Circuits and Systems
Mekelweg 4,
2628 CD Delft
The Netherlands
http://ens.ewi.tudelft.nl/

CAS-MS-2017-03

M.Sc. Thesis

Full-Custom Multi-Compartment Synaptic Circuits in Neuromorphic Structures

Xuefei You

Abstract

Neuromorphic engineering, aiming at emulating neuro-biological architectures in efficient ways, has been widely studied both on component and VLSI system level. The design space of neuromorphic neuron, the basic unit to conduct signal processing and transmission in nervous system, has been widely explored while that of synapse, the specialized functional unit connecting neurons, is less investigated.

In this thesis, a current-based phenomenological synapse model with power-efficient structures, consisting of efficient synaptic learning algorithms and multi-compartment synapses, has been proposed. A vertical insight is given into the design space of spike-based learning rules in regards to design complexity and biological fidelity. Due to various biological conducting mechanisms, the receptors, namely AMPA, NMDA and GABAa, demonstrate different kinetics in response to stimulus. The designed circuit offers distinctive features of receptors as well as the joint synaptic function. A better computation ability is demonstrated through a cross-correlation detection experiment with a recurrent network of synapse clusters. The analog multi-compartment synapse structure is able to detect and amplify the temporal synchrony embedded in the synaptic noise. The maximum amplification level is 2 times larger than that of single-receptor configurations. The final design implemented in UMC65nm technology consumes 1.92, 3.36, 1.11 and 35.22pJ per spike event of energy for AMPA, NMDA, GABAa receptors and the advanced learning circuit, respectively.

Full-Custom Multi-Compartment Synaptic Circuits in Neuromorphic Structures

Thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

in

Microelectronics

by

Xuefei You
born in Chengdu, China

This work was performed in:

Circuits and Systems Group
Department of Microelectronics Engineering
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Delft University of Technology

Delft University of Technology
Department of
Microelectronics Engineering

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled **"Full-Custom Multi-Compartment Synaptic Circuits in Neuromorphic Structures"** by **Xuefei You** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: August 30th, 2017

Chairman: _____
prof.dr.ir. Alle-Jan van der Veen

Advisor: _____
dr.ir. Rene van Leuken

Committee Members: _____
dr. Amir Zjajo

_____
dr. Carlo Galuzzi

Abstract

Neuromorphic engineering, aiming at emulating neuro-biological architectures in efficient ways, has been widely studied both on component and VLSI system level. The design space of neuromorphic neuron, the basic unit to conduct signal processing and transmission in nervous system, has been widely explored while that of synapse, the specialized functional unit connecting neurons, is less investigated.

In this thesis, a current-based phenomenological synapse model with power-efficient structures, consisting of efficient synaptic learning algorithms and multi-compartment synapses, has been proposed. A vertical insight is given into the design space of spike-based learning rules in regards to design complexity and biological fidelity. Due to various biological conducting mechanisms, the receptors, namely AMPA, NMDA and GABAa, demonstrate different kinetics in response to stimulus. The designed circuit offers distinctive features of receptors as well as the joint synaptic function. A better computation ability is demonstrated through a cross-correlation detection experiment with a recurrent network of synapse clusters. The analog multi-compartment synapse structure is able to detect and amplify the temporal synchrony embedded in the synaptic noise. The maximum amplification level is 2 times larger than that of single-receptor configurations. The final design implemented in UMC65nm technology consumes 1.92, 3.36, 1.11 and 35.22pJ per spike event of energy for AMPA, NMDA, GABAa receptors and the advanced learning circuit, respectively.

Acknowledgments

I would like to express my sincere gratitudes to my supervisors Rene van Leuken, Amir Zjajo and Sumeet Kumar for their patient assistances and meaningful advice on my project and life as well. Thank Mr.Rene for inspiring me to participate in the SAMOS conference, which is truly a precious experience for a master student. Thank Amir for consistent guidances and suggestions whenever I ran into problems. Thank Sumeet for giving me supports when I feel depressed.

I would like to thank my colleagues, Hemanth Singh, Johan Mes and Ester Stienstra. Thanks for the valuable times we spent, the jokes we talked about, the laughters we had in room HB17.090. They made my stay in this group full of joy and memory, and I will always remember the many moments "alright, we should go back to work".

I would also like to thank my dearest friend, Yuqian Zhang, who though is not here with me but can always understand and support me. We share similar experiences, but somehow she is always stronger in mind than me and becomes my spiritual support.

I am very grateful to my friends in TU Delft, Kangli Huang, He Zhang, Xianwei Zeng and the people that I can not list completely in this acknowledgement. Thank them for keep sharing the wonderful or funny things with me everyday, for caring for me in every way, for comforting me when I was down, filling my two-year in Netherlands with meaningful memories.

Also, I would like to thank Zhihao Zhou for his kind encouragement and company through my darkest times. Like he always says, being there is a kind of support, and that was what he truly did for me.

Finally, great thanks to my dad, mom and beloved brother for your unselfish love and company. The happy moments chatting with them really disperse my loneliness and keep reminding me that they will always be there for me no matter what happens.

Xuefei You
Delft, The Netherlands
August 30th, 2017

Contents

List of Figures

List of Tables

1 Introduction

Neuromorphic electronic system, a concept proposed by Carver Mead in 1990, is a milestone in the course of brain-like system development [13]. The system makes full use of continuum physics of transistors to emulate biological elements. The feasibility of neuromorphic engineering comes from the similarity of working mechanism between nerves and transistors as well as a massively parallel structure between neural and neuromorphic networks. More importantly, the transistor is able to display exponential dynamics observed in neural cells. Those desirable features make neuromorphic design a competitive method to achieve large-scale nerve emulation.

In the nervous system, information is exchanged and received between billions of nerve cells, i.e. *neurons* through electrical and chemical signal spikes. The connection of those neurons forms *spiking neural networks* (SNN). Strictly speaking, neurons are separate units connected by specialized functional units called *synapses*. Synapse plays an essential role in the formation of learning and memory, invoked by a special mechanism called *synaptic plasticity*. The anatomical periphery of two adjacent neurons connected by multiple synaptic terminals is displayed in Fig.1.1.

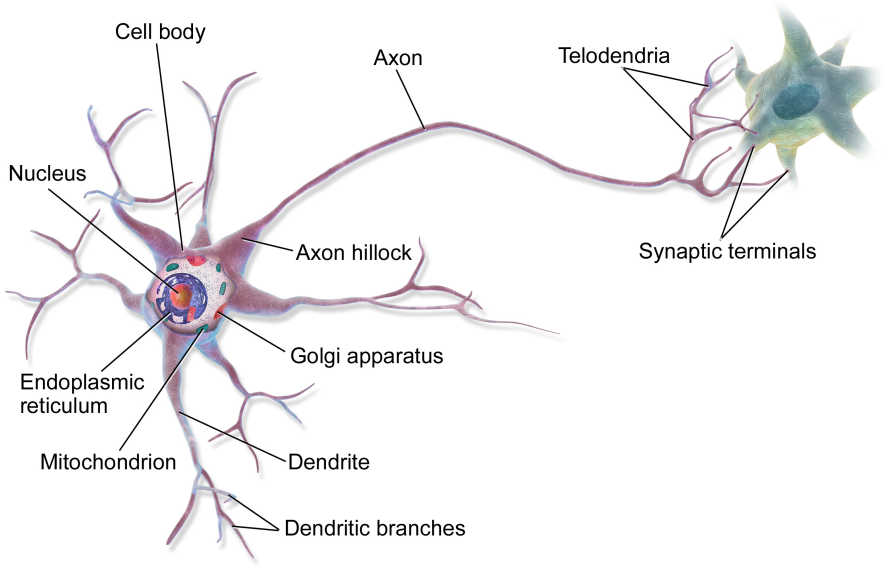


Figure 1.1: The anatomical periphery of two adjacent neurons connected by synapses.

In last few decades, considerable studies on neuromorphic circuits have paved the way for large-scale integration on hardware. Several analog and digital chips are validated depending on different application goals [14][7][15]. The design space of neuromorphic

neuron has been widely explored, ranging from biological to phenomenological models, or from real-time to accelerated-time and in some other design considerations. Synapse, a delicate but crucial unit, however, is usually simplified to various levels, restricting mapping fidelity. For this reason, this thesis mainly limits the scope to neuromorphic synaptic design in accordance with biological models.

1.1 Problem Statement

The first problem is about the synaptic learning configurations. In current literature, many spike-based synaptic plasticity rules have been proposed and built in hardware implementations. Each design focuses on specific applications or is derived from different biological models, which results in redundant or unclear design alternatives of the synapse. It is complex to understand or generate target design configurations under defined design space.

secondly, the synaptic structure is simplified to achive large integration in many cases. It should be noticed that synaptic plasticity is an abstraction of the learning process to so-called synapse structure. It is influenced by both the response of the presynaptic neuron to input trains and the interplay between adjacent neurons. Special membrane protein on dendrites called *neuroreceptor* further interprets the state information generated from synaptic plasticity depending on receptor types. A generic synapse structure does not capture the diverse dynamics of different types of receptors in biological synapses, which are essential for the realization of biophysically accurate neural behaviors in SNN [16].

1.2 Approach

The human brain is extremely energy-efficient in processing signals. Consisting of 10^{11} neurons and 10^{15} synapses, it only consumes around 20 W power, which leads to approximate 20 fJ of energy consumption per operation. An attempt has been made to emulate human-scale cortex model with software simulation. In total an exascale supercomputer and 0.5 GW power consumption are reported in [17], a factor of 10 million less efficient than the human brain. It is not easy to make the realization of human brain emulation with realistic power and area restrictions through software simulations.

Unlike Von-Neumann architecture-based computer, hardware implementation enables massively parallel structure, which is similar to that of biological neural networks. It gives a direct and accurate mapping from biological models to neuromorphic electronics. SpiNNaker project [18], a digital implementation of neural networks with 75-250 thousands of neurons and 38-86 millions of synapses, is reported to consume less than 1 W of power. The corresponding energy per synaptic event is about 10 n J/event, which is around 6 order of magnitude over the real brain. The hard limitation of digital

implementations of nerve elements (neurons and synapses) is the non-elegant way of interpreting the magnitude of transmission signals with only 1-bit representation.

The primitive analog feature of transistors could be well used to mimic the biological behavior of nerve cells. From the microscopic angle, transistor uses a population of electrons to achieve conductance, which is very much the same way as nerve membrane using a population of channels to conduct[13]. On the other hand, the exponential behavior of transistors in sub-threshold region is analogous to that of the relation between ionic conductance and corresponding membrane voltage [19]. Therefore, an analog implementation of nerve cells is a fundamental and potential opportunity for dedicated and efficient brain-like emulations in VLSI scale.

Various synaptic learning algorithms are achieved in hardware to compare their functionality and design space in biological fidelity aspects. The diverse types of receptors are also implemented in regards to their distinctive mechanisms to achieve a more powerful computation ability.

1.3 Goals

In this project, several main goals are to:

- Analyze multiple synaptic algorithms regarding key biological features and generate design space of each single type. A selection of these three models is extracted based on the design complexity and the biological fidelity.
- Build a multi-compartment synapse structure including three receptors: AMPA, NMDA, and GABAa for better emulation of biological synaptic behaviors.
- Characterize the functionality of each single learning models and choose one of them for further network integration.
- Build a unit cluster structure incorporating both the synaptic learning and the receptors and further use it to demonstrate the distinctive roles of receptors.
- Achieve synchrony detection verification of the multi-receptor synapse structure.

1.4 Contributions

The main contributions of this thesis are:

- Implementations and comparisons of multiple spike-based learning algorithms in UMC 65nm technology, and the corresponding analyses in a reconfigurable design flow from the biological fidelity and possible design space aspects.
- A novel neuromorphic synapse design with multi-compartment of receptors of low power and area consumptions.
- Proposal of a synchrony detection methodology using cross-correlogram.

- Achieve synchrony detection and amplification through the multi-compartment synapse structure.

1.5 Thesis Outline

In Chapter 2, the relevant biological phenomenon and mechanism of the brain will be introduced as the foundation of further design considerations. At the end of this chapter, extracted models of neural network, single synapse and its learning rules will be discussed to simplify the biological beahaviors to bridge between the biophysical and hardware models. Next, in Chapter 3, hardware implementations of three types of learning rules and synaptic receptors will be explained as well as a cluster network integration model, which will be used in the next chapters. Chapter 4 will present the characterizations of the building elements including the synaptic learning and the receptor circuits. Moreover in Chapter 5, the functionality of specific synchrony detection of the multi-compartment synapse structure is demonstrated. Finally, the conclusion and future work are given in Chapter 6.

2 Background and Models

This chapter gives a basic introduction to the phenomenon and mechanism of the brain. The elaborations will not be merely about biological facts but also target-oriented extractions of models from those facts, laying foundations for further neuromorphic design possibilities.

2.1 Neuron

Neuron, also known as *nerve cell*, is a basic functional unit in the brain, responsible for signal processing and transmission via electrical and chemical signals. Three types of specialized neurons are categorized. Sensory neurons, distributed mainly in sensory organs, collect sensory signals invoked by stimuli like touch, sound or any other affecting sources of stimuli and then send them back to *spinal cord* or brain, the central control system of mammals. Those signals are further transmitted through interneurons within the same region of the brain or spinal cord in the nervous system. Finally, motor neurons obtain order-carried signals from the brain or spinal cord to drive muscle contractions and glandular outputs. Among them, the interactions between interneurons play a crucial role in learning and decision-making in the brain [20].

A neuron usually consists of three specialized parts: *soma* (*cell body*), *dendrites* and *axon*. As shown in Fig.1.1, the spherical body with numerous branches is the soma and those surrounding branches are dendrites, forming "dendritic tree". Dendrites detect incoming signals from all the space within reach. When stimuli are large enough, an all-or-none electrochemical pulse named *action potential* is produced. The action potential is then transmitted rapidly through a long fiber *axon* to its terminal.

The action potential transmitted between soma and axon terminal follows an all-or-none principle. It is explained by the activation mechanism of an action potential shown in Fig.2.1. Once surpassing the threshold of excitation, the membrane voltage will boost to an extreme value due to a larger number of participating conducting channels. Even if a bigger stimulus is invoked, the potential amplitude can not exceed this height. Instead, those excess stimuli can only wait until next surpassing of threshold point to fire.

A neuron accumulates stimulus from multiple dendrites (N) that are connected to the pre-synaptic neurons at axon hillock, forming an N-to-1 relation. Sometimes, a single stimulus from the single neuron is high enough to produce action potentials. In other times, however, multiple neurons should fire almost simultaneously to launch one action potential (see Fig.2.2). Two types of action current are identified:

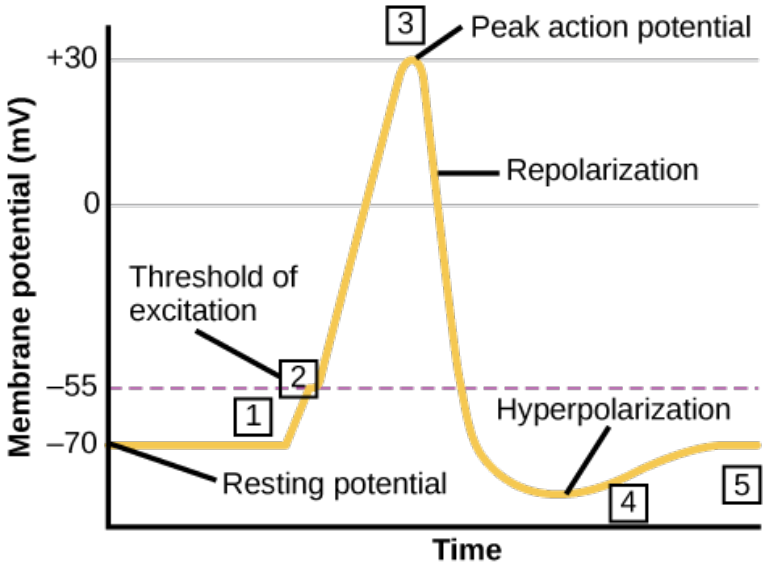


Figure 2.1: Action potential elaborations [1]. At point 1, the cell is stimulated and starts to depolarize. When the depolarization level surpasses certain threshold of excitation at point 2, a steep increase of Na^+ channels is observed, leading to a peak action potential at point 3. After that the membrane potential decays via the efflux of K^+ from the cell until a hyperpolarization occurs at phase 4. A resting potential state at point 5 is reached after a *refractory period*.

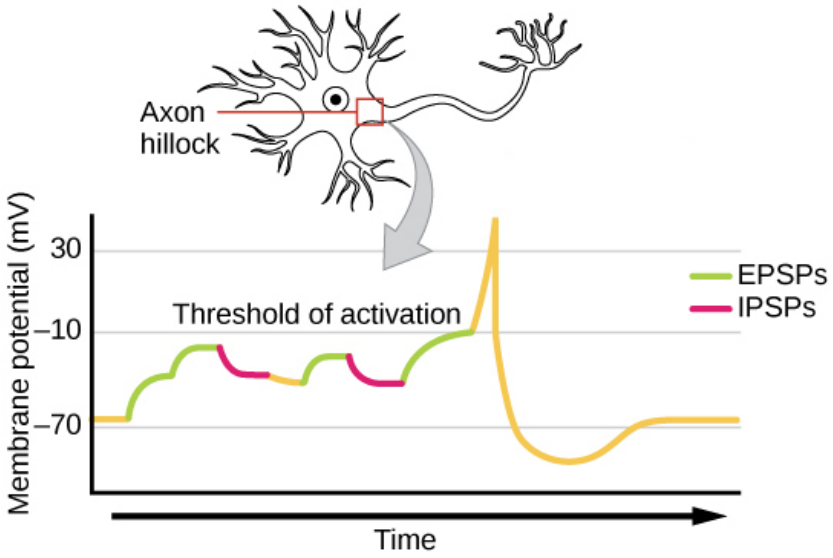


Figure 2.2: Signal summation to form an action potential [1]. Every jump of membrane potential represents the effect from a stimuli. In this specific case, an action porential is produced with the net function of six input spikes.

excitatory postsynaptic current (EPSC) and *inhibitory postsynaptic current* (IPSC), inducing potentiation and depression to membrane potential individually. The generated action potential is then transmitted through axon. At axon terminal, signals are separately passed to each single synapse, which will connect subsequent M neurons and form a 1-to-M relation. In total, an N-1-M network of signal paths is developed.

