experiments (Pfister and Gerstner 2006). The resulting circuit is shown in Fig. 8.3. This is the second minimal TSTDP circuit (Azghadi *et al.* 2013b).

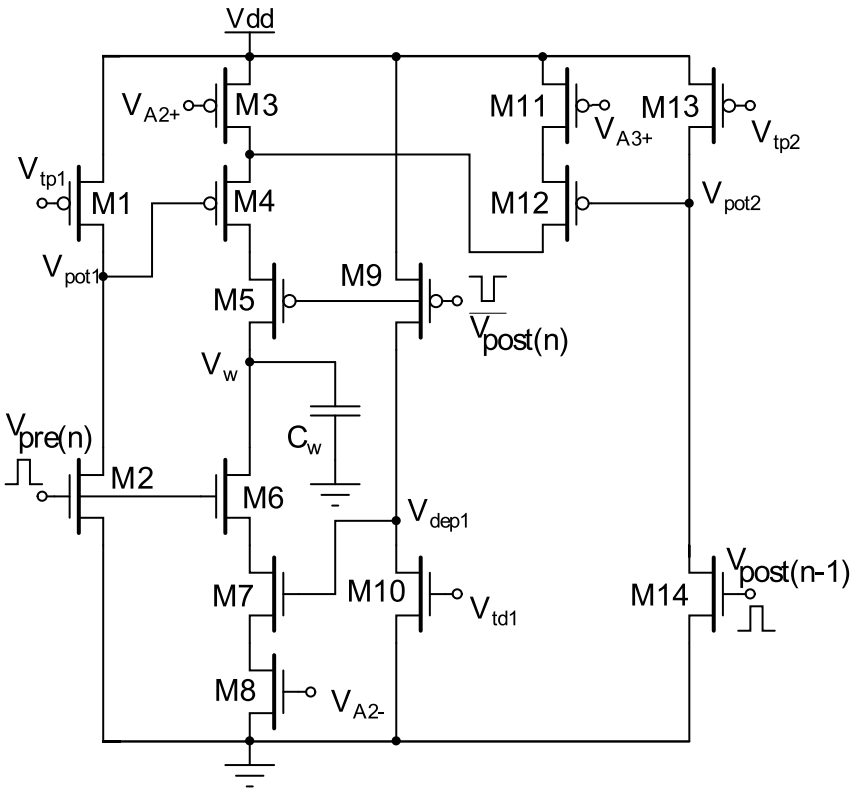


Figure 8.2. First minimal TSTDP circuit. This circuit corresponds to the first minimal TSTDP model shown in Eq. 8.2. Therefore, this circuit does not include the triplet depression section of the full TSTDP circuit shown in Fig. 8.1.

Experiments Data Sets

Since there are two versions of the minimal TSTDP rule each corresponding to a minimal TSTDP circuit, two sets of simulations are performed using the proposed minimal circuits. Each simulation set considers a specific set of data from the experiments. The first experimental data set that is utilised originates from hippocampal culture experiments that examine pairing, triplet and quadruplet protocols effects on synaptic weight change (Wang *et al.* 2005). This first data set consists of 13 data points obtained from Table 2 of Pfister and Gerstner (2006). These data points include (i) two data points and error bars for pairing protocol (ii) three data points and error bars for quadruplet protocol, and (iii) eight data points and error bars for triplet protocol. This data set shows the experimental weight changes, Δw, as a function of the relative spike timing Δt, Δt_1, Δt_2 and T under pairing, triplet and quadruplet protocols in hippocampal culture.

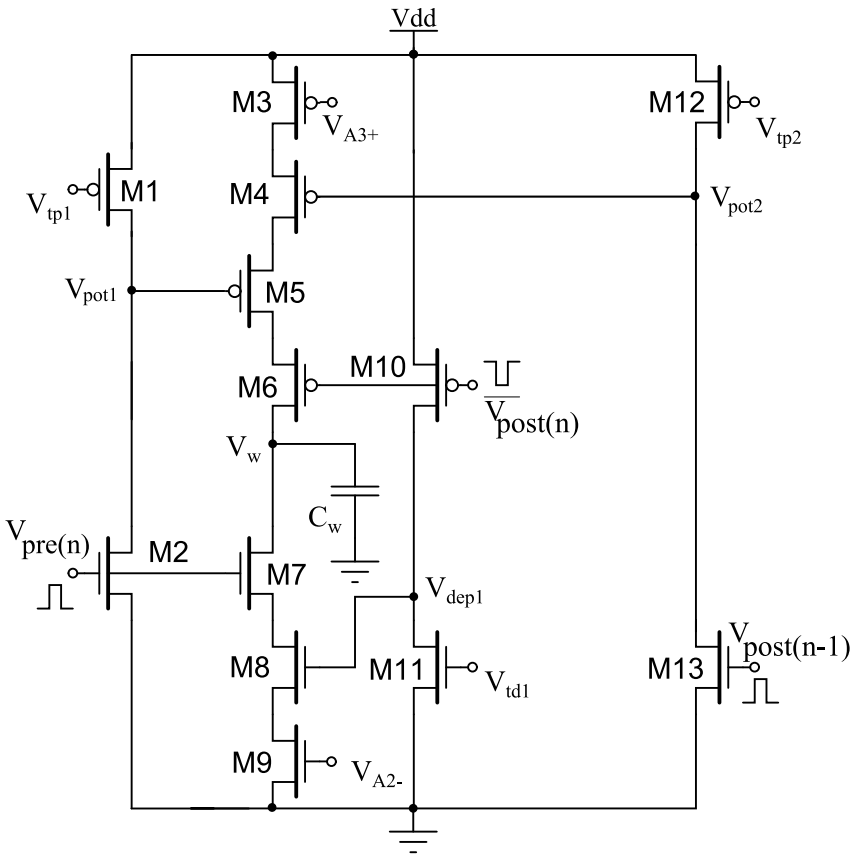


Figure 8.3. Second minimal TSTDP circuit. This circuit corresponds to the second minimal TSTDP model shown in Eq. 8.3. Therefore, this circuit does not include the triplet depression section of the full TSTDP circuit shown in Fig. 8.1, nor its pairing potentiation part (Azghadi *et al.* 2013b).

The second data set originates from experiments on the visual cortex, which investigated how altering the repetition frequency of spike pairings affects the overall synaptic weight change (Sjöström *et al.* 2001, Sjöström *et al.* 2008). This data set is composed of 10 data points—obtained from Table 1 of Pfister and Gerstner (2006)—that represent experimental weight change, Δw, for two different Δt's, and as a function of the frequency of spike pairs under a frequency-dependent pairing protocol in the visual cortex. Note that, the two mentioned data sets are those that were also used in the experiments performed in the previous chapters, when other instances of PSTDP as well as TSTDP circuits were verified for showing various synaptic plasticity experimental outcomes.

Circuit Simulation and Configuration

The minimised circuits are simulated in HSpice using the 0.35 μm C35 CMOS process by AMS. All transistors in the both designs (shown in Figs. 8.2 and 8.3) are set to 1.05 μm wide and 0.7 μm long. The weight capacitor value is set to 1 pF. It should be noted that the circuits are simulated in an accelerated time scale of 1000 times compared to real time, with all pulses having a 1 μs pulse width. This is the same approach that has been utilised in previous chapters and by previous synaptic plasticity circuit implementations such as Schemmel *et al.* (2006), Tanaka *et al.* (2009), Schemmel *et al.* (2010), Mayr *et al.* (2010) and Wijekoon and Dudek (2012). For the sake of simplicity when comparing simulation results to the biological experimental data, all shown results are scaled back to real time.

Furthermore, similar to previous chapters, the nearest-spike interaction of spikes is implemented in the proposed circuit that corresponds to the nearest-spike model of TSTDP rule presented in Pfister and Gerstner (2006). The circuit is examined under same protocols, using which the biological experiments presented in Sjöström *et al.* (2001), Froemke and Dan (2002), Wang *et al.* (2005), and the Matlab numerical simulations (performed in Chapter 2) were carried out.

Data Fitting Approach

Identical to the TSTDP computational simulations (Pfister and Gerstner 2006) and previous TSTDP circuit studies (Azghadi *et al.* 2011c, Azghadi *et al.* 2013a), which test the triplet model/circuit simulation results against the experimental data using a Normalised Mean Square Error (NMSE) for each of the data sets, the proposed circuit is verified by comparing its simulation results with the experimental data and ensuring a small NMSE value. The NMSE is calculated using Eq. 2.1.

In order to minimise the resulting NMSEs for the circuit and fit the circuit output to the reported experimental data in the literature, there is a need to adjust the circuit bias parameters and time constants. This is an optimisation process of the circuit bias voltages, which results in reaching a minimum NMSE value and so the closest possible fit to the experimental data.

Circuit Bias Optimisation Method

In order to minimise the NMSE function and achieve the highest analogy to the experimental data, the circuit bias voltages, which tunes the required parameters from the

models should be optimised. For this purpose, Matlab and HSpice were integrated in a way to minimise the NMSE resulted from circuit simulations using the Matlab built-in function `fminsearch`. This function finds the minimum of an unconstrained multivariable function using a derivative-free simplex search method. This is the same optimisation function that was also used in the simulation results of the previous TSTDP circuits.

8.4.2 Synaptic Plasticity Experiments with the Proposed TSTDP Minimal Circuits

Pairing Experiment (STDP Timing Window)

The first simulation that is performed using the first proposed minimal TSTDP circuit shown in Fig. 8.2, is reproducing the STDP learning window that demonstrates spike timing-dependent potentiation and depression, under pairing protocol (see Section 2.5.1). Fig. 8.4 shows how the proposed circuit can successfully perform the timing dependent weight modifications. This figure shows the normalised experimental data extracted from (Bi and Poo 1998) in blue. It suggests that the proposed circuit behaviour under a pairing (window) protocol can approximate the experimental data generated with the same protocol. Besides the blue experimental data, two other experimental values for $\Delta t = 10$ ms and $\Delta t = -10$ ms are shown with their standard error mean represented by black bars. These points are the first two points of the 13 data points of the aforementioned first (hippocampal) data set. These two points, were utilised to test and optimise the bias voltages of the first minimal TSTDP circuit. This is a similar approach to the method used in Pfister and Gerstner (2006). The circuit bias parameters for generating the STDP window are those corresponding to the hippocampal data set as shown in Table 8.1.

Quadruplet Experiment

The second simulation is performed using the first minimal TSTDP circuit and under quadruplet protocol. Fig. 8.5 demonstrates how the proposed circuit approximates the timing dependent weight modifications close to those for quadruplet experiment. In these results, the black data points are extracted from Wang *et al.* (2005), and the black deviation bars and data points are those that were used in Pfister and Gerstner (2006)

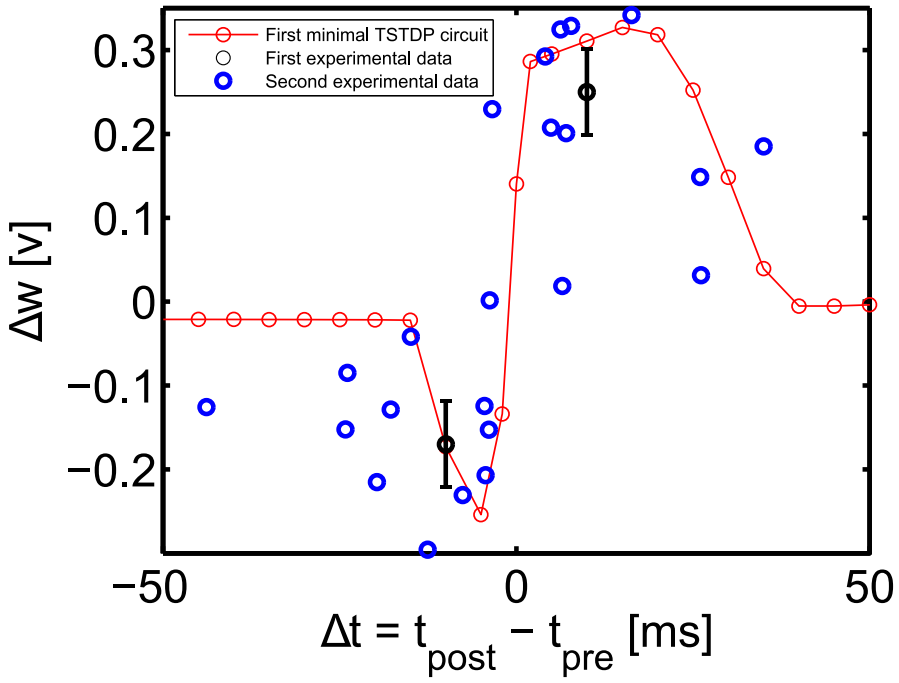


Figure 8.4. STDP timing window experiment in the hippocampal region can be approximated using the first minimal TSTDP circuit. Synaptic weight changes are produced under pairing protocol. The circuit bias parameters for generating the window approximation are those corresponding to the hippocampal data set shown in Table 8.1. The first experimental data set shown in black contains two data points with their standard error mean extracted from Pfister and Gerstner (2006), and the second experimental data set is part of the normalised experimental data extracted from Bi and Poo (1998).

for quadruplet experiments. The circuit bias parameters for generating the quadruplet approximation are those corresponding to the hippocampal data set shown in Table 8.1.

Triplet Experiment

The third experiment that is performed on the first minimal TSTDP circuit is the triplet experiment performed in the hippocampal region and reported in Wang *et al.* (2005) and Pfister and Gerstner (2006). Fig. 8.6 demonstrates how the proposed circuit approximates the timing dependent weight modifications close to those for triplet experiments. In the shown results, the black data and deviation bars are those that were used in Wang *et al.* (2005) and Pfister and Gerstner (2006) for triplet experiments. The circuit bias parameters for generating the triplet approximation are those corresponding to the hippocampal data set as shown in Table 8.1.

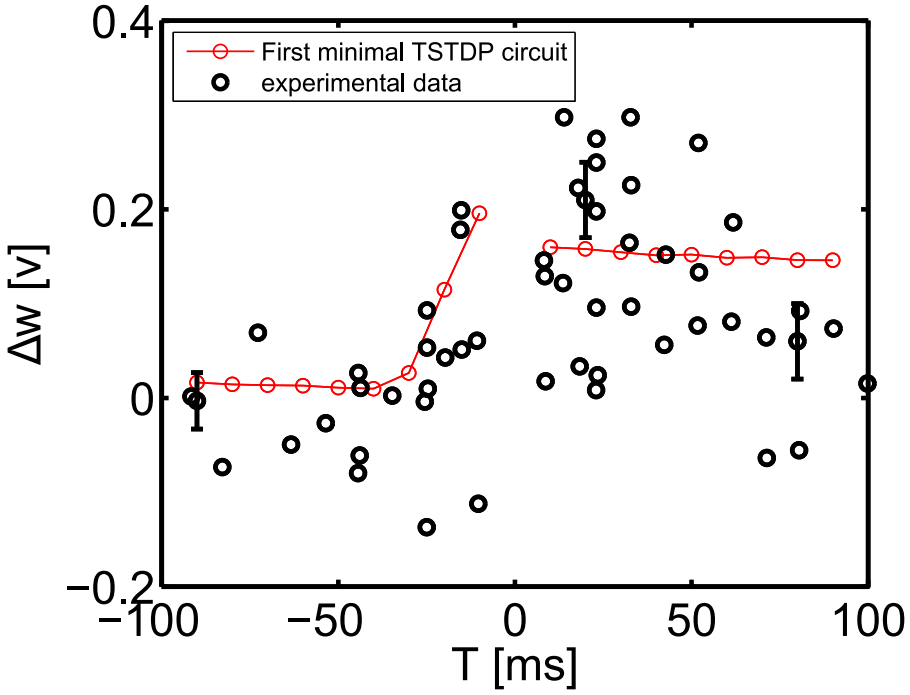


Figure 8.5. Quadruplet experiment in the hippocampal region can be approximated using the first minimal TSTDP circuit. Synaptic weight changes are produced under quadruplet protocol. The circuit bias parameters for generating the quadruplet approximation are those corresponding to the hippocampal data set as shown in Table 8.1. The experimental data shown in black were extracted from Wang *et al.* (2005) and Pfister and Gerstner (2006).

Simulation results show that the TSTDP circuit can distinguish between the pre-post-pre and post-pre-post spike combinations and show analogy to the experiments. However, the simulation results using the computational pair-based STDP model shown in Pfister and Gerstner (2006), as well as the results generated using different PSTDP circuits (Azghadi *et al.* 2011c, Azghadi *et al.* 2012b), demonstrate that the pair-based STDP models and circuits do not have the ability to distinguish among triplet combinations.

Considering Figs. 8.4 to 8.6, the first proposed minimal TSTDP circuit, can reach a good approximation of pairing, quadruplet, and triplet experiments, using a shared optimised set of bias voltages. Using these bias voltages a NMSE $= 2.04$ is obtained, when considering the 13 data points in the hippocampal data set. This is better than the minimal NMSE obtained using the minimal TSTDP computational model, as presented in Pfister and Gerstner (2006).

Table 8.1. Optimised biases for the minimal TSTDP circuits and two data sets. The presented values show optimised bias parameters for the minimal TSTDP circuits in order to reach the minimal NMSEs for the targeted set of data and experiments. The hippocampal (first) set of optimised bias parameters generate the results shown for pairing, quadruplet and triplet experiments. The visual cortex (second) set of optimised bias parameters are used to reach the minimal NMSE in frequency-dependent pairing experiment.

Data set	$V_{A_2^+}$ (V)	$V_{A_2^-}$ (V)	$V_{A_3^+}$ (V)	V_{tp1} (V)	V_{td1} (V)	V_{tp2} (V)	NMSE
Hippocampal (first)	3.2	0.32	2.7	2.75	0.35	2.65	2.04
Visual cortex (second)	0	0.29	2.7	2.7	0.17	2.86	0.39

Besides the above experiments, which are similar to the experiments performed by Pfister and Gerstner (2006), the proposed minimal circuit is additionally tested for all possible combination of spike triplets under the same protocol that used by Froemke and Dan (2002).

