Mixed Signal VLSI Circuit Implementation of the Cortical Microcircuit Models

A thesis submitted to the University of Manchester for the degree of

Doctor of Philosophy

in the Faculty of Engineering and Physical Sciences

2011

Jayawan Hasanka Bandara Wijekoon

School of Electrical and Electronic Engineering

CONTENT

Total words 46 628

Abstract

This thesis proposes a novel set of generic and compact biologically plausible VLSI (Very Large Scale Integration) neural circuits, suitable for implementing a parallel VLSI network that closely resembles the function of a small-scale neocortical network. The proposed circuits include a cortical neuron, two different long-term plastic synapses and four different short-term plastic synapses. These circuits operate in accelerated-time, where the time scale of neural responses is approximately three to four orders of magnitude faster than the biological-time scale of the neuronal activities, providing higher computational throughput in computing neural dynamics. Further, a novel biological-time cortical neuron circuit with similar dynamics as of the accelerated-time neuron is proposed to demonstrate the feasibility of migrating accelerated-time circuits into biological-time circuits.

The fabricated accelerated-time VLSI neuron circuit is capable of replicating distinct firing patterns such as regular spiking, fast spiking, chattering and intrinsic bursting, by tuning two external voltages. It reproduces biologically plausible action potentials. This neuron circuit is compact and enables implementation of many neurons in a single silicon chip. The circuit consumes extremely low energy per spike (8pJ). Incorporating this neuron circuit in a neural network facilitates diverse non-linear neuron responses, which is an important aspect in neural processing.

Two of the proposed long-term plastic synapse circuits include spike-time dependent plasticity (STDP) synapse, and dopamine modulated STDP synapse. The short-term plastic synapses include excitatory depressing, inhibitory facilitating, inhibitory depressing, and excitatory facilitating synapses. Many neural parameters of short- and long- term synapses can be modified independently using externally controlled tuning voltages to obtain distinct synaptic properties. Having diverse synaptic dynamics in a network facilitates richer network behaviours such as learning, memory, stability and dynamic gain control, inherent in a biological neural network.

To prove the concept in VLSI, different combinations of these accelerated-time neural circuits are fabricated in three integrated circuits (ICs) using a standard 0.35 µm CMOS technology. Using first two ICs, functions of cortical neuron and STDP synapses have been experimentally verified. The third IC, the Cortical Neural Layer (CNL) Chip is designed and fabricated to facilitate cortical network emulations. This IC implements neural circuits with a similar composition to the cortical layer of the neocortex. The CNL chip comprises 120 cortical neurons and 7 560 synapses. Many of these CNL chips can be combined together to form a six-layered VLSI neocortical network to validate the network dynamics and to perform neural processing of small-scale cortical networks.

The proposed neuromorphic systems can be used as a simulation acceleration platform to explore the processing principles of biological brains and also move towards realising low power, real-time intelligent computing devices and control systems.

Declaration

Except some materials of Section 4.2 in Chapter 4, no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgment

I would like to thank my supervisor Dr Piotr Dudek, in the University of Manchester, for providing stimulating suggestions, guidance and encouragement throughout this research work.

I express my gratitude to Dr Thomas Wennekers, and all my lab mates, including Dr Steve Carey, Mr. Kevin Brohan, Dr David Barr, Dr Christian Nilsen for their assistance in providing the friendly environment to work throughout my PhD research.

I would like to thank EPSRC (Engineering and Physical Sciences Research Council) for providing the studentship under COLAMN (*Novel Computing Architecture for Cognitive Systems based on the Laminar Microcircuitry of the Neocortex*) project to complete this task.

I would like to thank Prof Leslie Smith in the University of Stirling and Prof Trevor York in the University of Manchester for their valuable feedback in the PhD examination.

I cannot end without thanking my parents (*Heenbanda* & *Nilenthige*), wife (*Nidarsha*), daughter (*Jayani*) and brothers (*Piniwan* & *Layan*), for their encouragement and love, I have relied throughout my time at the academy.

The Author's List of Publications

- Wijekoon J.H.B and Dudek P., (2011) "Analogue CMOS Circuit Implementation of a Dopamine Modulated Synapse", IEEE International Symposium on Circuits and Systems, ISCAS 2011, Rio de Janeiro, Brazil, pp.877-880.

- Indiveri G. et al. including Wijekoon J. H. B, (2011) "Neuromorphic silicon neuron circuits" **Journal of Frontiers in Neuromorphic Engineering,** Vol 5, May 31, pp.1-23.

- Wijekoon J.H.B and Dudek P., (2009) "A CMOS circuit implementation of a spiking neuron with bursting and adaptation on a biological timescale", IEEE Biomedical Circuits and Systems Conference, BioCAS 2009, Beijing, China, pp.193-196.

- Wijekoon J.H.B and Dudek P., (2008) "Integrated Circuit Implementation of a Cortical Neuron", IEEE International Symposium on Circuits and Systems, ISCAS 2008, Seattle, Washington, USA, pp.1784-1788.

- Wijekoon J.H.B and Dudek P., (2008), "Compact Silicon Neuron Circuit with Spiking and Bursting Behaviour", **Journal of Neural Networks,** Vol 21, Number 2-3, pp.524-534.

- Wijekoon J.H.B and Dudek P., (2007), “Spiking and Bursting Firing Patterns of a Compact VLSI Cortical Neuron Circuit”, 20th International Joint Conference on Neural Networks, IJCNN 2007, Orlando, Florida, USA, pp.1332-1337 (Best Paper Award).

- Wijekoon J.H.B and Dudek P., (2006) “A simple analogue VLSI circuit of a cortical neuron”, IEEE International Conference on Electronics, Circuits and Systems, ICECS 2006, Nice, France, pp.1344-1347.

CHAPTER 1 : INTRODUCTION

1.1 Motivation

Biological nervous systems perform sophisticated functions vital to intelligent behaviour, such as formation of sensory perceptions, object- and event- representations, conscious thoughts, and motor control decisions. They do so with remarkably low energy consumption. These psychophysical functions are processed using massively parallel neural networks that are built with slow, imprecise and heterogeneous neural elements. Impressively, even with such fuzzy units these systems work robustly against noise and exhibit remarkable fault tolerance. These systems outperform modern computers in intelligent decision making ability. Therefore, understanding their fundamental processing principles will provide a huge step forward in science, and help to formulate engineering principles of building intelligent machines.

The Primate brain is a complex architecture produced by evolution that contains approximately one hundred billion neurons, where each neuron is connected up to tens of thousands of other neurons. Of all the brain regions, the neocortex is known to perform most of the psychophysical signal processing. However, among other difficulties, limitations in performing neuron-level recordings on animals make it impossible to understand the underlying computational principles of the cortical networks based solely on the available recordings' data. Therefore, there is an ongoing research effort to understand the principles of cortical information processing through simulating cortical networks in software, or taking it one step further, implementing brain-like circuits in electronic hardware (Arthur et al., 2007; Vogelstein et al., 2007; Schemmel et al., 2008).

Since understanding the primate brain's functionality is a challenging problem, a number of multidisciplinary collaborative research projects, such as COLAMN[1], FACETS[2], Blue Brain[3], SECO[4], Daisy[5], ALAVLSI[6], SyNAPSE[7],

[1] COLAMN: "A Novel Computing Architecture for Cognitive Systems based on the Laminar Microcircuitry of the Neocortex" Web link: http://colamn.plymouth.ac.uk/colamn-project/?page_name=Homepage

[2] FACETS: "Fast Analog Computing with Emergent Transient States", Web link: http://facets.kip.uni-heidelberg.de/

SpiNNaker[8], etc., have been initiated to conquer this challenge by researching into different functional areas of the brain from different perspectives while trying to build large scale networks in dedicated hardware. All these initiatives believe that this work might lead to the discovery of fundamental principles, underlying the remarkable computational abilities of the brain. This PhD thesis project has been carried out within the COLAMN project, and aims at formulating a basic set of VLSI (Very Large Scale Integration) circuit blocks that can be used to mimic the function of a neural network of the neocortex. The thesis provides compact circuit implementations of neuron level models. Furthermore, a prototype of cortical neural layer integrated circuit that has the structure similar to a small cortical layer of neocortex has been designed.

Mixed signal VLSI implementations have the potential of building neural systems with similar properties to those of biological systems. These systems can be used as an emulation platform to support the understanding of the processing principles of neural networks and also to pave the way towards realising potential low power real-time intelligent computing devices and control systems, including the devices that can interface with central nervous systems, or replace parts of the nervous system damaged through disease of injury (Vogelstein, 2007). Hence, despite the fact that it is not fully understood how brains process information, it would be worth implementing efficient, tailor made VLSI circuits that mimic the known cortical neural circuits and networks to reproduce their dynamics. Together with constructive feedback from and to other research disciplines, hardware implementable neural models are more likely to emerge, and implementation of an efficient and effective intelligent computing architecture (a brain like computer) may become possible.

---

[3] Blue Brain: "Blue Brain Project", Web Link: http://bluebrain.epfl.ch/

[4] SECO: "Self-Constructing Computing Systems", Web Link: http://www.seco-project.eu/

[5]Daisy: "Project to reverse-engineer the 'daisy architecture' (Neocortex's Uniform Architecture)", Web Link: http://daisy.ini.unizh.ch/

[6]ALAVLSI: "Attend-to-learn and learn-to-attend with neuromorphic, analogue VLSI", Web Link: http://alavlsi.ini.uzh.ch/

[7]SyNAPSE: "Systems of Neuromorphic Adaptive Plastic Scalable Electronics; sponsors by Defense Advanced Research Projects Agency (DARPA, USA)", Web Link:http://www.darpa.mil/Our_Work/DSO/Programs/Systems_of_Neuromorphic_Adaptive_Plastic_Scalable_Electronics_%28SYNAPSE%29.aspx

[8] SpiNNaker: "A Universal Spiking Neural Network Architecture", Web Link: http://apt.cs.man.ac.uk/projects/SpiNNaker/

If an intelligent processor can be implemented in hardware, it can be used to recognise complex patterns, perform complex motor control, perform autonomous learning, etc. It can also be used in applications that require robustness against noise and fault tolerance. These potential applications make the brain-inspired system an attractive alternative computing model which could be appropriate for designing systems in present and future integrated circuit technologies. Furthermore, in the long-term this line of research may help to understand the brain, potentially leading to the discovery of drugs for a variety of neuro-degenerative diseases such as Alzheimer's and Parkinson's.

1.2 Neuromorphic Engineering

Neuromorphic engineering is the discipline of developing electronic devices that mimic the operation of biological brains. Since the pioneering work of Carver Mead (Mead, 1989) on neuromorphic circuits, in the late 1980s, there has been a continuing interest in developing neuromorphic devices. In particular CMOS implementations of 'silicon neurons' (Mahowald et al., 1991; Linares-Barranco et al., 1991; Schultz et al., 1995; Patel et al., 1997; Simoni et al., 1999; Indiveri 2003; Young Jun Lee et al., 2004; Nakada et al., 2005; Rangan et al., 2010 and Schaik et al., 2010) and 'silicon synapses' (Hafliger et al., 1997; Bofill-i-Petit et al., 2004; Indiveri, 2006; Koickal et al., 2007; Tanaka et al., 2007) have been a subject of on-going development. Recently, a number of systems have been proposed (Arthur et al, 2007; Vogelstein et al., 2007; Schemmel et al., 2008) that attempt to integrate thousands of silicon neurons and synapses in a single chip to build neural networks.

1.3 The Strategy used in Designing the Proposed VLSI Neural Circuits

Problem solving approach used in computers and the way the primate brain solves problems seems to be fundamentally different. Computer processors are built with a set of logic gates and memory elements as precise constituents to perform Boolean logic operations using the logic alphabet "0" and "1". Over the years, the medium in which these machines are built (VLSI technology) has evolved mostly to optimise speed processing of these Boolean logic operations. On the other hand, the brain is built with heterogeneous neural elements that employ an imprecise (fuzzy), slow and non-linear processing approach. Hence mimicking neural circuits of the brain in VLSI is a challenging task.

Neural circuits perform neuronal communication electrically by utilising electro-chemical dynamics. The states of a neural circuit can be represented using analogue electrical potentials. The neural circuits can easily be modelled to form equivalent electrical circuits that use basic electronic elements. Therefore, the neural circuits could be mimicked in mixed-signal VLSI more closely with meaningful relation, and with efficient implementation to support the understanding of the neural dynamics. Furthermore, the circuits can be implemented at "accelerated-time" scale, which exploits the technological advantages of the VLSI technology while providing circuits with high computational throughput. Most of the neural circuits proposed in this thesis are designed to operate on "accelerated-time". The accelerated-time circuits operate approximately three or four orders of magnitude faster than the biological-time where the time scale of neural responses is identical with the time scale of the neuronal activities of the biological systems.

When building cortical networks in microelectronic technologies, the network size that is sufficiently large enough to observe or study the dynamics of a cortical network is a basic requirement. In mimicking a cortical network in hardware, compromise has to be made between the richness of the neural dynamics that can be included in the hardware and the size of the VLSI network that can be built. The cortical networks implemented in neuromorphic research (literature review will be presented in Chapter 3) use approximated basic neural models and in most cases, important non-linear dynamics are ignored (such as the complex, non-linear, oscillatory nature of the neurons, and facilitating and depressing or STDP neural dynamics).

The strategy of modelling the neural models using "*approximate hardware model*" is adopted in this thesis to build the cortical network in VLSI hardware. The basic properties of neural dynamics of the well established computational neural models are used as a *guide to arrive at phenomenological circuit models*, implementing *generic compact* VLSI circuits with *biologically plausible* neural dynamics that closely account for biological experimental facts. This will make it possible to construct sufficiently large VLSI neural networks, with rich non-linear dynamics, that could be used to study the cortical network behaviours. This strategy is formulated based on following principles:

- *Guide to arrive at a phenomenological circuit model*: As the neural elements and their dynamics are highly complex and heterogeneous, the computational models do not account for all the known experimental facts. Instead, the computational models approximate the neural dynamics in arriving at a simple meaningful mathematical model that closely account for some set of experimental results (Morrison et al., 2008). Furthermore, the computational models are derived so that they can be presented in a compact analytical form and /or implemented efficiently on digital computers. This does not always translate into efficient hardware implementation. Hence, the established computational neural models are used only as a qualititative guide to arrive at phenomenological circuit models.
- *Compact circuit implementations*: The silicon area consumed by the neural circuitry is a critical factor that decides the maximum possible size of the VLSI cortical network. Therefore, compact circuit implementations of basic neural circuits are a core requirement.
- *Generic circuit implementations*: The generic circuit element can be tuned to represent different types of a basic neural element (neuron or synapse) using a set of externally adjustable voltages. This makes the system flexible both in terms of circuit implementations and when configuring VLSI cortical networks for experiments.
- *Biologically plausible circuit implementations*: The richer neural dynamics of VLSI cortical neural network could resemble the biological cortical network more closely than other implementations that do not use most of the neural dynamics.

1.4 Thesis Structure

The thesis is divided into four parts: Introduction and Literature Review (Chapters 1 to Chapter 3), Core VLSI Neural Circuit Implementations (Chapters 4 to Chapter 6), Fabricated Neural Integrated Circuits (Chapters 7 to Chapter 9), and Discussions and Conclusions (Chapters 10 and Chapter 11).

Introduction & Literature Review

Following the motivation of this research and the strategy used in implementing VLSI neural circuits presented in this chapter, Part I provides a brief introduction to the

biological neurons, synapses and neocortical networks, and a literature review of their VLSI implementations. This gives the supporting background knowledge required for understanding the VLSI neural elements and their network implementations proposed in this thesis.

Chapter 1 The initial parts of this chapter provide the motivations of this PhD research and an introduction to Neuromorphic Engineering. Then the key strategy followed in developing neural elements and their network in VLSI is presented.

Chapter 2 The biological background of cortical neurons and their diversity, synapses and their short- and long-term plasticity rules are covered in the initial sections of the chapter. These provide insight into basic neural elements of cortical networks. Further, this chapter also presents the biological description of neocortical networks, structure of a functional column, and the layered structure of the neocortex. This introduces the background knowledge on the six-layer cortical neural network architecture of the neocortex that supports the understanding of VLSI Cortical Neural Network Architecture, and Cortical Neural Layer Chip presented at the later part of the thesis.

Chapter 3 This chapter presents a review of the state of the art mixed-signal microelectronic implementations that mimic neural circuits (neuromorphic circuits). These include a review of VLSI neurons, synapses and some neural network implementations. The review of synapses includes circuits that implement short- and long-term plasticity rules. An outline of neural network implementations of digital systems is also provided.

Core VLSI Neural Circuits

The section provides core neural circuit implementations. These include Neuron circuits, STDP (Spike-Time Dependent Plasticity) Synapse circuit, Dopamine Modulated STDP Synapse circuit and Short-Term Dynamic Synapse circuits. Except the Biological-Time VLSI Neuron circuit, all the other circuits operate on a three or four order of magnitude faster time scale (accelerated-time) than the biological neural circuits. The three integrated circuit (IC) implementations presented in Part III uses various combinations of these accelerated-time VLSI neural circuits.

Chapter 4 The circuit implementations of two VLSI cortical neuron circuits that operate on different time scales – the accelerated-time and the biological-time are given in this chapter. Both of these generic neuron circuits are capable of replicating many known types of cortical neurons simply by adjusting a few external voltages. The initial section of the chapter provides the computational model used as a guide to arrive at these neuron circuit models. The first neuron circuit presented is an accelerated-time neuron circuit, its operation and the mathematical model is reviewed. The circuit was proposed in my Mphil thesis (Wijekoon, 2007). Since this circuit is used in all the three fabricated ICs presented in Part III the circuit descriptions are briefly provided. The second neuron design presented is a redesign of the accelerated-time neuron to work on a biological-time scale. The circuit design, its operation and simulation results of this Biological-Time Neuron circuit are provided. Finally, merit and demerits of implementing neural circuits in accelerated- and biological-time scales are evaluated.

Chapter 5 The circuit implementations of two types of VLSI synapse circuits that obey different long-term synaptic dynamics– the STDP synapse and the dopamine modulated STDP synapse are given in this chapter. The initial section of the chapter provides the computational model used as a guide to arrive at these synapse circuit models. The first synapse circuit presented is the STDP synapse circuit, the circuit operation, simulation results and the mathematical model of the synapse are provided. The experimental result of this STDP circuit is given in Chapter 8. The second synapse circuit, the Dopamine Modulated STDP synapse circuit is presented with its circuit details, circuit operation, and simulation results. A generic synapse circuit that can be configured to operate in either in STDP or DA-modulated STDP dynamics fabricated in the STDP-DA Synapses Neuron IC, and the STDP synapse fabricated in Cortical Neural Layer IC will be presented in Chapter 9.

Chapter 6 The circuit implementations of four types of short-term plastic synapses that obey different short-term synaptic dynamics– the excitatory depressing, inhibitory facilitating, inhibitory depressing, and excitatory facilitating are given in this chapter. The initial part of the chapter introduces the computational model and its approximated mathematical formulation, which were used as a guide to arrive at these synapse circuit models. The synapse circuit, operation, and simulation results of each of these synapse circuits are also presented. These synapses are fabricated in Cortical Neural Layer IC presented in Chapter 9.

Fabricated Neural Integrated Circuits

This section provides a description of three Integrated Circuits (ICs) fabricated in CMOS technology. These ICs use combinations of core VLSI neural circuits presented in Chapter 4 to Chapter 6. These chips include Cortical Neuron Chip, STDP-DA Synapses Neuron Chip and Cortical Neural Layer Chip (CNL Chip).

Chapter 7 The accelerated-time cortical neurons are fabricated in a chip and the overview, test setup, and experimental results of the chip are presented in this chapter. The chip contains 202 neuron cells, with varied circuit parameters (transistor sizes and capacitances) to obtain circuit parameters for a generic neuron that could be configured to most of the neuron types, so that this neuron can be used in the next generation of ICs. The chip experimental results prove the functionality of the neuron circuit, and behaviours of various neuron types with their set of tuning variables are presented. This neuron circuit is used in the other two IC implementations.

Chapter 8 The accelerated-time neurons, STDP synapses and dopamine modulated STDP synapses are fabricated, and an overview, circuit implementations, test setup, and the experimental results of the chip are presented in this chapter. The chip contains two neuron cells, with 28 generic configurable synapse circuits that can be configured to operate either as STDP or dopamine modulated STDP synapse. As the STDP Synapse circuit consumes less circuit area than the Dopamine Modulated Synapse circuit, the STDP synapse is used in the larger scale CNL chip presented in Chapter 9. The chip experimental result proves the functionality of the STDP synapse and the results obtained from the chips are presented.

Chapter 9 This chapter proposes a Cortical Neural Network Architecture that could use many CNL chips to build a large VLSI cortical neural network. This chapter provides an overview, circuit implementations, and mathematical model of the fabricated CNL chip that contain 120 accelerated-time neurons, 2 100 STDP synapses and 5 460 short-term plasticity synapses. Finally, the Cortical Neural Layer Board, test setup, and the discussion and conclusions are given at the end of this chapter.

Discussion and Conclusions

Chapter 10 The initial section of this chapter presents estimates of VLSI cortical network size that can be built in a 0.35 µm standard CMOS IC, a 90 nm standard CMOS IC, a multi-chip approach, and in wafer-scale integration technology, using the core neural circuits used in the CNL chip. Further benefits of using a 3D integration technology to build the cortical network are discussed. Other factors that could provide problems and limitations in implementing network on these neuromorphic devices are also discussed. The later part of the chapter presents alternative technological approaches that could be used to mimic cortical networks such as organic electronics, novel neural devices, memristor as a synapse, and cell cultures. Finally, the higher abstractions of neural dynamics used to obtain brain-inspired architectures are briefly discussed.

Chapter 11 This chapter provides the conclusion to the thesis.

CHAPTER 2 : BIOLOGICAL NEURONS, SYNAPSES AND NEOCORTICAL NETWORK

The primate brain has a very complex structure that contains, approximately one hundred billion neurons where each neuron is connected up to tens of thousands of other neurons in a highly parallel layered architecture. In addition to the structural complexity of these networks, their neuronal responses are non-linear, and heterogeneous. The main constituents of these cortical networks are neurons and synapses. The circuit models of these constituents are proposed in this thesis, which may enable building of large-scale parallel VLSI network that closely resembles the microcircuits of the neocortex. Therefore, this chapter provides a brief description of a biological neuron, synapses and neocortical network as an introduction to understanding the circuit models and formation of a VLSI neural network. Further, important computational models of these constituents are also listed.

2.1 Biological Neuron

A neuron typically possesses a cell body (called soma), dendrite trees and an axon. The basic structure of a neuron is shown in Figure 2.1.

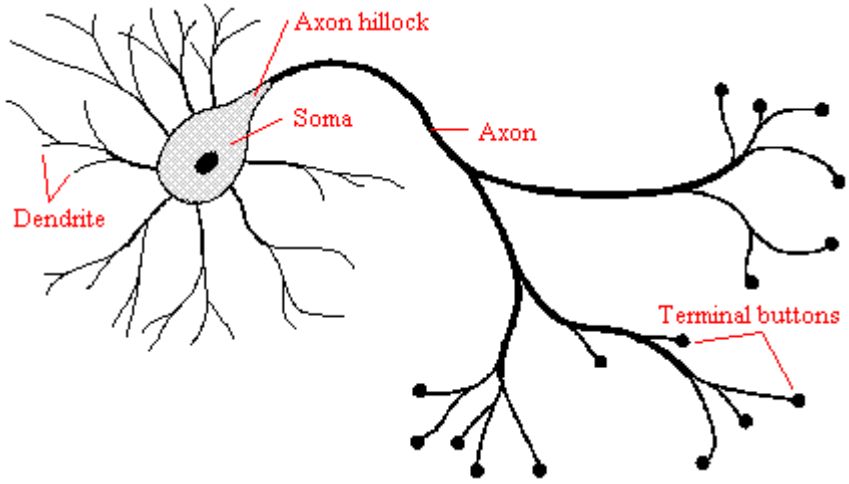


Figure 2.1 Basic structure of a neuron. [9]

[9] Picture taken from http://www.swarthmore.edu/NatSci/echeeve1/Ref/HH/index.htm

The neuron receives input signals from various spatial locations on the dendritic trees, and sometimes also on the soma. These spatiotemporal input signals are integrated onto the membrane capacitance of the soma. Once the integrated voltage (membrane voltage) reaches a threshold, a pulse (also called a firing event or a spike) is generated at the axon hillock. This spike propagates through the rest of the axon and onto adjacent cells through synapses. Simultaneously, a calcium signal slowly propagates backwards through the dendrite (back propagation) towards the input synapses, possibly "informing" them that the neuron has fired. This back propagation influences the dynamics of the input synapses. More details of generation of spikes and computational models of a neuron can be found in (Kandel et al., 2000).

2.1.1 Diversity of cortical neuron

The study of the brain reveals that cortical neurons are diverse in their behaviour and many neuron types have been identified based on their anatomy (or morphology, i.e. structure and organisation of a neuron) and ion channel distribution and composition within a neuron. Therefore, these neurons exhibit different electrical behaviour, transforming the same input signals into different firing patterns. Figure 2.2 shows morphologically different neuron cells found in the monkey's cerebral cortex. A few examples of signalling behaviours of some of the known diverse neuron cells and their morphologies are shown in Figure 2.3.

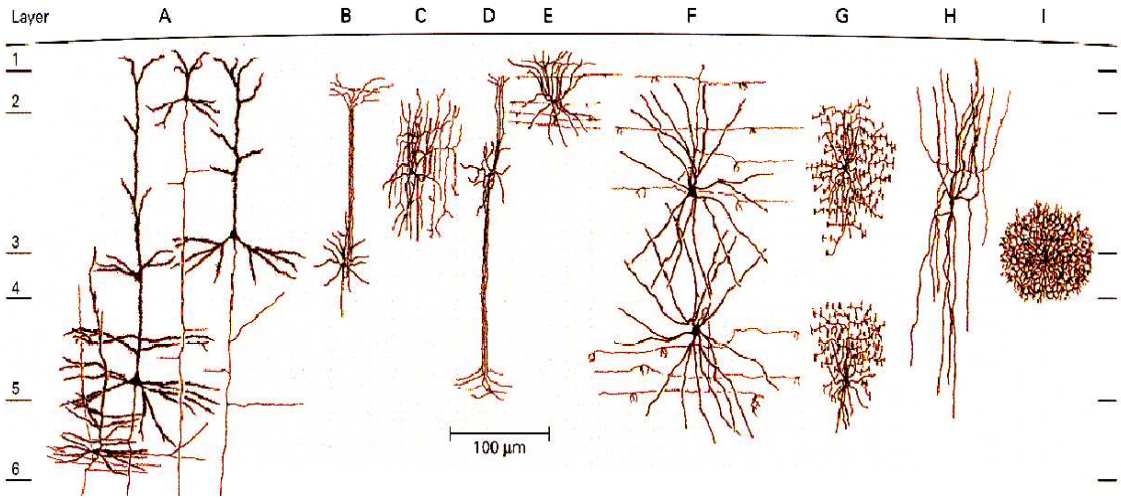
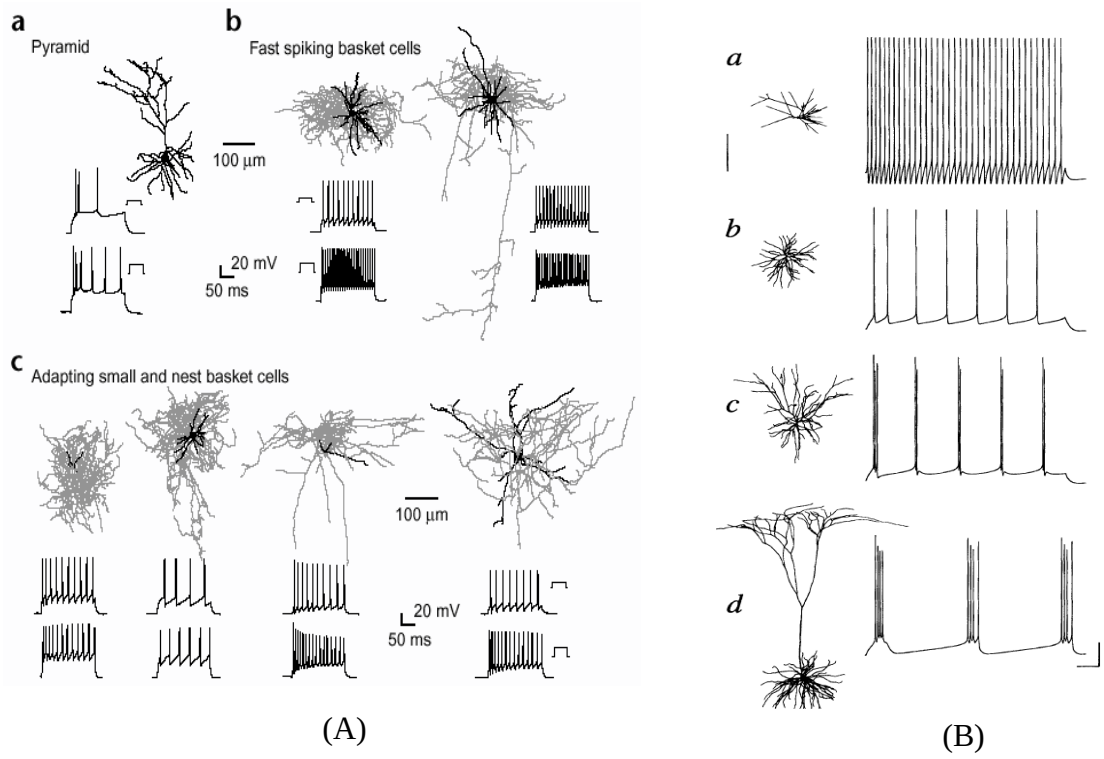


Figure 2.2 Morphological variety of cortical neurons found in monkey cerebral cortex. (A) Pyramidal cells. (B) Spiny stellate cells. (C) Bitufted cells. (D) Double bouquet cells. (E) Small basket cells. (F) Large basket cells. (G) Chandelier cells. (H) An undesignated cell, sometimes called a long stringy cell. (I) Neurogliaform cells. adapted from Well (2005).

A number of approaches to classifying neurons based on the electrophysiological recordings have been introduced (Connors et al., 1990; Markram et al., 2004; Nowak et al., 2003; Petilla Convention[10]; Toledo-Rodriguez et al., 2003). Many parameters, such as spike frequency, interspike-interval histogram, spike width, intra-burst frequency, adaptation index etc. can be used to classify the neurons. A summary of the basic classification important in designing the neuron circuits presented in Chapter 4 is given below.



(A) (B)

Figure 2.3 (A) Different morphological neurons and their spike patterns, (taken from Callaway et. al. 2000). (B) Distinct firing patterns in model neurons with identical channel distributions but different dendritic morphology, taken from Sejnowski (1996).

The neuronal response to a step stimulus of suprathreshold current (post-synaptic input current that causes action potentials) displays either spiking or bursting firing behaviour.

[10] Petilla Convention (2005)

Web Link: http://krasnow.gmu.edu/cng/petilla/

The spiking neurons are of two types: regular spiking (RS) and fast spiking (FS) (Nowak et al., 2003). The RS cells exhibit an accommodation (also known as adaptation) property: in a response to a supra-threshold current step they fire repeatedly, with a decreasing frequency, until the firing rate reaches a stable value, which depends on the input current. The RS cell class can be further sub-divided into two sub-types, the weak accommodating cells are called RS1 and strong accommodating cells are called RS2 (Toledo-Rodriguez et al., 2003). Examples of morphological cell that behave as RS1 type are neocortical layer II–VI pyramidal cells. The RS2 type cells are neocortical layer IV–VI pyramidal cells and spiny stellate cells (Connors et al., 1990). The FS cells fire repetitively at high frequency with little or negligible accommodation to a sustained supra-threshold current injection. The action potentials of FS cells exhibit faster rise rate, fall rate and distinct fast after-hyperpolarisation (Connors et al., 1990). Some neurons with FS behaviour commonly found in the cortex are, for example, neocortical small basket cells, nest basket cells, bitufted cells and large basket cells (Toledo-Rodriguez et al., 2003). The basic bursting cell types are chattering (CH) and intrinsic bursting (IB) (Nowak et al., 2003). The CH neurons usually display repetitive long clusters of spikes to a sustained supra-threshold current injection. The IB neurons respond to a step current injection with a cluster of three to five initial spikes followed by an after hyperpolarisation, and then by either single spikes or burst at more or less regular intervals (Toledo-Rodriguez et al., 2003). These types are observed in sub-populations of bitufted cells, bipolar cells and Martinotti cells in the neocortex (Connors et al., 1990).

Distinct firing patterns obtained from the reconstructed models of morphologically different neurons with identical channel distributions are given in Figure 2.3 (B) (Figure adopted from Sejnowski, 1996). A simple computational model that reproduces basic electrophysiological properties of known types of cortical neurons can be found in Izhikevich (2003). The model demonstrates some processing properties due to dendritic morphology and ion channel distribution, in addition to the neural dynamics on the cell body. The circuit implementation of an approximated model of this computation model is given in Chapter 4.

2.2 Biological Synapses

Neuron to neuron information transfer is carried out via a specialized element called a synapse. Synapse usually forms connections between the axon of a pre-synaptic neuron and a dendrite or cell body of a post-synaptic neuron. However, there exist synapses that directly connect dendrite to dendrite or dendrite to soma (Well, 2005). Although the synapses are heterogeneous between different brain areas and between different neuron types, synapses mainly can be classified into two types: a chemical synapse and an electrical synapse. The first type is mostly found in cortical networks and has complex synaptic dynamics which thought to be involved in learning, memory, and cortical plasticity (Morrison et al., 2008). Therefore, chemical synapse is a basic and important building block in neural computational and circuit models. Hence, the dynamics and circuit models of the chemical synapse are discussed in this thesis.

The chemical synapse transmits signals to another neuron by means of chemical reactions. Once the pre-synaptic neuron fires, the electrical spike sent down the axon terminal transmits to the adjacent dendrite through the synaptic cleft by converting the signal into a chemical signal. I.e. when an electrical signal arrives at the synapse, the neurotransmitters release into the synaptic cleft, some of these neurotransmitters are able to reach receptors at the dendrite spine –in which the chemical signal is converted back to an electrical signal. Then, this signal propagates to the soma of the adjacent cell. The electrical signal before the synapse is called a pre-synaptic signal and after the synapse is called post-synaptic signal. Figure 2.4 shows a sketch of a chemical synapse. The post-synaptic signal can be inhibitory (post-synaptic neuron's membrane is hyperpolarized) or excitatory (post-synaptic neuron's membrane is depolarized) depending on the type of neurotransmitters-receptor combination that facilitates the signal transmission between the two cells. During the synaptic transmission, the pre-synaptic action potential is shaped to carry extra information pertaining to the state of the synapse. At a given time, amplitude, rise time and fall time of the post-synaptic pulse is determined by the short-term plasticity of the synapse. These dynamics could change the strength of synaptic connection between neurons depending on the pre-synaptic activity. In some types of synapses, the long-term synaptic plasticity also contributes to this greatly. These dynamics even have the ability to form or eliminate its synaptic connection depending on the neural activities of the network. These short-term

and the long-term dynamics of synapses are discussed below and their some computational models are listed.

In addition to the aforementioned long-term synaptic dynamics, homeostatic changes of synapses could change the amplitudes of the synaptic response on a slow time scale of hours is called “synaptic scaling” as referred in Morrison et al., (2008) (Turrigiano et al. 1994). This can be useful to stabilise the neuronal firing rates (Morrison et al., 2008).

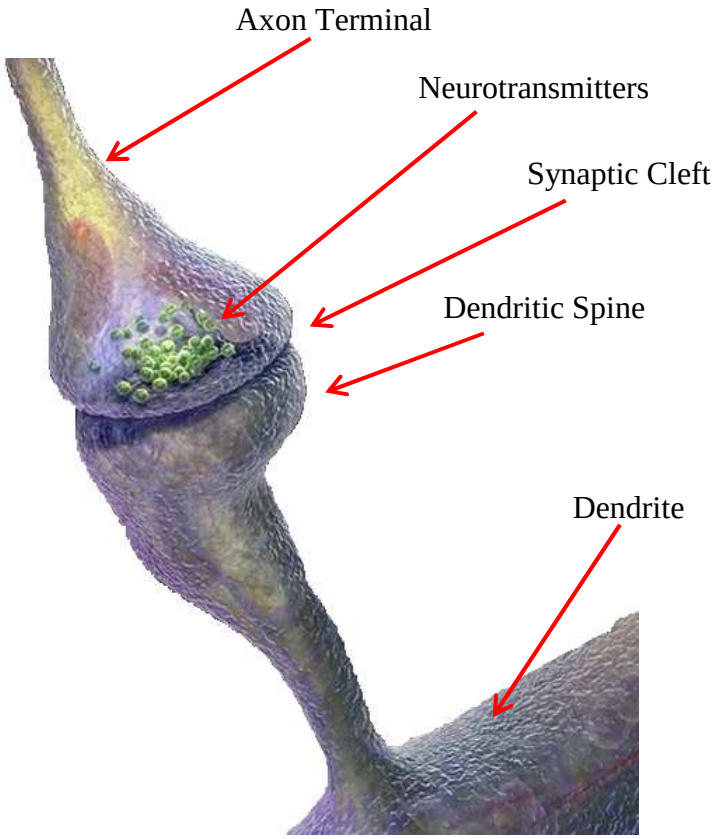


Figure 2.4 Basic structure of a chemical synapse[11]

2.2.1 Short-term dynamics of the synapse

The amplitude, rise time and fall time of the post-synaptic potential due to short-term plasticity depends on the properties of the constituents of the synapse. In addition to that, the temporal pattern of the incoming spike train also determines the amplitude of the post-synaptic pulse. Each successive incoming spike can cause the amplitude of the post-synaptic pulse to be either smaller (depression) or larger (facilitation) than the

[11] Picture modified from

Web Link: http://www.noeticsciences.co.uk/wp-content/uploads/2009/11/Synapse-Structure.jpg

previous one (Figure 2.5 shows facilitating and depressing dynamics). These dynamic temporal scales can range from 100 ms to about a second (Morrison et al., 2008) and the amplitude of the post-synaptic response recovers to close to normal values within less than a second (Markram et al., 1998; Thomson et al., 1993). Biological evidence on these dynamics are published in Gupta et al. (2000), Markram et al. (1998), and its computational models are proposed by Tsodyks et al. (2000), Abbott et al. (1997), and Thomson et al. (2007). The descriptions of the approximated models of these computational models used to implement short-term synaptic dynamics in VLSI circuits are given in Chapter 6.

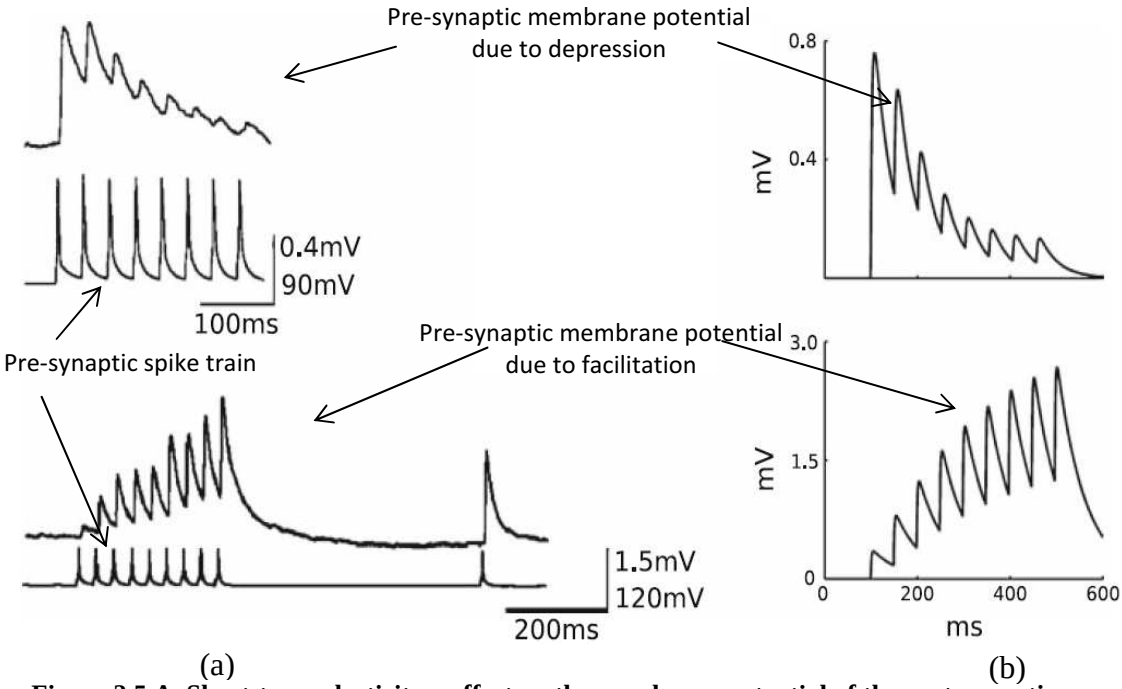


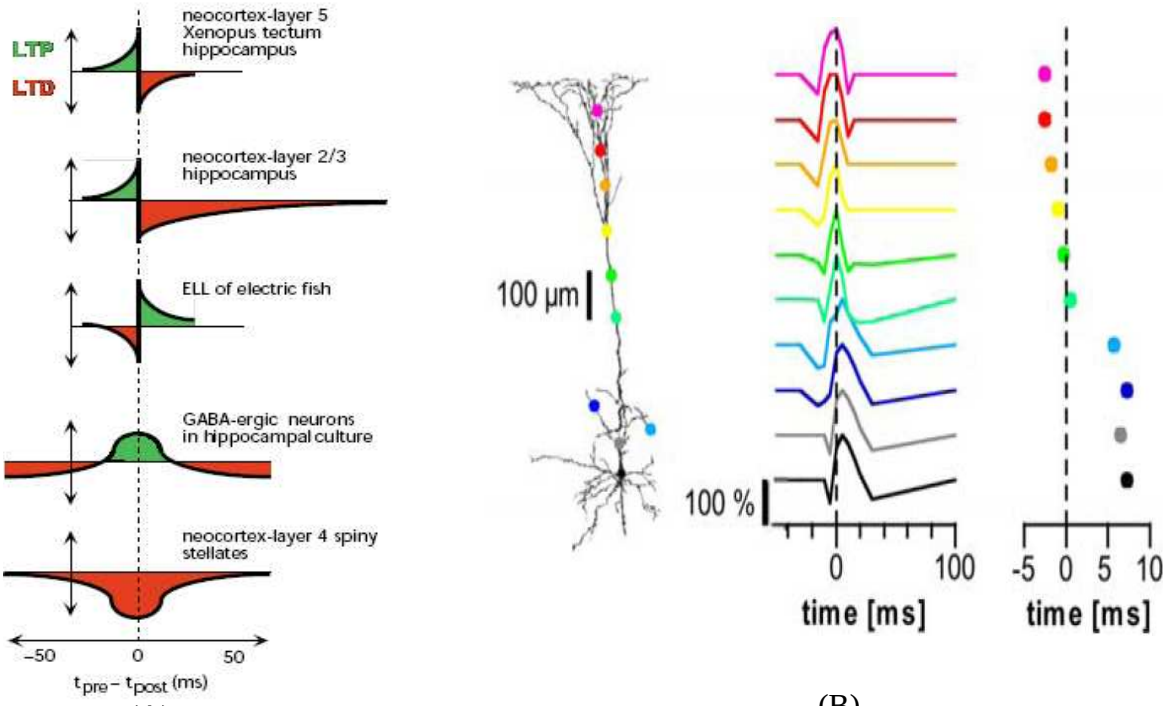
Figure 2.5 A. Short-term plasticity – effect on the membrane potential of the post-synaptic neuron due to pre-synaptic spike train (a) experimental results from rat cortex in slice Markram et al. (1998), (b) simulation results, Markram–Tsodyks model, Tsodyks et al. (2000); taken from (Morrison et al., 2008).

2.2.2 Long-term dynamics of the synapse – STDP synapses

In some synapses, the post-synaptic pulse is greatly influenced by the long-term plasticity dynamics, which depends on the actions of the post- synaptic neuron; i.e. if the post- synaptic neuron fires, the back propagating signal influence the input synapses to depress (reduce the strength of the synaptic transmission, and called long-term depression, LTD) or potentiate (increase the strength of the synaptic transmission, and it's called long-term potentiation, LTP). Amount of depression or facilitation depends on the time difference between pre- and post- spike firings (t_{pre} - t_{post}). This phenomenon

is called spike-time dependent plasticity (STDP) and plays a critical role in synaptic plasticity, which is the cellular mechanism for learning and memory. Experimentally observed STDP curve that defines LTP and LTD relationship with respect to time difference between pre- and post- spike firings is diverse and depends on the synapse and the neuron type (Abbott and Nelson, 2000; Bi and Poo, 2001). Some basic types of curves that are observed during experiments are shown in Figure 2.6 (taken from Abbott and Nelson, 2000). It is also observed that the dendritic distance from soma to synapse has effect on the shape of the STDP curve –it is shown in Figure 2.6 (taken from Letzkns et al., 2006). This is due to the dendritic-filtering of the back propagation signal (Letzkns et al., 2006; Saudargiene et al., 2005).

The standard STDP curve most popular in theoretical research on STDP is the topmost graph in Figure 2.6 –this is the mostly observed type of STDP in neocortical synapses. The computational model of the standard STDP rule can be found in Morrison et al. (2008). A mathematical model of a STDP curve that can implement compact STDP circuit and its circuit implementation are given in Chapter 5.



(A) (B)

Figure. 2.6. (A) different STDP curves found in synapses of different neurons (taken from Abbott and Nelson, 2000); (B) Model of the layer 5 pyramidal neuron showing the color coded location of synaptic inputs; Center, Color-coded STDP timing curves for synapses at the dendritic locations in the model. Right, Positive peaks of STDP timing curves (LTP) color-coded for each dendritic site show a shift from positive to negative spike timing with distance from soma. (taken from Letzkns et al., 2006)

2.2.3 Long-term dynamics of the synapse - Dopamine modulated STDP synapses

Apart from general form of STDP synapses discussed above, there exist synapses that the STDP is modulated by the extra-cellular Dopamine (DA) level. DA is a neuromodulator in the nervous system that regulates diverse populations of neurons. It originates from small groups of neurons in the mesencephalon (including the ventral tegmental area) and diencephalon areas of the brain. The neurons whose primary neurotransmitter is dopamine are called dopaminergic neurons. The brain areas where these neurons are present are known to carry normal brain functions such as working memory, reinforcement learning, and attention (Fellous et al., 2003). Even though these neurons are found in few brain regions only, their projections are generally highly diffuse and reach large portions of the brain (Fellous et al., 2003). The burst stimulation

of the dopaminergic neuron releases DA globally to many DA modulated synapses. This increases the extracellular DA concentration at the synapses enhancing their long-term potentiation (LTP) and/or depression (LTD) (Gurden et al., 2000; Otani et al., 2003). This effect of DA plays a major role, in particular, in reinforcement learning. The computational model of the DA modulated STDP synapse can be found in Izhikevich (2007) and its VLSI circuit implementation is given in Chapter 5.

2.3 Neocortical Network

The primate brain functions are carried out with a complex architecture that contains approximately hundred billion neurons where each neuron is connected to thousands or even tens of thousands of other neurons in a highly parallel layered architecture. Among these the largest network, the neocortex, confines billions of neurons to a few millimetres thick single folded sheet of neural tissue at the outer layer of cerebrum. The neocortex consists of a six-layer laminar structure and this organisation tends to be more homogenous throughout the neocortical tissue. About 80% of neurons in a neocortex are excitatory neurons, and others are inhibitory neurons (Somogyi, 1989, White, 1989; Peter et al., 1984). Anatomically, most of the excitatory neurons receive synaptic inputs from non-STDP excitatory and inhibitory depressing synapses and from excitatory STDP synapses. The inhibitory neurons receive inputs from inhibitory facilitating and excitatory depressing synapses, whereas some other inhibitory neuron types receive input spikes from excitatory facilitating and depressing synapses (Roth and Wennekers, 2009).

Most of the psychophysical signal processing of the brain is believed to be taken place in the neocortical brain areas (Well, 2005). Different areas of neocortex perform different psycho-physical functions. For example, the visual cortex, primary cortex, auditory cortex and in humans the ventrolateral prefrontal areas does vision, motor, hearing and complex language related processing respectively. However, processing of any psycho-physical phenomenon appears to have distributed functionality with many different cortical and non-cortical areas of the brain making important contributions to the processing of such function (Well, 2005). The experimental evidence suggests that neocortex could be divided up into small processing units called functional columns.

These functional columns seem to occupy lateral areas of a few tenths of a millimetre in diameter and extend down through the entire thickness of the neocortex.

2.3.1 Structure of a functional column

Though there is no such strict anatomical division of a functional column found in the neocortex, it is observed that there is a synchronised activation of neighbouring cells to process certain tasks. I.e. the neighbouring cells assemble together to perform certain tasks. Thus the hypothesis of generic functional cortical column is introduced. This generic functional column tends to consist of tens thousands of neurons with diverse behavioural properties. Each of these neurons connects to tens thousands of other neurons via synapses forming a column of processing unit. Between species these functional column only vary from 300 to 600 µm in diameter where as their brains differ in volume by a factor of 10^3. Functional columns are assumed to formed by cortical circuits, effectively 're-wiring' their lateral connections in response to control signals so that at least some neurons are capable of 'being part of' many different possible functional columns (Well, 2005). The circuit of the functional column is called cortical microcircuit of the cortex.

2.3.2 Neocortical Layers

As illustrated in the Figure 2.7 six layers cortical of architecture can be initially divided into 3 main layers: Supragranular layers (layer I, II and III), Granular layer (layer IV) and Infragranular layers (layer V and VI).

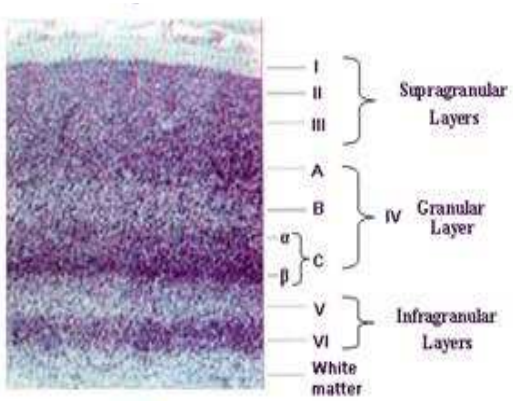


Figure 2.7 Cross section of a small area of neocortex showing anatomical division of six layers.[12]

[12] Picture taken from http://acces.inrp.fr

The Supragranular layers make up of layers I, II and III; the layer I occupies dendrites and axons coming from neurons in the deeper layers (from layers II and III pyramidal cells, the principle cell type in the cortex), therefore distal synaptic connections of those deeper layer neurons are formed in this layer. This layer also consists of few inhibitory neurons. The Layer II contains a mix of small pyramidal cells and some inhibitory neurons. It also contains apical dendrites coming from layer VI and layer V pyramidal cells. Majority of cells in layer III are small pyramidal cells. However this layer contains almost all the cell types found in neocortex. Layer IV, the granular layer contains spiny stellate cell and variety of inhibitory cells. The layer 4 receives most input from thalamus and is sub divided into 4 layers, labelled 4A, 4B, 4Cα, and 4Cβ. The Infragranular layers composed of layer V and VI. The layer V is composed of small number of inhibitory cells and many large pyramidal cells. Some pyramidal cell axonal outputs target the basal ganglia, brain stem, and spinal cord passing through the white matter with long axons projections. Special type of inhibitory cell, chandelier cells make synaptic connections only to the axons protruding from other neurons, are often found in layer V (Well, 2005). Layer VI is the final layer on top of white matter. Most of the cells in this layer are large pyramidal cells that project their axons back to the thalamus. It also contains class of inhibitory neurons cells whose axonal outputs make long projections across all layers of the neocortex (Well, 2005).

CHAPTER 3 : NEUROMORPHIC IMPLEMENTATIONS - NEURONS, SYNAPSES AND NETWORKS

This chapter presents the literature review of silicon neuron, synapse and neural network circuits. In evaluating these circuits, in addition to biological plausibility, compact implementation of circuit blocks is a core requirement, particularly considering the feasibility of building large cortical networks.

Some of the proposed neural circuits (e.g. Schemmel et al., (2008)) are operating at speeds far exceeding those of biological neural communication ("accelerated time"), and are intended to provide a computationally powerful simulation acceleration tool. This is motivated by the relative ease with which electronic circuits can operate at frequencies much higher than these typically observed in biological neural systems (e.g. typical mean firing frequencies of neurons in the cortex are in the order of 10 Hz and the time courses of membrane potentials have bandwidth limited to several kHz). Further, these circuit designs exploit the technological advantage of high speed optimised CMOS technologies rather than operating in non-optimised sub-threshold regime which is the case of "biological-time" circuit implementations. The technological constraints of the common communication infrastructure used for neuromorphic hardware, i.e. the address event representation (AER) framework (Boahen, 2000), as well as the desire to interface directly to sensors that operate on signals encountered in nature and at timescales similar to biology, lead to the situation that most of the silicon neural circuit proposed in the literature operate in biological real-time. The circuits that operate both of these time scales are outlined in this literature review.

3.1 VLSI Neurons

This section outlines the literature review of silicon neuron circuits. The latest review of the neuron circuits implemented by the neuromorphic research community is presented in Indiveri et al. 2011. Amongst the silicon neurons, several neuron models have been considered as a basis for circuit implementation, from integrate-and-fire (I&F) neurons (the I&F neurons integrate the input currents produced by the synapses and generate output spike trains with mean firing rates proportional to their input currents)

(Schemmel et al., 2010; Indiveri et al., 2006; Chicca et al., 2003; Haflinger et al., 1996), to non-linear conductance-based (Arthur et al, 2007; Vogelstein et al., 2007) and Hodgking-Huxley like (Zou et al., 2006; Farquhar et al., 2005) models. The latter are of particular interest, as they exhibit much richer dynamics and thus possible repertoire of spiking behaviours, both in the context of the network and individual responses to a fixed stimulus. However, these circuits use a larger number of transistors. Several other implementations have been proposed (Linares-Barranco et al., 1991; Patel et al., 1997; Young et al., 2004; Nakada et al., 2005) that are based on mathematical models that capture some of the features of the neuron's oscillatory behaviour. In evaluating these circuits the silicon area needed to implement the circuitry is an important consideration in addition to the heterogeneity of neural behaviours, as seen in biological neurons. Although the direct comparisons of different neuron circuits found in literature are difficult due to the lack of their implementation details; the transistor count is considered as an indication of the overall circuit area requirements.

Table 3.1 Summary of VLSI neuron circuits

Neuron model	**Approximate No. of transistors**	**Spiking pattern**	**Biological plausible spike pattern**	**Reference**
Conductance-based	27-30+	Simple spike	good	Mahowald et al. 1991
Integrate-and-fire	18-20	Simple spike	fair	Indiveri 2003
FitzHugh-Nagumo	20	Oscillatory	envelope	Linares-Barranco et al. 1991
Morris-Lecar	20	Oscillatory	envelope	Patel et al. 1997
Resonate-and-Fire	20	Oscillatory	pulse	Nakada et al. 2005
Hindmarsh-Rose	90	Bursting	fair	Young et al. 2004
Accelerated-time	**14**	**All main types**	**good**	**This thesis**
Biological-time	**23**	**All main types**	**good**	**This thesis**

3.2 VLSI Synapses

As explained in Chapter 2, the synapse transmits incoming pre-synaptic spike onto the membrane of the post-synaptic neuron as a current injection, with a variable gain (known as synaptic weight) that determines the strength of the connection between neurons. In implementing a synapse in VLSI, the basic short-term dynamics: synaptic integration, rise- and fall- time constants of the post-synaptic potential, facilitating and depressing properties of the synaptic weight are of major interest. In addition to these short-term dynamics, some synapses follow the long-term synaptic dynamic such as STDP or dopamine modulated STDP.

Typically, in neuromorphic circuits, the synaptic weight is stored in a capacitor (other implementations use a digital memory element –a register or an analogue floating gate transistor). This weight is used to generate the post-synaptic current when a pre-synaptic spike arrives at the synapse. Therefore, in order to implement short-term dynamics this capacitor needs charging or discharging accordingly– in the context of this report, the circuit that does this as well as generates the post-synaptic current, is referred to as the short-term plastic synapse circuit. The synapse circuit that equipped with both short- and long- term dynamics is called STDP synapse circuit. Following sections review some of the short-term plastic synapse circuit and STDP synapse circuit implementations. However, a VLSI implementation of a DA modulated synapse has not been reported in literature.

3.2.1 Short-term Plastic Synapse Circuit

A detailed review presented by Bartolozzi et al. (2007) discusses short-term plastic synapse circuits published in literature –starting from the primitive Pulse Current-Source Synapse circuit (Mead, 1989) up to the Diff-Pair Integrator Synapse circuit (Bartolozzi et al., 2007) by covering the following synapse circuits: Reset and Discharge (Lazzaro, 1994), Linear Charge-And-Discharge (Authur et al., 2004), Current-Mirror Integrator (Boahen, 1997) and Log-Domain Integrator (Merolla et al., 2004). Among these implementations, the Diff-Pair Integrator Synapse circuit performs linear integration of input spikes with tunable gain parameters and has one independently tunable time-constant parameter. As pointed out in the review, the other synapse circuits do not provide “proper” linear integration of spikes. This may lead to a

loss of incoming information to the synapse. That is, if another spike arrives during the period in which the first spike has an effect, the second spike is ignored, hence the information belonging to the second spike is lost. However, if the synapse is designed such that the post-synaptic current is injected in the form of a current impulse (i.e. if the rise- and fall- time of the post- synaptic potential are not considered), the incoming spikes to the synapse are not lost. Above mentioned synapse circuits are implemented as the sub-threshold (weak-inversion) CMOS circuits and use a capacitor to store the weight of the synapse. However, these implementations do not consider the effects of dendritic processing; Basic implementations that include dendritic integration are presented in Elias et al. (1995) and Rasche et al. (2001). All above mentioned implementations do not address the facilitating or depressing short-term dynamics as presented in biological models (Tsodyks et al., 2000; Dayan et al., 2001; Abbott et al., 1997). The basic Diff-Pair Integrator Synapse with both short-term depression and facilitation has been proposed in Liu (2003).

The synaptic circuits proposed in this thesis include short-term depressing and facilitating dynamics. The post-synaptic current is injected in the form of a current impulse where the magnitude of the impulse represents the synaptic weight. This reduces the circuit area consumed by the synapse circuit.

3.2.2 STDP Synapse Circuit

In a cortical network, STDP synapses provide activity driven rewiring of neurons with weighted connections. Hence, STDP synapse plays an important role in cortical neural networks that perform adaptability, learning and memorising. Among the various STDP synapse circuits presented in the literature, the basic STDP circuits proposed by Hafliger et al. (1997), Bofill-i Petit et al. (2004), Indiveri (2006), Koickal et al. (2007), and Tanaka et.al. (2007) are discussed below. Other implementations include a bimodal probabilistic plasticity STDP circuit based on membrane voltage level (Fusi et al., 2000; Badoni et al., 2006), a model of plasticity based on intracellular calcium levels (Rachmuth et al., 2003), and a mixed-signal STDP implementation (Schemmel et al., 2004).

STDP synapse circuit by Hafliger et al. (1997)

This is the first known neuromorpic synapse that implements time-dependent learning rule. It is a weight-dependent synapse implementation (where the values of LTP and/or LTD influence the value of the weight –this is an important criterion for synapse stability). However, only the potentiation aspect of STDP is considered in this circuit; i.e. it performs weight updates based on single pairs of pre- and post-synaptic spikes as shown in Figure 3.2. The circuit operates in biological time scale and is fabricated in 2 µm standard CMOS process. The schematic of the circuit is given in Figure 3.1. The circuit occupies larger silicon area and provides less functionality compared to other models discussed in this chapter.

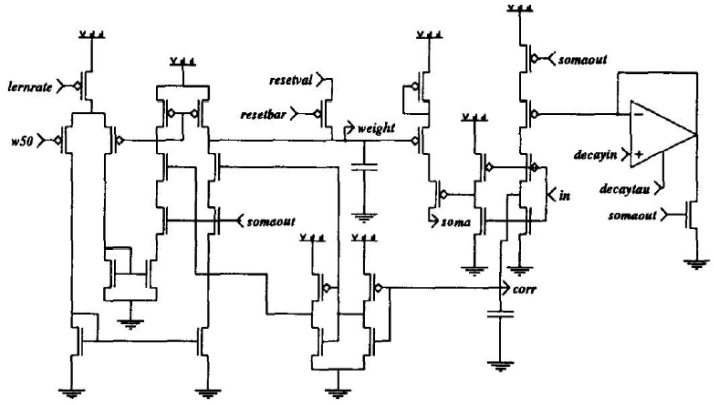


Figure 3.1 The CMOS synapse circuit of Hafliger et al. (1997); weight capacitor hold the weight, the *corr* capacitor stores the correlation signal representation. The magnitude of the weight increment and decrement are computed by a differential pair (upper left *w50*). These circuits are mirrored to the synaptic weight and gated by digital switches encoding the state of the correlation signal and of the somatic action potential. The correlation signal reset is mediated by a leakage transistor, decayin, which has a tonic value, but is increased dramatically when the output neuron fires; taken from Hafliger et al. (1997).

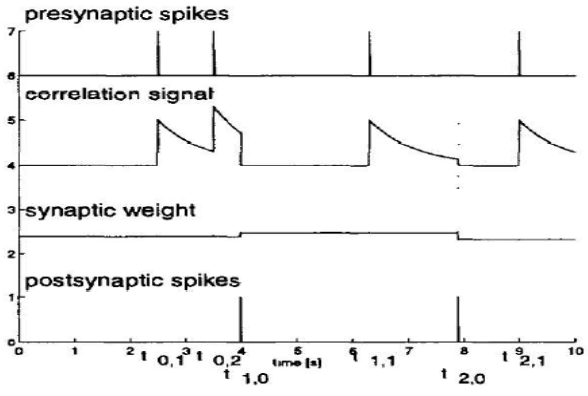


Figure 3.2 The learning rule explained by a snapshot of the simulation variables involved at one synapse; taken from Hafliger et al. (1997).

The STDP synapse circuit by Bofill-i-Petit et al. (2004)

This STDP circuit includes weight-dependent potentiation and depression, in which the degree of weight-dependence is tunable. These circuits operate in biological time scale and are fabricated in a standard 0.6 µm CMOS process. The I&F (Integrate and fire) neuron, the STDP synapse, and the short-term plastic synapse occupy 75 µm by 253 µm, 131.3 µm by 139.7 µm, and 73.2 µm by 21.3 µm ofchip area respectively. This implementation is relatively compact and has exponentially decaying STDP curves as shown in Figure 3.5 (a) and (b). The schematics details of the design are given in Figure 3.3, Figure 3.4 and Figure 3.5 (c).