2.2 Synapse

What is a synapse? Where is it located exactly? Synapse is an abstract concept proposed to describe the structure that enables signal transmission through neurons. Specifically, it is found at noncontinuous joints between adjacent neurons with the presynaptic part located in axon terminals and the postsynaptic part on dendrites (see Fig.2.3). The synapse, separated by a gap called *synaptic cleft*, is not an integral unit. This physical barrier for electrical signals carried by one neuron to be transferred to another causes a "short circuit" in an electrical circuit. Thus, another type of messengers is invoked to help transmit these blocked signals in a chemical way, *neurotransmitters* (also called transmitters).

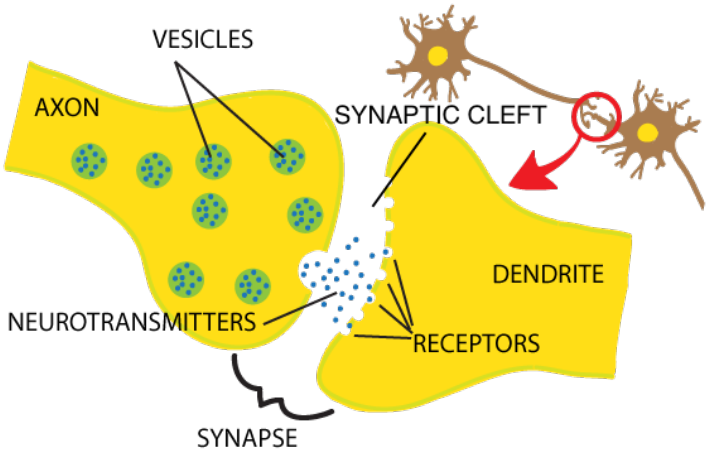


Figure 2.3: Detailed structure of synapse

When action potentials approach the axon terminals, ion channels on the membrane are opened, permitting small lipid bilayer vesicles in the axon, called *synaptic vesicles*, containing an enormous amount of neurotransmitters to efflux into the synaptic cleft. Fusion of vesicles with the membrane allows transmitters inside to be released into the synaptic cleft. Due to the concentration gradient, the transmitters will diffuse toward the dendrites on the postsynaptic neurons and bind to the corresponding receptors via *ligands* (there are other types of activation mechanisms though), inducing the activation of the particular receptor channels. Only then, the intracellular and extracellular ions are free to flow in between, generating action potentials in the postsynaptic neurons.

2.2.1 Synaptic Plasticity

While the synapse can transmit the input signals between the neurons, a particular learning rule is introduced along. The concept "synaptic weight" quantifies the learning output. The synaptic weight is potentiated or depressed depending on the analysis, or say learning of the current cell activities. This kind of weight adaptation ability is called *synaptic plasticity*. The underlying principle is the Hebbian theory [21]:

> Let us assume that the persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability. When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

To explain it more clearly, the synaptic weight of the synapses having the same trend as the local network is increased while that of the irrelevant ones is decreased. This selectivity gives a learning capacity to synapses. Meanwhile, in a long transmission round, the valued information is preserved in synapses, thus exhibiting a memory property of the synapses. Depending on the timescale, the synaptic plasticity is classified into two groups: short-term plasticity (STP) and long term plasticity (LTP), covering a time range of tens of milliseconds to a few minutes and from minutes to hours respectively [22]. Neuroscientists usually hypothesize learning rules of synapses and propose with specific models of plasticity that are able to explain the experimental observations. After that, those models will be build in software or hardware to further test their feasibility or be put into applications like sensory processing, robotics and brain-machine interface. In Chapter 2.3, several learning rules will be explained.

2.2.2 Synaptic Receptors

Depending on different ligand types, the effect of transmitter and receptor pairs on the postsynaptic neurons can be either excitatory or inhibitory, corresponding to positive and negative current flow to postsynaptic neurons. Different types of receptors display different temporal dynamics due to their distinctive conducting mechanisms (see Fig.2.4). A generic synapse structure does not capture the diverse temporal dynamics of different types of receptors in biological synapses, which are essential for realization of biophysically accurate neural behaviors in SNNs (details found in Sec.2.3.3). For this reason, two types of glutamatergic receptors (AMPA and NMDA) and one GABAergic receptor (GABAa) are discussed in details below as the foundation of further hardware implementations.

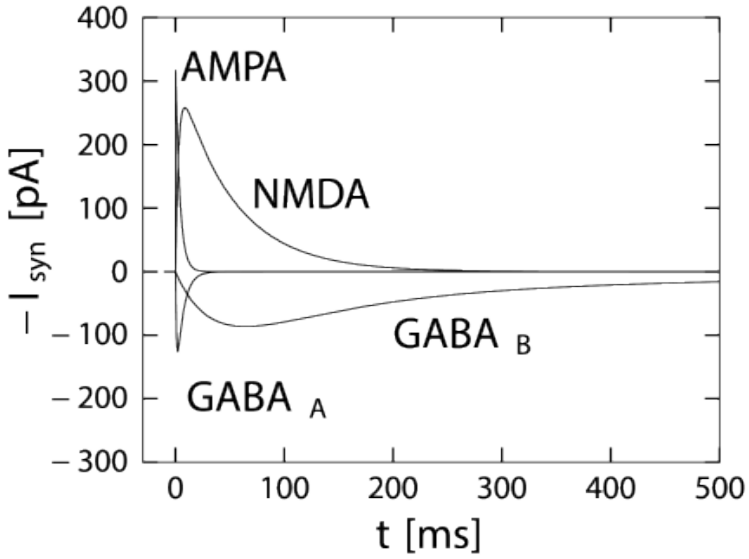


Figure 2.4: Temporal dynamics of four types of receptors [2]

2.2.2.1 AMPA receptor

The α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor, known as AMPA receptor, is one of the most common receptors in the nervous system. Mostly, the AMPA receptor is permeable to sodium (Na^+) via ion channels. Upon binding of transmitters on AMPA receptors, positively charged Na^+ enters the AMPA ion channels and depolarize the cell, thus inducing action potentials. However, the maximum conductance of AMPA receptors is limited by the intracellular calcium (Ca^{2+}) concentration. The prevention of calcium entry into the cell is reported to guard against excitotoxicity [23]. AMPA receptors open and close quickly due to a straightforward mechanism of channel opening and closing, and are thus responsible for fast signal transmission [24].

2.2.2.2 NMDA receptor

The ion channels of the N-methyl-D-aspartate receptor, also named NMDA receptor, is voltage-dependent, which is distinctive compared with other glutamatergic receptors. This dependency initially arises from the non-selectivity of its ion channels. When ligand-binding occurs, the non-selective ion channels are open to extracellular magnesium (Mg^{2+}) and zinc (Zn^{2+}), which will bind to specific sites on the receptor and block the channels for any other ions. To eliminate this blockage, a certain level of depolarization of the cell is necessary, usually through the influx of Ca^{2+} [25]. Once cleared, the ion channels introduce both Ca^{2+} and Na^+ into the target cell. At the same time, in response to the increased level of depolarization, more AMPA receptors are inserted into the membrane, creating more possibility of ion influx. Thus the conductance of NMDA receptor has a boost effect on the postsynaptic current.

To activate NMDA receptors, the presynaptic activities should introduce free transmitter to the dendrites while the postsynaptic depolarization should drive to open the ion channels on the receptors as a prerequisite. This kind of dual function of pre- and postsynapses implies the role of NMDA receptor in synchrony detection and biological emulation. On the temporal aspect, the NMDA receptors are typically three to six times slower than AMPA with regards to synaptic dynamics [11], which issues from a more complicated binding mechanism and small unchanneling speed.

2.2.2.3 GABA receptor

The gamma-Aminobutyric acid receptor, also called GABA receptor, is a primary inhibitory channel carrier in the nervous system. Two classes of the GABA receptors are defined according to different activation mechanisms, ligand-gated GABAa receptor, and protein-coupled GABAb receptor. A significant stimuli intensity is required to evoke the GABAb-mediated responses, which is hard to achieve in biological experiments or further obtain the estimation of GABAb receptors [11]. Thus here, we talk about the GABAa receptor for now.

The GABAa receptor is permeable to chloride (Cl^-). When activated, the GABAa receptor conducts Cl^- through the ion channels, causing the hyperpolarization of the cell and a lower possibility of neural firing. This inhibition function of the GABAa receptor is reported to be a prerequisite for balancing excitation and inhibition, thus stabilizing neural network [26]. The GABAa receptors have a similar temporal dynamics as AMPA, i.e. both the rise and the fall time of EPSCs are comparable. A comparison of temporal dynamics of above receptors is shown in Table 2.1.

Table 2.1: Biological dynamics for three receptors [11] [12]

	Polarity	Rise and Fall Times (ms)	Conduction Remarks
AMPA	+	0.4-0.8, 5	1-step, fast EPSC
NMDA	+	20, 100	2-step, voltage dependency, slow EPSC
GABAa	-	3.9, 20	1-step, fast IPSC

2.3 Model Extraction

To bridge from the biophysical systems to hardware implementations, models extracted from biophysical systems are discussed including the mesh network and an emphasis on the synaptic receptors and the synaptic learning rules.

2.3.1 Network

The nervous system is composed of repetitive units whether analyzed from cell level or cluster level or even higher levels. Fig.2.5 shows a $N \times M$ neural network model called a cluster model.

Information in the form of all-or-none spikes flows from axons to neurons modulated by connections between them, synapses (synaptic weight represented by w_{ij}). Thanks to the N-1-M network of signal paths mentioned in 2.1, a mesh distribution of synapses is observed in the cluster model. The binary input spikes $A_i(t)$ through a particular axon i at time t are transmitted to one row of synapses; an output neuron j integrates all the synaptic output in the same column over time. When its corresponding threshold is exceeded, an electrical spike is emitted, and then the membrane voltage is reset. The dynamics of the membrane voltage $V_j(t)$ of the output neuron j at time t can be summarized as below:

$$V_j(t) = V_j(t-1) - \lambda_j + \sum_{i=0}^{N-1} A_i(t) \cdot w_{ij} \tag{2.1}$$

$$if\ V_j(t) \geqslant \alpha_j,\ spike,\ reset \tag{2.2}$$

λ_j is the leakage of the output neuron. The summation term represents the integration of all the synapses connected to the target neuron. If the firing requirement is met, the cell generates one electrical spike and resets itself.

2.3.2 Synapse

In the synapse model, the specific transmitter activity in the synaptic cleft is ignored since a straightforward and clear representation of the synapse is expected for the current phase of emulations. One of the basic and direct models describing the synaptic conductance properties is the exponential decay model where the rising phase of the synaptic conductance is assumed to be instant [27], i.e. the release of transmitters, its corresponding diffusion across the cleft, the receptor binding, and channel opening all happen very fast. The conductance of the synapse at time t is then:

$$g_{syn}(t) = \bar{g}_{syn} \cdot e^{-(t-t_0)/\tau} \cdot \Theta(t - t_0) \tag{2.3}$$

where $\bar{g}_{syn}$ is the maximum conductance of the synapse, t_0 is the onset time of the presynaptic spike while τ is the decaying time constant, and $\Theta(x)$ is the Heaviside step function.

The exponential profile is a match with the relationship between the ionic conductance of a neuron and its membrane potential [19]. Regardless of the time boundary item $\Theta(t-t_0)$, $\bar{g}_{syn}$ represents the synaptic weight after certain learning while $e^{-(t-t_0)/\tau}$ induces biologically analogous exponential conversion. For some inhibitory synaptic currents (IPSCs), the exponential decay model is validated to outline their activities because the rising phase of these currents is much shorter compared with the decaying phase,

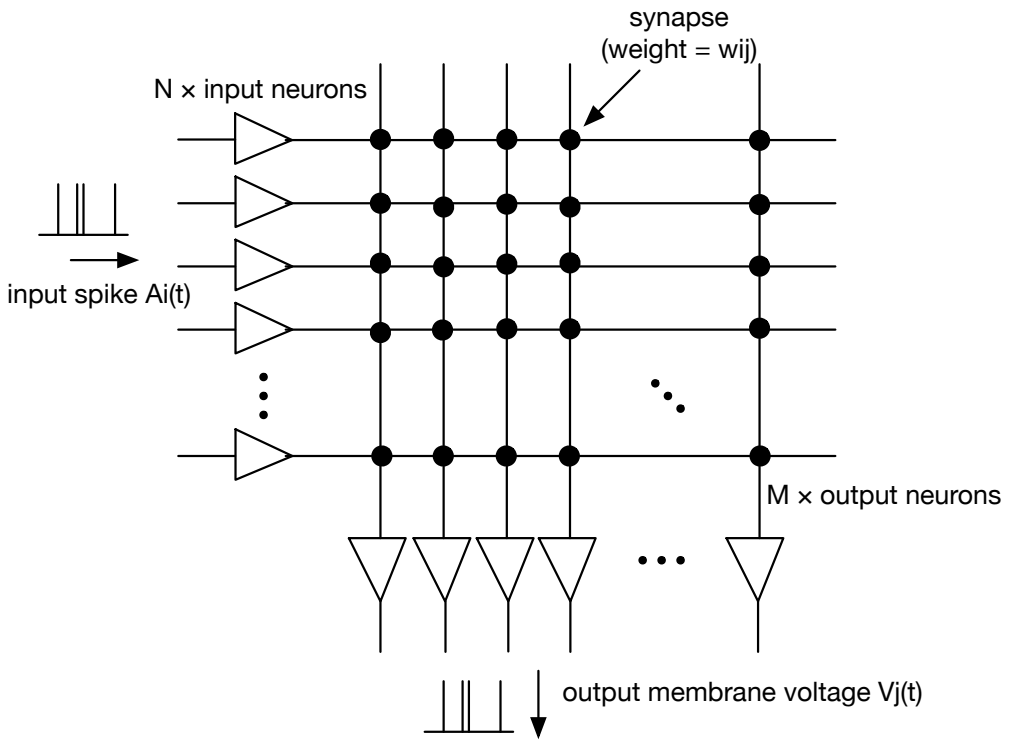


Figure 2.5: The cluster model of neural network. The triangles represent the N input axons and the M output neurons respectively. The synapses are represented by black dots. The straight lines display the interconnections between neural elements.

like the GABAa-induced currents (see Table 2.1). For the fast excitation contributor, AMPA receptors, the exponential decay model is also decent enough. However, when it comes to EPSC induced by NMDA receptors, which has comparable temporal dynamics in both rising and decaying phases, the model fails to emulate its behaviors. A more detailed model with two separate exponential components is introduced:

$$g_{syn}(t) = \bar{g}_{syn} \cdot f \cdot (e^{-(t-t_0)/\tau_{decay}} - e^{-(t-t_0)/\tau_{rise}}) \cdot \Theta(t - t_0) \tag{2.4}$$

The factor f is used to normalize the total amplitude of the sum to $\bar{g}_{syn}$, and τ_{decay} and τ_{rise} are the time constant for decaying and rising phase, respectively. The neuromorphic design for this model is more complex due to one extra rising phase consideration.

In Chapter 2.1, it is argued that a transistor-based analog implementation of nerve cells is a fundamental and potential method for dedicated and efficient brain-like emulations on VLSI scale. The transistor operating in the sub-threshold region is of the most interest mainly for two reasons. The first reason is its exponential relationship between the drain currents and the gate voltages analogous to the biological model. The second reason is its extremely low power consumption. The drain current of a transistor is

shown below:

$$I_{ds} = I_0 \cdot e^{\kappa_n V_g / U_T} \cdot (e^{-V_s/U_T} - e^{-V_d/U_T}) \tag{2.5}$$

where I_0 is the current value when V_{gs} equals the threshold voltage of the transistor, acting as a current scale control, and κ_n represents the sub-threshold slope for n-type MOSFET. V_g, V_d, V_s and U_T are the gate, drain, source voltages relative to the bulk and the threshold voltages of the transistor. When $V_{ds} \geqslant 4U_T \approx 100mV$, the transistor enters the sub-threshold saturation region where a pure exponential relationship between V_{gs} and I_{ds} is generated:

$$I_{ds} = I_0 \cdot e^{\kappa_n V_{gs} / U_T} \tag{2.6}$$

This is the region we expect to obtain an analogous behavior of biological neurons.

Regardless of the distinctions among different synapses, a general structure can be decomposed into two functional units: the synaptic learning block inducing the weight adaptation and the distinctive receptor compartments bringing in chemical ligand-gated channel control. In next subsection, the mathematic models of the synaptic learning rules will be explained.

2.3.3 Learning Rules

Synaptic plasticity models of various complexity levels [16] have been proposed in regards to various application requirements, ranging from the abstract ones to the more biologically realistic and detailed ones. In this chapter, two popular algorithms are explained.

2.3.3.1 Pair-Based STDP

Pair-based Spike timing dependent plasticity (PSTDP), an almost symmetrical pattern of Hebbian's theory, is a learning process that can adapt the synaptic weight according to temporal correlations between pre- and post spikes of a target synapse. These correlations should be within milliseconds time range in accord with biological temporal features: If the pre spike precedes the post spike, a potentiation of the synaptic weight occurs; In contrast, if a reversed sequence happens, depression is induced. A repeat of one of the two patterns evokes long-term effect, LTP and LTD respectively. A diagram of the STDP learning function is illustrated in Fig.2.6. Two factors of concern in this learning window are time constants (τ) and amplitudes (A) of two phases. The time constant indicates the temporal range where the correlation happens while the amplitude controls the adaptation level. The STDP rule is expressed as below:

$$\Delta w^+ = A^+ \cdot e^{-\Delta t/\tau_+} \qquad \Delta t > 0 \tag{2.7}$$

$$\Delta w^- = -A^- \cdot e^{\Delta t/\tau_-} \qquad \Delta t < 0 \tag{2.8}$$

where Δt is the temporal difference between a single pair of post- and pre spikes. A_+ and A_1 are the maximum amplitude while τ_+ and τ_- are time constants of the potentiation and the depression phase respectively. These parameters impact the area of the weight update curves during potentiation and depression. It is observed that stable learning is realized when the aggregate area of depression exceeds that of potentiation in the weight update function. On the contrary, weaker depression results in the extreme potentiation of synaptic weights and the eventual shorting of outputs to inputs. This behavior prevents the realization of any practical network transfer function.