Extra Triplet Experiment

As already mentioned, in 2002 Froemke and Dan proposed a suppression model for higher order spike trains and performed some experiments using the aforementioned extra triplet protocol. Their proposed suppression model can account for the required non-linearity in STDP experiments, when considering higher order of spike combinations. Fig. 8.7 shows that the first minimal TSTDP circuit, under the extra triplet protocol, and using the same set of parameters that were optimised for hippocampal experiments (shown in Table 8.1), is able to account for a similar behaviour to the experiments performed by Froemke and Dan in 2002 and for extra triplet patterns. Nonetheless, there is slight contrast between the achieved results using the TSTDP circuits and those produced under the suppressive model and reported in Froemke and Dan (2002). The result shown in the right bottom square of Fig. 8.7(a), which presents synaptic weight changes due to post-pre-post, demonstrates potentiation. This is in total agreement to the result shown in Fig. 8.6(b), which also shows potentiation for post-pre-post spike combination. However, the suppressive model results show a depression for this spike combination—Fig. 3b in Froemke and Dan (2002). Pfister and Gerstner (2006) discussed that this difference is due to the nature of the suppressive model, which gives rise to a depression when a post-pre-post spike triplet occurs, while

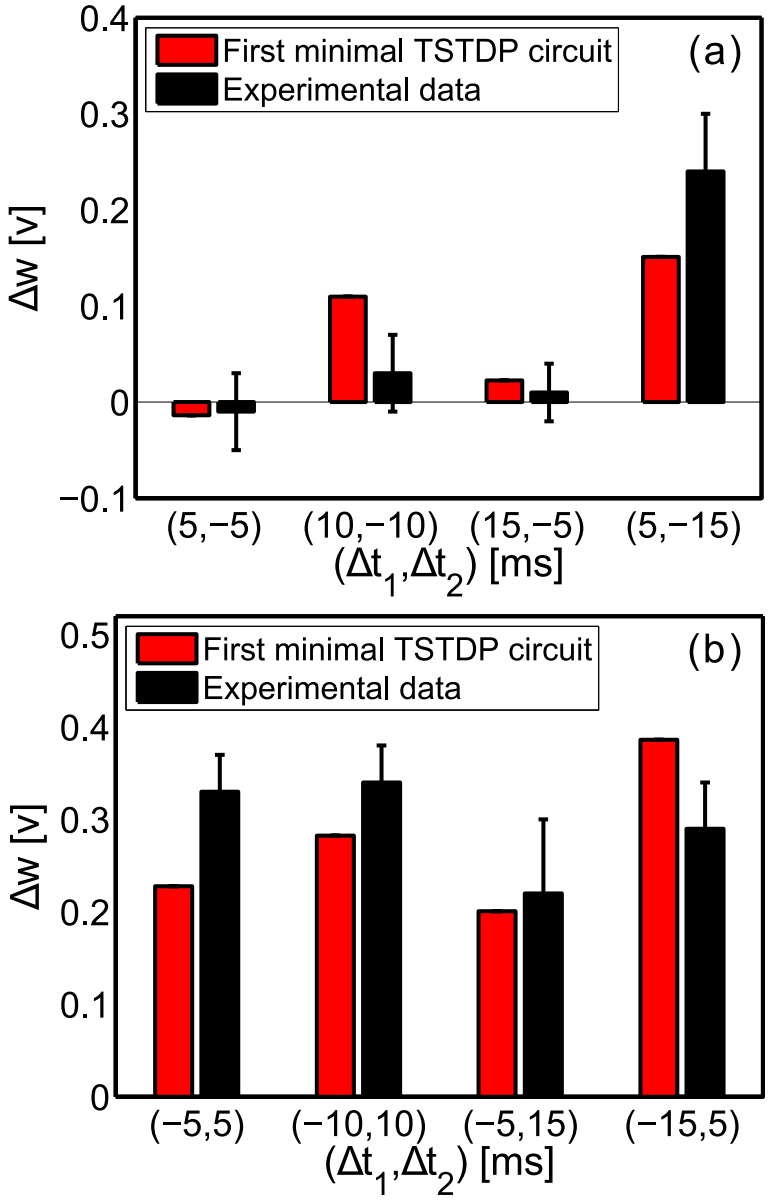


Figure 8.6. Triplet experiments in the hippocampal region can be approximated using the first minimal TSTDP circuit. Synaptic weight changes are produced under the triplet protocol. The circuit bias parameters for generating the triplet approximation are those corresponding to the hippocampal data set as shown in Table 8.1. The experimental data, shown in black and their standard deviations extracted from Wang *et al.* (2005) and Pfister and Gerstner (2006). (a) Simulation and experimental results for the pre-post-pre combination of spike triplets with various timings. (b) Simulation and experimental results for the post-pre-post combination of spike triplets with various timings.

clearly it leads to a potentiation in the TSTDP model. In a later study, Froemke *et al.* (2006) revised their model in order to address this issue.

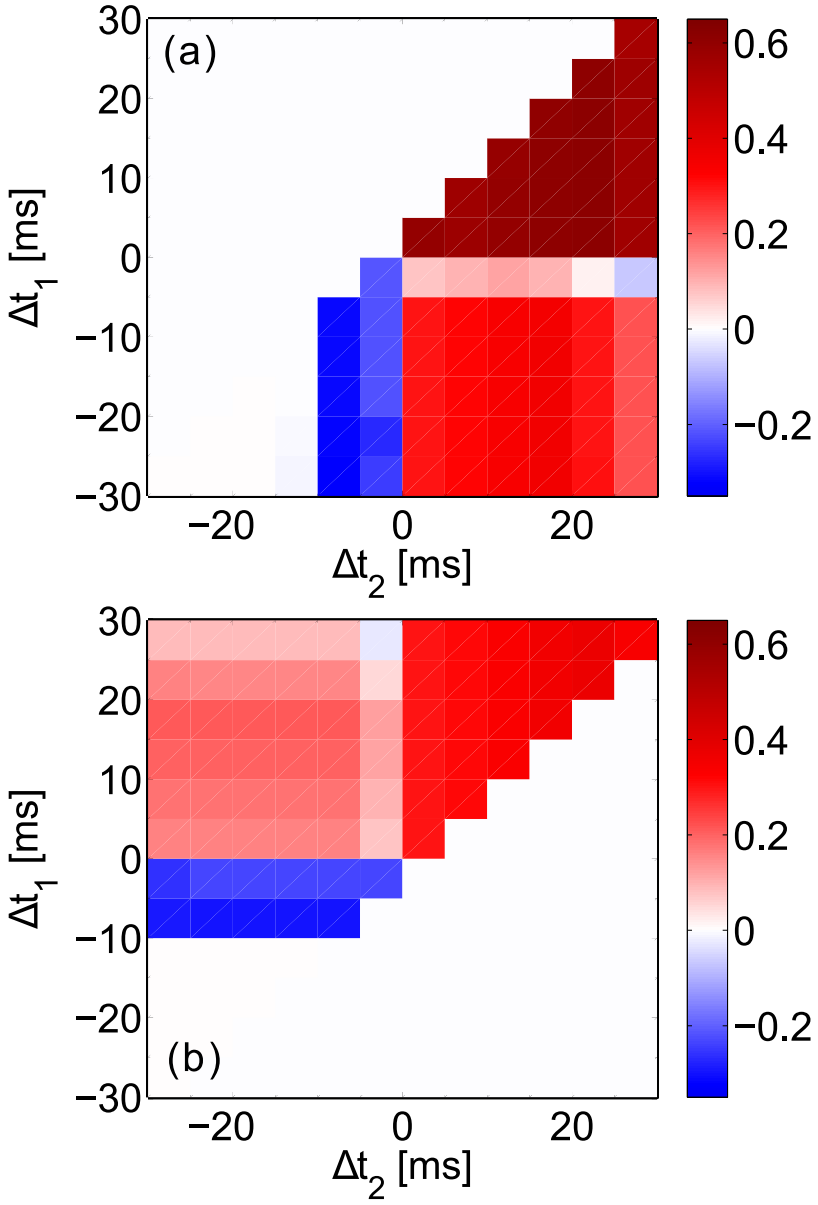


Figure 8.7. Extra triplet experiments using the suppression STDP model can be approximated using the first minimal TSTDP circuit. Synaptic weight changes in result of extra triplet protocol for (a) pre-post-post (top right triangle), post-post-pre (bottom left triangle) and post-pre-post (right bottom square) and (b) for pre-post-pre (top left square), pre-pre-post (top right triangle) and post-pre-pre (left bottom triangle) combination of spikes produced by the first minimal TSTDP circuit. The circuit bias parameters for generating the synaptic weight changes shown in this figure correspond to the hippocampal bias set shown in Table 8.1.

Frequency-dependent Pairing Experiment

As already mentioned, the frequency-dependent pairing experiments that were performed in the visual cortex, can also be replicated using a minimal TSTDP model. This model is simpler than the first minimal model and not only does not require the A_3^- parameter from the full triplet model, but also it does not need the A_2^+ parameter (see Eq. 8.3). Hence, the minimal circuit for generating this experiment is also simpler from the first minimal circuit as shown in Fig. 8.3. In order to approximate the outcome of frequency-dependent pairing experiments, which corresponds to the aforementioned visual cortex (second) data set, as reported in Sjöström *et al.* (2001) and Pfister and Gerstner (2006), a new set of synaptic parameters for the model and therefore a new set of bias voltages for second minimal circuit is required.

As shown in Fig. 8.8, the optimised biases (shown in Table 8.1) for the second minimal circuit can closely approximate the outcomes of experiments under frequency-dependent pairing protocol. The minimal obtained NMSE for this experiments was 0.39, which is close to the numerical simulation result of 0.34 reported in Pfister and Gerstner (2006). It is worth mentioning that the second minimal TSTDP circuit has only one transistor more than the simple PSTDP circuit proposed in Indiveri *et al.* (2006), but it has the ability to reproduce the frequency-dependent pairing experiments, while all neuromorphic PSTDP circuits, even with much higher number of transistors—see (Bofill-I-Petit and Murray 2004, Tanaka *et al.* 2009, Bamford *et al.* 2012b) for example—fail to replicate these experiments (Azghadi *et al.* 2011c).

BCM-like rate based experiment

In addition to the outcome of frequency-dependent experiments, the second minimal TSTDP circuit is also able to account for a BCM-like behaviour. By employing the same circuit and set of bias parameters, which were used to generate frequency-dependent pairing experiments shown in Fig. 8.8, a BCM-like experiment is also reproducible. Fig. 8.9 depicts the synaptic weight changes produced by the second minimal TSTDP circuit and under the aforementioned Poissonian protocol (Section 2.5.6). In this figure, three different curves show synaptic weight changes according to three different synaptic modification thresholds that demonstrate the points where LTD changes to LTP. The threshold is adjustable using the TSTDP rule parameters. In order to move the sliding threshold toward left or right, the $V_{\mathrm{A3+}}$ parameter can be altered as it is depicted in Fig. 8.9. The rate of random pre-synaptic Poissonian spike trains, ρ_{pre}, is

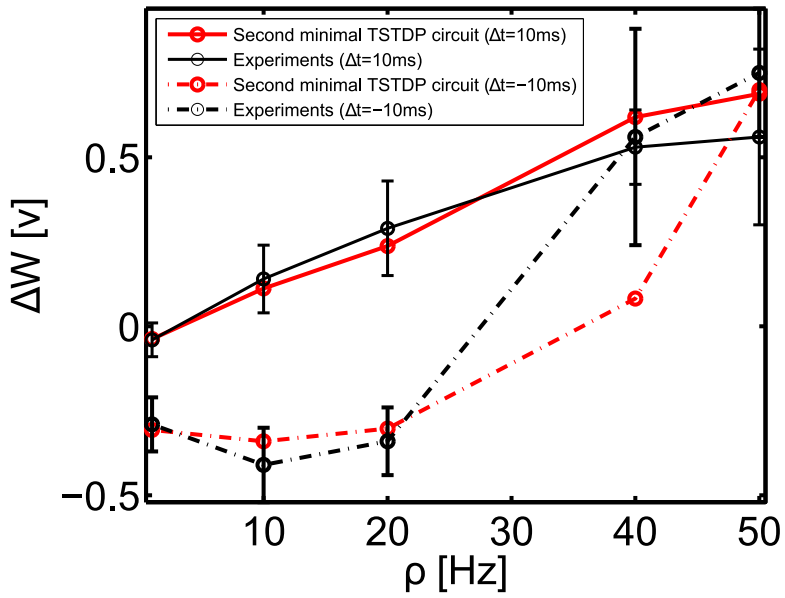


Figure 8.8. Frequency-dependent pairing experiment in the visual cortex region can be approximated using the second minimal TSTDP circuit. Simulation results are produced under frequency-dependent pairing protocol (Azghadi *et al.* 2013b). The circuit bias parameters for generating the synaptic weight changes shown in this figure correspond to the visual cortex (second) set of bias parameters shown in Table 8.1. The experimental data shown in black are extracted from Sjöström *et al.* (2001) and Pfister and Gerstner (2006).

equal to 10 Hz, and the trains with this spiking rate, are regenerated for each data point. Each data point shows the mean value of the weight changes for 10 various post-synaptic Poissonian spike trains and the error bars depict the standard deviations of the weight changes for each data points over 10 runs. In this experiment, similar to the experiment performed in Pfister and Gerstner (2006), the frequency of the post-synaptic spike, ρ_{post} is swept over a range of frequencies from 0 Hz up to 50 Hz, while the pre-synaptic spiking frequency, ρ_{pre}, is kept fixed at 10 Hz.

Although Pfister and Gerstner have used this methodology to show that their model is able to reproduce a BCM-like behaviour, in the original BCM experiments reported in Kirkwood *et al.* (1996), the synaptic weight changes were measured whilst the pre-synaptic and not the post-synaptic spike rate was swept (Cooper *et al.* 2004). In order to check that the proposed circuit could reproduce BCM-like behaviour, which is driven by pre-synaptic activity, the circuit simulation was repeated. Similar to the experiments presented in Chapter 7, we made this simple assumption that post-synaptic firing rate

is a linear function of the pre-synaptic firing rate, i.e. $\rho_{\text{post}} = A\rho_{\text{pre}}$ and for the sake of simplicity we let $A = 1$, i.e $\rho_{\text{post}} = \rho_{\text{pre}}$. Despite such a crude approximation, the circuit is successfully able to mimic BCM-like behaviour where weight changes were pre-synaptically driven, as illustrated in Fig. 8.10. In this figure, each data point shows the mean value of the weight changes for 10 different trials using random Poissonian pre- and post-synaptic spike trains for each trial, and the error bars depict the standard deviations of the associated weight changes over these 10 trials.

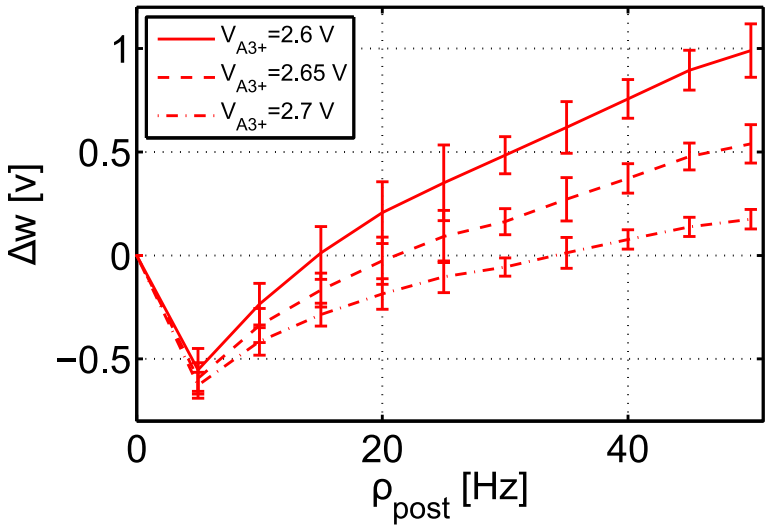


Figure 8.9. Post-synaptically driven BCM-like behaviour with sliding threshold feature can be approximated using the second minimal TSTDP circuit. Simulation results are produced under Poissonian protocol for BCM. The circuit bias parameters for generating the synaptic weight changes shown in this figure correspond to the visual cortex (second) set of bias parameters shown in Table 8.1. In this simulation, the pre-synaptic frequency, ρ_{pre}, was kept fixed at 10 Hz, and the post-synaptic frequency, ρ_{post}, was swept (see the text for more details).

All these experiments suggest that the proposed timing-based circuit has sufficient ability to replicate the outcome of other synaptic plasticity experiments, for BCM-like behaviour. In the next section we discuss and compare the proposed circuit and its counterparts from various circuit design as well as biological plausibility perspectives.

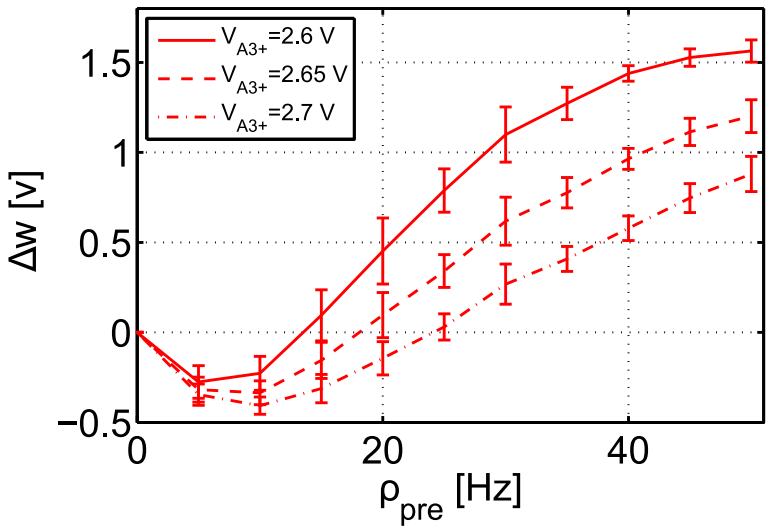


Figure 8.10. Pre-synaptically driven BCM-like behaviour with sliding threshold feature can be approximated using the second minimal TSTDP circuit. Simulation results are produced under Poissonian protocol for BCM. The circuit bias parameters for generating the synaptic weight changes shown in this figure correspond to the visual cortex (second) set of bias parameters shown in Table 8.1. In this simulation, the pre-synaptic frequency, ρ_{pre}, was swept, while the neuron is linear and $\rho_{\text{pre}} = \rho_{\text{post}}$ (see the text for more details).

8.5 Synaptic Plasticity Circuit Comparison

In order to measure the efficiency of the proposed analog neuromorphic circuit, it should be compared to its counterparts in terms of strength in reproducing the outcomes of various synaptic plasticity experiments. Besides, it is also essential to compare the proposed design with available synaptic plasticity circuits in various circuit design aspects such as required silicon real-estate, energy consumption, and process variation tolerance. In the following sections, we demonstrate that the proposed synaptic plasticity circuit outperforms most of its previous counterparts. In addition, it will be shown that the proposed circuit is much simpler, consumes less power and occupies smaller area in comparison to previous synaptic plasticity circuits. Furthermore, we show that the presented synaptic plasticity circuit is better than its counterparts in terms of process variation tolerance when a trade-off between complexity and performance is considered.

8.5.1 Synaptic Plasticity Ability for Reproducing Experimental Data

As already mentioned, the proposed design is able to regenerate the outcomes of a variety of synaptic plasticity experiments. These experiments are not reproducible by any of the previous circuits designed for PSTDP rule. However, they can be replicated using a number of previously proposed TSTDP circuits, as well as a few other synaptic plasticity designs. Table 8.2 shows a detailed comparison among investigated circuits, in terms of the ability to mimic various biological experimental outcomes.

This table demonstrates that all PSTDP and TSTDP circuits are able to account for a BCM-like behaviour. However, simulation results presented in Azghadi *et al.* (2012a) suggest that, using a TSTDP circuit, a better BCM-like behaviour is attainable and since there are more parameters available in the circuit, there will be a higher degree of control over the sliding threshold of the BCM rule. In addition, there is no evidence, if any of the circuits proposed in Mitra *et al.* (2009) or Meng *et al.* (2011) are capable of showing a BCM-like behaviour with sliding threshold feature.