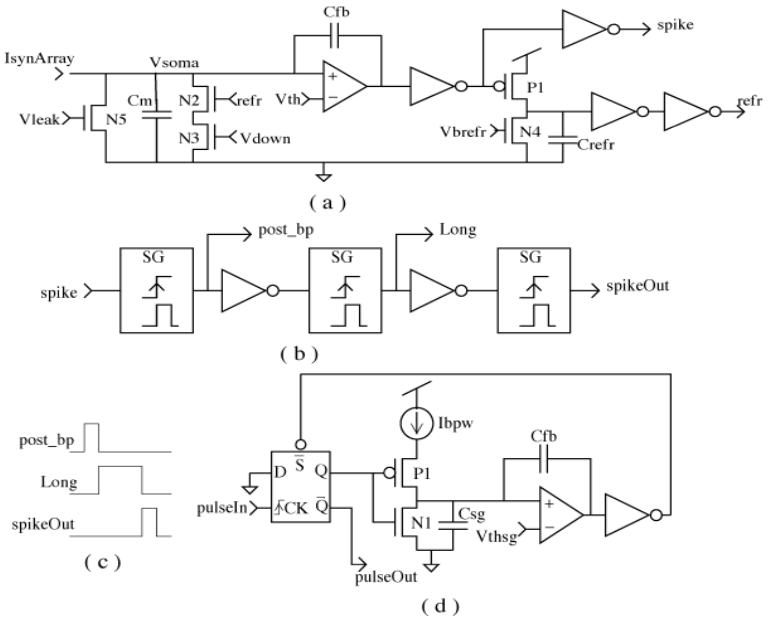


Figure 3.3 (a) Leaky I&F (Integrate and fire) neuron (b). A chain of spike generation circuits (SG) receives a spike signal from the I&F neuron. (c) Waveform generated by the chain of SG circuits. (d) Schematic details of the SG circuit; taken from Bofill-i-Petit et al. (2004).

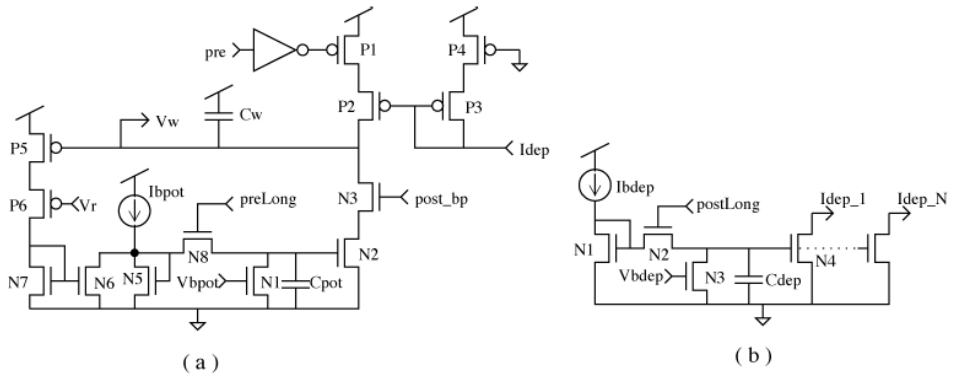


Figure 3.4 The STDP circuit (a) Circuit that detects causal spike correlations, (b) the depressing side of the learning curve; taken from Bofill-i-Petit et al. (2004).

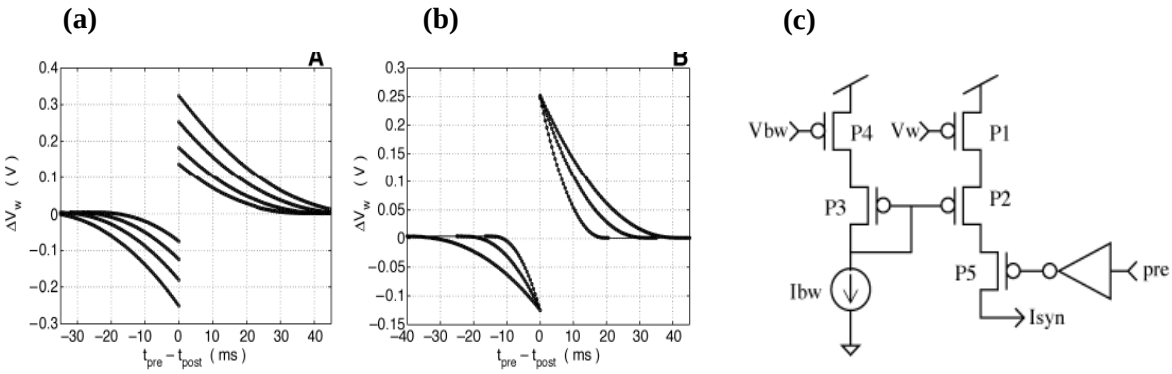


Figure 3.5 Graphs (a) and (b) are experimental STDP curves, showing the possibility of independent adjustment of curves; (c). Synapse output circuit: when pre-synaptic pulse, *pre*, is activated it injects a post-synaptic current proportionate to weight *Vw*; taken from Bofill-i-Petit et al. (2004).

The STDP synapse circuit by Tanaka et.al., (2007)

This implementation constructs a Hopfield-type neural network associative memory using a synapse circuit with STDP that has a symmetric time window. The circuit operates in accelerated time scale and is fabricated in the standard TSMC 0.25 µm CMOS technology. The STDP circuit consumes 6336 μm^2 of chip area.

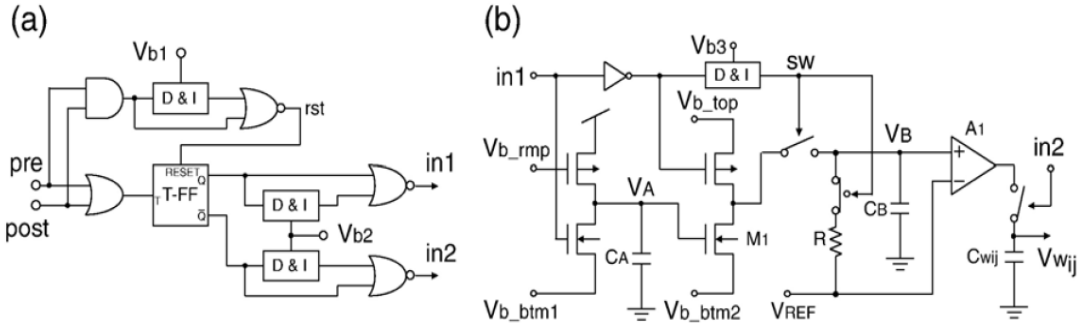


Figure 3.6 (a) Spike-detection and (b) weight-update parts; where, D&I –delay-and-inversion circuit; T-FF –Toggle flip-flop; taken from Tanaka et.al., (2007)

The STDP circuit by Indiveri et al. (2003, 2004)

The circuit is fabricated in standard 0.8 µm CMOS technology and the inhibitory and excitatory synapse measure 55 µm by 31 µm and 145 µm by 31 µm respectively, while the neuron circuitry occupies an area of 83 µm by 31 µm. The STDP synapse circuit

(Figure 3.7 centre) is weight dependent implementation, and it is more compact design compared to other implementations. However, the shape of the STDP curves (shown in Figure 3.8) are less easy to relate to the curves found in biology (Abbott et al., 2000). The STDP synapse circuit has some degree of flexibility in adjusting the curves as seen in Figure 3.8. The circuit operates in biological time scale and due to the continuous leakage of the weight capacitor, in long time scale the weight always becomes significantly biased to one side (either to zero-voltage or maximum-voltage depending on the topology). This effect has been reduced by introducing bi-stability circuit that drives the synaptic weight to one of two possible states on long time scales.

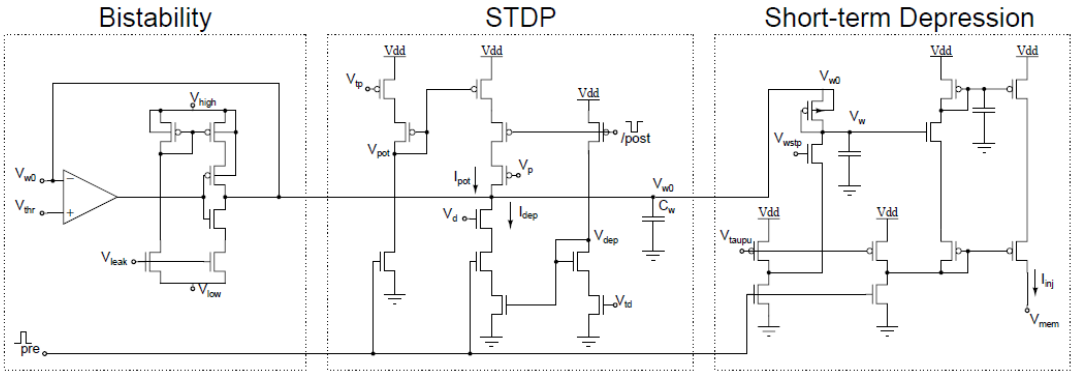


Figure 3.7 Synapse circuit. The bistability circuit compares the voltage *Vwo* to a threshold and drives it to one of two asymptotic values (*Vhigh* or *Vlow*). The STDP circuit increases (or decreases) *Vwo* with every post- (pre-) synaptic spike provided the pre- (post-) synaptic spike was emitted shortly before. The STD circuits implement short-term synapse weight *Vw* with every pre-synaptic spike, at a rate set by *Vwstd*. And the *Vw* is given to a current-mirror-integrator that generates a postsynaptic current and it's injected into the neuron; taken from Indiveri et al. (2003).

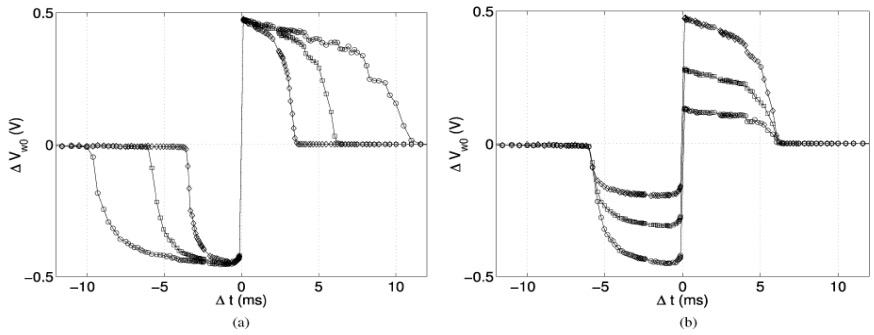


Figure 3.8 The STDP curve: the difference between pre- and post-synaptic spike times Δt = *tpre - tpost*. The curves in the left plot were obtained for different values of *Vtp, Vtd* , while the curves in the right plot were obtained for different values of *Vp* and *Vd* ; taken from Indiveri et al. (2003).

STDP circuit by Koickal et al. (2007)

This is a weight-independent STDP circuit, and the STDP dynamics are modelled as a pair of decaying exponentials. The circuit is fabricated in AMS 0.6um CMOS technology and operates in biological time scale. The circuit consumes a large area compared to Indiveri et al. (2004).

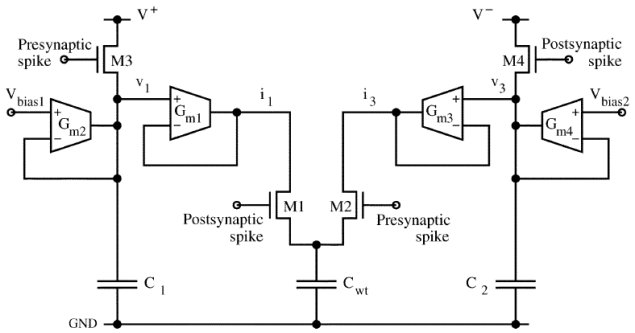


Figure 3.9 Simplified schematic of the STDP learning circuit formed by two symmetrical circuit blocks to implement the positive and negative phases of the learning function; taken from Koickal et al. (2007).

Among these STDP circuits, the Indiveri et al. (2003) weight dependent implementation of the approximated STDP circuit is a compact and simple design. The STDP circuit proposed in Chapter 5 is similar to the Indiveri et al. (2003) STDP circuit, however it operates on the accelerated-time scale.

3.3 VLSI Networks

Implementation of electronic cortical neural network systems that emulate the organisation and the function of the cortical networks of the nervous system has been a continuing interest in brain research (Schemmel et al., 2010; Furber et al., 2006; Indiveri et al., 2007; Merolla et al., 2007; Renaud et al., 2007). The fully digital circuits, and the mixed-signal circuits in which computation is shared between analog and digital hardware elements, have been used to reproduce these networks in hardware. In this section, literature review of selected neural network implementations in mixed-signal VLSI systems is summarised. Further, some cortical neural network implementations in digital systems are also outlined.

3.3.1 Cortical networks in mixed-signal VLSI system

These networks exploit continuously varying analogue signals to compute low-level biological dynamics. The basic neural elements can be implemented in analog circuits with heterogeneity and imprecise (noisy) signal communication. Further, these implementations occupy very small integrated circuit area. Hence, mixed-signal system has become an attractive platform to mimic neural dynamics. But, mixed-signal circuit implementation has complex design flow, and circuits are less flexible to be adapted to perform different tasks. As low-level neural processing principles are very different to conventional digital processing principles, customised analogue mixed-signal circuits could provide efficient implementations of neural circuits. It is also possible to use both analogue and digital techniques to optimise the performance of the full network.

The closeness of the VLSI cortical network dynamics to that of biological network depends on the types of neuron, synapse and connectivity models used in the network. Specially having synapse model with STDP dynamics is very crucial for cortical network plasticity. Large to medium size mixed-signal neural network implementations with STDP synapses include Schemmel et al. (2008), Giulioni et al. (2008), Indiveri et al. (2004, 2007). Other medium size mixed-signal neural network implementations without STDP synapses include Arthur et al. (2004), Merolla et al. (2007). All of these VLSI neural networks are discussed below briefly. Other implementations include small size spiking neural networks such as Renaud et al. (2007), Hasler et al. (2007), Hynna et

al. (2007), Binczak et al. (2006), Sorensen et al. (2004), Vogelstein et al. (2004), Le Masson et al. (2002), Jung et al. (2001), Liu et al. (2001) and Mahowald et al. (1991). Figure 3.10 illustrates implementation technologies (analog, digital and software) used to compute neural dynamics (taken from Renaud et al., 2007). It is seen that a few of these implementations are also supported with a firmware and/or software platform.

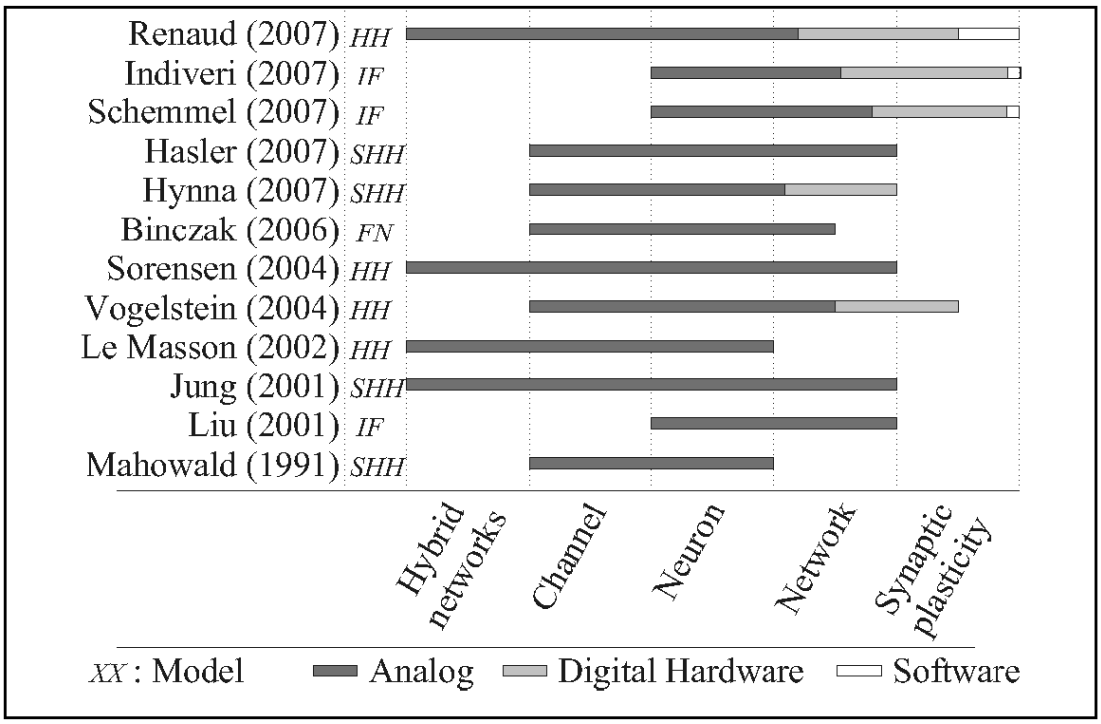


Figure 3.10 Computation distribution in various spiking neural network analog-based systems; Where HH- Hodgkin-Huxley model; IF- Integrate-and-Fire model; FN- FitzHugh-Nagumo model; SHH-HH inspired models where some of the conductance functions are simplified or fitted; taken from Renaud et.al., 2007.

3.3.1.1 Neural chip by Schemmel et al. (2008, 2010)

Under the FACETS (Fast Analog Computation with Emergent Transient States) project funded by European Commission, Schemmel et al. (2004, 2008 & 2010) have proposed configurable wafer-scale hardware system to emulate cortical networks in silicon. This is the largest mixed analog/digital integrated circuit network that has been fabricated in silicon. However, it has a lesser degree of faithfulness to biology when comparing with other VLSI neuromorphic multi-compartment neural network implementations (as shown in Renaud et al. 2007). This section summarise the network implementation given in Schemmel et al. (2004, 2008 & 2010).

Neuron and synapse Model complexity

The network uses an integrate-and-fire (I&F) neuron model that exhibits an exponential spike mechanism with adaptation, and current-injecting plastic synapses. On-chip analogue circuits are used to compute short-term synaptic depression and facilitation and to carry out the spike time dependent plasticity (STDP) measurements in each synapse. However, the weight update for STDP is performed on-chip digitally. The neurons operate on a typical time scale which is four to five orders[13] of magnitude faster in comparison to biological real time.

Network Size (No. of Neurons ≈ 450x512; No. of synapses ≈131 072x450)

The silicon wafers of approximately 450 chips are proposed. These are not cut apart into separate chips but left as a whole (wafer-scale integration). The basic chip elements of the hardware architecture are 10 mm by 5 mm network chips, each implementing 131 072 synapses which can be dynamically partitioned to up to 512 neurons. Such a wafers scale integration system has been proposed, but not fabricated. So far only a test containing 256 neurons has been demonstrated on a single chip.

Connectivity

The high bandwidth requirement for the neuronal connectivity is approached by wafer-scale integration. Additional metal layers, deposited onto the wafer in a post-processing step, allows to interface and inter-link the network chips with adequate connection density and thus to operate large-scale networks consisting of 10 000s neurons. The system used digital spike routing mechanism and the communication protocols are specially developed for this hardware architecture.

The quality of routing of the system reduces with the increase in homogeneous connectivity. Availability of large cortical network simulation results performed in this system is yet not clear. The average power consumption is expected to stay below 1 kW on a 20 cm wafer in a standard UMC 180 nm CMOS technology.

[13] http://facets.kip.uni-heidelberg.de/public/results/2ndYear/WP7/index.html

3.3.1.2 Neural chip (F-LANN) by Giulioni et al. (2008)

Giulioni et al. (2008) have implemented a 68.9 mm^2 chip in standard CMOS AMS 0.35 μm technology. This is one of the largest biological-time mixed-signal VLSI cortical neural network that has been fabricated in silicon that comprises of STDP synapses. It is also having a lesser degree of faithfulness to biology when comparing with other VLSI neuromorphic multi-compartment neural network implementations (as shown in Renaud et al., 2007) as it uses simple homogenous integrate and fire neurons. Following paragraph summaries the details of the network implementation given in Giulioni et al. (2008).

Neuron and synapse Model complexity

The neuron and synapse models used in this chip include the Integrate and Fire (I&F) neuron with spike-frequency adaptation and the bi-stable stochastic synapse with a STDP rule respectively. The synapse circuit model also has the "stop-learning" capability, which prevents the synaptic modification once the desired output of the network is reached.

Network Size (No. of Neurons ≈ 128; No. of synapses ≈ 16 384)

This reconfigurable network has 128 I&F neurons and 16 384 (128×128) bi-stable, STDP synapses. These synapses can be initialized and reconfigured. The system can read the synaptic state, at the hardware level, without disrupting the internal network activity.

Connectivity

The fully configurable synaptic matrix supports internal connectivity, external AER (Address Event Representation) connectivity, or combination of both. Each synapse may be set individually to an excitatory or inhibitory type, and synapse's initial weight can be set externally. Each neuron is connected to 128 synapses, and each synapse can accept input spikes from either internal or external neurons. Input spikes from external neurons are accepted in the form of AER events, which are addressed to the correct synapses using a decoder. Fixed weight inhibitory synapse circuits are used while excitatory synapses are plastic.

The neurons operate on a biological temporal scale. The size of the network can be increased by one order of magnitude by using multi chip network using AER infrastructure. Specifically, AER based PCI-AER board (Chicca et al., 2007; Dante et al., 2005) allows four chips to be connected together (e.g. to implement a recurrent network of 512 neurons with a uniform 25% connectivity) as given in Giulioni et al. (2008).

3.3.1.3 Neural Chip by Indiveri et al. (2007)

The chip is fabricated in a standard 0.35 μm CMOS technology and occupies an area of 6.1 mm^2. The basic elementary circuits of the network are designed to operate in the sub-threshold region to minimise the power consumption. The network operates in biological-time scale. It uses a simple I&F neuron model and has lesser degree of closeness to biology when comparing with other VLSI neuromorphic network implementations (as shown in Renaud et al., 2007). The details of the network implementation given in Indiveri et al. (2007) are summarised in this section.

Neuron and synapse Model complexity

The chip uses an integrate-and-fire neuron model. Few types of synapses are used, namely: STDP plastic synapse, excitatory non- STDP and inhibitory non-STDP synapse.

Network Size (No. of Neurons ≈ 16; No. of synapses ≈ 2 048)

Each chip comprises an array of 16 integrate and- fire (I&F) neurons and 2048 synapse circuits where each neuron is connected to 128 synapses. These 128 synapses include 120 synapse with STDP dynamics, 4 excitatory non-STDP and 4 inhibitory non-STDP synapses.

Connectivity

Since this is a simple small network it is easy to accommodate connectivity. The input spike patterns are provided to the synapses through the asynchronous AER interfacing circuits (Boahen, 2000).

The chip is fabricated in a standard 0.35 μm CMOS technology and occupies an area of 6.1 mm^2. However, the size of the network can be increased by about one order of magnitude by using multi chip network infrastructure such as in Fasnacht et al. (2008).

3.3.1.4 Other spiking neural network implementations in mixed-signal system

Other implementations presented in literature include: Neural Chip by Indiveri et al. (2004) that occupies silicon area of 16.8 mm^2, with 21 IF neurons, 129 synapses, including 56 STDP plastic synapses fabricated in standard AMS CMOS 0.6 μm technology (this is the older version of the chip presented in Indiveri et al. 2007); The "*Neurogrid"* neural chip implemented by Arthur and Boahen (2006) in 10 mm^2 of silicon area for a total of 9 216 neurons is fabricated in TSMC 0.25 μm standard CMOS technology. This implementation is further used in the neural Chip by Merolla and Boahen et al. (2007) with 8 192 neurons in each chip and expanded to 32 768 neurons in a network with four of the neural chips on a multi-chip board. They have demonstrated neuronal selectivity along position, spatial frequency and orientation properties of cortical network. However, their implementations do not include STDP learning synapses. Other sensory processing neuromorphic device implementations include "Silicon Retina" by Zaghloul et al. (2004 & 2006), and "Silicon Cochlea" by Liu et al. (2010).

3.3.2 Outline of spiking neural network implementations in digital system

Fully digital systems use conventional digital units such as RAMs, processors, digital logic, etc. to implement a neural architecture. These are implemented in custom VLSI chips and/or in off-the-shelf FPGA (Field Programmable Gate Array) devices.

Among these, the "SpiNNaker" project[14] has developed a chip that comprises 20 processing cores, each with ARM9 processor, local memory and DMA capability. According to the estimates, each ARM9 processor can model 1 000 Leaky I&F neurons, each with 1 000 inputs firing on average at 10 Hz, in biological real time. The synaptic data is held in an off-chip RAM (Furber et al., 2006). The processing cores are connected to its local peers via a Network-on-Chip (NoC). This provides inter-chip communication via links between SpiNNaker chips while utilising in-chip local high bandwidth communication. Using this approach the architecture can be extended to have thousands or millions of similar processing cores to build a massive cortical network in silicon.

[14] Web Link: http://intranet.cs.man.ac.uk/apt/projects/SpiNNaker/

The Memory Optimized Accelerator for Spiking Neural Networks (MASPINN) project (Schoenauer et al., 1998 and 2000) produced a neuro-accelerator board simulating one million neurons in real time. However, it does not include STDP learning, different types of neurons and synapses, which are a key noticeable fact in biological networks.

Other digital network implementations include the proposed Connectionist Network Supercomputer (CNS-1) architecture (Asanovic et al., 1993), RAPTOR2000 system (Porrmann et al., 2004), and systems by Agris at al. (2007) and Carrillo et al. (2008). All of these digital implementations of cortical network provide semi-customisation of homogeneous digital elements to mimic low-level heterogeneous neural elements. Hence this approach may not provide fully optimised mimicking of the nervous system. However, digital implementations use the well-optimised digital building blocks to design the system and in terms of programming for the different neural models, they are more flexible than the analogue mixed-signal implementations.

CHAPTER 4 : CORTICAL NEURON CIRCUITS

This chapter presents two cortical neuron circuit implementations that work in different time scales, the accelerated-time and the biological-time. These neuron circuits are capable of generating many types of the neuron behaviour, with diversity similar to that of biological neuron cells. These neuron circuits are inspired by the computational model proposed by Izhikevich (2003) and motivated by the desire to achieve, a single compact generic circuit that can easily be tuneable to a known cortical neuron type. The initial section provides the mathematical neuron model proposed by Izhikevich (2003).

All three fabricated ICs presented in this thesis use the Accelerate-Time Neuron circuit. The circuit design of the accelerate-time neuron was proposed in the Mphil thesis (Wijekoon, 2007), and for the purpose of completeness of this thesis, the accelerated-time neuron section of this chapter, present the neuron circuit, its operation, simulation results, and the mathematical model briefly. By extending the research done in the Mphil degree, in this PhD thesis the function of the neuron circuit is experimentally verified, and the results obtained from the fabricated neuron are given in Chapter 7 and Chapter 8.

The Biological-Time Neuron circuit is proposed at the end of this chapter and the circuit design, operation and the simulation results are provided. This Biological-Time Neuron circuit is implemented in a standard CMOS 0.35 µm technology, and the proposed circuit and the simulation results presented in this chapter were published in the Proceedings of the IEEE Biomedical Circuits and Systems Conference (Wijekoon et al., 2009).

Finally, summary of these neuron circuits and the merit and demerits of designing VLSI neural network in accelerated- vs. biological- time scales are discussed.

4.1 The Izhikevich Model of the Cortical Neuron

The Izhikevich (2003) neuron model is a simplified version of the Hodgkin-Huxley neuron model and has two state variables membrane potential (V) and membrane recovery (U). According to the model membrane potential, V, evolves as in the equation set given below:

$$\dot{V} = 0.04V^2 + 5V + 140 - U + I \qquad (4.1)$$

$$\dot{U} = a(bV - U) \qquad (4.2)$$

with the after spike resetting

$$\text{if } V \geq 30\,\text{mV}, \text{then} \begin{cases} V \leftarrow c \\ U \leftarrow U + d. \end{cases} \qquad (4.3)$$

where, *a*, *b*, *c* and *d* are dimensionless parameters.

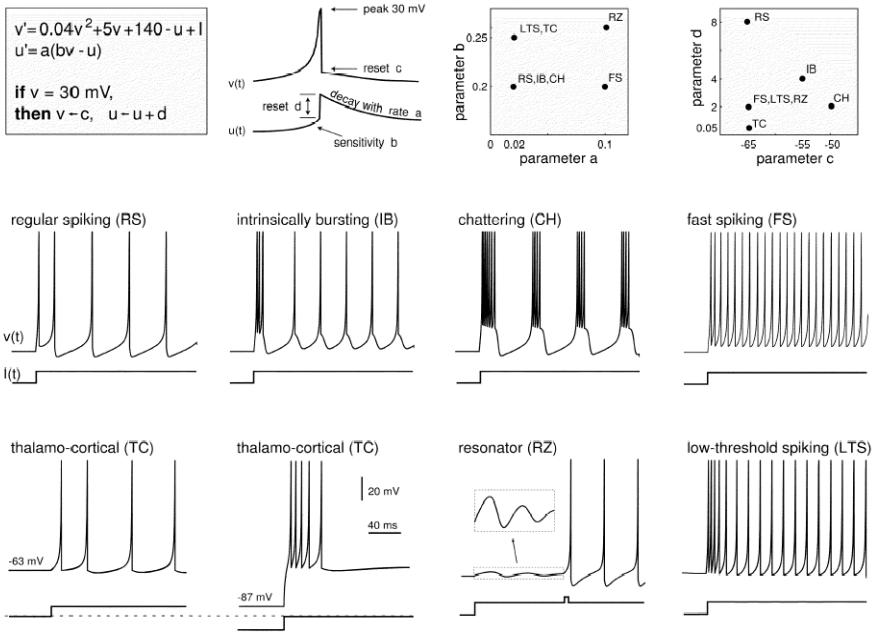


Figure 4.1. Types of neurons reproduced using the Izhikevich (2003) neuron model and their correspondent parameter values; taken from Izhikevich (2003).

Using the aforementioned simple set of formulas and resetting function various types of cortical neuron firing patterns can be reproduced. The reproduced firing patterns correspond to different a *b*, *c* and *d* values of parameters published by Izhikevich (2003) are shown in Figure 4.1.

It should be noted that this reset mechanism is more similar to the reset in the I&F model than a spike generating mechanism of biological sodium/potassium channels. However, the rich repertoire of behaviours, including adaptation and bursting, is a result of the dynamics of *V* and *U*, which can be qualitatively associated with the interplay between faster sodium/potassium dynamics and slower calcium dynamics. A similar mechanism for adaptation and bursting is also presented in an exponential I&F model Brette and Gerstner (Brette et al, 2005). A linear I&F model can also be extended to enable adapting and bursting behaviours, via mechanisms such as variable thresholds in

Gerstner's spike response model (Gerstner et al., 2002), and some additional dynamic variables such as "burst currents" proposed by Michalas and Niebur (2009).

The Accelerated-Time Neuron circuit design implements the qualititative behaviour of the Izhikevich neuron model in VLSI, whereas the Biological-Time Neuron circuit combines the simplicity of the I&F model with the slow-fast variable interactions present in the Izhikevich model to obtain a large variety of spiking behaviours in a simple circuit.

4.2 Accelerated-Time Neuron

The Accelerated-Time Neuron circuit, its operation and the approximated mathematical model is given in this section. Its experimental results can be found in Chapter 7. More elaborated details of this neuron circuit, including circuit operation, phase-plane analysis, derivation of mathematical model, and simulation results can be found in Mphil thesis (Wijekoon, 2007). Some of the materials presented here were published in Wijekoon et al. (2008). The circuit diagram of the neuron is shown in Figure 4.2.

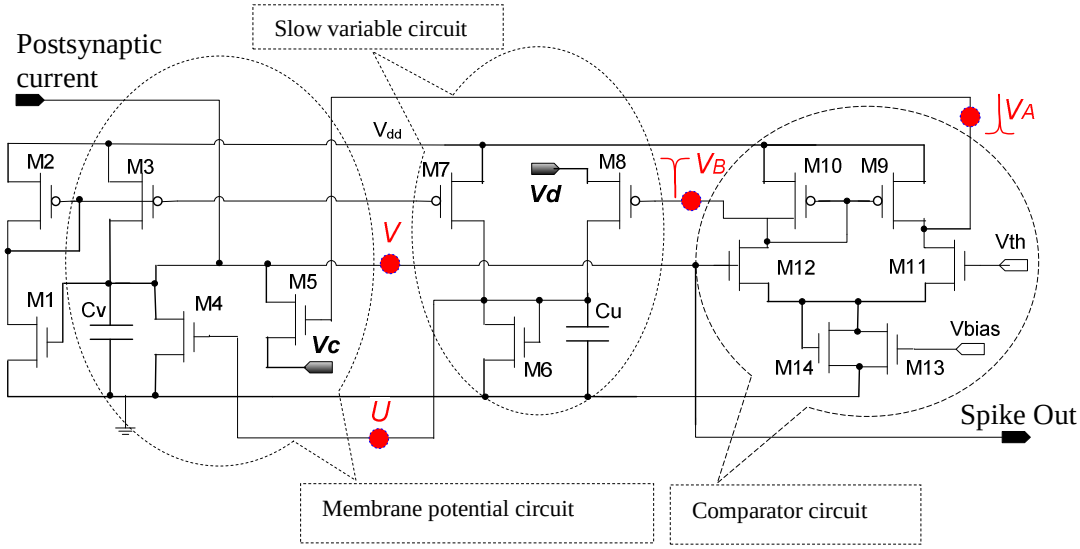


Figure 4.2. The compact silicon cortical neuron circuit.

4.2.1 The circuit operation

The implemented neuron model consists of two state variables: "membrane potential" (*V*) and "slow variable" (*U*), that are represented by voltages across capacitors C_V and C_U respectively. The circuit comprises of three functional blocks: membrane potential

circuit, slow variable circuit and comparator circuit (the transistors M1 and M2 are shared by the membrane potential and slow variable circuits). In the membrane potential circuit, the capacitor C_V integrates the postsynaptic input current, plus internal currents which depend on the state of the cell. Similarly, in the slow variable circuit the capacitor C_U integrates the currents that non-linearly depend on *U* and *V*. The comparator detects the spike and generates pulses (V_A and V_B) that perform the after-spike reset. Various spiking and bursting firing patterns are obtained by tuning two voltage parameters, *Vc* and *Vd*, which control the reset mechanism. Figure 4.3 shows example waveform of voltages, *V*, *U*, V_A and V_B.

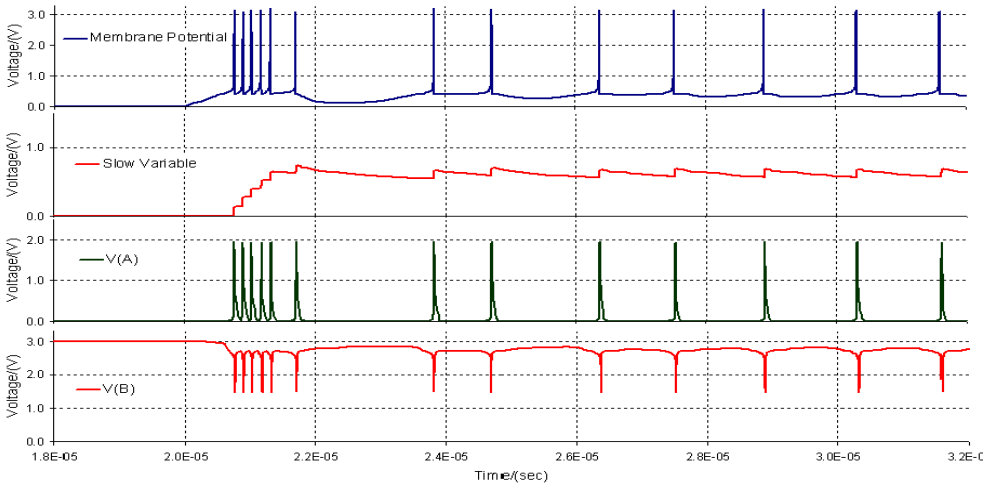


Figure 4.3 Example waveforms of the membrane potential (*V*), slow variable (*U*) and the reset pulses (*VA* and *VB*).

4.2.2 Simulation results

A Summary of firing patterns obtained by simulation the circuit using standard 0.35 µm CMOS technology libraries is shown in the Figure 4.4. Experimental results obtain from the fabricated neuron are given in Chapter 8.

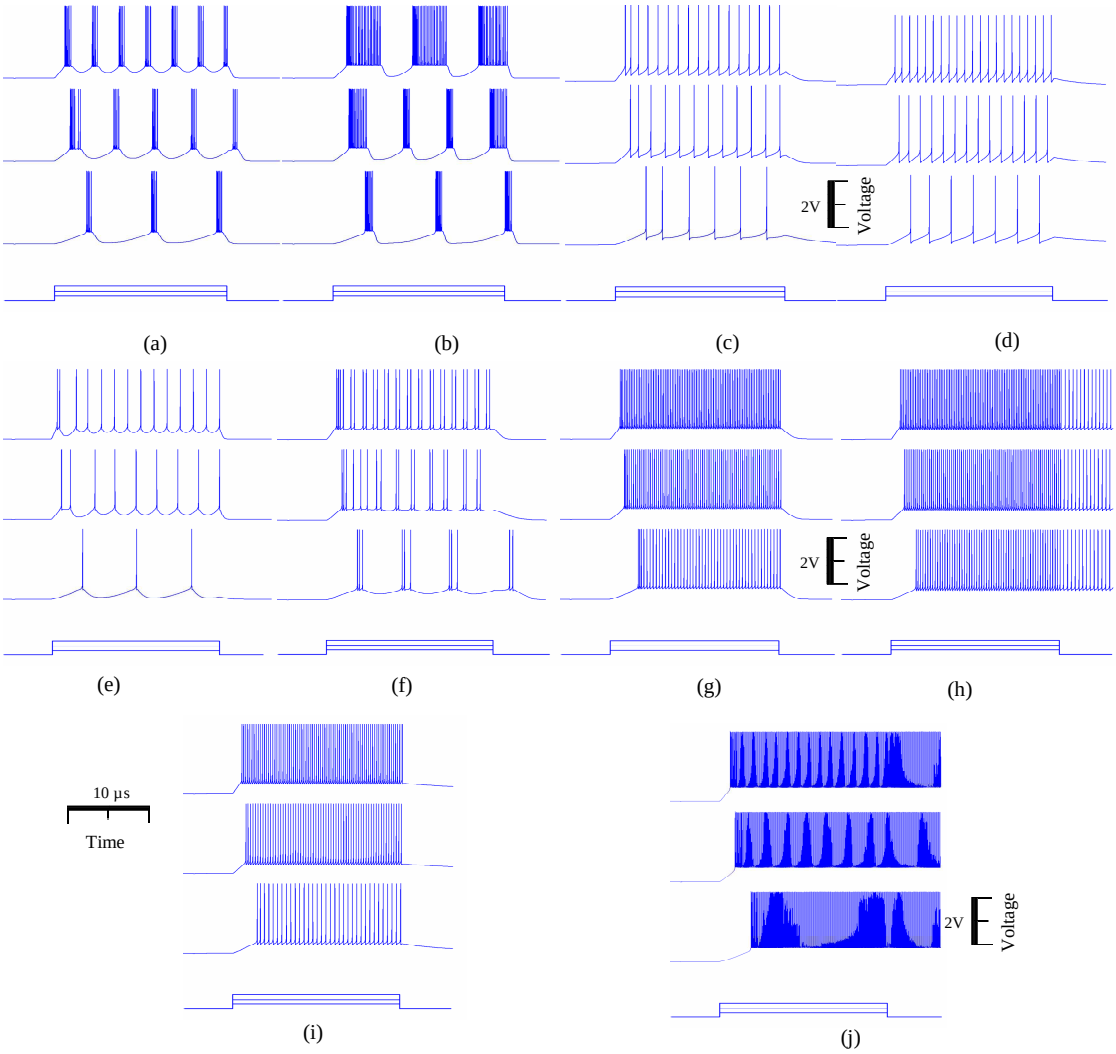


Figure 4.4 Spiking and bursting firing pattern behaviour to a increase in step post synaptic current (a) CH, (b) CH, (c) RS2, (d) RS1, (e) IB, (f) IB, (g) LTS, (h) FS, (i) FS and (j) FS for different *Vc* and *Vd* parameters. The plots shows responses to three increasing steps of dc-currents: 0.05µA, 0.1 µA, and 0.15 µA except plot (b) is 0.05 µA, 0.1 µA, and 0.12 µA and plot (e) is 0.05 µA, 0.1 µA, and 0.25 µA.

4.2.3 Mathematical model of the neuron circuit

The mathematical model of the neuron circuit is used in the Cortical Neural Layer (CNL) chip model discussed in Chapter 8. Hence, this section summarised the approximated mathematical equations of the neuron. According to the model membrane potential of the neuron, *V*, evolves as in the equation set given below. Each equation is corresponding to the circuit blocks discussed in the above section; i.e Equation 4.4,

Equation 4.5 and Equation 4.6 approximate the membrane potential circuit, slow variable circuit and comparator circuit dynamics respectively.

$$\dot{V}=\begin{cases}\dfrac{k}{C_v}\left\{\alpha\left[\dfrac{1}{2}\left(\dfrac{W}{L}\right)_{M1}(V-V_t)^2\right]-\beta\left[\dfrac{1}{2}\left(\dfrac{W}{L}\right)_{M4}(U-V_t)^2\right]+\dfrac{I}{k}\right\} & \text{when} \quad V\geq U-V_t \\ \dfrac{k}{C_v}\left\{\left(\dfrac{W}{L}\right)_{M4}\left((U-V_t)V-\dfrac{1}{2}V^2\right)+\dfrac{I}{k}\right\} & \text{otherwise (i.e.region 'A')}\end{cases} \qquad (4.4)$$

$$\dot{U}=\frac{k}{C_v}\left\{\alpha\left[\frac{1}{2}\left(\frac{W}{L}\right)_{M1}\left(\frac{L}{W}\right)_{M2}\left(\frac{W}{L}\right)_{M7}(V-V_t)^2\right]-\gamma\left[\frac{1}{2}\left(\frac{W}{L}\right)_{M6}(U-V_t)^2\right]\right\} \qquad (4.5)$$

$$\text{If} \quad V>V_{th} \quad \text{then} \begin{cases}V\leftarrow V_c \\ U\leftarrow U+D\end{cases} \qquad (4.6)$$

In the above equations, $V_{t,}$ is the nMOSFET threshold voltage. The value k is $\mu\times Cox$ of the nMOSFETs (μ - charge-carrier effective mobility, Cox -gate oxide capacitance per unit area) and C_V and C_U are membrane and slow variable capacitance values respectively. The $(W/L)_{Mx}$ is the gate width to length ratio of the MOSFET M_x. I is the postsynaptic current and V_C and D are externally tunable parameters. The α, β, γ, and the region 'A' depend on V_t , V and U as given in Figure 4.5,

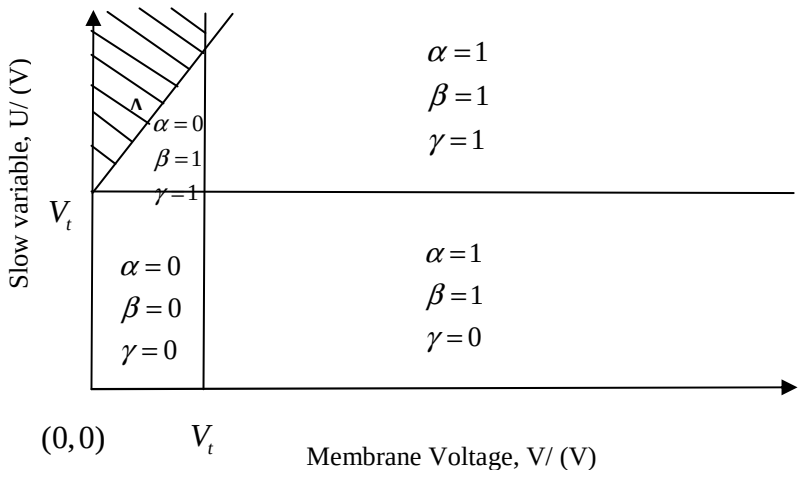


Figure 4.5 Parameter values for formulas 4.4-4.6.

4.3 Biological-Time Neuron

Similar to Accelerated-Time Neuron, the Biological-Time Neuron circuit is capable of generating many types of the cortical neuron behaviour, with diversity similar to that of biological neuron cells. The four tuning parameters *d*, *c*, *Uth* and *Vbisn* are used to configure the circuit to operate in a known type of neural behaviours, RS, FS, LTS, CH, IB and TC (these patterns are briefly defined in Chapter 2). Here, when considering the firing patterns the supra-threshold spike activities are considered and sub-threshold neuronal activities are not considered. The circuit is presented in Figure 4.6.

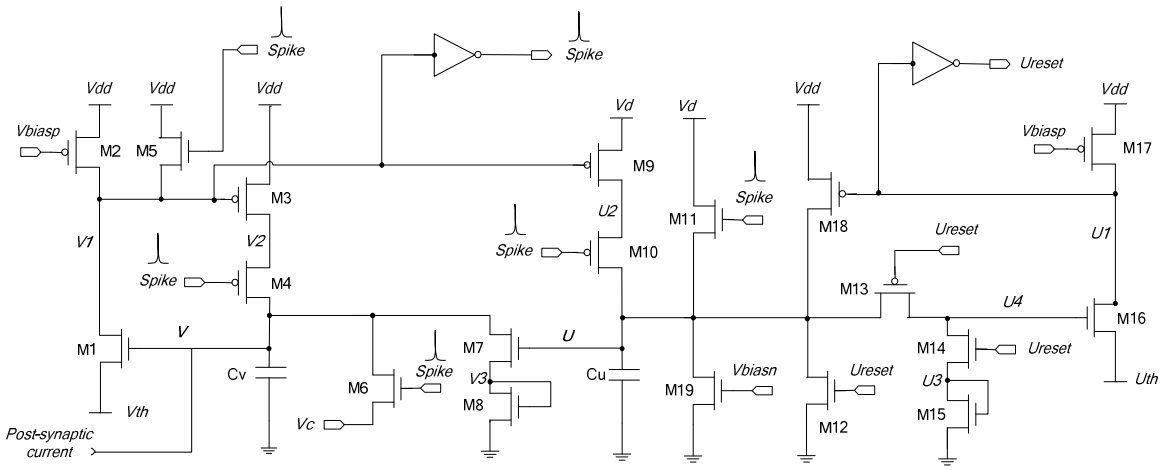


Figure 4.6 Schematic of the Biological-Time Neuron circuit

4.3.1 The circuit operation

The node voltages at *V* and *U* represent the state variables, membrane potential and the slow variable. The currents feeding into these nodes are integrated on capacitors C_V and C_U respectively. The currents are provided by two functional circuit blocks: membrane potential circuit (transistors M1 to M8) and slow variable circuit (transistors M9 to M19). The evolution of the membrane potential *V* is due to integration, on the capacitor C_V, of the post-synaptic current (which is assumed to be injected into that node through the synapse circuit) plus an exponential leakage current (generated via M7 and M8) which is determined by the value of the slow variable *U*. The spike threshold of *V* is determined by voltage *Vth*, and detected by M1, M2, and an inverter. M3-M5 help to control the duration of the spike and the reset dynamics. Initially M3 provides a positive feedback to quickly exceed the membrane potential threshold. As the spike is generated,

the feedback current is turned off by opening M4, so that the voltage at node *V* does not actually produce a significant voltage spike. Transistor M5 limits the output spike pulse duration, while membrane potential *V* is reset to the value of *Vc* via M6.

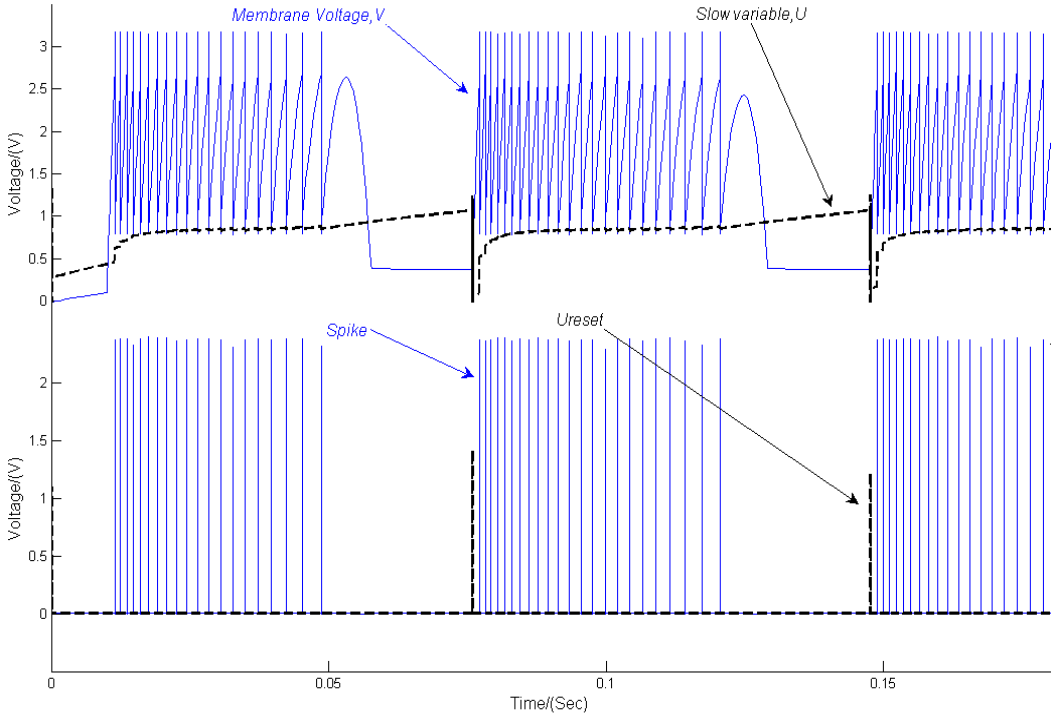


Figure 4.7 Waveforms of a typical CH firing pattern obtained from the circuit shown in Figure 4.6; top: membrane potential, *V*, and slow variable, *U*; bottom: output spikes and slow variable reset signal

As can be seen in Figure 4.7, the slow variable voltage *U* evolves at a much slower rate than the membrane voltage. The changes to the slow variable are primarily due to two reset mechanisms. After the spike is generated the value of *U* is pulled some amount towards V_d via M11. Additionally, when *U* reaches a threshold value, determined by the tuneable voltage U_{th}, a *Ureset* signal is generated (M16, M17 and the inverter) and *U* is reset to ground via M12. The arrangement of M13-M15 and M18 helps to control the timings in the reset circuit. Initially, M18 provides a positive feedback to quickly bring *U* above the threshold value. As a result the reset signal *Ureset* goes high. Since *U* evolves with very slow rates, the *Ureset* is generated using a slightly different topology than the spike generation circuit. By breaking the loop using M13 switch, the *Ureset* pulse can be generated, minimising the risk of settling on to a fixed DC value in the feedback loop. The *Ureset* signal is then used to reset *U* to zero via M12. Finally the

Ureset pulse is brought back to zero after the voltage at node *U4* is brought down through M14-M15. In addition to the two reset mechanisms, *U* continuously evolves as two currents are integrated on capacitor C_U, one is the current through M9, which depends on the membrane potential *V* (M10 is used to prevent large current flow during the spike), the second one is a leakage current, controlled by *Vbiasn* (M19).

4.3.2 Simulation results

The operation of the circuit and the circuit simulation results obtained using standard 0.35µm AMS CMOS technology, are presented in this section. The SPICE simulation results shown in Figure 4.8 illustrate membrane potential *V* during various types of cortical neuron firing patterns (CH, RS, IB, FS, LTS and TC). The spike pattern classification follows methods given in Nowak et al. (2003). The output spikes are produced at the times of membrane potential peaks. The four tuning voltage parameter values corresponding to the firing patterns are provided in Table 4.1. Figure 4.9 shows the trajectories in the state space corresponding to these firing patters.

As can be seen in Figure 4.8, the firing patterns obtained from the proposed circuit are in the same time scale as that of the biological neurons, the minimum refractory period is approximately 1 ms. The frequency of firing for a given step of post-synaptic stimulus typically ranges from below 1 Hz to 1 kHz and can be approximately configured to a desired frequency, simply tuning the neuron using an appropriate parameter set. All the waveforms in Figure 4.8 are obtained using a post-synaptic stimulus of 2 nA. It is observed that RS type neuron's inter-spike frequency can be configured to one typical of real RS inter-spike frequencies (Nowak et al., 2003); RS1 and RS2 sample waveforms are shown with 25 Hz and 100 Hz inter-spike frequencies. Similarly, the FS type neuron's inter-spike frequency can be configured to a frequency in the typical FS frequency range (Nowak et al., 2003) and two selected samples (FS1 and FS2) with different inter-spike frequencies are shown in Figure 4.8. The proposed neuron circuit can also be configured to obtain accommodating (spike frequency adaptation) or to a non-accommodation firing pattern. In CH type firing pattern, the inter-burst interval as well as number of spikes per burst can be configured easily as seen in CH1, CH2, and CH3 waveforms in Figure 4.8.

The layout of the circuit is shown in Figure 4.10. It consumes 70 µm x 70 µm of silicon area in a 0.35 µm CMOS technology. Here, Cv and Cu poly capacitors occupy a large

area of the layout. If non-linear gate oxide capacitance of the MOSFET is used as a capacitor, the silicon area can be reduced further.

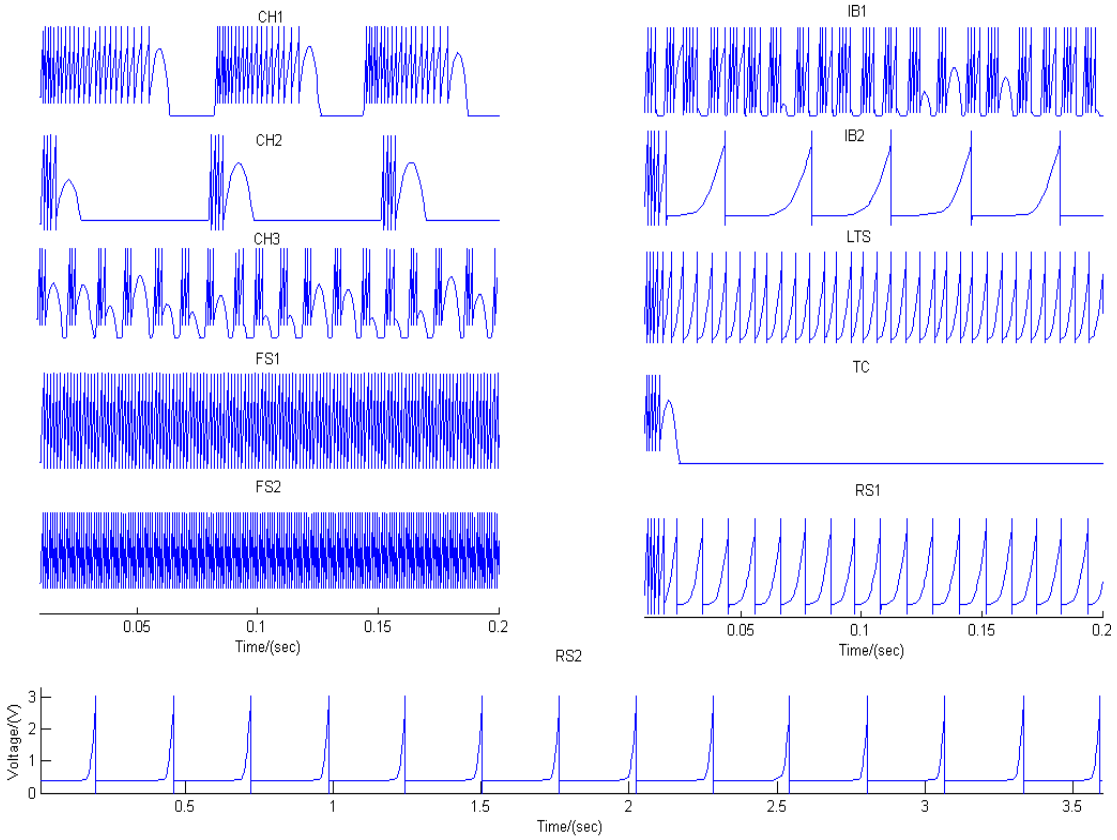


Figure 4.8 Membrane potential of the neural firing behaviours obtained from the neuron circuit in response to a step post-synaptic stimulus of 2 nA.

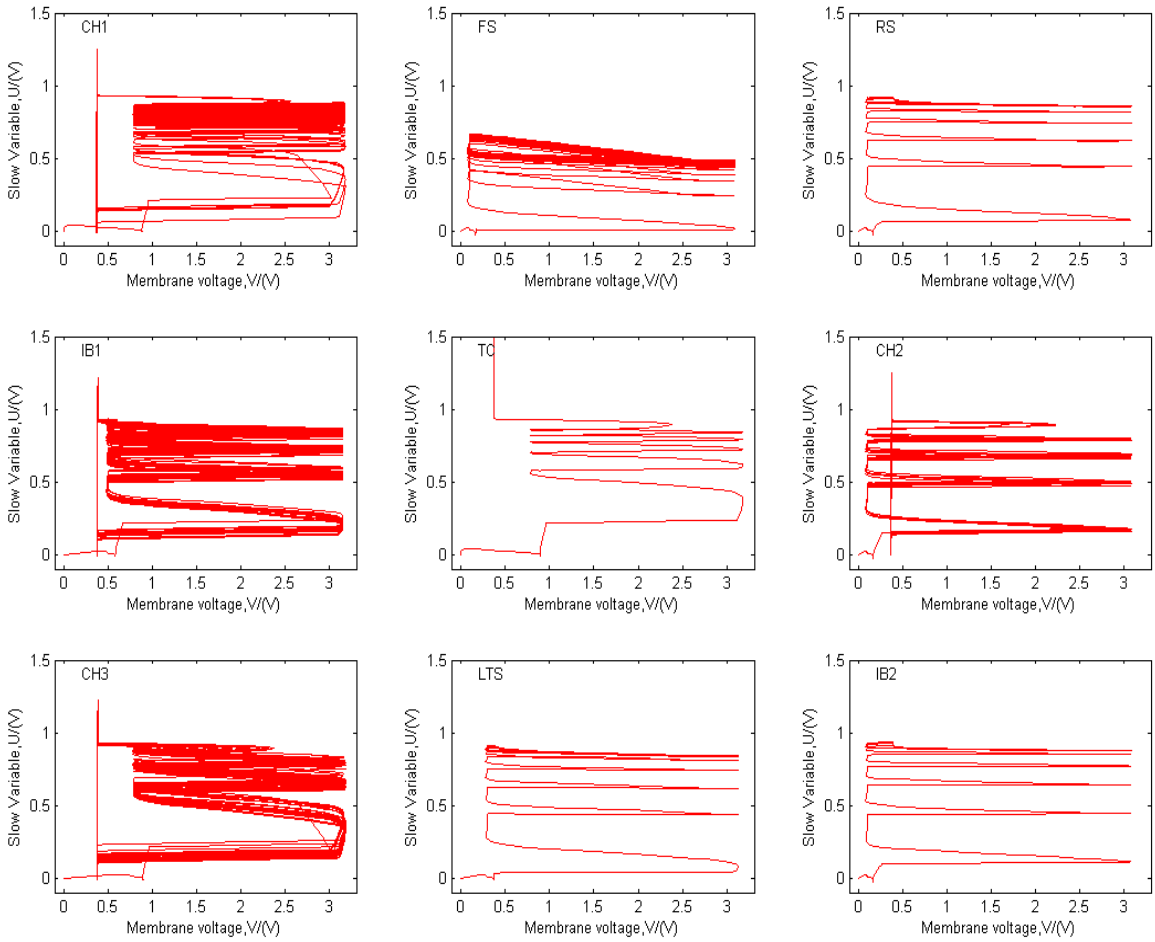


Figure 4.9 State trajectories of CH, IB, FS, TC, LTS & RS cells when a 2 nA of postsynaptic current step is injected. (The plots are drawn using data obtained from SPICE simulations)

Table 4.1: Tuning voltages used to obtain the firing patterns shown in Figure 4.8

Neuron Type	Tuning Parameter/(V)			
	Uth	Vbiasn	Vc	Vd
CH1	0.2	0.1	0.8	0.8
CH2	0.2	0.2	0.1	1.7
CH3	0.1	0.15	0.8	1.7
FS1	0.5	0.3	0.1	1.7
FS2	1.3	0.6	0.8	1.7
IB1	0.1	0.1	0.5	1.7
IB2	0.5	0.22	0.1	1.7
LTS	0.5	0.24	0.3	1.7
TC	2	0	0.8	0.9
RS1	0.5	0.22	0.1	1.7
RS2	1.3	0.22	0	2.5

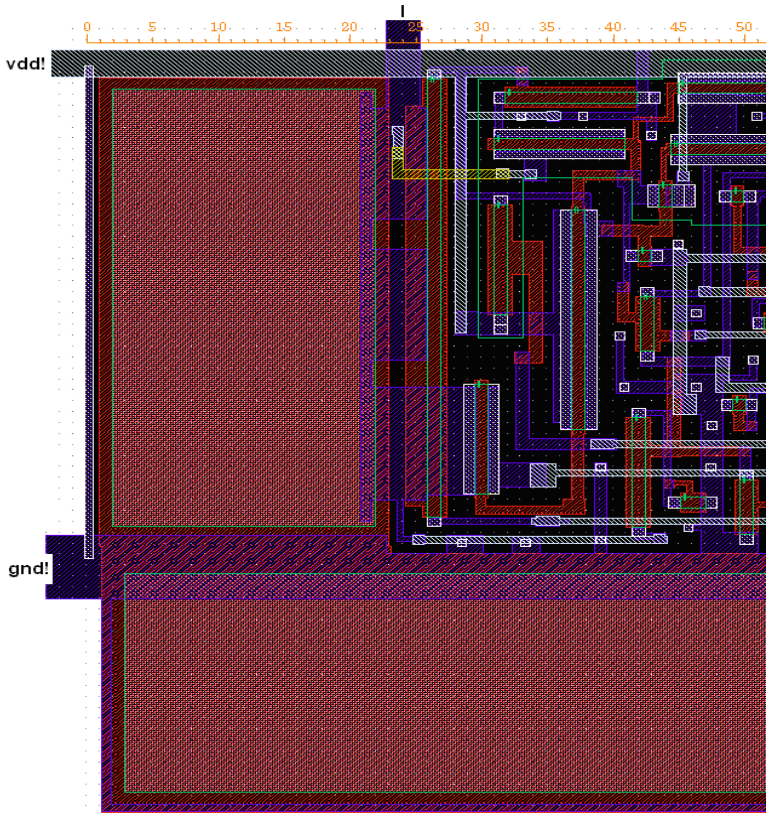


Figure 4.10 A layout of the proposed VLSI neuron circuit in a 0.35 µm CMOS technology.