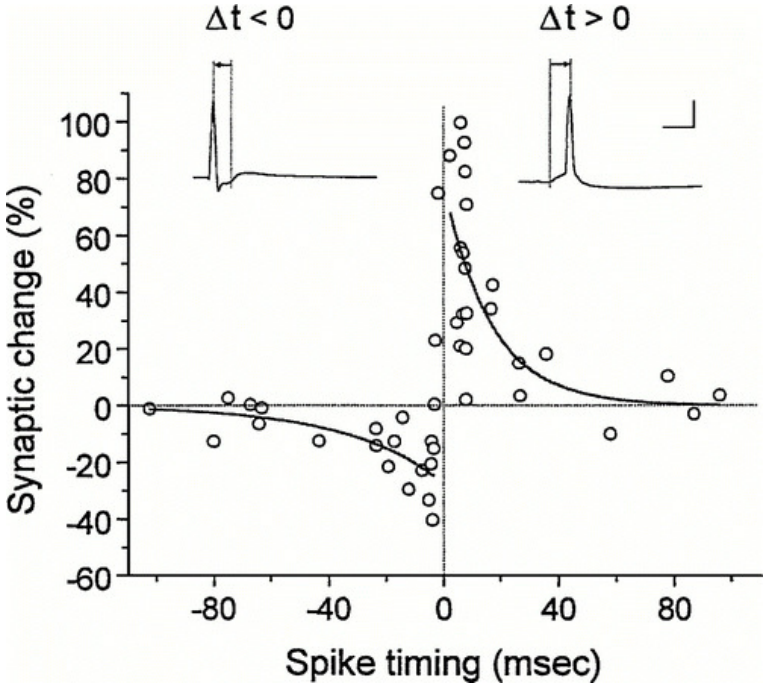


Figure 2.6: The learning window of STDP learning rule. The hollow circles are experimental data of EPSC amplitude percentage change at 20-30 min after repetitive stimuli of pre and post spikes at a frequency of 1Hz [3]. The spike timing is defined as the temporal interval between post and pre spikes. An exponential fit of those data points is outlined with two smooth curve (LTP and LTD). For LTP and LTD respectively, A=0.777 and 0.273; τ=16.8 and 33.7 ms.

In the biological experiments [9], it is found that LTP is apparently strengthened with relatively small initial weight while LTD does not display too much dependency. For this reason, certain initial weight dependence rules are supposed to be introduced to synapse models. What exact initial weight dependence rule do they follow? The additive [28], the multiplicative [29] update rules, or somewhere in between, also named power law update rule[30] are categorized. The detailed rules of the weight dependency are listed in Table2.2. The synaptic weight is denoted by w $(0 < w < 1)$, $\lambda \leq 1$ is the learning rate, α is an asymmetrical parameter, μ drives the combination point between additive and multiplicative models. For example, the learning function in Eq.(2.7) and (2.8) does not involve weight dependence in both A_+ and A_-, which gives it an additive update features. From the biological aspects, however, the depression phase turns out

to be in a good match with the additive rules while the potentiation phase can not be accurately determined [16].

Table 2.2: Weight dependency rules

	Additive	Multiplicative	Power Law
+	$A = \lambda$	$A(w) = \lambda(1-w)$	$A(w) = \lambda(1-w)^{\mu}$
-	$A = \lambda\alpha$	$A(w) = \lambda\alpha w$	$A(w) = \lambda\alpha w^{\mu}$

Different equilibrium distribution of synaptic weight are displayed using different weight dependency rules [16]. Additive update rule leads to a balanced bimodal distribution of synaptic weight while the multiplicative one to unimodal distribution. The former one is stable for long term period, which is essential for LTP. It involves strong competition among synapses. However, the latter one reconciles with the reported experimental data while long term stability is absent [16].

2.3.3.2 Triplet-Based STDP

Derived from the pair-based STDP, triplet-based STDP (TSTDP) incorporates the correlations among three consecutive spikes. The mathematical representation of TSTDP learning rule is given by:

$$\Delta w^{+} = e^{-\Delta t_1/\tau_{+}} \cdot (A_2^{+} + A_3^{+} \cdot e^{-\Delta t_2/\tau_y}) \tag{2.9}$$

$$\Delta w^{-} = -e^{\Delta t_1/\tau_{-}} \cdot (A_2^{-} + A_3^{-} \cdot e^{-\Delta t_3/\tau_x}) \tag{2.10}$$

τ_{+}, τ_{-}, τ_y, τ_x are time constants concerning triplet spikes (details shown in Fig.2.7). A_2^{+} and A_2^{-} are second-order potentiation and depression amplitude parameters while A_3^{+} and A_3^{-} are third-order amplitudes. Δt_1, Δt_2 and Δt_3 represent $t_{post}(n) - t_{pre}(n)$, $t_{post}(n) - t_{post}(n-1) - \epsilon$ and $t_{pre}(n) - t_{pre}(n-1) - \epsilon$ respectively with ϵ a small positive constant to ensure that the weight update uses the correct values occurring just before the target pre- or postsynaptic spike.

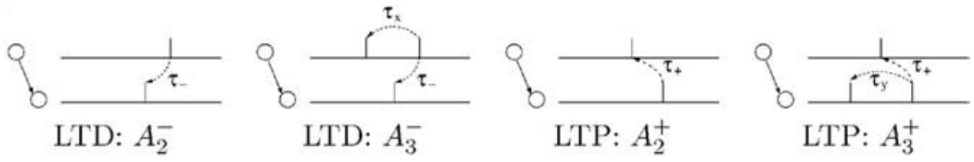


Figure 2.7: Time constant parameter interpretation of TSTDP

While the PSTDP can capture the basic learning algorithms of the synapse, i.e. the Hebbian's theory, it is shown to be a insufficient model to display the frequency effect and higher-order outcomes observed in biological models [31]. The PSTDP is a substantially linear model where the overall effect of consecutive spikes may sometimes

counteract with each other, thus restricting the frequency effect of the synapse. In contrast, this nonlinearity is demonstrated in the third-order spike patterns. The TSTDP is capable of reproducing triplet and quadruplet outcomes in experiments. Unfortunately, the design complexity increases with the higher biological similarity, leading to the redundancy in hardware implementation.

2.4 Conclusion

This chapter introduces the biological working principles of single nervous elements, neuron and synapse and the neural network. The synapse, as the focus of this thesis, is further elaborated from its functionality, synaptic plasticity, to its biological fidelity, receptors. Finally, mathematic models are extracted, ranging from the single synapse, synaptic learning rules to networks to bridge from the biophysical systems to hardware implementations. In the next two chapters, the hardware implementation space of the learning rules and the synaptic receptors is discussed.

3 Component Implementations

In the course of neuromorphic synapse exploration, a variety of synaptic architectures are proposed depending on specific application domain or derived from different biological data. This chapter plans to introduce three main types of synaptic learning rules discussed in Sec.2.3.3 in hardware implementation, from classic pair-based STDP to more advanced pair-based STDP and finally to the biologically realistic TSTDP. STDP. Several crucial points are worth to be discussed: how close is the designed architecture to the biophysically realistic models? What is the weight dependence of synaptic plasticity? What are the trade-offs? These are the main aspects to be discussed in the following pages.

The circuits are implemented in UMC 65nm technology. Considering the fact that the biological signals usually span over milliseconds to seconds range, one of the detrimental elements interfering the synaptic performance would be the spontaneous leakage of transistors and capacitors. For this reason, special low leakage transistor (LL) technology is chosen in all our designs. In addition, various supply voltage and threshold voltage alternatives are supported in this technology.

3.1 Learning Rule 1: Classic STDP

In classic STDP learning rule, the potentiation and depression display an almost symmetrical dynamics with each other, which should also be demonstrated in circuit. Shown in Fig.3.1(a), the circuit [4] has a main branch (M1-M6) responsible for pulling up or down the synaptic weight stored in a weight capacitor C_w. Additionally, two leaky integrators are implemented to offer the controllability of both potentiation and depression windows (V_{pot} and V_{dep}) in typical learning window (see Fig.2.6).

A complete circuit design is displayed in Fig.3.1(b). The potentiation current I_{pot} is initially drained from transistor M1, which is biased by V_{pot}. V_{pot} is rested at high rail voltage for most of the time but is pulled down via charging from M9 when a pre spike pulse approaches for a short period of time. After that, V_{pot} slowly discharges through M7-M8. The discharge occurs in an almost linear pattern if the channel length modulation effect is ignored. V_{tp} controls the discharge speed, thus determining the decay time constant of potentiation phase τ_+ in the learning window in Sec.2.3.3. To induce exponential dynamics, transistor M1 is maintained in saturated sub-threshold region where stable exponential dynamics happen ($V_{ds} > 100mV$). At certain point of the decay process, a post spike is transmitted and further activates the upper main branch. The generated current from upper branch is injected to C_w, causing an increment in

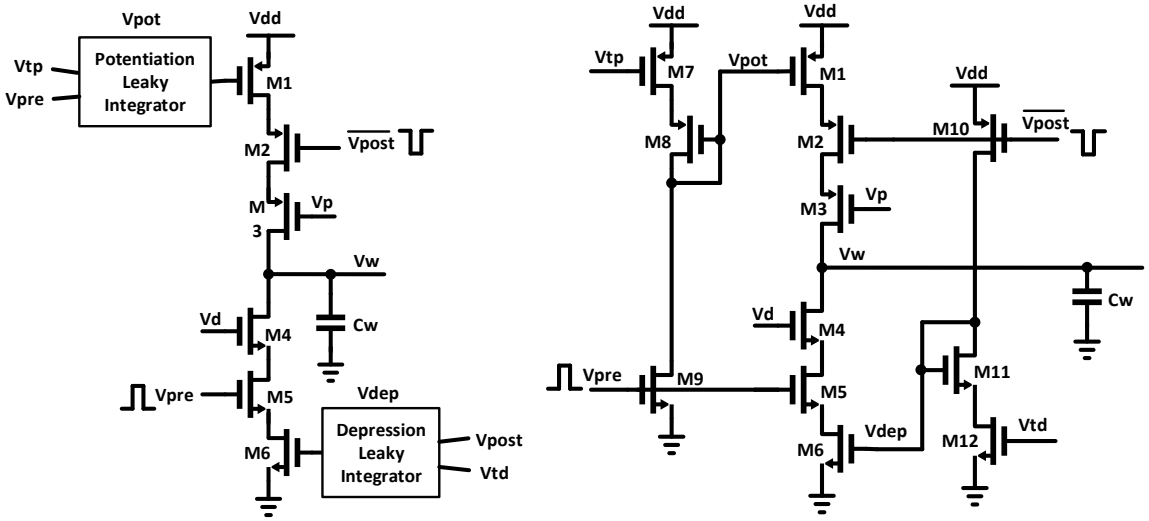


Figure 3.1: The simplified structure diagram of the classic PSTDP circuit [4] is shown in (a) with the leaky integrators represented by squares boxes. (b) shows the detailed circuit.

synaptic weight V_w. V_p acts as a controllable source, adjusting the amount of current to be allowed to flow into C_w and thus determining the amplitude (A_+) of the potentiation window. The same mechanism applies to the depression phase of the synapse with a complementary design.

This PSTDP circuit offers a low power consumption (activated only when spike occurs) and simplicity. Though regarded as a typical PSTDP model, the design faces one detrimental problem: the leaky integrators in both potentiation and depression phases can not guarantee a voltage range that drives M1 and M6 into the sub-threshold region, i.e. this circuit can not offer an exponential dynamics expected in real neural cells, thus reducing its biological fidelity. The detailed dimensional information of this circuit is listed in Table.3.1.

Table 3.1: Transistor and capacitor dimensions for classic PSTDP circuit

	Length(μm)	Width(μm)
M1	0.06	3.5
M2	3	0.08
M3	0.06	0.08
M4	0.06	0.08
M5	3	0.08
M6	3	0.08

	Length(μm)	Width(μm)
M7	0.06	0.08
M8	0.06	0.08
M9	0.06	0.08
M10	0.06	0.08
M11	0.06	0.08
M12	0.06	0.08

	Size(pF)
Cw	0.5

3.2 Learning Rule 2: Advanced STDP

The circuit [5] in Fig.3.2 partially solves the problem mentioned above. The general idea of pulling up or down the synaptic weight through injection or efflux of activated currents from the weight capacitor is similar shown in Fig.3.2(a). The main difference is that the amplitude control block is incorporated into a so-called *advanced leaky integrator* circuits (the blue dashed blocks in Fig.3.2(b)) representing both potentiation and depression learning windows. Besides, an additional weight dependence circuit (the red dashed line block in Fig.3.2(b)) in supplement to biological features.

Unlike the previous circuit in Fig.3.1, the bias voltage V_{pot} loaded to the main branch upon the arrival of pre spike is adjustable through a controllable current source I_{bpot}. Again, the post spike within the time range of the learning window activates I_{pot} and removes charges stored in C_w at certain point of time. It should be noted that this circuit follows a complementary design flow, i.e. the current efflux from C_w means an increment of synaptic weight. The time constant of the learning window is determined by M7 biased by V_{bpot}. Without the interference of the amplitude control on main branch, this circuit is able to achieve desirable exponential relation of synapses. The same analysis applies to the depression phase on the right side. A notable simplification is that the advanced leaky integrator activated by post spikes is shared by all the synapses connected to the target neuron. The depression currents for individual dendritic synapses are generated via duplicated current branches, allowing large area saving, in VLSI network.

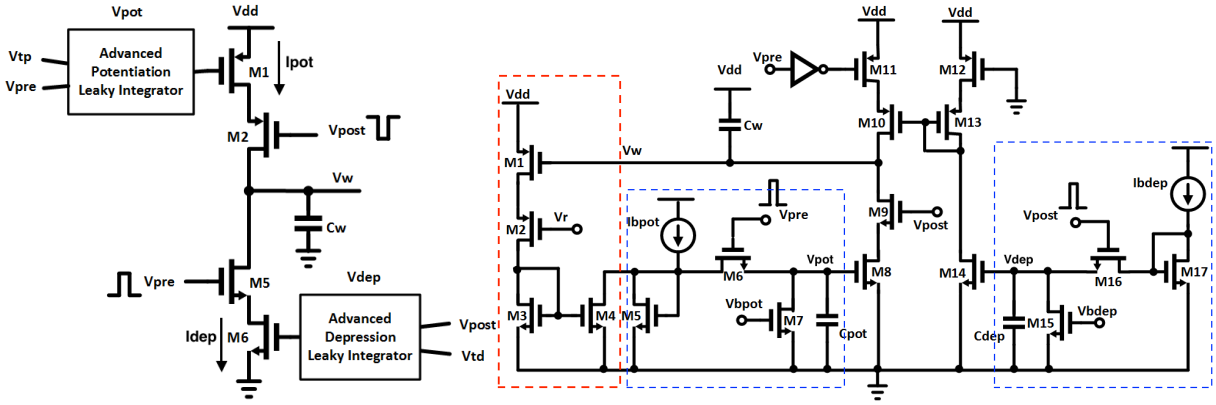


Figure 3.2: The simplified structure diagram of the advanced PSTDP circuit with weight dependence [5] is illustrated in (a) with the advanced leaky integrators represented by square boxes. (b) shows the circuit details.

The circuit includes the weight dependence feature (built by M1-M4) in addition to the basic STDP learning rule. M1 is a low-gain transistor operating in the strong inversion region while M2 is a high-gain transistor operating in the sub-threshold region. To enlarge the possible synaptic weight range, a low threshold voltage transistor is used

for M1. The bias voltage V_r tunes the weight dependence level. When V_w decreases, i.e. the synaptic weight increases, the current through M1-M2 increases. M1 will be driven to work in the linear region by M1, which means that a change of synaptic weight is converted linearly to a current subtraction from I_{bpot}. M5, a diode connected transistor, works in the sub-threshold as M5 and M8 share the same bias voltage upon the arrival of a presynaptic spike, and M8 has to be maintained in the sub-threshold region to offer an exponential relation. The linear decrement of current in M5 is thus mapped to M8 through a current mirror mechanism. In this way, an multiplicative weight dependence (see Table.2.2) is achieved by adding four extra transistors. The transistor and capacitor dimensions are listed in Table.3.2

Table 3.2: Transistor and capacitor dimensions for advanced PSTDP circuit

	Length(μm)	Width(μm)
M1	3	0.08
M2	0.06	2
M3	0.06	0.08
M4	0.06	0.08
M5	0.06	0.08
M6	0.06	0.08
M7	0.06	0.08
M8	3	0.08
M9	0.06	0.08

	Length(μm)	Width(μm)
M10	3	0.08
M11	0.06	0.08
M12	0.06	0.08
M13	3	0.08
M14	0.06	0.08
M15	0.06	0.08
M16	0.06	0.08
M17	0.06	0.08

	Size(pF)
Cw	0.5
Cpot	0.7
Cdep	0.6

An extra continuously activated current branch (M1-M3), especially in a design where most of the transistors conduct only upon the presence of pre or post spikes, will cause a much larger power consumption in comparison with the circuit without the weight dependence.

3.3 Learning Rule 3: Triplet-Based STDP

From Eq.(2.7)(2.8) and Eq.(2.9)(2.10), it can be observed that the second-order components in TSTDP model are the same as in the PSTDP model. The distinction of TSTDP arises from the third-order components which is summed together with the amplitude components (A_2^+ and A_2^-) before an exponential conversion. Based on this observation, the second-order design in Sec.3.2 can be well used in TSTDP implementations whereas extra third-order implementations need to be supplemented. In the classic PSTDP circuit, the amplitude control of both potentiation and depression learning window is blended into the main branch, namely the amplitudes can not be separated to be further added to the third-order components. The advanced PSTDP with separate tuning blocks of learning window is a feasible method.

A triplet-based STDP circuit [6] is shown in Fig.3.3. Four individual parts can be

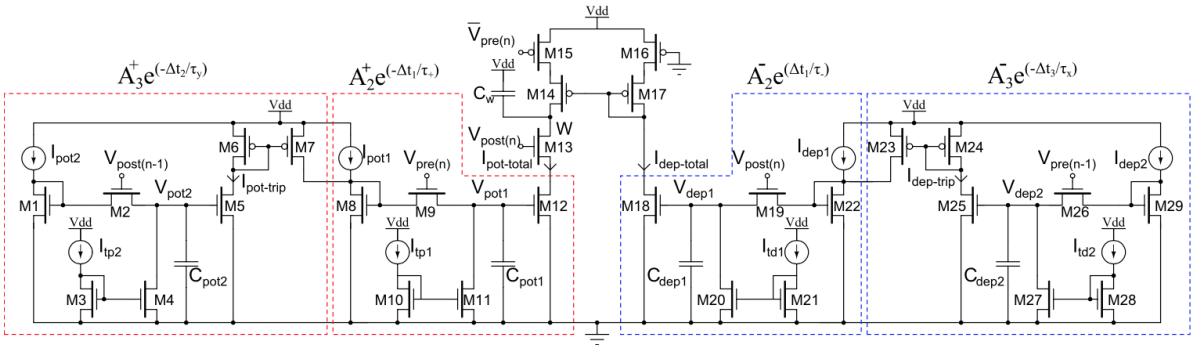


Figure 3.3: Triplet-based STDP circuit[6]. The red and blue dashed blocks represent potentiation and depression learning block respectively.

easily identified in the circuit in match with two second-order and two third-order components in triplet learning algorithm. The middle two parts are similar design as the advanced PSTDP circuit (details can be found in Sec.3.2) generating second-order potentiation and depression window activated by the current pre spike $V_{pre}(n)$ and post spike $V_{post}(n)$ respectively. The leftmost and rightmost parts are built by the same advanced leaky integrator but are activated by the previous pre spike $V_{pre}(n-1)$ and post spike $V_{post}(n-1)$. The resultant current from those two blocks are mirrored and summed with the corresponding amplitude control current I_{bpot} and I_{bdep}, introducing the third-order effects into the second-order ones. The component dimensions are listed in Table.3.3.