The table also summarises the ability of the proposed TSTDP circuit in reproducing other required experiments. Although a number of other synaptic plasticity circuits that are shown in the table, are also capable of qualitatively generating the required experiments (Mayr *et al.* 2010, Rachmuth *et al.* 2011), they need changes in their synaptic parameters or in their initial implementations, in order to be able to mimic biological experiments closely and with a small error. The table shows that the TSTDP designs proposed in Azghadi *et al.* (2011c), Azghadi *et al.* (2012b) and Azghadi *et al.* (2013a) as well as the proposed design in this chapter are able to account for all experiments using shared set of bias parameters. This is a useful feature of the synaptic plasticity circuit, to be able to reproduce as many experimental outcomes as possible, using a single set of parameters, and by means of least changes to the hardware. As a result, this new plasticity circuit can be used in developing large-scale networks of spiking neurons with high synaptic plasticity abilities.

When implementing a large-scale network of spiking neurons, the synaptic plasticity circuits should be as area- and power-efficient as possible. This leads to the essential requirements of a large scale neuromorphic design, which include low power consumption and small area occupation. Despite these essential needs, most of the previously available synaptic plasticity VLSI designs do not meet these requirements. Some of these designs mimic the required biological functions well, but at the same time

Table 8.2. Comparison of various synaptic plasticity VLSI circuits.

✓ indicates that the outcomes of experiments can be closely mimicked using the circuit.
✓* indicates that the related study has not investigated the corresponding experiment, but according to its plasticity rule, it can most likely reproduce the expected experiment, though using a different set of plasticity parameters.
✓** indicates that the related study has not investigated the corresponding experiment, but according to its plasticity rule, it might be able to reproduce the expected experiment.
× indicates that the outcomes of experiments cannot be generated using the circuit.

Experiment / Plasticity Circuit	STDP window	Pairing frequency	Triplet	Quadruplet	BCM
PSTDP (Bofill-I-Petit and Murray 2004)	✓	×	×	×	✓
PSTDP (Cameron *et al.* 2005)	✓	×	×	×	✓
PSTDP (Indiveri *et al.* 2006)	✓	×	×	×	✓
PSTDP (Schemmel *et al.* 2006)	✓	×	×	×	✓
PSTDP (Arthur and Boahen 2006)	✓	×	×	×	✓
PSTDP (Koickal *et al.* 2007)	✓	×	×	×	✓
PSTDP (Tanaka *et al.* 2009)	✓	×	×	×	✓
PSTDP (Ramakrishnan *et al.* 2011)	✓	×	×	×	✓
PSTDP (Cassidy *et al.* 2011)	✓	×	×	×	✓
PSTDP (Bamford *et al.* 2012b)	✓	×	×	×	✓
PSTDP (Cruz-Albrecht *et al.* 2012)	✓	×	×	×	✓
SDSP (Mitra *et al.* 2009)	✓*	✓**	✓**	✓**	✓**
Voltage-based BCM (Mayr *et al.* 2010)	✓	✓	✓	✓*	✓
Iono-neuromorphic (Meng *et al.* 2011)	✓*	✓**	✓**	✓**	✓**
Iono-neuromorphic (Rachmuth *et al.* 2011)	✓*	✓**	✓**	✓**	✓
TSTDP (Azghadi *et al.* 2011c)	✓	✓	✓	✓	✓
TSTDP (Azghadi *et al.* 2012b)	✓	✓	✓	✓	✓
TSTDP (Azghadi *et al.* 2013a)	✓	✓	✓	✓	✓
New low energy and compact TSTDP circuit	✓	✓	✓	✓	✓

are large and power hungry such as the designs presented in Bofill-I-Petit and Murray (2004), Mayr *et al.* (2010), Meng *et al.* (2011), Rachmuth *et al.* (2011), Azghadi *et al.* (2012b) and Azghadi *et al.* (2013a). Some other designs such as the synaptic plasticity circuits presented in Indiveri *et al.* (2006), Tanaka *et al.* (2009), Bamford *et al.* (2012b) and Cruz-Albrecht *et al.* (2012), have improved power and area features, but do not have most of the required biological abilities. Therefore, a circuit with low power and area consumption and at the same time with high synaptic plasticity capabilities is required. The design presented in this chapter aims at reaching these goals. This design

has high synaptic weight modification ability, while it is low power and occupies small silicon area.

8.5.2 Area and Power Consumption

Since the proposed design only uses a small number of transistors to reach its required synaptic plasticity features, compared to many previous designs with inferior or equal synaptic capability, the area and power consumption in this design are lower than all previous designs with similar capabilities, and close to other designs with much lower synaptic strength. Table 8.3 compares the proposed design, with a number of the previous synaptic plasticity designs available in the literature, in terms of complexity (required number of transistors and capacitors), which has a direct relation with the needed silicon area, and their estimated power consumption.

Power consumption of a synaptic plasticity circuit is directly linked to its synaptic biasing parameters such as its synaptic time constants e.g. V_{tp1}, V_{td1}, V_{tp2}, V_{td2}, as well as its synaptic amplitude parameters e.g. $V_{A_2^+}$, $V_{A_2^-}$, $V_{A_3^+}$, $V_{A_3^-}$. In addition, consumed power is in a direct relation with the supply power, as well as the spike pulse width. Therefore, in order to have a fair comparison among synaptic plasticity circuits, they should all be compared under similar conditions. The presented results in the last six rows of Table 8.3, depict the simulation results for various circuits under similar conditions. The synaptic parameters, for all these synaptic circuits are firstly optimised to reach the best NMSEs for the hippocampal data set. The optimisation process determines the value of synaptic biasing parameters, which significantly influence the power consumption of these circuits. For instance, the high power consumption observed in the TSTDP circuit proposed in Azghadi *et al.* (2013a) is due to large time constants required for reaching a small NMSE $= 1.74$, which results in transistors being on for longer period of time and this leads to high power consumption. Table 8.3 reports the energy consumption per spike for a number of the mentioned designs. The energy consumption is measured on both pre-synaptic and post-synaptic spikes. Due to differences in depression and potentiation biasing parameters, different energy consumptions are measured for pre- and post-synaptic spikes, but the larger one is reported in Table 8.3.

The energy consumption per spike for the first three designs in Table 8.3, are extracted from related papers. These circuits are PSTDP circuits, which do not possess the high

Table 8.3. Area and power comparison for various synaptic plasticity circuits.

* The biases are optimised for the hippocampal (first) data set to reach minimal NMSEs and then the energy consumptions are measured.

** The PSTDP and TSTDP designs presented in the last six rows of this table are all simulated using a 3.3 V supply voltage, while other designs use equal or lower supply voltages.

*** This design has been implemented in a 90 nm CMOS process with a supply voltage of 0.6 V.

Comparison Measure / Plasticity Circuit	Transistor No.	Capacitor No.	Energy per spike**	NMSE*
PSTDP (Bamford *et al.* 2012b) with weight dependence	15	5	0.3 pJ	>10
PSTDP (Cruz-Albrecht *et al.* 2012)***	>100	4	0.37 pJ	>10
PSTDP (Tanaka *et al.* 2009)	18	3	42 pJ	>10
Voltage-based BCM (Mayr *et al.* 2010)	>100	2	NA	NA
Iono-neuromorphic (Meng *et al.* 2011)	>100	2	NA	NA
Iono-neuromorphic (Rachmuth *et al.* 2011)	>100	2	NA	NA
PSTDP (Bofill-I-Petit and Murray 2004) without weight dependency part	15	3	1.5 pJ	10.76
PSTDP (Indiveri *et al.* 2006)	12	1	3 pJ	11.3
TSTDP (Azghadi *et al.* 2011c)	26	1	0.03 pJ	3.46
TSTDP (Azghadi *et al.* 2012b)	44	7	1.5 pJ	2.25
TSTDP (Azghadi *et al.* 2013a)	37	5	60 pJ	1.74
Proposed minimal TSTDP	18	1	0.02 pJ	2.04

biological plausibility available in TSTDP circuits including the low power TSTDP design presented in this chapter. Although two of these designs are low power and consumes very low energy per spike, they require a high number of transistors/capacitors that require large silicon area. Note that in the best case, the NMSE of these designs that implement the same STDP rule as the design presented and simulated in Azghadi *et al.* (2011c) and Azghadi *et al.* (2012b), will be >10, which is not acceptable as a fitting error.

In addition, there is no energy consumption information available for the other three designs shown in the fourth to sixth rows of the table. Two of these designs are biophysically-based synaptic plasticity circuits, which are bulky detailed VLSI circuits implemented with more than 100 transistors, and the other one that implements the voltage-based BCM rule, imposes an inevitable interference with the neuron circuit and also needs more than 100 transistors for the design (Mayr *et al.* 2010, Mayr and Partzsch 2010).

Considering both area and power consumption, under similar conditions to other synaptic plasticity circuits, Table 8.3 suggests that the proposed design outperforms all other designs in terms of energy consumption, silicon real estate, and biological accuracy.

In addition to operating the transistors in the subthreshold region of operation, which makes the proposed circuit low-power, the accelerated time scale is another factor that results in a lower energy consumption, compared to other designs, which are implemented on real time scales. This is due to the fact that the static current, which is usually the dominant power consumption cause, is reduced (Wijekoon and Dudek 2012).

This allows the proposed design to be a suitable learning and computational component for large scale and low power neuromorphic circuits with high biological capability. However, one should keep in mind that, any analog VLSI design will be affected by the mismatch due to fabrication imperfections. Therefore, besides area and energy consumption, mismatch may also be taken into account when considering design of an analog synaptic plasticity circuit for learning and computational purposes.

8.5.3 Process Variation and Transistor Mismatch

As already mentioned in Chapter 5, apart from power consumption and silicon area, transistor mismatch is another challenge that is always associated with all analog VLSI

designs, especially designs for synaptic plasticity circuits (see Section 5.4.2). The functionality of these circuits are dependent on the synaptic parameters and changes in the values of these parameters, which can occur due to process variations, result in deviations from the synaptic circuit expected behaviour. These deviations can bring about degradation of synaptic plasticity capability. The mismatch may be taken into account from two different design perspectives. First, is a mismatch that occurs between the targeted design and the implemented design, and results in the physically implemented transistor to be different from the designed one. Second, is a mismatch that occurs among the transistors all over the fabricated design. These transistors suppose to have similar behaviour and functionality inter- or intra-chip. The design of large neuromorphic circuits become challenging due to these mismatches.

Transistor mismatch becomes more challenging when the transistor works in its subthreshold region of operation. This is due to the changes to the threshold of the transistor, which affect its subthreshold current characteristics. Due to the exponential behaviour and also low power consumption of transistors in their subthreshold regime, many spiking neural circuits, including neurons and synaptic weight change components are implemented in this region. In addition, many neuromorphic VLSI designs employ mismatch susceptible components such as current mirrors and differential pairs in their current- or voltage-mode structures. Therefore, these neural systems are seriously susceptible to device mismatch (Azghadi *et al.* 2012b, Azghadi *et al.* 2013a, Mayr *et al.* 2010, Poon and Zhou 2011).

A variety of mismatch minimisation techniques and approaches were introduced and discussed in Section 5.4.2. Each of the discussed approaches has its own advantages and limitations. For instance the approach used in Meng *et al.* (2011) and Rachmuth *et al.* (2011) requires especially designed process tolerant circuits with negative feedbacks and source degeneration features, which lead to higher design complexity. In addition, the fine-tuning approach, which has been successfully utilised in Azghadi *et al.* (2013a), is not applicable for large-scale neuromorphic circuits. Nonetheless, this approach could be used for a set of circuits with shared synaptic parameters across the chip, or even inter-chips, in order to reach the required functionality (Gao *et al.* 2012).

Apart from the techniques to reduce the mismatch and/or alleviate its effect, in order to have a process tolerant design, it is essential to use less components susceptible to mismatch including current mirrors (Azghadi *et al.* 2012b, Azghadi *et al.* 2013a), differential pairs (Douglas *et al.* 1995), and OTAs (Cruz-Albrecht *et al.* 2012, Koickal *et al.* 2007).

The low energy and compact design proposed in this chapter does not use any of these components. Hence, it is less susceptible to process variations than many previous designs.

Fig. 8.11 shows the variation in NMSE for the first (Hippocampal) data set, if a rigorous case mismatch scenario occurs in the fabrication. In the applied scenario, all transistors in the design independently go under a 1000 Monte Carlo (MC) threshold voltage variation, with three standard deviations from their typical process technology threshold voltage. This may cause deviations in the threshold voltage of any transistors up to 30 mV. This level of variation in the thresholds of transistors is very unlikely to occur. This variation scenario was used in a previous design proposed in Azghadi *et al.* (2013a), where under the same protocol the worst case NMSE can go up to 306 (See Fig. 7.12). Therefore, the proposed design is much more robust compared to the previous designs and that is because of not using of process variation susceptible circuit modules, such as current mirrors, which are extensively used in the previous designs (See Fig. 7.1, as well as Fig. 6.5). Note that the circuit bias parameters for all 1000 MC runs are fixed and correspond to the parameters for Hippocampal experiments shown in Table 8.1. However, as the results presented in Azghadi *et al.* (2013a) show, the bias parameters can be justified again and bring the circuit back to a significantly low NMSE.

Identical to the mismatch analysis performed in Fig. 8.11, the proposed TSTDP circuit is subjected to another variation analysis, this time using the second minimal TSTDP circuit and while stimulated under the frequency-dependent pairing protocol (see Fig. 8.8), in order to measure the variation effect. Fig. 8.12 represents 1000 MC runs, and the NMSE deviation, for the mismatch scenario explained earlier. The NMSE obtained using the new proposed circuit is significantly smaller than that of the designs presented in Azghadi *et al.* (2012b) and Azghadi *et al.* (2013a) and shown in Chapters 6 and 7.

According to Figs. 8.11 and 8.12, in both cases of mismatch analysis, more than 60% of NMSEs are very close to the best reached NMSEs in simulations. In addition, even the worst NMSEs shown in these figures that are due to severe unlikely mismatch, are still better than PSTDP circuit NMSEs even without considering variation in them.

Furthermore, it should be noted that, the applied variation scenario considers independent changes in the design. This means that the threshold voltage of every single transistor in the design changes independently, which is not likely in the case of

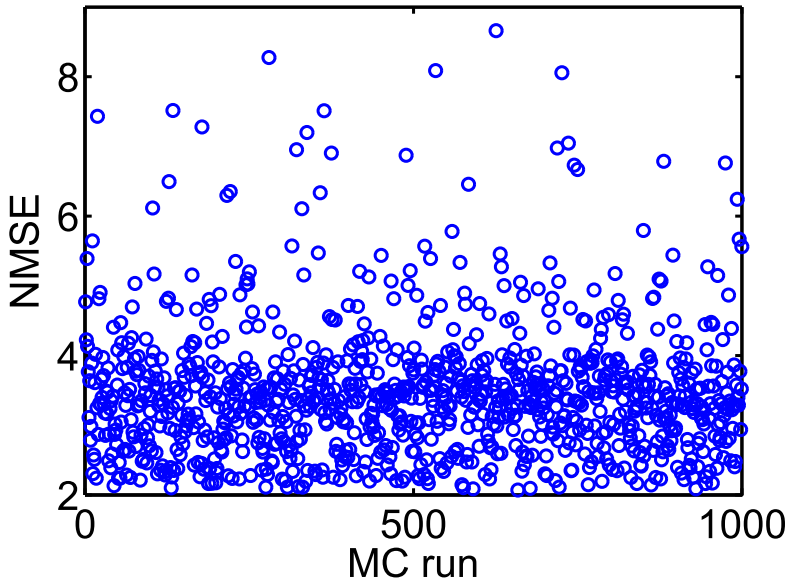


Figure 8.11. Transistor mismatch effects on the first minimal design. This figure shows 1000 Monte Carlo (MC) runs. In each run, the threshold voltage of all transistors are independently varied, based on a three-sigma deviation. The NMSE in each MC run shows the fitting error of the design, which is affected by transistors threshold deviations. Simulation results are produced under pairing, triplet and quadruplet protocols and using the first minimal TSTDP circuit. The circuit bias parameters correspond to those for the hippocampal region shown in Table 8.1.

closely positioned transistors in the proposed compact design. Considering this fact, a mismatch tolerant synaptic circuit design is expected after fabrication. However, these independent changes can happen globally and in the replicates of the proposed plasticity circuit across the chip, in the case of a large scale neuromorphic design. This means that shared fine-tuning for various sets of synaptic circuits, which are positioned in a close neighbourhood on the chip, could be an effective way of tackling the mismatch problem (Gao *et al.* 2012).

In general, Figs. 8.11 and 8.12, suggest that the proposed circuit is not heavily affected by process variation, and an acceptable synaptic behaviour compatible with several synaptic plasticity protocols is expected after fabrication. This feature along with low power consumption, small area requirement, and high biological accuracy, make the proposed circuit an ideal synaptic plasticity component that can be utilised in large-scale neuromorphic systems. These systems will have higher capability to mimic more biological experiments, while enjoying a compact structure, which consumes little

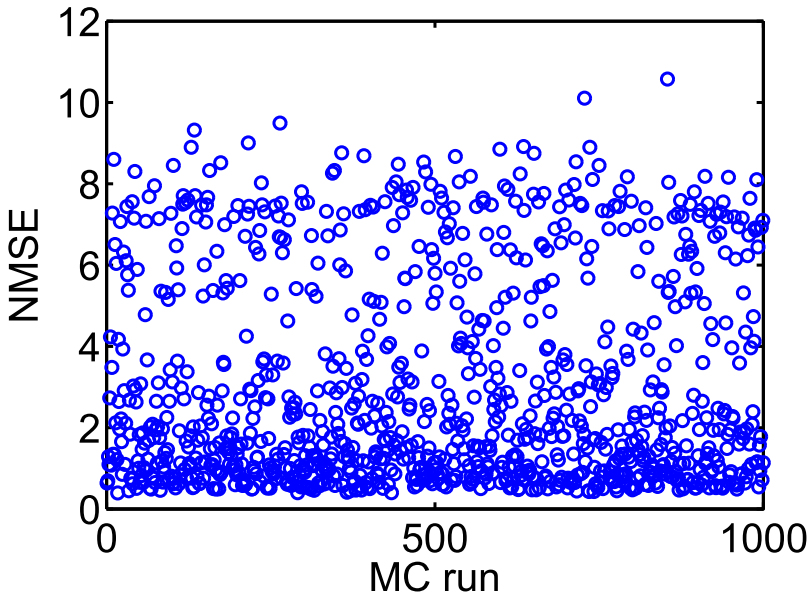


Figure 8.12. Transistor mismatch effects on the second minimal design. This figure shows 1000 Monte Carlo (MC) runs. In each run, the threshold voltage of all transistors are independently varied, based on a three-sigma deviation. The NMSE in each MC run shows the fitting error of the design, which is affected by that run deviated transistors thresholds. Similar to Fig. 8.8, simulation results are produced under frequency-dependent pairing protocol and using the second minimal TSTDP circuit. The circuit bias parameters correspond to those for the visual cortex region shown in Table 8.1.

power. This is significant progress toward developing biologically plausible systems on scales approaching that of the brain.