4.4 Discussion

Both CMOS cortical neuron circuits replicate many known types of spiking neural behaviours by adjusting a few external voltages. These circuits provide a much richer repertoire of spiking patterns than a simple integrate and fire model, while using only one additional state variable. The circuits provide simple, compact and easily configurable universal cortical neurons, with potential applications in the development of massively parallel analogue VLSI neuromorphic chips that closely resemble the circuits of the neocortex. In addition, the Biological-Time Neuron can be used in the context of interfacing electronic neural circuits with biological systems.

As seen in the literature (in Chapter 3) both accelerated and biological-time implementations are popular in neuromorphic circuits and both approaches have their own merits and de-merits. The summary of the merits ("+") and de-merits ("-") of implementing both accelerated and biological-time VLSI devices are listed below.

4.4.1 Merits (+) and de-merits (-) of implementing biological-time devices:

+ Require low communication bandwidth: Communication bandwidth between VLSI chips and within VLSI chip circuits is very low. This enables large number of neuron connections. Well established Address-Event Representation (AER) protocol can be used.
+ Circuits can be easily interface with biological systems, i.e. the silicon neurons can be interconnected with the biological cells to form “hybrid networks” (LeMasson et al., 2002).

- Experiment takes longer time duration to perform compared with accelerated-time implementations.
- Decay and rising timings of signals are very slow; hence advantages of standard CMOS technology, which is optimised for speed signals, are not fully exploited.
- Circuits require operation in weak inversion region of transistors and large capacitors are required to store analogue voltage values. This consumes large-silicon area.
- Weak inversion region of operation can cause large mismatch effects on the characteristics of a circuit which resulted in more variations on the circuit’s characteristics.
- Scaling of the circuit to fabricate in advanced deep-submicron technology is difficult for the circuit with transistors operating in sub-threshold region.

4.4.2 Merits (+) and de-merits (-) of implementing accelerated-time devices:

+ Real time long-duration simulations can be performed within very small time duration (example: 5 year real time simulation of a cortical network can be observed in 48 hours if time scale is 10^3 times faster) or extensive parameter searches of an experiment are possible.
+ Technological advantage of speed optimised CMOS technologies can be exploited.
+ Less power consumption per experiment (very slow biological time experiment take long time to perform an experiment, hence leakages currents and refreshing voltage states may consume more power)

- As accelerated time processing is used, communication bandwidth limits the size of the largest connectivity matrix of a network. The conventional AER protocol cannot be used. However, using mixed mode circuits, significant increase of neuron connectivity can be obtained as demonstrated by Schemmel et al. (2008), using circuits that operate four orders of magnitude faster than biological-time.

In the rest of this thesis, the accelerated time neuronal implementations are considered in designing a large cortical network, which provides higher computational throughput of the neuro-mimetic computing device. Hence, extensive analysis of the biological-time neuron is not considered in this thesis. In implementing a neural network chip, the rest of the neural circuits proposed in this thesis operate three orders of magnitude faster than the biological-time (rather than four orders as considered here). This will ease the implementation of communication of a network that has a larger inter neuron connectivity matrix. Hence, the parameter set of the accelerated-time neuron is selected such that the neural dynamics are three orders of magnitude faster than biological-time.

CHAPTER 5 : LONG-TERM DYNAMIC SYNAPSE CIRCUITS

This chapter proposes STDP (Spike-Time Dependent Plasticity) Synapse and Dopamine (DA) Modulated STDP Synapse circuits that operate in accelerated-time. The STDP plasticity rule is a crucial feature of a cortical network and is believed to be the neuronal mechanism for the learning and memory of a network, whereas the DA modulated STDP plasticity rule is believed to be the mechanism for the reinforcement learning in a cortical network. The plasticity rules, the STDP and the DA modulated STDP are explained in Section 5.1, and the computational model used to implement the DA Modulated Synapse circuit is given in Section 5.2. The Section 5.3 and 5.4 provide the circuit operation and simulation results of the STDP Synapse circuit, and the DA Modulated Synapse circuit respectively. To prove the concept in hardware, Integrated Circuits (ICs) have been fabricated in a standard 0.35 µm CMOS technology. The fabricated Integrated Circuit that contains both STDP Synapse circuits and DA Modulated Synapse circuits is called the STDP/DA-STDP Synapses Neuron (*STDP/DA Neuron*) chip presented in Chapter 8. The synapses of this chip can be configured to work as a DA modulated synapse or as a STDP synapse without the DA modulation. The STDP synapses are included in the Cortical Neural Layer (CNL) chip and its details are presented in Chapter 9. Linearly approximated discrete mathematical models of the STDP synapse circuits are presented in Section 5.3. The model could be used to simulate the approximate behaviour of the CNL chip in software. Some of the DA Modulated Synapse circuit materials presented in this chapter have been accepted for publication in the proceedings of IEEE International Symposium on Circuits and Systems (Wijekoon et al., 2011).

5.1 STDP and DA modulated STDP

In some synapses, synaptic weight is changed by the timing difference between pre- and post- synaptic activity. This plays an important role in synaptic plasticity, which is believed to be the mechanism for learning and memory in a biological system. In general, if the pre-synaptic spike proceeds post-synaptic spike of the synapse the synaptic weight is increased (Long-Term Potentiation, LTP), whereas if the pre-synaptic spike follows the post-synaptic spike the synaptic weight is decreased (Long-Term

Depresion, LTD). The magnitude change in synaptic weight depends on the temporal difference between pre- and post- synaptic firings. The curve that provides the amount of synaptic weight modifications with respect to temporal differences of the pre/post spikes is called the STDP curve. As mentioned in Chapter 2 a variety of STDP curves have been observed experimentally (Abbot et al., 2000). The synapses that obey this STDP weight modification (STDP synapse) are mostly excitatory synapses. The Silicon area consumed by a synapse is a crucial factor as it could limit the size of the network that can be implemented in the VLSI hardware; hence the STDP curve that can be implemented with a small number of transistors is used.

Some special type of synapses are believed to exist in the neocortex, in which the synaptic modification due to STDP is modulated by the level of extracellular dopamine concentration. That is, the extracellular DA level regulates the LTP and LTD modification on the synaptic weight (Fellous et al., 2003; Izhikevich, 2007). These synapses are called dopamine modulated synapse and a computational model it is proposed by Izhikevich (2007) is given in next section.

5.2 Computational Model of DA Modulated Synapse

The DA modulated circuit is broadly based on the model presented in (Izhikevich, 2007), describing dopamine modulated synapse, where the LTP and LTD components of the spike-timing-dependent plasticity (STDP) are modulated by DA present during the critical window of a few seconds after the post synaptic spike. According to the model, the strength of the synapse, *s*, evolves as per the following three equations (Izhikevich, 2007):

$$\dot{c} = -c / \tau_c + STDP(\tau)\delta(t - t_{pre/post}) \qquad (5.1)$$

$$\dot{s} = cd \qquad (5.2)$$

$$\dot{d} = -d / \tau_d + DA(t) \qquad (5.3)$$

In the above equations c is the synaptic eligibility trace (ET); $\delta(t)$ is the Dirac-delta function that provides a step-increase or -decrease of c depending on pre- and post-synaptic neuron firing times, t_{pre} and t_{post}; the function $STDP()$ describes the spike-timing-dependent change of the ET (typically, the change has a positive value when post-synaptic spike follows a pre-synaptic spike within a small time interval, negative value when post-synaptic spike precedes the pre-synaptic one, and decays to zero for

larger pre- and post- synaptic spike time differences); *d* represents the extracellular concentration of DA, *DA(t)* is the amount of the DA released due to the activities of the dopaminergic neurons. Time constants, τ_c = 1 s and τ_d = 0.2 s. Dynamics that are described by these equations are further explained in (Izhikevich, 2007). Figure 5.1 and its caption are taken from the Izhikevich (2007) well explain the dynamics of the model.

This model addresses a solution to the distal reward/credit assignment problem using DA modulation of STDP; only nearly coincident spiking patterns occurring in the time period before the reward are reinforced by the reward, whereas uncorrelated spikes occurring before the reward, and correlations when no reward is present, are ignored by the network. The spike coincidences produce relevant changes in the slowly decaying eligibility traces, and the eligibility traces control changes in the synaptic strength, making the greatest influence when the reward signal (DA activity) is strong.

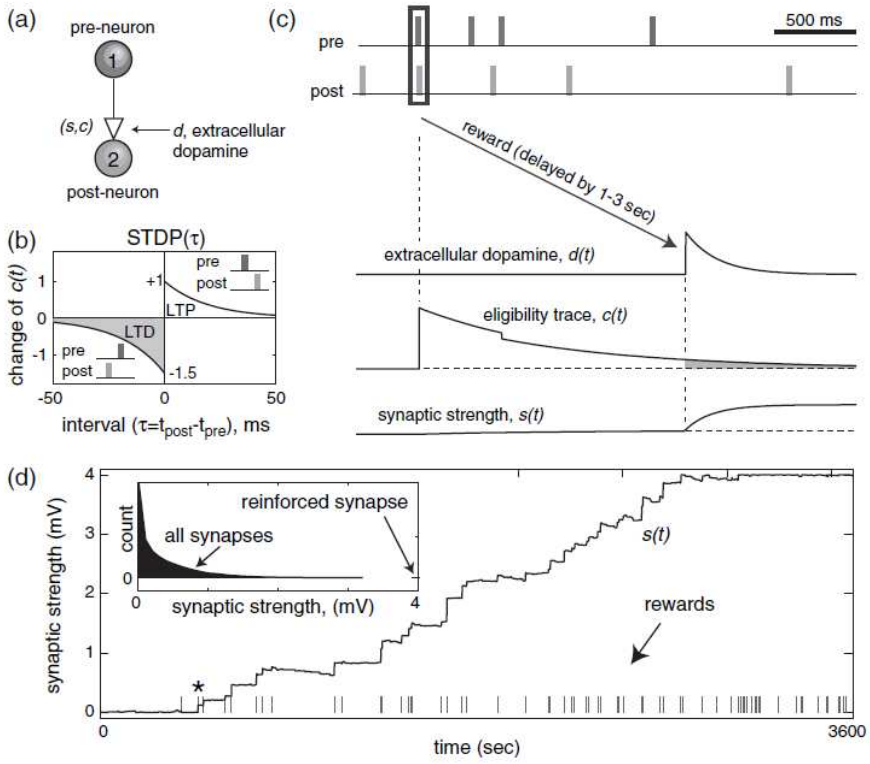


Figure 5.1 (a) The dynamics of each synapse is described by synapse strength s and eligibility trace c, which are gated by the extracellular DA d. The STDP rule that induce changes to the variable c is shown in (b). These changes result in modification of the synaptic strength, s, only when extracellular DA is present (d > 0) during the critical window of a few seconds while the eligibility trace c decays to zero. (c) The magnification of the region in (d) marked by *. To reinforce coincident firings of 2 coupled neurons, deliver a reward (step-increase of variable d) with a random delay (between 1 and 3 s) each time a postsynaptic firing occurs within 10 ms after a pre-synaptic firing (marked by a rectangle in c). This rare event increases c greater than any random firings of the same neurons during the delayed period. (d) Consistent rewarding of each such event results in the gradual increase of synaptic strength, s, which increases the probability of coincident firings and brings even more reward. The time course of a typical unreinforced synapse (not shown here) looks like a random walk near 0. The inset shows the distribution of all synaptic weights in the network. The reinforced synapse is potentiated to the maximal allowable value 4 mV (42 out of 50 experiments) whereas the other synapses are not.(figure and caption taken from Izhikevich, 2007)

5.3 STDP Synapse Circuit

The accelerated time STDP Synapse circuit comprises the STDP circuit and Synaptic Current Generator (ISYN) circuit as shown in Figure 5.2 and Figure 5.3 respectively. This section presents the circuit operation, simulation results and approximated mathematical model of the STDP Synapse circuit. The STDP synapse circuit has been fabricated in two ICs: the STDP/DA Neuron chip and the CNL chip. However, STDP circuit in STDP/DA-STDP Neuron chip uses the complementary circuit topology of the STDP circuit explained here (as explained in Section 5.4). The experimental results of a STDP Synapse circuit in the STDP/DA Neuron Chip are presented in Chapter 7.

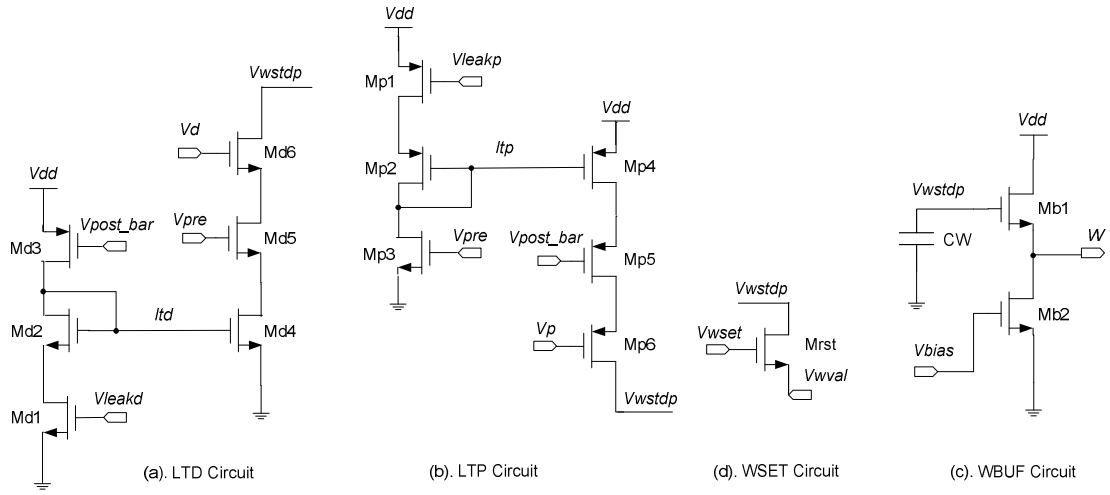


Figure 5.2 STDP circuit; a) LTD, b) LTP, (c) WSET, (d) WBUF sub-circuits.

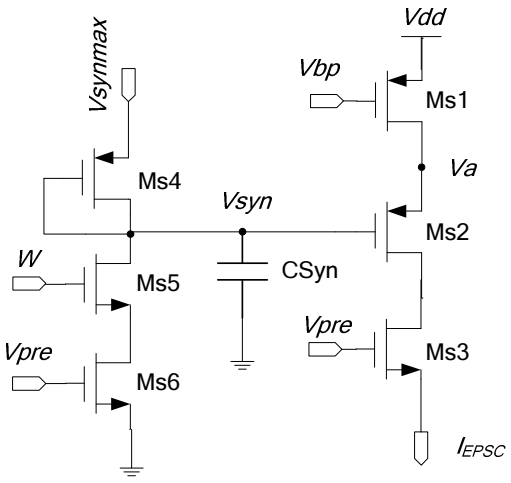


Figure 5.3 Excitatory Synaptic Current Generator (ISYN) circuit

5.3.1 Operation of the STDP circuit

The STDP circuit modifies the synaptic weight, *w* according to the STDP rule. The circuit comprises of 4 sub-circuits namely Long-Term Depression (LTD) circuit, Long-Term Potentiation (LTP) circuit, Synaptic Weight Set (WSET) circuit, and Synaptic Weight Buffer (WBUF) circuit. The LTD and LTP circuit topologies have initially been proposed by Indiveri (2003) to operate in biological time scale. However, appropriately sizing the transistor and capacitors and by shifting the operating point of some transistors, the LTD and LTP circuits are designed to operate in the accelerated time scale. The weight of the STDP circuit, *Vwstdp* is stored in the capacitor C_W. Firings of the pre- and post- synaptic neurons induce the changes to the synaptic weight using the LTD and/or LTP circuits to implement the STDP rule.

The signal *Vpre* is the pre-synaptic firing signal whereas the *Vpost* is the post-synaptic firing signal. The signal *Vpost_bar* is the inverted post-synaptic firing signal. Once the post-synaptic neuron fires, the gate capacitor of Md2 is charged to *Vdd*, the supply voltage, *Vdd*, by switching-on the transistor Md3. The gate capacitor is then continuously discharged with a "leakage" current, through transistors Md1 and Md2. The amount discharge from the gate capacitor of Md2 is approximately proportional to the time after the last post-synaptic neuron firing. The maximum time duration for the capacitor to discharge to a voltage low enough to force the gate voltage of the transistor Md4 to reach its cut-off region of operation is equivalent to the '*LTD time window'*. This is controlled by the voltage *Vleakd*. If a pre-synaptic spike follows the post-synaptic firing within the LTD time window, the LTD circuit reduces the charge in the capacitor Cw by switching-on the current path through the transistors Md4, Md5 and Md6. The voltage *Vd* limits the maximum current through these transistors. The LTP circuit has a complementary topology to the LTD circuit. Figure 5.4 shows the effect on *ltp*, *ltd*, *Vwstdp* and *w* to twenty pre- and post- synaptic spike pairs (shown in Figure 5.4 (a)) - initially, pre-synaptic spike follows the post-synaptic spike and then the post-synaptic spike follows the pre-synaptic spike.

The WSET circuit is used to set the weight of the STDP circuit, *Vwstdp* to the externally set voltage *Vwval* when a pulse is provided at node *Vwset*. The WBUF circuit buffers the synaptic weight and provides it to the ISYN circuit (Figure 5.3) that generates the excitatory post-synaptic current. When the pre-synaptic spike arrives at the synapse, the

Ms6 of the ISYN is switched-on and the current through the transistors Ms5 and Ms6 removes an amount of charge, Δq_{wstdp} from the capacitor Csyn. This reduces the voltage *Vsyn* to generate I_{EPSC}. The value of Δq_{wstdp} is removed approximately proportional to the value of the buffered synaptic weight (*w*). The Ms4 transistor continuously charges *Vsyn* to the resting voltage of *Vsyn* (*Vsynmax*). The *Vwmax* can be set externally. The I_{EPSC} and *Vsyn* values for various synaptic weights, *w* are shown in Figure 5.5- this is obtained by providing continuous pre-synaptic spike train to *Vpre* while slowly varying *w* from zero to 2V.

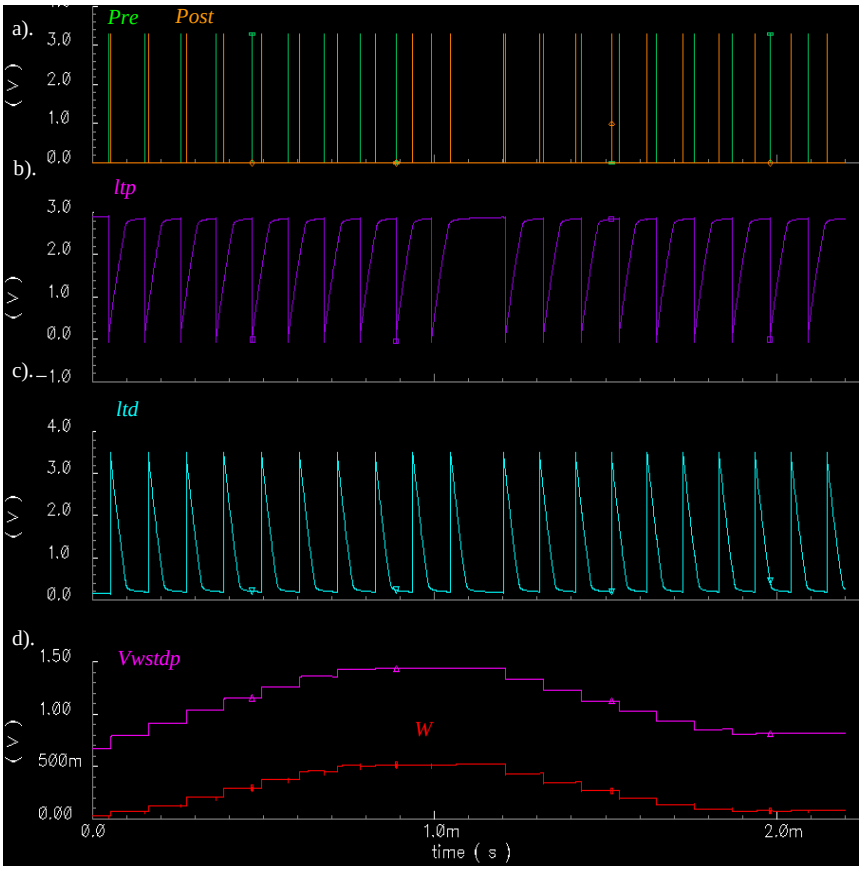


Figure 5.4 (a) *pre* and *post* synaptic spikes; (b) *ltp* and (c) *ltd* node voltages of the STDP circuit; (d) LTP and LTD effects on *Vwstdp* and buffered synaptic weight (*w*).

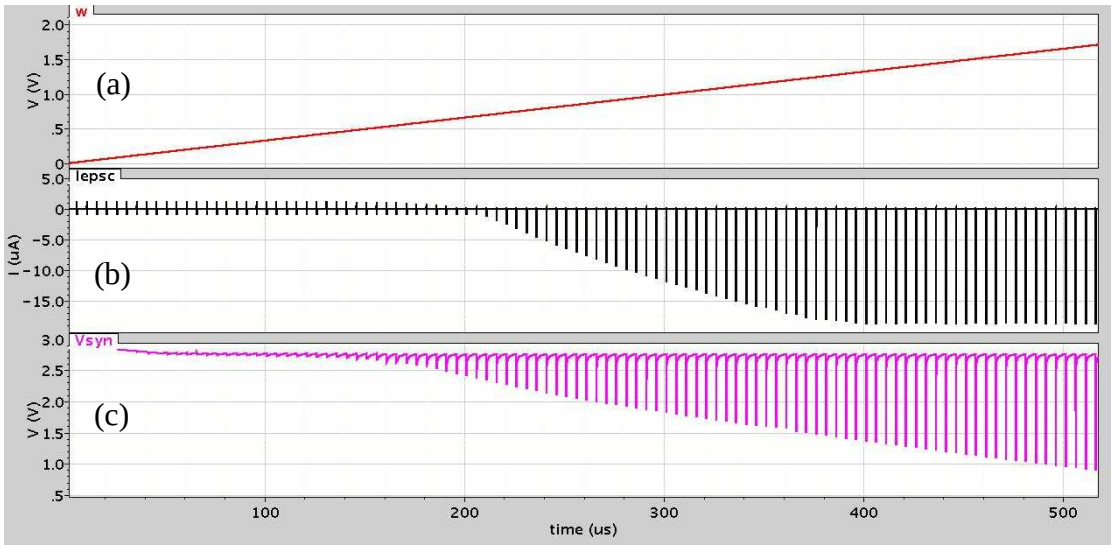


Figure 5.5 Responses of ISYN circuit shown in Figure 5.3 : The synaptic weight *w* is varying (as in (a)) while providing continours pre-synaptic spike train to the ISYN circuit (b) I_{EPSC} and (c) *Vsyn* responses are obtained. *Vsynmax* = 3.3 V;*Vbp*= 0 V;

5.3.2 Simulation results and the layout of the STDP synapse circuit

The STDP synapse circuit is simulated in a standard 0.35um CMOS technology and results are presented in Figure 5.6 to Figure 5.7. The synaptic weight capacitor (Cw ≈ 180 fF) the gate capacitance of the MCw is used The layout of the STDP synapse circuit used in the Cortical Neural Layer (CNL) chip discussed in Chapter 9 is shown in Figure 5.9.

Figure 5.6 shows the STDP curves generated from the STDP circuit shown in Figure 5.2. The STDP curves can be adjusted using control voltages *Vp*, *Vd*, *Vlkp* and *Vlkd*. Figure 5.5 (a) and (b) shows the effect on the magnitude of weight modification (Δw) when *Vd* and the *Vp* are varied respectively, whereas Figure 5.6 (a) and (b) shows how LTP and LTD time windows can be varied by varying the voltages *Vlkd* and *Vlkp* respectively. It should be noted that Δw also depends on the value of *w* and the plots in Figure 5.6 and Figure 5.7 are generated when *w* is at its mid value. In a practical circuit, the device mismatch will also affect these characteristics. By increasing (reducing) the capacitance of the capacitor Cw the magnitude of synaptic modification can be reduced (increased). Here these values are selected such that the continuous repetitions of 50–60 pre/post synaptic spike pairs are needed to reach the maximum weight of the synapse (as in biological experiments Morrisons et al., 2008). Typically LTP and LTD time

windows are in the range of 20 ms to 40 ms and hence in accelerated time these are 20 µs to 40 µs. Shape of the STDP curve use is different to the standard STDP curve most commonly used in the theoretical neuroscience. However, as mentioned in Chapter 2 a variety of STDP curves have been observed experimentally (Abbot et al., 2000) and as long as the STDP curve provide a tuneable STDP plasticity rule and the circuit uses a small number of transistors, it should be a good candidate to use in larger VLSI cortical networks.

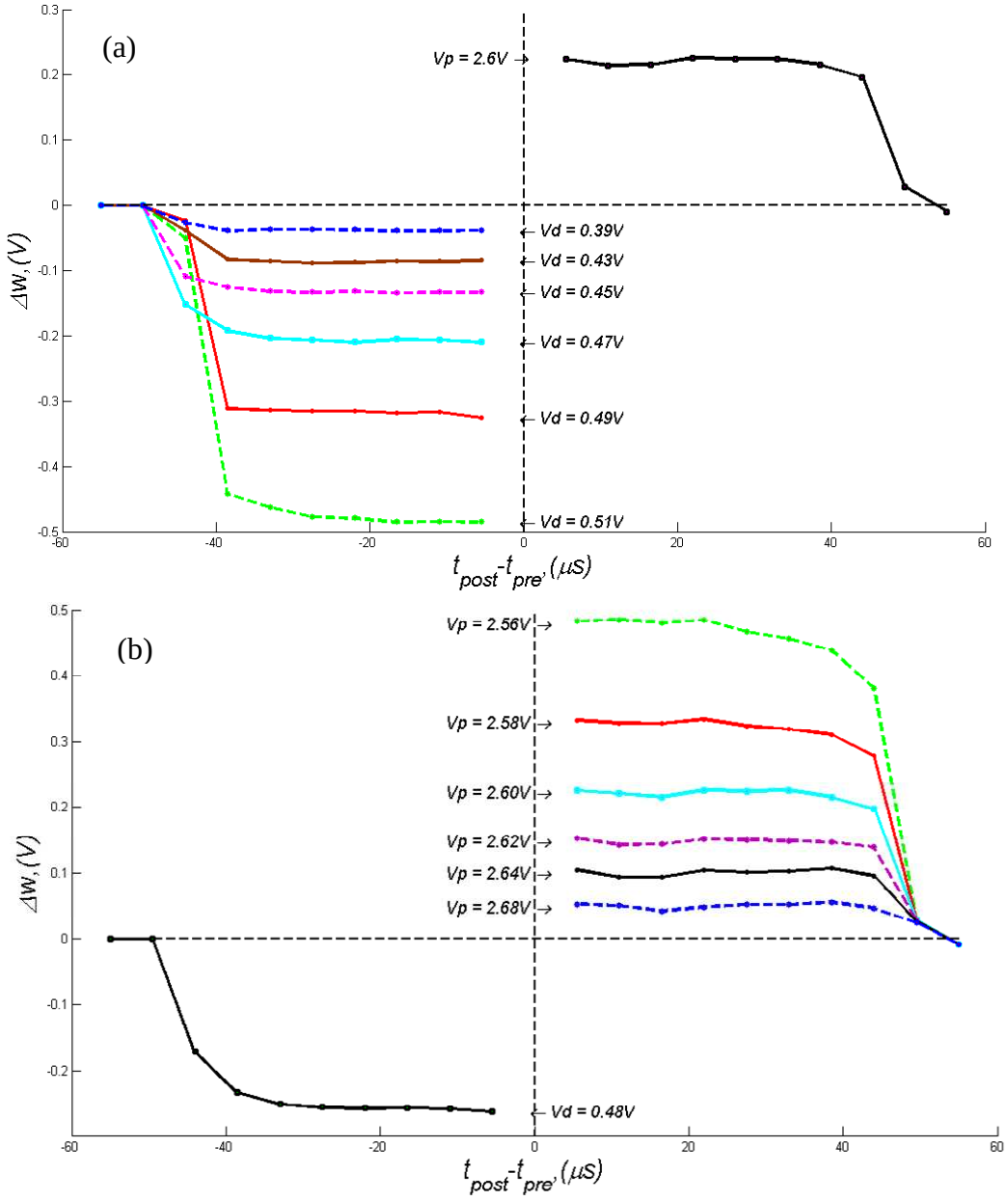


Figure 5.6 (a) LTD and (b) LTP curves generated for various values of *Vd*, and *Vp* respectively.

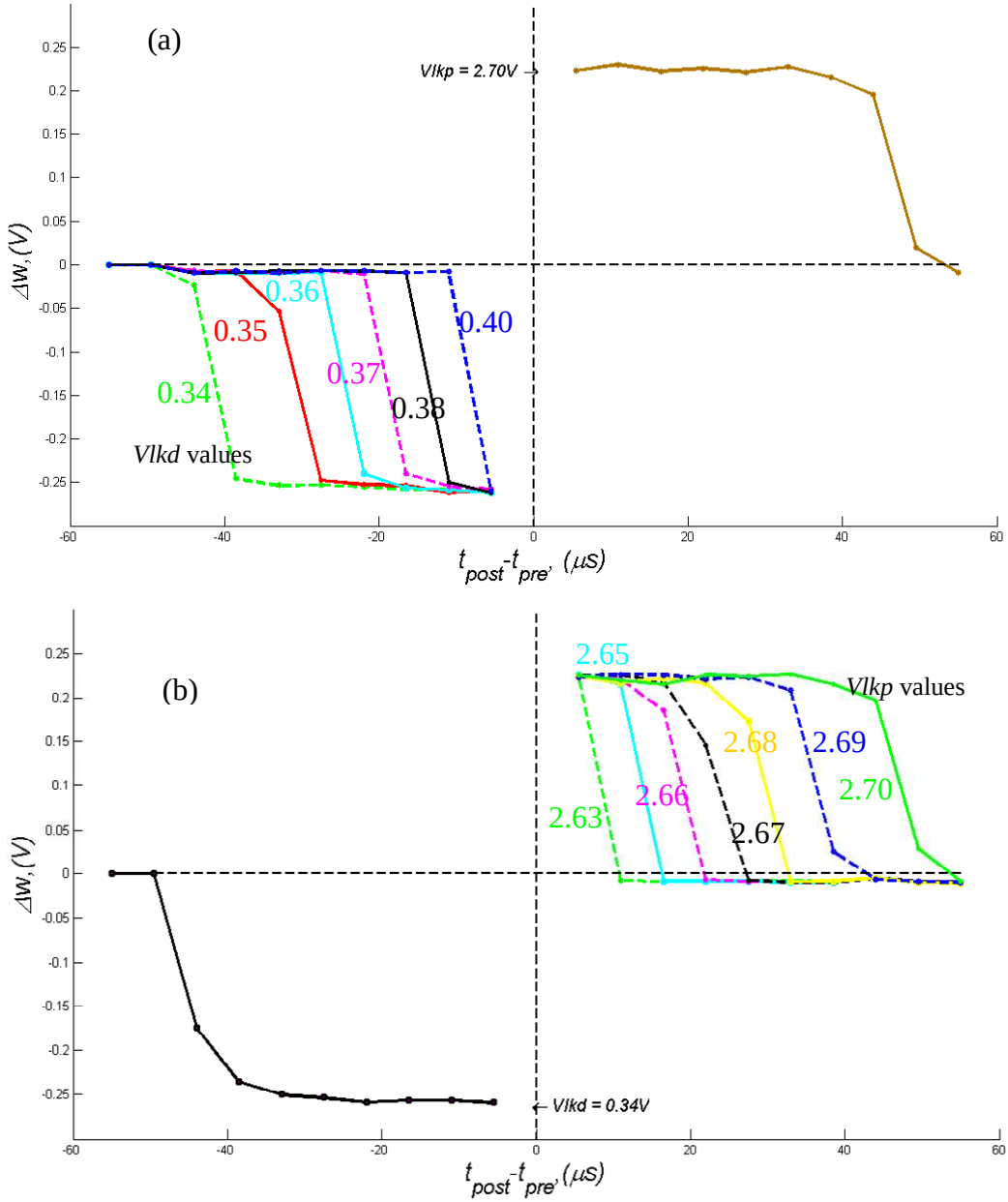


Figure 5.7 (a) LTD and (b) LTP curves generated for various values of *Vlkd*, and *Vlkp* respectively.

The mismatch effect of the circuit has been simulated using 1000 Monte Carlo iterations and Figure 5.8 shows the mean curve and the standard deviation of a typical synaptic weight trace to stimulus of 20 pairs of pre- and post- synaptic spikes shown in Figure 5.4(a), i.e. synaptic weight response to twenty pre- post- synaptic spike pairs- initially the pre-synaptic spike follows the post-synaptic spike, and then the post-synaptic spike follows the pre-synaptic spike. Here, the mismatch models of the AMS 0.35 µm standard CMOS technology are used. It is seen from Figure 5.8 that the variability due to the mismatch is within the operational region of the STDP circuit. Variability is an inherent property of a biological system. Hence, as long as the curves follow the STDP

rule and are within the operating range of the synaptic weight this variability could possibly be exploited as an advantage in a VLSI cortical network, analogous to the variability of the biological synaptic weights.

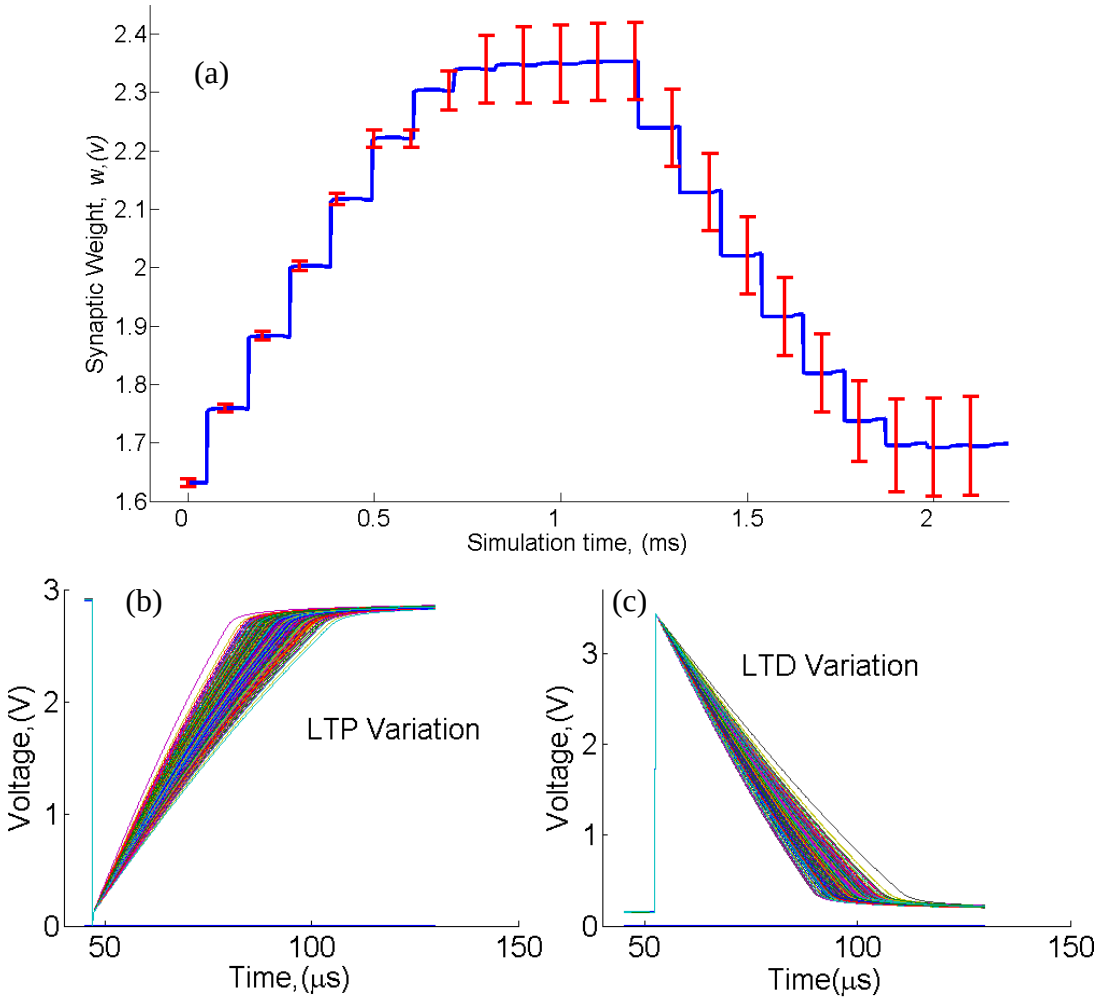
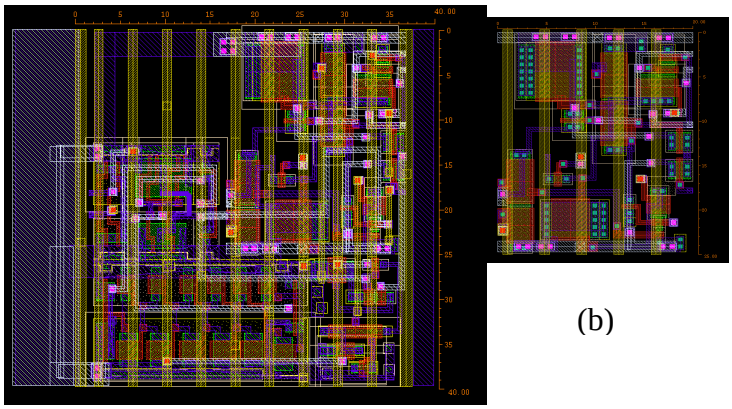


Figure 5.8 Mismatch analysis of the STDP curve (a) Synaptic weight trace showing mean (blue plot) and the standard deviation(red plot) b) Variation of signal *ltp* and (c) *ltd.*



(b)

(a)

Figure 5.9 Layout of the (a) STDP Synapse circuit and (b) STDP circuit

As a capacitor is used to store the synaptic weight, most common problem of this type of STDP circuit implementation is the continuous leakage of the synaptic weight. The leakage time constant of the synaptic weight is approximately 41 ms (in accelerated time). This value depends on the value of the weight and average value is taken. As this circuit operates in 10^3 faster accelerated-time the leakage problem has been reduced.

5.3.3 Mathematical model of the STDP synapse circuit

Approximated mathematical model of the STDP circuit is given below.

Synaptic weight, w

The weight of the synapse, *w* is evolving as in equation 5.4.

$$\dot{w} = -\frac{w}{\tau_l} + STDP(\Delta t_{pre/post}).\delta(t - t_{pre/post}) \qquad (5.4)$$

Where $\tau_l \approx$ 41 ms (an approximated time constant of the leakage of weight, *w* due to the leakage of *Vwstdp* node in the circuit shown in Figure 5.2); $\Delta t_{pre/post} = t_{post} - t_{pre}$.

STDP curve, STDP(τ)

As seen in Chapter 5, the STDP curve can be approximated to a piecewise linear curve as given in equation 5.5.

$$STDP(\Delta t_{pre/post}) = \begin{cases} \Delta w_p & \text{when} \quad 0 < \Delta t_{\text{pre/post}} < t_{lin} \\ \left(\Delta w_p \times (\Delta t_{\text{pre/post}} - t_{lin}) \Big/ (t_{wp} - t_{lin}) \right) & \text{when} \quad t_{lin} < \Delta t_{\text{pre/post}} < t_{wp} \\ -\Delta w_d & \text{when} \quad 0 > \Delta t_{\text{pre/post}} > -t_{lin} \\ \left(\Delta w_d \times (\Delta t_{\text{pre/post}} + t_{lin}) \Big/ (t_{wd} - t_{lin}) \right) & \text{when} \quad -t_{lin} > \Delta t_{\text{pre/post}} > -t_{wd} \\ 0 & \text{o.w} \end{cases} \quad (5.5)$$

Where, Δw_p and Δw_d can be set to a value between 0 V to 1.2 V. t_{lin} is the start time of the linear region of the STDP curve, default value is 20 µs. t_{wp} and t_{wd} are time windows of long-term potentiation and long-term depression respectively. These can be adjusted to any value between 1 µs and 70 µs ($t_{lin} \leq t_{wp}, t_{wd}$).

Vsyn and I_{EPSC} of the ISYN circuit

An intermediate state variable *Vsyn* is used to generate the post synaptic current, I_{EPSC}. The ISYN circuits (Figure 5.3) in the CNL chip the voltage *Vsynmax* is connected to the supply voltage *Vdd*. The voltage *Vsyn* depends on the amount charge and/or discharge on to the capacitor C_{syn}. The diode connected transistor Ms4 is in saturation if the *Vsyn* is below its resting voltage (*Vdd-vt*). The Ms5 is assumed to be in saturation as the transistor Ms4 pulls the voltage *Vsyn* to (*Vdd-vt*) at a higher rate whenever *Vsyn* voltage is reduced. In the practical implementation, the *Vsyn* reduction due to a pre-synaptic spike occurs within 3 ns duration. However in the equation 5.6, the total reduction of *Vsyn* due to the pre-synaptic spike is considered as an instantaneous reduction at the arrival time of the pre-synaptic spike. Therefore, the magnitude of the drain current of the Ms5 transistor, $I_{Ms5} = K_d (w - v_t)^2 \Delta t_{pre}$ considered as a rate of reduction in *Vsyn* voltage (V/s). Approximated vales of state *Vsyn* and I_{EPSC} are given below.

$$\dot{V}_{syn} = K_c (Vdd - V_{syn} - V_t)^2 - K_d (w - V_t)^2 \Delta t_{pre} \delta(t - t_{pre}) \quad (5.6)$$

$$I_{EPSC} = \begin{cases} \max\{I_{\max}, \mathrm{K_p}(V_{Ms2gs} - Vt)^2\} & \text{when pre-neuron fires \& } V_{Ms2gs} > V_t \\ 0 & \text{otherwise} \end{cases} \quad (5.7)$$

Where, $\mathrm{K_c} = \frac{1}{2\mathrm{C_{syn}}}\left(\frac{\mathrm{W}}{\mathrm{L}}\right)_{\mathrm{Ms4}} \mu_\mathrm{p}\mathrm{C_{ox}}$, $\mathrm{K_d} = \frac{1}{2\mathrm{C_{syn}}}\left(\frac{\mathrm{W}}{\mathrm{L}}\right)_{\mathrm{Ms5}} \mu_\mathrm{n}\mathrm{C_{ox}}$, Δt_{pre} spike duration and it is 3 ns; $\mathrm{K_p} = \frac{1}{2}\left(\frac{\mathrm{W}}{\mathrm{L}}\right)_{\mathrm{Ms2}} \mu_\mathrm{p}\mathrm{C_{ox}}$; $\mathrm{C_{syn}} \approx$ 20 fF ; V_{Ms2gs} is the gate to source voltage of the transistor Ms2, which is equal to (*Va*-*Vsyn*). The CMOS process parameters $\mathrm{C_{ox}}$, V_t, μ_n, and μ_p are gate oxide capacitance per unit area, threshold voltage of transistor, charge-carrier effective mobility of nMOSFET, and pMOSFET respectively. From AMS standard 0.35 μm CMOS technology process parameters μ_n, μ_p, $\mathrm{C_{ox}}$, and v_t value are 370 $\mathrm{cm^2/VS}$, 126 $\mathrm{cm^2/VS}$, 4.54 $\mathrm{fF/\mu m^2}$, and V_t,=0.5 V respectively. The user can scale the I_{EPSC} current for a given value of *w*, using externally controllable voltage *Vbp* (as shown in Figure 6.17 in Appendix A). This controllable voltage could also be used to limit the maximum I_{EPSC}.

5.4 Dopamine Modulated Synapse Circuit

The DA STDP Synapse Circuit approximately implements the dynamics of the dopamine modulated synapse model proposed by Izhikevich (2007). The DA Modulated Synapse circuit comprises of three sub-circuits: the Eligibility-Trace (ET) circuit, the Synaptic Strength circuit and the Synaptic Current Generator circuit; the circuit schematics are shown in Figure 5.10 and Figure 5.11. The DA modulated synapses receive DA signal from the common DA Generator Circuit shown in Figure 5.12.

5.4.1 Operation of the DA Modulated Synapse Circuit

The Eligibility-Trace circuit of the DA Modulated Synapse generates the 'eligibility traces' (ET) according to the STDP rule where the eligibility potentiates or depresses depending on the sequence of pre-/post- synaptic spiking activity. The Synaptic Strength circuit ensures strengthening or weakening of the synaptic strength (synaptic weight) depending on the eligibility trace and the reward which is signalled by DA. DA Generator circuit provides the DA signal such that its amplitude represents the rewarding status of the network. Finally, the Synaptic Current Generator circuit generates an excitatory post-synaptic current approximately proportional to square of the synaptic strength. Detailed description of each sub-circuit is given below.

Elegibility-Trace Circuit

The Eligibility-Trace circuit is shown in Figure 5.10 and it generates potentiating and depressing parts of the ET separately using two sub-circuits: Long-Term Eligibility Potentiation (LTEP) circuit and Long-Term Eligibility Depression (LTED) circuit. The design of these circuits are similar to the STDP Synapse circuit proposed in Section 5.3 of this chapter, however complementary topology of the STDP circuit is used here. The capacitors Cwp and Cwd store the potentiation (*Vetp*) and depression (*Vetd*) information of the ET respectively. Firings of the pre- and post- synaptic neurons induce changes to the *Vetp* and *Vetd*, implementing the STDP rule. If the DA is present, these synaptic changes will result in modification of the synaptic strength, *S*, (produced in the Synaptic

Strength circuit) during the critical window of a few milliseconds (equivalent to a few seconds in biological time) before the *Vetp* and *Vetd* decay to zero.

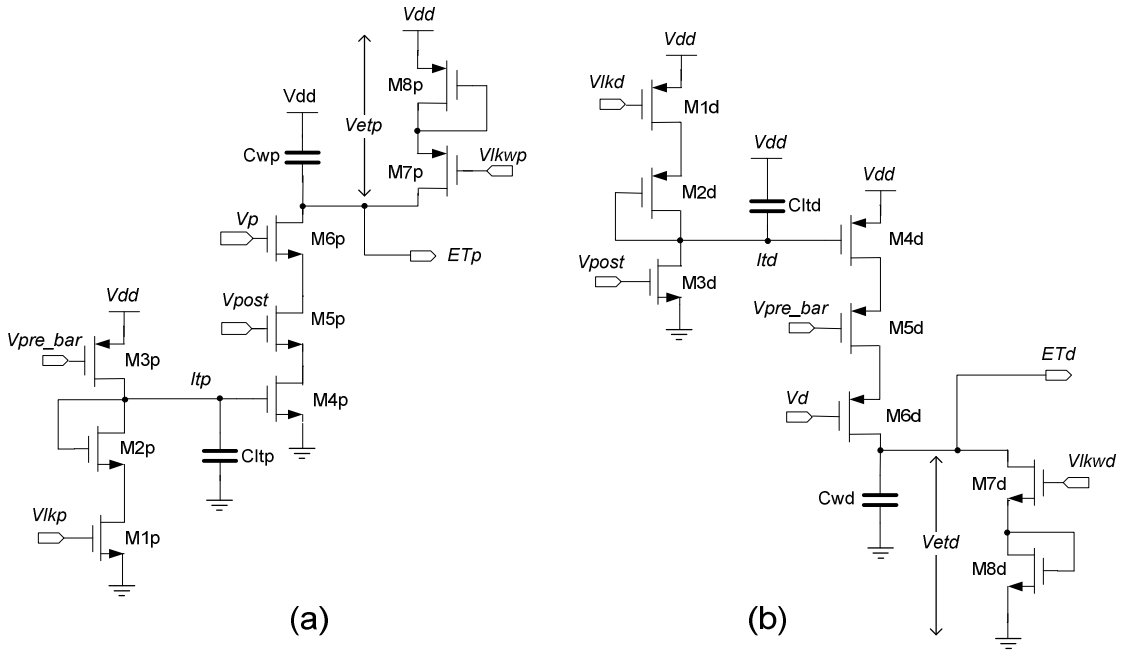


Figure 5.10 Eligibility-Trace circuit: (a). LTEP circuit; (b). LTED circuit

The signal *Vpre_bar* is the inverted pre-synaptic firing signal (*Vpre*) where as the signal *Vpost* is the post-synaptic firing signal. Once the pre-synaptic neuron fires, the capacitor Cltp is charged to *Vdd* by switching-on the transistor M3p. The capacitor is then continuously discharged with a "leakage" current, through transistors M1p and M2p. The amount of charge removed from Cltp is approximately proportional to the time after the last pre-synaptic neuron firing. The maximum time duration for the capacitor to discharge to a voltage low enough to force the gate voltage of the transistor M4p to reach its cut-off region of operation is equivalent to the *'LTP time window'*. This is controlled by the voltage *Vlkp*. If a post-synaptic spike follows the pre-synaptic firing within the LTP time window, the LTEP circuit increases the charge in the capacitor Cwp by switching-on the current path through the transistors M6p, M5p and M4p. The voltage *Vp* limits the maximum current through these transistors. The *Vetp* decays to *Vdd* through M7p-M8p. The speed of decay is controlled by the voltage *Vlkwp*.

The LTED circuit has a complementary topology to the LTEP circuit. Both outputs signals of the ET circuit, *ETp* and *ETd*, are provided to the Synaptic Strength circuit to

produce the synaptic strength change. When the pre-synaptic neuron fires, the synaptic strength, *S,* regulates the amount of post-synaptic current *(PSC)* injected to the post-synaptic membrane.

Synaptic Strength Circuit

The Synaptic Strength circuit is shown in Figure 5.11, and it receives the eligibility-trace signals the *ETp* and *ETd* from the ET circuit, and the DA pulse signal, *Vda,* and its inverted signal, *Vda_bar,* from the DA Generator circuit. The circuit parts for the synaptic strength potentiation and depression, shown in Figure 5.11 (a), are complementary. When considering the potentiation part of the circuit, during the time *Vda_bar* is at logic low, the potential divider (transistors M1p, M3p-M4p) creates a potential at *Vsp* proportional to the *Vetp* voltage. The pulse width of the *Vda_bar* signal is proportional to the amount of DA. Hence, the amplitude and the width of the signal *Vsp* of the Synaptic Strength circuit carries the *Vetp* and DA level information respectively. The M6p transistor can operate either in the sub-threshold or in the linear range depending on the externally controlled voltage, *Vsmp*. If the transistor M6p (M6d) is biased to operate in the sub-threshold region, the charge through the M6p and M7p (M6d and M7d) is proportional to the product of the DA level and the exponential of the eligibility traces, *Vetp* (*Vetd*). Hence in this case, the net charge increase at the capacitor, Cs is proportional to the product of the DA-level and $e^{(Vetp-Vetd)}$.

If the transistor is biased to operate in the linear region, then the charge through the M6p and M7p transistors (charging the capacitor Cs) is proportional to the product of the DA level and the *Vetd* voltage. The depression circuit M1d-M7d works in the same way to discharge the capacitor MCs. Hence the net charge increase at the capacitor, MCs which stores the synaptic strength, is proportional to the product of the DA-level and the *Vetp*-*Vetd* voltage difference.

Once the pre-synaptic neuron fires, the synaptic output circuit creates a post-synaptic current (*PSC*) as a function of the synaptic strength, *S.* The externally controlled *Vbp* voltage limits the PSC current flow to the membrane of the post-synaptic neuron.

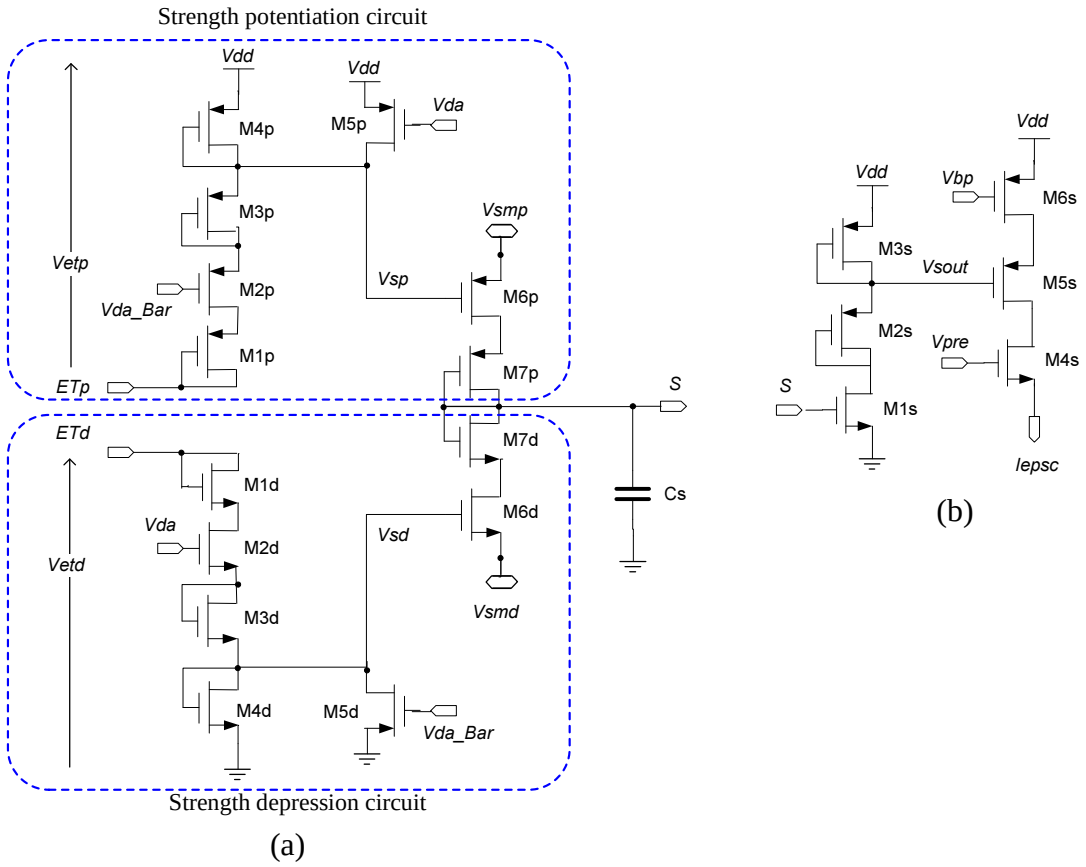


Figure 5.11 (a) Synaptic Strength circuit, (b) Synaptic Current Generator circuit.

DA Generator Circuit

The DA Generator circuit is shown in Figure 5.12 and it provides the DA pulse signal (*Vda*) to the DA Modulated Synapses circuit in order to update the strengths of the synapses. The level of extracellular DA (which is represented by the voltage *Veda*) is increased by a burst of spikes provided at the gate of the M1 transistor. This burst of spikes is assumed to be provided from a bursting neuron output as a consequence of the reward prediction clue or reward-triggering action. After the burst, the DA level decays towards *Vdd* through the transistor M3. The time constant of the decay can be controlled using the voltage *Vlkda*. The DA level is buffered to the node, *Vdab* using the source follower (M4 and M5) and is provided to the transmission gate TR1. The transmission gate is switched periodically using an externally controlled clock signal, *Vda_clk*. When the TR1 is 'ON' the parasitic capacitance at the node *Vdat* is charged to the voltage at node *Vdab*, and the transistors M9 and the M10 are switched 'OFF' and 'ON'

respectively. This creates the rising edge of the DA pulse (*Vda*). Then the parasitic capacitor is discharged through the transistors M6 and M7. The speed, at which this capacitor is discharged, is controlled by the voltage *Vlk*. If the *Vlk* is kept at a fixed voltage, the time taken to discharge the capacitor such that the transistors M9 and the M10 are switched 'ON' and 'OFF' respectively, is approximately proportional to the level of DA (*Veda*). Therefore, the pulse width of the *Vda* signal is proportional to the DA level. The buffer at the output is used to provide faster rise and fall times for the *Vda* signal.

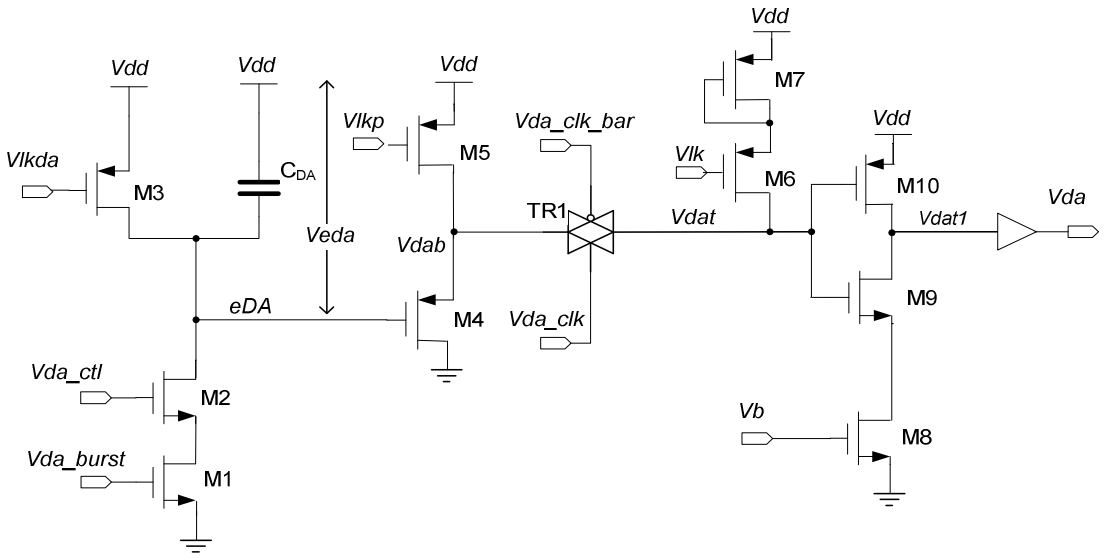


Figure 5.12 DA Generator Circuit of DA modulated synapses.

5.4.2 Simulation results

The DA Modulated circuit is simulated in a standard 0.35 µm CMOS technology and results are presented in Figure 5.13 to Figure 5.17. This synapse is designed and fabricated such that it can be configured to work as a DA modulated synapse or as a STDP synapse without the DA modulation. The Figure 5.18 shows the layout of the configurable STDP/DA-STDP Synapse circuit.

Figure 5.13 and Figure 5.14 show the STDP curves generated from the Eligibility Trace circuit shown in Figure 5.10. The STDP curves can be adjusted using control voltages *Vp*, *Vd*, *Vlkp*, and *Vlkd*. Figure 5.13 (a) and (b) shows the effect on the magnitude of changes to the voltage *Vetp* (*ΔVetp*) and the *Vetd* (*ΔVetd*) when *Vp* and the *Vd* is varied respectively. Figure 5.13 (a) and (b) shows how LTP and LTD time windows can be

varied by varying the voltages *Vlkp* and *Vlkd* respectively. It should be noted that *ΔVetp* and *ΔVetd* also depend on the value of *Vetp* and *Vetd* respectively; the plots in Figure 5.13 and Figure 5.14 are generated when the *Vetp* and *Vetd* are at their mid values.

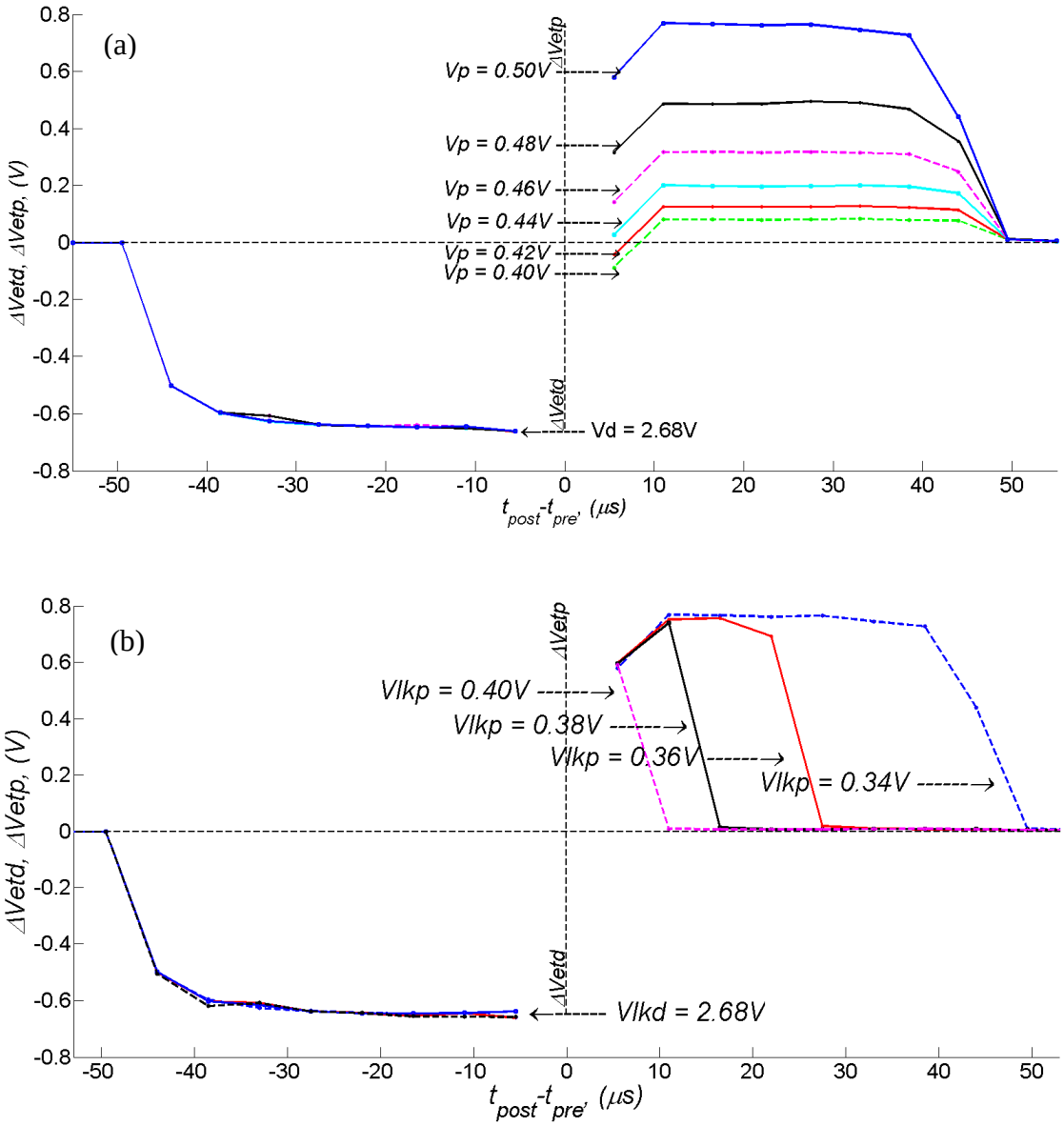


Figure 5.13 STDP curves generated using the ET circuit; Plots show the changes to the *Vetp* and *Vetd* as a function of a time interval between pre- and post-synaptic spike; ΔVetp curves with variation of control voltages (a) *Vp*, and (b) *Vlkp*.