Table 3.3: Transistor and capacitor dimensions for TSTDP circuit

	Length(μm)	Width(μm)
M1	0.06	0.08
M2	0.06	0.08
M3	0.06	0.08
M4	0.06	0.08
M5	03	0.08
M6	0.06	0.08
M7	0.06	0.08
M8	0.06	0.08
M9	0.06	0.08
M10	0.06	0.08
M11	0.06	0.08
M12	0.06	0.9
M13	0.06	0.08
M14	3	0.08
M15	0.06	0.08

	Length(μm)	Width(μm)
M16	0.06	0.08
M17	3	0.08
M18	0.5	0.08
M19	0.06	0.08
M20	0.06	0.08
M21	0.06	0.08
M22	0.06	0.08
M23	0.06	0.08
M24	0.06	0.08
M25	3	0.08
M26	0.06	0.08
M27	0.06	0.08
M28	0.06	0.08
M29	0.06	0.08

	Size(pF)
Cw	0.5
Cpot1	0.7
Cpot2	0.7
Cdep1	0.6
Cdep2	0.6

The incorporation of the triplet spikes enlarges the biological fidelity of the synapse at an expense of a doubled area and power consumption, which should be taken into account in large-scale network design.

3.4 Synaptic Weight Storage

Multiple options exist for the storage of synaptic weights. Traditional capacitive storage employs bulky capacitors to lower the effect of leakage. While this is mitigated by the use of digital memories [32], they require current-mode ADCs and DACs for each neuron, adding complexity. Similarly, floating gate memories [33] offer an effective means for long term synaptic weight storage due to their non-volatility. However, the precise programming of synaptic weights is challenging. The synapse can also incorporate additional mechanisms for information storage. One of these is bistability [4] (shown in Fig.3.4), which due to its low area and low power consumption, is a comparably efficient storage medium. Synaptic weights drift towards one of two voltage rails depending on their relative value compared to the bistability threshold. These dynamics lend great robustness to the state storing in synapses against stochastic background events [34].

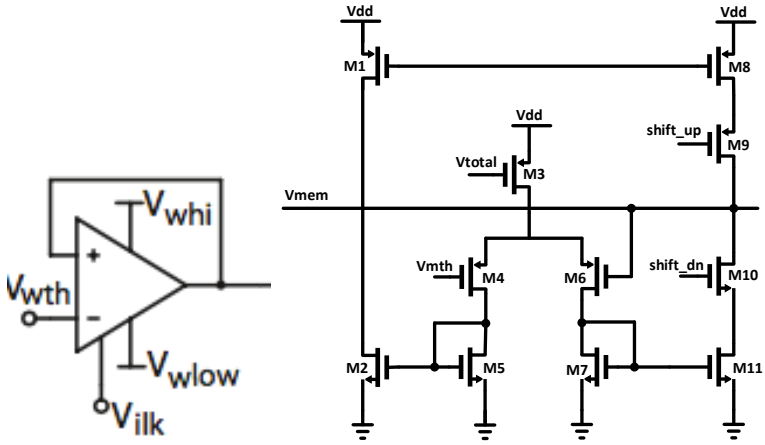


Figure 3.4: Analog bistability circuit [7]

3.5 Synaptic Receptors

The synaptic weight update is converted to the synaptic currents, which will be integrated by the postsynaptic neuron. This is achieved through the receptors, i.e. functional integrators transferring the weight information to EPSCs or IPSCs upon upcoming stimuli. Diverse structures are applied to capture the distinctive dynamics of receptors.

3.5.1 AMPA Receptor

A differential-pair integrator (DPI) [8] structure is applied to emulate the fast rising and decaying dynamics of AMPA receptor shown in Fig.3.5(a). The receptor potential state V_{syn} stored in C_{syn} is decremented fast with every efflux of charges from M2, M3 and M4 path upon the arrival of presynaptic spikes, and then decays linearly towards the high rail from M5. The EPSC is obtained via an output transistor M6 operating in the sub-threshold region. Four characteristics of this circuit should be highlighted:

- The differential structure of M1 and M4 offers a "seesaw" like control over the currents through these two branches. When the synaptic weight is maintained, the current efflux in path M2, M3 and M4 can be adjusted through bias M1. The driving ability of the synapses to the postsynaptic neurons is thus flexible depending on diverse configurations between the target neuron and the corresponding dendritic synapses.
- The decaying time constant τ_{AMPA} is tunable through the bias V_τ in dealing with the various experimental data obtained in different experiments or receptor locations in the brain.
- The circuit follows a linear differential equation dynamic, enabling direct summation of identical receptor sources.
- The circuit has a compact structure and a low power consumption because the main circuit mostly conducts only in presence of the presynaptic spikes, which lasts for no more than 2 ms.

3.5.2 NMDA Receptor

Unlike the single exponential dynamics used for AMPA receptor, the charging phase of NMDA receptor can not be ignored due to its relatively large portion in the whole temporal range. Thus a double exponential function should be displayed in NMDA receptor design as well as its distinctive weight dependence. In Fig.3.5(b), the presynaptic spike enables a instantaneous current influx into C_{rise} in the rising phase, the amplitude of which is controlled by V_w. The bias V_{taur} determines the discharge speed of C_{rise}. During this controllable period of time, the transistor M4 is always active, inducing the voltage drop of V_f. After that, capacitor C_{syn} begins to discharge through M8 biased by V_{tauf}, adjusting the falling time constant. In this way, controllable double exponential dynamics are generated.

To incoporate the distinctive voltage dependence of NMDA receptors, a differential pair is added to the circuit, forming a comparison between V_{mem} and V_{mth}. When the postsynaptic neuron is depolarized, V_{mem} surpasses V_{mth}, introducing valid current flux into C_{syn}. On the contrary, if V_{mth} surpasses V_{mem}, no or only small fraction of current is induced to generate EPSCs.

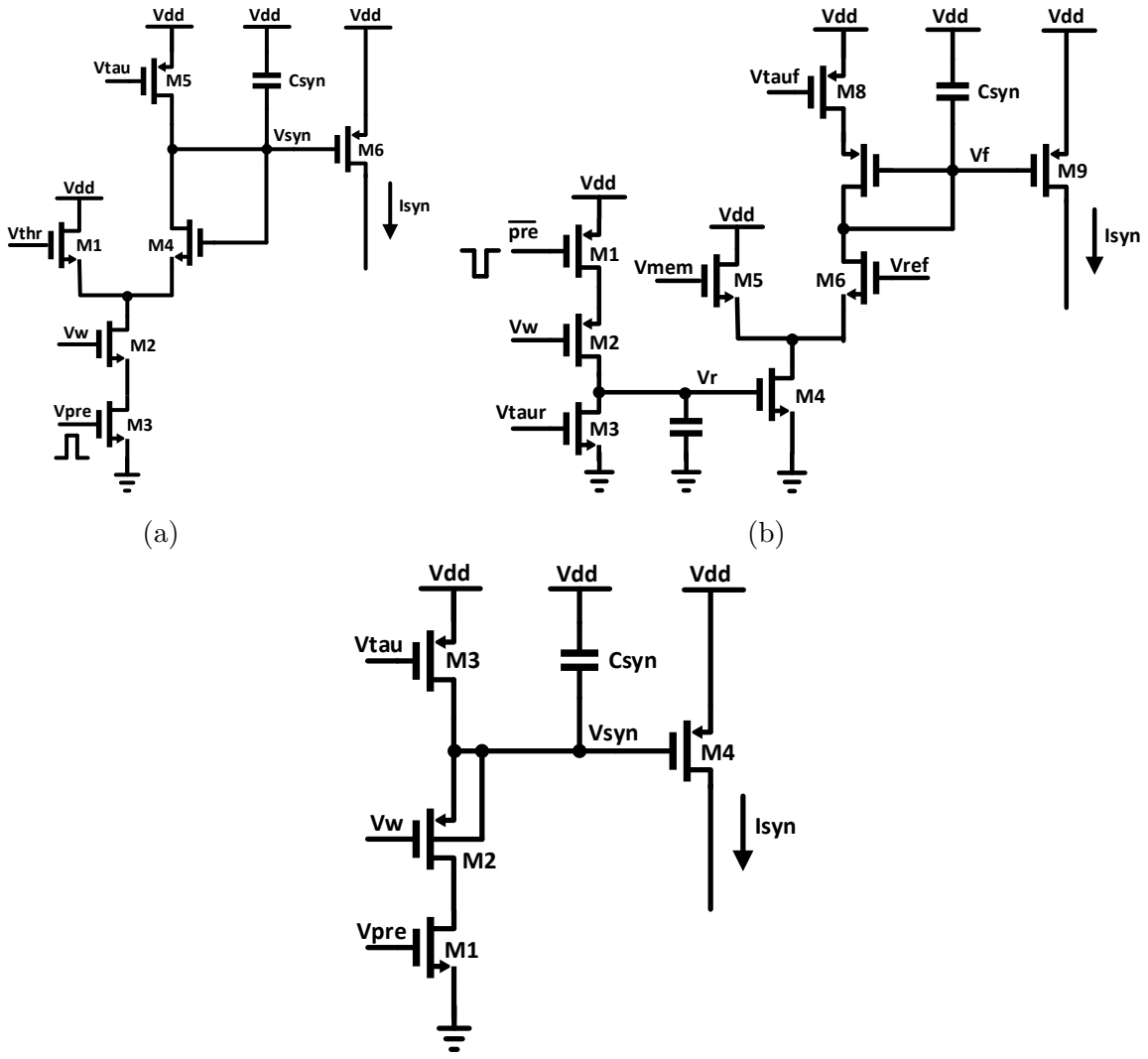


Figure 3.5: Receptor implementations according to their distinctive dynamics [8]. (a) AMPA receptor using DPI structure; (b) NMDA receptor using two-stage conduction mechanism; (c)GABAa using log-domain integrator structure.

3.5.3 GABA Receptor

Due to the similarity between the dynamics of AMPA and GABAa receptors (except for the polarity), a complementary design of AMPA synapse (DPI synapse) can be efficiently applied to induce IPSCs. Three control voltages are needed. However, since the inhibitory synapses do not exhibit learning properties, the inhibitory level is independent on the synaptic weight, for which the synaptic weight value is maintained the same. It is not necessary to have two control voltages over the inhibitory level (V_{thr} and V_w). A log-domain integrator can be well chosen as the implementation of the GABAa receptor for its simplicity as well as a linear dynamics [8] (see Fig.3.5(c)). The

transistor and capacitor dimensions are shown in Table.3.4.

Table 3.4: Transistor and capacitor dimensions for receptors

AMPA	Length(μm)	Width(μm)
M1	0.06	0.08
M2	3	0.08
M3	0.06	0.08
M4	0.06	0.08
M5	0.06	0.08
M6	3	0.08

	Size(pF)
Csyn	1

NMDA	Length(μm)	Width(μm)
M1	0.06	0.08
M2	0.06	0.08
M3	3	0.08
M4	0.06	0.08
M5	0.06	0.08
M6	3	0.08
M7	0.06	0.08

	Size(pF)
Csyn	1

GABAa	Length(μm)	Width(μm)
M1	0.06	0.08
M2	3	0.08
M3	0.06	0.08
M4	0.06	0.08

	Size(pF)
Csyn	1

3.6 Conclusion

In this chapter, three types of learning algorithms implemented in hardware are discussed. The analysis focuses on the reconfiguration flow of the designs. Next, a bistability circuit implementation is introduced to achieve a balanced weight distribution. Finally, the details of three types of receptors are explained. In next chapter, the characterizations of the built elements in this chapter will be conducted.

Component Charaterizations 4

In the previous chapter, three types of spike-timing-based synaptic learning circuits are introduced and implemented in analog hardware: the classic STDP, the advanced STDP and the triplet STDP circuits. In this Chapter, characterizations of those architectures will be conducted. First of all, the chosen learning algorithm should be able to process signals in biologically plausible time range, i.e. of the order of tens of milliseconds. Secondly, the learning should display an exponential dynamics observed in real neurons. Finally, in pursuit of brain-like large integrations, both power and area consumptions should be maintained as low as possible.

Additionally, the distinctive properties of three types of receptors, AMPA, NMDA and GABAa, are demonstrated individually as well as the joint function of combinations in a unit network called a *cluster* of neural network. These components will be further integrated into a two-layer recurrent neural network to demonstrate the functionality of multi-compartment synapse structure.

4.1 Learning Rule 1: Classic STDP

As described in Sec.3.1, the circuit adjusts the amplitudes and time constants of the STDP learning window in both potentiation and depression phases through V_{tp}, V_{td} and V_p and V_d biasing voltages, which influences the learning levels and ranges of correlation between spikes. It gives flexibility of various configurations depending on experimental needs or the development of brain explorations. In Fig.4.1, as an example, several biasing voltages are given to V_{tp} and V_{td}, generating different time constants of the leaky integrator voltages V_{pot} and V_{dep}. The charging and discharging processes in the leaky integrator circuits are stored in the parasitic capacitors which are of fF level, which sets an upper bound to the time constants that can be obtained through the leaky integrator blocks. Another detrimental problem here is that transistor M1 and M6 are supposed to operate in the saturated sub-threshold region to offer the plausible exponential dynamics observed in real neurons. The leaky integrator circuit is not able to maintain V_{pot} and V_{dep} in the preferred region, i.e. below the threshold voltage of M1 and M6. Thus, the exponential dynamics are absent in the learning algorithm. Fig.4.5 shows the synaptic weight evolvement and the membrane voltage distribution with Poisson distributed presynaptic and postsynaptic input signals of 200 Hz. In all our experiments, pulse width is set as 100 μm. The neuron implementation will be introduced in Sec.4.4.

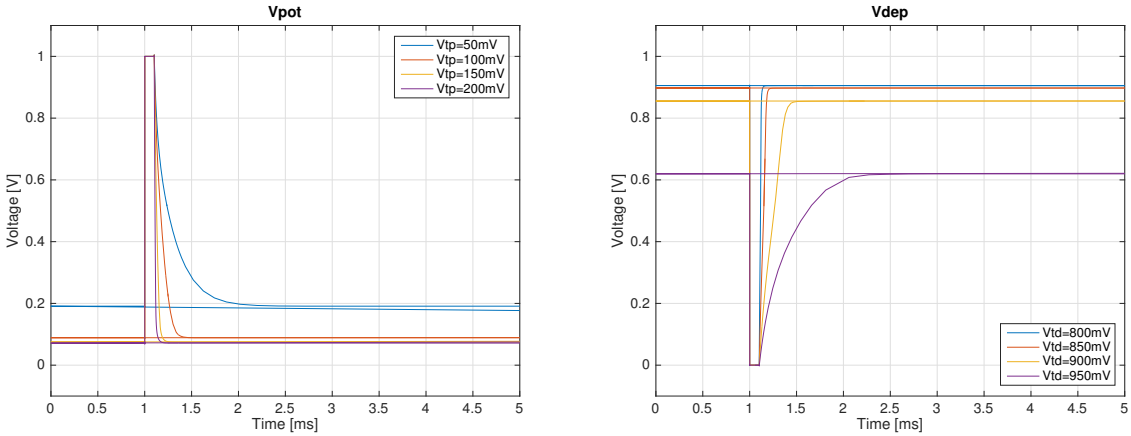


Figure 4.1: Time constant control of V_{pot} and V_{dep} in the classic STDP circuit via biasing voltage V_{tp} and V_{td}

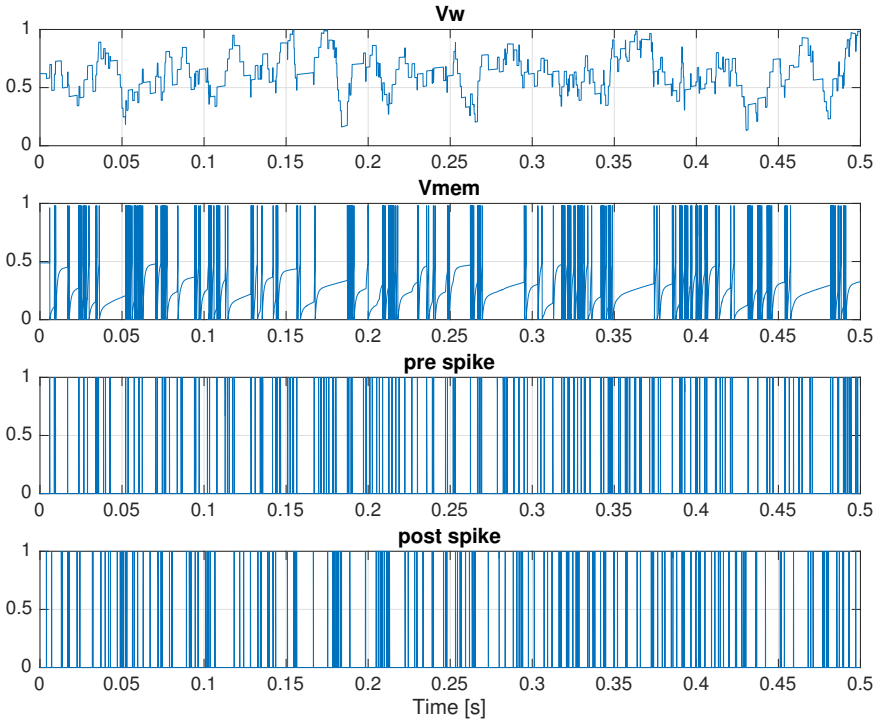
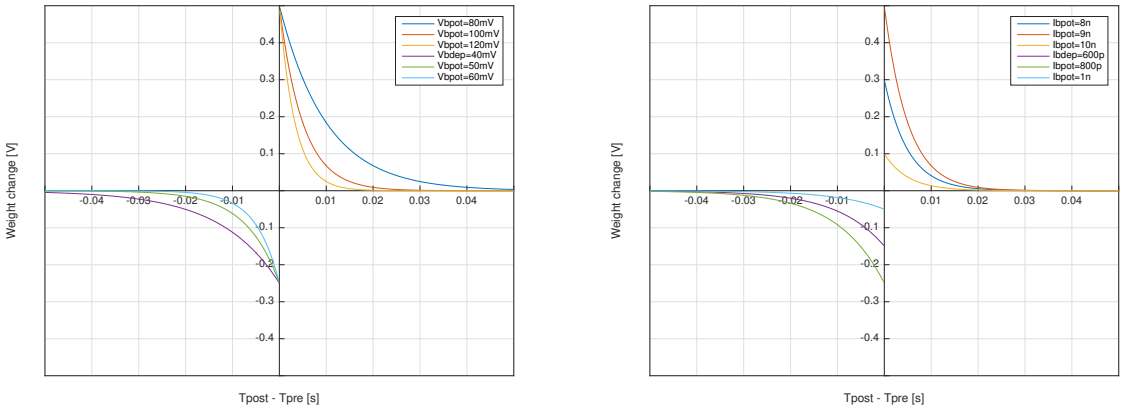


Figure 4.2: Sample synaptic weight evolvement and the membrane voltage distribution of the classic STDP circuit.