8.6 Discussion

Despite the performance advantages that the proposed circuit presents, it has a number of limitations that need to be considered when integrating it within a network configuration. As Fig. 8.1 demonstrates, in order to induce weight changes using the triplet circuit, current pre- or post-synaptic spike, i.e. $V_{\mathrm{pre(n)}}$ or $V_{\mathrm{post(n)}}$, as well as the immediate previous pre- or post-synaptic spike, i.e. $V_{\mathrm{pre(n-1)}}$ or $V_{\mathrm{post(n-1)}}$, are needed. This results in the need for introducing a delay into the design that provides the circuit with a delayed version of pre- and post-synaptic spike trains. Note that this is a limitation that all previous TSTDP circuits also have.

In our simulation setup, we have delayed the input pre- and post-synaptic spike trains, generated in software, for one spike width of 1 μs, and produced the required delayed spike trains, i.e. $V_{\text{pre(n-1)}}$ and $V_{\text{post(n-1)}}$. However, in the physical implementation of TSTDP circuits, the mentioned delay element should be combined with either neuron or synapse circuit, in order to produce the required delayed spike trains. Since the density of neurons is significantly lower than that of synapses in a neuromorphic system, it is therefore preferred to integrate the required delay element into the neuron design, hence saving precious silicon real estate and power. Another viable method for implementing a delay into the system is to delay the spike while transmitting it via an Address Event Representation (AER) protocol in the system. Since in the AER, only spike time stamps are transferred, the spike time for any specified value, can be easily delayed. Because the AER is an unavoidable part of any neuromorphic system, it is beneficial to use AER instead of any extra circuitry (whether part of the neuron or synapse) for introducing the required delay times into the system.

Another limitation in the proposed circuit is the use of a large weight capacitor, in order to retain the synaptic weight for required period of times, needed for adopted experimental protocols. Although this capacitor is much smaller than the weight capacitor used in the design proposed in Chapter 7, considering a large-scale neuromorphic system, there is a need to further minimise the size of capacitor. The utilised capacitor can be implemented using Metal Oxide Semiconductor Capacitors (MOSCAPs), which approximately consumes up 20×20 μm^2 of silicon real estate. Therefore, compared to the full TSTDP circuit body that is composed of 18 transistors all with 1.05 μm width and 0.35 μm length, the capacitor takes up about 90 % of the whole area required for the TSTDP circuit.

In a recent study we have shown that a similar version of the proposed low energy and compact circuit can use a 50 fF capacitor instead of the large 1 pF one, while retaining its ability to reproduce the STDP learning window, and the triplet and quadruplet experimental data (Azghadi *et al.* 2013c). This becomes possible if a modified version of the experimental protocols is used. This modified protocol considers only one pair, triplet or quadruplet of spikes, instead of the original protocols that use 60 spike sets with a frequency of 1 Hz (Pfister and Gerstner 2006). The design presented in Azghadi *et al.* (2013c), cannot account for the frequency-dependent pairing experiments, or other complicated experiments shown in this chapter, and is suitable only for experiments with high spike frequencies.

On the contrary, the utilised experimental protocols in this chapter introduce 60 pairs, triplet, or quadruplet of spikes with frequency of 1 Hz, into the TSTDP circuit, and the resulting weight change is the summation of the weight changes of all these 60 spike sets. Therefore, the synaptic weight change after each of these spike sets should be strongly preserved during the rest period before the arrival of the next spike set, or for longer times when there is no spike. As already discussed in Section 5.4.8, due to the capacitor leakage, the synaptic weight stored on the capacitor, will leak away resulting in the learnt weight will be eventually altered/lost. This is the reason why a 1 pF capacitor in the design in this chapter is employed to minimise this loss. Similarly, many of the previous designs (Indiveri *et al.* 2006, Bofill-I-Petit and Murray 2004, Bamford *et al.* 2012b), which only able to produce synaptic weight changes for the STDP protocol, with only one spike pair, also utilised large capacitors, for the same reason.

However, with large capacitors, and even accelerated time, the leakage current still has a significant effect on the stored synaptic weight value. In the performed simulations throughout this thesis, the voltage difference between the synaptic weight values stored on the capacitor, at the start of the experiments and just after the experiment is finished, is reported. During the experiment, the leakage is not significant and can be compensated for, using the parameter tuning performed for the STDP circuit under test. However, after the experiment is finished, namely when there is no spike coming, the updated weight stored on the capacitor will leak away in less than a second. For an example, see the STDP measurement results from a similar accelerated-time neuromorphic chip reported in Wijekoon and Dudek (2012).

One of the possible approaches that can be employed beside any STDP circuit, including the proposed TSTDP circuits in this thesis is the use of a bistability circuit as discussed in Section 5.4.8. However, even with the use of a bistable mechanism, the final synaptic weight ought to be in a nonvolatile storage element for later use. Therefore, there is always need for long-term synaptic weight storage. There exist a number of nonvolatile weight storage methods in neuromorphic engineering such as (i) memory cells (Azghadi *et al.* 2013d), (ii) floating gate (Ramakrishnan *et al.* 2011), and (iii) memristive devices (Zamarreño-Ramos *et al.* 2011, Azghadi *et al.* 2013d), which could be utilised for this task. For further details see Section 5.4.8.

8.7 Chapter Summary

A low-power, compact, and tunable neuromorphic circuit with high synaptic plasticity capabilities was proposed in this chapter. Experimental results demonstrated how the proposed circuit can mimic the outcomes of various biological synaptic plasticity experiments. The presented design was compared with many previous synaptic plasticity circuits, including those proposed in previous chapters, in terms of power and area consumption, learning ability, and tolerance to transistor mismatch and process variation. The comparison shows that the proposed circuit possesses significant synaptic plasticity capabilities, which may potentially lead to neuromorphic systems with higher learning and computational performance. Besides, the comparison suggests that, this new design is low power and occupies small area, which are essential features in neuromorphic circuits. Furthermore, according to the performed comparisons, the new design shows better performance against the inevitable process variation in the VLSI fabrication process, compared to its counterparts. Because of all these features, this design can substitute previous synaptic plasticity circuit modules, e.g. the PSTDP device presented in Seo *et al.* (2011), and therefore can significantly improve the learning power, as well as the performance of previously developed neuromorphic systems in terms of power consumption, silicon real estate, and variation tolerance. The potential applications, future research directions, as well as the outlook of the learning circuits presented in this thesis are discussed in the next concluding chapter.

Chapter 9

Conclusion, Future Work, and Outlook

THIS chapter summarises the presented study and gives concluding remarks on the research carried out in this thesis. It highlights the original contributions the proposed research makes to the field of neuromorphic engineering, and discusses how the research conducted in this thesis extends the state of the art to elevate the ongoing research for realising a large-scale neuromorphic system with capabilities close to that of the brain. The chapter also provides ideas for future research to further boost neuromorphic engineering. It also states the author's outlook of the filed of neuromorphic engineering and learning in spiking neural networks.

9.1 Introduction

Neuromorphic engineering is concerned with the use of inherent properties of silicon devices to build low power and extremely compact systems for real-world engineering tasks, with similar performance to that of the biological systems (Mead 1989, Mead 1990). In a typical neuromorphic system, silicon neurons and synapses form a spiking neural network that can mimic the behaviour of biological spiking neural networks (Indiveri and Horiuchi 2011, Indiveri *et al.* 2011, Azghadi *et al.* 2014c). It is widely believed that synapses are the main apparatus for learning and computation in spiking neural networks (Sjöström *et al.* 2008). Synapses through their specific efficacy alteration rules, so-called synaptic plasticity process, control the way neurons behave (fire spikes). There exist a variety of neuron, synapse, and synaptic plasticity models, which have been successfully implemented in VLSI (Indiveri *et al.* 2011, Indiveri *et al.* 2006, Giulioni *et al.* 2009, Rachmuth *et al.* 2011, Mayr *et al.* 2010) and utilised in various applications (Arthur and Boahen 2006, Koickal *et al.* 2007, Mitra *et al.* 2009, Giulioni *et al.* 2009, Seo *et al.* 2011). The simulation of a number of synaptic plasticity rules have shown that these rules are efficient and powerful in mimicking biology and performing specific applications (Pfister and Gerstner 2006, Clopath and Gerstner 2010, Graupner and Brunel 2012). However, the implementation of such rules into the neuromorphic engineering field has yet to be fully explored. This thesis is aimed at exploring various unexplored synaptic plasticity rules and their applications in engineering tasks, when implemented in silicon.

This thesis discusses the design, implementation, application, and challenges of various spike-based synaptic plasticity rules in silicon, especially those that have not been explored yet (Azghadi *et al.* 2014c). It provides the reader with an insight on the previous and current states of VLSI synaptic plasticity circuits that have been utilised in different applications and proposes new VLSI designs and implementations for a novel STDP learning rule, that has not been presented in previous studies. The thesis also shows how this timing-based rule is able to give rise to a rate-based learning behaviour observed in previous studies (Pfister and Gerstner 2006). Furthermore, for the first time this timing-based rule is utilised to carry out a pattern classification task in a neuromorphic system (Azghadi *et al.* 2014b). The original contributions presented in different chapters of this thesis, which extend the state-of-the-art neuromorphic engineering research for implementing a high performance VLSI spiking neural network, capable of performing engineering tasks, are discussed in the following sections.

9.2 Implementing Spike Timing- and Rate-Based Synaptic Plasticity Rules on the IFMEM Device for Pattern Classification

As part of this thesis, a programmable hybrid analog/digital neuromorphic circuit, called IFMEM, that can be used to build compact low-power neural processing systems, was calibrated and programmed to operate correctly over a wide range of input frequencies; a feature that is essential for many applications, where interfacing to real-world sensors and systems is required (Liu and Delbrück 2010), or even when higher processing speed is required (Schemmel *et al.* 2006, Mayr *et al.* 2010, Azghadi *et al.* 2013a). The chip programmability is also essential for implementing various synaptic plasticity rules that is one of the main targets of this thesis.

9.2.1 Original Contributions

- Several novel experiments and measurements were performed on the IFMEM device to showcase the biological plausibility of the silicon neurons, as well as neuron-synapse combinations available on the device through specific software programs developed for these purposes. The conducted measurements demonstrate how specific behaviours can be generated by programming the chip and optimising neural parameters of the silicon neurons and synapses (Azghadi *et al.* 2013d). This feature is needed to utilise the chip to carry out various synaptic plasticity experiments as well as several other applications including a classification task using the TSTDP learning algorithm. The results related to this part that describe how the chip is programmed, as well as a description of the IFMEM device are presented in Azghadi *et al.* (2013d).

- Pair-based STDP rule was successfully implemented using silicon neurons and programmable synapses on the IFMEM chip. In order to test the correct functionality of the implemented STDP learning rule, the spiking neural network on the IFMEM chip was utilised to generate the well-known STDP learning window presented in both biological experiments (Bi and Poo 1998, Wang *et al.* 2005) as well as in computational studies (Song *et al.* 2000, Pfister and Gerstner 2006). The produced window correctly follows both experimental and computational data.

- Additionally, in order to further verify the functionality of the implemented STDP rule in a spiking neural network, a simple network comprised of several input synapses and one post-synaptic neuron were set in the same way they have been used in a computational experiment performed in Song *et al.* (2000). The results demonstrate that the implemented hardware neural network, can reproduce a competitive Hebbian learning behaviour similar to the one observed in computational STDP experiments (Song *et al.* 2000, Azghadi *et al.* 2014b).

- The triplet-based STDP learning algorithm was implemented on the IFMEM neuromorphic hardware. Following computational experiments presented in Izhikevich (2003) and Pfister and Gerstner (2006), a rate-based BCM learning behaviour was produced using the TSTDP learning rule, implemented on the IFMEM device (Azghadi *et al.* 2014b). The results show the excellent agreement between the implemented neural network outputs and the outcomes of BCM computational experiments (Pfister and Gerstner 2006).

- A perceptron like neural network was set up on the IFMEM device and the synapses and neurons were programmed and tuned in a way that the device acts as a pattern classification tool. The utilised learning algorithm for the performed pattern classification task was the triplet-based STDP learning rule, which is shown to be useful for learning and classification (Gjorgjieva *et al.* 2011) of rate-based patterns. Obtained results show the high performance of the TSTDP rule in real-time classification of complex correlated rate-based patterns (Azghadi *et al.* 2014b).

  The preformed research in this part provides good view of the STDP and TSTDP rules and their properties and features, which are essential when designing VLSI STDP synapses. The above mentioned original contributions were described in detail in Azghadi *et al.* (2014c).

9.2.2 Future Work

- The use of the AER representation for receiving inputs, computing with spikes, and transmitting signals in output, makes the IFMEM device an ideal computational platform for building embedded neuromorphic event-based computational systems that process events generated by neuromorphic sensory systems (Liu and Delbrück 2010). Therefore, in a future work the programmed IFMEM

chip that has been already adapted with the TSTDP (or can be adapted to any other learning rule), can be interfaced to a neuromorphic sensory system such as the Dynamic Vision Sensor (DVS) presented in Lichtsteiner *et al.* (2008).

- In the performed experiments, only PSTDP, TSTDP and BCM learning rules were implemented and tested using the IFMEM device. However, in future studies, any synaptic plasticity rule of choice including complex and detailed biophysically grounded rules (Shouval 2011) as well as other simple or complicated phenomenological rules (Clopath and Gerstner 2010, Graupner and Brunel 2012, Uramoto and Torikai 2013) can be implemented and tested on the IFMEM device. This will provide us with a good comparison of various synaptic plasticity rules performance in carrying out different applications while they are interfaced to silicon neurons and have been utilised in a network configuration. However, a current limitation of the IFMEM device is its limited number of neurons and synapses (32 neurons and 1K synapses), which restricts the extension of the required neural network size. To address this limitation, currently new programmable neuromorphic devices are being developed in the NCS group of INI. For example see the characteristics of the newly developed MNR256R1 chip briefly described in Chapter 7.

9.3 Spike-based Synaptic Plasticity in Silicon: Design, Implementation, Application and Challenges

This thesis provides insight into opportunities and challenges of the implementation of various synaptic plasticity rules in silicon. It reviews, describes, discusses and proposes various analog VLSI designs for different spike timing- and rate-based synaptic plasticity rules and highlights their limitations and benefits (Azghadi *et al.* 2014c). Below is a summary of several original contributions made on this aspect.

9.3.1 PSTDP VLSI Learning Circuits

The PSTDP learning algorithm has been implemented by various groups and under different design strategies (Bofill-I-Petit and Murray 2004, Cameron *et al.* 2005, Indiveri *et al.* 2006, Koickal *et al.* 2007, Tanaka *et al.* 2009, Bamford *et al.* 2012b). In

this thesis, for the first time a number of these PSTDP designs were tested if they are able to generate several synaptic plasticity experiments including the pairing experiment for generating STDP learning window (Bi and Poo 1998, Wang *et al.* 2005), triplet (Wang *et al.* 2005), quadruplet (Wang *et al.* 2005) and frequency-dependent pairing experiments (Sjöström *et al.* 2001). In addition, two various PSTDP learning circuits, as representative for the class of PSTDP circuits, are also tested for their abilities to mimic a rate-based BCM learning behaviour with sliding threshold feature (Bienenstock *et al.* 1982, Cooper *et al.* 2004).

Original Contributions

- Various synaptic plasticity experimental results presented in Azghadi *et al.* (2011c) and Azghadi *et al.* (2012b) show that the two sample PSTDP designs presented in Bofill-I-Petit and Murray (2004) and Indiveri *et al.* (2006) are not able to account for triplet, quadruplet and frequency-dependent pairing experiments, but they can successfully generate the STDP learning window (Song *et al.* 2000), as reported in relevant papers (Bofill-I-Petit and Murray 2004, Indiveri *et al.* 2006).

- The previous VLSI implementation of the PSTDP rule presented in Indiveri *et al.* (2006) was simplified to reduce area and power consumption. The result of this study is presented in Azghadi *et al.* (2011b). This design is unable to account for the mentioned triplet, quadruplet, and frequency-dependent pairing experiments, since it also implements the PSTDP rule, which according to the computational studies cannot account for these experiments (Pfister and Gerstner 2006).

- For the first time a PSTDP learning circuit was utilised to generate a rate-based BCM learning behaviour under a Poissonian protocol described in Section 2.5.6. The results presented in Azghadi *et al.* (2012a) demonstrate that the PSTDP learning circuit, similar to the PSTDP learning computational model (Izhikevich and Desai 2003), can generate a rate-based BCM learning behaviour.

Future Work

- The mentioned higher order spike experimental protocols, i.e. triplet, quadruplet and frequency-dependent pairing protocols, were all used to stimulate the previous and current PSTDP VLSI circuits (Azghadi *et al.* 2011c, Azghadi *et al.* 2012b) and show that these circuits are unable to account for several biological experiments, due to their weakness in processing higher order spike combinations.

However, in future studies, one might investigate the ability of some other spike-based synaptic plasticity models/VLSI designs in reproducing the outcomes of various timing-based (Wang *et al.* 2005, Pfister and Gerstner 2006) as well as hybrid rate/timing-based biological experiments (Sjöström *et al.* 2001). One of the main synaptic plasticity rules discussed and reviewed in this thesis is the SDSP learning model (Fusi *et al.* 2000, Brader *et al.* 2007), which has two various VLSI implementations presented in Fusi *et al.* (2000) and Mitra *et al.* (2009). A future research direction is to investigate the ability of this rule and its variant VLSI implementations for mimicking the outcome of complicated timing- and rate-based experiments.

9.3.2 TSTDP VLSI Learning Circuits

The first VLSI designs for the triplet-based STDP learning circuit were proposed in this thesis. These circuits were devised to overcome the deficiencies of PSTDP circuits in synaptic plasticity experiments. Below sections provide a summary of original contributions made in this relation.

Original Contributions

- The first VLSI design for the TSTDP learning rule was proposed. The new proposed voltage-mode circuit presented in Azghadi *et al.* (2011d), is able to account for many biological experiments, where the previous PSTDP circuits clearly fail. It was first shown that this circuit is able to mimic the outcomes of a wide range of synaptic plasticity experiments including timing-based, hybrid rate/timing-based, and rate-based synaptic plasticity experiments (Azghadi *et al.* 2011d).

- In another study (Azghadi *et al.* 2011c), the proposed voltage-mode TSTDP and a previous voltage-mode PSTDP VLSI design proposed by Indiveri *et al.* (2006) were optimised and simulated under same experimental protocols and conditions, to reproduce the outcome of various synaptic plasticity experiments. The comparison of the results show that the TSTDP design significantly outperforms the PSTDP design in closely mimicking the experimental data (Azghadi *et al.* 2011c).

- Furthermore, the proposed voltage-mode TSTDP circuit was used (Azghadi *et al.* 2011a) to mimic a similar behaviour to the outcomes of a rate-based BCM experiment (Izhikevich and Desai 2003, Pfister and Gerstner 2006). The achieved results demonstrate that this circuit closely mimics the sliding threshold behaviour of the BCM rule (Azghadi *et al.* 2011a). In addition, the performance of the previous voltage-based PSTDP circuit presented in Indiveri *et al.* (2006) in reproducing the BCM-like behaviour was also compared to the proposed TSTDP circuit. The comparison shows that the TSTDP circuit has higher ability in mimicking the required BCM-like behaviour (Azghadi *et al.* 2012a).