Characteristics of the Synaptic Strength Circuit (Figure 5.11) are shown in Figure 5.15. The amount of change in strength (*ΔS*) is plotted as a function of *Vetp*/*Vetd* value and *Vdap* pulse width for a single update of strength. The frequency of update can be set by changing the DA clock frequency (using the *Vda_clk*). Further, the amount of increase

or decrease in *S* per single update, for various *Vetp* and *Vetd* values can be changed independently by tuning *Vsmp* and *Vsmd* values respectively. The amount of change in *S* also depends on the actual value of *S*, the plots are shown for a mid-value of *S*= 1V. The Figure 5.15 shows that the change in synaptic strength is dependent both on the eligibility trace value and DA level, similar to the product in Equation 5.2.

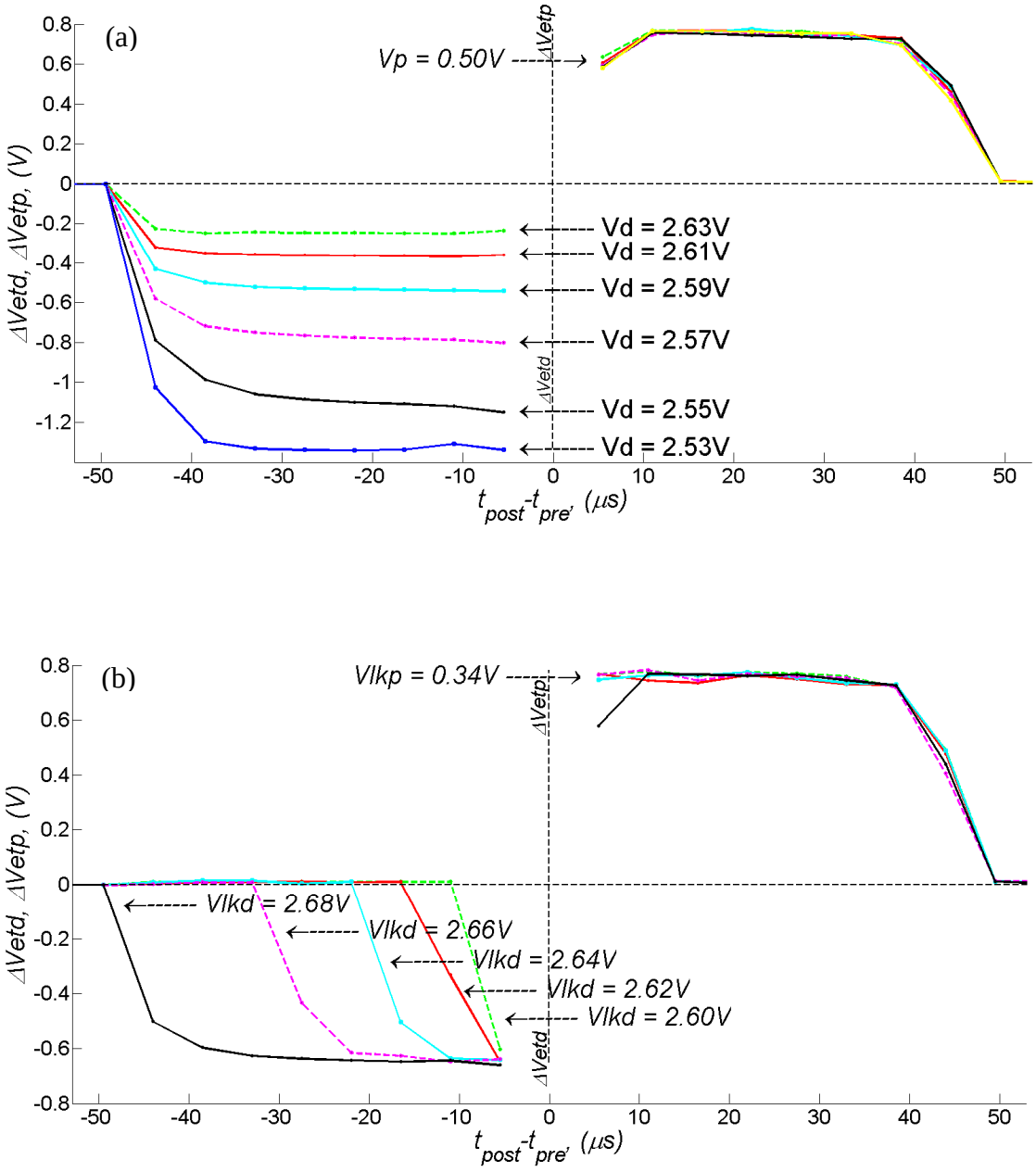


Figure 5.14 STDP curves generated using ET circuit; Plots show the changes to the *Vetp* and *Vetd* as a function of a time interval between pre- and post-synaptic spike; *ΔVetd* curves with variation of control voltages (c) *Vd* and (d) *Vlkd*.

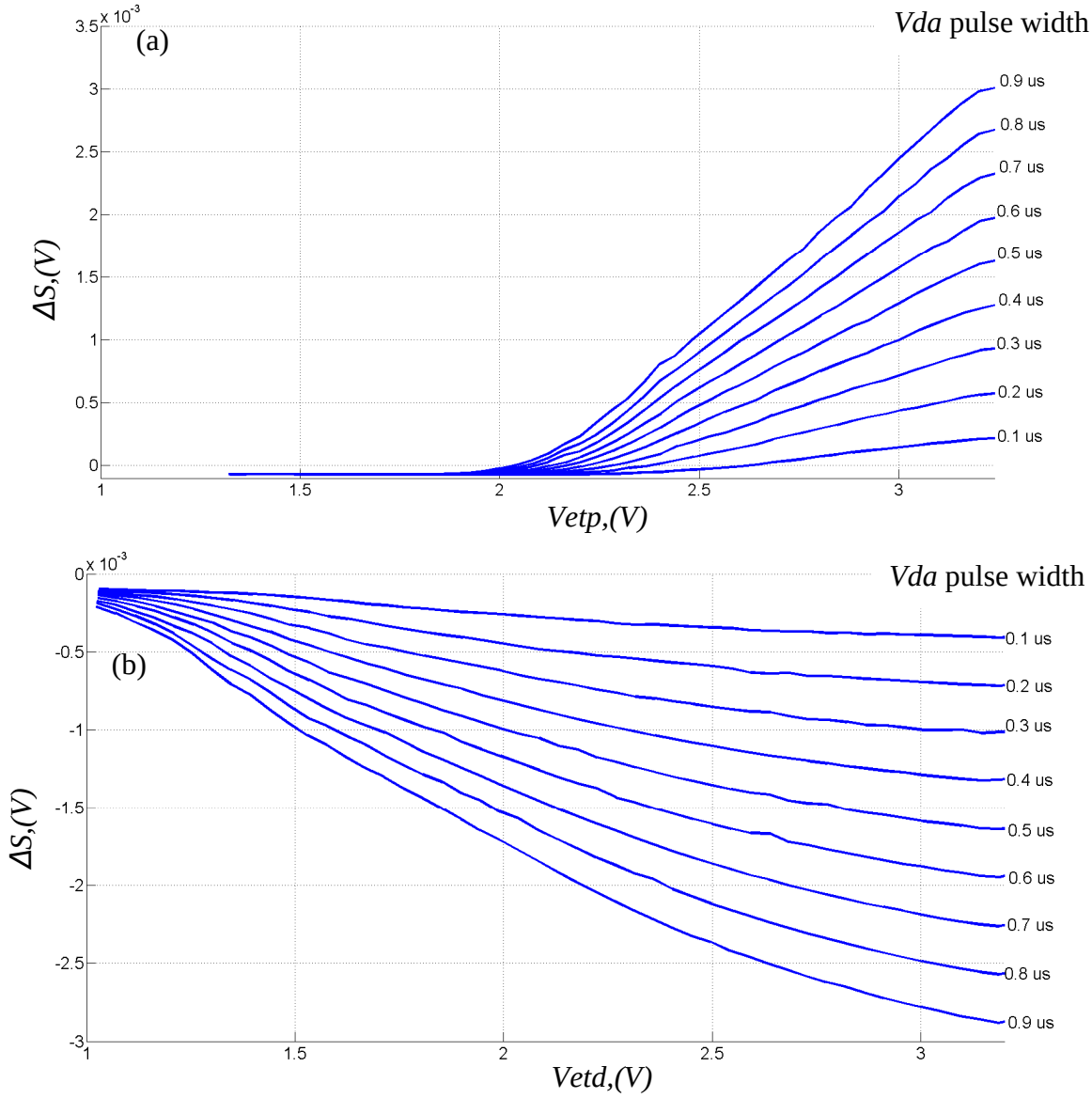


Figure 5.15 Changes in synaptic strength S for different (a) *Vetd* and *Vda* pulse widths, (b) *Vetp* and *Vda* pulse widths.

Figure 5.16 shows the waveforms of the responses of *Vda*, synaptic strength (*S*), the internal voltages of the Synaptic Strength circuit, *Vsd*, *Vsp* to a given sinusoidal inputs of *Vdea*, *ETp* and *ETd*. Periodically, high amplitude eligibility potentiation input is provided to the Synaptic Strength circuit than that of the eligibility depression input, while slowly varying the DA (*Veda*) level provided to the DA Generator circuit. It is seen that the strength of the synapse is increased when higher *Vetp* value and the DA pulse (*Vda*) are present. It is seen that the DA pulses are generated by the DA Generator circuit when the *Veda* is at a higher value (i.e. where the *eDA* node voltage of the DA Generator circuit is at a lower value).

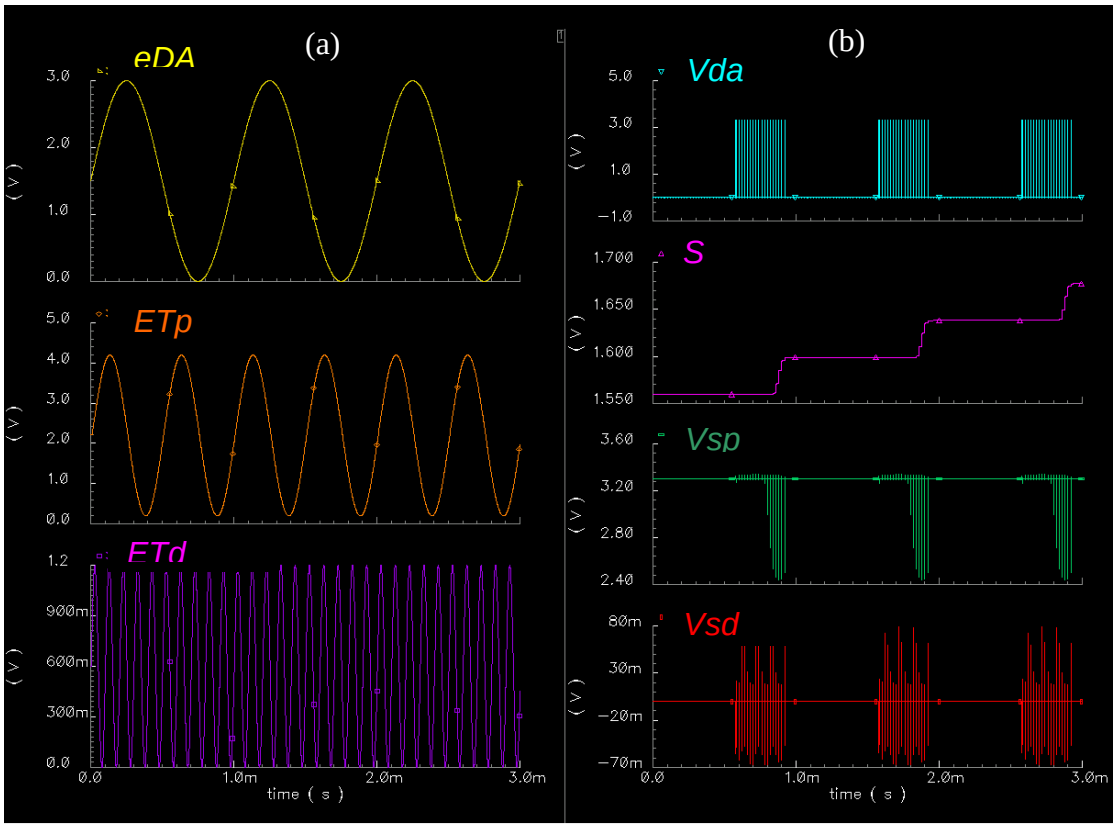


Figure 5.16 Responses of the Synaptic Strength circuit; (a) Top to bottom graphs: *Veda*, *Vwp* and *Vwd* sinusoidal inputs provided to the DA Generator circuit and the Strength circuit; (b) Top to bottom graphs: *Vda* pulses, synaptic strength *S* and internal voltages of the Synaptic Strength circuit, *Vsd* and *Vsp* responses the sinusoidal inputs given in (a).

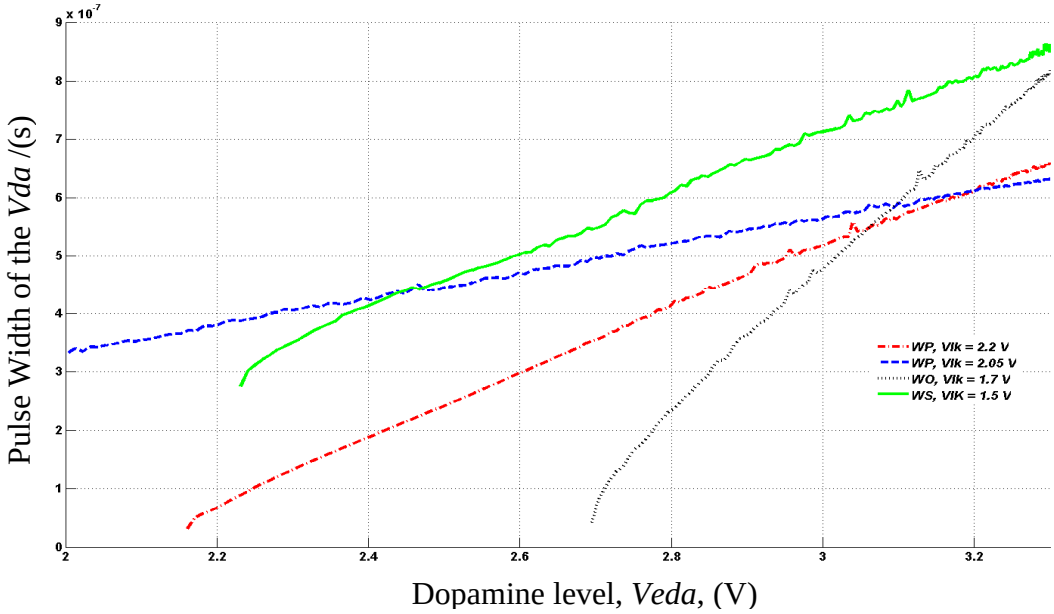


Figure 5.17 Variation of the DA pulse width with respect to the DA level (*Veda)* generated from the DA generator circuit shown in Figure 5.12; four graphs correspond to four process corners: Worst Speed (WS), Worst Power (WP), Worst Zero (WZ) and Worst One (WO).

The DA Generator circuit generates the *Vdap* pulse widths proportional to the DA level. The variation of the *Vdap* pulse width with respect to the DA level (*Veda*) is shown in Figure 5.17 for the four worst case process corners: worst speed (WS), worst power (WP), worst zero (WZ) and worst one (WO). This illustrates that the DA generator circuit produces at least 0.3 µs pulse width difference for the maximum and minimum DA levels.

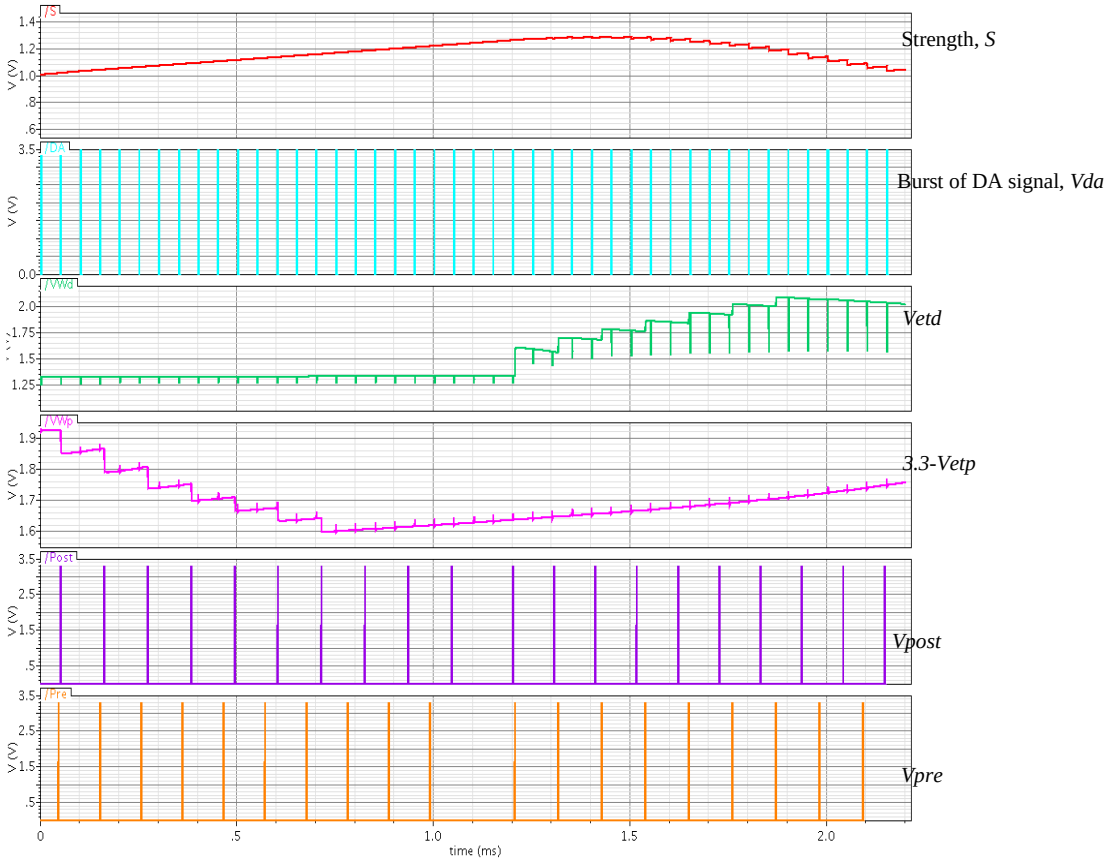


Figure 5.18 Changes to *Vetp*, *Vetd* and synaptic strength *S* when post-synaptic spike (*Vpost*) follows pre-synaptic spike (*Vpre*) and pre-synaptic spike follows post-synaptic spike.

Figure 5.18 shows the effect on *Vetp*, *Vetd*, and *S* when pre-synaptic spike follows the post-synaptic spike and the post-synaptic spike follows the pre-synaptic spike, while bursting (repeating four spike with higher inter burst frequency) spike train of DA pulse signal is given to the DA Generator circuit to increase the DA level of the synapse.

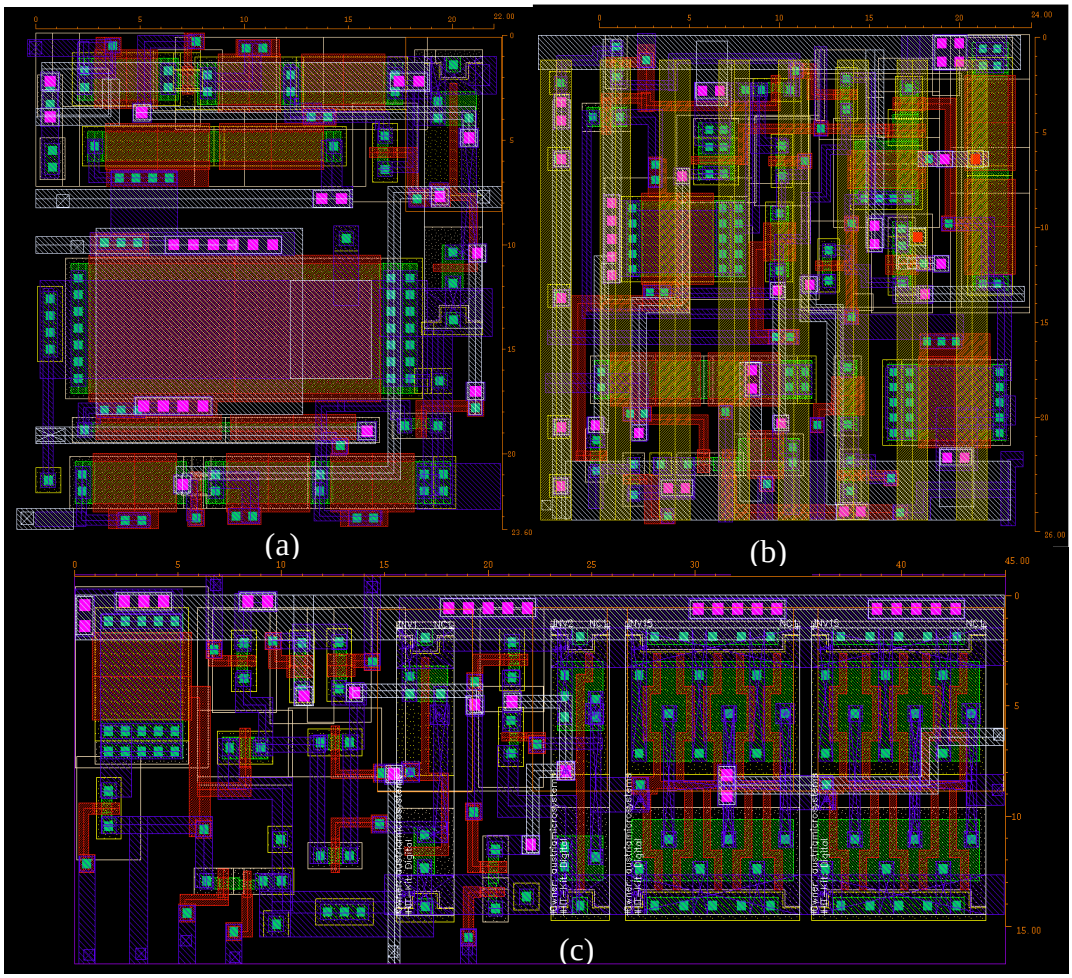


Figure 5.19 Layout of the (a) Synaptic Strength circuit, (b) STDP circuit, and (c) DA Generator circuit.

5.5 Discussion and Conclusion

Circuit implementations of a STDP synapse and a Dopamine modulated STDP synapse are presented in this chapter. The dopamine-modulated synapse circuit implements a model similar to the one proposed in Izhikevich (2007), where eligibility traces are used to provide the dynamics required to facilitate the learning of synaptic strength based on spike-time-dependent plasticity rule and a distal reward signal. The circuit has applications in VLSI implementations of biologically-plausible neural networks.

To prove the concept in hardware the STDP/DA Neuron chip and the CNL chip have been fabricated in a standard 0.35 µm CMOS technology. The STDP/DA Neuron chip contains 28 STDP and DA modulated STDP synapses with a DA generator circuit, and two cortical neuron circuits. The size of the synapse cell layout is 26 µm x 50 µm and

these synapses can be configured to work as DA modulated synapse or as a STDP synapse without the DA modulation. The experimental results obtained from the chip are given in Chapter 8. The configurable synapse circuit typically consumes between 2 µW and 5 µW power at *Vpre* and *Vpost* synaptic spike rates of 200 kHz (i.e high neural activity level), but it could be as high as 40 µW, depending on the synapse state, parameters and spike rates. The DA Generator circuit, which is shared by many synapses, consumes up to 600 µW of power (worst case).

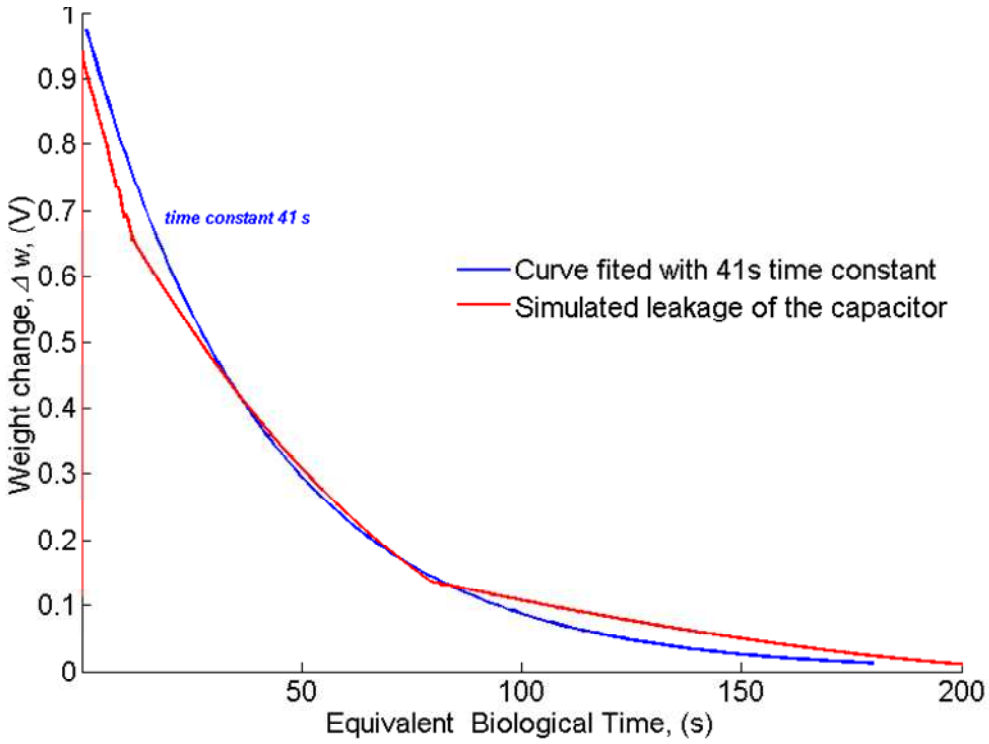


Figure 5.20 Leakage of the synaptic weight of the STDP circuit: in Spice circuit simulation (red curve) and 41s time constant decay plot (blue curve).

As a capacitor is used to store the synaptic weight, most common problem of the STDP circuit implementation is the continuing leakage of the synaptic weight. However, in DA modulation synapse this effect is used to an advantage as the eligibility leakage is a requirement (as per in Equation 5.1). In the STDP synapses, this influences the dynamics of longer time simulations, however, as the circuits operate in 10^3 faster time scale this effect is less detrimental than in biological real time VLSI synapse implementations (Indiveri, 2003). The leakage of the synaptic weight is equivalent to approximately 40 seconds time constant in biological time. Figure 5.20 shows the weight change due to leakage of the capacitor in equivalent biological time scale.

Although this value depends on the value of the weight, highest weight leakage value, corresponding to highest synaptic weight, is taken for the purpose of arriving at the highest leakage. The leakage effect can be minimised by providing a regular artificial spike pair (pre- and post- synaptic spikes) to compensate the leakage.

It is seen that the circuit variability due to the mismatch is within the operational region. All the curves obtained from the Monte Carlo simulation follow the STDP rule and are within the operating range of the synaptic weight. This variability could possibly be exploited as an inherent property of a VLSI cortical network that is analogous to the variability of the synaptic weights in biological systems.

CHAPTER 6 : SHORT-TERM DYNAMIC SYNAPSE CIRCUITS

The chapter propose inhibitory and excitatory synapse circuits that have either facilitating or depressing short-term synaptic dynamics. In arriving at these synapse circuits, the basic properties of short-term dynamics of the computational neural model by Abbott et al. (1997) is used as a guide to arrive at a phenomenological model that implements compact VLSI circuit with suitable plasticity rules. Hence, the initial section introduces the computational neural model by Abbott et al. (1997). Simplified mathematical formulation of facilitating or depressing dynamics used in the synapse circuit models are given in Section 6.2, which is the approximated dynamics used for implementing short-term dynamics of the proposed synapse circuits.

These synapse circuits have been implemented in a standard 0.35 µm CMOS technology. The circuit operation and the simulation results of these circuits are presented in this chapter. In practical implementation, these short-term dynamic plasticity rules can be switched off or switched on, by biasing using appropriate voltages, as discussed in the circuit operation Section 6.3.

The neurons that excite other neurons are called excitatory neurons, and these neurons are equipped with excitatory synapses. Similarly, the inhibitory neurons inhibit connected neurons using inhibitory synapses. Having inhibitory and excitatory neurons in a network can provide stable network activities. The synaptic facilitation and depression of synapses in a network provide a dynamic gain-control mechanism. A single neuron in the cortex receives approximately 10 000 synaptic inputs, where each input could have a wide variety of different spike rates ranging from less than 1 Hz to more than 200 Hz (Abbott et al., 1997). Hence, the information carried by a slowly firing input synapse may be ignored by random fluctuations in the activity of a synapse firing at high rates. This is avoided by having depressing synapses that effectively decrease the gain of high-rate firing as compared with slowly firing inputs (Abbott et al., 1997). Further, the continuously firing facilitating synapse on a network could become dominant over the other rest of the inputs to the neuron (e.g. continuous spike train to an inhibitory facilitating synapse could be silence the activities of the post-synaptic neuron). Though these dynamics are important properties to have in a cortical

network, none of the network implementations discussed in Chapter 3 includes the depressing and facilitating synapses.

In addition to the other circuits, the four types of synapses presented in this Chapter are used in the fabricated CNL IC presented in Chapter 9. The operations of these four types of synapses are similar and the mathematical model of one of the VLSI short-term synapse circuit is presented in Appendix A.

6.1 The Abbott Model of the Short-Term Synaptic Plasticity

In the Abbott et al. (1997) model, the product of the maximum conductance ($\overline{g_s}$), fraction of open post-synaptic channels (P_s), and fraction of pre-synaptic sites that are releasing a neurotransmitter (P_{rel}) is used to obtain the synaptic conduction (g_s) as shown in equation 6.1. The factor P_{rel} incorporates the facilitating or depressing effect on the short-term dynamics. The facilitating synapse can be modelled as the pre-synaptic process that alters the P_{rel} shown in equation 6.2. The f_F controls the degree of the facilitation ($0 \leq f_F \leq 1$). Similarly, depression is modelled as shown in equation 6.3 and the f_D controls the degree of depression ($0 \leq f_D \leq 1$). The P_{rel} decays exponentially with a time constant τ_P, aiming at the 'resting' level P_0.

$$g_s = \overline{g}_s P_s P_{rel} \qquad (6.1)$$

P_{rel} for facilitating synapse

$$\frac{dP_{rel}}{dt} = \frac{P_0 - P_{rel}}{\tau_P} + f_F(1 - P_{rel})\delta(t - t_{pre}) \qquad (6.2)$$

P_{rel} for depressing synapse

$$\frac{dP_{rel}}{dt} = \frac{P_0 - P_{rel}}{\tau_P} - f_D(1 - P_{rel})\delta(t - t_{pre}) \qquad (6.3)$$

Where t_{pre} is the arrival time of the pre-synaptic spike; $\delta(t)$ is the Dirac-delta function; f_F and f_D controls the degree of facilitation and depression (with $0 \leq f_F, f_D \leq 1$) respectively.

Other important computational models of short-term dynamics include Thomson et al. (2007) that uses a similar approach as of Abbott et al. (1997), and Tsodyks et al. (2000). The Tsodyks et al. (2000) proposed detail mathematical model that model the interplay

between recovered, active, and inactive states of synaptic resources to closely fit the experimental data of synaptic plasticity and depression. However, the model proposed in Abbott et al. (1997) is a very descriptive simple set of mathematical equations that can be used abstract the qualititative behaviour of the depression and facilitation dynamics to design a silicon area efficient synapse circuits.

6.2 A Simplified Model of Short-Term Dynamics

Abbortt et al. (1997) short-term dynamic synapse model (facilitating and depressing dynamics) has been simplified to provide a model that is implemented in hardware. In order to understand the circuits and their mathematical formulations better, these generic facilitation and depression parameters used in the circuit description are elaborated here. The pictorial representations of these parameters are given in Figure 6.1. In the circuit implementations, the synaptic weight change are considered instead of referring to the synaptic conductance as in the case of Abbott's model (It is the parameter that regulates the post-synaptic current injection). Hence, $\overline{g_s}$ and $\overline{g_s}P_sP_{ral}$ are resting weight (wr_d or wr_f) and the instantaneous weight (w) of the synapse, respectively. The amount of spike-induced facilitation ($\overline{g_s}P_s f_F(1-P_{rel})\delta(t-t_{pre})$) or depression ($\overline{g_s}P_s f_D(1-P_{rel})\delta(t-t_{pre})$) is qualititatively modelled with the weight dependent Δwf (degree of facilitation) and Δwd (degree of depression) respectively. Here, the P_s of the Abbott's model that generates the shape of the post-synaptic conduction is not considered as the post-synaptic current injection is implemented simply as a short current spike of a few nanoseconds of duration and the location of the synapse on the dendritic tree is not modelled (however, some morphological effect of the dendritic trees are considered in designing the neuron circuit presented in Chapter 4). Furthermore, rather than decaying the weight to the resting weight exponentially, a linearly decaying is used. The dynamic gain-control mechanism implemented on hardware consider only the facilitating and depressing effects in the similar range of magnitude and time as that of the computational model, and although they may be important, the finer details are not modelled to reduce the transistor count of the implemented circuit. The generic discrete mathematical equations of facilitating and depressing dynamics are given in equations 6.4 and 6.5 respectively.

Weight (w) of the facilitating synapse, wf evolves as follows,

$$w_f(t+\Delta t) = \begin{cases} \min\left\{w(t)+\Delta w_f, w_{f\max}\right\} & \text{if pre-neuron fires} \\ \max\left\{w(t)-\Delta w_{\alpha f}, wr_f\right\} & \text{otherwise} \end{cases} \qquad (6.4)$$

Where, Δt is the time step, *wrf* is the resting weight of the facilitating synapse, Δwf is the degree of facilitation, and $\Delta w\alpha f$ the step decay, providing recovery towards the resting weight, *wrf*; these can be set externally. *wfmax* is the maximum value of the facilitated weight.

Weight (*w*) of the depressing synapse, *wd* evolves as given below,

$$w_d(t+\Delta t) = \begin{cases} \max\left\{w(t)-\Delta w_d, w_{d\min}\right\} & \text{if pre-neuron fires} \\ \min\left\{w(t)+\Delta w_{\alpha d}, wr_d\right\} & \text{otherwise} \end{cases} \qquad (6.5)$$

Where, *wrd* is the resting weight of the depressing synapse, Δwd is the degree of depression, and $\Delta w\alpha d$ the step decay, aiming the resting weight, *wrd*; these can be set externally. *wdmin* is the minimum value of the depressed weight.

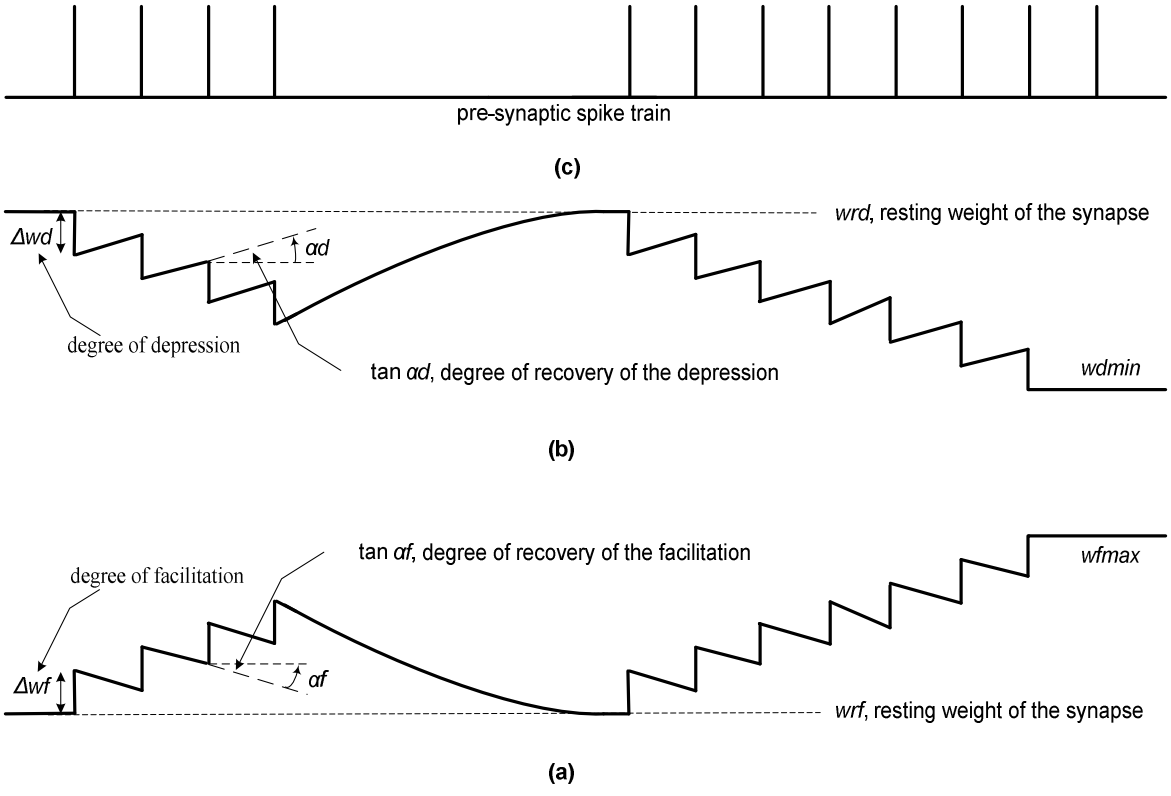


Figure 6.1 (a) Facilitation, (b) Depression dynamics of the synapse to the pre-synaptic spike train shown in (c).

6.3 Synapse Circuits and Their Operations

This section provides circuits of the proposed Excitatory Depressing Synapse (EDS) circuit, Inhibitory Facilitating Synapse (IFS) circuit, Inhibitory Depressing Synapse (IDS) circuit, and Excitatory Facilitating Synapse (EFS) circuits and their operation. In obtaining a different combination of excitatory or inhibitory, and facilitation or depression dynamics, same basic circuits and their complementary circuits are used with different source follower circuits. Therefore, the EDS circuit is presented in detail and other synapse circuits and their implementations are summarised.

Approximate mathematical equations of the EDS circuit are given in Appendix A.

6.3.1 Excitatory Depressing Synapse (EDS) – Circuit

The EDS circuit is shown in Figure 6.2. The circuit comprises of Excitatory Weight Depressing circuit and Excitatory Synaptic Current Generator circuit as shown in Figure 6.2 (a) and (b) respectively.

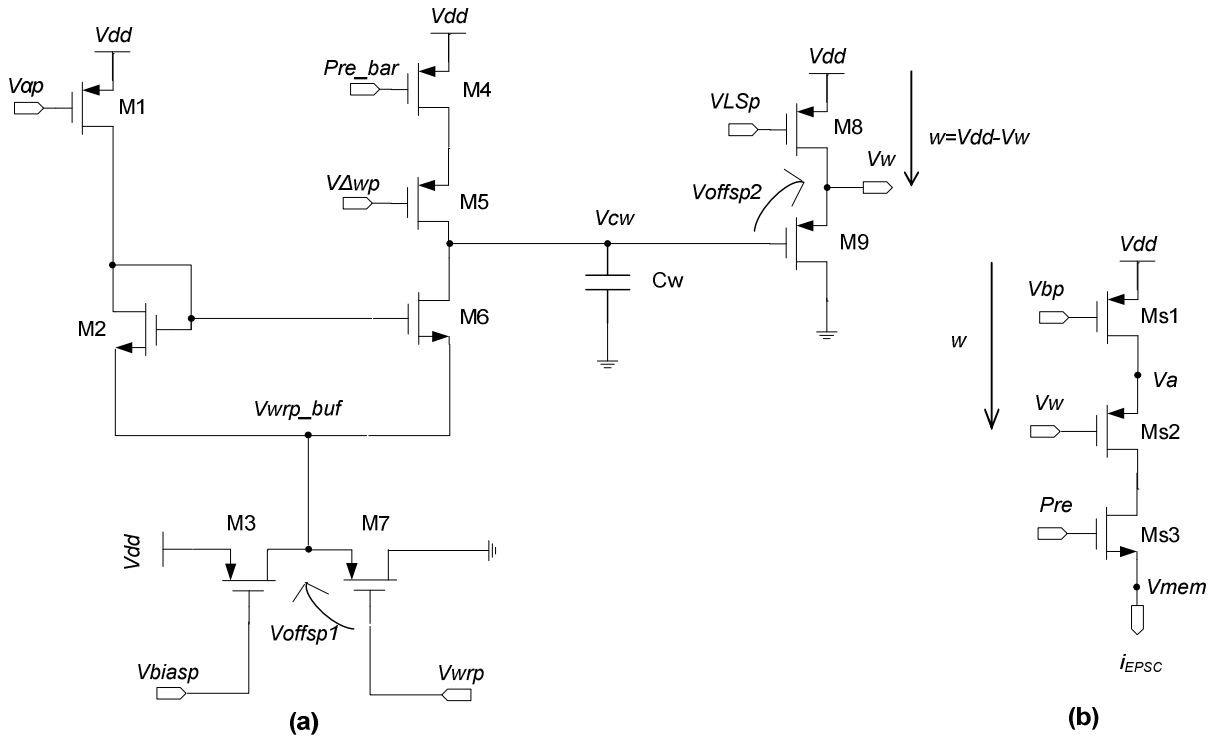


Figure 6.2 Excitatory Depressing Synapse circuit; (a) Excitatory Weight Depressing circuit, (b) Excitatory Synaptic Current Generator (EX-ISYN) circuit.

6.3.1.1 Operation of the EDS circuit

The circuit in Figure 6.2 (a) mimics the short-term depression dynamics of a synapse and produces depressing synaptic weight to the EX-ISYN circuit to generate the synaptic current, when a pre-synaptic spike (*Pre*) arrives. The depressing weight of the synapse, *w* (i.e. *wd* in equation 6.5), is represented by the voltage at the node *Vw* with reference to *Vdd*. It can have a value between the resting weight of the synapse (*wrd*), and the zero voltage (*wdmin*) depending on the short-term neural activity of the pre-synaptic neuron. The *Vw* follows the voltage across the capacitor Cw with an off-set (*Voffsp2*) as the source follower circuit (transistors M8-M9) buffers the voltage across the capacitor on the node *Vw*.

The pre-synaptic spike is signalled by a short pulse on the *Pre* input signal, and its inverted signal is *Pre-bar*. Once the pre-synaptic neuron fires, finite charge, Δq is added to the capacitor Cw through the transistors M4 and M5. Consequently, the weight *w* is depressed. The externally controllable gate voltage, $V\Delta wp$ controls the amount of Δq charge added to the capacitor (i.e. the degree of depressing of the synapse). Depending on the chosen operating range of $V\Delta wp$, the amount of charge added can also depend on the weight of the synapse (when the transistor M5 is in the linear region). The current mirror circuit (transistors M1, M2 and M6) continuously discharges the capacitor towards the voltage *Vwrp_buf*. The source follower circuit (transistors M3 and M7) buffers the voltage *Vwrp* onto the *Vwrp_buf* node. So that *Vwrp_buf = Vwrp+Voffsp1*. Hence, the resting weight (*wrd*) can be set by the externally controlled bias voltage *Vwrp* (as the *wrd = Vdd - wrp - Voffsp1- Voffsp2*). Due to the source follower circuit, the externally provided voltage node *wrd* draws negligible current. Therefore, the synapses that have the same resting weight can be easily provided to with a common reference voltage. The rate of discharge towards the *w* to resting weight (i.e. the degree of recovery of the depressing synapse) is controlled by the gate voltage of transistor M1, $V\alpha p$. By biasing the voltage $V\Delta wp$ to the supply voltage *Vdd*, the depressing synaptic dynamics can be switched off completely, and the synapse can be use as a simple weight dependent excitatory synapse.

When the pre-synaptic spike arrives at the gate of transistor Ms3 of the EX-ISYN circuit shown in Figure 6.2 (b), the circuit generates an excitatory post-synaptic current (i_{EPSC}) approximately proportional to the square of the synaptic weight. The i_{EPSC} current for a

given value of *w*, can be scaled using externally controllable voltage *Vbp*, as shown in Figure 6.17 in Appendix A. The *Vbp* could also be used to limit the maximum i_{EPSC}, depending on the operational region of the Ms1 transistor (higher tuning values of *Vbp* as seen in Figure 6.17(a) in Appendix A).

In summary, to configure the short-term dynamics of the excitatory depressing synapse *VΔwp* (controls the degree of depressing of the synapse), *Vwrp* (sets the resting weight of the synapse), *Vαp* (controls the degree of recovery of the depressing synapse) and *Vbp* (scale the value of the i_{EPSC} or set the maximum cut-off value for i_{EPSC}) can be set externally. The *Vbiasp* and *VLSp* are used to bias the source follower circuits.

6.3.2 Inhibitory Facilitating Synapse (IFS) – Circuit

The IFS circuit comprises of Inhibitory Weight Facilitating circuit and Inhibitory Synaptic Current Generator (IN-ISYN) circuit as shown in Figure 6.3 (a) and (b) respectively.

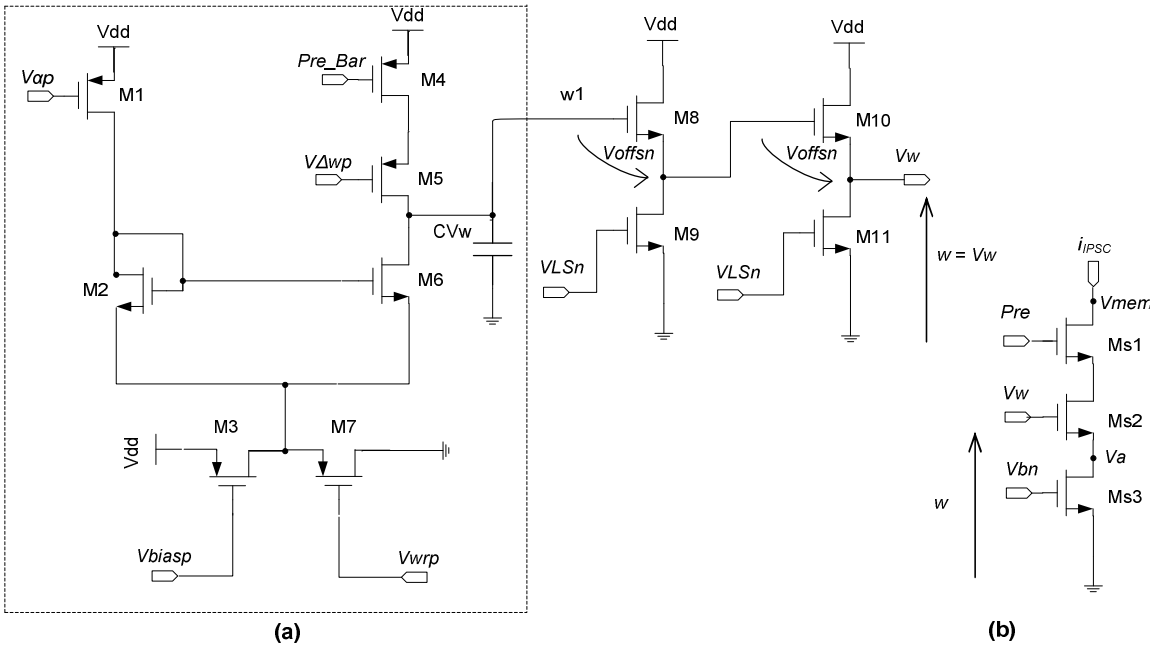


Figure 6.3 Inhibitory Facilitating Synapse circuit; (a) Inhibitory Weight Facilitating, (b) Inhibitory Synaptic Current Generator (IN-ISYN) circuit.

6.3.2.1 Operation of the IFS circuit

The circuit in the Figure 6.3 (a) is designed by adding two n-type source follower circuits as an output level-shifter to the output of the Figure 6.2 (a) instead of the p-type output source follower circuit (the dotted box in Figure 6.3(a); the circuit description is given in Section 6.3.1.1). However, the *Vw* output of the IF circuit is treated as a non-inverted synaptic weight (referenced to zero voltage) rather than inverted (referenced to *Vdd* voltage as in the EDS circuit). Consequently, depressing dynamics of Figure 6.2 (a) in the EDS circuit become facilitating dynamics. The two NMOS source follower circuits (M8-M9 and M10-M11) are used to shift the output voltage, *Vw,* to a lower voltage range to generate the required inhibitory current value. The shifted voltage is then given to the gate of transistor Ms2 of the IN-ISYN circuit to generate inhibitory post-synaptic current.

When the pre-synaptic spike arrives at the gate of transistor Ms1 of the IN-ISYN circuit shown in Figure 6.3 (b), the circuit generates an inhibitory post-synaptic current approximately proportional to the square of the synaptic weight, *w*. The external control biasing voltage *Vbn* can be used to scale the inhibitory post-synaptic current or to limit the maximum current value of the inhibitory post-synaptic current.

In summary, $V\Delta wp$ (controls the degree of facilitating of the synapse), *Vwrp* (sets the resting weight of the synapse), $V\alpha p$ (controls the degree of recovery of the facilitating synapse) and *Vbn* (scale the value of the i_{IPSC} or sets maximum cut-off value for i_{IPSC}) can be set externally. The *Vbiasp* and *VLSn* are used to bias the source follower circuits.

6.3.3 Inhibitory Depressing Synapse (IDS) – Circuit

The IDS circuit is shown in Figure 6.4, it comprises of Inhibitory Weight Depressing circuit and Inhibitory Synaptic Current Generator (IN-ISYN) circuit as shown in Figure 6.4 (a) and (b) respectively.

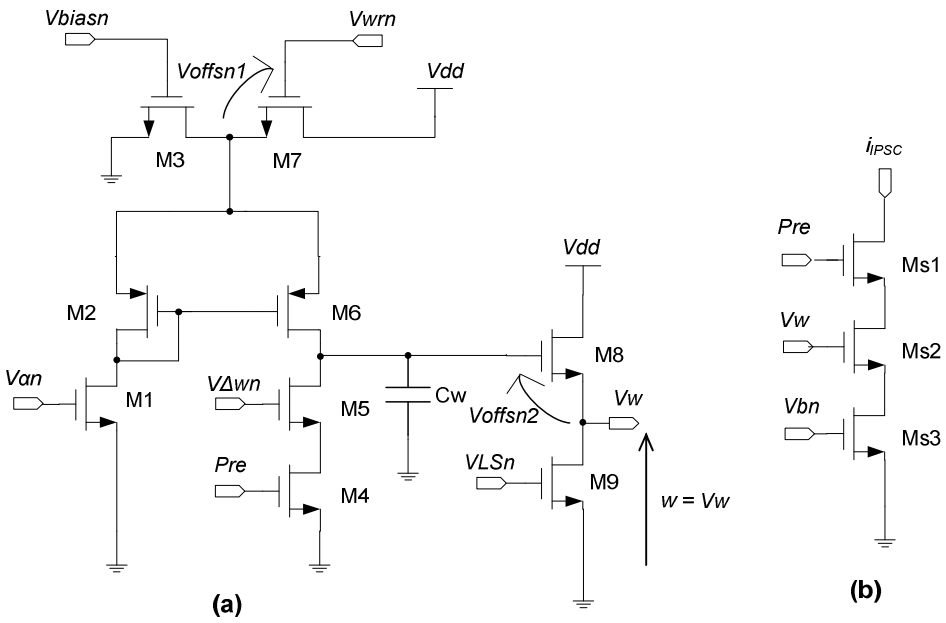


Figure 6.4 Inhibitory Depressing Synapse circuit; (a) Inhibitory Weight Depressing, (b) Inhibitory Synaptic Current Generator (IN-ISYN).

6.3.3.1 Operation of the IDS circuit

The circuit shown in Figure 6.4 is the complementary circuit of the EDS circuit. Hence the circuit shown in Figure 6.4 (a) mimics the approximated synaptic weight dynamics of a short-term depressing synapse, functionally same as the circuit shown in Figure 6.2 (a) of the EDS circuit. However, as these two circuits are complementary, the circuit shown in Figure 6.4 (a) produces a non-inverted synaptic weight output in contrast to an inverted in circuit in Figure 6.2 (a). This non-inverted output from ID circuit is provided to the IN-ISYN circuit to generate an inhibitory post-synaptic current. Usage of complementary circuit to generate a non-inverted synaptic weight minimises the power and the number of transistors used in the inhibitory depressing synapse than using the circuit shown in Figure 6.2 (a) with an analogue inverter circuit.

The IN-ISYN circuits operation is given in IFS Section 6.3.2.1

Similar to EDS circuit the short-term dynamics of the inhibitory depressing synapse are configured using control voltages *VΔw*n (controls the degree of depressing of the synapse), *Vαn* (controls the degree of recovery of the depressing synapse), *Vwrn* (sets the resting weight of the synapse), and *Vbn* (scales the value of the i_{IPSC} or set the maximum cut-off value for i_{IPSC}) can be set externally. The *Vbiasn* is a fixed biasing

voltage used for the source follower (M3 and M7 transistor) circuit. The *VLSn* is used to bias the source follower (M8 and M9 transistor) circuit.

6.3.3.2 Additional IDS circuit implementation – Somatic IDS

In the circuit implementation discussed in Chapter 9 additional IDS circuit called "somatic IDS" has been implemented as shown in Figure 6.5. This synapse circuit is used to provide a high depressed weight to the same bias voltage value that sets the resting weight of the synapse and consequently the higher inhibitory post-synaptic current. This can be considered to model an inhibitory depressing synapse that connects directly to cell body (soma) since such synapses produce higher inhibition to the same input spike train than distal inhibitory depressing synapses. In using this circuit arrangement, the same tuning parameters (including the bias voltage that set the resting weight value) used for IDS circuit described above can be used, so that the there is no requirement for extra external bias voltages.

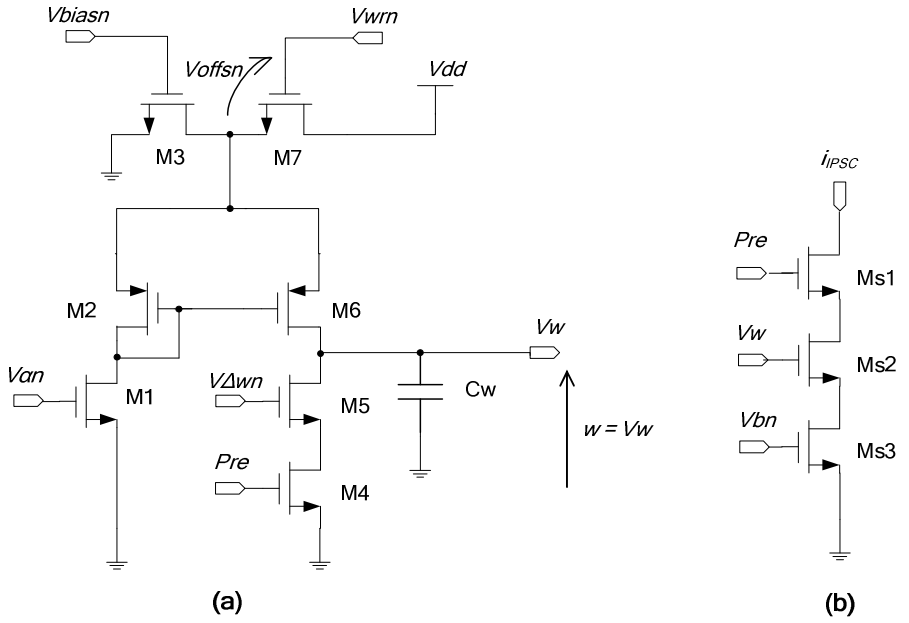


Figure 6.5 Somatic Inhibitory Depressing Synapse circuit; (a) Somatic Inhibitory Weight Depressing, (b) Inhibitory Synaptic Current Generator (IN-ISYN).

The Somatic IDS circuit is same as the circuit shown in Figure 6.4 (a) without output source follower circuit (M8-M9). As there is no level-shifting to reduce the weight of

the synapse, *w*, Somatic-IDS has higher weight and hence provides high inhibitory post-synaptic current when a pre-synaptic spike fires. Somatic IDS equations for synaptic weights and inhibitory post-synaptic current equations are the same as the equations for IDS except that the *Voffsn2* is not subtracted from *wrd* and *Vw* (*wd*) as in the IDS equations.

6.3.4 Excitatory Facilitating Synapse (EFS) – Circuit

The EFS circuit comprises of Excitatory Weight Facilitating circuit and Excitatory Synaptic Current Generator (EX-ISYN) circuit as shown in Figure 6.6 (a) and (b) respectively.

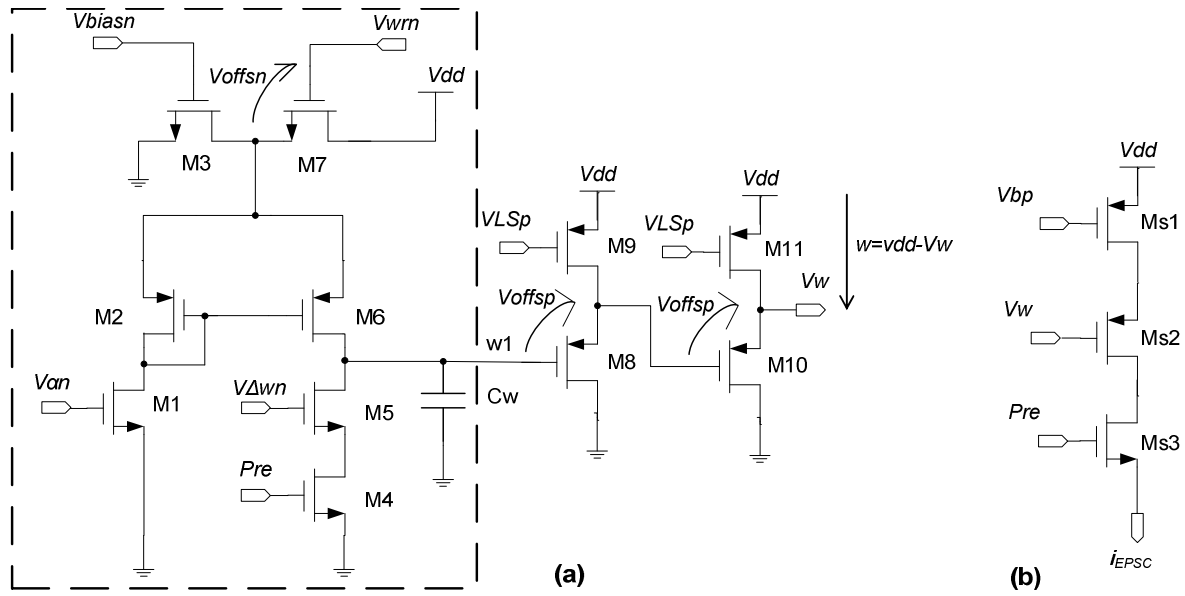


Figure 6.6 Excitatory Facilitating Synapse circuit; (a) Excitatory Weight Facilitating circuit, (b) Excitatory Synaptic Current Generator (EX-ISYN) circuit.

6.3.4.1 Operation of the EFS circuit

The circuit shown in Figure 6.6 (a) essentially the same as the circuit shown in Figure 6.4 (a) circuit, however it uses two pMOSFET source follower circuits (M8-M11 and M10-M11) as an output level-shifter instead of having single nMOSFET source follower.

Similar to IDS circuit the short-term dynamics of the EFS are configured using control voltages: *VΔw*n (controls the degree of facilitation of the synapse), *Vαn* (controls the degree of recovery of the synapse), *Vwrn* (sets the resting weight of the synapse), and

Vbp (scale the value of the i_{EPSC} or set the maximum cut-off value for i_{EPSC}) can be set externally. The *Vbiasn* and *VLSp* are used to bias the source follower circuits.

6.4 Simulation Results of the Synapse Circuits

This section provides simulation results of the proposed Excitatory Depressing Synapse (EDS) circuit, Inhibitory Facilitating Synapse (IFS) circuit, Inhibitory Depressing Synapse (IDS) circuit, and Excitatory Facilitating Synapse (EFS) circuits. These synapses are simulated in standard 0.35 µm CMOS technology.

6.4.1 Excitatory Depressing Synapse - Simulation Results

The simulation results of Excitatory Weight Depressing circuit are given in Figure 6.7 and 6.8.

In order to demonstrate the weight depression, the behaviour of the circuit for different values of the main tuning parameters that configure the short-term dynamics of the excitatory depressing synapse, is simulated. An input pre-synaptic spikes (*Pre*) signal consisting of a 4 ms burst of spikes at 10 kHz rate, followed by no spike activity is used.

Figure 6.7 (a) shows the variation of *Vw* (depressed weight *wd* = *Vdd* - *Vw*) for seven values of $V\Delta wp$ starting from 2.3 V (higher degree of depression) to 3 V (low degree of depression). It is observed from the seven graphs that the variable degree of depression can be obtained by controlling the voltage $V\Delta wp$ appropriately. The mid range of the parameter are plotted here however, it is also possible configure the circuit to fully depress for a single pre-synaptic spike if lower $V\Delta wp$ value is used. Similarly, Figure 6.7 (b) shows the variation of *Vw* for different values of $V\alpha p$ starting from 2.74 to 2.78 V. It is observed that higher to lower linear degree of recovery can be obtained by controlling the narrow range of voltage $V\alpha p$ appropriately. Figure 6.7 (c) shows the variation of *Vw* for six mid range values of *Vwrp* starting from 1.6 V (corresponds to lower resting weight) to 0.6 V (corresponds to higher resting weight). Hence, it is observed from the graphs that higher to lower resting weight can be obtained by controlling the voltage *Vrwp* appropriately.

The circuit has been simulated to observe the weight depression of the EDS circuit resulting from different pre-synaptic input frequencies. The *Vw* for pre-synaptic input spike trains with different inter-spike intervals (ISI) starting from 10 µs to 300 µs are shown in Figure 6.8. It is seen that the lower ISI (i.e. high frequency pre-synaptic input) produces higher and quicker depression than the higher ISI.