4.2 Learning Rule 2: Advanced STDP

Unlike the leaky integrator used in the classic STDP circuit, the one in the advanced STDP circuit has an extra capacitor, increasing the load diversity of charging and discharging phases. The time constants are well controlled via the leaky branch M7 and M15 biased by V_{bpot} and V_{bdep} (see Fig.4.3(a)). The range reaches several tens of miliseconds. The amplitudes of the learning window, on the other hand, are controlled by active current sources I_{bpot} and I_{bdep} (Fig.4.3(b)). For pure temporal coding, an important factor in the stability of any STDP approach is the configuration of maximum facilitation amplitudes A^+ and A^-. These parameters impact the area of the weight update curves during potentiation and depression. It is observed that stable learning is realized when the aggregate area of depression exceeds that of potentiation in the weight update function. On the contrary, weaker depression results in the extreme potentiation of synaptic weights and the eventual shorting of outputs to inputs. This behavior prevents the realization of any practical network transfer function. Compared with the last learning architecture, the advanced STDP circuit has a better controllability over the learning window parameters, and thus is possible to offer the plausible exponential dynamics as well as a more stable performance.



(a) (b)

Figure 4.3: The learning window parameter control in the advanced STDP circuit. (a) Time constants adjusted by V_{bpot} and V_{bdep}; (b) Amplitude adjusted by I_{bpot} and I_{bdep}.

Fig.4.4 demonstrates the functionality of the extra weight dependence block in the advanced STDP circuit. The input signal pairs induce a stable increment of the synaptic weight value. As the weight adjustment level decreases (V_r increases), this increment becomes less effective, which indicates a lack of adaptability of synaptic weight. This explains the bimodal weight distribution in a long period of time reported in [16] for additive weight update rules where the weight dependence is absent. In contrast, the synapse with certain weight dependence, i.e. the multiplicative update rules, shows a unimodal distribution of synaptic weight [16]. An example weight evolvement and the corresponding membrane voltage distribution with Poisson distributed presynaptic and

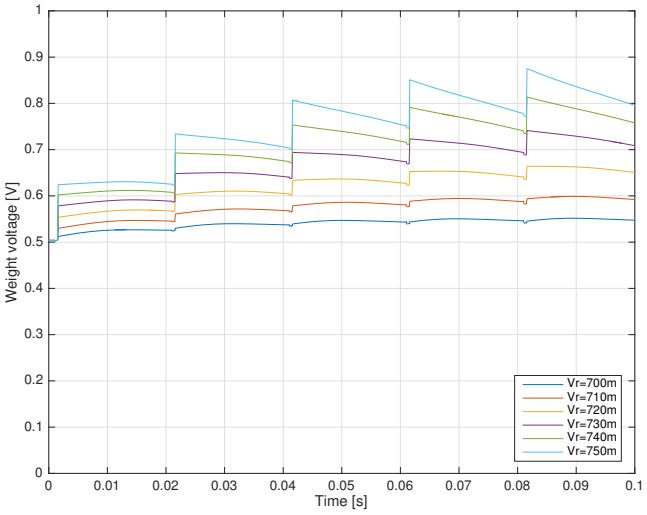


Figure 4.4: Weight dependence property of the advanced STDP circuit. In this example, the presynaptic and postsynaptic signals are of 50 Hz frequency with 1 ms delay. V_r determines the weight dependence level.

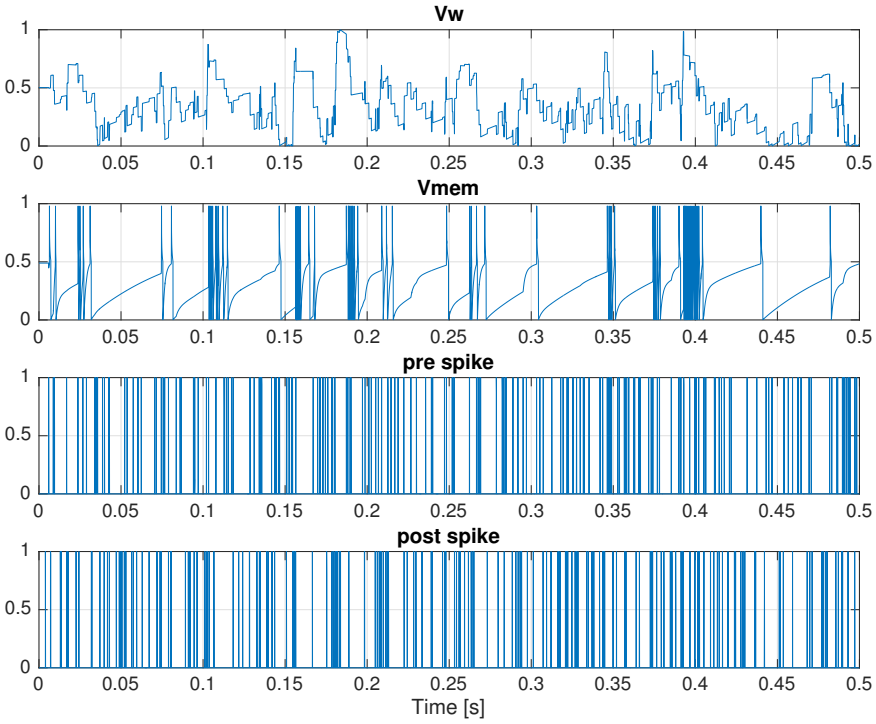


Figure 4.5: Sample synaptic weight evolvement and the membrane voltage distribution in the advanced STDP circuit.

postsynaptic input signals of 200 Hz are displayed in Fig.4.5.

4.3 Learning Rule 3: Triplet-Based STDP

TSTDP synapses, while well suited for pair-based temporal coding, can additionally support triplet and quadratic dynamics, which offers additional nonlinear dynamics to the system, which is reported to be biologically realistic [9]. Fig.4.6 shows the basic pair-based learning window dynamics with exponential fits.

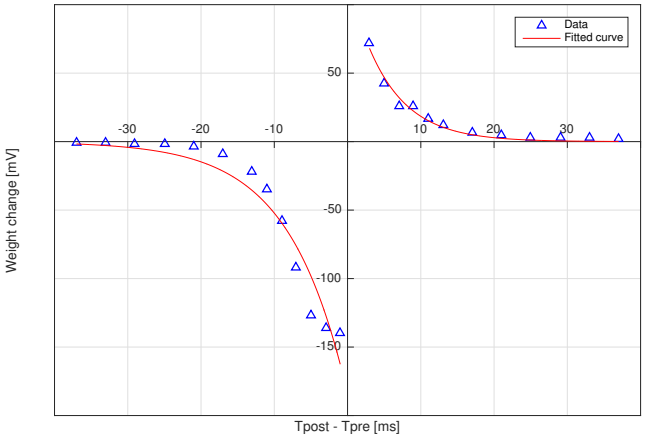
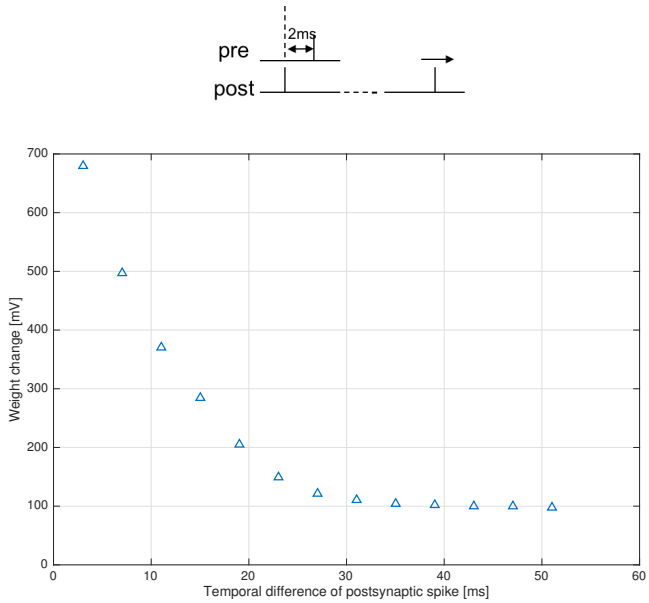


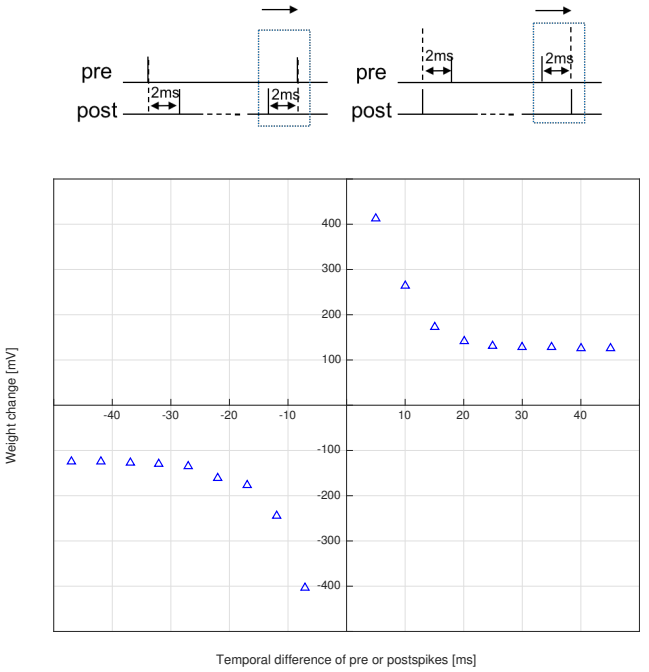
Figure 4.6: Pair-based learning window of TSTDP circuit. Triangular marks are extracted from TSTDP circuit while red curve shows the exponential fit of these data points. The generated curve matches with classic learning window reported in [9].

Fig.4.7(a) reports the weight change induced in the synapse as a function of temporal difference between two post-spikes in a triplet. The influence of spike pairs is negated in this analysis using a fixed 2 ms temporal difference in all experimental runs. As observed, the closer the two post-spikes are, the larger is the effected potentiation, in the synapse. Such third-order spike interactions can be observed for temporal differences under 30 ms. Beyond this, the impact of third order spike interactions wanes, leaving the base potentiation caused by the 2 ms spike pairs, shown as the flat portion of the curve. Similar experiments are applied to a quadrant except that the shifting component is now another pair of pre and postspikes (see Fig.4.7(b)). The forth-order effect lasts until 20-30 ms for both phases. After this, constant weight changes are observed as the flat portion, caused by the fixed temporal difference pre-post pairs.

A parallel comparison of above three learning rule circuits is reported in Table.4.1. Note that the energy features are obtained under 200 Hz input spike trains. The areas are an approximate normalized values (the area of classic STDP is set as 1) for preliminary comparison. The classic STDP circuit does not exhibit basic exponential dynamics though the area and energy features are the best among three. The triplet



(a)



(b)

Figure 4.7: Synaptic weight change due to temporal difference between pre or postspikes in a (a) triplet and (b) quadrant. The inserts above interpret the temporal settings of input spikes. In (a), the x-axis represents the temporal difference between two postspikes while that in (b) are between post and prespike depending on its in depression or potentiation phases. The dashed blocks in (b) inserts indicate the temporal shifting block.

STDP has higher-order dynamics, achieving a better fidelity. The corresponding costs are almost double power and area features as the advanced STDP circuit. In the following experiments, the advanced STDP is chosen for integration due to its best trade-off between various properties. For other certain applications, other models are still possible.

Table 4.1: Comparison of three learning rules

	Classic STDP	Advanced STDP	Triplet STDP
Exponential Dynamics	✗	✓	✓
2nd-Order Dynamics	✓	✓	✓
Higher-Order Dynamics	✗	✗	✓
Weight Dependence	✗	✓	✓
Time Constant Range (ms)	$0 \sim$ several	$0 \sim 100$	$0 \sim 100$
Energy per Spike (pJ)	21.64	35.22	82.07
Normalized Area	1	3	5

4.4 Synaptic Receptors

The synaptic time constant is regulated by transistor M5 biased via V_{tau} to cover various temporal range for AMPA receptors. The possible time constant dynamics are displayed in Fig.4.8(a), ranging from several to tens of miliseconds. Similarly, the time constants for NMDA receptor in both rising and falling phases are adjustable via two seperate bias voltage V_{taur} and V_{tauf}, displayed in Fig.4.8(b)(c). Double exponential dynamics are generated, which gives a better fidelity with biological neuron cells.

The weight dependence of NMDA receptor is demonstrated through a comparison of Vmem and a reference voltage V_{ref}. A sequence of presynaptic spikes are introduced to synapse. The V_{mem} is a step signal from 0 to 500 mV (larger than V_{ref}) onset time at 40 ms (Though in reality, the membrane voltages should be in spike forms, here the setting is made to examine the specific function of NMDA receptor). It can be observed in Fig.4.8(b) that a growing output current start to emerge from the onset of V_{mem}. A linear increase of EPSC amplitudes can be found at each stimuli, as commonly found in AMPA and GABAa receptors. When stimulus are densely distributed, single NMDA EPSC fails to return to resting line before the next stimuli comes due to large decaying time constants, resulting in a summation behavior of previous activities.

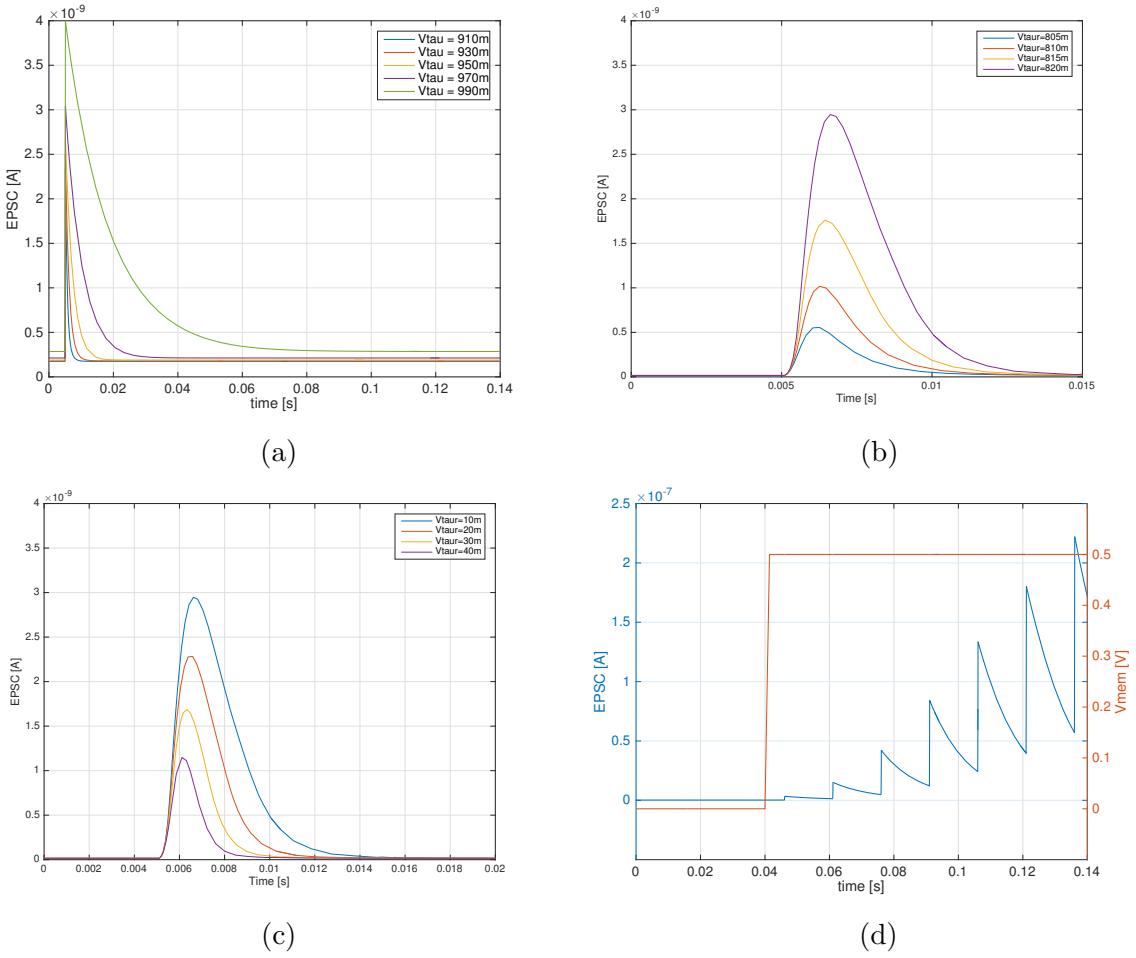


Figure 4.8: Single receptor characterization. (a) τ_{AMPA} control by V_{tau}; (b) Rising phase time constant control by V_{taur} of NMDA receptor; (c) Falling phase time constant control by V_{tauf} of NMDA receptor; (d) Voltage dependence demonstration of NMDA receptor.