- Further investigations on the proposed voltage-mode TSTDP circuit showed that this circuit, similar to its PSTDP counterpart, cannot account for the observed exponential-like weight changes in the original PSTDP experiments (Bi and Poo 1998) or in the computational modelling of PSTDP rule (Song *et al.* 2000). Therefore, a new current-mode synaptic plasticity circuit was proposed that built upon the PSTDP design proposed by Bofill-I-Petit and Murray (2004) and can efficiently account for the required exponential behaviour (Azghadi *et al.* 2012b). Similar to the first TSTDP circuit design, this circuit was also verified for its ability in reproducing the outcomes of various experiments such as triplet, quadruplet, pair-based, frequency-dependent pair-based and BCM-like experiments. The simulation results presented in Azghadi *et al.* (2012b) testify to the ability of this new design in efficiently reproducing the outcomes of all these experiments. In Azghadi *et al.* (2012b), we also compared the performance of the previous PSTDP and the proposed voltage- and current-mode TSTDP circuits in reproducing the experimental data, and highlighted the higher ability of the new TSTDP design.

- Extra investigations and simulations on both proposed TSTDP circuits presented in Azghadi *et al.* (2011d) and Azghadi *et al.* (2012b) show that none of these circuits are able to closely replicate the outcomes of some other synaptic plasticity experiments performed on all possible combinations of spike triplets Froemke and Dan (2002). In addition, these designs cannot account for BCM experimental data presented in Kirkwood *et al.* (1996), which drives the TSTDP circuit pre- and not post-synaptically. Therefore, a new high performance design for the TSTDP rule was proposed (Azghadi *et al.* 2013a), which not only accounts for all the previous experiments, but is also able to reproduce the outcomes of these more

complicated new experiments. The new proposed TSTDP circuit has a significantly better performance in mimicking the various experiments in comparison to the two previously developed TSTDP circuits, and shows lower data fitting error (i.e. NMSE) as discussed in Chapter 7. It is shown how the new TSTDP circuit is able to account for all previously mentioned experiments with higher performance compared to the previous PSTDP and TSTDP designs, and also can closely mimic the experimental data in new experiments such as experiments involved with various spike triplet combinations, as well as pre-synaptic and post-synaptic driven rate-based BCM-like experiments (Azghadi *et al.* 2013a), where the previous TSTDP and PSTDP designs do not show suitable performance and cannot mimic the experiments effectively.

- The new TSTDP circuit uses subthreshold transistors to reach the required exponential behaviour that is needed in the TSTDP computational model (Pfister and Gerstner 2006). Therefore, the circuit is inherently prone to process variations and device mismatch due to the imperfect fabrication processes (Poon and Zhou 2011, Azghadi *et al.* 2014c). In order to investigate the susceptibility of the proposed design against device mismatch, the design underwent a severe device mismatch verification in 1000 MC simulation runs. The presented results in Azghadi *et al.* (2013a) show that although the circuit is susceptible to process variation, the effect of variations can be mitigated through a post-fabrication calibration technique to bring the circuit back to its desired behaviour even in the presence of severe variations (Azghadi *et al.* 2013a).

- Since this new design enjoys a high-performance structure and is able to efficiently reproduce all the required synaptic plasticity experimental data, it was chosen to be fabricated in silicon. A proof of concept TSTDP circuit device was successfully fabricated and tested as part of this thesis. The chip measurement results presented in Section 7.6, show the correct functionality of the circuit that was fabricated in an AMS 0.18 μm CMOS technology.

- The presented simulation results for the high-performance TSTDP circuit are performed with the circuit including a 10 pF weight capacitor, which occupies a very large portion of the proposed circuit. This large capacitor is needed to maintain the synaptic weight value for long period of time, required to replicate the experimental data, under the same circumstances as those utilised in the original

biological experiments (Pfister and Gerstner 2006). In order to address the problem of the very large capacitor, a new compact TSTDP circuit that uses a small 50 fF weight capacitor was proposed (Azghadi *et al.* 2013c). Due to the small capacitor used in this design, it is quite suitable for short term plasticity or when processing inputs with high frequency. However, for lower spike frequencies and long term plasticity, the latest updated value across the capacitor will be lost during the circuit operation. In addition, this design is not able to account for the hybrid rate/timing-based experiments presented in Sjöström *et al.* (2001), due to its limited capacitor size (Azghadi *et al.* 2013c).

- The mentioned limitations in the design presented in Azghadi *et al.* (2013c) resulted in utilising a large capacitor of 1 pF size. This design only uses a 1 pF capacitor and 18 transistors, and in comparison with the previous TSTDP design presented in Azghadi *et al.* (2013a), which uses 37 transistors, a 10 pF weight capacitor, as well as four 100 fF time constant capacitors, needs a significantly smaller silicon area. The presented results in Azghadi *et al.* (2013c) show that the new compact circuit can account for the hybrid rate/timing-based experiments presented in Sjöström *et al.* (2001).

- Further investigations on this new compact circuit show that a 14-transistor minimal version of this TSTDP circuit is able to account for the full set of experiments reviewed in Section 2.5. This minimal circuit which is developed based on the minimal TSTDP rule (Pfister and Gerstner 2006) has only two transistors more than its smallest PSTDP counterpart presented in Indiveri *et al.* (2006). The proposed design is not only more compact than all other previous TSTDP (Azghadi *et al.* 2011c, Azghadi *et al.* 2012b, Azghadi *et al.* 2013a) and many of the previous PSTDP designs (Bofill-I-Petit and Murray 2004, Cameron *et al.* 2005, Koickal *et al.* 2007, Tanaka *et al.* 2009, Bamford *et al.* 2012b) in the literature, but also it consumes lesser power than all these designs (Azghadi *et al.* 2014a). For further details please refer to Table 8.3.

- The proposed compact circuit is also investigated and compared to other designs in terms of tolerance to mismatch and process variation (Azghadi *et al.* 2014a). Monte Carlo simulation results show that the proposed design, due to its circuit structure, is much more stable than its previous counterparts (Azghadi *et al.*

2011d, Azghadi *et al.* 2012b, Azghadi *et al.* 2012b) in terms of vulnerability to transistor mismatch, which is a significant challenge in analog neuromorphic design (Azghadi *et al.* 2014a).

Future Work

- Presented variation analysis on the proposed TSTDP designs in Azghadi *et al.* (2012b), Azghadi *et al.* (2013a), Azghadi *et al.* (2013c), and Azghadi *et al.* (2014a) suggest that similar to all other neuromorphic subthreshold analog VLSI designs, the performance of these designs can also significantly degrade due to the inevitable device mismatch and fabrication imperfections (Poon and Zhou 2011). Although it was shown in Azghadi *et al.* (2013a) that the designs can be fine-tuned after fabrications to reach their initial targeted performance, this approach is not applicable in large-scale neuromorphic systems with millions of synapses (Poon and Zhou 2011). Therefore, working toward implementing synaptic plasticity circuits including TSTDP circuits that are prone to device mismatch is an essential direction for future research.

  As discussed in Section 5.4.2, various approaches have been proposed to tackle the process variation and mismatch problems in neuromorphic systems. Among these approaches the mismatch minimisation technique utilising wide-dynamic range devices (Rachmuth and Poon 2008, Rachmuth *et al.* 2011), as well as the off-chip event-based compensation strategies (Choudhary *et al.* 2012) such as the use of AER mappers and routers (e.g., probabilistic) to re-distribute events in a way to compensate for mismatch effects, are the viable methods for reducing the effect of mismatch in large-scale neuromorphic systems.

  In a future work, wide dynamic range devices can be utilised to implement the circuit and therefore minimise the mismatch effect (Rachmuth *et al.* 2011). In addition, off-chip event-based mismatch compensation strategies (Choudhary *et al.* 2012) are other viable methods useful for utilising the proposed synaptic plasticity element in a large-scale neuromorphic system.

- Synaptic weight storage is another significant challenge in the design of various synaptic plasticity circuits. It was discussed in Section 5.4.8 that a variety of approaches are available to be utilised in synaptic circuits for synaptic weight

storage. However, each of these approaches has its own limitations and advantages. The proposed synaptic circuits employed the accelerated time design strategy to minimise the effect of leakage on the synaptic weight capacitor. This approach has been utilised in many previous neuromorphic designs such as the design presented in Schemmel *et al.* (2006), Tanaka *et al.* (2009), Schemmel *et al.* (2010), Mayr *et al.* (2010), and Wijekoon and Dudek (2012). Although this approach can be useful in applications were high synaptic plasticity update speed is required, it has the limitation of requiring higher bandwidth for spike communication (Schemmel *et al.* 2010). In addition, a neuromorphic system utilising this technique cannot simply be interfaced to sensory systems with biologically plausible time constants.

A future research direction is to implement the proposed TSTDP circuits in the biological time scale and employ a bistable circuitry for stabilising the synaptic weight (Indiveri *et al.* 2006, Mitra *et al.* 2009). However, one must keep in mind that the bistable circuit is a volatile weight storage technique and the final synaptic weight needs to be stored in a non-volatile memory for future references before the system is powered down (Azghadi *et al.* 2014c).

- The synaptic weight in the proposed circuit is updated with the arrival of each spike which leads to charging/discharging the weight capacitor, according to the TSTDP rule. However, if there is no spike coming, or when the learning phase has finished, the final synaptic weight must be stored for later use. Therefore, there is an essential need for a non-volatile memory element to store the latest weight. Many neuromorphic systems utilise memory cells and DACs to store and restore the synaptic weight when required (Arthur and Boahen 2006, Seo *et al.* 2011, Pfeil *et al.* 2012, Azghadi *et al.* 2014b). In addition, Ramakrishnan *et al.* (2011) have used a floating gate device to store the synaptic weight in a non-volatile fashion on their single transistor synaptic device. Furthermore, in a recent study we have used the non-volatile characteristic of memristor to implement a programmable DAC (Azghadi *et al.* 2013d).

  In future research, the TSTDP learning algorithm can be investigated using similar techniques for storing the synaptic weight. Therefore, one may utilise SRAM cells along with ADC circuits and record the latest weight in the memory once there is no more input spikes or after the learning phase finished or need to be

stopped (Mitra *et al.* 2009). Also, one might utilise memristor features to implement the TSTDP rule, in a similar way to that utilised in Zamarreño-Ramos *et al.* (2011), for implementing PSTDP. Furthermore, the TSTDP rule can be implemented using two instances of the single transistor floating gate synaptic device presented in Ramakrishnan *et al.* (2011).

- The developed knowledge gained through the course of the presented project resulted in a reliable VLSI chip including a prototype TSTDP circuit that has been tested and shown desired behaviour while being stimulated with artificial input spikes. This circuit is now ready to be interfaced to silicon neurons and other spike-based devices including neuromorphic sensory systems (Lichtsteiner *et al.* 2008, Liu and Delbrück 2010), for engineering applications such as the pattern classification task carried out in the IFMEM device (Azghadi *et al.* 2014b).

9.4 Outlook

As already mentioned, synaptic plasticity is believed to be responsible for acquiring computational capabilities, learning and memory in the brain. It is critical to understand the underlying mechanisms of the plasticity rules and their computational role before utilising them for learning and processing in real-world applications. Recent advances in VLSI technology, combined with progress in experimental neuroscience and neuromorphic circuit design techniques, have led to useful implementations of these rules in hardware. However, most of these implementations can only be applied to demonstrate proofs of principles. To successfully apply neuromorphic circuits in real-world applications, potentially replacing or enhancing some of the conventional technology and approaches being used today, requires the development of large-scale neuromorphic systems that go beyond single chip, or single core solutions (Hasler and Marr 2013). One of the most challenging tasks that needs to be addressed to achieve this is therefore the inter-chip, or inter-module communication. Currently, both single-wafer and multi-core or multi-chip solutions based on asynchronous logic are being investigated (Scholze *et al.* 2011, Imam *et al.* 2012, Merolla *et al.* 2013). In addition promising emerging technologies such as 3D VLSI and memristors (Zamarreño-Ramos *et al.* 2011, Eshraghian *et al.* 2012, Sheridan and Lu 2014) may provide efficient solutions to this problem.

Another open challenge that is hindering progress in the design of large scale neuromorphic systems is the lack of appropriate EDA tools to assists neuromorphic designers in the design, verification, and testing phases. As already mentioned in Chapter 5, currently there are several promising design automation tools for generating asynchronous logic circuits that are helpful for designing interconnecting circuits in large-scale neuromorphic systems, but further developments for mixed analog/digital design tools is needed. The area requirement for synaptic weight storage is another challenge for large-scale neuromorphic systems. This can be addressed with the use of newly developed resistive memory elements, which are integrable with CMOS technology, occupy small area, and consume little power (Indiveri *et al.* 2013). However, these resistive elements are susceptible to variations and suffer from low yields, which should be effectively addressed before utilising them in large-sale systems.

All these and other mentioned challenges are currently being addressed by an active and enthusiastic research community. The small group of neuromorphic engineers that was once limited to a dozen research laboratories around the world in the mid 90s is now flourishing, with many more groups spread around the whole globe, and with increasing support from both research funding organisations and strong industrial microelectronic groups.

In general, with the many efforts and initiatives that are being started in the field of neuromorphic engineering, the future of this field is very promising, and the ongoing research on implementations of learning mechanisms in neuromorphic systems is likely to lead to systems that can be used in real-world applications in the near future.

Appendix A

Extra Investigations on the Ability of the Proposed High-performance TSTDP Circuit for Producing BCM-like Behaviour

THE influence of high and low pre-synaptic rates on the BCM-like behaviour produced using the proposed high-performance TSTDP circuit that was presented in Chapter 7, is investigated. It is also verified if the desired BCM-like behaviour is possible when a neuron is integrated with the TSTDP synapse. The Izhikevich neuron model as well as a linear Poisson neuron model are simulated along with a TSTDP synapse and the resulting weight changes under the Poissonian protocol mentioned in Section 2.5.6 are recorded. The simulation results show that, in all cases, a well-shaped BCM-like behaviour with distinguishable sliding thresholds can be achieved.

A.1 Introduction

According to the literature, STDP and BCM rules are related and a BCM-like weight modification behaviour with sliding threshold that depends on the rate of the pre- or post-synaptic spike trains, is an emergent property of the STDP rules such as PSTDP (Izhikevich 2003) and TSTDP (Pfister and Gerstner 2006, Gjorgjieva *et al.* 2011). In both cases of these timing-based plasticity rules, a BCM-like behaviour emerges when synaptic weight modification changes are reported against changes in the post-synaptic spike train rates. In this case, the sliding threshold of the BCM rule depends on the parameters of the utilised STDP model and can be modified accordingly as described in Section 7.4.6 for TSTDP and in Izhikevich (2003) for PSTDP. The BCM protocol in this case, which involves in sweeping the post-synaptic rate and recording the changes in the synaptic weight accordingly, is mainly used in computational modelling of the synaptic plasticity rules (Izhikevich 2003, Pfister and Gerstner 2006, Gjorgjieva *et al.* 2011). However, in the original BCM experiments performed by Kirkwood *et al.* (1996), the pre-synaptic spike train rate is swept, while the post-synaptic firing rate is determined by the current synaptic weight and the dynamics of the neuron. Hence, in order to test the response of the proposed circuit that implements a TSTDP model, while it is pre-synaptically driven, a neuron model is required. Here, it is shown that using two different neuron models, a BCM-like behaviour is achievable, when the TSTDP model (circuit) is pre-synaptically (in contrary to the post-synaptically) driven.

In addition, in the performed post-synaptically driven simulations in this thesis (see Figs. 7.8 and 8.9), the pre-synaptic firing rate was always kept fixed at 10 Hz, while the post-synaptic rate swept over a range of frequency. One might wonder, if an increase or a decrease in the pre-synaptic firing rate has any effect on the BCM-like behaviour of the proposed TSTDP rule. Therefore, here we investigated how the changes in the pre-synaptic firing rate in a post-synaptically driven TSTDP synapse affects the desired BCM-like behaviour. These extra investigations gives us deeper insight into these emerging properties of the proposed TSTDP circuits and is useful in utilising these circuits in various tasks such as pattern selection (Gjorgjieva *et al.* 2011).

A.2 Post-synaptically Driven BCM-like Behaviour

Fig. A.1 illustrates the outcome of our circuit simulations when subject to the same Poissonian protocol as used by Pfister and Gerstner (2006) and described in 7.4.6. In

this figure, each data point at each post-synaptic frequency (ρ_{post}), is the average value of the weight changes for ten different realisations of post-synaptic and pre-synaptic Poissonian spike trains, where the error bar indicates the standard deviation.

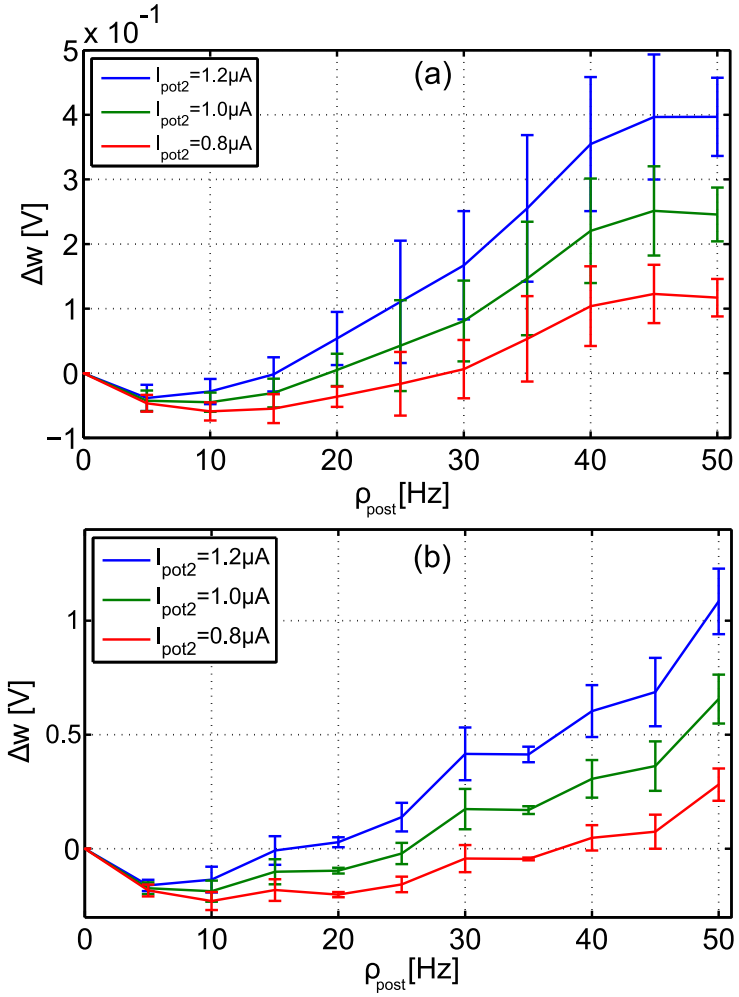


Figure A.1. The proposed high-performance TSTDP circuit can generate BCM-like behaviour for various pre-synaptic spike rates. The three different curves show the synaptic weight changes according to three different synaptic modification thresholds. The thresholds that are controlled by the current I_{pot2}, demonstrate the points where LTD changes to LTP. The rate of pre-synaptic spike trains, ρ_{pre}, used in (a) and (b) was 5 and 15 Hz, respectively. Each data point shows the mean value of the weight changes for 10 different trials and the error bars depict the standard deviations of the weight changes for each value of ρ_{post}.