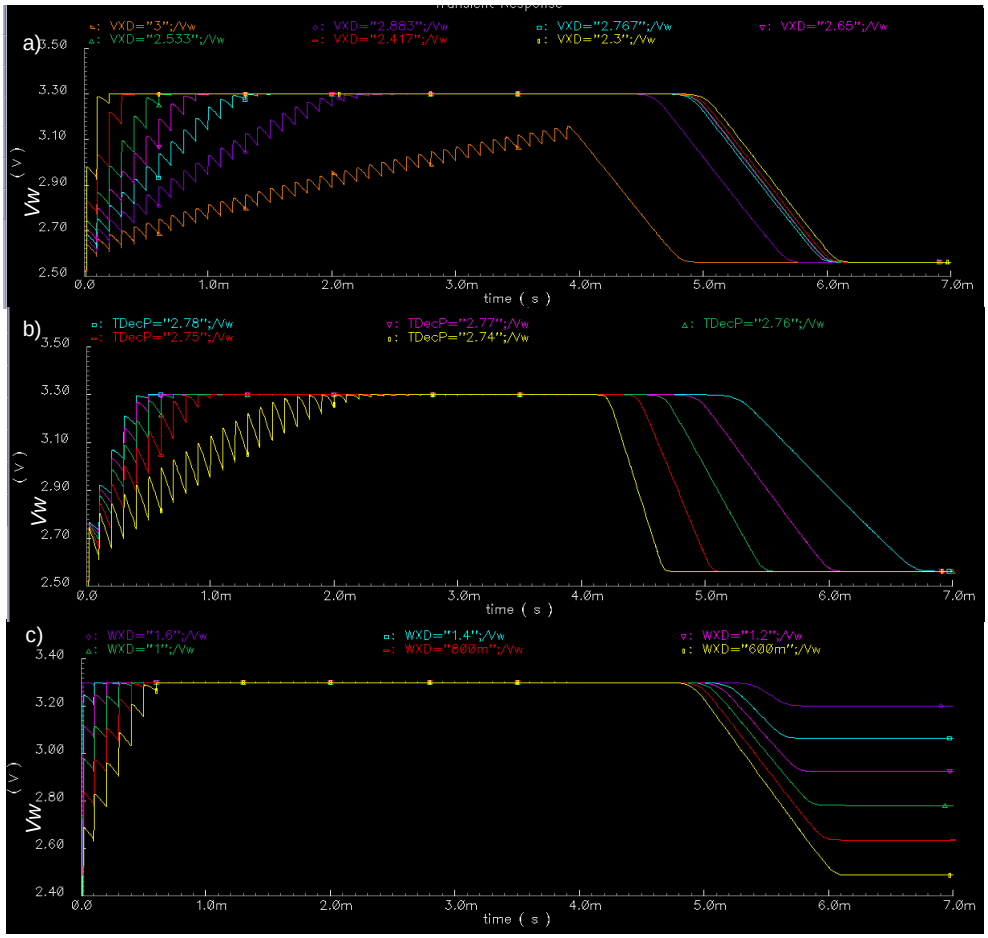


Figure 6.7 Simulated EDS circuit dynamics; Variation of *Vw* to a different values of control voltages a) degree of depression *VΔwp* (labelled as VXD), b) degree of recovery *Vαp* (labelled as TDecP), and c) resting weight of the synapse *Vwrp* (labelled as WXD) the pre-synaptic input is a 4 ms burst of 10 kHz spikes followed by a silent period; instantaneous synaptic weight is *wd=Vdd-Vw*, *Vdd*=3.3 V.

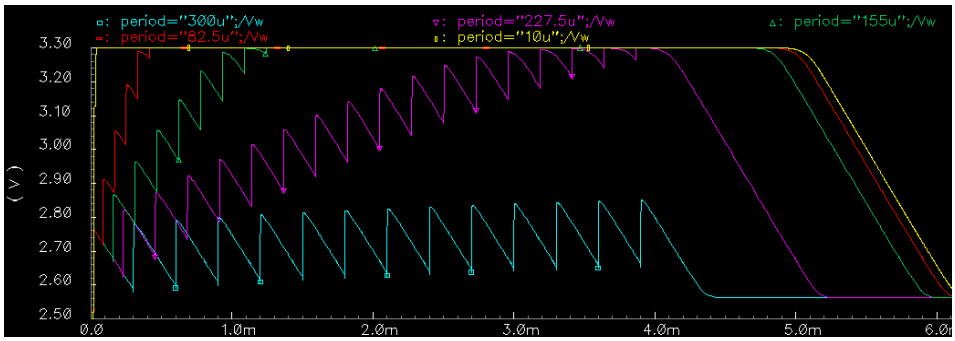


Figure 6.8 The response (*Vw*) of EDS circuit to 4 ms pre-synaptic input spike train with different inter-spike intervals (period) followed by a 4 ms of silent period.

6.4.2 Inhibitory Facilitating Synapse - Simulation Results

The simulation results of the Inhibitory Weight Facilitating circuit are shown in Figure 6.9 to 6.10.

Similar to EDS circuit, in order to demonstrate the effect on the faciliting weight (*w*), different values of the main tuning parameters that configure the short-term dynamics of the IFS $V\Delta wp$ (controls the degree of facilitation), $V\alpha p$ (controls degree of recovery), and *Vwrp* (sets the resting weight of the synapse) are simulated. An input pre-synaptic spikes (*Pre*) signal consisting of a 4 ms burst of spikes at 10 kHz rate, followed by no spike activity is used.

Figure 6.9 (a) shows the variation of *Vw* (*w*) for eight values of $V\Delta wp$ starting from 2.3 V (lower degree of facilitation) to 3 V (higher degree of facilitation). It is also possible configure the circuit to fully facilitate from a couple of pre-synaptic spikes if a higher $V\Delta wp$ value is used. Similarly, Figure 6.9 (b) shows the variation of *Vw* for different values of $V\alpha p$ starting from 2.74 V (higher degree of recovery) to 2.77 V (low degree of recovery). Figure 6.9 (c) shows the variation of *Vw* for six mid range values of *Vwrp* starting from 1.6 V (corresponds to higher resting weight) to 0.6 V (corresponds to lower resting weight).

The circuit has been simulated to observe the facilitation effect of the IFS circuit to different pre-synaptic input frequencies. The *Vw* for pre-synaptic input spike trains with different inter-spike intervals (ISI) starting from 10 µs to 300 µs are shown in Figure 6.10 It is seen that the lower ISI (high frequency pre-synaptic input) produces higher and quicker depression than for the higher ISI.

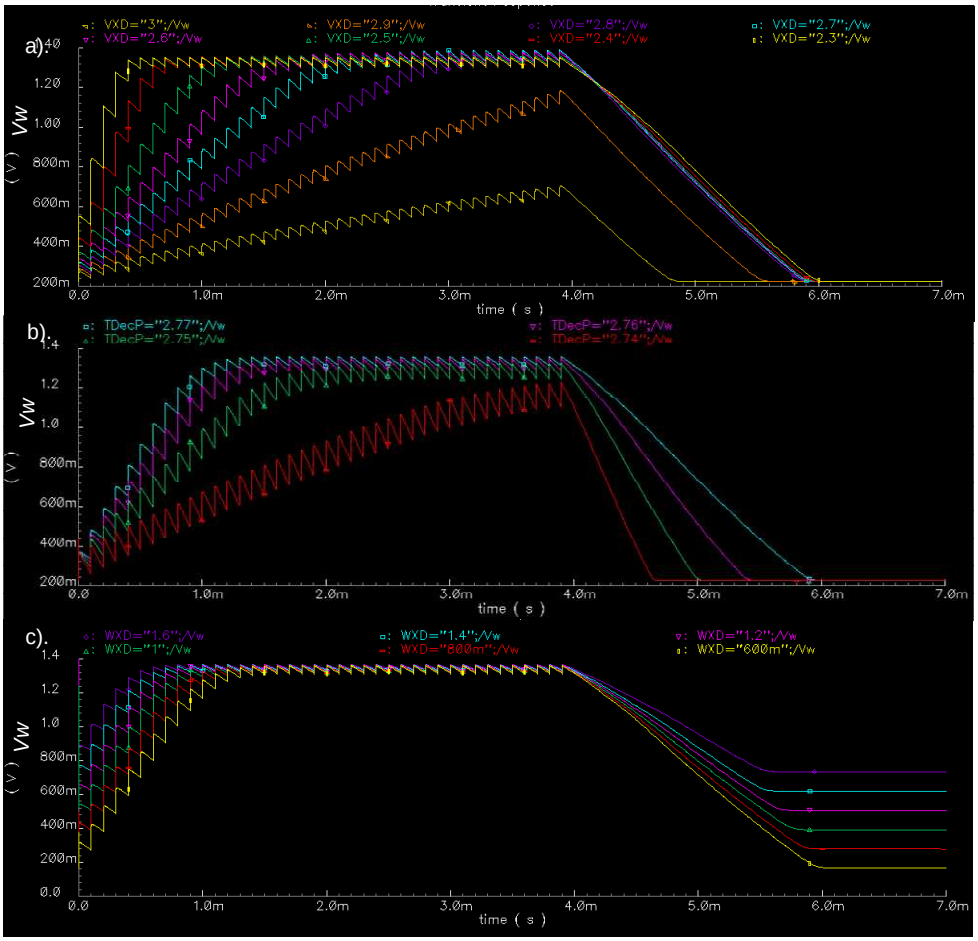


Figure 6.9 IFS dynamics; Variation of weight of the synapse (*Vw*) to a different values of control voltages a) degree of facilitation (*VΔwp*) (labelled as VXD) , b) degree of recovery (*Vαp*) (labelled as TDecP), and c) resting weight (*Vwrp*) (labelled as WXD) when a pre-synaptic input of 4 ms of 10 kHz spikes followed by a silent period is provided.

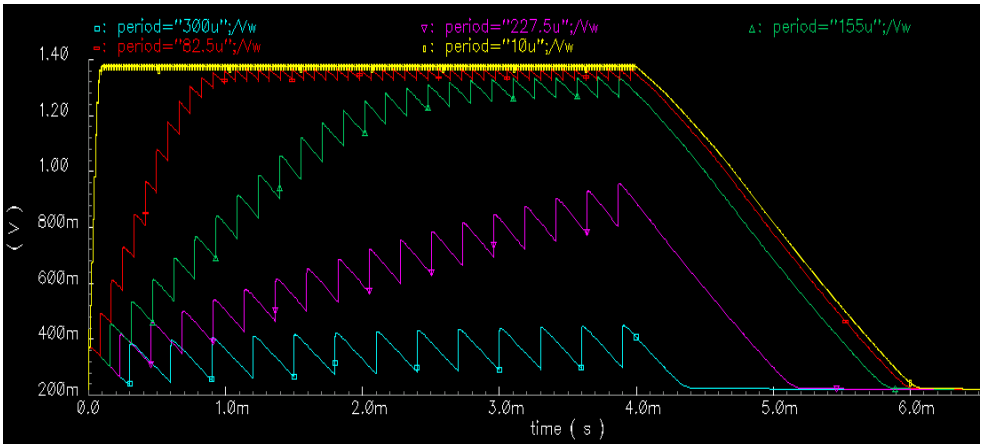


Figure 6.10 Responses *Vw* of IFS circuit to 4 ms pre-synaptic input spike train with different inter-spike intervals (*period*) followed by a 4 ms of silent period.

6.4.3 Inhibitory Depressing Synapse - Simulation Results

The IDS circuit has been simulated and the simulation results are shown in Figure 6.11 to 6.13.

The effect on the depressed weight, *w* for different values of the main tuning parameters that configure the short-term dynamics of the inhibitory depressing synapse are simulated and the results are shown in Figure 6.11.

Figure 6.11 (a) shows the variation of synaptic weight (*Vw*) for six values of mid-range $V\Delta wn$ voltages starting from 1 V (corresponds to higher degree of depression) to 0.53 V (corresponds to low degree of depression).

It is observed from the six graphs that higher to lower degree of depression can be obtained by controlling the voltage $V\Delta wn$ appropriately. However, it is also possible configure the circuit to fully depress for a single pre-synaptic spikes if higher $V\Delta wn$ value is used. Figure 6.11 (b) shows the variation of synaptic weight for different values of $V\alpha n$ starting from 0.415 V (corresponds to low degree of recovery) to 0.44 V (corresponds to high degree of recovery). Figure 6.11 (c) shows the variation of synaptic weight for five mid range values of *Vwrn* starting from 2.1 V (corresponds to lower resting weight) to 3.2 V (corresponds to higher resting weight).

Figure 6.12 shows IDS's inhibition with depressing effect on the neural activity of a neuron when the IDS synaptic current plus a 0.1 µA continuous synaptic current stimulus are given to a post-synaptic neuron which is configured to RS type. The pre-synaptic input to the IDS is a repetition of the stimulus that has a 10 kHz spike train for 4 ms followed by a silent period of 4 ms. It is seen that the effect of continuous input spike train on the inhibitory depressing synapse could become less significant on the post-synaptic activities thereby other synaptic input sources responses are not silence due to the continues spiking activity of the inhibitory depressing synapse (In a network, this leads to dynamic gain-control mechanism depending on the input pattern of the synapses).

The circuit has been simulated to observe the weight depression of the IDS circuit to different pre-synaptic input frequencies. The depressing synaptic weight change for pre-synaptic input spike trains with different inter-spike intervals (ISI) starting from 10 µs

to 300 µs are shown in Figure 6.13. It is seen that the lower ISI (high frequency pre-synaptic input) produces higher and quicker depression than for the higher ISI.

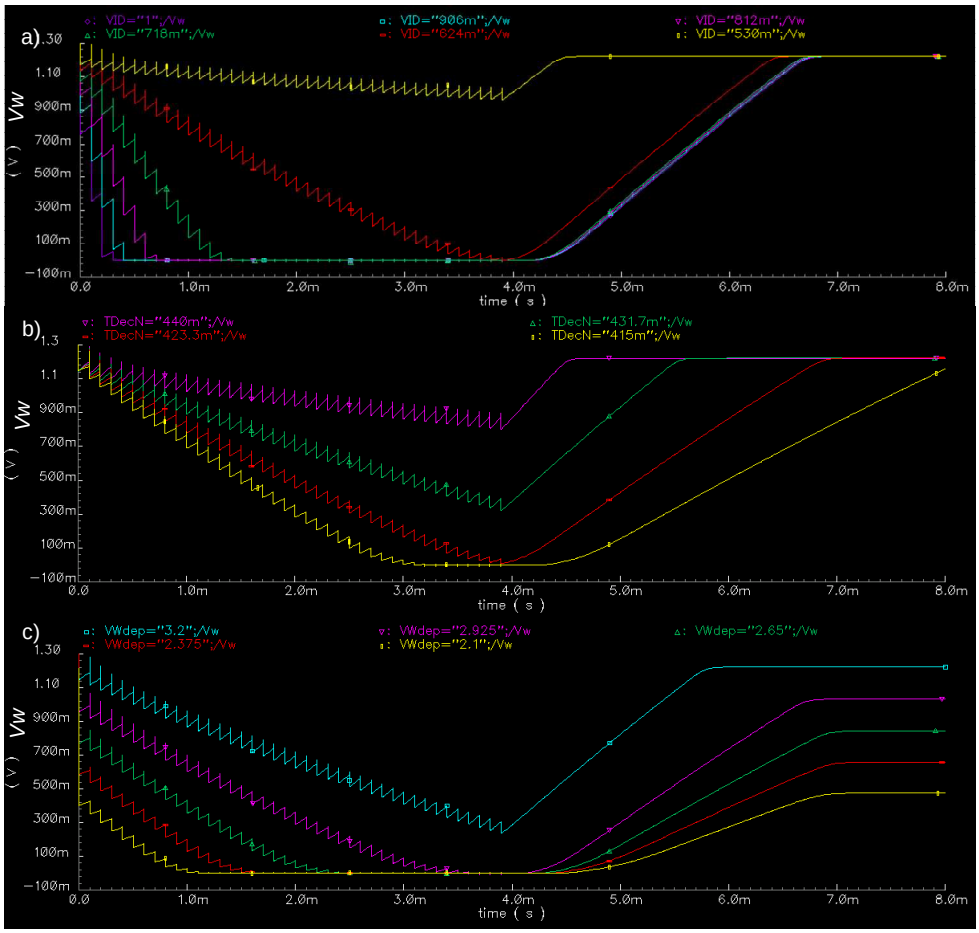


Figure 6.11 IDS dynamics; Variation of synaptic strength, *w (Vw)* to a different values of a) degree of depression control voltage, *VΔwn* (labelled as VID) b) degree of recovery control voltage, *Vαn* (labelled as TDecN), and c) *Vwrn* (labelled as VWdep) resting voltage control voltage when a pre-synaptic input of 4 ms of 10 kHz spike train followed by a 4 ms of silent period is provided.

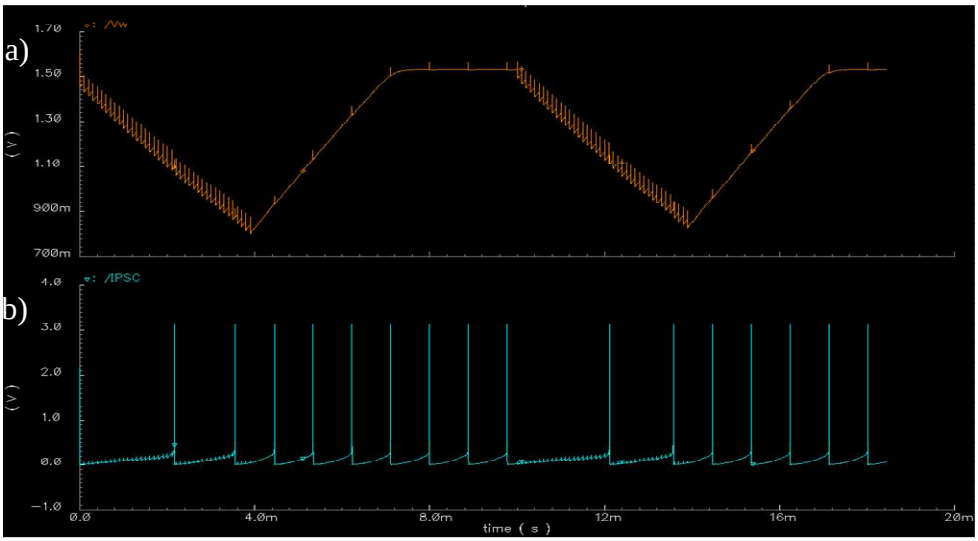


Figure 6.12 IDS synapse inhibiting RS neural activity; a) Synaptic weight, *w* generated to a pre-synaptic input of a repetition of the stimulus that has a 10 kHz spike train for 4 ms followed by a silent period of 4 ms, b). Neuron's RS spike activity has depressing inhibition effect due to depressing synapse, the neuron is configured to RS and an extra stimulus of 0.1 µA of constant pre-synaptic current is given.

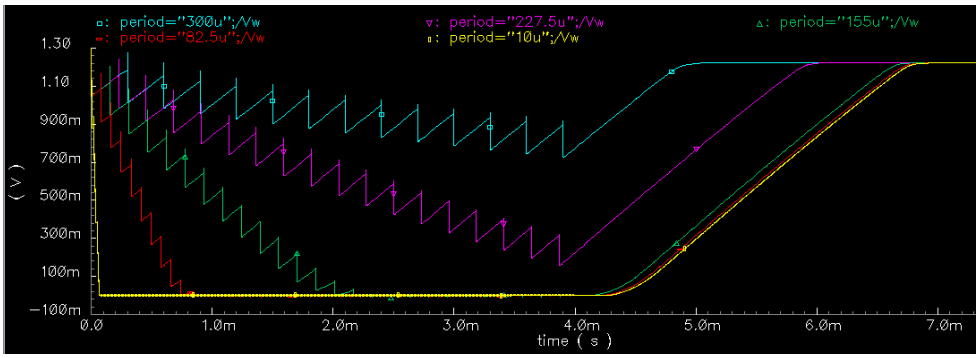


Figure 6.13 Distal-IDS's *w* responses to 4 ms pre-synaptic input spike train with different inter-spike intervals (*period*) followed by a 4 ms of silent period.

6.4.4 Excitatory Facilitating Synapse - Simulation Results

EFS circuits simulation results are shown in Figure 6.14 to 6.16.

The effects on the facilitating weight, *w* (*Vdd-Vw*) for different values of the main tuning parameters that configure the short-term dynamics of the excitatory facilitating synapse are simulated and the results are shown in Figure 6.17.

Figure 6.14 (a) shows the variation of *Vw* for six values of mid-range *VΔwn* voltages starting from 1 V (corresponds to higher degree of facilitation) to 0.5 V (corresponds to

low degree of facilitation). However, it is also possible configure the circuit to fully facilitate from a single pre-synaptic spikes if higher $V\Delta wn$ value is used. Figure 6.14 (b) shows the variation of *Vw* for different values of $V\alpha n$ starting from 0.42 V (corresponds to low degree of recovery) to 0.45 V (corresponds to high degree of recovery). Figure 6.14 (c) shows the variation of *Vw* for five mid range values of *Vwrn* starting from 1.8 V (corresponds to higher resting weight) to 3 V (corresponds to lower resting weight).

Figure 6.15 shows EFS's excitation with facilitation effect on the neural activity of a neuron when the EFS's synaptic current is given to a post-synaptic neuron which is configured to RS type. The pre-synaptic input to the EFS is a repetition of the stimulus that has a 10 kHz spike train for 4 ms followed by a silent period of 4 ms. It is seen that the effect of continues input spike train on the excitatory facilitating synapse could become more significant on the post-synaptic activities thereby other synaptic input sources responses could be ignored due to continues spiking activity of the inhibitory facilitating synapse.

The circuit has been simulated to observe the weight facilitation of the EFS circuit to different pre-synaptic input frequencies. The facilitating synaptic weight change for pre-synaptic input spike trains with different inter-spike intervals (ISI) starting from 10 µs to 300 µs are shown in Figure 6.16. It is seen that the lower ISI (high frequency pre-synaptic input) demonstrate higher and quicker facilitation than for higher ISI.

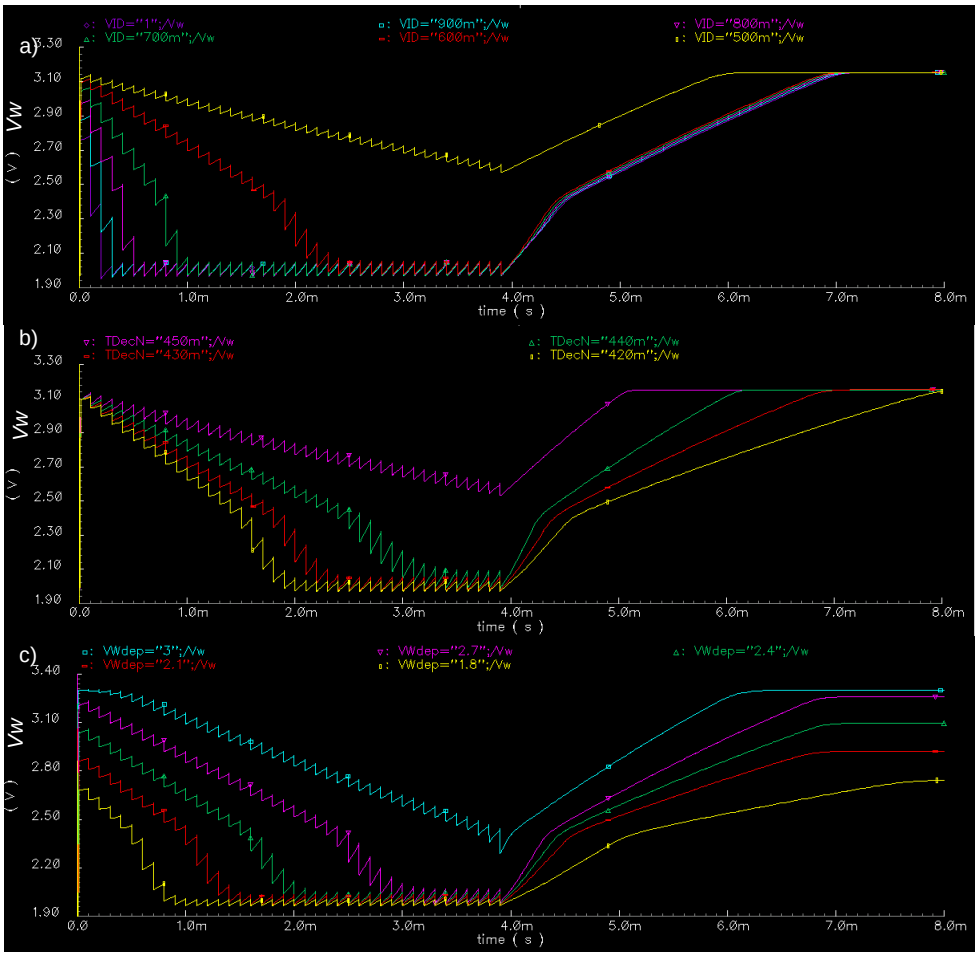


Figure 6.14 EFS dynamics; Variation , *Vw* (=*Vdd-Vw*) to a different values of a) degree of facilitating control voltage, *VΔwn* (labelled as VID) b) degree of recovery control voltage, *Vαn* (labelled as TDecN), and c) *Vwrn* (labelled as VWdep) resting voltage control voltage when a pre-synaptic input of 4 ms of 10 kHz spikes followed by a 4 ms of silent period is provided.

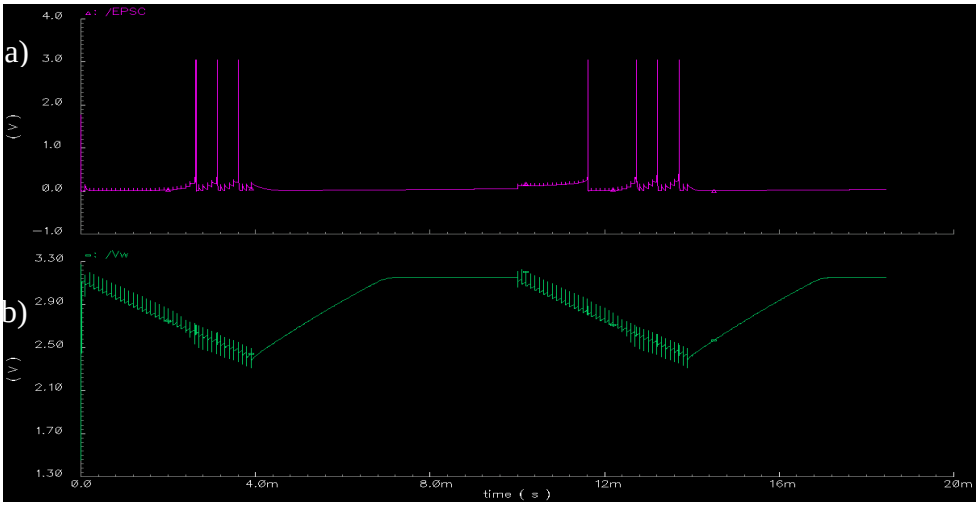


Figure 6.15 EFS synapses post-synaptic neural activity; a). Neuron's spike activity generated from a facilitating synaptic strength, *w* shown in b); b) Synaptic strength, *w* generated to a pre-synaptic input of a repetition of the stimulus that has a 10 kHz spike train for 4 ms followed by a silent period of 4 ms.

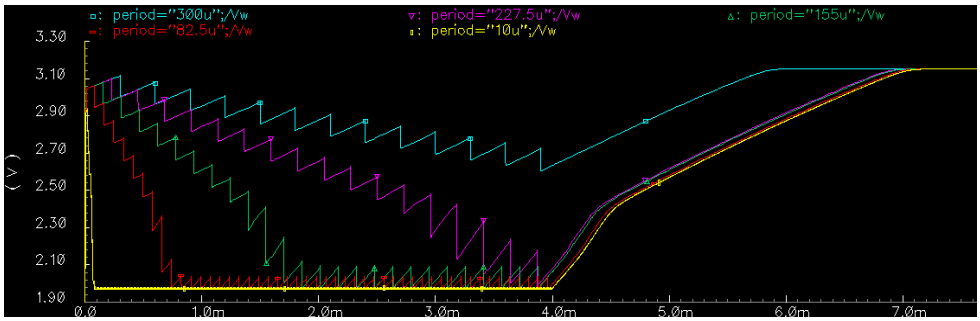


Figure 6.16 EFS's *Vw* responses to 4 ms pre-synaptic input spikes with different inter-spike intervals (*period*) followed by a 4 ms of silent period.

6.5 Discussion and Conclusion

Compact implementations of the Excitatory Depression, Inhibitory Facilitating, Inhibitory Depression and Excitatory Facilitating synapse circuits are proposed in this chapter. These synapse circuits have been fabricated in the Cortical Neural Layer (CNL) chip, and the detailed implementation description can be found in Chapter 9. The mathematical models of these synapses are formulated to be used in the simple approximated mathematical model of the CNL chip discussed in Chapter 9.

The presented circuits demonstrate depressing and facilitating dynamics qualititatively similar to the computational model proposed by Abbott et al. (1997), while making approximation to achieve compact circuit implementations.

Each synapse circuit has four tunable parameters and two or three biasing voltages. Tunable parameters include control voltages of the degree of facilitation or depression, the degree of decay, the resting weight and the post-synaptic current scaling/limiting voltage (*Vbn* or *Vbp*). The simulation results of facilitation and depression effect for these tunable variables have been presented. The degree of depression or facilitation of a synapse can be controlled using the bias voltage $V\Delta wp$ or $V\Delta wn$. However, this can be weight dependent depending on the operational range of the weight (If the transistor M5 of the respective circuit is in the linear region of operation).

In different biological synapses short-term dynamic effect have been observed in time scale ranges from 100 ms to 1 s (Morrison et al.; 2008) possibly due to the exponential decay. However, it is observed that the linear degree of recovery range is in the slower end of the biological synapse's recovery time range, and it can only be controlled within

a narrow range (≈ 0 to 30 mV) of the control voltage. Therefore, these synapses represent sub-set of the synapses from the highly heterogeneous synapses.

Although the degree of depression or facilitation and the decay can be tuned individually for different synapses, in a practical implementation, several same types of synapse circuits in a group might share same tuning parameter. Hence, the variability, mismatch and the supply voltage drop could cause the range of curves to be available within the same type of synapse. Biological neural responses are also highly heterogeneous and have considerable variability across the same type of neural elements, and the network dynamics are possibly exploiting these properties.

It is also possible to switch off depressing and/or facilitating dynamics of synapses completely (by biasing the voltage $V\Delta wp$ to the supply voltage *Vdd* or voltage $V\Delta wn$ to the analogue ground). If depressing/facilitating dynamic is switched off the synapse becomes a generic weight dependent excitatory or inhibitory synapse.

The excitatory and inhibitory synapses are designed with simple three transistor circuits (Synaptic Current Generator circuits IN-ISYN or EX-ISYN) that source or sink weight dependent current to or from the membrane of the neuron, depending on the synapse type. The amount of current source or sink can be approximately proportional to the square of the weight of the synapse. However, this simple three transistor circuit can operate such that post-synaptic current is proportional to the post-synaptic neuron's membrane voltage and the synaptic weight, if the weight connected transistor of the Synaptic Current Generator circuits is in the linear range of operation.

The rise- and fall- time of the post-synaptic potentials are not modelled by these circuits as the width of the current pulse is only a few nanoseconds (shaping of the corresponding post synaptic potential pulse adds extra circuitry. The tuning of the rise- and fall-time has a negligible effect on membrane integration as the time scales are so small). Further, effect of the rise- and fall-time can be neglected as the dendritic location of the synapse is not considered in modelling the neural circuits. However, the dendritic delays can be incorporated by introducing a delay to the pre-synaptic spike during the off-chip spike routing. Neuron circuits designed in Chapter 4 account for some dendritic morphological effects on the spike response, although the full non-linear filtering due to exact location of the synapse has not been considered. The dendritic dynamics also could be introduced by modelling the detailed dendritic compartment model (Elias et

al., 1995; Rasche et al., 2001); however, this will consume larger silicon area and hence limit the size of the network considerably.

CHAPTER 7 : CORTICAL NEURON CHIP

This chapter presents a Cortical Neuron Chip that contains the Accelerated-Time Cortical Neuron circuits (Chapter 4). The purpose of the chip was to experimentally verify the spiking behaviour of a single cell. The neuron circuit has been used in other ICs presented in this thesis.

Neuron is a key element in neural processing, and cortical network consists of many types of neurons. These neuron types exhibit distinct nonlinear neural responses to the same set of input stimulus. Therefore, having different types of neurons is an important aspect in neural processing. In implementing a VLSI neural network incorporating the diverse neuron responses similar to the biological neuron responses are of the essence to produce brain like computation.

The initial sections of this chapter present an overview of the chip and the test setup. The experimental results presented at the end of the chapter confirm that the neuron circuit is capable of generating many types of the cortical neuron behaviour, with diversity similar to that of biological neuron cells. Some of the experimental results presented in this chapter have been published in the Journal of Neural Networks (Wijekoon et al., 2008b), and the Proceedings of the IEEE International Symposium on Circuits and Systems (Wijekoon et al., 2008a).

7.1 Chip Overview

A prototype test integrated circuit, the Cortical Neuron chip that contains 202 neuron cells, with varied circuit parameters (transistor sizes and capacitances) was fabricated in a 0.35 µm CMOS technology. These neuron circuits were used to obtain the best combinations of neuron circuit parameters that are capable of reproducing most of the firing patterns of neurons using two tuning parameters (see Chapter 4 for more details of the neuron circuit). The size of the cell that reproduces most of the firing patterns is 40 µm by 70 µm. The size of the chip is approximately 3 mm by 2 mm, and it has 84 pins.

A photograph showing the chip layout, as well as individual cells is shown in Figure 7.1. The cells are individually accessible and do not form any network. In addition to the neuron cells, the chip contains multiplexers, buffers and simple synaptic circuitry to generate excitatory and inhibitory postsynaptic currents. The different

neurons are provided with three different types of output buffers to feed the membrane potential signal to the output pads, these types include single stage buffering with two nMOSFETs, double stage buffering with two nMOSFETs and operational amplifier (OpAmp) buffering. The circuit also contains a multiplexing unit that selects one neuron at a time. Some cells are designed with an additional external membrane potential resetting circuit using a single transistor. More circuit design details of this chip are presented in Wijekoon (2007).

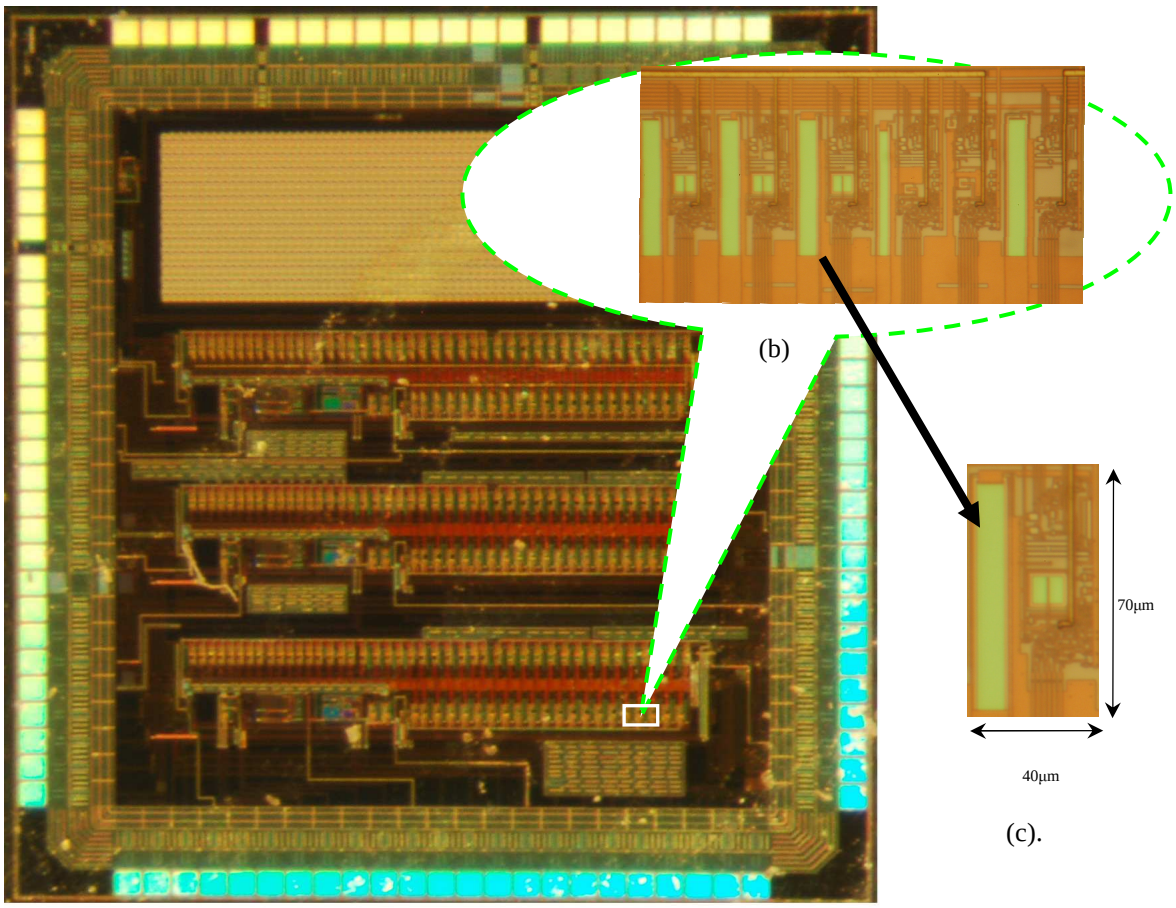


Figure 7.1 Photograph of the fabricated device: (a) chip with 202 neurons having different circuit parameters; (b) six different neuron cells; (c) a single neuron including an output buffer and control circuit.

7.2 Test Setup

The test setup used to record the experimental results presented in Section 7.3 is shown in Figure 7.2. The test setup includes the chip, the Address Generator Circuit, the

Programmable Digital Pulse Generator, the Programmable Voltage Supplies, and an Oscilloscope. The synaptic input is supplied using the Digital Pulse Generator, and an internal circuit converts this pulse to a synaptic current. The spike rate of the pre-synaptic signal can be programmed on the Digital Pulse Generator. The biasing parameters are set using Programmable Voltage Supplies. The neuron cells are individually accessible using the test address and do not form any network. The test address is generated using the Address Generator circuit that includes seven digital switches to provide the seven bits address manually. The spike output of the selected neuron can be observed using a digitising oscilloscope. The results presented in the next section are recorded from an on-chip OpAmp buffered output of a neuron.

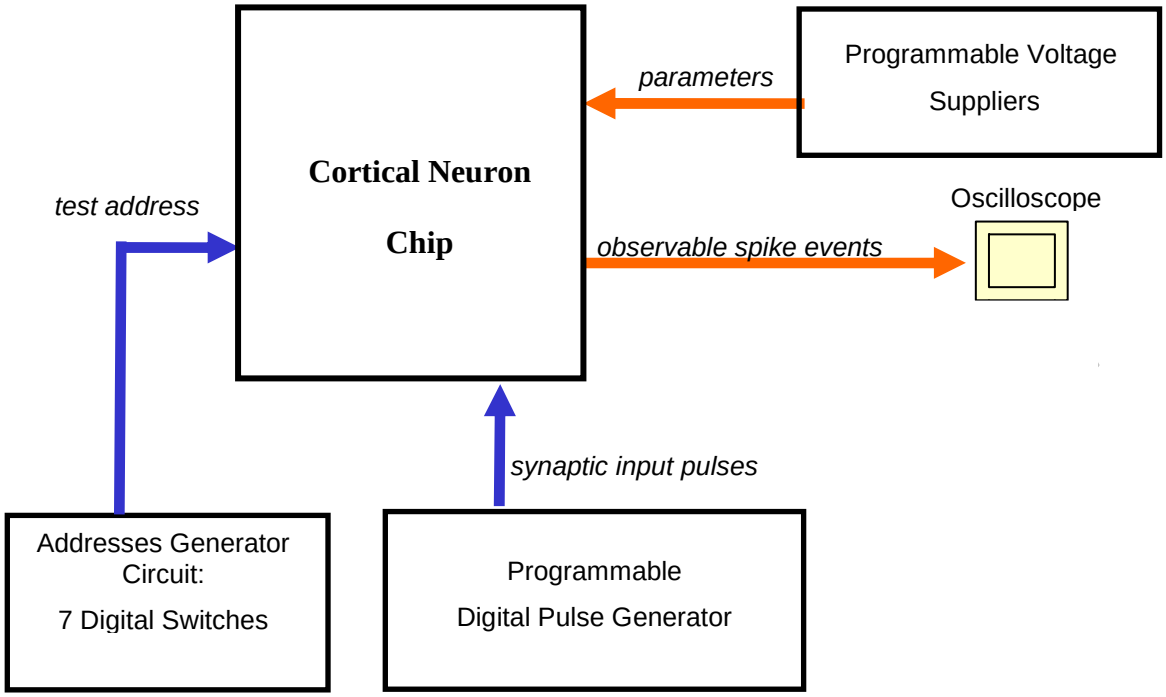


Figure 7.2 Test setup of the Cortical Neuron circuit.

7.3 Experimental Results

The experimental results presented in this section are recorded from a single Cortical Neuron circuit that is capable of reproducing most of the firing patterns of neurons, using two tuning parameters (*Vc* and *Vd*). The Cortical Neuron circuit is shown in

Figure 4.2 of Chapter 4. The transistor sizes, capacitances of the circuit and biasing voltages used to obtain the results are: (W/L)M1= (2.3/1), (W/L)M2= (2.3/1), (W/L)M3= (2.3/1), (W/L)M4= (1.3/22), (W/L)M5= (5.3/1), (W/L)M6= (1.3/18), (W/L)M7 = (1.3/14), (W/L)M8= (1.3/1), C_v=0.1 pF, C_u=1pF, *Vth* =1.70 V, *Vdd* = 3.3 V, and *Vbias* = 0.6 V. Where (W/L)Mi is the Width to Length ratio of the transistor Mi and lengths are in µm.

Different responses of the circuit to a postsynaptic input current step of 0.1 µA are shown in Figure 7.4 to Figure 7.9 and their respective parameters of the tuning voltages *Vc* and *Vd* are provided in Figure 7.3. The circuit operates approximately 10^3 to 10^4 times faster than the biological real-time, depending on the selected area of the parameter space of *Vc* and *Vd*. For comparison purposes, the scaled time domain is considered in order to adopt biological classifications methods given in Nowak et al. (2003). The circuit mimics various types of cortical neuron firing patterns: fast spiking (FS), regular spiking (RS), low-threshold spiking (LTS), intrinsic bursting (IB) and chattering (CH). Brief definitions of each of these firing patterns are presented in Chapter 2. The FS firing patterns recorded from the circuit are shown in Figure 7.4 and Figure 7.5. The RS, LTS, IB and CH firing patterns recorded are shown in Figure 7.6, Figure 7.7, Figure 7.8 and Figure 7.9 respectively.

The adaptation index measures the accommodation of the firing pattern, i.e. the progressive decrease in firing frequency despite the maintained depolarization. The adaptation index is calculated as $100\times(1-F_{ad}/F_1)$, where F_1 corresponds to the firing rate of the first inter-spike interval, and F_{ad} is the adapted firing rate (Nowak et al., 2003). The approximate values of delay between the start of the supra-threshold current injection and the first spike of the spike train, adaptation index and frequency of spiking values for each of RS, LTS and FS type firing patterns are provided in Table 7.1.

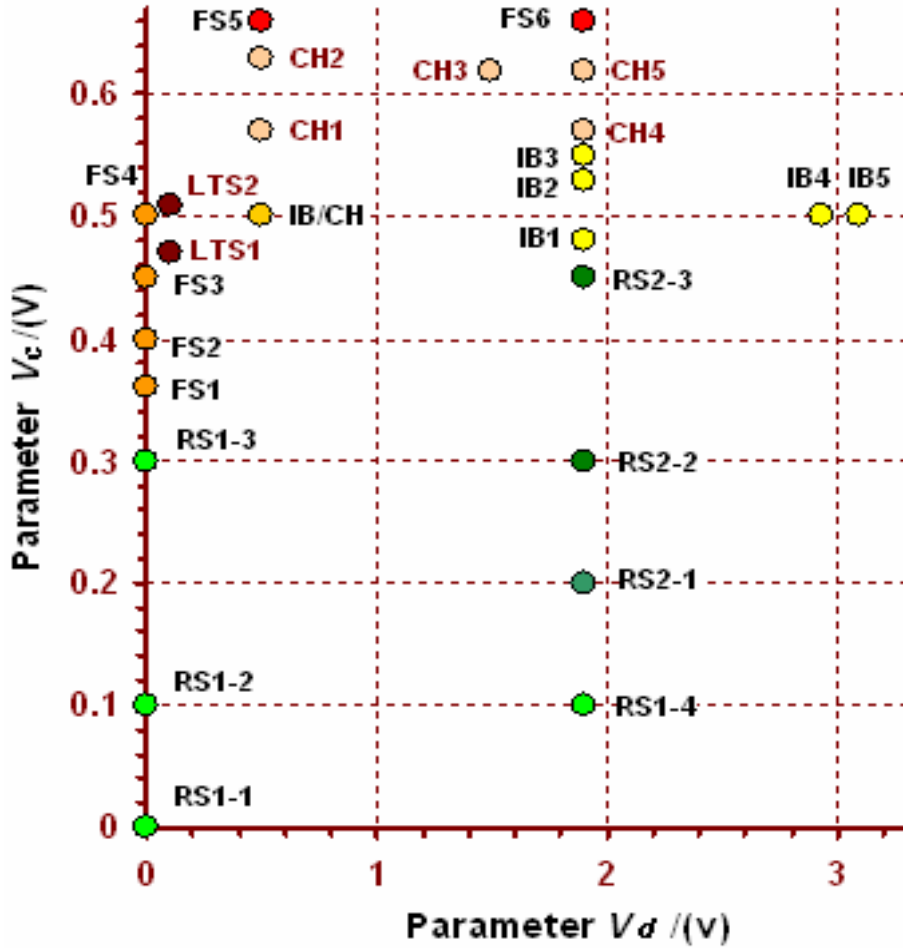


Figure 7.3 Parameters *Vc* and *Vd* that were used to obtain the cortical neuron firing patterns given in Figure 7.4 to Figure 7.9.

As seen in Figure 7.3 and Figure 7.4, the FS1, FS2, FS3 and the rest of the firing patterns where *Vd*=0 and 0.36 V<*Vc*<0.5 V behave as a FS type and their frequency of spiking lies in between 200 kHz and 800 kHz. It can be seen that all the firing patterns across *Vd*=0 V are weak-accommodating resulting in either RS1 or FS type firing patterns. The FS4 type neuron continues its spiking even after the supra-threshold current is removed, however, it shuts down if the inhibitory postsynaptic current is provided. In Figure 7.3, the approximate parameter space area where 0.2 V<*Vd*<3.25 V and *Vc*<0.4 V results in the RS neuron type and when *Vc* increases from 0 to 0.4 V, the frequency of spiking and adaptation index values increase. The parameter space area where 0.2 V<*Vd*<3.25 V and 0.45 V<*Vc*<0.56 V results in IB type firing, and different IB firing patterns can be obtained by varying *Vd* and *Vc* appropriately. Similarly, the area where 0.2 V<*Vd*<3.25 V and 0.56 V<*Vc*<0.65 V produces CH behaviour, and

various numbers of spikes in a burst and inter-bursting frequencies can be obtained by varying *Vd* and *Vc*. In the same *Vd* region, when *Vc* is greater than 0.65 V the cell produces a delayed FS firing pattern with higher firing frequency. Variations of firing patterns of the selected RS, IB, CH, and FS cell types with the variation of *Vc* across *Vd*=1.9 V illustrate the sensitivity of the firing patterns and their properties to the tuning variable *Vc* (Wijekoon et al., 2008a).

Table 7.1: Neural properties of RS, LTS and FS firing patterns provided in Figure 7.5 to Figure 7.7

Label	Delay[15] (≈ in μs)	Adaptation Index (≈ in %)	Frequency of spiking (≈ in kHz)	Type
FS1	4	21	220	FS
FS2	3	22	280	FS
FS3	1	4	400	FS
FS4	<1	2	1000	-
FS5	14	22	5500	FS
FS6	16	35	6300	FS
RS1-1	17	0	50	RS-1
RS1-2	13	13	70	RS-1
RS1-3	12	24	130	RS-1
RS1-4	16	22	51	RS-1
RS2-1	15	35	65	RS-2
RS2-2	14	44	90	RS-2
RS2-3	14	72	190	RS2
LTS1	9	72	300	LTS
LTS2	8	65	480	LTS

[15]Delay between the start of the supra-threshold stimuli and the initial spike

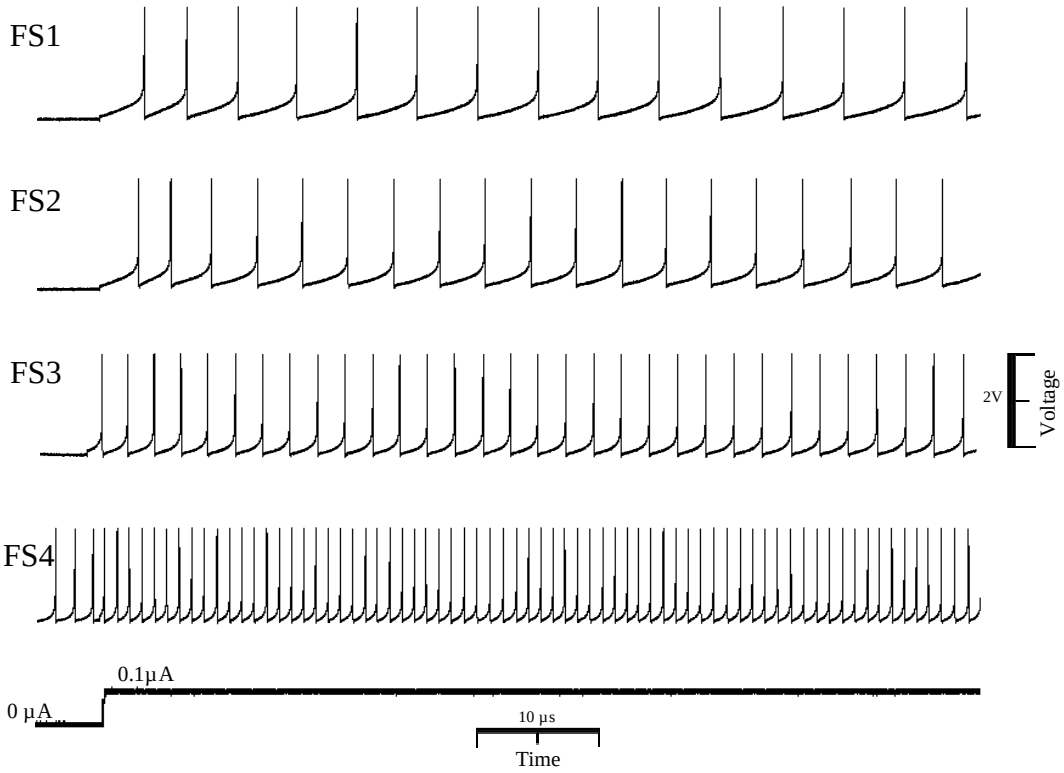


Figure 7.4 Experimental waveforms of FS cells. Each plot shows voltage response of the fabricated circuit to a 0.1 µA step current. Parameters *Vc* and *Vd* of each response are provided in Figure 7. 3.

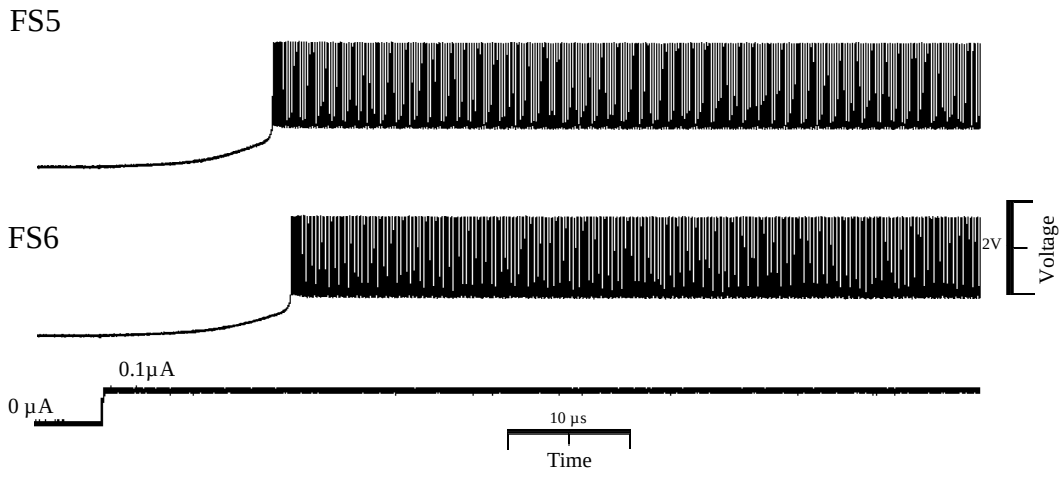


Figure 7.5 Experimental waveforms of vary fast spiking cells. Each plot shows voltage response of the fabricated circuit to a 0.1 µA step current. Parameters *Vc* and *Vd* of each response are provided in Figure 7. 3.

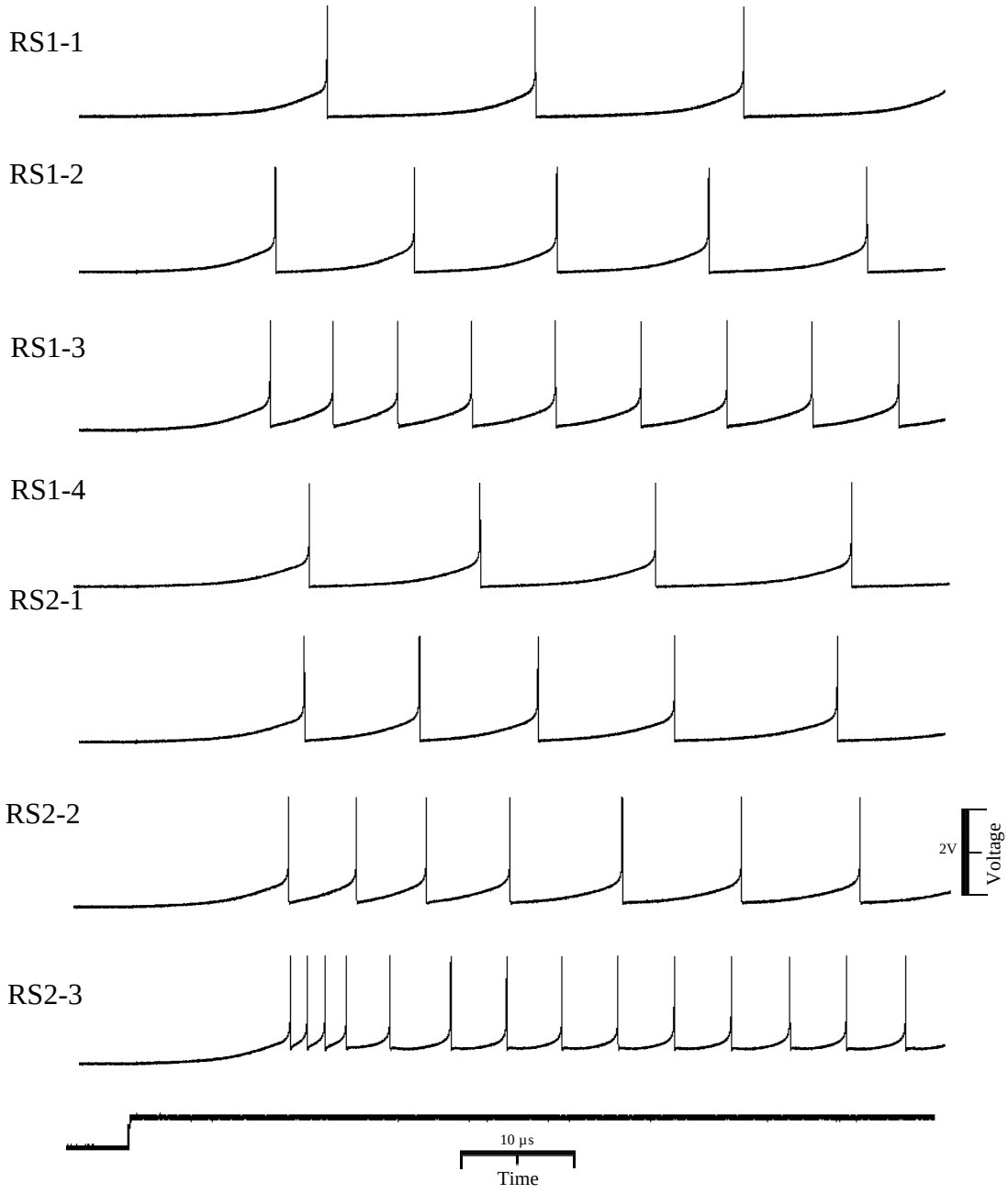


Figure 7.6 Experimental waveforms of RS1 and RS2 cells. Each plot shows voltage response of the fabricated circuit to a 0.1 µA step current. Parameters *Vc* and *Vd* of each response are provided in Figure 7.3.

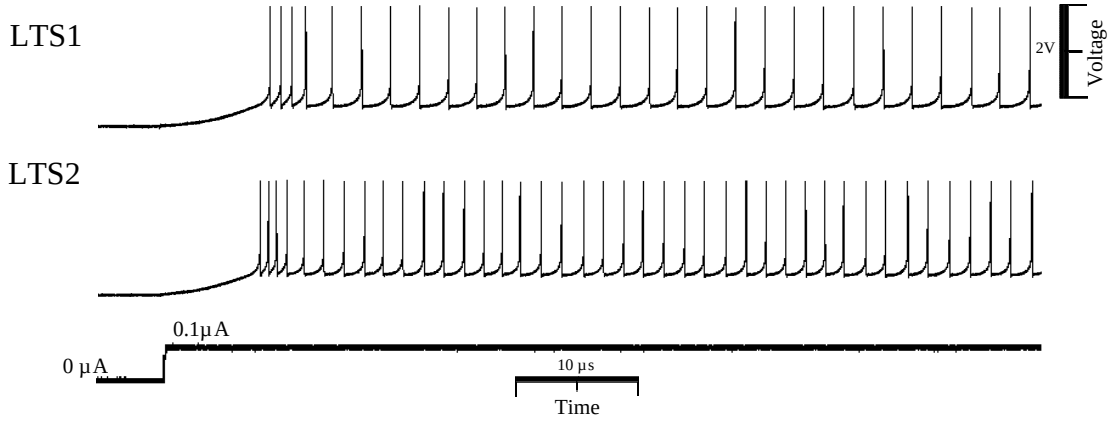


Figure 7.7 Experimental waveforms of LTS cells. Each plot shows voltage response of the fabricated circuit to a 0.1 µA step current. Parameters *Vc* and *Vd* of each response are provided in Figure 7.3.

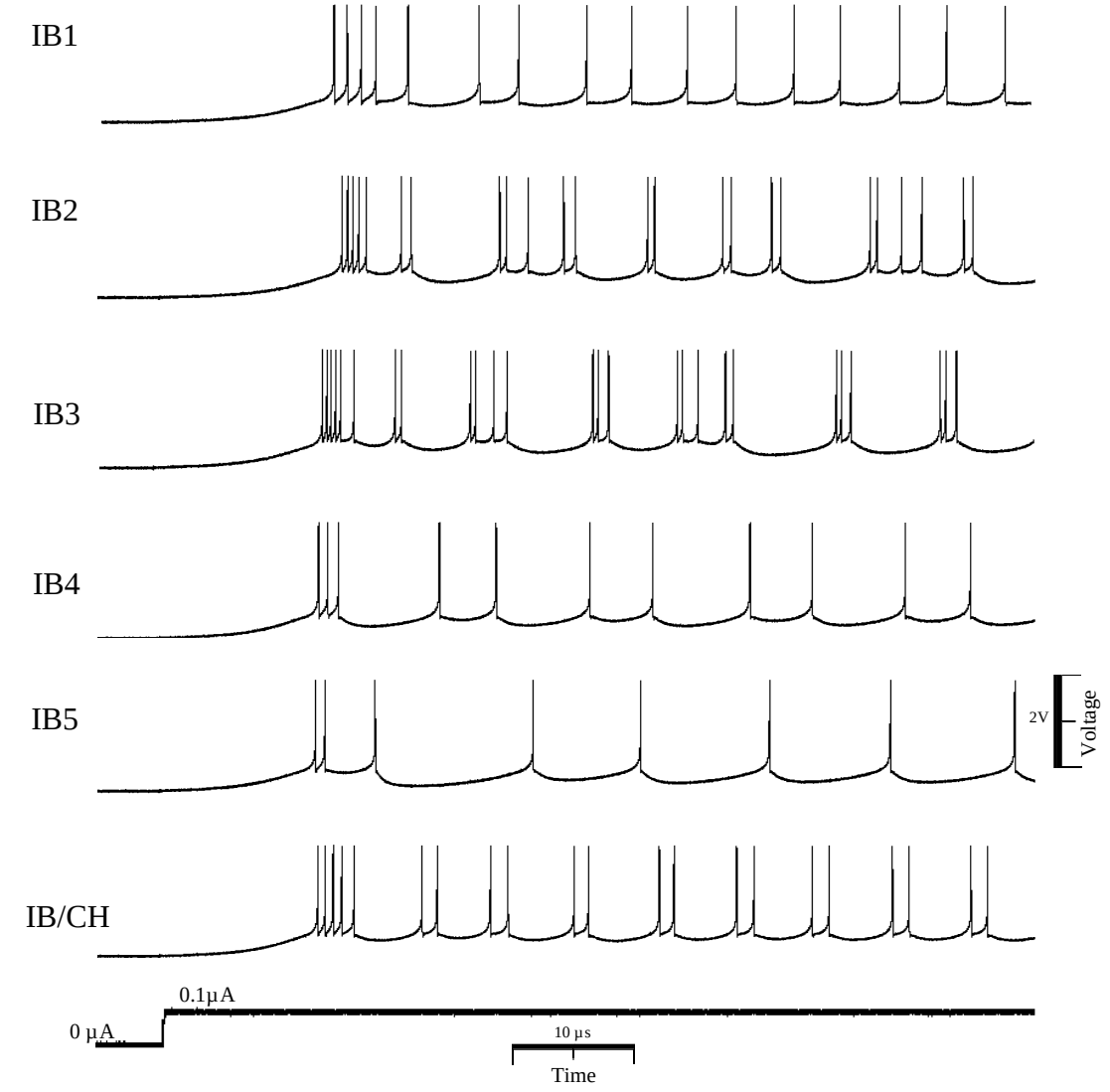


Figure 7.8 Experimental waveforms of IB cells. Each plot shows voltage response of the fabricated circuit to a 0.1 µA step current. Parameters *Vc* and *Vd* of each response are provided in Figure 7. 3.