4.4.1 Environment Settings

To characterize the joint function of three receptors as well as the synaptic learning functionality, an integration of the receptors, a synaptic learning block and a neuron are implemented to build a unit neural network, called a *cluster*, that really learns in unsupervised pattern. While the focus of this thesis is mainly the synapse, there are still necessities for a specific network for several reasons:

- Neural network is a signal mesh based on electrical spikes. The adaptation ability of the synaptic weight and the function of receptors can be exhibited in the form of spikes in this mesh, which is a clear and straightforward way of representation.
- The incorporation of pre and postsynaptic neural activities is the foundation for synaptic learning (see Sec.2.3.3). It is necessary to obtain both pre and post neuron activities to achieve unsupervised synaptic learning.

- The synchrony detection of synapses is examined between outputs of different neuron layers, which calls for a network configuration.

The top-level architecture of this network is shown in Fig.4.9. The STDP learning circuit incorporates the current spike activities of the presynaptic and the postysyanptic neuron activities to induce the synaptic weight adaptation. This resultant weight is then transmitted to receptor array (except for the GABAa receptor) to achieve distinctive dynamics. As a weight-dependent receptor, the NMDA receptor updates according to the membrane state, which is a feedback signal from the postsynaptic neuron. This joint function of three receptors is achieved through the summation of different ions. In the biophysical neuron, the ions flowing through different receptor channels are gathered together in soma body. The overlapping function of those ions are tested at the hillock, which determines whether a action potential is produced. Therefore, a summation form of individual responses of different receptors is determined.

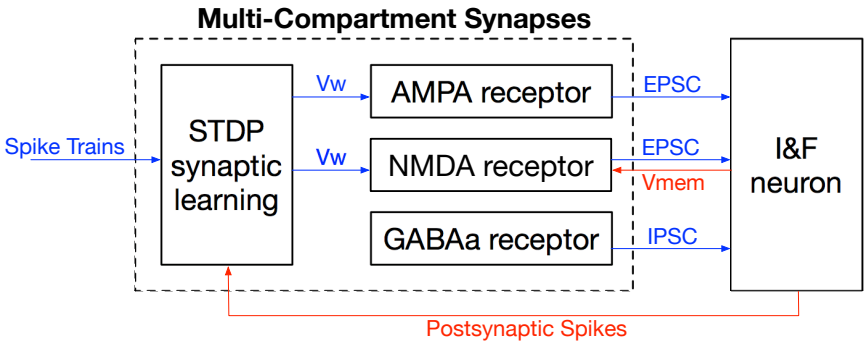


Figure 4.9: Top-level architecture of a cluster network. Multi-compartment synapses consist of three distictive receptors and a learning circuit inducing adaptable learning results to receptors. The blue line represents the forward transmission signals while the red ones are feedbacks.

The complete cluster circuit is illustrated in Fig.4.10. The advanced learning circuit with biologically realistic weight dependence as well as a low power and energy consumption feature is introduced in Sec.3.2, and the receptor architecture is discussed in Sec.3.5. For the neuron architecture, the classic *Integrated and Fire* (I&F) neuron model proposed in [10] is applied (see Fig.4.11). The membrane state V_m is stored in C_m, which is charged by the EPSCs or IPSCs generated by the dendritic synapses. The voltage gain is made with two cascade inverters with a threshold voltage V_{mth}. This voltage gain block and two capacitor C_{fb} and C_m form a positive feedback loop. When $V_{mem} > V_{mth}$, the positive feedback drives abruptly the V_{mem} towards the high rail. Meanwhile, the digital output V_{out} of this neuron is turned to "on" state, which activates the reset branch. When V_{mem} is below V_{mth} after the reset leakage, V_{out} recovers to "off" state, inducing a abrupt decrease of V_{mem}. In this way, the action potentials are produced through charging and discharging process of the membrane voltage.

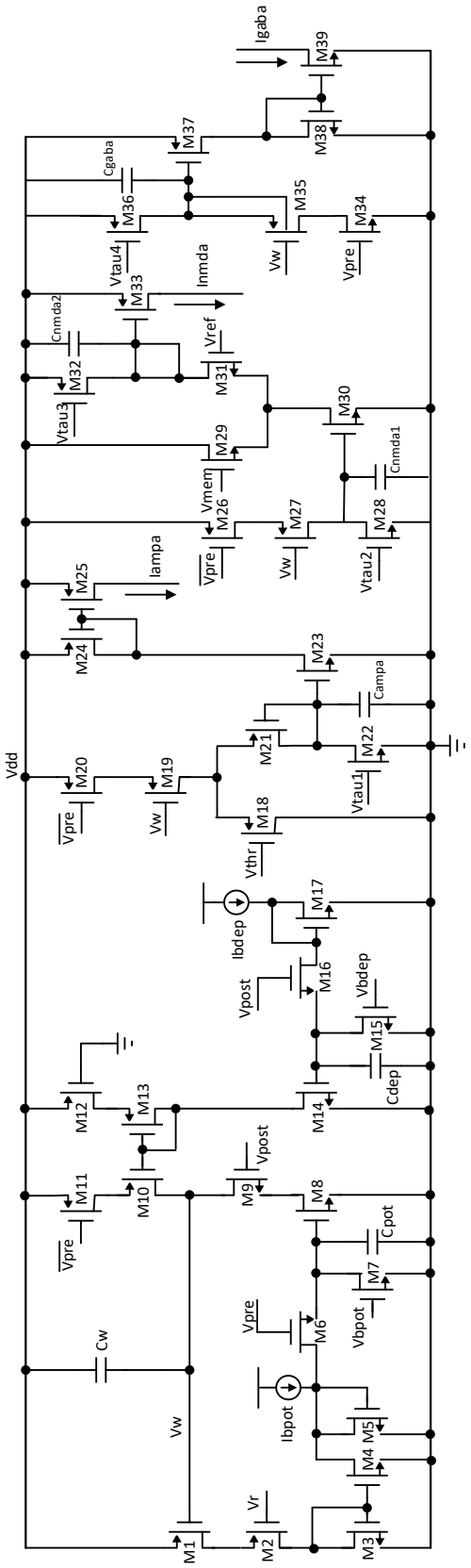


Figure 4.10: The cluster circuit details

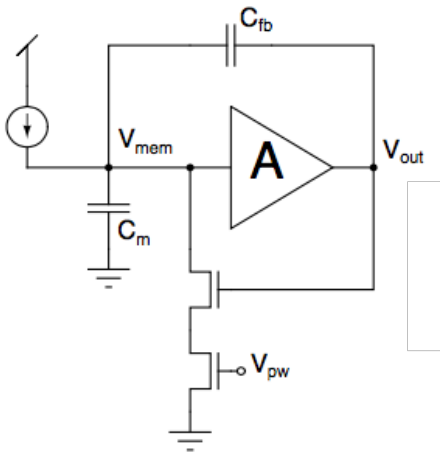


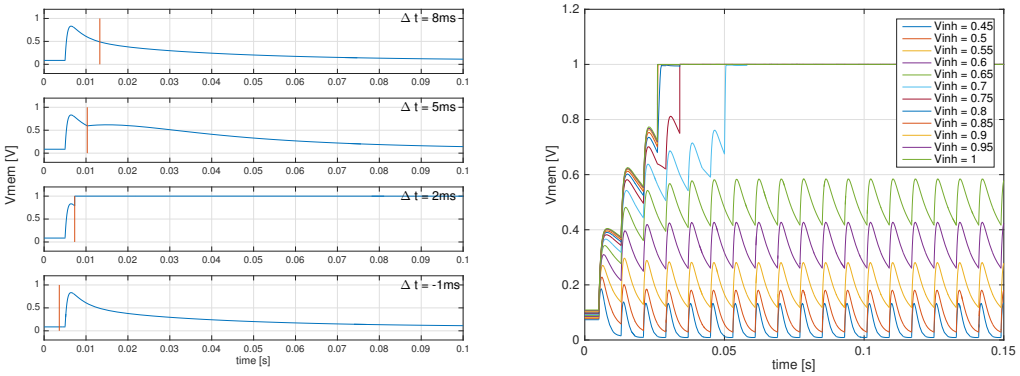
Figure 4.11: Classic I&F neuron circuit [10]

4.4.2 Results

In Fig.4.12(a), AMPA currents are introduced at the onset time of 5 ms, and NMDA currents are injected at different times. Two cases needs to be discussed individually. In the first case where AMPA stimuli precedes NMDA, the excitatory function of NMDA receptors can be demonstrated by the increase of V_{mem} (Δt= 2, 5, 8 ms). As NMDA stimuli approaches AMPA stimuli, larger V_{mem} is detected by NMDA synapse, which gives a greater voltage amplification. However, if delivered in reversed sequence (Δt= -1 ms), no modification is observed. This can be principally explained by the cooperation mechanism of those two receptors, i.e. AMPA receptors usually act as preliminary depolarization of post neurons by inducing small amount of ions (Na^+) into cells. When depolarization threshold is surpassed, NMDA receptors are activated, which allows substantial incursion of ions (both Na^+ and Ca^{2+}) and bigger electrical stimuli are produced. Thus, it is implied that NMDA receptors are not self-initiated. However once activated, the NMDA receptor acts as a major contribution to electrical signal transmission in neuron system.

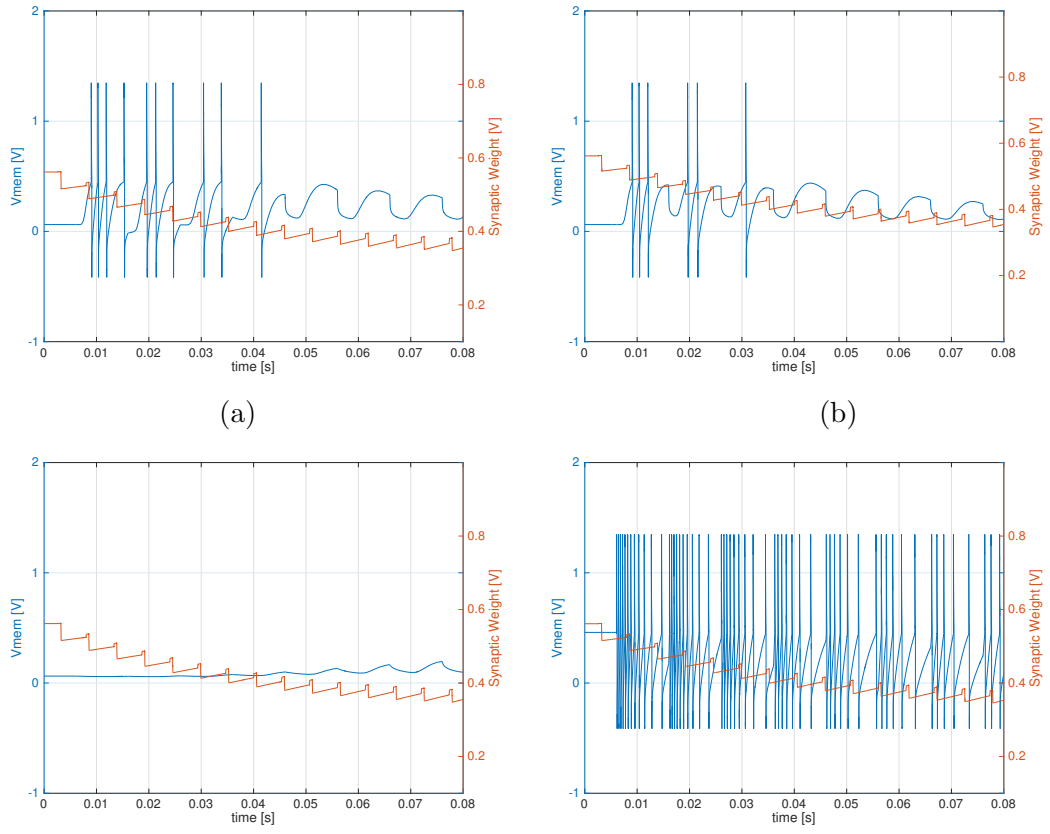
In Fig.4.12(b), the contribution of inhibitory synapses to synapse integration is identified. Various levels of inhibition are applied to the system while the setting of excitatory synapses are maintained. When inhibition behavior is larger than certain level ($Vinh \leqslant 0.65V$), neuron system operates normally. Conversely, if the inhibition level decreases, excitation prevails, driving membrane state to the upper boundary, and consequently information may be lost during this process. This result suggests that inhibitory synapses are of great importance in balancing membrane activities, especially in the case of NMDA receptors where long-term summation of multiple receptors may exist.

The joint functionality of receptors can also be indicated in a spike pattern. Input spikes at a rate of 100Hz with prespikes precede postspikes for 1 ms are introduced to synaptic learning circuit, inducing consecutive depression to synaptic weight. In



(a) (b)

Figure 4.12: The role demonstration of receptors. (a) AMPA and NMDA cooperative function. The AMPA current are induced at a onset time of 5ms while that for NMDA receptors varies (labeled with red vertical lines). Δt represents the interval between NMDA and AMPA activations, ranging from -1 to 8 ms; (b) The balance function of GABAa receptors.



(a) (b)

(c) (d)

Figure 4.13: The role demonstration of receptors in spike patterns. (a) Spike response of three receptor; (b) Spike response without NMDA receptor; (c) Spike response without AMPA receptor; (d) Spike response without GABAa receptor.

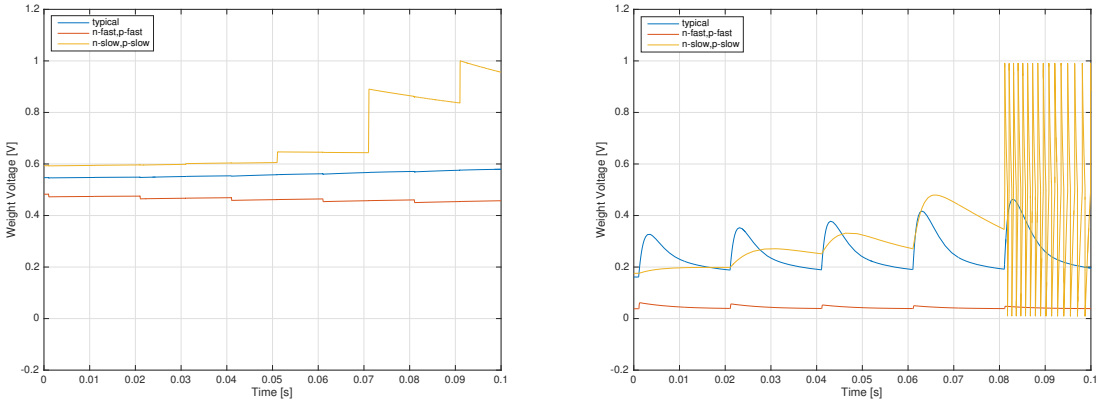
presence of three receptors (Fig.4.13(a)), ten membrane spikes are generated. A gradually sparser distribution of the spikes is observed along with the decline of synaptic weight. When synaptic weight reaches lower bound, the network fails to produce any spike trains. In Fig.4.13(b), the function of NMDA receptor is inhibited. Though some post activities occur, the temporal intervals to generate equal number of spikes as Fig.4.13(a) are larger, and the amount of spike clusters is lower due to a lack of long-term dynamics. NMDA receptor acts as a supplement to synaptic excitation. On the contrary, when the function of AMPA receptor is forbidden, no postspike trains are observed (shown in Fig.4.13(c)). This result is coherent with that of biological experiments observed in hippocampal region [35]. Synapse with only NMDA receptors, also called *silent synapse*, will only transmit information when the postsynaptic neuron is depolarized, caused by synchrony pairing of other synapses with AMPA receptors. Otherwise, a minimal current will be produced by this silent synapse. In last experiment shown in Fig.4.13(d), the GABAa receptor is blocked. A burst of postspikes are produced even though the spike dynamics should decline with a decreasing synaptic weight. The network fails to transmit learning information carried by synapses. Hence, GABAa receptor is essential to create stable signal transmission in SNNs.

The process verification is conducted with regards to corner (fast-fast, slow-slow), supply voltage (0.8-1.2 V) and temperature variation (27-47 $°C$) analysis with corresponding results shown in Fig.4.14. In Fig.4.14(a)(b), the synaptic weight and the corresponding membrane activities under ff and ss corner tests show some drifts from the typical circumstance, which does not interfere with the basic adaptation functionality of synapses since the activity tendency change, instead of the exact values, is of concern. Techniques like the global tuning in DPI blocks can be used to balance the system from drift. The supply voltage and temperature variations show similar effect on the system functionality from Fig.4.14(c)-(f).

In summary, each type of receptor has its distinctive role in neural system to ensure accurate and stable processing of input signals. Despite the phenomenological models we proposed, the circuits are able to reproduce essential synaptic behaviors observed in biological experiments [11] with a compact structure. Due to the sub-threshold region operation and a high threshold voltage implementation, the receptors have low energy consumption of only 1.92, 3.36, and 1.11 pJ/spike respectively (tested under input frequency of 100 Hz), achieving a balance between biological complexity and configuration diversity. The layout of the multi-receptor circuit is displayed in Fig.4.15 with a total dimension of $36.8\mu m \times 40.2\mu m$ (the momcaps take about 95% of the total area).