The demonstrated results were produced using the bias currents that correspond to the visual cortex data set (see Table 7.1 for these values). The three different curves presented in Fig. A.1(a-b) display three different weight modification thresholds. These curves are in the results of three different values for I_{pot2} currents that correspond to

three different values of A_3^+. These thresholds are related to the post-synaptic firing rate, ρ_{post}, for the rates up to 50 Hz, akin to previously reported results in Pfister and Gerstner (2006). The simulation results show that if the mean pre-synaptic firing rate decreases to 5 or increases to 15 Hz (in comparison to 10 Hz in the original experiments), the post-synaptically driven BCM-like behaviour can be still preserved.

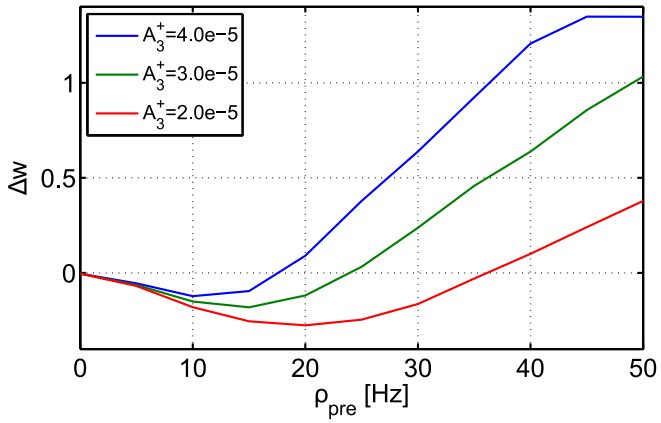


Figure A.2. Pre-synaptically driven BCM-like behaviour from Matlab simulations for the linear Poisson neuron model. This figure shows the synaptic weight changes produced by the minimal TSTDP model (shown in Eq. 2.5), when integrated with a linear Poissonian neuron. The three different curves represent three different BCM thresholds, which are controlled by A_3^+ parameter of the TSTDP rule.

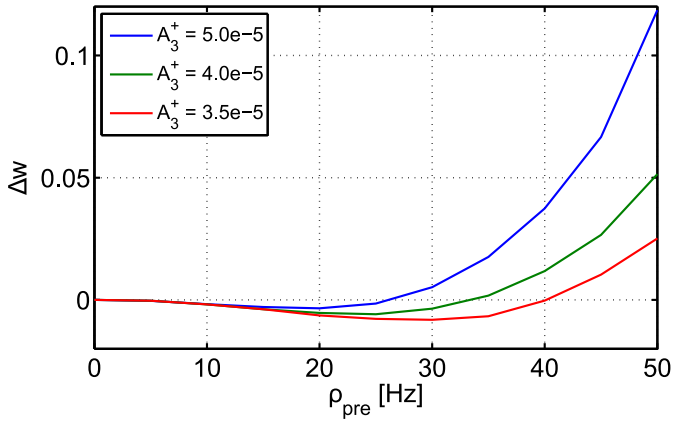


Figure A.3. Pre-synaptically driven BCM-like behaviour from Matlab simulations for the Izhikevich's neuron model. This figure shows the synaptic weight changes produced by the minimal TSTDP model (shown in Eq. 2.5), when integrated with an Izhikevich neuron. The three different curves represent three different BCM thresholds, which are controlled by A_3^+ parameter of the TSTDP rule.

A.3 Pre-synaptically Driven BCM-like Behaviour

Additional Matlab simulations were performed to assess if pre-synaptically driven minimal TSTDP model leads to BCM-like synaptic weight changes using both the linear Poisson neuron model and the Izhikevich neuron model. We found that in the case of increasing the pre-synaptic activity, the resulting synaptic weight changes followed a BCM-like profile, regardless of which neuron model was used. Such a pre-synaptically driven BCM-like profile of synaptic change occurs for each above stated neuron model and the results of these simulations are presented in Figs. A.2 and A.3. Note that the three different curves in these figures correspond to different thresholds controlled by the A_3^+ parameter.

Bibliography

ABARBANEL-H. D., HUERTA-R., AND RABINOVICH-M. (2002). Dynamical model of long-term synaptic plasticity, *Proceedings of the National Academy of Sciences*, **99**(15), pp. 10132–10137.

ABBOTT-L. F., AND NELSON-S. B. (2000). Synaptic plasticity: taming the beast, *Nature Neuroscience*, **3**, pp. 1178–1183.

AFIFI-A., AYATOLLAHI-A., AND RAISSI-F. (2009). Implementation of biologically plausible spiking neural network models on the memristor crossbar-based CMOS/nano circuits, *European Conference on Circuit Theory and Design*, pp. 563–566.

AL-SARAWI-S., ABBOTT-D., AND FRANZON-P. (1998). A review of 3-D packaging technology, *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging*, **21**(1), pp. 2–14.

AMIT-D. J., AND FUSI-S. (1994). Learning in neural networks with material synapses, *Neural Computation*, **6**(5), pp. 957–982.

ANANTHANARAYANAN-R., ESSER-S. K., SIMON-H. D., AND MODHA-D. S. (2009). The cat is out of the bag: cortical simulations with 109 neurons, 1013 synapses, *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, DOI: 10.1145/1654059.1654124*.

ANDREOU-A. G., BOAHEN-K. A., POULIQUEN-P. O., PAVASOVIC-A., JENKINS-R. E., AND STROHBEHN-K. (1991). Current-mode subthreshold MOS circuits for analog VLSI neural systems, *IEEE Transactions on Neural Networks*, **2**(2), pp. 205–213.

ARENA-P., FORTUNA-L., FRASCA-M., AND PATANÉ-L. (2009). Learning anticipation via spiking networks: application to navigation control, *IEEE Transactions on Neural Networks*, **20**(2), pp. 202–216.

ARENA-P., FORTUNA-L., FRASCA-M., PATANÉ-L., AND SALA-C. (2007). Integrating high-level sensor features via STDP for bio-inspired navigation, *IEEE International Symposium on Circuits and Systems*, pp. 609–612.

ARTHUR-J., MEROLLA-P., AKOPYAN-F., ALVAREZ-R., CASSIDY-A., CHANDRA-S., ESSER-S., IMAM-N., RISK-W., RUBIN-D., MANOHAR-R., AND MODHA-D. (2012). Building block of a programmable neuromorphic substrate: A digital neurosynaptic core, *The 2012 International Joint Conference on Neural Networks, DOI: 10.1109/IJCNN.2012.6252637*.

ARTHUR-J. V., AND BOAHEN-K. (2004). Recurrently connected silicon neurons with active dendrites for one-shot learning, *2004 IEEE International Joint Conference on Neural Networks*, Vol. 3, pp. 1699–1704.

ARTHUR-J. V., AND BOAHEN-K. (2006). Learning in silicon: Timing is everything, *Advances in Neural Information Processing Systems*, pp. 75–82.

AZGHADI-M. R., AL-SARAWI-S., ABBOTT-D., AND IANNELLA-N. (2013a). A neuromorphic VLSI design for spike timing and rate based synaptic plasticity, *Neural Networks*, **45**, pp. 70–82.

AZGHADI-M. R., AL-SARAWI-S., ABBOTT-D., AND IANNELLA-N. (2013b). Pairing frequency experiments in visual cortex reproduced in a neuromorphic STDP circuit, *2013 IEEE International Conference on Electronics, Circuits, and Systems*, pp. 229–232.

AZGHADI-M. R., AL-SARAWI-S., IANNELLA-N., AND ABBOTT-D. (2011a). Emergent BCM via neuromorphic VLSI synapses with STDP, *5th Australian Workshop on Computational Neuroscience*, p. 31.

AZGHADI-M. R., AL-SARAWI-S., IANNELLA-N., AND ABBOTT-D. (2011b). Physical implementation of pair-based spike-timing-dependent plasticity, *EPSM-ABEC 2011 Conference*, Vol. 34, art. no. 141.

AZGHADI-M. R., AL-SARAWI-S., IANNELLA-N., AND ABBOTT-D. (2012a). Design and implementation of BCM rule based on spike-timing dependent plasticity, *Proc. IEEE 2012 International Joint Conference on Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2012.6252778*.

AZGHADI-M. R., AL-SARAWI-S., IANNELLA-N., AND ABBOTT-D. (2012b). Efficient design of triplet based spike-timing dependent plasticity, *Proc. IEEE 2012 International Joint Conference on Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2012.6252820*.

AZGHADI-M. R., AL-SARAWI-S., IANNELLA-N., AND ABBOTT-D. (2013c). A new compact analog VLSI model for spike timing dependent plasticity, *2013 IFIP/IEEE 21st International Conference on Very Large Scale Integration (VLSI-SoC)*, pp. 7–12.

AZGHADI-M. R., AL-SARAWI-S., IANNELLA-N., AND ABBOTT-D. (2014a). Tunable low energy, compact and high performance neuromorphic circuit for spike-based synaptic plasticity, *PLoS ONE*, **9**(2), p. art. no. e88326.

AZGHADI-M. R., AL-SARAWI-S., IANNELLA-N., INDIVERI-G., AND ABBOTT-D. (2014b). Spike-based synaptic plasticity in silicon: Design, implementation, application, and challenges, *Proceedings of the IEEE*, **102**(5), pp. 717–737.

AZGHADI-M. R., BONYADI-M. R., AND SHAHHOSSEINI-H. (2007). Gender classification based on feedforward backpropagation neural network, *Artificial Intelligence and Innovations 2007: from Theory to Applications*, Springer US, pp. 299–304.

AZGHADI-M. R., BONYADI-M. R., HASHEMI-S., AND MOGHADAM-M. E. (2008). A hybrid multiprocessor task scheduling method based on immune genetic algorithm, *Proceedings of the 5th International ICST Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*, pp. 561–564.

AZGHADI-M. R., KAVEHEI-O., AL-SARAWI-S., IANNELLA-N., AND ABBOTT-D. (2011c). Novel VLSI implementation for triplet-based spike-timing dependent plasticity, *Proceedings of the IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 158–162.

AZGHADI-M. R., KAVEHEI-O., AL-SARAWI-S., IANNELLA-N., AND ABBOTT-D. (2011d). Triplet-based spike-timing dependent plasticity in silicon, *The 21st Annual Conference of the Japanese Neural Network Society*, Japanese Neural Network Society, pp. P34–35.

AZGHADI-M. R., MORADI-S., AND INDIVERI-G. (2013d). Programmable neuromorphic circuits for spike-based neural dynamics, *11th IEEE International New Circuit and Systems (NEWCAS) Conference, DOI: 10.1109/NEWCAS.2013.6573600*.

AZGHADI-M. R., MORADI-S., FASTNACHT-D., OZDAS-M. S., AND INDIVERI-G. (2014c). Programmable spike-timing dependent plasticity learning circuits in neuromorphic VLSI architectures, *ACM Journal on Emerging Technologies in Computing Systems, submitted*.

BADOUAL-M., ZOU-Q., DAVISON-A. P., RUDOLPH-M., BAL-T., FRÉGNAC-Y., AND DESTEXHE-A. (2006). Biophysical and phenomenological models of multiple spike interactions in spike-timing dependent plasticity, *International Journal of Neural Systems*, **16**(02), pp. 79–97.

BAMFORD-S. A., MURRAY-A. F., AND WILLSHAW-D. J. (2010). Synaptic rewiring for topographic mapping and receptive field development, *Neural Networks*, **23**(4), pp. 517–527.

BAMFORD-S. A., MURRAY-A. F., AND WILLSHAW-D. J. (2012a). Silicon synapses self-correct for both mismatch and design inhomogeneities, *Electronics Letters*, **48**(7), pp. 360–361.

BAMFORD-S., MURRAY-A., AND WILLSHAW-D. (2012b). Spike-timing dependent plasticity with weight dependence evoked from physical constraints, *IEEE Transactions on Biomedical Circuits and Systems*, **6**(4), pp. 385–398.

BARTOLOZZI-C., AND INDIVERI-G. (2007). Synaptic dynamics in analog VLSI, *Neural Computation*, **19**(10), pp. 2581–2603.

BARTOLOZZI-C., MITRA-S., AND INDIVERI-G. (2006). An ultra low power current–mode filter for neuromorphic systems and biomedical signal processing, *Proc. IEEE Biomedical Circuits and Systems Conference, (BioCAS)*, pp. 130–133.

BELATRECHE-A., MAGUIRE-L., AND MCGINNITY-M. (2006). Advances in design and application of spiking neural networks, *Soft Computing*, **11**(3), pp. 239–248.

BELHADJ-B., TOMAS-J., BORNAT-Y., DAOUZLI-A., MALOT-O., AND RENAUD-S. (2009). Digital mapping of a realistic spike timing plasticity model for real-time neural simulations, *Proc. XXIV Conference on Design of Circuits and Integrated Systems (DCIS)*, pp. 326–331.

BENUSKOVA-L., AND ABRAHAM-W. C. (2007). STDP rule endowed with the BCM sliding threshold accounts for hippocampal heterosynaptic plasticity, *Journal of Computational Neuroscience*, **22**(2), pp. 129–133.

BIENENSTOCK-E., COOPER-L., AND MUNRO-P. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex, *The Journal of Neuroscience*, **2**(1), pp. 32–48.

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

BILL-J., SCHUCH-K., BRÜDERLE-D., SCHEMMEL-J., MAASS-W., AND MEIER-K. (2010). Compensating inhomogeneities of neuromorphic VLSI devices via short-term synaptic plasticity, *Frontiers in Computational Neuroscience, DOI: 10.3389/fncom.2010.00129*.

BLISS-T. V., AND COLLINGRIDGE-G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus, *Nature*, **361**(6407), pp. 31–39.

BOAHEN-K. A. (2000). Point-to-point connectivity between neuromorphic chips using address events, *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, **47**(5), pp. 416–434.

BOFILL-A., MURRAY-A. F., AND THOMPSON-D. P. (2001). Circuits for VLSI implementation of temporally-asymmetric hebbian learning, *Advances in Neural Information Processing Systems (NIPS)*, Vol. 9, pp. 1091–1098.

BOFILL-I-PETIT-A., AND MURRAY-A. (2004). Synchrony detection and amplification by silicon neurons with STDP synapses, *IEEE Transaction on Neural Networks*, **15**(5), pp. 1296–1304.

BONYADI-M. R., AZGHADI-S. M. R., AND HOSSEINI-H. S. (2007). Solving traveling salesman problem using combinational evolutionary algorithm, *Artificial Intelligence and Innovations 2007: from Theory to Applications*, Springer US, pp. 37–44.

BRADER-J., SENN-W., AND FUSI-S. (2007). Learning real-world stimuli in a neural network with spike-driven synaptic dynamics, *Neural Computation*, **19**(11), pp. 2881–2912.

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

BUONOMANO-D. (2000). Decoding temporal information: A model based on short-term synaptic plasticity, *The Journal of Neuroscience*, **20**, pp. 1129–1141.

CAMERON-K., BOONSOBHAK-V., MURRAY-A., AND RENSHAW-D. (2005). Spike timing dependent plasticity (STDP) can ameliorate process variations in neuromorphic VLSI, *IEEE Transactions on Neural Networks*, **16**(6), pp. 1626–1637.

CARLSON-K. D., DUTT-N., NAGESWARAN-J. M., AND KRICHMAR-J. L. (2013). Design space exploration and parameter tuning for neuromorphic applications, *Proceedings of the Ninth IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis*, CODES+ISSS '13, IEEE Press, Piscataway, NJ, USA, pp. 20:1–20:2.

CARLSON-K., NAGESWARAN-J., DUTT-N., AND KRICHMAR-J. (2014). An efficient automated parameter tuning framework for spiking neural networks, *Frontiers in Neuroscience*.

CASSIDY-A., ANDREOU-A. G., AND GEORGIOU-J. (2011). A combinational digital logic approach to STDP, *The 2011 International Symposium on Circuits and Systems*, pp. 673–676.

CASSIDY-A., DENHAM-S., KANOLD-P., AND ANDREOU-A. (2007). FPGA based silicon spiking neural array, *IEEE Biomedical Circuits and Systems Conference*, pp. 75–78.

CHICCA-E., BADONI-D., DANTE-V., D'ANDREAGIOVANNI-M., SALINA-G., CAROTA-L., FUSI-S., AND DEL GIUDICE-P. (2003). A VLSI recurrent network of integrate-and-fire neurons connected by plastic synapses with long-term memory, *IEEE Transactions on Neural Networks*, **14**(5), pp. 1297–1307.

CHICCA-E., WHATLEY-A., LICHTSTEINER-P., DANTE-V., DELBRÜCK-T., DEL GIUDICE-P., DOUGLAS-R., AND INDIVERI-G. (2007). A multi-chip pulse-based neuromorphic infrastructure and its application to a model of orientation selectivity, *IEEE Transactions on Circuits and Systems I*, **5**(54), pp. 981–993.

CHOI-T. Y. W., SHI-B. E., AND BOAHEN-K. A. (2004). An on-off orientation selective address event representation image transceiver chip, *IEEE Transactions on Circuits and Systems I: Regular Papers*, **51**(2), pp. 342–353.

CHOUDHARY-S., SLOAN-S., FOK-S., NECKAR-A., TRAUTMANN-E., GAO-P., STEWART-T., ELIASMITH-C., AND BOAHEN-K. (2012). Silicon neurons that compute, *in* A. Villa., W. Duch., P. Érdi., F. Masulli., and G. Palm. (eds.), *Artificial Neural Networks and Machine Learning – ICANN 2012*, Vol. 7552 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 121–128.

CHUA-L. (2011). Resistance switching memories are memristors, *Applied Physics A*, **102**(4), pp. 765–783.

CLOPATH-C., AND GERSTNER-W. (2010). Voltage and spike timing interact in STDP—a unified model, *Frontiers in Synaptic Neuroscience, DOI: 10.3389/fnsyn.2010.00025*.

CLOPATH-C., BÜSING-L., VASILAKI-E., AND GERSTNER-W. (2010). Connectivity reflects coding: a model of voltage-based STDP with homeostasis, *Nature Neuroscience*, **13**(3), pp. 344–352.

COOPER-L., INTRATOR-N., BLAIS-B., AND SHOUVAL-H. (2004). *Theory of Cortical Plasticity*, World Scientific Pub Co Inc.

CRUZ-ALBRECHT-J. M., YUNG-M. W., AND SRINIVASA-N. (2012). Energy-efficient neuron, synapse and STDP integrated circuits, *IEEE Transactions on Biomedical Circuits and Systems*, **6**(3), pp. 246–256.

DASGUPTA-B., AND SCHNITGER-G. (1994). The power of approximating: a comparison of activation functions, *Mathematical Research*, **79**, pp. 641–641.

DAVISON-A. P., AND FRÉGNAC-Y. (2006). Learning cross-modal spatial transformations through spike timing-dependent plasticity, *The Journal of Neuroscience*, **26**(21), pp. 5604–5615.