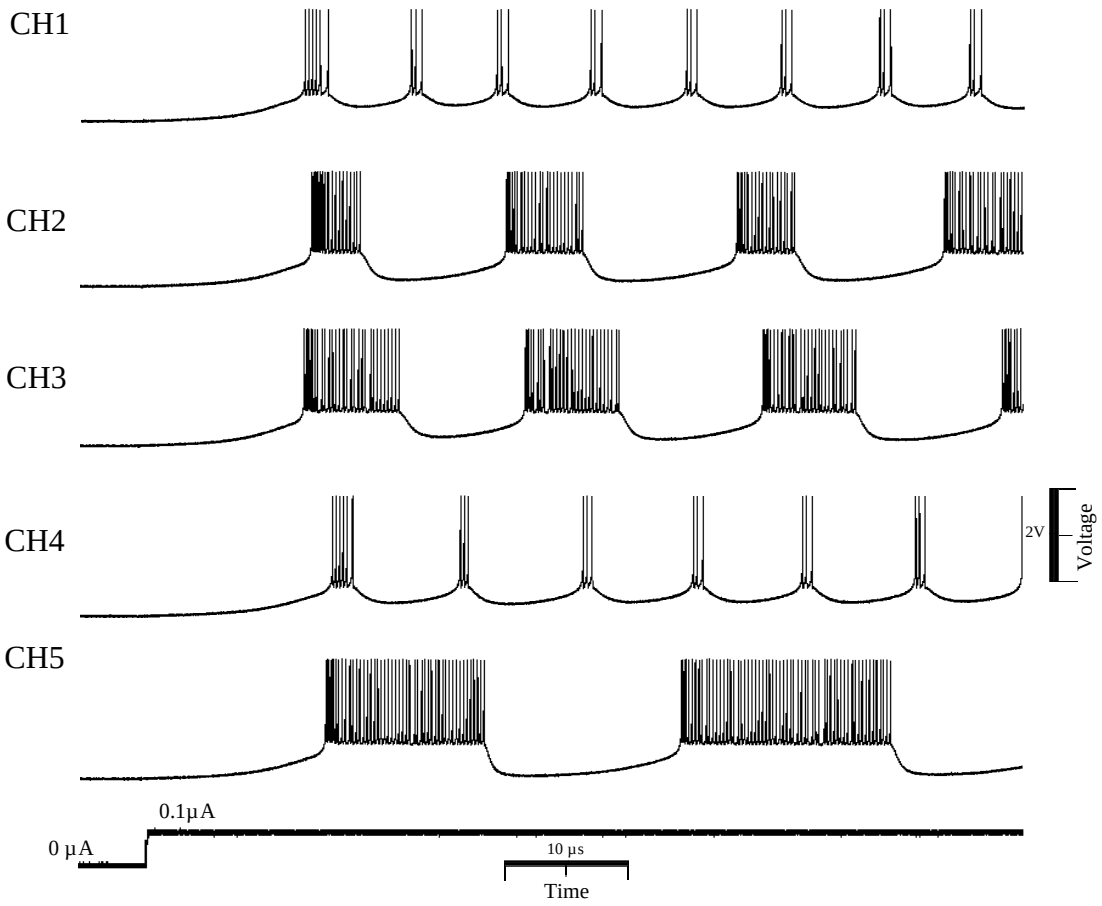


Figure 7.9 Experimental waveforms of CH cells. Each plot shows voltage response of the fabricated circuit to a 0.1 µA step current. Parameters *Vc* and *Vd* of each response are provided in Figure 7. 3.

The power consumption of the circuit is approximately proportional to the average spiking frequency. As shown in Figure 7.10 when the postsynaptic current is less than 0.1 µA, the power consumption of a circuit can vary between 0.1-65 µW. The energy per spike provides a figure of merit that allows a fair comparison of power consumption with respect to the circuit's computational performance. In the circuit the energy consumption per spike is 8.5-9.0 pJ (value obtained via post-layout simulations). For comparison, the I&F circuit described in (Indiveri, 2003) consumes 3-15 nJ/spike. The high energy efficiency of implementation is a result of the higher operating frequency, biasing with low dc currents, and the circuit topology that minimizes the current paths that do not directly contribute to the implementation of the circuit dynamics (i.e. charging and discharging of C_u and C_v) . However, it has to be remembered that a complete neural system will need to include synapse models and spike communication mechanisms, which are likely to dominate the energy requirements. It can be also noted,

for comparisons, that a simulation of the Izhikevich neuron (Izhikevich, 2003) on a conventional digital hardware platform consumes somewhere in the range of 1 µJ per spike.

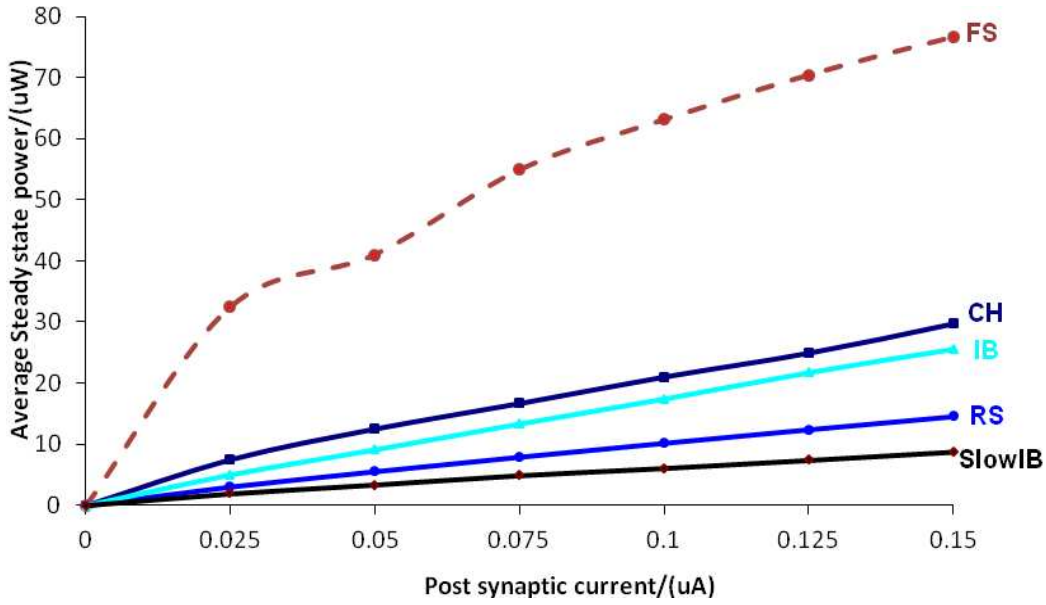


Figure 7.10 Average steady state power consumption with the variation of postsynaptic current for different firing patterns.

7.4 Conclusion

This CMOS Cortical Neuron circuit replicates many known types of spiking neural behaviours by adjusting two external voltages. The Cortical Neuron circuit provides a much richer repertoire of spiking patterns than a simple integrate and fire model. The circuit provides simple, compact and easily configurable universal cortical neurons, with potential applications in the development of large VLSI neuromorphic chips that closely resemble the circuits of the neocortex. Hence this neuron circuit is used in other IC implementations presented in this thesis.

In order to use this circuit in the other ICs neuron time acceleration should be same with all the types of firing patterns. Table 7.2 matches the biological neuron types with the corresponding set of VLSI neuron types obtained from the parameter space given in Figure 7.3 that has three orders of magnitude faster timing. The biological neurons spike timings are adopted from Nowak et al. (2003). Any FS neuron types that have less than 0.5 V *Vc* voltage values could be used as the faster FS type (to get the frequencies of VLSI neuron firing patterns see Table 7.1). Any RS and IB type given in the

experimental result matched with the required magnitude increase in time. Although the CH type matched with the required time scaling, some CH type VLSI neuron configuration have the refractory period less than 1 µs.

Table 7.2. Three orders of magnitude accelerated time VLSI neuron to biological time neuron type mapping ; ISI - inter-spike interval, Inter burst - frequency or ISI between two burst of spikes, Burst - frequency or ISI between two spikes in a bursts of spikes.

Description			**FS**	**RS**	**IB**		**CH**	
					Burst	**Inter burst**	**Burst**	**Inter burst**
Biological Neurons	**Frequency (Hz)**	Max	330	100	500	130	1000	100
		Avg	130	45	400	40	500	70
		Min	66	20	130	20	200	40
	ISI (ms)	Min	3	10	2	7.5	1	10
		Avg	7.5	22	2.5	24	2	14
		Max	15	50	7.5	50	5	25
VLSI neuron			***Vc*< 0.5 FS**	**All RS**	**All IB**		**All CH but refractory period is low**	

Therefore, overall all the VLSI neuron firing patterns can be mapped with three orders of magnitude faster time scaling as given in Table 7.2. However, some CH neurons' inter burst spikes could be ignored, if maximum network routing delay is set to the refractory period of the neuron (spike routing delays are discussed in Chapter 9.6).

CHAPTER 8 : STDP-DA SYNAPSES NEURON CHIP

This chapter presents the STDP-DA Synapses Neuron prototype chip. The chip accommodates accelerated time STDP Synapse circuit, the Dopamine Modulated STDP Synapse circuit (Chapter 5) and the Cortical Neuron circuit (Chapter 4) to verify the functionality of these circuits experimentally. Mainly, this chip tests the STDP Synapse circuit together with the Neuron circuit so that these neural elements can be accommodated in the CNL chip presented in Chapter 9.

The STDP plasticity rule is an important feature of a cortical network for the learning and memory formation of a network. Recently developed DA modulated STDP plasticity rule (Izhikevich, 2007) is also known to perform learning, in particular, reinforcement learning in a neural network. Hence, encompassing these two promising rules in a VLSI circuit could have the potential for the formation of learning and memory in a VLSI based cortical network.

The initial sections of this chapter present an overview of the chip, and the neural and auxiliary circuit implementations. This is followed by the chip test setup and the experimental results sections. Some of the materials discussed in this chapter have been accepted for publication in the Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS (Wijekoon et al., 2011).

8.1 Chip Overview

The prototype test integrated circuit, contains twenty eight STDP /DA-STDP synapses with a global DA generator circuit, and two cortical neuron circuits, fabricated in a standard 0.35µm CMOS technology. The STDP /DA-STDP synapse can be configured to work as a STDP synapse or as a DA-modulated STDP synapse and the size of the synapse cell layout is 26 µm by 50 µm. The area of the chip is 1.8 mm by 1.8 mm with 44 pins. Figure 8.1 shows the layouts of the chip and the photographs of the fabricated chip. Although the actual circuit area is approximately 0.5 mm by 0.8 mm, a large chip area is used to accommodate 44 pins needed by the circuits. This prototype chip is fabricated to test the functionality of the STDP Synapse circuit and DA-Modulated STDP Synapse circuit along with the Neuron Circuit. Accordingly, some of the internal states of all the synapses can be observed externally and at any given time, the internal

states of two synapses can be observed along with the inverted spike outputs of both neurons. The STDP or Eligibility traces functionality of the synapse can be tested and calibrated. However, due to the pin constrains of the test chip, internal states of DA Generator circuit and the DA pulse (*Vda*) signals are not provided to any pins of the chip, and hence these signals cannot be observed externally. More details of the configurations and functions are presented in the next section.

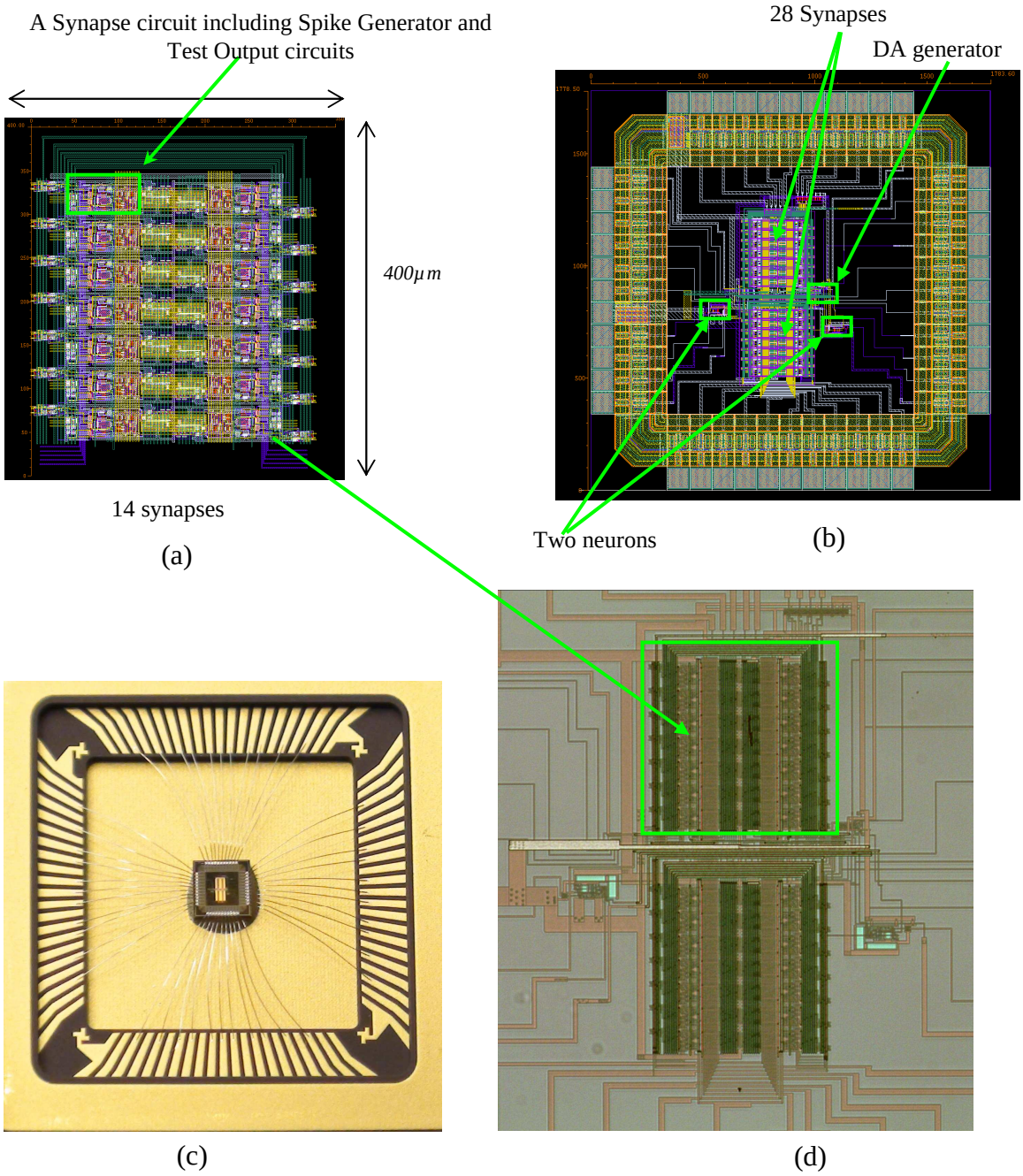


Figure 8.1 STDP-DA Synapse Neuron chip (a) layout of a 14 Synapse circuit, (b) STDP chip layout, (c) picture of the packaged chip, (d) picture of a fabricated chip.

8.2 Circuit Implementations

This section provides the circuit implementation details of the core neural circuit elements of the STDP-DA Synapses Neuron chip and their composition, the auxiliary circuit details and their configurations. The core neural circuits include Synapses, Neurons and Dopamine Generator circuits. The pre-synaptic inputs and spike outputs are made accessible to an off-chip device preferably an FPGA. If necessary, a small network can be formed by configuring the connections of the network using this off-chip device. Therefore, the auxiliary circuits include the periphery circuit used to route these inputs and outputs to the bond pads (pins of the chip). The chip also employs a Synapse Debugging circuit to test the functionality and to calibrate synapses.

8.2.1 Neural Circuits

All the 28 DA Modulated synapses of the STDP-DA Synapses Neuron chip share a single Dopamine Generator circuit. The extracellular DA level (*eDA*) can be provided as an analogue voltage through a dedicated chip pin. Figure 8.2 shows the synapses and neurons composition of the STDP-DA Synapse Neuron Chip and Figure 8.3 shows the arrangement of the neural circuit elements using block diagrams.

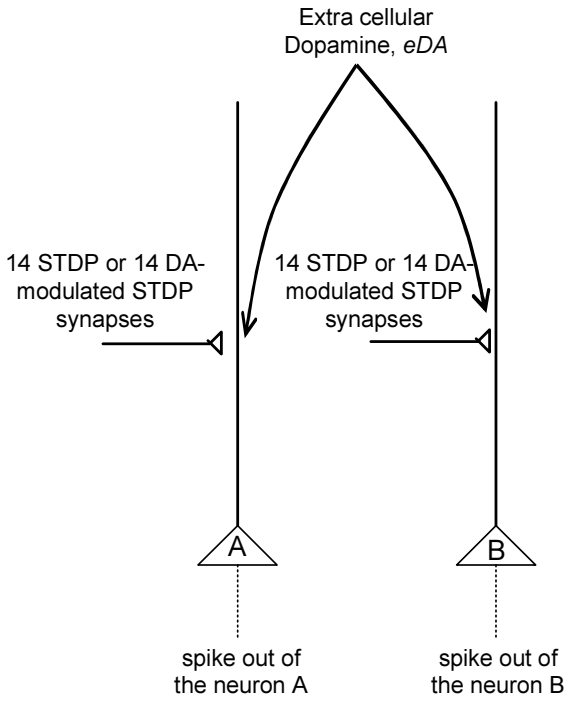


Figure 8.2 Schematic of the synapse and neuron composition of the STDP-DA Synapses Neuron chip.

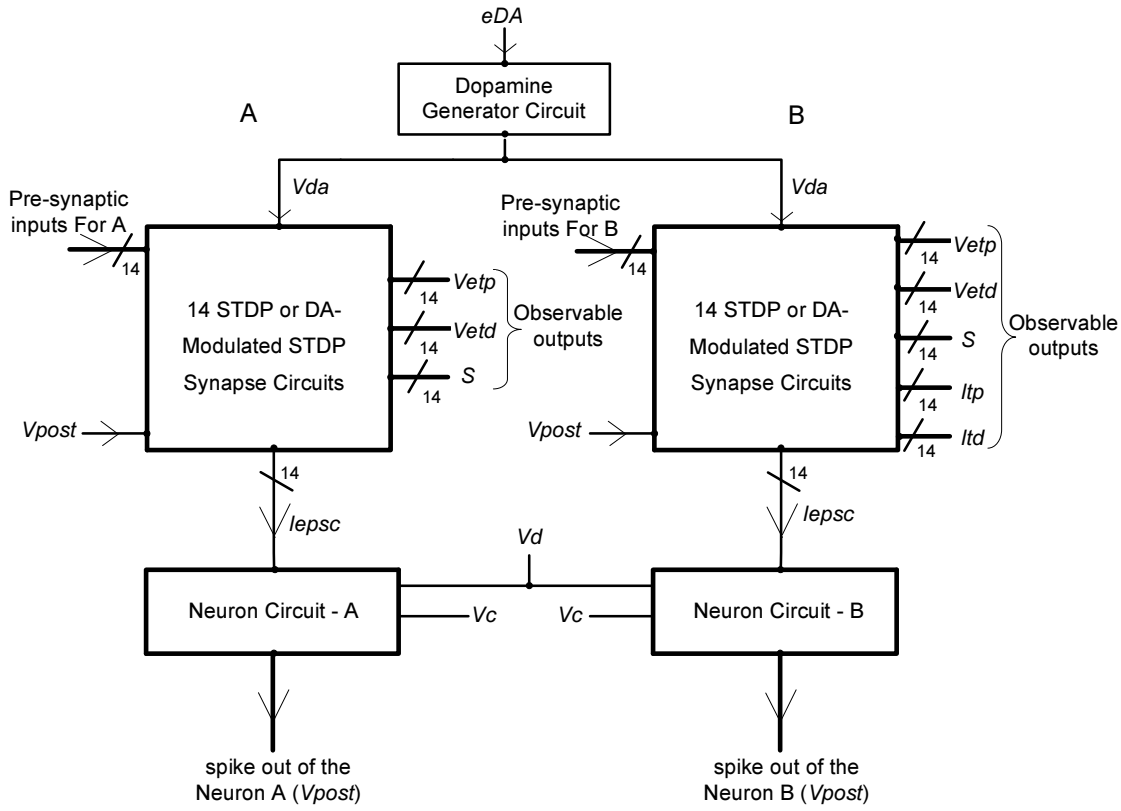


Figure 8.3 Synapse and Neuron composition in STDP-DA Synapses Neuron Chip: block diagrams of the neural circuit arrangement.

8.2.1.1 Synapse Circuit

The 28 synapses can be configured to work in a DA-modulated STDP mode or in a basic STDP mode. Each synapse comprises of an Eligibility Trace (ET) circuit, a Synaptic Strength circuit and a Post-Synaptic Current Generator circuit.

The ET circuit and the Synaptic Strength circuit are shown in Figure 5.10 and Figure 5.11 respectively. The ET "leakage" circuit parts of the ET circuit, and the *Vstdp_en* transmission gate used to configure the type of the synapse are shown in Figure 8.4. The *Vstdp_en* can be provided externally. In the usual mode of operation, the DA-modulated synapse mode, the *Vstdp_en* flag is set to logic low (0 V) and the voltages *Vlkwd* and *Vlkwp* are provided with appropriate analogue values that suit the synapses to operate as the DA modulated STDP synapses.

In order to work as an STDP synapse, the signals *Vstdp_en* flag, and voltages *Vlkwd* and *Vlkwp* are set such that the ET circuit is modified to work as STDP circuit. This is done by switching off the additional circuits of the ET circuits and by combining the *ETp* and

ETd nodes of the ET circuit to act as the synaptic weight node, which is connected to the Post Synaptic Current Generator circuit. The Post-Synaptic Current Generator circuit is shown in Figure 8.5.

By setting the *Vstdp_en* flag to logic high, the transmission gate shown in Figure 8.4 (b) connects the *ETp* node of the ET circuit to the *ETd* node. This forms a common synaptic weight node, *Vwstdp* (complementary topology of the STDP circuit shown in Figure 5.2). Simultaneously, the M2 transistor of the Post-Synaptic Current Generator circuit shown in Figure 8.5 is switched on and M4 is switched off. Consequently, the weight of the STDP circuit is buffered to the *Vw* node of the Synaptic Current Generator circuit rather than to the strength, *S*, of the DA-Modulated Synapse circuit (the transistors M1-M3 when M2 is switched on and the transistors M1, M4 and M5 when M4 is switched on form source follower circuits). The isolation of the additional "leakage" circuits (Transistors M8p -M7p and M8d-M7d of the ET circuit) is achieved by supplying the supply voltage (*Vdd*) to the *Vlkwp* and 0 V(*gnd*) to the *Vlkwd*, so that the ET circuit becomes the STDP circuit.

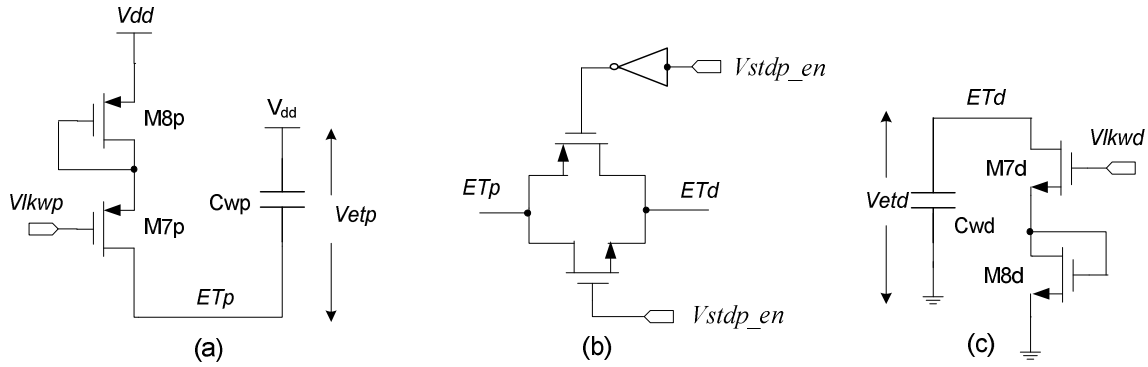


Figure 8.4 Eligibility trace "leakage" circuit parts of the ET circuit shown on (a) and (c) ;The Transmission Gate use to connect *ETp* node to *ETd* node is shown in (b).

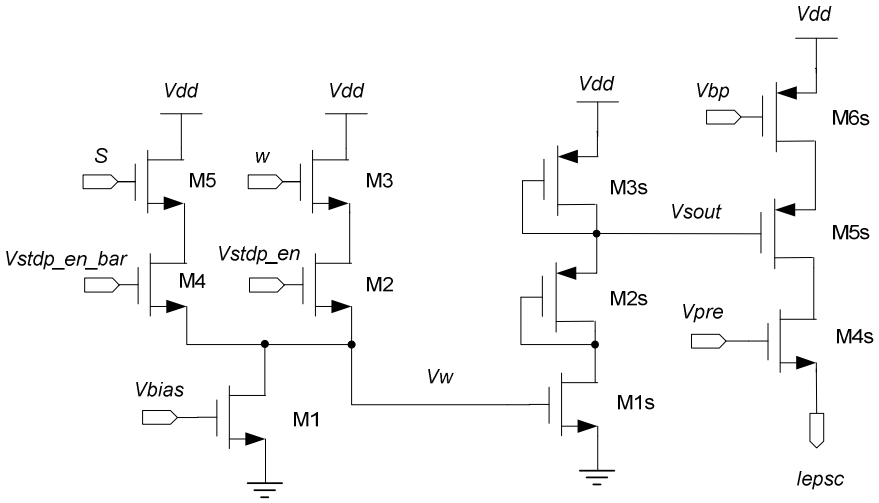


Figure 8.5 Post-Synaptic Current Generator circuit in STDP-DA Synapses Neuron chip.

8.2.1.2 Neuron Circuit

The accelerated time neuron (Chapter 4) is used in this chip. The tuning voltages *Vc* and *Vd* are used to configure the neuron to a given type of spiking behaviour. As the limited numbers of pins are available on the chip, the two neurons in the chip use a common *Vd* chip pin (Figure 8.3). It is also seen in Chapter 4 that the basic neuron types can be obtained for different values of *Vc*, while *Vd* is at a constant voltage (e.g. at *Vd* = 1.9 V for different values of *Vc* most of the neuron types can be obtained; see parameter space Figure 7.3 of Chapter 7). Therefore, this does not severely limit the number of possible spiking patterns. Neuron outputs are buffered using inverted buffer circuits, and the inverted spike outputs are provided to the chip pins. Therefore, these outputs can be read directly by an external device.

8.2.1.3 Dopamine Generation Circuit

A Dopamine Generator circuit is used to generate a DA signal for all the synapses globally, and the level of the DA can be controlled by the external supply voltage, *eDA*. Here, the *eDA* voltage with referenced to *Vdd* is considered as the DA level. The DA Generator circuit implemented in the chip is given in Figure 8.6.

The decay of the DA level and the DA injection using a burst of spikes are not implemented as in the Dopamine Generator circuit presented in Figure 5.13 of Chapter

5. This DA level needs to decay with time constant of 0.2 ms (in accelerated time) and the level should be able to be increased using a burst of input spikes (i.e. digital pulses) that result in the injection of dopamine as a reward signal. In order to reduce the pin requirements of the chip, the capacitor/transistor circuit that does this is implemented off chip. The operation of the DA Generator circuit is described in Chapter 5.

The Clock signal (*Vda_clk*) of the DA Generator circuit that generates the *Vda* pulses also needs to be provided by an off-chip device. The width of the pulse can be tuned using the leakage voltage, *Vlk*. Aforementioned, the internal states of DA Generator circuit and the DA pulse (*Vda*) output signals are not accessible to the off-chip pins due to the pin constrains and hence these signals cannot be observed externally. However, the ways that could verifying its function is discussed in Section 8.5.

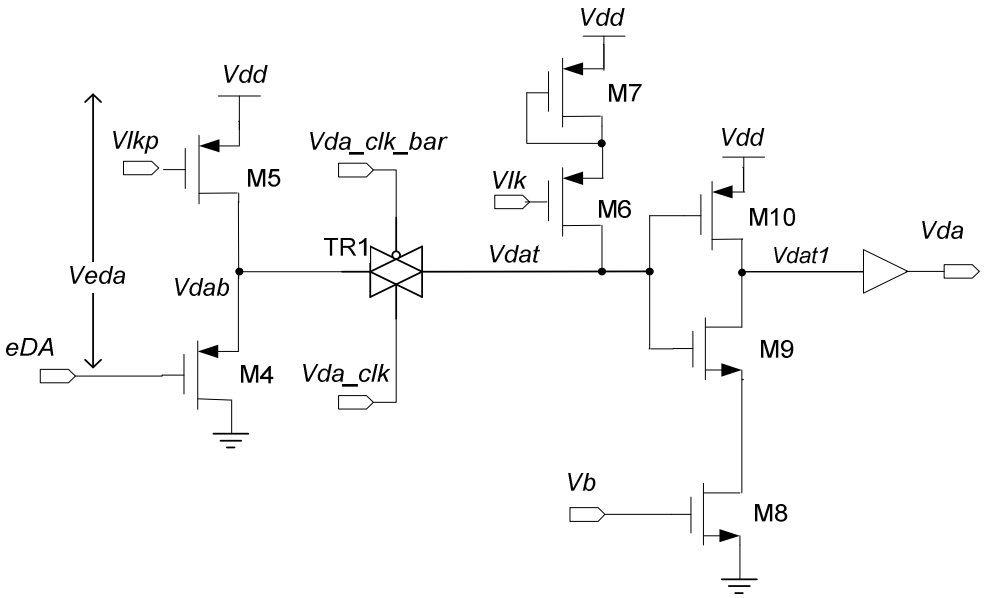


Figure 8.6 DA Generator circuit in the STDP-DA Synapses Neuron chip.

8.2.2 Auxiliary Circuits

The core neural elements discussed above need to be configured as a network, in which their characteristics can be set using the tuning parameters, and their functionality can be measured externally. These functions are facilitated by auxiliary circuits fabricated on the chip. Some of the auxiliary circuits discussed in this section include the

Pre-synaptic addressing and Spike Generation circuits, the Synaptic weight pre-setting circuit, and a circuit for observing & debugging of synapses.

8.2.2.1 Pre-synaptic addressing and the Spike Generation circuit

The pre-synaptic input (*Pre*) of each synapse is connected to the output of the Spike Generator circuit. Figure 8.7 (a) shows the Spike Generator circuit and on the chip this circuit is physically located next to the synapse layout to provide non-attenuated spikes. Using a five bit pre-synaptic address, the Spike Generator circuit for a target synapse can be activated to generate a pre-synaptic spike.

When a synapse is addressed, the incoming address is decoded, which enables the Spike Generator circuit of the target synapse. Once the address decoder enables the Spike Generator, a 5 ns *Pre* pulse is sent to the target synapse. The Spike Generator circuit consists of a Four Input NAND gate, a Delay (DLY) circuit, an Inverter and a NOR gate. Figure 8.7 (b) shows the input, intermediate, and output signals of the Spike Generator circuit when the Spike generator circuit is enabled twice.

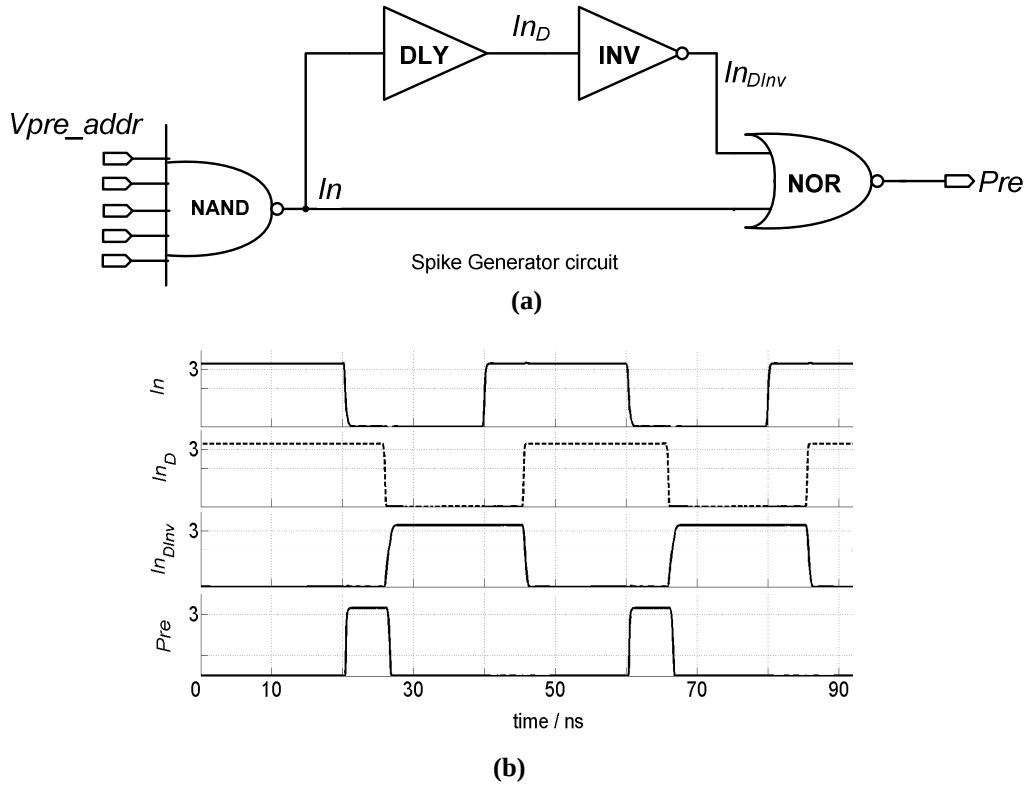


Figure 8.7 Spike Generator (a) circuit (b) intermediate signals.

8.2.2.2 Synaptic weight pre-setting

For basic STDP synapse mode weight of the synapse can be set of reset by providing appropriate voltages to *VlkWd* and *VlkWp* at the network configuration stage (i.e. before the network emulation). For the DA modulated STDP mode of operation the eligibility traces (voltage across the capacitors Cwp and Cwd) can also be set or reset using the bias voltages *VlkWd* and *VlkWp* and the *STDP_En* switch at the network configuration stage. The leakage circuit for the eligibility traces and the transmission gate are shown in the Figure 8.4.

8.2.2.3 Observing and debugging of synapses

As seen in Figure 8.3 the internal states of the synapses can be observed externally. These include *Vetp*, *Vetd*, and *S* for the synapses of the neuron A, and the *Vetp*, *Vetd*, *S*, *ltp*, and *ltd* for the synapses of the neuron B. At any given time, all these internal states of a synapse from Neuron A and a synapse from Neuron B can be observed. A four bit test address is used to select observable internal states for a particular synapse. The address is provided to the 32:1 multiplexer, using the dedicated test address bus of the chip. Depending on the test address, the multiplexer switches the outputs to observe the state of the targeted synapse.

For STDP functionality, both synapse types require the output of the targeted neuron (i.e. the post-synaptic neuron's spike) as an input. The post-synaptic signal can be provided either with the output of the targeted neuron or with an artificial external post-synaptic signal, which can be used in order to debug the synapse circuits.

8.3 Test Setup

The test setup includes the STDP-DA Synapses Neuron chip, a computer (Host PC), a Digital to Analogue Converter (DAC) and an Oscilloscope (Figure 8.8). This setup can be used to verify the function of the synapses and neurons. The DAC is used to supply the tuning parameter voltages that set the characteristics of the synapses and the neuron types. These bias voltages can initially be programmed using the Host PC. The FPGA is

used to configure the chip, implement the connectivity of the neural network and to facilitate spike routing (a Xilinx Spartan 3 FPGA[16] is used here).

Perhaps it is impossible to build a network that demonstrates truly useful network behaviour using only two neurons. However, the setup can be used to build two neuron networks or to have many *virtual neurons* along with the two neurons. In addition to the post-synaptic spikes of the two neurons, spikes can be generated on the FPGA (as spikes from a *virtual neuron*) using a set of rules or by reading an input file from the computer. Therefore, pre-synaptic spike can be provided to the chip using the FPGA either depending on the spike event received from the two neurons or by using spikes from a *virtual neuron*. Furthermore, the required observable outputs can also be set using the FPGA, by sending a five bit test address to the test address bus of the CNL chip. Once network emulation starts, the observable internal state of the synapses, neurons spike activities and pre-synaptic addresses can be fed to a digitising oscilloscope to observe and record the waveforms. Simultaneously, the network spike activities can be recorded on to the FPGA board memory.

[16] Xilinx Spartan 3 Web link: http://www.xilinx.com/support/documentation/spartan-3.htm

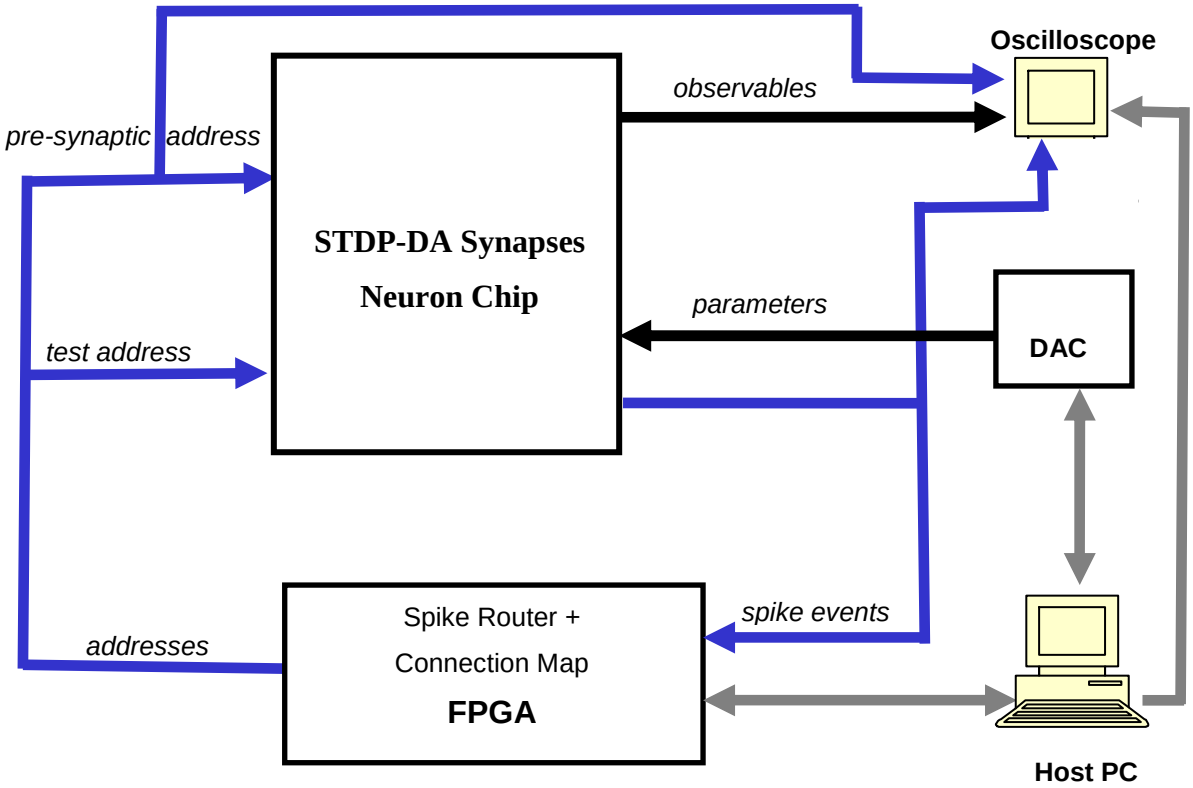


Figure 8.8 Test setup of the STDP-DA Synapses Neuron circuit.

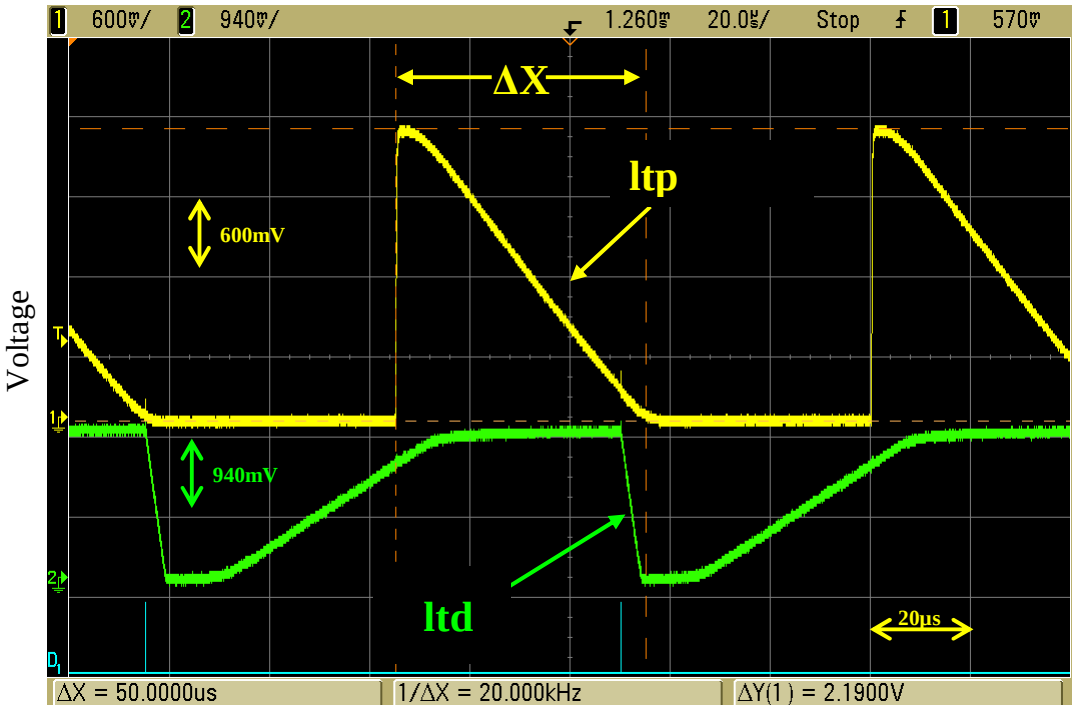
8.4 Experimental Results

This section provides experimental results obtained from the chip that verify the STDP functionality and the cortical neuron functionality.

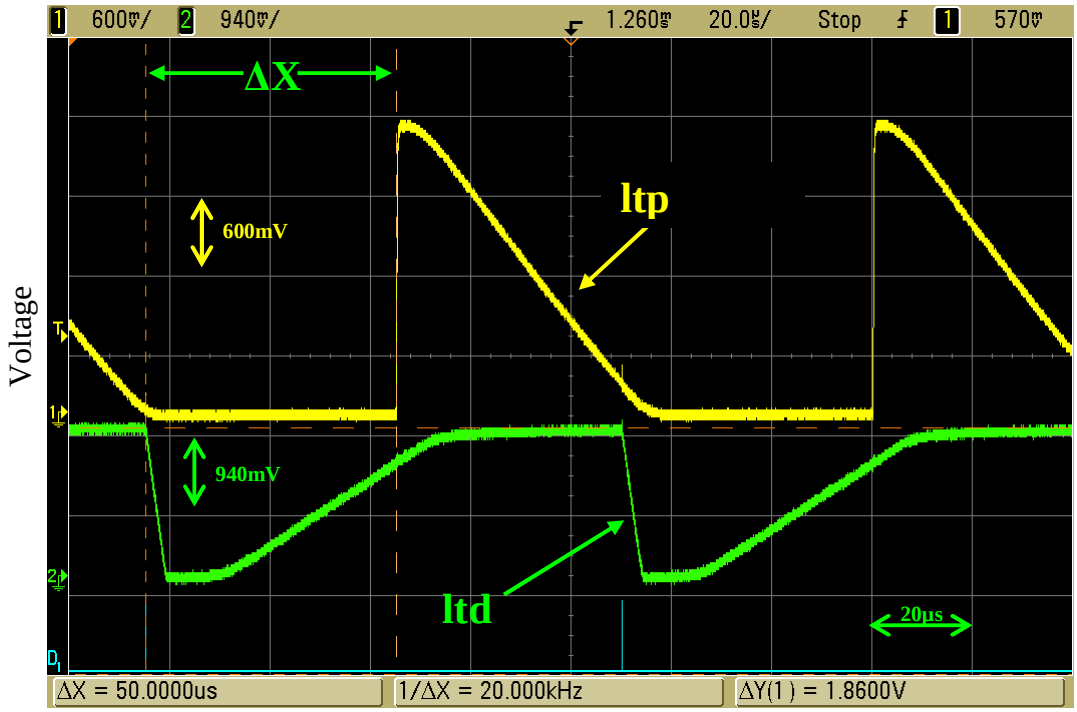
8.4.1 STDP Synapse

Initially, the LTP time window and the LTD time window can be adjusted using *Vleakp* and *Vleakd* respectively (As explained in Section 5.3). Figure 8.9 shows measurements of the time windows after setting it to 50 µs. The signal *ltp* reaches the maximum voltage when pre-synaptic spike (*Pre*) fires and the *ltd* reaches its minimum when a post-synaptic spike (*Post*) arrives. As the *ltp* and *ltd* signals are provided to the chip pins after the NMOS source follower buffer stage, hence 0 V to 0.6 V an approximate voltage range of the output waveforms could not be observed. The lower waveform part of the *ltd* signal is distorted as shown in Figure 8. The slop in the charging phase of the *ltd* is due to the input capacitance of the source follower circuit.

Figure 8.10 shows the long-term plasticity effect on synaptic weight (*w*), for many pre- and post-synaptic spike pairings. Figure 8.11 provides synaptic weight change when post-synaptic spike follows the pre-synaptic spike for 10 occurrences within approximately 1 ms (accelerated time) duration and when post-synaptic spike precedes the pre-synaptic spike for 10 occurrences within 1 ms duration. It is seen that when *Pre* follows *Post* the synaptic weight is potentiated, whereas *Post* precedes *Pre*, the synaptic weight is depressed, implementing STDP rule in the synapse.



(a)



(b)

Figure 8.9 The *ltp* and *ltd* signals showing the history of the pre- and post-synaptic firing timing respectively, (a) LTP time window measurement (b) LTD time window measurements; their are set to 50 µs using the tuning voltages of the synapse, *Vlkp* and *Vlkd*.

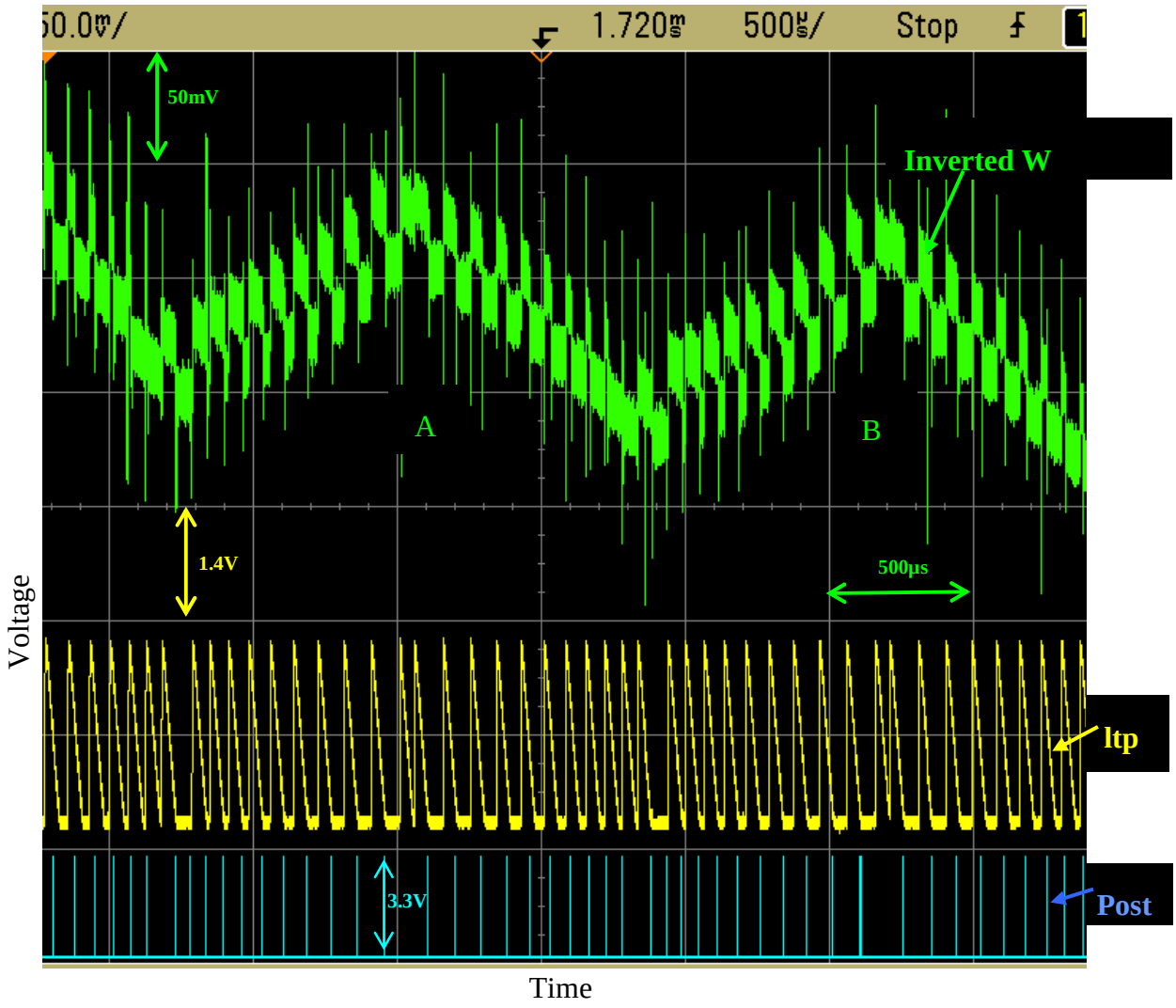


Figure 8.10 The weight of the synapse variation for different pre- and post- synaptic firing sequences; enlarge waveforms clearly showing the long-term potentiation and depression due to the spike sequence are shown in Figure 8.11.

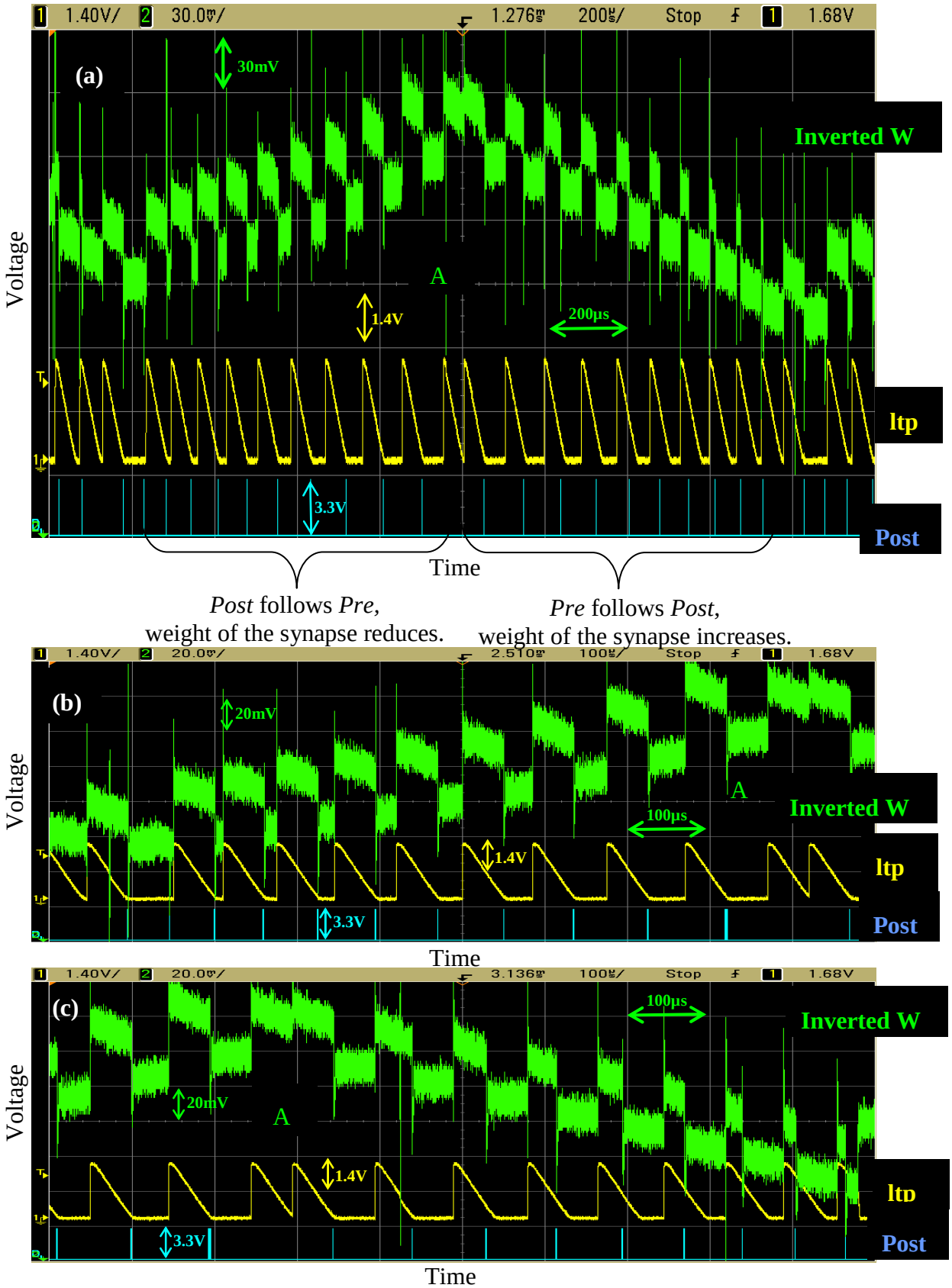


Figure 8.11 STDP synapse weight variations (a) depressing and potentiation synaptic weight due to the Pre and Post firing sequence, (b) weight depressing when Pre follows post, (c) weight potentiating when Pre precede. Note: here inverted weight is recorded.

8.4.2 Neurons

By providing the same continuous pre-synaptic spikes to many synapses, synaptic currents are injected to a neuron regularly. For the same synaptic current injection, the neuron's tuning parameters *Vc* and *Vd* are set to different firing patterns, CH, RS and IB types. These firing patterns (inverted) are shown in Figure 8.12.

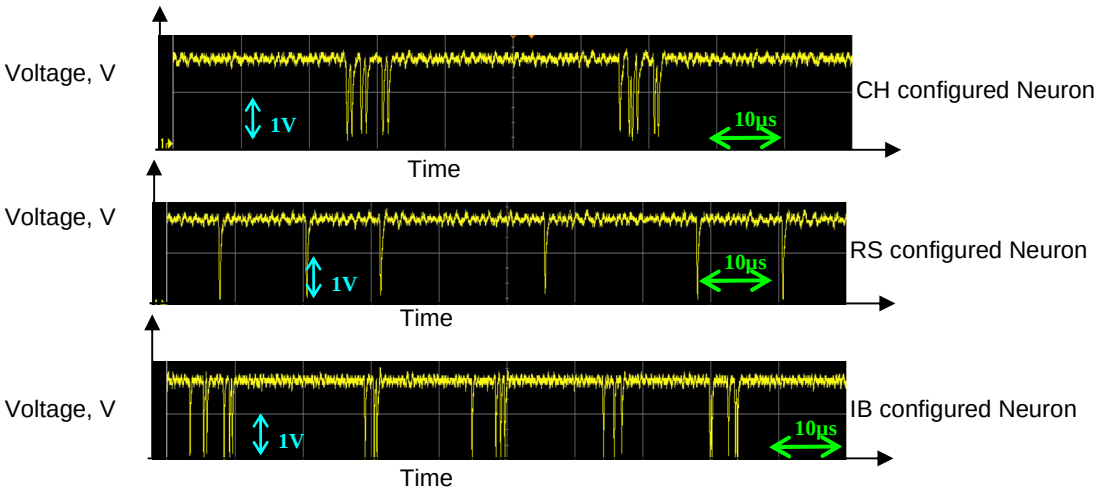


Figure 8.12 Different firing patterns obtained for the same pre-synaptic stimulus. Here inverted spike output of the neuron is recorded.

Power consumption

The synapse circuit typically consumes between 2 µW and 5 µW power at *Pre* and *Post* synaptic spike rates of 200 kHz (i.e high neural activity level), but it could be as high as 40 µW, depending on the synapse state, parameters and spike rates. The DA generator circuit, which is shared by many synapses, consumes up to 600 µW of power (worst case).

8.5 Discussion and Conclusion

8.5.1 Discussion

Noise of the observable signals

The results presented in this chapter are taken from a manually wired circuit board and as seen in Figure 8.10 and Figure 8.11 observable waveforms are noisy. This could be reduced by rebuilding the test setup with a printed circuit board (PCB) that includes high frequency filtering using capacitor banks and by properly shielding and reducing

the length of the wires of the observable analogue signals. Specially, slowly varying, narrow voltage swing observable signals such as synaptic strength signals are noisier to observe.

DA-Modulated Synapse circuit

In order to test the DA-Modulated STDP Synapse circuit *Vetp*, *Vetd*, and *S* can be observed externally however any internal signal related to the DA signal cannot be observed. The noise in the system makes it difficult to observe *S* which is a slow varying signal. Under these conditions, DA-modulated synapse can be verified by using the synapses in a large network and observing the reinforcement learning in the network. Once the CNL chip is ready the STDP-DA Synapses Neuron PCB can be connected together to test the DA-Modulated Synapse functionality. Once the DA-modulated chip functionality is verified, this board can be use in the Cortical Neural Network Architecture proposed in Chapter 9 along with the CNL chips to facilitate reinforcement learning in a large network.

Network Connectivity capacity of the Chip

The two neurons along with the DA-modulated synapses can be use in a network (e.g. with the CNL chip discuss in Chapter 9) initially to test the DA-modulated synapse functionality and then to facilitate the reinforcement learning dynamics on a large VLSI network. However, there is a limitation to which these DA-modulated synapses can be connected.

The pre-synaptic addressing is provided to the chip using the five bit addresses bus of the chip. At any given time only one synapse can be addressed, and each pre-synaptic routing consumes two clock cycles.

Therefore, assuming that the FPGA device facilitates the spike routing for the network and its clock frequency is 100 MHz, if all the incoming spikes are from fast neurons and are continuously firing at the spike frequency of 200 kHz, and each incoming spike is routed to a maximum of two synapses (K =2), then the maximum number of spikes that can be routed within the refractory period of the neuron (1 µs) is approximately 125. I.e if each incoming spike connect only to two synapses (simple connectivity) then 125 neurons (at high spike activity level) can be connected to the chip in order to form a network. The equations used to calculate this figure are given below.

Average spiking rate of all the neurons = 0.2 MHz

Average spike rate of N neurons = 0.2 x N MHz

Number of synapses addressed per incoming spike = K

Clock rate of the FPGA = CLK_{FPGA}

Number of cycles for pre-spike addressing = CYC

Number of spikes can be served = CLK_{FPGA}/ (200 000 x K x CYC)

8.5.2 Conclusion

It is verified from the experimental results that the SDTP Synapse circuit and the Synapse circuits along with the neuron operate as expected. Therefore, these circuits are used in the CNL chip presented in Chapter 9. the DA-Modulated Synapse functionality need to be tested by demonstrating reinforcement learning in a network, which require sufficiently large VLSI network. The CNL board (Chapter 9) can be used to test the reinforcement feature of the Chip.

CHAPTER 9 : VLSI CORTICAL NEURAL NETWORK AND CORTICAL NEURAL LAYER CHIP

A prototype microelectronic Cortical Neural Layer (CNL) integrated circuit that could closely represent the neuron and synapse type composition of a layer of the neocortex is fabricated using elementary circuits proposed in Chapters 4 to Chapters 6. This chip incorporates more biologically plausible heterogeneous neural elements than the other similar size VLSI neural network implementations found in literature. In order to realise a larger cortical network in microelectronic hardware, a VLSI Cortical Neural Network (VCNN) architecture that combines many CNL chips together is proposed. The next section presents an overview of the VCNN architecture, and Section 9.5 provides details of CNL board used in the VCNN architecture. The CNL chip overview, circuit implementations, and model are presented in the rest of the chapter. In analysing the feasibility of building a large neural network in VLSI, estimation of the network size that can possibly be implemented using similar composition of neural elements as of CNL chip, in a wafer-scale integration is provided in Chapter 10.

Performing neuron-level recordings on animals is very limited both in the number of observable neuronal activities and in experimental time the neural tissue can be kept alive. The proposed architectures could provide a neural accelerator platform that can test some computational and neurobiological models of network. These platforms can also perform extensive parameter searches of an experiment as it works in three orders of magnitude faster than biology. These experiments could help to improve the understanding of the underlying principles of cortical processing.

9.1 VLSI Cortical Neural Network Architecture (VCNN) - Overview

Neocortical neural tissue is composed of anatomically repeating six-layered neural network. Each layer is composed of neurons that may receive spikes through synapses from neurons in the same layer, from the other layers, from external cortical or sensory afferents, or any combination of these. The composition of neurons and synapses in a layer is highly heterogeneous and different layers are composed of a variety of neuron and synapse type combination (for more details see Chapter 2).

The proposed VCNN architecture combines six CNL boards together to build a six-layered VLSI cortical network that could closely resemble a small scale network of the neocortex. A CNL board comprises of a CNL chip, and a dedicated FPGA device. The CNL Chip implements generic neural layers. Each chip can be pre configured to represent a layer of the cortical network by configuring neuron and synapse type composition appropriately (more details of the CNL chip are given in Section 9.3). The FPGA is used to pre configure the CNL chip and to observe the spike activities of the network. Most importantly, the network connectivity configuration and spike routing within the CNL chip and between CNL boards are facilitated by the FPGA (more detail of CNL board is given in Section 9.5). Therefore, the spike routing of the VCNN architecture is carried out using the six distributed routers (each implemented in the FPGA of a CNL board), and according to the connectivity matrix of a given network, the router's lookup tables (LUTs) of the routers that define the inter neuron connectivity, can be pre configured.

This small six-layered VLSI cortical network comprises of 720 cortical neurons of different cortical neuron types and 45 360 short- and long- term plastic synapses. This architecture operates three orders of magnitude faster than the biological real time.