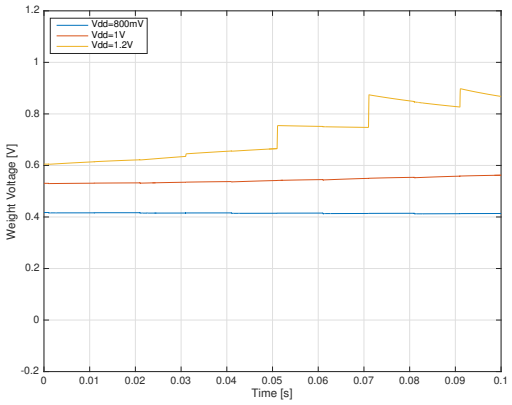
4.5 Conclusion

In this chapter, the characterizations of the synaptic learning architectures discussed in Chapter 3 are present. Several aspects are highlighted to make a comparison: different order of dynamics, the weight dependence, time constant range, energy and area features. The advanced STDP prevails the other two for its widely-inclusive functionality

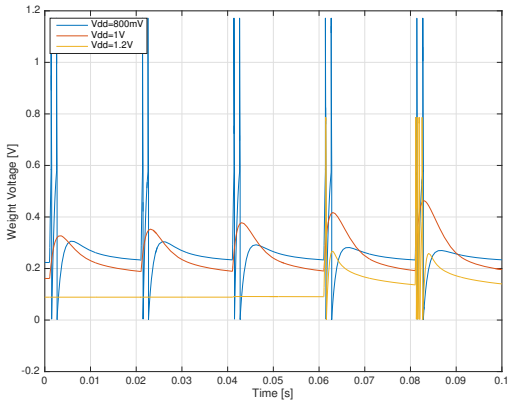


(a)

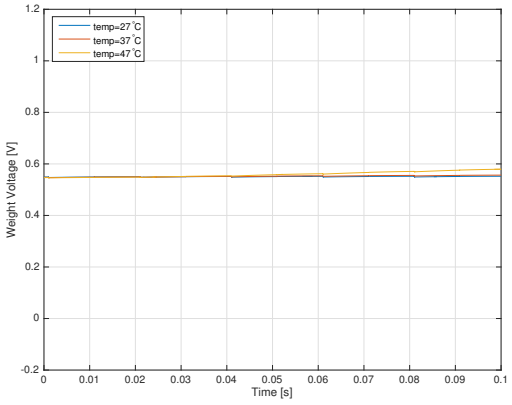
(b)



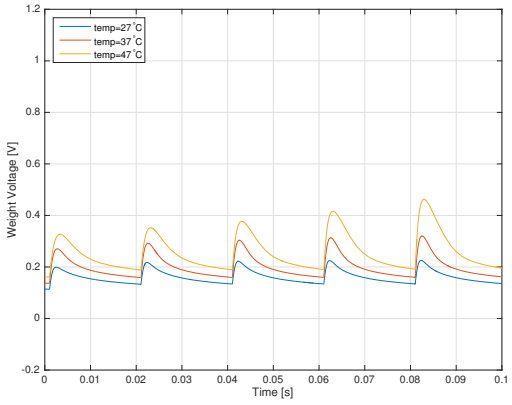
(c)



(d)



(e)



(f)

Figure 4.14: The process verification. (a)(b) corner analysis; (c)(d) supply voltage variation analysis; (e)(f) temperature variation anlysis.

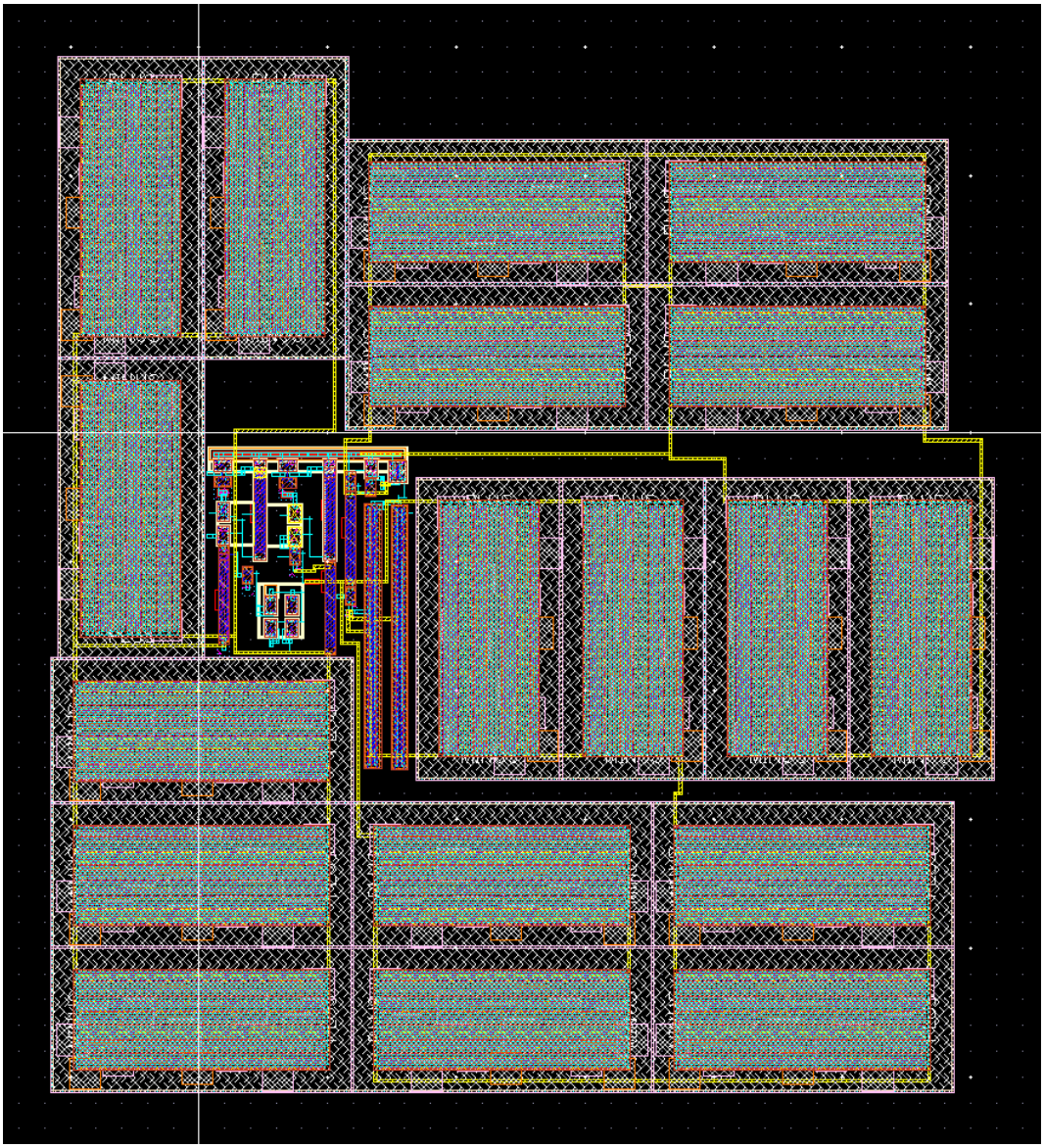


Figure 4.15: The schematic layout of the multi-receptor circuit.

and better trade-offs between functionality and resource consumptions. After that, the distinctive function of multiple receptors are demonstrated both individually and comprehensively. Moreover, the process analysis verifies that the synaptic functionality is seldom interfered after simple global tuning techniques. The cluster structure built will be further employed in next chapter to detect the synchrony features of spike trains obtained through the multi-compartment synapse structure.

Neural Network with Multi-Receptor Synapses

5

5.1 Synchrony Detection Tool: Cross-Correlograms

The method used in these experiments is called $cross - correlograms$. Sample cross-correlograms are displayed in Fig.5.1. The left plot shows no dependence while the right one shows strong correlation. It is a visualization of cross-correlation between two spike trains, i.e. the similarity of two series as a function of the temporal displacement of one relative to the other. In this cross-correlogram, the temporal differences between every single pair of spikes are summed for certain temporal bin. A peak present in the cross-correlogram indicates a correlated relation at this certain temporal bin between target spike groups. For discrete signals, the cross-correlation is defined as [36]:

$$(f \star g)[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m]g[m+n] \tag{5.1}$$

where f^* denotes the complex conjugate of f, and n is the displacement bins, which correspond to the temporal difference between two target spikes.

Figure 5.1: Sample cross-correlograms

The magnitude of the correlation in the cross-correlogram indicates the causality level between sequential neural units. If two spike trains are strongly correlated, a large correlation amplitude displays in the cross-correlogram; it is an efficient way to characterize the computation ability of artificial neural network, and especially the SNN where the weight update information is encoded into the temporal difference between spikes. When valid learning updates happen, the spikes generated are more closely distributed(regardless of the transmission delay), leading to a higher cross-correlation between two spike trains. Therefore, in this experiment, the cross-correlogram is applied to demonstrate the computation ability of the synapses.

5.2 Environment Settings

In this section, a two-layer recurrent network displayed in Fig.5.2 is framed to explore the parallel and hierarchical synchrony detection and amplification function of the multi-compartment synapse. It is presumed that the system will detect the spike-timing synchrony embedded in a noisy environment and amplify this correlation from layers.

The same simulation environment (network configuration, input and noise patterns) is applied to three receptor settings: multi-receptor, AMPA-receptor and NMDA receptor shown in the dashed blocks in Fig.5.2. Each setting consists of three neural clusters, which includes one classic $I\&F$ neuron and four functional synapses with either multiple or single receptor implementations as described in Sec.4.4. A correlated Poisson distributed spike train C1 is introduced to the first two synapses out of four of every cluster while another correlation C2 is added between two C1 which adds an additional correlation between N1 and N2 (also N4 and N5, N7 and N8). The rest of the synapses obtain Poisson distributed spike trains as noise. The Poisson distributed spike trains are generated from Matlab. More correlated spike trains are more likely to coincide in the defined learning window of STDP learning, which will cause more valid weight update events. The cross-correlogram discussed in Sec.5.1 is used to demonstrate the temporal synchrony between two output spike patterns in whether the same or different layers of neurons.

5.3 Input Patterns

Noises have a great impact on the response dynamics of neural system. Among the many sources of noises existing in neurons, the synaptic noise is the main contribution [37]. On one hand, the chemical synapses discussed in this thesis release packets of neurotransmitter at the axon terminals depending on the history firing of both the pre- and postsynaptic neurons probabilistically; on the other hand, the learning induces long-term effect on the postsynaptic neurons, which would transmit the resultant spike trains to every spacial locations they can reach and further form a recurrent network. This non-unidirectional transmission may pass on those spikes to a cell as noises. The summation of the thousands of synaptic inputs of one neuron forms irregular fluctuations on the neural response, ranging from completely random Poisson inputs to periodic inputs [37]. Thus in our experiment, the Poisson distribution is applied to describe the noise distribution. The probability of k events in an interval $P(x)$ is given by the equation:

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \tag{5.2}$$

where λ is the average number of events per interval, k ranges from 0 to the temporal interval concerned. The Poisson distribution with five different λ values are illustrated in Fig.5.3.

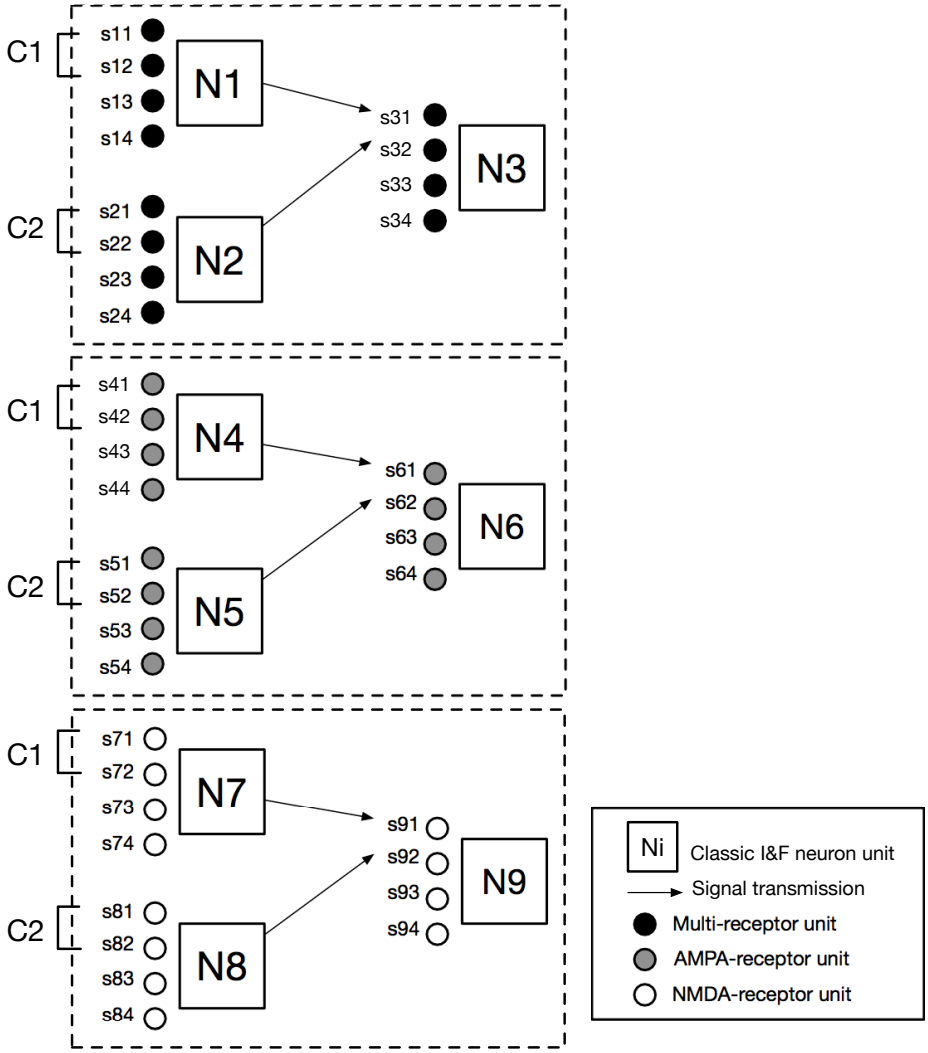


Figure 5.2: Top-Level diagram of the two-layer recurrent testing network. The symbol interpretations are listed in the box on the right. Each neuron Ni are connected with four synapses $si1 - si4$, forming a cluster unit discussed in Sec.4.4. Clusters belonging to different dashed block includes different types of synaptic receptor configurations. Input C1 is a Poisson distributed spike train of 40 Hz. Input C2 is correlated with C1, and this correlation can be in any correlation form. Here a delay of 2 ms is used. The rest of the synapses receive Poisson distributed spike trains of 15Hz.

In previous learning window simulation in Chap.3, the learning width is chosen as 100 ms [9], which corresponds to a frequency of 10 Hz. To induce valid learning process, the average input signal frequency should range from 10 Hz to 1000 Hz considering that the refractory period is 1 ms. The same analysis applies to the noise signals. In our experiments, 40 and 15 Hz Poisson distributed signals are employed as an example to demonstrate the functionality of the system. Other values of signal and noise frequencies are possible.

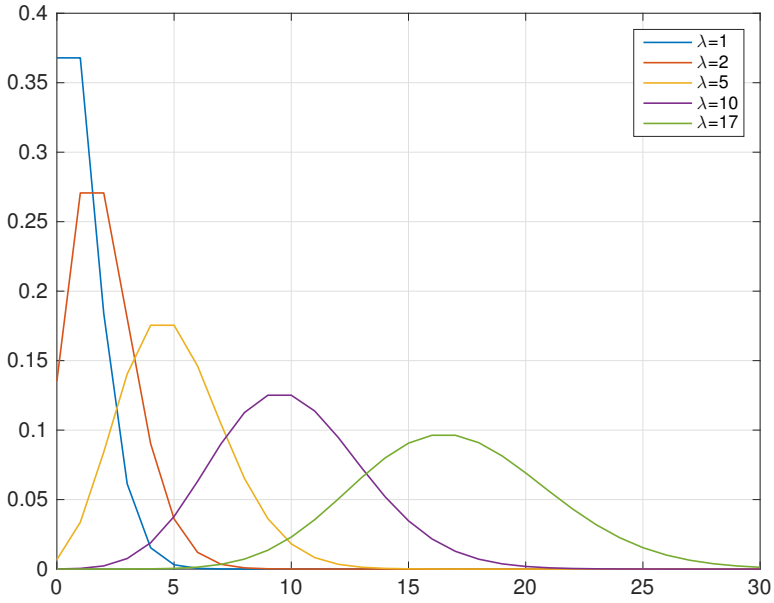


Figure 5.3: The Poisson distribution with five different λ values.

The correlated input spikes induced by C1 and C2 correlation can be in several forms: it can be perfectly simultaneous, or one precedes the other, or one leads the other. For our case, the spike trains within the same cluster is perfectly simultaneous while between the clusters, a delay is added to C2 to form correlation with C1. An example of cross-correlation with different delay values is given in Fig.5.4. Larger delay time decreases the possibility of coincidence between two spike trains within target time range, thus reducing the peak correlation amplitude. Along with that, a temporal shift is generated, and is proportional to the delay introduced to C2. As we want to compare the synchrony detection and amplification of different receptor configurations, the exact magnitudes of the correlation levels are of no interest. Therefore, as long as the magnitude of correlation peak is above the noise horizon, any delay time is possible for simulation. In our experiments, the decay time is set as 2 ms. An example of input signals C1, C2 and two noise signals are displayed in Fig.5.5.

5.4 Synchrony Detection

Fig.5.6 shows the normalized cross-correlogram results from the two-layer recurrent testing network described above. It compares in parallel the synchrony level of three different receptor configurations: multi-receptor, AMPA-receptor and NMDA receptor. In every plot, the temporal range of interest is 0.1 s, which covers the range of the learning window reported in [3]. The correlation levels are normalized for better comparison. As observed in Fig.5.4, the shift of correlation spikes implies the temporal difference

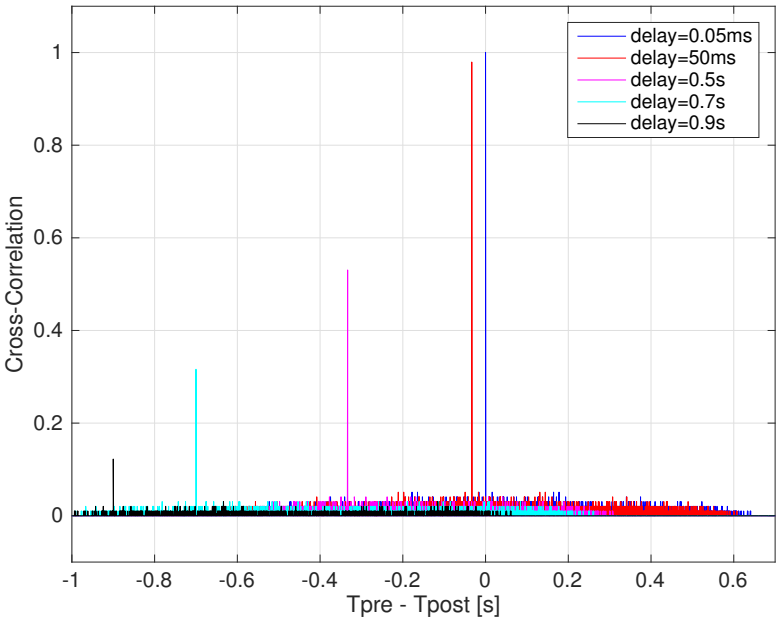


Figure 5.4: The effect of delay of spike trains on cross-correlation.