DAYAN-P., AND ABBOTT-L. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, Taylor & Francis.

DEISS-S., DELBRÜCK-T., DOUGLAS-R., FISCHER-M., MAHOWALD-M., MATTHEWS-T., AND WHATLEY-A. (1994). Address-event asynchronous local broadcast protocol, World Wide Web page. http://www.ini.uzh.ch/~amw/scx/aeprotocol.html.

DELBRÜCK-T., AND VAN SCHAIK-A. (2005). Bias current generators with wide dynamic range, *Analog Integrated Circuits and Signal Processing*, **43**(3), pp. 247–268.

DELBRÜCK-T., BERNER-R., LICHTSTEINER-P., AND DUALIBE-C. (2010). 32-bit configurable bias current generator with sub-off-current capability, *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1647–1650.

DEMMING-A., GIMZEWSKI-J. K., AND VUILLAUME-D. (2013). Synaptic electronics, *Nanotechnology, doi:10.1088/0957-4484/24/38/380201*.

DOUGLAS-R., MAHOWALD-M., AND MEAD-C. (1995). Neuromorphic analogue VLSI, *Annual Review of Neuroscience*, **18**, pp. 255–281.

D'SOUZA-P., LIU-S.-C., AND HAHNLOSER-R. H. (2010). Perceptron learning rule derived from spike-frequency adaptation and spike-time-dependent plasticity, *Proceedings of the National Academy of Sciences*, **107**(10), pp. 4722–4727.

DUDEK-S. M., AND BEAR-M. F. (1992). Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade, *Proceedings of the National Academy of Sciences*, **89**(10), pp. 4363–4367.

EBERHARDT-S., DUONG-T., AND THAKOOR-A. (1989). Design of parallel hardware neural network systems from custom analog VLSI 'building block' chips, *International Joint Conference on Neural Networks*, pp. 183–190.

EBONG-I., AND MAZUMDER-P. (2012). CMOS and memristor-based neural network design for position detection, *Proceedings of the IEEE*, **100**(6), pp. 2050–2060.

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

ESHRAGHIAN-K. (2010). Evolution of nonvolatile resistive switching memory technologies: The related influence on hetrogeneous nanoarchitectures, *Transactions on Electrical and Electronic Materials (TEEM)*, **11**(6), pp. 243–248.

ESHRAGHIAN-K., CHO-K.-R., KAVEHEI-O., KANG-S.-K., ABBOTT-D., AND KANG-S.-M. S. (2011). Memristor MOS content addressable memory (MCAM): Hybrid architecture for future high performance search engines, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, **19**(8), pp. 1407–1417.

ESHRAGHIAN-K., KAVEHEI-O., CHAPPELL-J., IQBAL-A., AL-SARAWI-S., AND ABBOTT-D. (2012). Memristive device fundamentals and modeling: Applications to circuits and systems simulation, *Proceedings of the IEEE*, **100**(6), pp. 1991–2007.

FASNACHT-D., WHATLEY-A., AND INDIVERI-G. (2008). A serial communication infrastructure for multi-chip address event system, *International Symposium on Circuits and Systems, (ISCAS), 2008*, IEEE, pp. 648–651.

FITZHUGH-R. (1961). Impulses and physiological states in theoretical models of nerve membrane, *Biophysical Journal*, **1**(6), pp. 445–466.

FORTUNA-L., FRASCA-M., AND XIBILIA-M. G. (2009). *Chua's Circuit Implementations, Yesterday, Today and Tomorrow*, World Scientific.

FROEMKE-R., AND DAN-Y. (2002). Spike-timing-dependent synaptic modification induced by natural spike trains, *Nature*, **416**(6879), pp. 433–438.

FROEMKE-R., TSAY-I., RAAD-M., LONG-J., AND DAN-Y. (2006). Contribution of individual spikes in burst-induced long-term synaptic modification, *Journal of Neurophysiology*, **95**(3), pp. 1620–1629.

FRYE-J., ANANTHANARAYANAN-R., AND MODHA-D. S. (2007). Towards real-time, mouse-scale cortical simulations, *CoSyNe: Computational and Systems Neuroscience*, p. 106.

FURBER-S. B., LESTER-D. R., PLANA-L. A., GARSIDE-J. D., PAINKRAS-E., TEMPLE-S., AND BROWN-A. D. (2013). Overview of the spinnaker system architecture, *IEEE Transactions on Computers*, **62**(12), pp. 2454–2467.

FUSI-S., ANNUNZIATO-M., BADONI-D., SALAMON-A., AND AMIT-D. J. (2000). Spike-driven synaptic plasticity: theory, simulation, VLSI implementation, *Neural Computation*, **12**(10), pp. 2227–2258.

GAO-P., BENJAMIN-B., AND BOAHEN-K. (2012). Dynamical system guided mapping of quantitative neuronal models onto neuromorphic hardware, *Journal of Neurophysiology*, **59**(10), pp. 2383–2394.

GERSTNER-W., AND KISTLER-W. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*, Cambridge University Press.

GERSTNER-W., KEMPTER-R., VAN HEMMEN-J., AND WAGNER-H. (1996). A neuronal learning rule for sub-millisecond temporal coding, *Nature*, **383**(6595), pp. 76–78.

GIULIONI-M., PANNUNZI-M., BADONI-D., DANTE-V., AND DEL GIUDICE-P. (2009). Classification of correlated patterns with a configurable analog VLSI neural network of spiking neurons and self-regulating plastic synapses, *Neural Computation*, **21**(11), pp. 3106–3129.

GJORGJIEVA-J., CLOPATH-C., AUDET-J., AND PFISTER-J. (2011). A triplet spike-timing–dependent plasticity model generalizes the Bienenstock–Cooper–Munro rule to higher-order spatiotemporal correlations, *Proceedings of the National Academy of Sciences*, **108**(48), pp. 19383–19388.

GOLDBERG-D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Vol. 412, Addison-Wesley Reading Menlo Park.

GORDON-C., AND HASLER-P. (2002). Biological learning modeled in an adaptive floating-gate system, *IEEE International Symposium on Circuits and Systems*, Vol. 5, pp. 609–612.

GRASSIA-F., BUHRY-L., LÉVI-T., TOMAS-J., DESTEXHE-A., AND SAÏGHI-S. (2011). Tunable neuromimetic integrated system for emulating cortical neuron models, *Frontiers in Neuroscience, DOI: 10.3389/fnins.2011.00134*.

GRAUPNER-M., AND BRUNEL-N. (2012). Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location, *Proceedings of the National Academy of Sciences*, **109**(10), pp. 3991–3996.

GUETIG-R., AHARONOV-R., ROTTER-S., AND SOMPOLINSKY-H. (2003). Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity, *The Journal of Neuroscience*, **23**(9), pp. 3697–3714.

HAFLIGER-P. (2007). Adaptive WTA with an analog VLSI neuromorphic learning chip, *IEEE Transactions on Neural Networks*, **18**(2), pp. 551–572.

HAFLIGER-P., AND KOLLE RIIS-H. (2003). A multi-level static memory cell, *Proceedings of the 2003 International Symposium on Circuits and Systems*, Vol. 1, pp. I–25–I–28.

HAFLIGER-P., MAHOWALD-M., AND WATTS-L. (1997). A spike based learning neuron in analog VLSI, *Advances in Neural Information Processing Systems*, **9**, pp. 692–698.

HAMILTON-T. J., AND TAPSON-J. (2011). A neuromorphic cross-correlation chip, *2011 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 865–868.

HAMILTON-T. J., AND VAN SCHAIK-A. (2011). Silicon implementation of the generalized integrate-and-fire neuron model, *2011 IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pp. 108–112.

HAMILTON-T. J., JIN-C. T., VAN SCHAIK-A., AND TAPSON-J. (2008). An active 2-D silicon cochlea, *IEEE Transactions on Biomedical Circuits and Systems*, **2**(1), pp. 30–43.

HAMILTON-T. J., NELSON-N. M., SANDER-D., AND ABSHIRE-P. (2009). A cell impedance sensor based on a silicon cochlea, *IEEE Biomedical Circuits and Systems Conference*, pp. 117–120.

HASLER-J., AND MARR-H. B. (2013). Finding a roadmap to achieve large neuromorphic hardware systems, *Frontiers in Neuroscience, DOI: 10.3389/fnins.2013.00118*.

HAYKIN-S. (1994). *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR.

HEBB-D. (2002). *The Organization of Behavior: A Neuropsychological Theory*, Lawrence Erlbaum.

HODGKIN-A. L., AND HUXLEY-A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve, *The Journal of Physiology*, **117**(4), p. 500.

HOLLER-M., TAM-S., CASTRO-H., AND BENSON-R. (1989). An electrically trainable artificial neural network (ETANN) with 10240 'floating gate' synapses, *International Joint Conference on Neural Networks*, pp. 191–196.

HSIEH-H.-Y., AND TANG-K.-T. (2012). A spiking neural network chip for odor data classification, *2012 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, pp. 88–91.

HYNNA-K., AND BOAHEN-K. (2001). Space-rate coding in an adaptive silicon neuron, *Neural Networks*, **14**(67), pp. 645–656.

IANNELLA-N., AND TANAKA-S. (2006). Synaptic efficacy cluster formation across the dendrite via STDP, *Neuroscience Letters*, **403**(1-2), pp. 24–29.

IANNELLA-N., LAUNEY-T., AND TANAKA-S. (2010). Spike timing-dependent plasticity as the origin of the formation of clustered synaptic efficacy engrams, *Frontiers in Computational Neuroscience, DOI: 10.3389/fncom.2010.00021*.

IMAM-N., AKOPYAN-F., ARTHUR-J., MEROLLA-P., MANOHAR-R., AND MODHA-D. (2012). A digital neurosynaptic core using event-driven QDI circuits, *2012 18th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*, pp. 25–32.

INDIVERI-G. (2003). A low-power adaptive integrate-and-fire neuron circuit, *Proceedings of the 2003 International Symposium on Circuits and Systems*, Vol. 4, pp. 820–823.

INDIVERI-G., AND HORIUCHI-T. (2011). Frontiers in neuromorphic engineering, *Frontiers in Neuroscience, DOI: 10.3389/fnins.2011.00118*.

INDIVERI-G., CHICCA-E., AND DOUGLAS-R. (2006). A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity, *IEEE Transactions on Neural Networks*, **17**(1), pp. 211–221.

INDIVERI-G., LINARES-BARRANCO-B., HAMILTON-T., VAN SCHAIK-A., ETIENNE-CUMMINGS-R., DELBRÜCK-T., LIU-S., DUDEK-P., HÄFLIGER-P., RENAUD-S., SCHEMMEL-J., CAUWENBERGHS-G., ARTHUR-J., HYNNA-K., FOLOWOSELE-F., SAIGHI-S., SERRANO-GOTARREDONA-T., WIJEKOON-J., WANG-Y., AND BOAHEN-K. (2011). Neuromorphic silicon neuron circuits, *Frontiers in Neuroscience, DOI: 10.3389/fnins.2011.00073*.

INDIVERI-G., LINARES-BARRANCO-B., LEGENSTEIN-R., DELIGEORGIS-G., AND PRODROMAKIS-T. (2013). Integration of nanoscale memristor synapses in neuromorphic computing architectures, *Nanotechnology*, **24**(38), p. art. no. 384010.

INDIVERI-G., LIU-S., DELBRÜCK-T., AND DOUGLAS-R. (2009). Neuromorphic systems, *Encyclopedia of Neuroscience*, pp. 521–528.

INDIVERI-G., STEFANINI-F., AND CHICCA-E. (2010). Spike-based learning with a generalized integrate and fire silicon neuron, *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1951–1954.

IZHIKEVICH-E., AND DESAI-N. (2003). Relating STDP to BCM, *Neural Computation*, **15**(7), pp. 1511–1523.

IZHIKEVICH-E. M. (2001). Resonate-and-fire neurons, *Neural Networks*, **14**(6), pp. 883–894.

IZHIKEVICH-E. M. (2003). Simple model of spiking neurons, *IEEE Transactions on Neural Networks*, **14**(6), pp. 1569–1572.

IZHIKEVICH-E. M. (2004). Which model to use for cortical spiking neurons?, *IEEE Transactions on Neural Networks*, **15**(5), pp. 1063–1070.

JO-S. H., CHANG-T., EBONG-I., BHADVIYA-B. B., MAZUMDER-P., AND LU-W. (2010). Nanoscale memristor device as synapse in neuromorphic systems, *Nano letters*, **10**(4), pp. 1297–1301.

KASABOV-N., FEIGIN-V., HOU-Z.-G., CHEN-Y., LIANG-L., KRISHNAMURTHI-R., OTHMAN-M., AND PARMAR-P. (2014). Evolving spiking neural networks for personalised modelling, classification and prediction of spatio-temporal patterns with a case study on stroke, *Neurocomputing*, **134**, pp. 269–279.

KASABOV-N. K. (2014). Neucube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data, *Neural Networks*, **52**, pp. 62–76.

KEMPTER-R., GERSTNER-W., AND VAN HEMMEN-J. (1999). Hebbian learning and spiking neurons, *Physical Review E*, **59**(4), pp. 4498–4514.

KHOSLA-D., HUBER-D. J., AND KANAN-C. (2014). A neuromorphic system for visual object recognition, *Biologically Inspired Cognitive Architectures, DOI: 10.1016/j.bica.2014.02.001*.

KIRKWOOD-A., RIOULT-M., AND BEAR-M. (1996). Experience-dependent modification of synaptic plasticity in visual cortex, *Nature*, **381**(6582), pp. 526–528.

KISTLER-W. M., AND HEMMEN-J. L. (2000). Modeling synaptic plasticity in conjunction with the timing of pre-and postsynaptic action potentials, *Neural Computation*, **12**(2), pp. 385–405.

KOICKAL-T., HAMILTON-A., TAN-S., COVINGTON-J., GARDNER-J., AND PEARCE-T. (2007). Analog VLSI circuit implementation of an adaptive neuromorphic olfaction chip, *IEEE Transactions on Circuits and Systems I*, **54**(1), pp. 60–73.

KUDITHIPUDI-D., MERKEL-C., SOLTIZ-M., ROSE-G., AND PINO-R. (2014). Design of neuromorphic architectures with memristors, *in* R. E. Pino. (ed.), *Network Science and Cybersecurity*, Vol. 55 of *Advances in Information Security*, Springer New York, pp. 93–103.

LAUGHLIN-S., AND SEJNOWSKI-T. (2003). Communication in neuronal networks, *Science*, **301**(5641), pp. 1870–1874.

LEE-S.-J., KAVEHEI-O., HONG-Y.-K., CHO-T.-W., YOU-Y., CHO-K., AND ESHRAGHIAN-K. (2010). 3-D system-on-system (SoS) biomedical-imaging architecture for health-care applications, *IEEE Transactions on Biomedical Circuits and Systems*, **4**(6), pp. 426–436.

LICHTSTEINER-P., POSCH-C., AND DELBRÜCK-T. (2008). A 128 ×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor, *IEEE Journal of Solid-State Circuits*, **43**(2), pp. 566–576.

LIKHAREV-K., AND STRUKOV-D. (2005). CMOL: Devices, circuits, and architectures, *Introducing Molecular Electronics*, pp. 447–477.

LINARES-BARRANCO-B., AND SERRANO-GOTARREDONA-T. (2003). On the design and characterization of femtoampere current-mode circuits, *IEEE Journal of Solid-State Circuits*, **38**(8), pp. 1353–1363.

LISMAN-J., AND SPRUSTON-N. (2010). Questions about STDP as a general model of synaptic plasticity, *Frontiers in Synaptic Neuroscience, DOI: 10.3389/fnsyn.2010.00140*.

LISMAN-J. E. (1985). A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase, *Proceedings of the National Academy of Sciences*, **82**(9), pp. 3055–3057.

LIU-S.-C., AND DELBRÜCK-T. (2010). Neuromorphic sensory systems, *Current Opinion in Neurobiology*, **20**(3), pp. 288–295.

LIU-S.-C., DELBRÜCK-T., KRAMER-J., INDIVERI-G., AND DOUGLAS-R. (2002). *Analog VLSI: Circuits and Principles*, MIT Press, Cambridge, MA.

MARKRAM-H., LÜBKE-J., FROTSCHER-M., AND SAKMANN-B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs, *Science*, **275**(5297), pp. 213–215.

MARTIN-S., GRIMWOOD-P., AND MORRIS-R. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis, *Annual Review of Neuroscience*, **23**(1), pp. 649–711.

MASQUELIER-T., AND THORPE-S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity, *PLoS Computational Biology*, **3**(2), p. art. no. e31.

MASQUELIER-T., AND THORPE-S. J. (2010). Learning to recognize objects using waves of spikes and spike timing-dependent plasticity, *International Joint Conference on Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2010.5596934*.

MASQUELIER-T., GUYONNEAU-R., AND THORPE-S. J. (2008). Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains, *PLoS ONE*, **3**(1), p. art. no. e1377.

MAYR-C., AND PARTZSCH-J. (2010). Rate and pulse based plasticity governed by local synaptic state variables, *Frontiers in Synaptic Neuroscience, DOI: 10.3389/fnsyn.2010.00033*.

MAYR-C., NOACK-M., PARTZSCH-J., AND SCHUFFNY-R. (2010). Replicating experimental spike and rate based neural learning in CMOS, *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 105–108.

MAYR-C., PARTZSCH-J., NOACK-M., AND SCHUFFNY-R. (2013). Live demonstration: Multiple-timescale plasticity in a neuromorphic system, *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 666–670.

MCCULLOCH-W. S., AND PITTS-W. (1943). A logical calculus of the ideas immanent in nervous activity, *The Bulletin of Mathematical Biophysics*, **5**(4), pp. 115–133.

MEAD-C. (1989). *Analog VLSI and Neural Systems*, Addison-Wesley.

MEAD-C. (1990). Neuromorphic electronic systems, *Proceedings of the IEEE*, **78**(10), pp. 1629–1636.

MENG-Y., ZHOU-K., MONZON-J., AND POON-C. (2011). Iono-neuromorphic implementation of spike-timing-dependent synaptic plasticity, *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, pp. 7274–7277.

MEROLLA-P., AND BOAHEN-K. (2004). A recurrent model of orientation maps with simple and complex cells, *Advances in Neural Information Processing Systems 17 (NIPS 2004)*, pp. 1995–2002.

MEROLLA-P., ARTHUR-J., AKOPYAN-F., IMAM-N., MANOHAR-R., AND MODHA-D. S. (2011). A digital neurosynaptic core using embedded crossbar memory with 45 pJ per spike in 45 nm, *2011 IEEE Custom Integrated Circuits Conference (CICC), DOI: 10.1109/CICC.2011.6055294*.

MEROLLA-P., ARTHUR-J., ALVAREZ-R., BUSSAT-J.-M., AND BOAHEN-K. (2013). A multicast tree router for multichip neuromorphic systems, *IEEE Transactions on Circuits and Systems I: Regular Papers*, **61**(3), pp. 820–833.

MITRA-S., FUSI-S., AND INDIVERI-G. (2009). Real-time classification of complex patterns using spike-based learning in neuromorphic VLSI, *IEEE Transactions on Biomedical Circuits and Systems*, **3**(1), pp. 32–42.