9.1.1 System Implementation of VCNN Architecture

This section discusses the VCNN architecture briefly. Figure 9.1 shows a VCNN architecture setup that includes six CNL boards, computer (Host PC), Digital to Analogue Converter (DAC) and Oscilloscope. Initially, using the Host PC the FPGAs are programmed to implement the connectivity of a given network to facilitate spike routing. The DAC PCI card is used to supply the tuning parameter voltages that set the characteristics and weights of the synapses and the neuron types. These voltages can initially be programmed using the PC. The network connectivity and the tuning parameter voltages define the network on the VCNN architecture. The network emulation can start after configuring the network and the internal synaptic activities of the network can be recorded using the digitising oscilloscope. The spike activities can be recorded in the FPGA memory and could be acquired on to the PC using the USB interface while or/and after running the emulation of the network. The spike data can be processed within the PC to analyse the network activities.

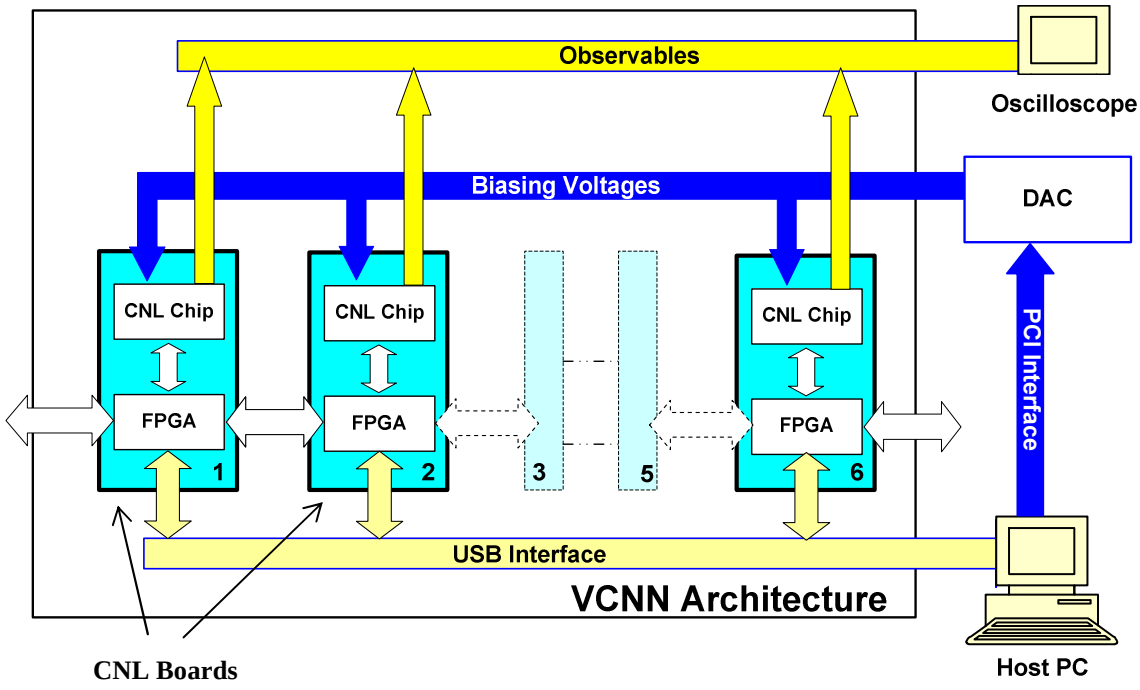


Figure 9.1 VCNN Architecture with six CNL boards.

Though the integration of six CNL boards is considered in the basic VCNN architecture, in practice, a few tens of CNL boards chips could be assembled to form an architecture where each cortical layer can be configured using many CNL boards, so that a larger six layer VLSI network can be constructed.

9.2 Cortical Neural Layer (CNL) Chip – Overview

The CNL Chip containing 120 cortical neurons, and 7 560 synapses has been fabricated in a standard 0.35 µm CMOS technology. The chip comprises generic neuron and synapse circuits with configurable neuronal connections. The neurons of the chip can be configured to different known types of neurons (discussed in Chapter 4). The chip is also equipped with different short-term and long term dynamics synapse circuits that include inhibitory, excitatory, facilitating and depressing and STDP dynamics (discussed in Chapter 5 and Chapter 6). The size of the chip is 24 mm^2 (6.78 mm by 3.58 mm), and it has 180 pins. Figure 9.2 and Figure 9.3 show the layout of the chip and a photograph of the fabricated chip respectively. Most of the neurons' outputs are available in parallel from the chip pins. Some neurons' outputs are accessible serially

and a few of the neurons are internally wired. The pre-synaptic spike inputs can be provided externally by addressing the synapses using the address bus of the chip. The internal states of the selected synapses can be calibrated and/or observed externally. More descriptions of the fabricated circuits are given in next section.

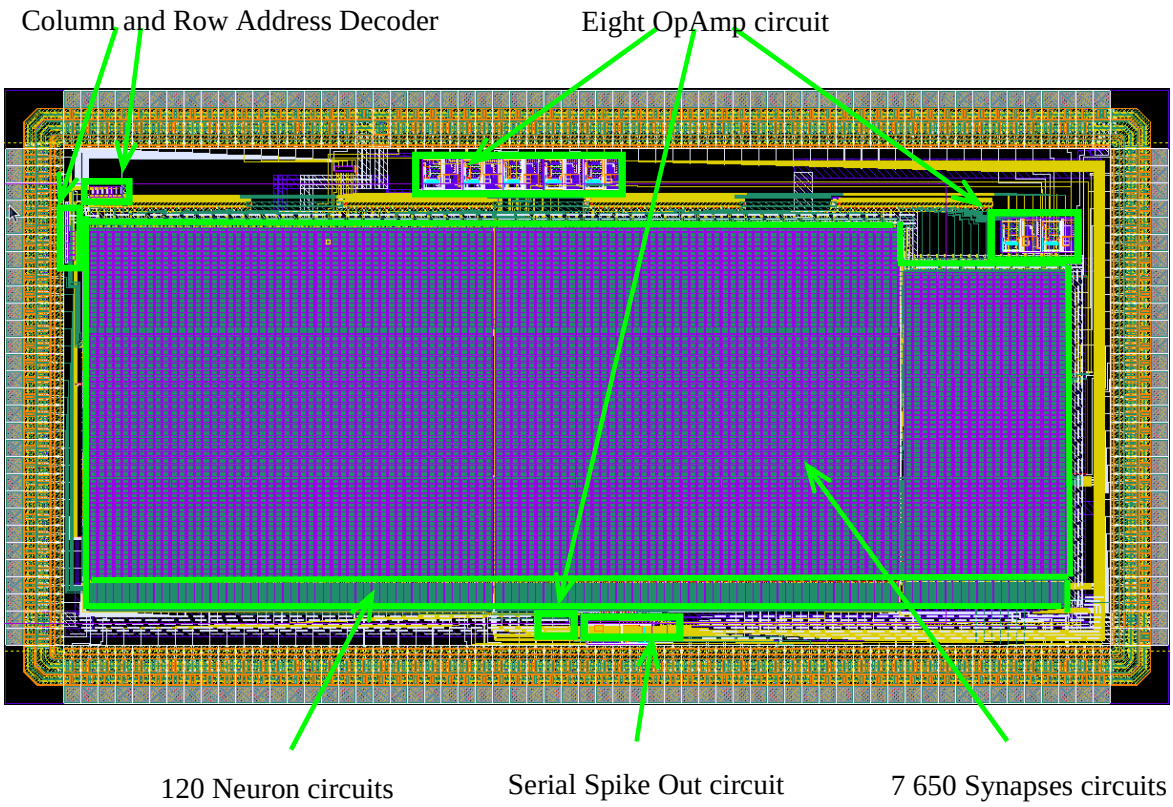


Figure 9.2 (a) Layout of the CNL Chip: 7 560 synapses, and 120 neurons and auxiliary circuits.

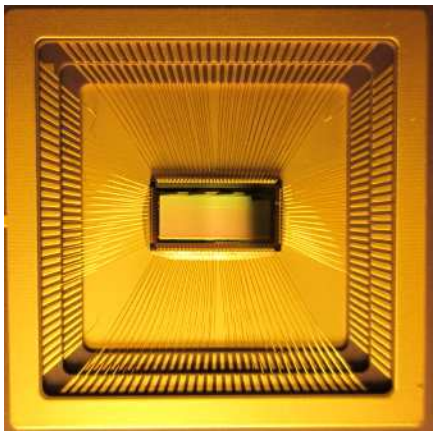


Figure 9.3 Picture of the fabricated Cortical Neural Layer Chip.

9.3 Neural element composition on the chip

The CNL chip can be configured to have a heterogeneous neuron and synapse type combination, such that it could closely represent the neuron and synapse type composition of the cortical layer of the neocortex. The neural elements occupy on the chip in two separate blocks: Block-A and Block-B.

9.3.1 The analogy to the neocortex

About 80% of the neurons in a cortical network the neocortex are excitatory neurons, and others are inhibitory neurons (Somogyi et al., 1989; White, 1989; Peter et al., 1984). Anatomically, these two types of neurons are equipped with different types of input and output synapse combination. By considering this, the CNL chip is designed to represent the excitatory and inhibitory neurons in the Block-A and Block-B neurons respectively. Figure 9.4 shows the neuron and synapse composition of the CNL chip, considering the inhibitory and excitatory representation of the blocks.

The output spike from an excitatory neuron excites the membrane potentials of post-synaptic neurons using excitatory synapses. Anatomically, most of these neurons receive synaptic inputs from non-STDP excitatory and inhibitory depressing synapses and from excitatory STDP synapses (Roth et al., 2009). Therefore, if the output of a Block-A neuron is connected to excitatory synapses, then the Block-A neuron closely represents an excitatory neuron of a cortical network.

The output spikes from an inhibitory neuron inhibit the membrane potentials of the post-synaptic neurons using inhibitory synapses. It is also known that, some of the inhibitory neurons receive inputs from inhibitory facilitating and excitatory depressing synapses, whereas some other inhibitory neuron types receive input spikes from excitatory facilitating and depressing synapses (Roth et al., 2009). Therefore, by choosing an appropriate input and output synapses combination the Block-B neurons can be configured to represent either of these two types of the inhibitory neurons.

Although the Block-A and Block-B neurons are equipped with a specific type of synaptic inputs to represent closely the anatomy of the excitatory and the inhibitory cortical neurons respectively, it is the selection of output synapse type that purely determines the excitation or inhibition effect on the membrane potential of the post-synaptic neuron (i.e. excitatory or inhibitory functionality of a neuron). Hence,

irrespective of the location of the neurons in Block-A or Block-B, the user has the freedom to configure any of these neurons to work as excitatory or inhibitory neurons, by projecting the neuron output to appropriate synapses.

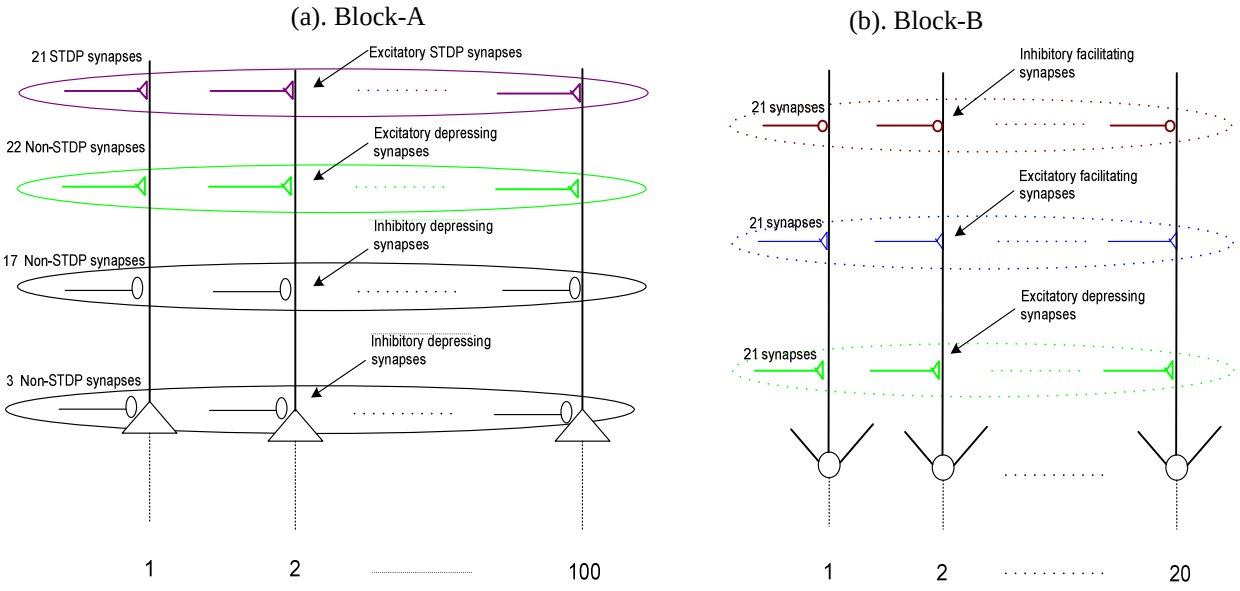


Figure 9.4 Description of a generic layer of VLSI cortical network model; 100 excitatory neurons and 20 inhibitory neurons with various types of input synapses.

9.3.2 The neural circuit composition on the chip

The Block-A consists of 100 neurons and 6 300 synapses. Each of the neurons in this block receives inputs from 43 excitatory depressing synapses (21 STDP and 22 Non-STDP excitatory depressing synapses) and 20 inhibitory depressing non-STDP synapses (3 somatic and 17 distal inhibitory synapses). The Block-B consists of 20 neurons and 1 260 synapses. Each of the Block-B neurons receives inputs from 63 non-STDP synapses. The 63 synapses comprise an equal number of excitatory facilitating, inhibitory facilitating, and excitatory depressing synapses. Layout of the Cortical Neural Layer Chip showing the physical location of the synapses is shown in Figure 9.5.

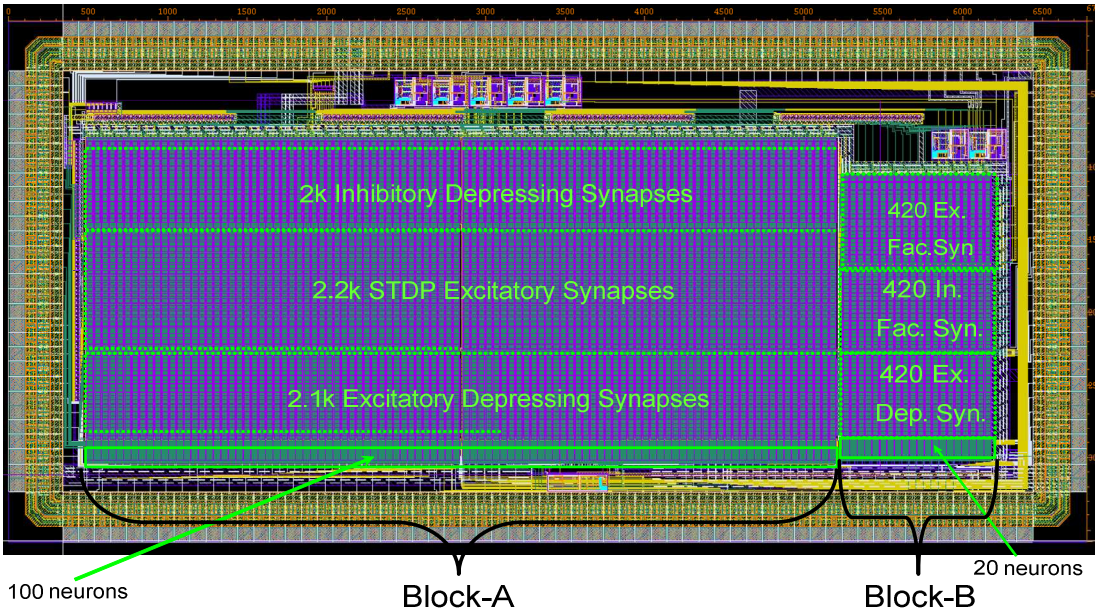


Figure 9.5 Layout of the Cortical Neural Layer Chip.

9.4 Circuit Implementations

The circuit implementation details of the CNL chip are given in this section. These include core neural circuits and auxiliary circuits of the chip. The auxiliary circuits provide circuits to configure and debug the neural elements and to interface the inputs and output responses of these neural elements with other microelectronic devices. The pre-synaptic inputs and spike outputs are accessible by an off-chip device (Xilinx Virtex 5 FPGA[17]) and the network connections are configurable using this off-chip device. The CNL chip composition of the neural circuits and its auxiliary circuits that include network configuration details and observable outputs are discussed in this section.

9.4.1 Neural Circuits

As mentioned above, the neural circuits uses in the CNL chip include Accelerated Time Cortical Neuron circuit (see Chapter 4), STDP Synapse Circuit (see Chapter 5) and four types of Short-Term Dynamic Synapse circuits (XD, IF, ID, and XF Synapse circuits; see Chapter 6). Each of these synapse circuit on the CNL chip include Spike Generator

[17] Xilinx Virtex 5 FPGA Web Link: http://www.xilinx.com/products/virtex5/

circuit (SG circuit; discussed in Section 9.4.2.1) in addition to their basic circuits described in their respective Chapters. Synapse circuit receives pre-synaptic inputs from this Spike Generator circuit. Each Spike Generator circuit has two inputs (row and column address enable) and once these are enabled using pre-synaptic address, a spike is generated to its connected Synapse circuit.

The same type synapse circuits in a block share the same set of tuning voltages; e.g. all the XD Synapse circuits in Block-A share one set of biasing voltages to set the parameters: the $V\Delta wp$ (controls the degree of depressing of the synapse), $V\alpha p$ (controls the degree of recovery of the depressing synapse) and *Vbp* (sets maximum cut-off value of EPSC) except the *Vwrp* of the synapse. The method used to set the parameter *Vwrp* of the synapses that set the resting weights is discussed in Section 9.4.2.2. The XD Synapse circuits in Block-B shares deferent set of voltages of the same parameter set.

Neural Circuits in a Block-A Neuron Unit

Each of the Block-A Cortical Neuron circuit receives inputs from 20 Inhibitory Depressing Synapse circuits, 21 Excitatory STDP synapse circuits and from 22 Excitatory Depressing Synapse circuits that form a Block-A Neuron Unit as shown in Figure 9.6. One hundreds of these common Neuron Units are generated to form the Block-A of the CNL chip.

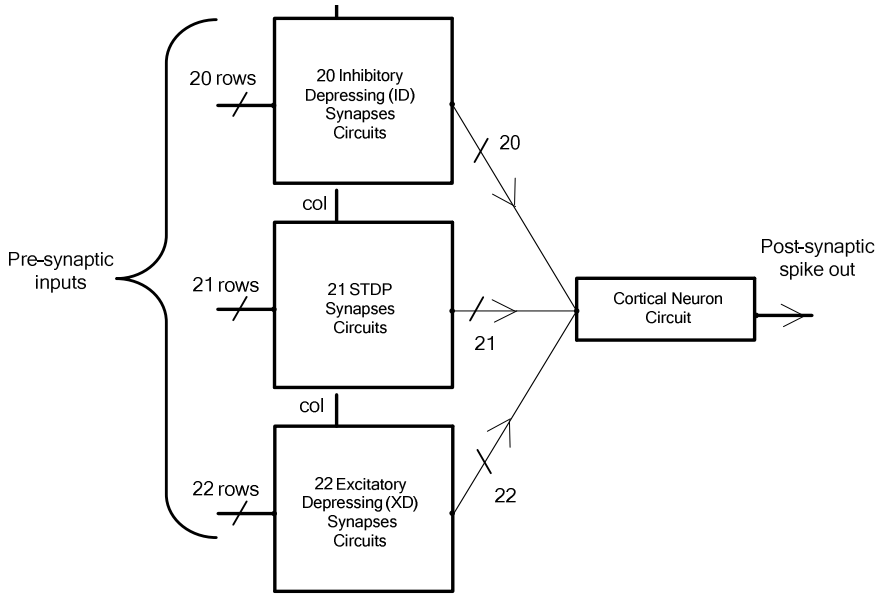


Figure 9.6 Block-A Neuron Unit and its input synapses.

Neural Circuits in a Block-B Neuron Unit

Each of the Block-B Cortical Neuron circuit receives inputs from three types of Short-Term Dynamic Synapse circuits. As shown in Figure 9.7, they are 21 Excitatory Facilitating Synapse circuits, 21 Inhibitory Facilitating Synapse circuits, and 21 Excitatory Depressing Synapse circuits. The Block-B neurons use the same Excitatory Depressing Synapse Circuits as in the Block-A neurons. A common structure of a Neuron Unit that comprises a Cortical Neuron circuit and its Short-Term Dynamic Synapse circuits that provide inputs is constructed and 20 of these Neuron Units are generated to form the Block-B of the CNL chip.

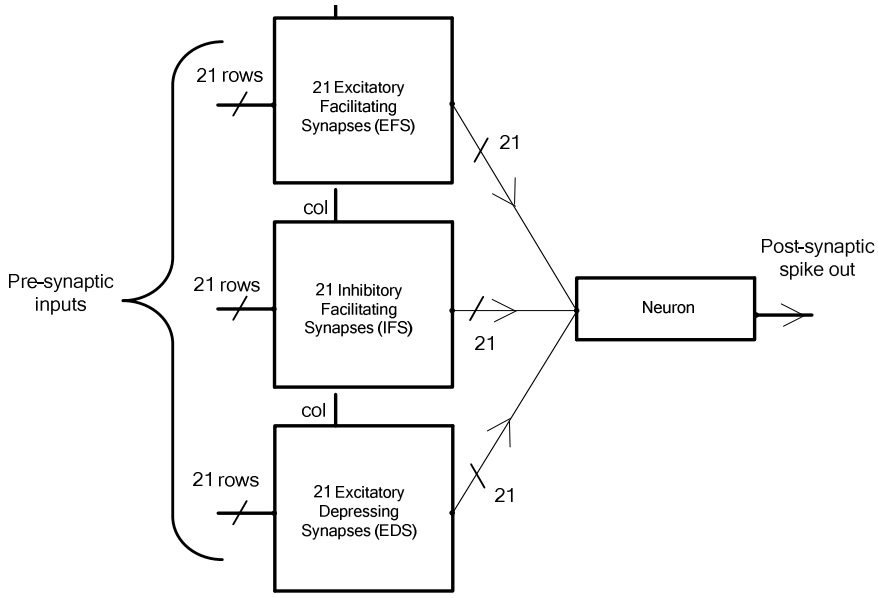


Figure 9.7 Block-B Neuron Unit and its input synapses.

9.4.2 Auxiliary Circuits

The core circuits of the CNLC are used as primitives when emulating a cortical network. The network configurations are set using an external device. Hence the auxiliary circuits of the CNLC facilitate interfacing neural signals to the external devices and the configurations and debugging circuits of the core circuits. The supplementary circuits include, Pre-Synaptic Spike Decoder, Neural circuit Configuration circuits, Post-Synaptic Spike Output circuit, and Debugging circuits. The pre-synaptic spike decoder circuit decodes the externally provided synaptic address

(*Pre_spike* address) to generate pre-synaptic spike (*Pre*) signal to the desired synapse/s. The neurons' outputs are available to the external device with the help of the post-synaptic spike output circuits. Further, the synapses and neurons are configured using bias voltages and some resistor divider circuits. The CNL chip also supports on-chip neural connectivity. The debugging circuit allows some of the internal voltages of a few synapses to be observed externally as well as to set the characteristic of the synapses. The following sections explain the operation of Pre-Synaptic Spike Decoder, Neural circuit Configuration circuits, Post-Synaptic Spike Output circuit, and Debugging circuits.

9.4.2.1 Pre-Synaptic Spike Decoder (PSSD) Circuit

As the neural network connections are configured off the chip, the Pre-Synaptic Spike Decoder (PSSD) circuit is used to route the input spikes to the intended synapses, decoding the incoming *Pre_spike* addresses. A synapse address is made up of the row and column numbers corresponding to the place the synapse occupies on the chip. Hence, the column number represents its post-synaptic neuron's address. The PSSD is comprised of Column Address Decoder circuit and Row Address Decoder circuit that share a common address bus (*Addr <1:14>*). At a given time, these decoders can select one or many synapses depending on the *Pre_spike* input. Figure 9.8 shows the block diagrams of these decoders. A crossbar grid layout that has 120 columns and 64 rows of metal conductors is formed by the output terminals of these decoders. At each cross point, a two input NAND gate receives the incoming column (*Cb_c*) and row (*Cb_r*) outputs of the decoders. The output of the NAND gate at a cross point is given to a Spike Generator (SG) circuit (except on the row 0, as addressing row 0 along with a column address is used to reset a neuron or group of neurons). Furthermore, the first 100 SG circuits in the row 43 are used to generate external post-synaptic spikes (*Post* signals) to debug the STDP circuits. Figure 9.9 (a) shows the SG circuit. The SG circuit is constructed using a Delay circuit, Inverter, NOR and NAND gate. The layout of the SG circuit is located close to its synapse layout to provide non-attenuated pre-synaptic spikes. When the decoders enable a SG circuit, the circuit generates approximately 5 ns *Pre* pulse to its synapse. Figure 9.9 (b) shows the input, output and intermediate signals of the SG circuit.

As mentioned above, the PSSD circuit can address a SG or a group of SG circuits simultaneously using a *Pre_spike* address. The *Pre_spike* address is defined with the doubled size of the synaptic address. It includes synaptic address bits and their corresponding “don’t care” address bits. If a “don’t care” address bit is set to logic one then its corresponding synaptic address bit is ignored. Hence, two synaptic addresses that match excluding the ignored bit’s (ignored bit is at "don’t care") are addressed simultaneously. Similarly, many “don’t care” address bits can be set with logic one to send pre-synaptic spikes to a larger group of synapses simultaneously. The Don’t Care Addressing (DCA) circuit is used to provide the addressing and Figure 9.10 shows the circuit of one bit DCA element of the PSSD circuit and its truth table. The outputs of the DCA (DCA<0:6> or DCA<0:5>) are given to the address de-multiplexer circuit (DEMUX7 or DEMUX6) to generate the crossbar signals.

Address a synapse or a group of synapses

An external device, preferably an FPGA, provides the *Pre_spike* address to the address bus of the CNL chip in three stages. Firstly, the Data Flip-Flops (DFF) of the Column Address Decoder circuit latch the column address of the *Pre_spike* by providing a pulse to the strobe signal, *Stb1_c* after providing the column address onto the address bus. Secondly, both *Stb_r* and *Stb_c2* signals of the PSSD are given a short pulse soon after the row address of the *Pre_spike* address is given to the bus. This starts decoding the column and row numbers of the address using Column Address Decoder circuit and Row Address Decoder circuit respectively. Finally, the decoded addresses are sent to the crossbar simultaneously to enable the intended SG circuit/s by providing a pulse to the *Stb*. Figure 9.11 shows the timing diagram of the *Pre_spike* addressing.

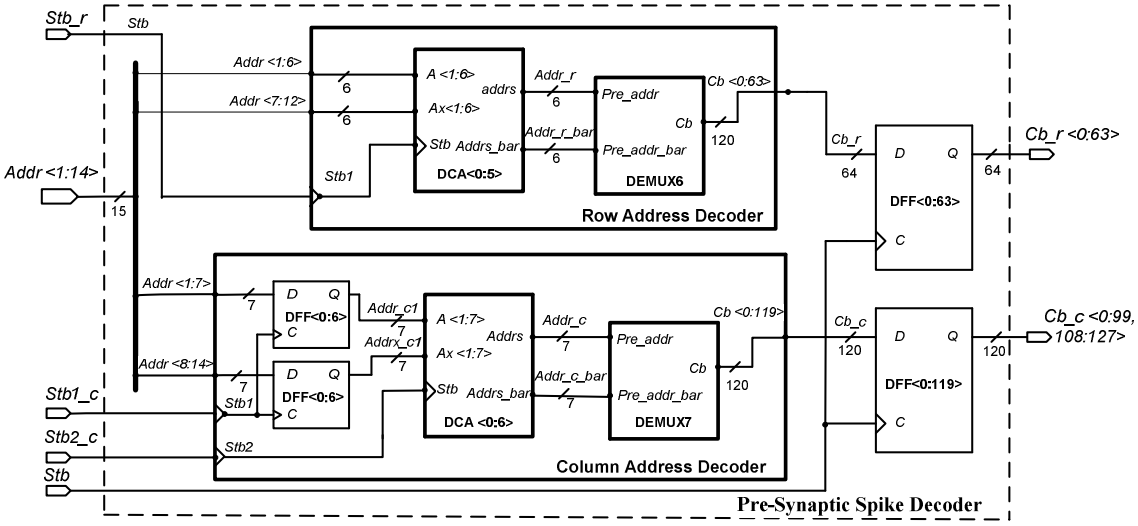


Figure 9.8 Block diagram of the Pre-Synaptic Spike Decoder (PSSD) circuit.

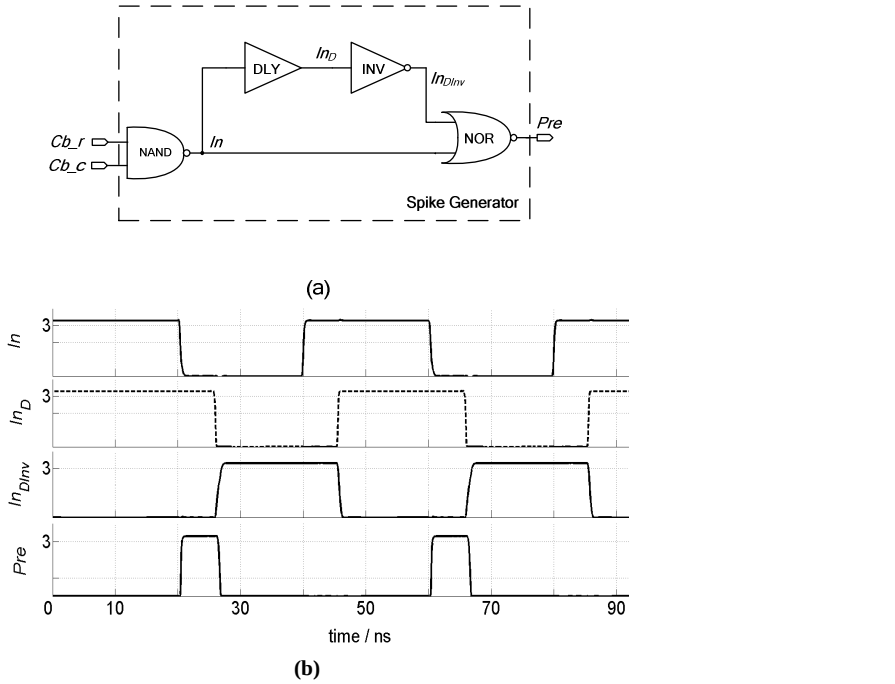
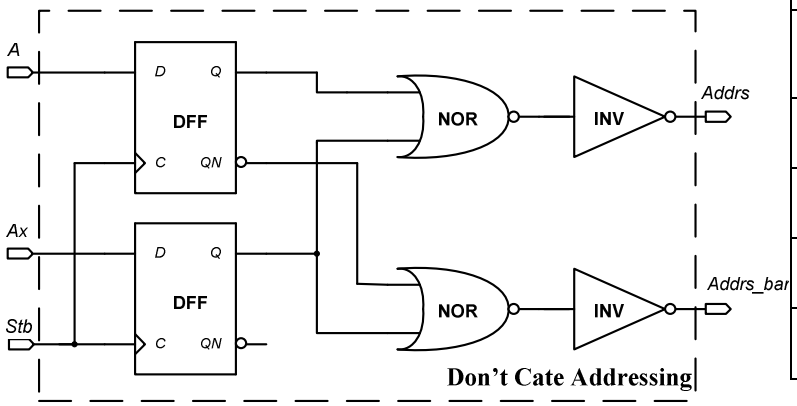


Figure 9.9 (a) Spike Generator (SG) circuit and (b) its timing diagram of the SG circuit.



(a)

Input		*Addrs_bar*	*Addrs*
A	*Ax*		
0	0	0	1
1	0	1	0
0	1	1	1
1	1	1	1

(b)

Figure 9.10 (a) Don't Care Addressing (DCA) circuit, and (b) its truth table; Array of DCA circuits is used in Row Address Decoder and Column Address Decoder circuits of the PSSD circuit.

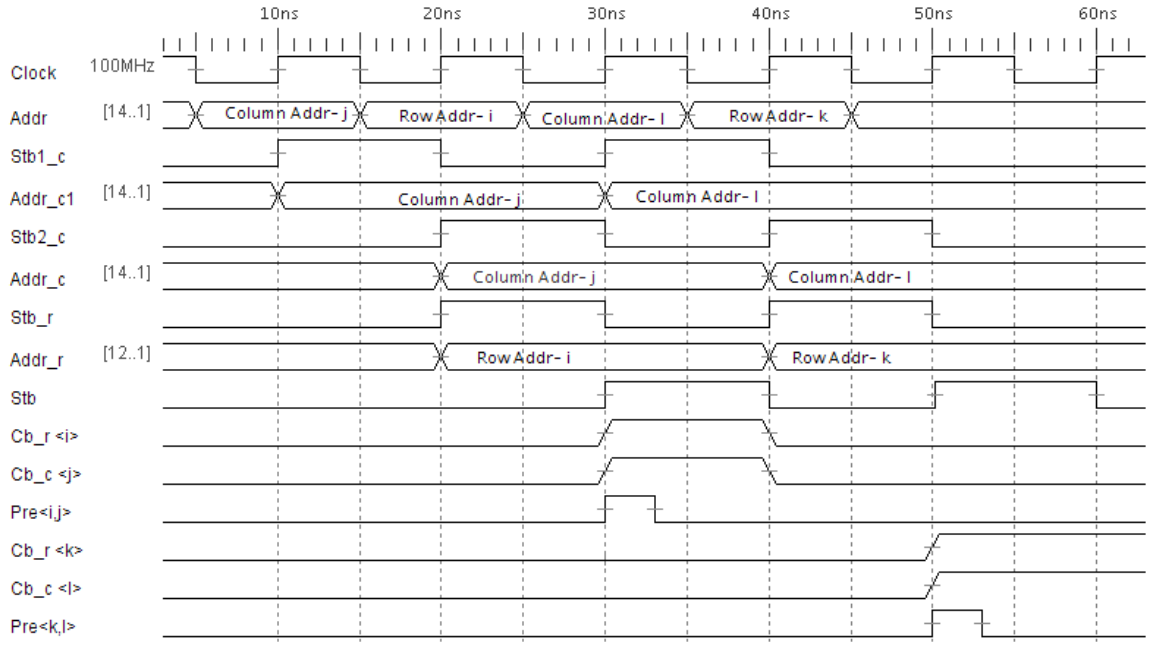


Figure 9.11 Timing diagram of the PSSD shown in Figure 9.8.

9.4.2.2 Neural Circuit Configuration Circuits

The neural circuit configurations include configuration of resting weight of the synapses, groupings of neurons and on-chip neuron projections (i.e. internal network connections). Brief descriptions of these are given below.

Configuration of Resting Weight of the Synapses

In this section, synaptic weight, *W* is referred to the voltage that sets the resting weight control voltage of a non-STDP synapse (*Vwrp* or *Vwrn*). The non-STDP synapses require setting their synaptic resting weights when mapping a cortical network on the chip. These synapses include 5 356 synapses, which belong to seven groups. The XD and ID synapses of Block-A and XD, IF and XF synapses of Block-B receive their pre-synaptic inputs from an off-chip device. Hence, these synapses are called *externally connected* non-STDP synapses. The synapses of the other two groups (XD-i and XD-e) are used to route spikes internally- they are called *internally connected* synapses. The synaptic weight configurations of these two types are given below.

Resting weight configuration of the externally connected non-STDP synapses

Synaptic weights of the synapses of a group can have linearly distributed weights along a column. However, along a row the weights remain constant. The circuit shown in Figure 9.12 is used to provide the linearly distributed weights. The externally provided voltages W_B and W_A are used to set the linear distribution as seen in Figure 9.12 (b). The nominal resistance of the resistor R is equal to 6.425 kΩ. The Poly-2 resistors are used in the resistor divider to achieve a higher resistance in a compact design. Once the two ends of the distribution are set, the weight of the ith synapse (W_i) can be obtained from the following equation.

$$\left.\begin{aligned} W_i &= W_A + \Delta w \times (i - \frac{1}{2}) \qquad \text{for} 1 \le i \le 20 \\ W_{21} &= W_B ; \quad \Delta w = (W_B - W_A) / 20 \end{aligned}\right| \qquad (9.1)$$

Where, W_A and W_B are externally set voltages; For ID synapses group, W_{21} is irrelevant as it has only 20 synapses.

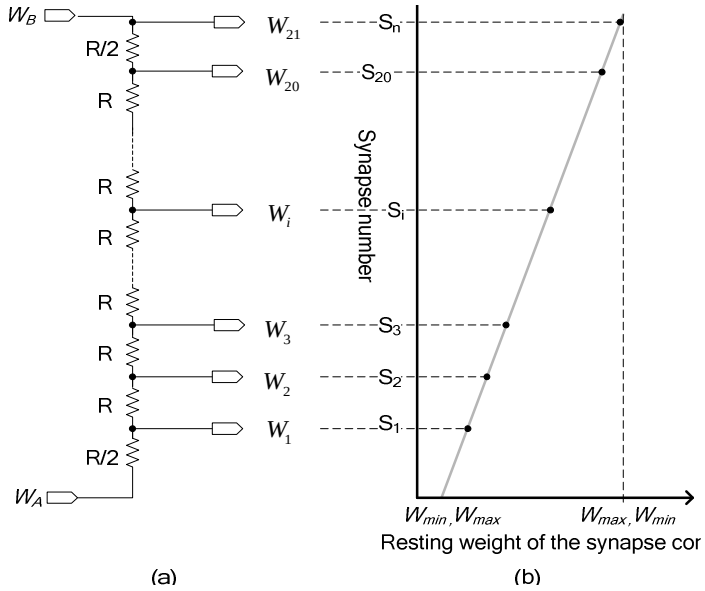


Figure 9.12 (a) The Resistor divider circuit is used by each of the XD, ID, XD, IF and XF synapse group, R = 6.425 kΩ, (b) the generated weight distribution line.

Resting weight configuration of the internally connected synapses

The XD-i synapses that connect Block-B neurons to Block-A have a linear weight distribution along the row (the chip has 64 synapses such that occupy on the row one of the CNL chip). The resistor divider circuit used by XD-i is given in Figure 9.13. The weight of the j[th] synapse (W_j) in the XD-i group can be calculated using the equation given below.

$$W_j = \begin{cases} W_C + \Delta w \times j & \text{for} \quad 1 \le j \le 32 \\ W_C + \Delta w \times (j - 32) & \text{for} \quad 33 \le j \le 64 \end{cases} \qquad (9.2)$$

Where, $\Delta w = (W_D - W_C)/32$ and W_C and W_D are externally set voltages;

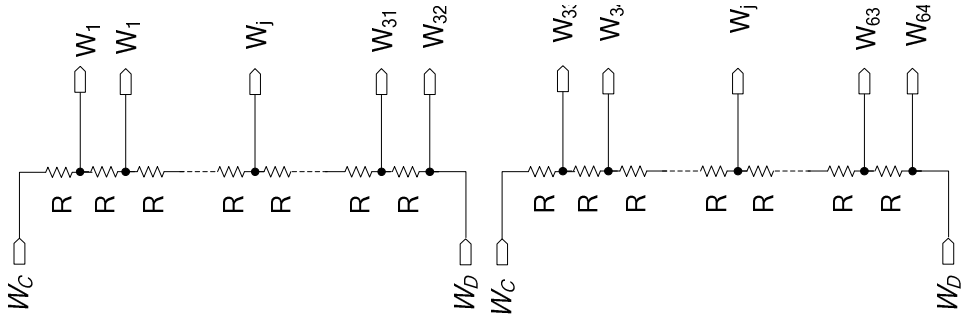


Figure 9.13 Resistor divider circuit used by XD-i synapse type in Block-A to receive a linear weight distribution across the column: R = 4.25 kΩ.

Further, all the XD-e SYNAPSES (32 synapses on row one of the CNLC) that connect Block-A to Block-A neurons can only have the same weight value that can be set externally.

Groupings of neurons

Neuron circuit can be configured to different cortical neuron types (RS, CH, IB, etc.) by tuning the parameters (*Vc* and *Vd*) of a neuron appropriately as given in Chapter 4. Independent configuration of 120 neurons requires extra circuits and consumes extra chip area. Neurons in a cortical network are composed of larger groups of the same neuron type. Hence, 120 neurons are grouped into 13 groups where each group can be configured to a given neuron type independently. These groups are made up of different numbers of neurons so that a group that contains the closest number of neuron to a required number in a type can be constructed by combining many groups. Table 9.1 shows the number of neurons in a group and their addresses.

Table 9.1: Independently configurable groupings of neurons in Block-A and Block-B

Block-A			Block-B		
Group size	No. of groups	Addresses of the neurons	Group size	No. of groups	Addresses of the neurons
20	3	15 to 34	8	1	108 to 115
		40 to 59			
		65 to 84			
10	2	00 to 09	6	1	116 to 121
		90 to 99			
5	4	10 to 14	4	1	122 to 125
		35 to 39			
		60 to 64	2	1	126 to 127
		85 to 89			

On-chip neuron projections

Though most of the network connections are configured off-chip, 1.2% synapses are hard wired internally to test the prototype with internal and external connections. Eventually, this could provide more options to configure network connections. The outputs of 16 neurons out of 120 are connected internally. The hard wired connections are of two types, and they are given below. These connections can be switched -off or -on depending on the network specifications.

Block-A to Block-A neuron projections

Outputs of eight Block-A neurons (with addresses 16 to 19 and 66 to 69) project internally to thirty two other neurons in Block-A via excitatory depressing synapses (XD-e). As given in Table 9.2, each output projects to four consecutive neurons. These synapses can have equal weights that can be set externally.

Block-B to Block-A neuron projections

Outputs of eight Block-B neurons (with addresses 100 to 107) project internally to sixty four neurons in Block-A, via XD-i synapses. As given in Table 9.2, each output projects to eight Block-A neurons. These synapses' weight can have a linear distribution as discussed in Section 9.4.2.2.

Table 9.2: Hardwired projections of neurons

No.	Neuron projections Block-A to Block-A		Neuron projections Block-B to Block-A	
	From neuron	**To neurons**	**From neuron**	**To neurons**
1	16	0 to 3	100	18 to 25
2	17	4 to 7	101	26 to 33
3	18	8 to 11	102	34 to 41
4	19	12 to 15	103	42 to 49
5	66	50 to 53	104	68 to 75
6	67	54 to 57	105	76 to 83
7	68	58 to 61	106	84 to 91
8	69	62 to 65	107	92 to 99

9.4.2.3 Post-Synaptic Spike Out circuit

The spike outputs of 104 neurons in the CNL chip are sent to an off-chip device, whereas the other 16 neurons' outputs connect internally. 84 of the 104 outputs send to output chip pins directly, through the output buffer circuits. These can be read by an off-chip device in parallel. These neurons include the neurons with addresses 0 to 15, 20 to 54, 79 to 99, and 108 to 119. The rest of the twenty outputs and two parallel connected outputs are sent serially. The two signals that outputs through both parallel and serial ports are use to debug the Serial Spike Out circuit. Description of the Serial Spike Out circuit that is used to interface output of the twenty two neurons is given below.

Serial Spike Out (SSO) Circuit

The twenty two neurons that can read their outputs using SSO circuit include the neuron addresses 55 to 65 and 70 to 80. The SSO circuit has three terminals, *Out_bit*, *Clk* and *Se*. Spike read-out operation of the SSO circuit is explained below.

Spike Read-Out from SSO Circuit

An external device can read the neurons' outputs serially from the *Out_bit* terminal providing a suitable clock (*Clk*) and scan-enable (*Se*) signal to the SSO circuit. The SSO circuit comprises of 22 bits Parallel Input Serial Output (PISO) shift register, 22 Set/Reset Latch (SRL) and Reset Pulse Generator circuits as shown in Figure 9.14. When a neuron fires, SRL is set and latched until the Data Flip Flop (DFF) on the PISO shift register updates with the SRL state within the refractory period of the neuron. The SSO circuit works synchronously with the rising edge of the clock and has two phases; the spikes update phase (*Se* = '0') and the serial scan phase (*Se* = '1'). At the spikes update phase, the PISO shift register updates with the spike data received from SRLs, whereas in the scan phase the external device can read the twenty two spikes serially. Soon after the scan phase is started, the SRLs are reset using the Reset Pulse Generator circuit. The Reset Pulse Generator circuit generates a reset pulse at the rising edge of the *Se* signal. The circuit comprises of Buffer, Delay circuit, Inverter, and AND gate as shown in Figure 9.15.

The 25 MHz clock and repetitive scan-enable signals as shown in the timing diagram in Figure 9.16 can be used to read the spikes periodically. It is also possible to use a higher clock rate than 25 MHz. However, the scan phase requires 21 clock cycles, and for the rest of the time the device should be put into the update phase. The update phase should last for at least one clock cycle within 1 μs so as not to lose any spikes.

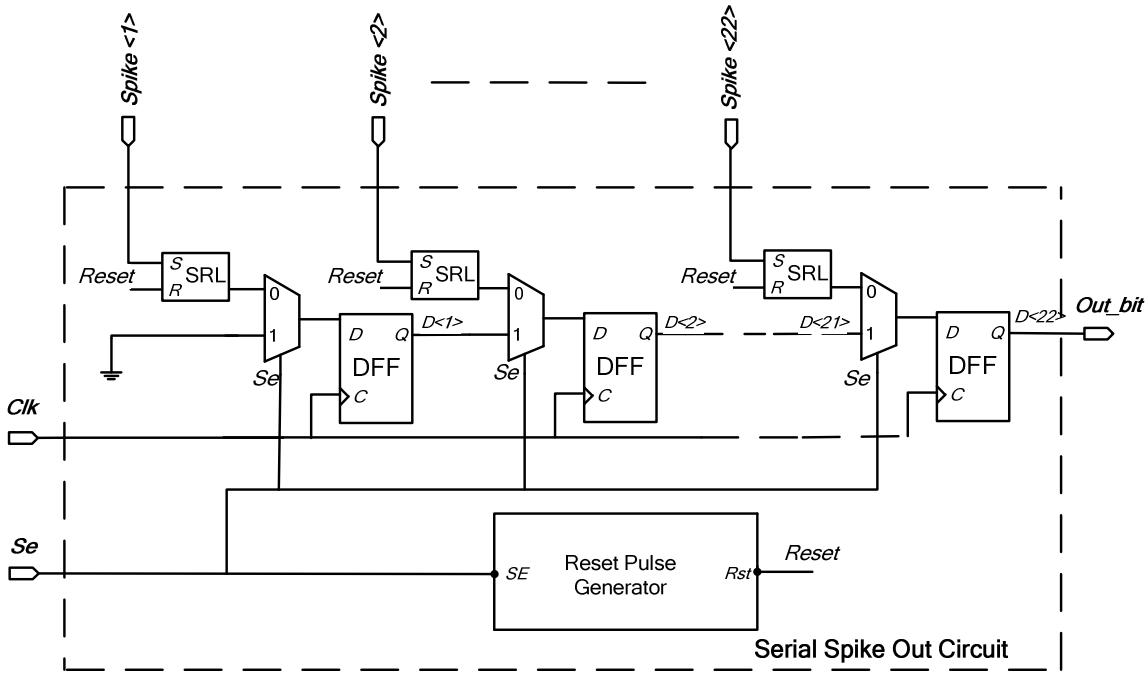


Figure 9.14 Block diagram of the serial spike out unit.

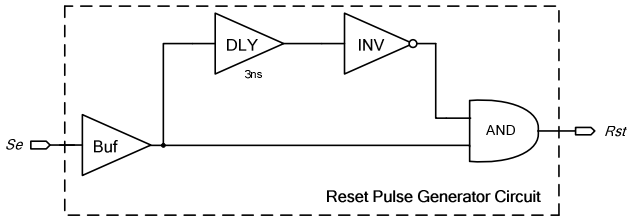


Figure 9.15 Reset Pulse Generator circuit.

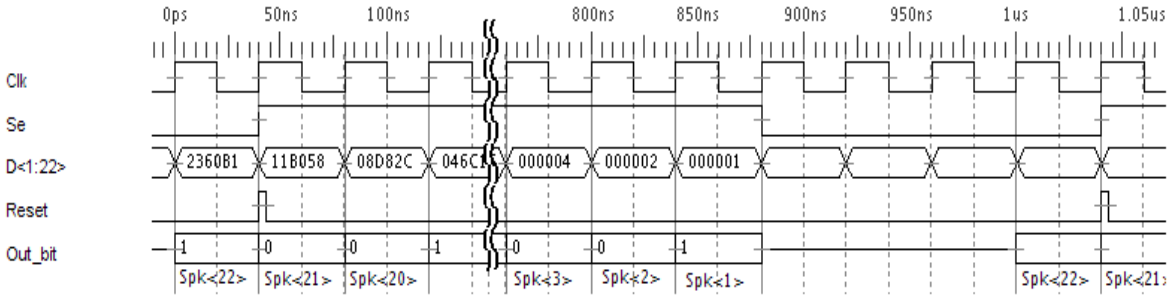


Figure 9.16 Timing diagram to obtain 22 neurons' spike output data from SSO circuit serially.

9.4.2.4 The debugging circuits

The debugging circuits are used to test the synaptic states for different bias voltages and parameters as well as to observe their states while emulating a cortical network.

The testing of a STDP synapse may include observing the synaptic dynamics with an externally provided post-synaptic signal (*Post*) rather than using the output of the post-synaptic neuron. It is also possible to reset the weight of the STDP synapse. The observable internal signals of a STDP synapse include *ltp*, *ltd*, *Wstdp*, *w*, and *Vsyn* (see Chapter 5). For non-STDP synapses the facilitating or depressing weight signals (*Vw*) can be observed off the chip. These signals can be used to obtain the on-chip characteristics of the synapses by sweeping the parameters of the synapses, and to facilitate the testing circuits if required.

Due to the limitation of the chip area and the output pins, the signals of only a few selected synapses of only eight neurons from each Block can be observed off the chip. The observable synapses associated with the Block-A and Block-B neurons are the neurons with the addresses 50 to 57 and 112 to 119, respectively. Further, internal signals of each synapse from each main synapse type are observable. That is, from each of ID/IDS, STDP and XD types in the Block-A neuron, and from each of IF and XF/XD types of the Block-B neuron. The addresses of these synapses that belong to each neuron can be obtained from Table 9.3. The cross marks in Table 9.3 indicate the synapses and their observable synaptic signals for a given neuron address.

Each of the observable signals is initially buffered using a source follower circuit (a two transistor NMOS or PMOS circuit). Then the eight inputs from similar synapses of eight neurons of a block are multiplexed using an 8:1 multiplexer. Finally, the multiplexed output is given to an OpAmp before providing the output to the chip pins. The OpAmp circuit works in a unity gain voltage follower configuration.

To observe the signal/s of a synapse, the test address (*Tadd<0:2>*) of the synapse's neuron should be sent to the address of the 8:1 multiplexer. This is done using the common address bus (*Addr*). Similar to pre-synaptic addressing, the three test address bits along with the two flags, the *Post_sel* and *Wrst* are also provided to the address bus, followed by a pulse to the Strobe signal, *Stb_t*. This latches the bits on to five DFFs (Data Flip Flops). The three DFFs that latch the address bits are connected to the address of the 8:1 multiplexer. Hence the address to switch the output signals is

decoded. Once a test address is provided, all the synaptic signals of the Block-A and Block-B neurons belonging to the test Address can be observed simultaneously. The test addresses corresponding to the neurons addresses can be obtained from Table 9.3. Setting the *Post_sel* flags (*Tadd<3>*) switches the *Post* signal of the STDP synapses to an externally provided *Post* signal (the external *Post* signal is provided using the row 43 address in a similar way as the *Pre* signal is provided) rather than to its post-synaptic neuron's output, to facilitate debugging and the measurements of the STDP circuit. The Flag *Wrst* (*Tadd<4>*) is used to reset the STDP weight to zero.

Table 9.3: Observable signals of synapses, their address and respective test address

Block-A Neurons												
Synapse Type	**Synapse Row**	**Test Address (*Tadd<0:2>*)**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Output Buffer Type**	**Output OpAmp Label**
		Neuron (col) / Test Signal	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**		
ID	**63**	*Vw*	x	x	x	x					PMOS	*Vw1*
IDS	**44**	*TVwID*					x	x	x	x		
STDP	**42**	*ltp*	x	x	x	x	x	x	x	x	PMOS	*Ltp*
		ltd	x	x	x	x	x	x	x	x	NMOS	*Ltd*
		Wstdp	x	x	x	x	x	x	x	x	PMOS	*Wstdp*
		w				x	x				NMOS	*Vsyn_w*
		Vsyn	x	x	x			x	x	x		
XD	**21**	*Vw*	x	x	x	x	x	x	x	x	NMOS	*Vw2*
Block-B Neurons												
Synapse Type	**Synapse Row**	**Test Address (*Tadd<0:2>*)**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Output Buffer Type**	**Output Signal Label**
		Neuron (col) / Test Signal	**112**	**113**	**114**	**115**	**116**	**117**	**118**	**119**		
IF	**63**	*Vw*	x	x	x	x	x	x	x	x	PMOS	*Vw3*
XF	**42**	*Vw*	x	x	x	x	x	x			NMOS	*Vw4*
XD	**21**	*Vw*							x	x		

9.5 CNL Chip- Model

An approximate mathematical model of the CNL chip is formulated to simulate approximated network behaviour of the chip. Figure 9.17 shows the block diagram of

the mathematical simulation setup of the CNL chip. The model has been implemented in MatLAB[18]. The Simulation Core uses mathematical models of neuron and synapses. It comprises of the same number of neurons and synapses and with the same structure as that of the CNL chip. The mathematical model of the neuron is given in Chapter 4 and the synapse mathematical models are given in Chapter 5 and Chapter 6. The pre-synaptic inputs to the Simulation Core are provided using the matrix **PreSpike [63 x120]**. The spike outputs from the neurons are continuously updated on to a matrix **Fire [120x1]**. The parameter matrix, **Para [60x1]** is used to tune the characteristic of the synapses and the neuron types. The user can set this matrix according to the characteristic of neural elements of a given network. The incoming spikes from the neurons of the Simulation Core are routed to the synapses of the Simulation Core using the Route Spikes function. The Route Spikes function uses the Connectivity Matrix map that defines the network connectivity of a given network. Hence, before simulating a network, the user needs to set up the Connectivity Matrix and the Parameter Matrix that define the given network.

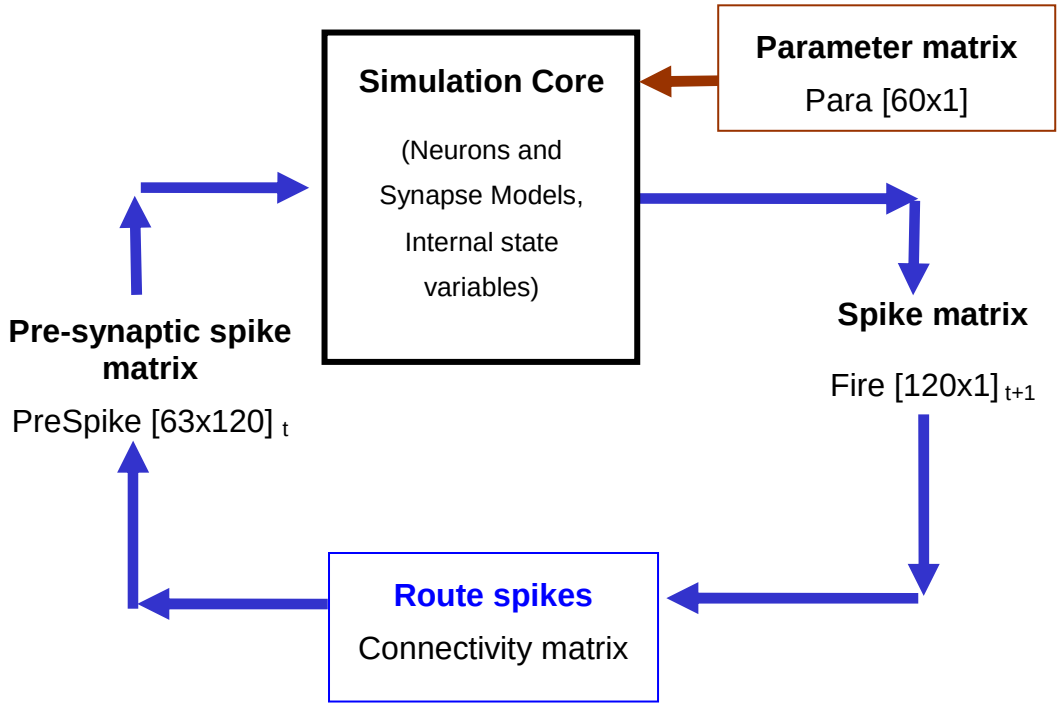


Figure 9.17 Cortical Neural Layer Simulation Core.

[18] MATLAB Web Link: http://www.mathworks.co.uk/

Some simple models so far tested by the users include Synfire Chains (Grossberg, 1969). Winner-take-all network (Redgrave et al., 1999), and AND, OR and XOR gates (Agmon-Snit et al., 1998).

9.6 The CNL Board

The CNL board consists of the CNL chip board and a Xilinx Vertex 5 FPGA board (Xilinx[16]; Opal Kelly[19]). Figure 9.18 shows the block diagram of the CNL board. The 180 pin CNL Chip is bonded on to a printed circuit board (PCB), and an FPGA board is connected onto the PCB to form the CNL board. The analogue biasing voltages that tune the parameters of the CNL chip are routed to connectors on the board. The tuning parameters can be provided using externally programmable voltage supplies to pre configure the characteristics of the synapses and neuron types on the VLSI cortical neural layer. The eight analogue output pins of the CNL chip are used to observe the internal states of the synapses are also wired onto connectors on the board. All the digital input and output pins of the CNL chip are connected to the FPGA directly. These include 14 bits address bus, 5 bits strobe signals, 84 bits parallel spike outputs of 84 neurons and 3 bits to access serial spike outputs of 22 neurons. The address bus including strobe signals are used to send pre-synaptic spikes to the synapses, to select observable internal states of the synapse and to test the STDP synapses (see Section 9.4.2.1 and Section 9.4.2.4 for more details). At a given time pre-synaptic spikes can be sent to one or many synapses that allow dense network connectivity with reduced latency.

[19]Opal Kelly Xilinx Virtex 5 FPGA board Web Link: http://www.opalkelly.com/products/xem5010/

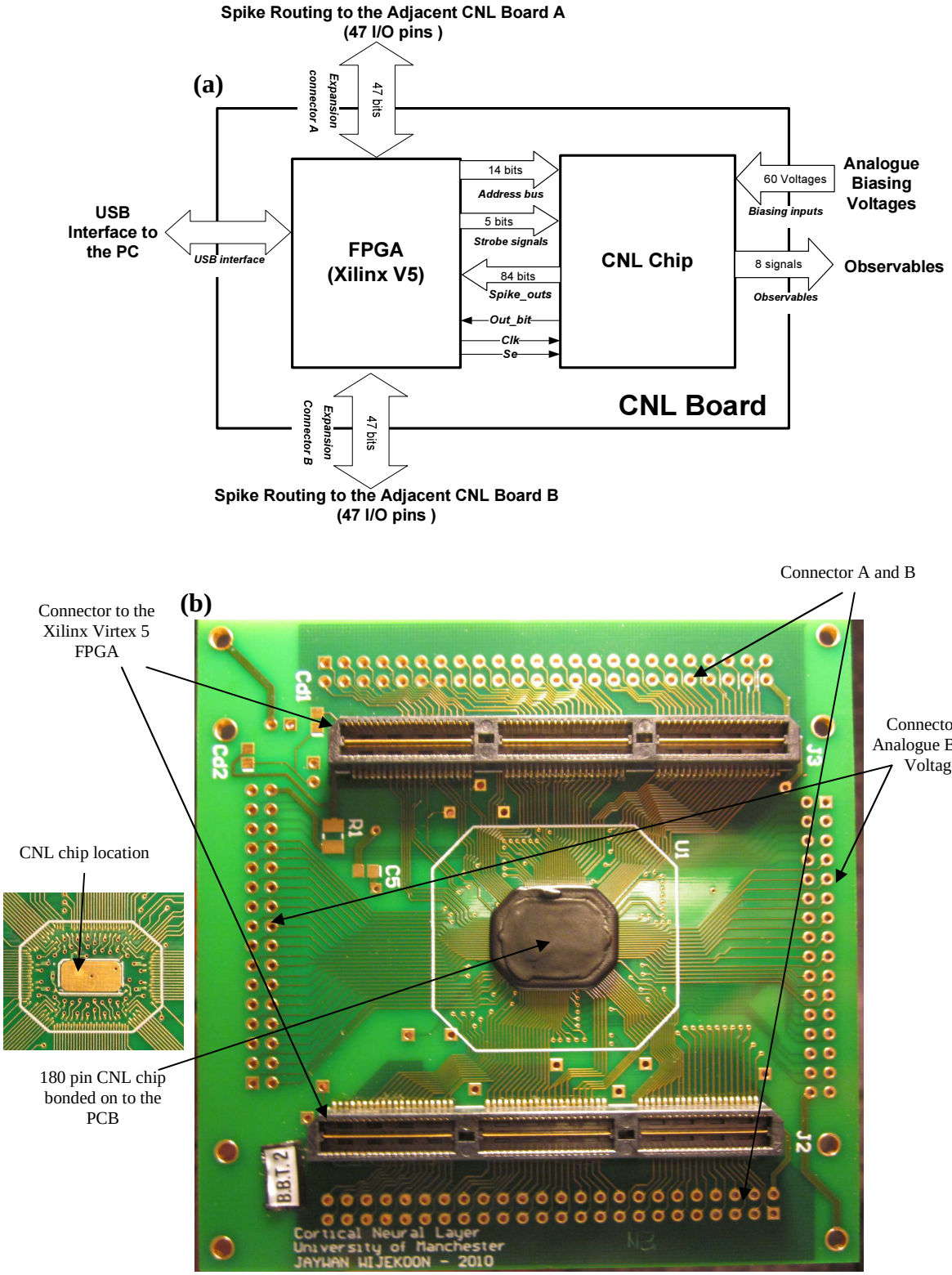


Figure 9.18 Cortical Neural Layer Board. (a) Block diagram (b) Actual Board (work-in-progress CNL board, this work is continuing under EPSRC funding)

9.6.1 Operation of the CNL board

In emulating a network in the CNL board, the network needs to be configured. This requires pre configuring the network connectivity and presetting of the tuning parameters of the network. The Routing Module that performs the spike routing and the LUT that defines the connectivity of a given network, are implemented on the FPGA. The LUT needs pre configuring according to the connectivity matrix of the network. The tuning parameters that set the properties of the neural elements need to be supplied by the programmable voltage suppliers. These configurations define a network on the CNL board. Furthermore, required observable outputs can also be set using the address bus of the CNL chip.