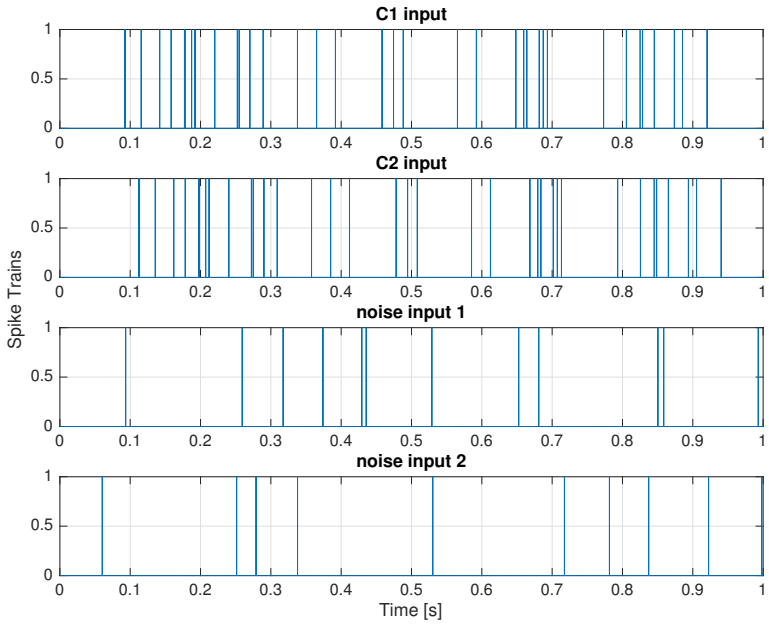
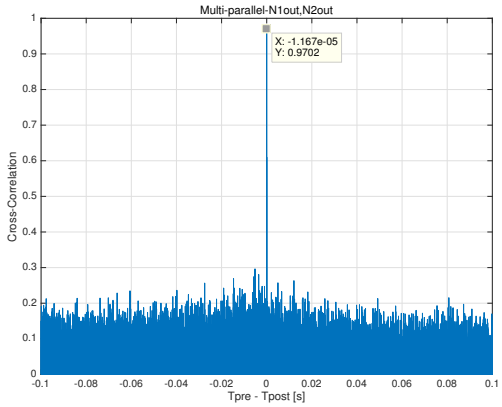
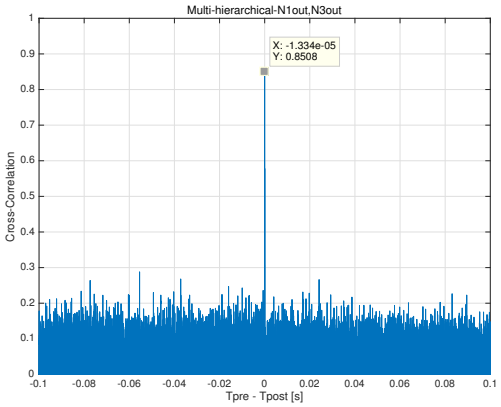


Figure 5.5: Example input spike trains

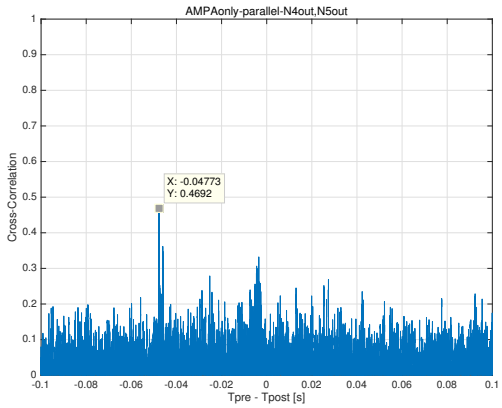
between two spike trains. On the other hand, the amplitude of the correlation peak denotes the correlation level, i.e. the total amount of the coincidence occurring at this target lag point.



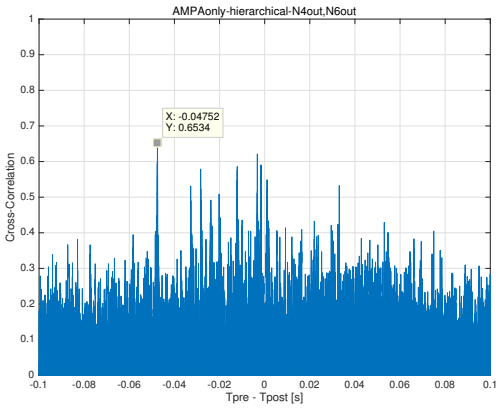
(a) Multi-parallel-N1outN2out



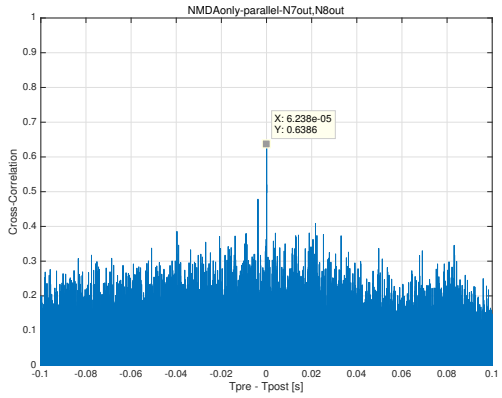
(b) Multi-hierarchical-N1outN3out



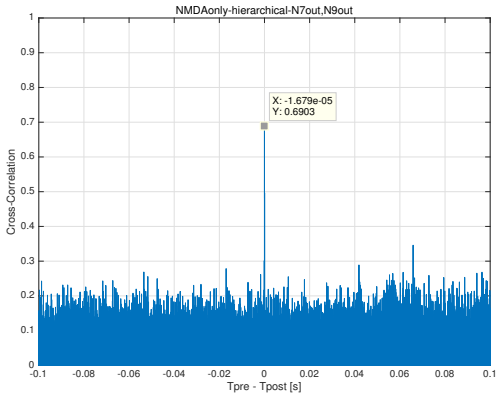
(c) AMPAonly-parallel-N4outN5out



(d) AMPAonly-hierarchical-N4outN6out



(e) NMDAonly-parallel-N7outN8out



(f) NMDAonly-hierarchical-N7outN9out

Figure 5.6: Normalized cross-correlogram results from the two-layer recurrent testing network. (a)(b), (c)(d) and (e)(f) are the parallel and hierarchical cross-correlation plots of multi-receptor, AMPA-receptor and NMDA-receptor configurations respectively.. The annotation above each figure tells "receptor configuration-correlation source type-neuron numbers" For example, "Multi-parallel-N1outN2out" means the parallel correlation of multi-receptor settings between neuron clusters N1 and N2.

The histograms in Fig.5.6(a)(b) evaluate the cross-correlations between parallel clusters N1, N2, and hierarchical clusters N1, N3. A large level of correlation is observed at close to zero time point for both cases. This indicates a strong synchrony between both parallel and hierarchical spike trains after synaptic learning process with multi-receptor settings.

In Fig.5.6(c)(d), the synchrony level is decreased almost by half. In the synapse unit, the advanced STDP learning circuit (see Sec.3.2), which incorporates both the presynaptic and postsynaptic spike activity to induce learning, presents certain level of synchrony detection [5]. However, the large discrepancy between the two correlation level does not originate from the learning circuit but from the multi-receptor synapses, since both of the systems have the same learning implementations. Additionally, the background noise is observed as well as several sub-peaks occurring near the origin in the hierarchical relations, which implies a relatively poorer stability performance. Along with the amplitude decay, a peak shift occurs. The delay between the inputs to the clusters is passed through layers while that of the multi-receptor synapses is mitigated.

Finally, Fig.5.6(e)(f) characterize the synchrony detection function of NMDA-receptor network. Both parallel and hierarchical pairs have similar correlation plots as multi-receptor network but with reduced amplitudes (around 60-70% as that of multi-receptor).

The detection and amplification level of the cross-correlation function of various spike train pairs using both single and multiple receptors is shown in Table.5.1. The maximum amplification level of multi-receptor configuration is almost 2 times as that of single-receptor ones.

Table 5.1: Normalized cross-correlation comparison

	AMPA-Receptor		NMDA-Receptor		Multi-Receptor	
	location (s)	amplitude	location (s)	amplitude	location (s)	amplitude
Parallel	-0.04773	0.4692	-6.238e-5	0.6386	-1.167e-5	0.9702
Hierarchical	-0.04752	0.6534	-1.679e-5	0.6903	-1.333e-5	0.8508

In summary, the analog multi-compartment synapse structure is able to detect and amplify the temporal synchrony embedded in the synaptic noise. The maximum amplification level is 2 times larger than that of single-receptor configurations. Moreover, the circuit shows efficient learning ability; the consecutive neural clusters generate almost synchronized output spike patterns in the presence of delay in inputs signals, i.e. it takes shorter time for system with multiple-receptor to achieve synchrony. Analysis indicate that this ability originates from the NMDA-receptor as NMDA-receptor displays similar correlations, except for a decrement in the amplitude of correlation level. Clearly, AMPA and NMDA receptors have a collaborate relation in inducing efficient synchrony detection for synapse structures.

5.5 Conclusion

This chapter characterizes the synchrony detection function of the multi-compartment synapse structure. First of all, the method cross-correlogram is described as a way to visualize the correlation relation between spike trains. After that, experimental environment set-up is explained. Input patterns to the system is then introduced. Finally, the synchrony function is demonstrated via the comparison experiments between various receptor configurations in the recurrent network consisting of clusters of neural and synapse units.

6 Conclusion and Future Work

6.1 Conclusion

In this thesis, a novel synapse structure incorporating an advanced STDP learning algorithm and multi-compartment receptor elements has been proposed and built in UMC65nm technology. The intrinsic analogous dynamics between transistors and neurons offer great fidelity possibilities and power features.

Three synaptic learning circuits are proposed: the classic STDP, the advanced STDP and the triplet STDP circuits. The classic STDP circuit does not exhibit basic exponential and real-time dynamics through the area and power features are the best among three. The triplet STDP is able to capture diverse experimental observations obtained from real neural system. The corresponding area and power consumptions are almost doubled compared with the advanced STDP circuit. The advanced STDP circuit achieves a great trade-off between biological fidelity and resource consumptions. Wide temporal range of synaptic learning windows up to 100 ms is possible. Additionally, a weight dependence feature expands its biological properties. The energy consumed per spike event is merely 35.22 pJ.

The multi-compartment synapse design gives biologically accurate modeling of chemical synapse, which increases the computation ability of synapses. Each type of receptor forming the synapse structure has its distinctive role in neural system to ensure accurate and stable processing of input signals. More importantly, the cross-correlated input patterns can be well detected and amplified (in maximum 2 times larger than single-receptor configuration) through layers of this multi-compartment synapse structure. Better synchrony indicates a higher efficiency in signal processing and better computation ability in recurrent neural systems. Moreover, the receptors have extremely low energy consumption of only 1.92, 3.36, and 1.11 pJ/spike respectively (tested under input frequency of 100 Hz), representing a good balance between biological complexity and configuration diversity.

In summary, this thesis provides a new vertical insights into the detailed structure of synapses from the biological fidelity aspects. The synchrony detection and amplification functionality has been demonstrated through the cross-correlation verifications with power-efficient configurations of the learning algorithm and receptors.

6.2 Future Work

This project orientation, the multi-compartment synapse configuration, is relatively new in hardware neuromorphic design domain. That also means wide ranges of research direction are possible. During this project, many interesting and potential topics are found for future work.

Whether for the synaptic learning or the receptor structures, the membrane voltage is incorporated to achieve conduction and adaptation. The membrane voltage is in the form of spikes, which displays rapid dynamics. It is expected to have a relatively slow-changing parameter to mediate synaptic updates for stability considerations, like the calcium concentration, which exhibits integrated effect of current activities of target neuron.

From biological aspect, the signal transmission through synapse consists of neurotransmitter release activated by presynaptic spike only, and the ligand-binding between transmitters and the corresponding receptors. While the multi-receptor concept proposed in this thesis matches with biological phenomenon, a learning algorithm that mimics the transmitter release process can be further investigated.

Bibliography

[1] OpenStax College, "How neurons communicate," 2013. [Online]. Available: https://cnx.org/contents/cs_Pb-GW@5/How-Neurons-Communicate

[2] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

[3] G.-q. Bi and M.-m. Poo, "Synaptic modification by correlated activity: Hebb's postulate revisited," *Annual review of neuroscience*, vol. 24, no. 1, pp. 139–166, 2001.

[4] G. Indiveri, E. Chicca, and R. Douglas, "A vlsi array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity," *IEEE transactions on neural networks*, vol. 17, no. 1, pp. 211–221, 2006.

[5] A. Bofill-i Petit and A. F. Murray, "Synchrony detection and amplification by silicon neurons with stdp synapses," *IEEE Transactions on Neural Networks*, vol. 15, no. 5, pp. 1296–1304, 2004.

[6] M. R. Azghadi, S. Al-Sarawi, D. Abbott, and N. Iannella, "A neuromorphic vlsi design for spike timing and rate based synaptic plasticity," *Neural Networks*, vol. 45, pp. 70–82, 2013.

[7] N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumislawska, and G. Indiveri, "A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses," *Frontiers in neuroscience*, vol. 9, 2015.

[8] C. Bartolozzi and G. Indiveri, "Synaptic dynamics in analog vlsi," *Neural computation*, vol. 19, no. 10, pp. 2581–2603, 2007.

[9] G.-q. Bi and M.-m. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type," *Journal of neuroscience*, vol. 18, no. 24, pp. 10 464–10 472, 1998.

[10] C. Mead and M. Ismail, *Analog VLSI implementation of neural systems*. Springer Science & Business Media, 2012, vol. 80.

[11] A. Destexhe, Z. F. Mainen, and T. J. Sejnowski, "Kinetic models of synaptic transmission," *Methods in neuronal modeling*, vol. 2, pp. 1–25, 1998.

[12] S. H. Wu, C. L. Ma, and J. B. Kelly, "Contribution of ampa, nmda, and gabaa receptors to temporal pattern of postsynaptic responses in the inferior colliculus of the rat," *Journal of Neuroscience*, vol. 24, no. 19, pp. 4625–4634, 2004. [Online]. Available: http://www.jneurosci.org/content/24/19/4625

[13] C. Mead, "Neuromorphic electronic systems," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–1636, 1990.

[14] B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen, "Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 699–716, 2014.

[15] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura *et al.*, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668–673, 2014.

[16] A. son, M. Diesmann, and W. Gerstner, "Phenomenological models of synaptic plasticity based on spike timing," *Biological cybernetics*, vol. 98, no. 6, pp. 459–478, 2008.

[17] P. Kogge, "The tops in flops," *IEEE Spectrum*, vol. 48, no. 2, 2011.

[18] E. Stromatias, F. Galluppi, C. Patterson, and S. Furber, "Power analysis of large-scale, real-time neural networks on spinnaker," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*. IEEE, 2013, pp. 1–8.

[19] M. R. Azghadi, N. Iannella, S. F. Al-Sarawi, G. Indiveri, and D. Abbott, "Spike-based synaptic plasticity in silicon: Design, implementation, application, and challenges," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 717–737, May 2014.

[20] J. Wu, *Introduction to neural dynamics and signal transmission delay*. Walter de Gruyter, 2001, vol. 6.

[21] G. L. Shaw, *Donald Hebb: The Organization of Behavior*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1986, pp. 231–233. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-70911-1_15

[22] K. Gerrow and A. Triller, "Synaptic stability and plasticity in a floating world," *Current Opinion in Neurobiology*, vol. 20, no. 5, pp. 631 – 639, 2010, neuronal and glial cell biology New technologies. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0959438810001078

[23] D. Y. Kim, S. H. Kim, H. B. Choi, C.-k. Min, and B. J. Gwag, "High abundance of glur1 mrna and reduced q/r editing of glur2 mrna in individual nadph-diaphorase neurons," *Molecular and Cellular Neuroscience*, vol. 17, no. 6, pp. 1025–1033, 2001.

[24] S. R. Platt, "The role of glutamate in central nervous system health and disease–a review," *The Veterinary Journal*, vol. 173, no. 2, pp. 278–286, 2007.

[25] A. M. VanDongen, *Biology of the NMDA Receptor*. CRC Press, 2008.

[26] S. H. Wu, C. L. Ma, and J. B. Kelly, "Contribution of ampa, nmda, and gabaa receptors to temporal pattern of postsynaptic responses in the inferior colliculus of the rat," *Journal of Neuroscience*, vol. 24, no. 19, pp. 4625–4634, 2004.

[27] E. De Schutter and I. ebrary, *Computational Modeling Methods for Neuroscientists*. MIT Press, 2009, vol. Computational neuroscience series.

[28] S. Song, K. D. Miller, and L. F. Abbott, "Competitive hebbian learning through spike-timing-dependent synaptic plasticity," *Nat Neurosci*, vol. 3, no. 9, pp. 919–926, 09 2000.

[29] J. Rubin, D. D. Lee, and H. Sompolinsky, "Equilibrium properties of temporally asymmetric hebbian plasticity," *Phys. Rev. Lett.*, vol. 86, pp. 364–367, Jan 2001. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.86.364

[30] R. Gütig, R. Aharonov, S. Rotter, and H. Sompolinsky, "Learning input correlations through nonlinear temporally asymmetric hebbian plasticity," *Journal of Neuroscience*, vol. 23, no. 9, pp. 3697–3714, 2003. [Online]. Available: http://www.jneurosci.org/content/23/9/3697

[31] J.-P. Pfister and W. Gerstner, "Triplets of spikes in a model of spike timing-dependent plasticity," *Journal of Neuroscience*, vol. 26, no. 38, pp. 9673–9682, 2006.

[32] S. Moradi and G. Indiveri, "An event-based neural network architecture with an asynchronous programmable synaptic memory," *IEEE transactions on biomedical circuits and systems*, vol. 8, no. 1, pp. 98–107, 2014.

[33] P. Hasler, "Continuous-time feedback in floating-gate mos circuits," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 48, no. 1, pp. 56–64, 2001.

[34] J. M. Brader, W. Senn, and S. Fusi, "Learning real-world stimuli in a neural network with spike-driven synaptic dynamics," *Neural computation*, vol. 19, no. 11, pp. 2881–2912, 2007.

[35] D. Liao, N. A. Hessler, and R. Malinow, "Activation of postsynaptically silent synapses during pairing-induced ltp in ca1 region of hippocampal slice," *Nature*, vol. 375, no. 6530, p. 400, 1995.

[36] D. J. Amit and S. Fusi, "Learning in neural networks with material synapses," *Neural Computation*, vol. 6, no. 5, pp. 957–982, 1994.

[37] N. Brunel, F. S. Chance, N. Fourcaud, and L. Abbott, "Effects of synaptic noise and filtering on the frequency response of spiking neurons," *Physical Review Letters*, vol. 86, no. 10, p. 2186, 2001.