MITRA-S., INDIVERI-G., AND ETIENNE-CUMMINGS-R. (2010). Synthesis of log-domain integrators for silicon synapses with global parametric control, *International Symposium on Circuits and Systems (ISCAS)*, pp. 97–100.

MORADI-S., AND INDIVERI-G. (2011). A VLSI network of spiking neurons with an asynchronous static random access memory, *2011 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 277–280.

MORADI-S., AND INDIVERI-G. (2014). An event-based neural network architecture with an asynchronous programmable synaptic memory, *IEEE Transactions on Biomedical Circuits and Systems*, **8**(1), pp. 98–107.

MORADI-S., IMAM-N., MANOHAR-R., AND INDIVERI-G. (2013). A memory-efficient routing method for large-scale spiking neural networks, *2013 European Conference on Circuit Theory and Design (ECCTD), DOI: 10.1109/ECCTD.2013.6662203*.

MORRIS-C., AND LECAR-H. (1981). Voltage oscillations in the barnacle giant muscle fiber, *Biophysical Journal*, **35**(1), pp. 193–213.

MORRISON-A., DIESMANN-M., AND GERSTNER-W. (2008). Phenomenological models of synaptic plasticity based on spike timing, *Biological Cybernetics*, **98**(6), pp. 459–478.

MOSTAFA-H., CORRADI-F., OSSWALD-M., AND INDIVERI-G. (2013). Automated synthesis of asynchronous event-based interfaces for neuromorphic systems, *2013 European Conference on Circuit Theory and Design (ECCTD), DOI: 10.1109/ECCTD.2013.6662213*.

NEFTCI-E., AND INDIVERI-G. (2010). A device mismatch compensation method for VLSI neural networks, *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 262–265.

NEFTCI-E., BINAS-J., RUTISHAUSER-U., CHICCA-E., INDIVERI-G., AND DOUGLAS-R. J. (2013). Synthesizing cognition in neuromorphic electronic systems, *Proceedings of the National Academy of Sciences*, **110**(37), pp. E3468–E3476.

NERE-A., OLCESE-U., BALDUZZI-D., AND TONONI-G. (2012). A neuromorphic architecture for object recognition and motion anticipation using burst-STDP, *PloS ONE*, **7**(5), p. art. no. e36958.

O'CONNOR-D. H., WITTENBERG-G. M., AND WANG-S. S.-H. (2005). Graded bidirectional synaptic plasticity is composed of switch-like unitary events, *Proceedings of the National Academy of Sciences of the United States of America*, **102**(27), pp. 9679–9684.

OLIVERI-A., RIZZO-R., AND CHELLA-A. (2007). An application of spike-timing-dependent plasticity to readout circuit for liquid state machine, *International Joint Conference on Neural Networks*, pp. 1441–1445.

PAINKRAS-E., PLANA-L., GARSIDE-J., TEMPLE-S., GALLUPPI-F., PATTERSON-C., LESTER-D., BROWN-A., AND FURBER-S. (2013). Spinnaker: A 1-W 18-core system-on-chip for massively-parallel neural network simulation, *IEEE Journal of Solid-State Circuits*, **48**(8), pp. 1–11.

PELGROM-M., DUINMAIJER-A., AND WELBERS-A. (1989). Matching properties of MOS transistors, *IEEE Journal of Solid-State Circuits*, **24**(5), pp. 1433–1439.

PERSHIN-Y. V., AND DI VENTRA-M. (2012). Neuromorphic, digital, and quantum computation with memory circuit elements, *Proceedings of the IEEE*, **100**(6), pp. 2071–2080.

PETERSEN-C. C., MALENKA-R. C., NICOLL-R. A., AND HOPFIELD-J. J. (1998). All-or-none potentiation at CA3-CA1 synapses, *Proceedings of the National Academy of Sciences*, **95**(8), pp. 4732–4737.

PFEIL-T., POTJANS-T. C., SCHRADER-S., POTJANS-W., SCHEMMEL-J., DIESMANN-M., AND MEIER-K. (2012). Is a 4-bit synaptic weight resolution enough?–constraints on enabling spike-timing dependent plasticity in neuromorphic hardware, *Frontiers in Neuroscience, DOI: 10.3389/fnins.2012.00090*.

PFISTER-J., AND GERSTNER-W. (2006). Triplets of spikes in a model of spike timing-dependent plasticity, *The Journal of Neuroscience*, **26**(38), pp. 9673–9682.

PILLOW-J. W., SHLENS-J., PANINSKI-L., SHER-A., LITKE-A. M., CHICHILNISKY-E., AND SIMONCELLI-E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population, *Nature*, **454**(7207), pp. 995–999.

POON-C., AND ZHOU-K. (2011). Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities, *Frontiers in Neuroscience, DOI: 10.3389/fnins.2011.00108*.

RACHMUTH-G., AND POON-C.-S. (2008). Transistor analogs of emergent iono-neuronal dynamics, *HFSP Journal*, **2**(3), pp. 156–166.

RACHMUTH-G., SHOUVAL-H., BEAR-M., AND POON-C. (2011). A biophysically-based neuromorphic model of spike rate- and timing-dependent plasticity, *Proceedings of the National Academy of Sciences*, **108**(49), pp. E1266–E1274.

RAMAKRISHNAN-S., HASLER-P., AND GORDON-C. (2011). Floating gate synapses with spike-time-dependent plasticity, *IEEE Transactions on Biomedical Circuits and Systems*, **5**(3), pp. 244–252.

RAZAVI-B. (2002). *Design of Analog CMOS Integrated Circuits*, Tata McGraw-Hill Education.

ROSE-R., AND HINDMARSH-J. (1989). The assembly of ionic currents in a thalamic neuron I. The three-dimensional model, *Proceedings of the Royal Society of London. B. Biological Sciences*, **237**(1288), pp. 267–288.

ROWCLIFFE-P., AND FENG-J. (2008). Training spiking neuronal networks with applications in engineering tasks, *IEEE Transactions on Neural Networks*, **19**(9), pp. 1626–1640.

SATYANARAYANA-S., TSIVIDIS-Y., AND GRAF-H.-P. (1992). A reconfigurable vlsi neural network, *IEEE Journal of Solid-State Circuits*, **27**(1), pp. 67–81.

SCHEMMEL-J., BRUDERLE-D., GRUBL-A., HOCK-M., MEIER-K., AND MILLNER-S. (2010). A wafer-scale neuromorphic hardware system for large-scale neural modeling, *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1947–1950.

SCHEMMEL-J., GRUBL-A., MEIER-K., AND MUELLER-E. (2006). Implementing synaptic plasticity in a VLSI spiking neural network model, *The 2006 International Joint Conference on Neural Networks, DOI: 10.1109/IJCNN.2006.246651*.

SCHMUKER-M., PFEIL-T., AND NAWROT-M. P. (2014). A neuromorphic network for generic multivariate data classification, *Proceedings of the National Academy of Sciences, DOI: 10.1073/pnas.1303053111*.

SCHOLZE-S., SCHIEFER-S., PARTZSCH-J., HARTMANN-S., MAYR-C. G., HÖPPNER-S., EISENREICH-H., HENKER-S., VOGGINGER-B., AND SCHFFNY-R. (2011). VLSI implementation of a 2.8 Gevent/s packet based AER interface with routing and event sorting functionality, *Frontiers in Neuroscience, DOI: 10.3389/fnins.2011.00117*.

SEEBACHER-E. (2005). CMOS technology characterisation for analog/RF application, *11th Workshop on Electronics for LHC and Future Experiments*, CERN, pp. 33 – 41.

SEO-J.-S., BREZZO-B., LIU-Y., PARKER-B. D., ESSER-S. K., MONTOYE-R. K., RAJENDRAN-B., TIERNO-J. A., CHANG-L., AND MODHA-D. S. (2011). A 45 nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons, *The 2011 IEEE Custom Integrated Circuits Conference (CICC), DOI: 10.1109/CICC.2011.6055293*.

SERRANO-GOTARREDONA-T., MASQUELIER-T., PRODROMAKIS-T., INDIVERI-G., AND LINARES-BARRANCO-B. (2013). STDP and STDP variations with memristors for spiking neuromorphic learning systems, *Frontiers in Neuroscience, DOI: 10.3389/fnins.2013.00002*.

SHAH-N. T., YEUNG-L. C., COOPER-L. N., CAI-Y., AND SHOUVAL-H. Z. (2006). A biophysical basis for the inter-spike interaction of spike-timing-dependent plasticity, *Biological Cybernetics*, **95**(2), pp. 113–121.

SHEIK-S., CHICCA-E., AND INDIVERI-G. (2012a). Exploiting device mismatch in neuromorphic VLSI systems to implement axonal delays, *The 2012 International Joint Conference on Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2012.6252636*.

SHEIK-S., COATH-M., INDIVERI-G., DENHAM-S. L., WENNEKERS-T., AND CHICCA-E. (2012b). Emergent auditory feature tuning in a real-time neuromorphic VLSI system, *Frontiers in Neuroscience, DOI: 10.3389/fnins.2012.00017*.

SHERI-A., HWANG-H., JEON-M., AND GEUN LEE-B. (2014). Neuromorphic character recognition system with two PCMO memristors as a synapse, *IEEE Transactions on Industrial Electronics*, **61**(6), pp. 2933–2941.

SHERIDAN-P., AND LU-W. (2014). Memristors and memristive devices for neuromorphic computing, *Memristor Networks*, Springer, pp. 129–149.

SHOUVAL-H. Z. (2011). What is the appropriate description level for synaptic plasticity?, *Proceedings of the National Academy of Sciences*, **108**(48), pp. 19103–19104.

SHOUVAL-H. Z., BEAR-M. F., AND COOPER-L. N. (2002). A unified model of NMDA receptor-dependent bidirectional synaptic plasticity, *Proceedings of the National Academy of Sciences*, **99**(16), pp. 10831–10836.

SHOUVAL-H. Z., WANG-S. S.-H., AND WITTENBERG-G. M. (2010). Spike timing dependent plasticity: a consequence of more fundamental learning rules, *Frontiers in computational neuroscience, DOI: 10.3389/fncom.2010.00019*.

SJÖSTRÖM-P. J., RANCZ-E. A., ROTH-A., AND HÄUSSER-M. (2008). Dendritic excitability and synaptic plasticity, *Physiological Reviews*, **88**(2), pp. 769–840.

SJÖSTRÖM-P., TURRIGIANO-G., AND NELSON-S. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity, *Neuron*, **32**(6), pp. 1149–1164.

SMITH-G. D., COX-C. L., SHERMAN-S. M., AND RINZEL-J. (2000). Fourier analysis of sinusoidally driven thalamocortical relay neurons and a minimal integrate-and-fire-or-burst model, *Journal of Neurophysiology*, **83**(1), pp. 588–610.

SMITH-L. S. (2006). Implementing neural models in silicon, *Handbook of Nature-Inspired and Innovative Computing*, Springer, pp. 433–475.

SOLEIMANI-H., AHMADI-A., AND BAVANDPOUR-M. (2012). Biologically inspired spiking neurons: Piecewise linear models and digital implementation, *IEEE Transactions on Circuits and Systems I: Regular Papers*, **59**(12), pp. 2991–3004.

SONG-S., MILLER-K., AND ABBOTT-L. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity, *Nature Neuroscience*, **3**, pp. 919–926.

STRUKOV-D., SNIDER-G., STEWART-D., AND WILLIAMS-R. (2008). The missing memristor found, *Nature*, **453**(7191), pp. 80–83.

TANAKA-H., MORIE-T., AND AIHARA-K. (2009). A CMOS spiking neural network circuit with symmetric/asymmetric STDP function, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **E92-A**(7), pp. 1690–1698.

URAMOTO-T., AND TORIKAI-H. (2013). A calcium-based simple model of multiple spike interactions in spike-timing-dependent plasticity, *Neural Computation*, **25**(7), pp. 1853–1869.

VITTOZ-E. (1985). The design of high-performance analog circuits on digital CMOS chips, *IEEE Journal of Solid-State Circuits*, **20**(3), pp. 657–665.

VOGELSTEIN-R. J., MALLIK-U., CULURCIELLO-E., CAUWENBERGHS-G., AND ETIENNE-CUMMINGS-R. (2007a). A multichip neuromorphic system for spike-based visual information processing, *Neural Computation*, **19**(9), pp. 2281–2300.

VOGELSTEIN-R. J., MALLIK-U., VOGELSTEIN-J. T., AND CAUWENBERGHS-G. (2007b). Dynamically reconfigurable silicon array of spiking neurons with conductance-based synapses, *IEEE Transactions on Neural Networks*, **18**(1), pp. 253–265.

VREEKEN-J. (2003). Spiking neural networks, an introduction, *Technical Report UU-CS*, (2003-008), pp. 1–5.

WANG-H., AND WAGNER-J. J. (1999). Priming-induced shift in synaptic plasticity in the rat hippocampus, *Journal of Neurophysiology*, **82**(4), pp. 2024–2028.

WANG-H., GERKIN-R., NAUEN-D., AND BI-G. (2005). Coactivation and timing-dependent integration of synaptic potentiation and depression, *Nature Neuroscience*, **8**(2), pp. 187–193.

WANG-R. M., HAMILTON-T. J., TAPSON-J. C., AND VAN SCHAIK-A. (2014a). A mixed-signal implementation of a polychronous spiking neural network with delay adaptation, *Frontiers in neuroscience, DOI: 10.3389/fnins.2014.00051*.

WANG-Z., ZHAO-W., KANG-W., ZHANG-Y., KLEIN-J.-O., RAVELOSONA-D., AND CHAPPERT-C. (2014b). Compact modelling of ferroelectric tunnel memristor and its use for neuromorphic simulation, *Applied Physics Letters*, **104**(5), pp. 053505:1–053505:5.

WESTE-N., AND ESHRAGHIAN-K. (1994). *Principles of CMOS VLSI design: a systems perspective*, 2nd edn, Addison-Wesley.

WESTE-N., AND HARRIS-D. (2005). *CMOS VLSI Design: A Circuits and Systems Perspective*, Addison-Wesley.

WIJEKOON-J. H., AND DUDEK-P. (2012). VLSI circuits implementing computational models of neocortical circuits, *Journal of Neuroscience Methods*, **210**(1), pp. 93–109.

WILSON-H. R. (1999). Simplified dynamics of human and mammalian neocortical neurons, *Journal of Theoretical Biology*, **200**(4), pp. 375–388.

WILSON-H. R., AND COWAN-J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons, *Biophysical Journal*, **12**(1), pp. 1–24.

YAO-E., HUSSAIN-S., BASU-A., AND HUANG-G.-B. (2013). Computation using mismatch: Neuromorphic extreme learning machines, *2013 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 294–297.

YOUNG-J. M., WALESZCZYK-W. J., WANG-C., CALFORD-M. B., DREHER-B., AND OBERMAYER-K. (2007). Cortical reorganization consistent with spike timing—but not correlation-dependent plasticity, *Nature Neuroscience*, **10**(7), pp. 887–895.

YU-S., GAO-B., FANG-Z., YU-H., KANG-J., AND WONG-H.-S. P. (2013). A low energy oxide-based electronic synaptic device for neuromorphic visual systems with tolerance to device variation, *Advanced Materials*, **25**(12), pp. 1774–1779.

ZAMARREÑO-RAMOS-C., CAMUÑAS-MESA-L. A., PÉREZ-CARRASCO-J. A., MASQUELIER-T., SERRANO-GOTARREDONA-T., AND LINARES-BARRANCO-B. (2011). On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex, *Frontiers in Neuroscience, DOI: 10.3389/fnins.2011.00026*.

ZHANG-L., LAI-Q., AND CHEN-Y. (2010). Configurable neural phase shifter with spike-timing-dependent plasticity, *IEEE Electron Device Letters*, **31**(7), pp. 716–718.

ZUCKER-R., AND REGEHR-W. (2002). Short-term synaptic plasticity, *Annual Review of Physiology*, **64**, pp. 355–405.

List of Acronyms

GA Genetic Algorithm

ANN Artificial Neural Network

VLSI Very Large Scale Integration

OTA Operational Transconductance Amplifier

HH Hodgkin-Huxley

IF Integrate and Fire

LIF Leaky Integrate and Fire

IFB Integrate-and-Fire-or-Burst

QIF Quadratic Integrate and Fire

MOSFET Metal Oxide Semiconductor Field Effect Transistor

SNN Spiking Neural Network

EPSC Excitatory Post-Synaptic Current

IPSC Inhibitory Post-Synaptic Current

LTP Long Term Potentiation

LTD Long Term Depression

BCM Bienenstock Cooper Munro

ISI Inter Spike Interval

STDP Spike Timing Dependent Synaptic Plasticity

TSTDP Triplet-based Spike Timing Dependent Plasticity

PSTDP Pair-based Spike Timing Dependent Plasticity

DPI Differential Pair Integrator

CMOS Complementary Metal Oxide Semiconductor

FPGA Field Programmable Gate Array

DAC Digital to Analog Converter

AER Address Event Representation

SRAM Static Random Access Memory

BMI Brain Machine Interface

Index

Biography

S. Mostafa Rahimi Azghadi was born in Mashhad, Iran, in 1982. In 2006, he graduated from Sadjad University of Mashhad, with a Bachelor's Degree in Computer Engineering (majoring in hardware) with the first class rank. He obtained his Master of Engineering in Computer Architecture with first class honours, from Shahid Beheshti University (SBU), Tehran, Iran, in 2009 under the supervision of Prof. Keivan Navi.

In 2010, he was awarded an Australian International Postgraduate Research Scholarship (IPRS) and the Adelaide University Scholarship (AUS) to pursue his PhD under Dr Said Al- Sarawi, Dr Nicolangelo Iannella, and Prof. Derek Abbott, in the School of Electrical & Electronic Engineering, The University of Adelaide. In 2012 and 2013, he was a visiting scholar in the Neuromorphic Cognitive System group, Institute of Neuroinformatics, University and Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, hosted by Prof. Giacomo Indiveri. During his candidature, Mr Rahimi received several International awards and scholarships including, Japanese Neural Network Society travel award (2011), Brain Corporation fellowship for Spiking Neural Networks (2012), IEEE Computational Intelligence Society travel award (2012), Research Abroad scholarship (2012), the D. R. Stranks fellowship (2012), the Doreen McCarthy Bursary (2013), the IEEE SA Section Student Travel Award (2013), and the Simon Rockliff Scholarship for outstanding postgraduate mentorship from DSTO (2013). Mr Rahimi has served as a reviewer for a number of recognised journals and conferences including IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS, PLOS ONE, and IEEE ISCAS AND BIOCAS. He is a member of Iranian National Elite Foundation and the Institution of Engineers Australia. He is also graduate student member of the IEEE, IEEE computational intelligence society, the Australian Society for Medical and Biological Engineering, and the Australian Associations of Computational Neuroscientists and Neuromorphic Engineers.

S. Mostafa Rahimi Azghadi
mostafa@eleceng.adelaide.edu.au

Scientifc Genealogy of

S Mostafa Rahimi Azghadi

Formalised supervisor relationship
Mentoring relationship
Nobel prize

"If I have seen further it is by standing on the shoulders of Giants."
Isaac Newton