The CNL spike outputs that include both parallel and serial spike outputs are connected to the input channels of the FPGA. Once a neuron fires, the input stage of the FPGA detects an arrival of a spike on its input channel. This generates an address event that triggers the request to perform routing of spikes. The FPGA Routing Module accepts an incoming request and performs the routing of the spike to its connected synapses by generating the pre-synaptic spike address to the CNL chip. Once the request is served (routing is performed) the request is acknowledged and the next incoming request on the queue is performed by the Routing Module. In performing these routing operations, the predefined network connectivity table, the LUT, is used to obtain the connected synapse addresses of a neuron. Simultaneously, these spike activities can be recorded on the memory and read by the PC using the USB interface to analyse the spikes. Figure 9.19 shows the test setup of the CNL board.

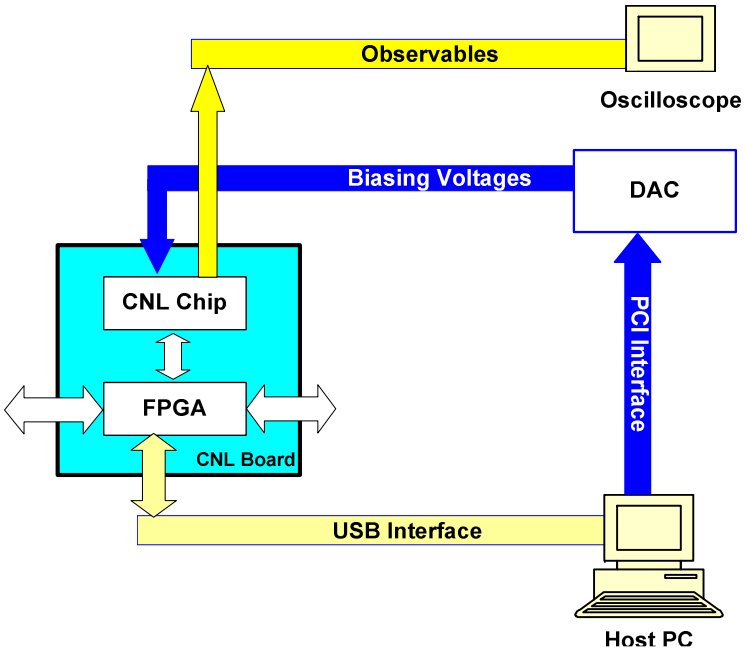


Figure 9.19 Cortical Neural Layer Board.

9.7 Discussion and Conclusions

The fabricated generic neural layer prototype IC assembles the basic neural elements with diverse properties and composition as of the neural layer of the neocortex in a small scale. A VLSI Cortical Neural Network architecture that combines many CNL chips to build a small scale cortical network is proposed, and it could provide a platform to study the network behaviours and learning. Approximated mathematical model of the CNL chip is formulated to simulate the approximated behaviour of a network in software. This helps to understand and possibly to reduce the network mapping problems before configuring the network on the CNL ICs. In addition to the configurations of neural dynamics of neurons and synapses discussed above, the delays of signal propagation due to the location of the synapse on the dendrite can be modelled on the FPGA using delay blocks.

Mapping cortical network on to the CNL chip

The CNL chip is equipped with many configuration options such as on chip and off chip network connections, different types of synapse and neuron configuration, etc. On the other hand, it also has limitations such as characteristics of synapse or neurons are set in groups, synapse can only have fixed linear pattern of weight distribution configuration, etc. These require an extra effort in mapping a given network on to the CNL chip.

Therefore, automating the network mapping task will ease the usability of the CNL chip. This could be done by formulating a generic algorithm that implements most obvious problems of resource mappings using a rule set, and complex and non-trivial resource mappings using the trial and error method. In arriving at optimised mapping parameter set using trial and error techniques require performance figures of network mapping in order to judge the performance of mapping. The performance figures may depend on many factors such as the network specifications, the expected behaviour of the network, etc. Example of performance figures could be the percentage of connections mapped, percentage of neuron types satisfied, percentage of STDP synapses used, the spike routing time, closeness of the output results of the network, etc. Many performance figures need to be considered in arriving at optimised mapping parameter set and these figures could have weightings depending on the level of significance on the network performance. Hence the approach like the Balance Scorecard[20] could be used. For example, in a network percentage of STDP synapse mapped on to the chip could be more important than a percentage of excitatory depression synapses mapped, depending on the network behaviour (e.g. STDP learning) tested on the chip which requires higher weight to the earlier factor than later.

Cortical Network Connectivity and Spike Routing Latency

The fact that a spike from a neuron takes time to reach its connected synapses is referred here as the spike routing latency. The network connections are implemented external to the CNL chip, and the time taken to route a spike needs to be within the acceptable range. The timing delay of the spike outputs of the CNL chip to reach the FPGA is short as the spikes are recorded in parallel. However, depending on the network configuration, the spike congestion due to higher network activity could slow down the spike routing from the FPGA to the synapses. If the activity level of the network is higher than the spike routing speed of the router on the FPGA, there is a chance that incoming spikes will be queued on the FPGA memory, resulting in longer spike routing latency. This problem could increase further if many CNL chips are connected together as it increases the number of spikes to be routed. In reducing the FPGA routing load, the

[20] Balance Scorecard Web Link:
http://www.balancedscorecard.org/BSCResources/AbouttheBalancedScorecard/tabid/55/Default.aspx

CNL chip use the "don't care" addressing of the synapses that allows simultaneously addressing of many synapses using one cycle of addressing (discussed in 9.3.2.1). The chip also implements neuron to synapses fixed connections that reduce the external spike routing load. These internal connections facilitate one neuron to many synapse projections (discussed in 9.3.2.5). The spike routing latency of internal connections is significantly lower than the external routing latency.

Spike routing delay shorter than the refractory period of the neuron (1 µs in accelerated time) is required to provide integration of post-synaptic currents on to the membrane of the post-synaptic neuron properly. The calculation of number of spikes (at full load) that can be routed within the refractory period of the neuron is provided below.

Assuming that all the neurons are fast spiking neuron type and spike continuously at 200 MHz (200 Hz in biological time), i.e. at full load, where all the neurons are spiking approximately at their highest spike rate:

Average spiking rate of all the neurons = 0.2 MHz

Average spike rate of N neurons = N x0.2 MHz

Clock Speed of the FPGA = CLK_{FPGA}

Number of cycles per pre-synaptic spike addressing = CYC

Number of input neurons that can be served

within the refractory period = CLK_{FPGA}/ (200 000 CYC)

If FPGA clock rate is 100 MHz and spike addressing takes two clock cycles, assuming one pre-synaptic addressing per spike is needed to define the connectivity of the network (i.e. all the synapse groups can be selected using the "don't care" bits address patterns), and not considering the serial reading delays of the serially read spikes.

250 fast spiking, 200 Hz (biological time) continuously firing neurons can be served by the router. This is sufficient to combine six CNL boards to form a VCNN architecture, where one third from the full network connectivity (total number of neurons 720) could be satisfied even with the full load activity level. The full load activity level is highly

unlikely as cortical network comprises of RS, IB and CH firing patters, FS neurons could not fire continuously for a longer period, and the activity levels of the networks are believed to be well below 10%.

Importance of the STDP synapse and its memory retention,

The STDP synapse implementation proposed in this thesis lack the long-term memory retention but provides a compact circuit implementation. Other approaches of implementing STDP synapses that retain memory for a long period include use of floating gate transistor synapses (Hasler et al., 1999) and digital memory synapses (Schemmel et al., 2008) but these implementations require larger silicon area. The memristor nano-scale device operates similar to a STDP synapse and can keep the memory for a long period. The emerging VLSI technologies that incorporate fabrication of memristor device could make easier to fabricate massively parallel large cortical neural networks in hardware.

Experimenting cortical network models

The proposed network architectures can be used to configure small neural network models and the scaled down versions of a large scale network models (Redgrave et al., 1999; Riesenhuber et al., 1999; Grossberg, 1969; Agmon-Snit et al., 1998 Traub et al., 2005; Häusler et al., 2006; Goldberg et al., 2004; Basalyga et al, 2010; Stein et al., 2008). The Matlab model of the CNL chip has been provided to computational neuroscientists in the COLAMN project (Institute of Computational Neuroscience, Plymouth), and to postgraduate students in the School of Electronic and Electrical Engineering and the School of Life Science in the University of Manchester to test models in numerical simulation before implementing them on to the hardware.

Finally, the results obtained using the hardware model could be benchmarked with the computational models and then with the biological recordings. This will provide insight into the closeness of the neural dynamics of the circuit implementations. Further, these could help to predict some biological behaviours of the network by observing the dynamics of the hardware implemented models.

CHAPTER 10 : MIMICKING CORTICAL NEURAL NETWORK IN HARDWARE –A DISCUSSION

This chapter discusses the feasibility of implementing larger cortical neural network architectures in hardware. The complex non-linear nature of neural response, heterogeneity of the neural elements, the complexity of neuron connectivity and the practical limitation of maximum silicon area of a chip, limit the maximum size of a cortical network in silicon. Implementations of complex non-linear computational models consume larger silicon area of a chip. Therefore, there is a trade-off between the extent to which the VLSI circuit element can be made biologically plausible and the size of the cortical network that can be implemented in VLSI.

Continuing the effort of implementing larger cortical network in hardware, the initial section of this chapter provides estimations of VLSI cortical network sizes that can possibly be built utilising the latest technologies. This includes VLSI cortical network size estimates in a large chip fabricated in CMOS technology 90 nm, in multi-chip approach, and in wafer-scale integration technology that uses the accelerated-time core VLSI neural circuits used in CNL chip. Here, the estimate only down to the 90 nm CMOS technology is used as the design of analogue mixed signal circuits in the technologies that have smaller feature sizes is challenging, and the network size estimations may be not realistic. Further benefit of using 3D integration technology to build the cortical network is also discussed. Although a network could be implemented in these estimated sizes, other factors that could provide problems and limitations are also discussed. The later part of the chapter presents alternative approaches that could be used to mimic cortical networks. Finally, the higher abstractions of neural dynamics used to obtain brain-inspired computing models are outlined.

10.1 Estimates of VLSI Cortical Network Size

The proposed cortical networks include richer neural dynamics than the other VLSI network implementations found in the literature. This section analyses the feasibility of implementing larger scale cortical neural network in VLSI using the accelerated time neuron and synapse circuits proposed in Chapter 4 to Chapter 6. The CNL chip has been fabricated in a standard 0.35 µm CMOS technology, and includes 120 neurons and

7 560 synapses. The CNL chip uses generic neural elements that can be configured to a variety of neuron and synapse types of a cortical network layer. Further, a VLSI Cortical Neural Network architecture that accommodates multi-chips to build a larger network that comprises 2 400 neurons and 151 200 synapses has been proposed in Chapter 9. By continuing this analysis, this section provides possible size estimations of the VLSI cortical network implementations in 120 mm^2 chip area, (corresponding to a relatively large die size, which can nevertheless be fabricated with a good yield on a modern standard CMOS 90 nm process technology), multi-chip integration and wafer-scale integration. Further, the advantages of using 3D-VLSI integration are also briefly discussed.

Using a straightforward area scaling from the implemented CNL chip, it is estimated that the 120 mm^2 VLSI chip in 0.35 µm technology can accommodate up to 600 neurons and 38 000 synapses. Migration of 0.35 µm technology analogue circuits into 90 nm technology require redesigning the analogue neural circuits proposed in Chapter 4 to Chapter 6. Redesigning these circuits requires taking into account high sub-threshold leakages and mismatch problems that are inherent in deep sub-micron circuit implementations. Here, a conservative area scaling factor is used to estimate the equivalent silicon area consumption in deep sub-micron implementation, and it is estimated that in a 90 nm CMOS technology, 120 mm^2 chip can accommodate approximately 5 000 neurons and 300 000 synapses. The multi-chip approach that uses twenty 120 mm^2 VLSI chips in 90 nm technology could accommodate 100 000 neurons and 6 million synapses. A hypothetical deep submicron wafer-scale integration of a system on a 12” wafer can accommodate approximately 2.8 million neurons and 180 million synapses (This assumes very low scaling factor include additional penalty for increased routing). These network size estimation calculations are provided in Appendix B. The largest network of wafer-scale and multi-chip integration requires high bandwidth communication architecture (GHz range) to communicate between neurons. Brain has evolved in three-dimensional space, where close by neurons have been connected to each other via synapses with remarkably dense connectivity patterns to form a 3-dimensional cortical network. Though the VLSI networks can be implemented on a large scale the complexity of inter-neuron connectivity limits the network models that can be configured to be emulated by these microelectronic devices. Therefore, the larger networks require the use of digital technology to route spikes

between distal parts of the network in order to reduce the inter-neuron communication problem. Architectures like Network on Chip (Plana et al., 2007) could be used to facilitate the multi-chip communication requirements. The larger wafer-scale integration network proposed by Schemmel et al. (2008, 2009) uses hierarchical digital communication architecture to support the high bandwidth of inter-neuron communication.

On the other hand, the 3D-VLSI techniques are becoming available that stack many layers of silicon circuits to form a three dimensional (3D) chip[21]. These technologies can be used to reduce the inter neuron connectivity constraint. The size of the network that can be implemented in a chip can be increased several times due to the increase in silicon area. The VLSI 3-D integration technologies are at their early stage, though they provide higher density of wiring and fast signal propagation. Currently, these technologies cannot provide the full random 3-dimensional connectivity such as that of neurons in the neural tissue, although the density of connectivity between circuit elements has been improved greatly. Since the circuit is very densely packed, power dissipation may impose a limit on the size of the network that could be implemented using these technologies.

10.2 Limitations of VLSI Cortical Network

Although the efforts are made to propose larger VLSI cortical networks, they can only be built with relatively basic models, limited in size, and with reduced flexibility. Further, the contemporary science also lacks the full understanding of the cortical network dynamics. Therefore, cortical network experiments in larger implementation of VLSI cortical network could encounter the following difficulties:

- Calibrating the VLSI cortical network

Hardware neural network model is built with approximations and with reduced flexibility. On the other hand, biological neural responses are highly heterogeneous. Hence, the optimal configuration that should be used in an experiment is non trivial and in most of the cases, setting-up the initial conditions and calibrating all the neural

[21]3D-IC Alliance http://www.3d-ic.org/

elements to an optimal set of parameters that suits the given network model may not be trivial. Comparing VLSI network results and biological network results in order to calibrate the neural elements is also not straightforward due to the highly non-linear nature of the relation between the characteristics of the neural elements and the dynamics of the network.

- Mapping a cortical network model onto the hardware

Mapping larger cortical network models onto the hardware may require approximations, simplifications, and scale reductions. It is not a straight forward task to find the best mapping that satisfies the hardware constrains. The necessary modifications may not be valid and due to the non-linear dynamics it may not be possible to find the best approximated, or/and scaled network model that could be mapped to the hardware.

As the neural elements are heterogeneous and complex in their connectivity, the unavailability of standard method of classifying neural elements and simplifying connectivity patterns makes the approximating and/or reducing the network size a challenging task.

There are attempts by the neuroscience community to standardise the neural classification methods (e.g. Pettila convention[22], see Markram, (2006)). The standardising of the neural classification methods is difficult due to the unavailability of the pool of known data and as some types of neurons and synapses are yet to be found. Recently pooling of known neuron and synapse data have been begun (e.g. Neuromorpho.org[23]) and data is publicly available, which might help to formulate universal classification methods for the neural elements once a sufficiently large data set is accumulated. Methods of simplifying network connectivity need better understanding of the cortical network dynamics, and these could evolve hand in hand with the evolution of the hardware emulating platforms.

---

[22] Petilla Convention (2005)

Web Link: http://krasnow.gmu.edu/cng/petilla/

[23] Neuromorpho.org

Web Link: http://neuromorpho.org/neuroMorpho/index.jsp

- Variability and noise of the hardware

Though it seems possible to compensate for the variability and noise of the hardware by properly configuring the cortical network, computational principles that facilitate fault tolerance in a network are yet unknown. Therefore, at this stage the variability and noise in VLSI hardware could not be properly analysed to obtain the desired cortical network dynamics. However, by emulating different cortical networks on hardware the effect of the variability and noise on cortical network dynamics could be studied.

- Acquiring and processing of data

There is a limit to which the VLSI network data could be observed. Access to all the internal variable of the neural elements may not be available due to the hardware constrains and the chips usually provide only the spike outputs of the neurons rather than instantaneous status of the state variables of all neurons and synapse. Even so, these data may reach a very larger volume, which requires high computing power to analyse.

- Benchmarking of VLSI hardware results with the biological recordings

The procedure for validating the closeness of the network results against the result of computational models and then against the biological data is not obvious and there is no standard method for doing so this is true even for validating computational models against the biological data. However, it can be hoped that the standard method for benchmarking hardware neural accelerator platforms become available with the evolution of the hardware neural accelerator platforms.

- Lack of availability of promising cortical network models

Most of the known cortical network models address the cortical networks of early sensory processing stages. These preliminary stages of sensory processing perform simple feature detections and do not perform promising cognitive processing. For example, the most researched area of the neocortex is the area V1 which belongs to the

visual cortex; area V1 performs preliminary stages of visual processing that include directional selectivity, orientation selectivity, binocular disparity, etc. The intelligent complex processing is believed to be done in higher cortical areas of the brain and how the information is processed is still a mystery in Neuroscience.

Furthermore, due to the limitations of acquiring data from biological systems, the available models represent a small part of the cortical networks, with many assumptions and approximations. These network models provide only simple oscillation or wave propagation behaviours and do not demonstrate useful intelligent information processing phenomena (Traub et al., 2005; Häusler et al., 2006; Goldberg et al., 2004; Basalyga et al, 2010; Stein et al., 2008). However, experimenting with different hypothetical network models may help to improve the understanding of the cortical network dynamics.

10.3 Alternative Approaches

This section discusses some alternative approaches to mimicking the cortical networks in hardware platforms. These include the need for alternative IC fabrication technologies, the utilisation of a memristor device as a synapse, and an approach based on cell cultures. Finally, the scientific approach that can be used to develop brain-inspired computing models at a higher abstraction of neural network dynamics is outlined.

10.3.1 Alternative IC fabrication technologies

As discussed in Chapter 1, there is a need for a VLSI technology that can mimic cortical neural network architecture optimally. The circuit should be able to accommodate cells with 3-D connections, while the requirements on the speed, preciseness and the minimum feature size can be more relaxed. The possible alternative approach that could provide these properties includes 3D VLSI technologies, such as Plastic (or organic) VLSI that uses organic materials to build 3D or 2D circuit elements. Further, invention of a dedicated neural element as a device rather than constructing these using transistors would reduce the complexity of large scale network implementations (Linares-Barranco et al., 2009; Yajie et al., 2007).

10.3.2 Memristor as a synapse

A fabrication of the memristor devices is an emerging technology that could substitute the STDP synaptic circuit with a nano-scale device (Linares-Barranco et al., 2009). The memristor theory was first formulated by Chua (1971). The memristor is a two dimensional circuit element that has the characteristic of resistance change due to the history of its current flow and the potential difference across the terminals that can be described as a functional relationship between charge and magnetic flux. This characteristic can be used to implement STDP rule with non leaky long-term memory retention while occupying only a few square nanometres of chip area. For comparison, the proposed STDP circuit (Chapter 5) occupies six orders of magnitude larger circuit area while not providing adequate long-term memory retention. The memristor technologies are being developed (e.g. Hewlett-Packard lab) at many research institutes and are not yet available as a generic technology to implement circuits. Once these devices are available, replacing the STDP circuits of CNL with memristors would make a large-scale massively parallel cortical network in hardware more feasible, as some of the area constrains in implementing cortical network in VLSI technologies would be relaxed.

10.3.3 Cell culture

Another approach includes culturing of biological neural networks to study the network behaviours (Wagenaar et al., 2006; Stegenga et al., 2008). This uses biological neural tissues taken from animal brains, cultured on top of multi-electrode arrays that allow communication to and from the cultured neural network. These networks are typically flat single layer sheets, which are limited to networks with two-dimensional connectivity.

10.3.4 Higher abstractions of neural dynamics

Although this thesis uses neural level abstraction to build and understand biological systems, it is an open question, which level of abstraction neural circuits should be modelled to yield a practical computational architecture. The “bottom-up” approach that links the neuron-level activities to higher-level actions such as decision making, storing

memories, experiencing the world, emotions, etc. as well as its opposite “top-down” approach is popular in brain science as a way to reverse engineer the brain.

Most of the models that abstract higher level neural network dynamics are efficient and cost-effective to implement in hardware, though they tend to deviate away from the biological plausibility. On the other hand, as presented in this thesis, neural level abstraction of circuits requires higher computational power to simulate or emulate small neural network, though, in comparison, it is more biologically plausible. It should be also noted that whether biologically plausible or not, if the brain-inspired system could perform intelligent processing, it would be an excellent achievement.

Some of the higher level abstracted models of cortical processing presented in the literature include Kalman filter neural model (Rao et al., 1996), the Bayesian neural model (Lee et al., 2003), Factor graphs (Bishop, 2006), LaminART model (Grossberg, 2007) and LISSOM Model (Miikkulainen et al., 2005). These models demonstrate some basic functions such as autonomous following of an object, predicting future probabilities of an action, object recognition, etc. These models are not capable of performing sophisticated intelligent functions as in biological systems but are capable of far more complex functionalities than the neuron-level models.

Once the underlying principles of brain computation are known, the abstract models implemented in hardware could be more efficient and cost effective.

CHAPTER 11 : CONCLUSION

Generic compact VLSI implementations of neural circuit elements that can be used to mimic the functions of a cortical network have been designed and fabricated in a standard 0.35µm CMOS technology. The VLSI network can be used as an emulation platform to research into the potential capabilities of a cortical network in performing real-world psychophysical tasks. The accelerated-time network implementation saves time in performing an experiment (e.g. hours of biological network simulation could be performed in seconds), providing considerable savings in the case where parameter sweeps are an essential part of the experiment. The key contributions of this thesis include:

- implementation of silicon cortical neuron with the lowest reported energy consumption per spike, in a generic and compact form;
- mixed signal VLSI implementations of the Izhikevich neuron, first time in the neuromorphic research, both in accelerated time and biological time implementations;
- first hardware implementation of the dopamine modulated synapse;
- implementation of novel, compact, short-term plastic VLSI synapse circuits;
- implementation of the configurable mixed signal VLSI cortical network integrated circuit with the most diverse neural dynamics that include diverse nonlinear neuronal responses and most of the short- and long- term plastic synapse types.

The accelerated-time VLSI neural circuits designed and fabricated include a compact cortical neuron circuit, two different long-term plastic synapse circuits and four different short-term plastic synapse circuits. Further, a biological-time cortical neuron circuit with similar dynamics as of the accelerated-time neuron is designed to demonstrate the feasibility of migrating accelerated time circuits into a biological-time domain implementations, which could be used to build biological time cortical network that has applications such as the real- time, sensory signal processing.

The neuron circuit is capable of replicating many known types of cortical neurons, simply by tuning two external voltages. The neuron reproduces biologically plausible

action potentials. The spiking and bursting firing patterns observed in cortical neurons such as regular spiking, fast spiking, chattering and intrinsically bursting and other complex activity patterns can easily be reproduced. The circuit uses only 14 transistors and is extremely compact. It consumes about 8pJ per spike and hence consumes low energy per experiment. Therefore, this circuit is an attractive candidate for building a massively parallel VLSI cortical neural network that incorporates diverse nonlinear neural responses, which seems essential for producing brain like computation.

The STDP and the Dopamine Modulated STDP synapse circuits that demonstrate long-term plasticity dynamics have been designed and fabricated in hardware. The STDP dynamics of the STDP circuit follows an approximated STDP curve to arrive at compact design. STDP synapse's amount of weight change due to LTP and LTD and the time windows of the LTP and LTD can be configured independently. This circuit can be incorporated in a cortical network to facilitate the learning and memory of a network. However, as the circuit holds the synaptic weight using a capacitor the operational time of an experiment is limited. The dopamine modulated synapse circuit is implemented based on the computational model proposed by Izhikevich (2007). This circuit has been designed by extending the STDP synapse circuit to facilitate regulation of eligibility traces based on the dopamine concentration. The changes in an eligibility trace due to LTEP or LTED, and the time windows of the LTEP or LTED can be configured independently, and the dopamine concentration can be generated globally using an external voltage bias or using a burst of spikes. This circuit can be used to provide the reinforcement learning in a VLSI cortical network. In the case of DA modulated synapse, the use of capacitors to hold the memory traces does not directly limit the operational time of an experiment. These long term synapse circuits can only be used in small-scale cortical network implementations.

Excitatory depressing, inhibitory facilitating, inhibitory depressing, and excitatory facilitating synapse circuits that demonstrate short-term plasticity dynamics have been designed and fabricated in VLSI hardware. The strength of depression or facilitation and the time constant of the recovery can be configured independently using externally controlled tuning voltages. The post-synaptic current can be scaled using an externally adjustable bias voltage. Accommodating inhibitory and excitatory synapses in a network provides stable network activities with rich network dynamics. Incorporating

facilitation and depression of the synapse circuits in a VLSI neural network provides dynamic gain-control inherent in the biological cortical networks.

To prove the concept in VLSI, different combinations of these accelerated-time neural circuits have been fabricated in a standard 0.35 µm CMOS technology. These include the Cortical Neuron Chip, STDP-DA Synapses Neuron Chip, and Cortical Neural Layer Chip (CNL chip). The former two ICs are designed to test the function of the basic neural elements, and the CNL chip is designed to facilitate cortical network emulations. The Cortical Neuron Chip has been fabricated with 202 cortical neurons, and the neuron function is experimentally verified. The STDP-DA Synapses Neuron Chip has been fabricated with two cortical neurons and 28 STDP/Dopamine modulated synapses to test the functionality of long-term dynamics synapses. These two fabricated chips have been tested, and the functions of cortical neuron and STDP synapse have been experimentally verified. The CNL chip that has the neural circuit composition similar to the cortical layer of the neocortex has been designed with 120 cortical neurons and 7650 synapses, and its design and implementation details have been provided. Further, the approximated mathematical models of the chip elements have been formulated to build a chip simulation platform that could test an approximated behaviour of the cortical network implemented on the CNL chip in software.

A Cortical Neural Network Architecture that utilises several CNL chips to build a cortical network of neocortex has been proposed. Finally, estimations of VLSI cortical network sizes that could possibly be built in the latest silicon technologies have been provided. The estimations suggest that a wafer-scale integration of a system on a 12" wafer in 90 nm technology could accommodate approximately 2.8 million neurons and 180 million synapses, if the composition of the basic circuit blocks is similar to the CNL chip. The largest network of the wafer-scale integration requires a high-bandwidth communication architecture to communicate between neurons. At full load, the largest fraction of the power consumption of the system would be consumed by the communication architecture (most likely to be in kilowatts range). Therefore, design of the communication architecture needs a great attention. The limitations in configuring cortical networks in large neuromorphic hardware have been discussed.

Building a large-scale network in VLSI that mimics the full cortical network of a primate brain requires novel compact synapse devices with long memory retention and

low power, and a dense synaptic wiring mechanism. In order to accommodate a larger network in a portable integrated circuit, the size of the synapse circuit should be on the nanometre scale or even smaller. In order for the power consumption not to exceed kilowatts of power, the synapse device should only consume a few pico Joules of energy per synaptic transfer operation. Integrating these devices in a 3D integrated technology reduces the dense inter-neuron connectivity problem. As the precision of a synaptic transmission in a neural network is not critical, attention on accuracy of the process parameters of the nanometre technology, where the synaptic devices are fabricated, can be relaxed. Investigating into a nanometre synapse device in 3D VLSI technology that satisfies the above criteria would be a promising research direction in implementing the next generation of neuromorphic devices.

Although the neuromorphic devices are at an early stage of evolution, these systems can be used as an emulation platform to support understanding of the processing principles of the cortical network. However, the lack of promising cortical network models makes it difficult to utilise the ability of neuromorphic hardware in intelligent processing tasks. This VLSI cortical network design exercise has emphasised the physical hardware constraints that computational models should take into account in formulating computational models of cortical networks. Finally, the design and implementation exercises and experiments in VLSI cortical network help to develop intuitive understanding of the models and behaviours which will pave the way towards developing future technologies that could build low-power real-time intelligent control systems for real-life applications, including the nature-inspired intelligent computing machines.

REFERENCES

Abbott L. F., Nelson S. B. (2000), “Synaptic plasticity: taming the beast”, Nat. Neurosci., pp.1178–1183.

Abbott L. F., Varela J. A., Sen K., Nelson S. B. (1997), “Synaptic depression and cortical gain control. Science”, pp.220–223.

Agis R., Ros E., Diaz J., Carrillo R., Ortigosa E. M.(2007), “Hardware event-driven simulation engine for spiking neural networks”, International Journal of Electronics, 94(5), 469-480.

Agmon-Snit H., Carr, C. E., Rinzel J., (1998), “The role of dendrites in auditory coincidence detection”. Nature 393, 268–272.

Arthur J. V. and Boahen K., (2004), “Recurrently Connected Silicon Neurons with Active Dendrites for One-Shot Learning”. In IEEE Int. Joint Conf. on neural Networks, Vol. 3, pp. 1699-1704.

Arthur J.V. and Boahen K. A.,(2007), "Synchrony in Silicon: The Gamma Rhythm", IEEE Trans. on Neural Nets. vol.18, no.6, pp1815-1825.

Arthur, J. V., and Boahen K. (2006), "Learning in Silicon: Timing is Everything." In: Advances in Neural Information Processing Systems 18:75-82

Asanovic K et al. (1993), “CNS-1 Architecture Specification”, EECS Department, UC Berkeley, Technical Report No. UCB/CSD-93-747.

Badoni D., Giulioni M., and Dante V., (2006), “An aVLSI recurrent network of spiking neurons with reconfigurable and plastic synapses”, In ISCAS, 2006.

Bartolozzi C., and Indiveri G., (2007), “Synaptic dynamics in analog VLSI. Neural Computation”, 19, pp.2581–2603.

Bi G., Poo M., (1998) "Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic" cell type, J. Neurosci., pp.10464–10472.

Binczak S., Jacquir S., Bilbault J. M., Kazantsev V. B., Nekorkin V. I. (2006), “Experimental study of electrical FitzHugh–Nagumo neurons with modified excitability”, Neural Networks, vol.19, pp.684-693.

Boahen K. A. (2000), “Point-to-point connectivity between neuromorphic chips using address-events,” IEEE Transactions on Circuits and Systems II, vol. 47, no. 5, pp. 416–34.

Boahen K., (1997), “Retinomorphic vision systems: Reverse engineering the verbratate retina. Ph.D. thesis, California Institute of Technology, Pasadena, CA.

Bofill-i Petit A. and Murray A. F. (2004), “Synchrony detection and amplification by silicon neurons with stdp synapses”. Neural Networks, IEEE Transactions on, vol.15, pp.1296–1304.

Brette R. and Gerstner W., (2005), “Adaptive Exponential Integrate-and-Fire Model as an Effective Description of Neuronal Activity”, Journal of Neurophysiology, vol 94, pp. 3637-3642.

Carrillo R. R., Ros E., Boucheny C., Coenen D (2008), “A real-time spiking cerebellum model for learning robot control”, Biosystems, 94, pp. 18-27.

Chen Y., Hall S, McDaid L., Buiu O., Kelly P. M., (2007), “Analog Spiking Neuron with Charge-Coupled Synapses”. In: eds. Proc. of World Congress on Engineering . World Congress on Engineering, Florida. pp.1-6.

Chicca E. , Badoni D., Dante V., Andreagiovanni M., Salina G., Carota L., Fusi S., Del Giudice P., (2003), "A VLSI recurrent network of integrate-and-fire neurons connected by plastic synapses with long-term memory", IEEE Trans. on Neural Networks, vol. 14, pp. 1297-1307.

Chicca E., Dante V., Whatley A., Lichtsteiner P., Delbruck T., Indiveri G., Giudice P. D., and Douglas R. J. (2007), “A multi-chip pulse-based neuromorphic infrastructure and its application to a model of orientation selectivity”, IEEE Transactions on Circuits and Systems I, vol. 54(5), pp. 981–993.

Chua, L., O., (1971), "Memristor—The Missing Circuit Element", IEEE Transactions on Circuit Theory CT-18 (5): 507–519,

Connors, B. W., & Gutnick, M. J, (1990), "Intrinsic firing patterns of diverse neocortical neurons", Trends in neoscience, vol 13, no. 3, pp. 99-104.

Dante V., Giudice P. D., and Whatley A. M. (2005), “The neuromorphic engineer news letter.

Douglas R, Mahowald M, Mead C, (1995), “Neuromorphic analogue VLSI” Annual Review of Neuroscience, Vol. 18, pp 255-281.

Elias J. G. and Northmore D. P. M., (1995), “Switched capacitor neuromorphs with wide-range variable dynamics”, IEEE Transactions on Neural Networks, 6(6), pp.1542–1548.

Farquhar E., Hasler P., (2005), "A bio-physically inspired silicon neuron", IEEE Transactions on Circuits and Systems, 52:(3), 477-488.

Fasnacht, D. B. and Whatley, A. M. and Indiveri, G. (2008), “A Serial Communication Infrastructure for Multi-Chip Address Event Systems”, IEEE International Symposium on Circuits and Systems (ISCAS) 648-651.

Fellous J. M. and Suri R. E. (2003), "The Roles of Dopamine, Hand book of brain theory and neural networks", Arbib M.A. 2nd (ed), pp. 361-365, Cambridge, MA:MIT Press.

Furber S. B., Temple S. and Brown A. (2006), “On-Chip and Inter-Chip Networks for Modelling Large-Scale Neural Systems”, ISCAS06.

Fusi S., Annunziato M., Badoni D., Salamon A., and Amit D.J., (2000), “Spike-Driven Synaptic Plasticity: Theory, Simulation, VLSI Implementation”. Neural Computation, vol.12, pp. 2227–2258, 2000.

Gerstner W. and Kistler W.M. (2002), "Spiking Neuron Models - Single Neurons, Populations, Plasticity", Cambridge Univ. Press

Giulioni, M. and Camilleri, P. and Dante, V. and Badoni, D. and Indiveri, G. and Braun, J. and Del Giudice, P. (2008), “A VLSI network of spiking neurons with plastic fully configurable "stop-learning" synapses” , IEEE International Conference on Electronics, Circuits and Systems, ICECS 2008 678-681, IEEE (Eds.).

Grossberg S., (1969), “Some networks that can learn, remember, and reproduce any number of complicated space-time patterns”, Journal of Mathematics and Mechanics 19, 53–91.

Grossberg S., (2007), “Towards a unified theory of neocortex: laminar cortical circuits for vision and cognition” Progress in brain Research, VOL 165, pages 79-104

Gupta A, Wang Y,Markram H. (2000), "Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex", Science, pp.273–278.

Gurden H., Takita M., Jay T. M., (2000), "Essential role of D1 but not D2 receptors in the NMDA receptor-dependent long-term potentiation at hippocampal- prefrontal cortex synapses in vivo", J Neurosci. 20:RC106.

Hafliger P., Mahowald M., and Watts L., (1997), “A spike-based learning neuron in analog VLSI”. in MC Mozer, MI Jordan, and T Petsche, editors, Advances in Neural Information Processing Systems, vol. 9, pp 692.

Haflinger P., Mahowald M., and Watts L., (1996), "A spike based learning neuron in analog VLSI", Advances in neural informationprocessing systems, vol. 9, 1996.

Hasler P., Koziol S., Farquhar E. (2007), “Transistor channel dendrites implementing HMM classifiers”, ISCAS. 27-30 May, pp.3359 – 3362.

Hasler P., Minch B. A., Dugger J., and Diorio C., (1999), "Adaptive Circuits and Synapses Using pFET Floating-Gate Devices", in Gert Cauwenbergs, editor, Learning in Silicon, Kluwer Academic, 1999, pp. 33-65.

Hynna K. M., Boahen, K. (2007), “Silicon neurons that burst when primed”, ISCAS, 27-30 May, pp. 3363-3366.

Indiveri G (2003), A low-power adaptive integrate-and-fire neuron circuit.” IEEE Int. Symp. Circuits and Systems, ISCAS, pp IV820-823.

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 Trans. on Neural Netw., vol. 17, no. 1, pp. 211-221.

Indiveri G., (2003), “Neuromorphic bistable VLSI synapses with spike-timing-dependent plasticity”, In Advances in Neural Information Processing Systems.

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, pp.211–221.

Indiveri G., Chicca E., Douglas R., (2004), “A VLSI reconfigurable network of integrate–and–fire neurons with spike–based learning synapses”, ESANN'2004 proceedings - European Symposium on Artificial Neural Networks Bruges, 28-30, pp. 405-410.

Indiveri G., Linares-Barranco B., Hamilton T. J., Van Schaik A., Etienne-Cummings R., Delbruck T., Liu S.C., Dudek P., Häfliger P., Renaud S., Schemmel J., Cauwenberghs G., Arthur J., Hynna K., Folowosele F., Saighi S., Serrano-Gotarredona T., Wijekoon J. H. B, Wang Y. and Boahen K., (2011), "Neuromorphic Silicon Circuits", Frontiers in Neuroscience, 5:73. pp 1-23.

Indiveri G., Mitra S, Fusi S. (2007), "Spike-based learning in VLSI networks of integrate-and-fire neurons", in Proc. IEEE International Symposium on Circuits and Systems ISCAS. Issue 27-30 May, pp. 3371 – 3374.

Izhikevich E. M. (2004), "Which Model to Use for Cortical Spiking Neurons?", IEEE Trans. Neural Networks 15, pp. 1063-1070.

Izhikevich, E. M. (2003), "Simple model of spiking neurons." IEEE Transactions on Neural Networks 14(6): 1569-72.

Izhikevich, E. M. (2007), "Solving the Distal Reward Problem through Linkage of STDP and Dopamine Signaling", Cereb. Cortex ,vol.17: pp. 2443-2452.

Jung R, Brauer E. J, Abbas J. J. (2001), “Real-time interaction between a neuromorphic electronic circuit and the spinal cord”, IEEE Trans. On Neural Systems and Rehab. Eng., vol. 9(3), pp.319–326.

Kandel E. R., Schwartz J. H., Jessell T. M., (2000.), "Principles of neural science", McGraw-Hill, New York.

Koickal T. et al. (2007), "Analog VLSI Circuit Implementation of an Adaptive Neuromorphic Olfaction Chip", circuits and Systems I: regular Papers, IEEE Transactions on 54, pp. 60-73.

LeMasson G., Renaud S., Debay D., Bal T. (2002), “Feedback inhibition controls spike transfer in hybrid thalamic circuits”, Nature, vol. 4178, pp. 854-858.

Letzkus J. J et al., (2006), "Learning Rules for Spike Timing-Dependent Plasticity Depend on Dendritic Synapse", Journal of Neuroscience, October 11, 26, pp. 10420 –10429.

Linares-Barranco B et al. (1991), A CMOS implementation of FitzHugh-Nagumo neuron model. IEEE Journals of Solid-State Circuits, vol. 26, no. 7, pp 956-965.

Linares-Barranco B., Serrano-Gotarredona T., (2009), " Memristance can explain Spike-Time-Dependent-Plasticity in Neural Synapses", Nature Precedings.

Liu S C., Schaik A van, Minch B., & Delbruck T,(2010), “Event-based 64-channel binaural silicon cochlea with Q enhancement mechanisms”, IEEE International Symposium on Circuits and Systems, Paris, France, pgs 2027-2030.

Liu S. C., Kramer J., Indiveri G., Delbrück T., Burg T., Douglas R. (2001), “Orientation-selective aVLSI spiking neurons”, Neural networks, vol. 14, pp. 629-643.

Liu Shih-Chii, (2003), “Analog VLSI Circuits for Short-Term Dynamic Synapses”, EURASIP Journal on Applied Signal Processing 7, pp. 620–628

Mahowald M, Douglas R (1991), "A silicon neuron". Nature, vol. 354, no 6354, 19-26, pp 515-518.

Markram H et al (2004), "Interneurons of the neocortical inhibitory system". Nature Reviews Neuroscience, vol. 5, pp 793-807.

Markram H., Wang Y., Tsodyks M., (1998), “Differential signaling via the same axon of neocortical pyramidal neurons”, Proc Natl. Acad. Sci. USA, pp. 5323–5328.

Mead C. A. (1989), "Analog VLSI and Neural Systems", Addison-Wesley, Reading, MA.

Merolla P, Arthur J, Shi B. E., and Boahen K. (2007), “Expandable Networks for Neuromorphic Chips, IEEE Transactions on Circuits and Systems I”, vol 54, No 2. pp. 301-311.

Merolla P. and Boahen K., (2003), “A Recurrent Model of Orientation Maps with Simple and Complex Cells”. In Advances in Neural Information Processing Systems, Vol. 16, pp995-1002, MIT Press.

Mihalas S and Niebur E (2009), "A generalized linear integrate-and-fire neural model produces diverse spiking behaviors", Neural Computation 2009; 21(3):704-18

Morrison, A. et al. (2008), "Phenomenological models of synaptic plasticity based on spike timing." Biological Cybernetics, pp. 459-478

Nakada K, Asai T, and Amemiya Y, (2004), "Analog CMOS implementation of a bursting oscillator with depressing synapse. in Proc. ISSNIP '04, pp 503-506.

Nakada K, Asai T, and Hayashi H, (2005), "A silicon Resonate-and-fire neuron based on the volterra system". Int. Symp. on Nonlinear Theory and its Applications, pp 82-85.

Nelson S (2008), "Pyramidal neurons: dendritic structure and synaptic integration" Nature Reviews Neuroscience, vol. 9, pp 206‑221.

Nowak L. G. et al., (2003), "Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses", J Neurophysiol,vol. 89, pp.1541-1566.

Otani S., Daniel H., Roisin M. P., Crepel F., (2003), "Dopaminergic modulation of long-term synaptic plasticity in rat prefrontal neurons". Cereb Cortex. 13, pp.1251-1256.

Patel G N and DeWeerth S P (1997), "Analogue VLSl Morris-Lecar neuron" Electronics Letters, vol. 33, n.12, pp 997-998.

Peter A and Jones E. G.,(1984), " Cellular Components of the Cerebral Cortex", New York: Plenum Press. Cerebral cortex, vol. 1.

Peter A. and Jones E G (1984), "Cellular Components of the Cerebral Cortex", New York: Plenum Press. Cerebral cortex, vol. 1.

Plana L. A., Furber S. B., Temple S., Khan M., Yebin S., Jian W., Shufan Y., (2007), "A GALS Infrastructure for a Massively Parallel Multiprocessor", IEEE Design & Test of Computers Volume: 24, Issue: 5, pp 454-463.

Porrmann M., Franzmeier M., Kalte H., Witkowski U., Rückert U. (2004), "A Reconfigurable SOM Hardware Accelerator", Proc. ESANN'2002 - European Symposium on Artificial Neural Networks, Bruges, Belgium, 24-26 April, pp. 337-342.

Rachmuth G. and Poon C. S., (2003), "Design of a Neuromorphic Hebbian Synapse Using Analog VLSI". In Proceedings of the 1st international IEEE EMBS Conference on Neural Engineering.

Rangan V., Ghosh A., Aparin V., Cauwenberghs G. (2010), "A subthreshold aVLSI implementation of the Izhikevich simple neuron model", Conference Proceedings of the International Conference of IEEE Engineering in Medicine and Biology Society.

Rasche C. and Douglas R. J., (2001), "Forward- and backpropagation in a silicon dendrite". IEEE Transactions on Neural Networks, Vol. 12, pp.386–393.

Redgrave, P., Prescott, T. J., Guenry, K., (1999), "The basal ganglia: a vertebrate solution to the selection problem? " Neuroscience 89 (4), 1009 – 1023.

Renaud S., Tomas J., Bornat Y., Daouzli A., and Sa¨ıghi S. (2007), “Neuromimetic ICs with analog cores: an alternative for simulating spiking neural networks,” in ISCAS, pp. 3355–3358.

Riesenhuber, M., Poggio, T., (1999), “Hierarchical models of object recognition in cortex”, Nature Neuroscience 2 (11), 1019–1025.

Ros E., Ortigosa E. M., Agís R, Arnold M., Carrillo R. (2006), “Real time computing platform for spiking neurons (RT-Spike), IEEE Transactions on Neural Networks”. 17(4), pp. 1050-1063.

Roth A. E., and Wennekers T., (2009), Internal meetings COLAMN project

Saudargiene, B. Porr and Wörgötter F., (2005), Synaptic modifications depend on synapse location and activity: a biophysical model of STDP, 5th International Workshop on Neural Coding, Vol. 79, Issues 1-3, pp. 3-10 .

Schaik A. V., Jin C., Hamilton T. J. (2010), A log-domain implementation of the Izhikevich neuron model, ISCAS’ 10.

Schemmel et al. (2010), “Modeling synaptic A Wafer-Scale Neuromorphic Hardware System for Large-Scale Neural Modeling”, Proceedings of the 2010 IEEE International Symposium on Circuits and Systems, Paris, France (2010):1947-1950.

Schemmel J, Fieres J., and Meier K. (2008), “Wafer-Scale Integration of Analog Neural Networks” IEEE International Joint Conference on Neural Networks, Issue , 1-8 June, pp 431 – 438.

Schemmel J., Brüderle D., Grübl A., Meier K., Müller E. (2007), “Modeling synaptic plasticity within networks of highly accelerated I&F neurons”, ISCAS. 27-30 May, pp.3367 - 3370

Schemmel j., Meier K., Muller E., (2004), “A New VLSI Model of Neural Microcircuits Including Spike Time Dependent Plasticity”, Proceedings of the 2004 Int. Joint Conference on Neural Networks, IEEE Press, pp. 1711-1716.

Schoenauer T., Atasoy S., Mehrtash N., Klar H. (2000), “ Simulation of a digital neuro-chip for spiking neural networks” Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on Volume 4, Issue , 2000 Page(s):490 - 495 vol.4

Schoenauer T., Mehrtash N., Jahnke A., Klar H. (1998), “MASPINN: Novel Concepts for a NeuroAccelerator for Spiking Neural Networks”, Proc. VIDYNN’98, Stockholm, June 22-26.

Schultz S R, and Jabri M A (1995), "Analogue VLSI ’integrateand-fire’ neuron with frequency adaptation2. Electronic Letters, vol. 31, no. 16, pp. 1357–1358.

Simoni M F, and DeWeerth S P (1999), Adaptation in a VLSI model of a neuron. IEEE Transactions on circuits and systems-II: Analog and digital signal processing, vol. 46, no. 7, pp 967-970.

Somogyi P,(1989), "Synaptic organisation of GABAergic neurons and GABA-A receptors in the lateral geniculate nucleus and visual cortex", in Neural Mechanisums of visulal cortex, in Neural Machanisms of visual Perception. Proceedings of ratina research foundation Symposia.

Somogyi P.,(1989), “Synaptic organisation of GABAergic neurons and GABA-A receptors in the lateral geniculate nucleus and visual cortex”, in Neural Mechanisums of visulal cortex, in Neural Machanisms of visual Perception. Proceedings of ratina research foundation Symposia.

Sorensen M., DeWeerth S., Cymbalyuk G., Calabrese R. L. (2004), “Using a hybrid neural system to reveal regulation of neuronal network activity by an intrinsic current”, J. Neurosci., vol.24, pp. 5427-5438.

Stegenga J, Feber J. L., Marani E., Rutten W. L. C., (2008), "Analysis of Cultured Neuronal Networks Using Intraburst Firing Characteristics". IEEE Transactions on Biomedical Engineering 55 (4): 1382–1390.

Tanaka H., Morie T., Aihara K., (2007), “A CMOS circuit for STDP with a symmetric time window”, International Congress Series 1301, pp. 152–155.

Thomson A. M., Deuchars J., West D. C., (1993), "Large, deep layer pyramidpyramid single axon EPSPs in slices of rat motor cortex display paired pulse and freuquency-dependent depression, mediated presynaptically and self-facilitation mediated postsynaptically", Journal of Neurophysiol vol. 70, pp.2354–2369.

Toledo-Rodriguez M, et al. (2003), "Neocortex: Basic neuron types. in Hand book of brain theory and neural networks", 2nd e., M.A. Arbib (ed), pp719-725, Cambridge, MA:MIT Press.

Tsodyks M., Uziel A., Markram H., (2000), “Synchrony generation in recurrent networks with frequency-dependent synapses”. Journal Neuroscience vol. 20, RC1 (1–5)

Vogelstein R. et.al. (2007), "Dynamically Reconfigurable Silicon Array of Spiking Neurons With Conductance-Based Synapses", IEEE Transactions on Neural Networks, 18.

Vogelstein R.J., Malik U., Cauwenberghs G. (2004), “Silicon spike-based synaptic array and address-event transceiver”, Proceedings of ISCAS’04, vol.5, pp.385-388.

Wagenaar D. A., Pine J., Potter S. M., (2006), "Searching for Plasticity in Dissociated Cortical Cultures on Multi-Electrode Arrays". Journal of Negative Results in BioMedicine: 516–35.

Well R. B., (2005), “Cortical Neuron and Circuits: A tutorial introduction” April.

White E.L.,(1989), “Cortical Circuits: Synaptic Organization of the Cerebral Cortex, Structure, Function, and Theory”, Boston, MA: Birkhäuser, 1989.

White, E.L., Cortical Circuits (1989), "Synaptic Organization of the Cerebral Cortex, Structure, Function", and Theory, Boston, MA: Birkhäuser.

Wijekoon J. H. B. (2007), “Novel analogue VLSI circuit of a cortical neuron”, M.Phil. thesis. School of Electrical and Electronic Engineering, The University of Manchester, UK.

Wijekoon J. H. B. and Dudek P., (2009), "A CMOS circuit implementation of a spiking neuron with bursting and adaptation on a biological timescale", BioCAS 09 (accepted).

Wijekoon J. H. B. and Dudek P.,(2006), “A simple analogue VLSI circuit of a cortical neuron”, IEEE International Conference on Electronics, Circuits and Systems, ICECS 2006, pp.1344-1347, December.

Wijekoon J. H. B. and Dudek P.,(2007), Spiking and Bursting Firing Patterns of a Compact VLSI Cortical Neuron Circuit, International Joint Conference on Neural Networks, IJCNN 2007, Orlando, Florida, August.

Wijekoon J. H. B. and Dudek P.,(2008a), "Integrated Circuit Implementation of a Cortical Neuron", IEEE International Symposium on Circuits and Systems, ISCAS 08, pp 1784-1788.

Wijekoon J. H. B. and Dudek P.,(2008b), "Compact Silicon Neuron Circuit with Spiking and Bursting Behaviour", Neural Networks, Vol 21, Number 2-3, pp 524-534, March/April.

Young Jun Lee et.al (2004), "Low power real time electronic neuron VLSI design using subthreshold technique". IEEE Int. Symp. Circuits and Systems, vol. 4, pp IV-744-747.

Zachary F. Mainen, Terrence J.Sejnowski (1996), "Influence of dendric structure on firing pattern in model neocortical neurons” Letter to nature, Vol. 382.25.

Zaghloul K. A., Boahen K., (2004), “Optic nerve signals in a neuromorphic chip II: Testing and results” IEEE transactions on bio-medical engineering, Apr. 51(4):667-75.

Zaghloul K. A., Boahen K., (2006), “A silicon retina that reproduces signals in the optic nerve” J. Neural Eng. 3 257.

Zhang L. I. et al., (1998), "A critical window for cooperation and competition among developing retinotectal synapses". Nature, vol.395, pp. 37–44.

Zou Q. et.al., (2006), "Real-time simulations of networks of Hodgkin–Huxley neurons using analog circuits", Neurocomputing, Volume 69, Issues 10-12, Pages 1137-1140.

APPENDIX A: Short-Term Dynamic Synapse Equations

Approximated mathematical equations of the EDS circuit are derived below. Since, IFS, IDS, and EFS circuits have the same circuit topology as of the EDS, the equations for IFS, IDS and EFS circuits can be derived in the same way.

Excitatory Depressing Synapse (EDS) Model

Recovery of w per time step, $\Delta w_{\alpha d} / \Delta t$

Considering the current mirror circuit (transistors M1, M2 and M6) of the EDS circuit, the rate of recovery of the depressing synapse, $\Delta w_{\alpha d} / \Delta t = \beta / Cw \times I_{dM1}$

Where, I_{dM1} is the drain current through the M1 transistor and the β is the current gain of the current mirror. Cw is the capacitance of the capacitor Cw. Δt is time-step of the simulation. Assuming transistor M1 operate in saturation region,

$$\Delta w_{\alpha d} = k_r (2.8 - V_{\alpha p})^2 \qquad (6.6)$$

Where $k_r = \frac{\Delta t}{2} \frac{\beta}{C_w} \left(\frac{W}{L} \right)_{M1} \mu_p C_{ox}$ is a constant, $\left(\frac{W}{L} \right)_{M1}$ is width to length ratio of the transistor M1; CMOS process Parameters C_{ox}, μ_p, and V_t are gate oxide capacitance per unit area, charge-carrier effective mobility of pMOSFET, and threshold voltage of pMOSFET transistor respectively.

Depression of w per pre-synaptic spike, Δw_d

Amount of depression per pre-synaptic spike, $\Delta w_d = \Delta q_c / C_w$

Amount charge added to the capacitor Cw per spike, $\Delta q_c = \int_o^{t_{sw}} i_{dM5} \times dt$

$$\Delta w_d = \frac{1}{C_w} \int_o^{t_{sw}} i_{dM5} \times dt$$

By considering the drain current of M5, i_{dM5} amount of depression approximately,

$$\Delta w_d = \begin{cases} \frac{k_d}{2} (2.8 - V\Delta w_p)^2 & \text{when M5 is in saturation region} \\ k_d (2.8 - V\Delta w_p)(w - Voffsp2) & \text{when M5 is in linear region} \end{cases} \qquad (6.7)$$

Where $k_d = \frac{t_{sw}}{C_w}\left(\frac{W}{L}\right)_{M5}\mu_p C_{ox}$, $\left(\frac{W}{L}\right)_{M5}$ is width to length ratio of the transistor M5; i_{dM5} is the drain current through the M5 transistor (Figure 6.2 (a)) during the pre-synaptic spike pulse duration, t_{sw} (≈ 3 ns), *Voffsp2* is off-set of the M8-M9 source follower (Shown in Figure 6.2 (a) and this value can be changed using *VLSp*; the default value of the offset can be assumed as 0.4V).

Excitatory Wight Depressing (ED) circuit's output voltage, Vw

Approximated discrete mathematical model of the EDS circuit's output voltage *w(t)* evolves same as the mathematical model equation given in equation 6.5. This can be rewritten with the circuit parameters.

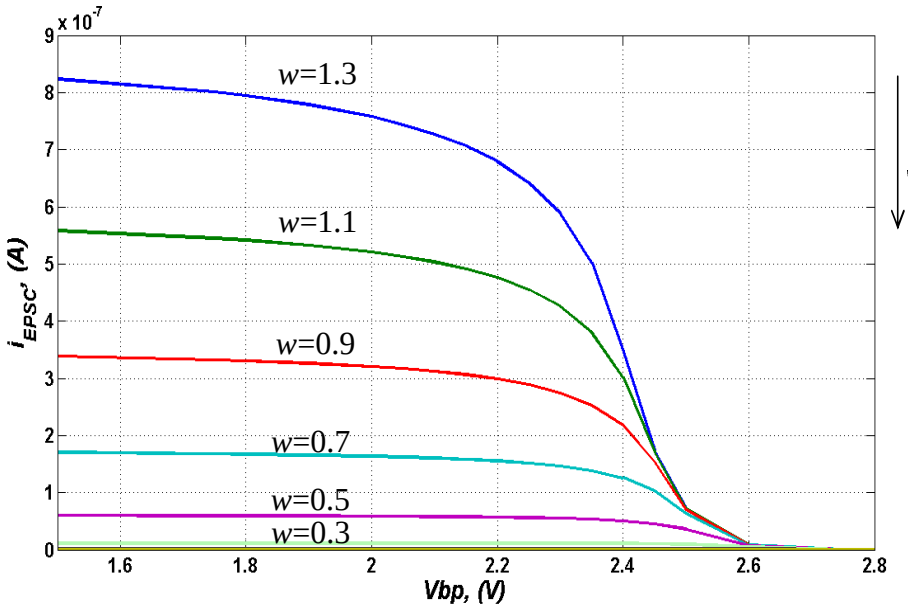
$$w_d(t+\Delta t) = \begin{cases} \max\{w(t)-\Delta w_d, w_{d\min}\} & \text{if pre-neuron fires} \\ \min\{w(t)+\Delta w_{\alpha d}, wr_d\} & \text{otherwise} \end{cases} \quad (6.5)$$

Where, $\Delta w_{\alpha d}$ and Δw_d are given from the equation 6.6 and 6.7 respectively. The externally control tuning voltages, $V_{\alpha p}$ and $V\Delta wd$ sets the degree of the decay and the depression respectively; The design parameter, $\frac{\beta}{C_w}\left(\frac{W}{L}\right)_{M1}$ used in the circuit implementation of Chapter 9 is 1.5; μ_p, C_{ox}, *and* V_t value from AMS standard 0.35 μm CMOS technology process parameters is 126 cm^2/VS, 4.54 fF/μm^2 and V_t,=0.5 V respectively. w_{dmin} = 0, and wr_d = (*3.3-Vwrd-Voffsp1-Voffsp2*), wr_d is the resting of the synapse. The biasing voltage *Vwrp* is used to set the wr_d of the circuit. The voltages *Voffsp1*, and *Voffsp2* is off-set of the M3-M7 and M8-M9 source follower (shown in Figure 6.2 (a) and these offset can be assumed as 0.4V) respectively.

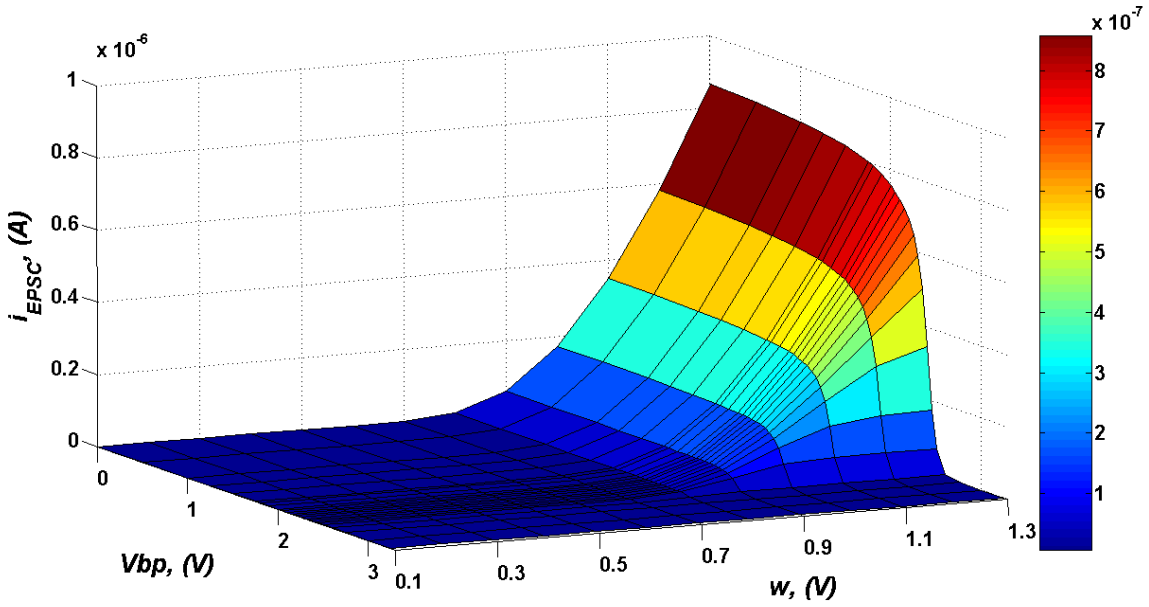
EX-Isyn circuit's excitatory post-synaptic current (EPSC), i_{EPSC}

The amount of post-synaptic current injection, i_{EPSC}, caused by a pre-synaptic spike depends on the synaptic strength, *w*. However, user can scale the i_{EPSC} current for a given value of *w* using externally controllable voltage *Vbp*, as shown in Figure 6.17. This controllable voltage could also be used to limit the maximum i_{EPSC}, depending on the operational region of the Ms1 transistor (for the higher tuning values of *Vbp* as seen in Figure 6.17(a)). Each post-synaptic current injection lasts for a period of a few nanoseconds (Pre-Gen circuit used in the CNL chip in Chapter 9 use approximately 3 ns

pulse). The excitatory post-synaptic current can be obtained/modelled as per the graph shown in Figure 6.17(a)).



(a) (c)



(b)

Figure 6.17: Excitatory Synaptic Current Generator (EX-ISYN) circuit's i_{EPSC} values for different *Vbp* and synaptic weight, *w*: (a) in a 2D plot, (b) in a 3D plot; (c) the EX-ISYN circuit.

APPENDIX B: Estimation of Cortical Network Size in VLSI

Current CNL Chip (120 neurons unit)

The CNL chip is fabricated in 0.35µm CMOS technology:

Total number of neurons = 120

Number of excitatory neurons per unit 100

Number of inhibitory neurons per unit 20

Total number of synapses per unit = 7 560

Number of STDP/DA-STDP synapses = 2 100

Number of Non-STDP synapses = 5 460

Area of the 120 neurons unit in 0.35µm CMOS technology = 24 mm^2

Approximate number of neurons and synapses in 120 mm^2 chip

120 mm^2 chip in 0.35µm CMOS technology:

Number of neurons = 120/24 x 120 ≈ 600

Number of synapses = 120/24 x 7 560 ≈ 37 800

Using sub-micron technology (90 nm) 120 mm^2 chip

In 90nm technology; assuming effective technological migration scaling factor 8* (theoretical area multiplication factor 15): Number of neurons = 600 x 8 ≈ 5 000

Number of synapses = 37 800 x 8 ≈ 300 000

* Even though the theoretical area multiplication factor is 15, analogue circuit cannot be scales in the same factor. However, since spike routing to pre-synapses and the auxiliary circuit use digital circuit elements these can be scaled with higher scaling factor than the pure analogue circuit scaling factor.

Using multi-chip sub-micron technology (using 90 nm, 120 mm^2 chips)

If twenty 120 mm^2 90 nm chips are used to form multi-chip network, the size of the network could be: Number of neurons = 5 000 x 20 ≈ 100 000

Number of synapses = 300 000 x 20 ≈ 6 000 000

Size of the network with hypothetical wafer scale integration

Wafer scale integration size of the network, if wafer diameter is 12" (30 cm); (wafer area 730 cm^2)

Assuming extra overhead of 25 cm^2 area

Number of neurons = 600 x 8 x705/120 x100 ≈ 2 800 000

Number of synapses = 37 800 x 8 x 705/120 x100 ≈ 180 000 000