

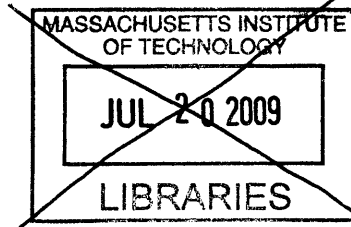
# Analog VLSI Circuit Design of Spike-Timing-Dependent Synaptic Plasticity

by

Joshua Jen C. Monzon

S.B. in Electrical Engineering and Computer Science, MIT, 2007

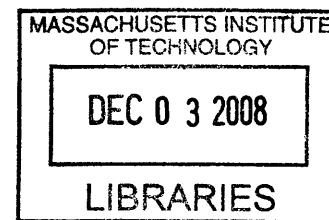
Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Electrical Engineering and Computer Science at the Massachusetts Institute of Technology.



[Done]  
May 23, 2008

**ARCHIVES**

Copyright 2008 Joshua Jen C. Monzon. All rights reserved.



The author hereby grants to M.I.T. permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole and in part in any medium now known or hereafter created.

Author \_\_\_\_\_  
Department of Electrical Engineering and Computer Science  
May 23, 2008

Certified by \_\_\_\_\_  
Principal Research Scientist, Department of Health Sciences Technology  
Thesis Supervisor

Accepted by \_\_\_\_\_  
Arthur C. Smith  
Professor of Electrical Engineering  
Chairman, Department Committee on Graduate Theses

10/13/2014

## Abstract

Synaptic plasticity is the ability of a synaptic connection to change in strength and is believed to be the basis for learning and memory. Currently, two types of synaptic plasticity exist. First is the spike-timing-dependent-plasticity (STDP), a timing-based protocol that suggests that the efficacy of synaptic connections is modulated by the relative timing between presynaptic and postsynaptic stimuli. The second type is the Bienenstock-Cooper-Munro (BCM) learning rule, a classical rate-based protocol which states that the rate of presynaptic stimulation modulates the synaptic strength. Several theoretical models were developed to explain the two forms of plasticity but none of these models came close in identifying the biophysical mechanism of plasticity. Other studies focused instead on developing neuromorphic systems of synaptic plasticity. These systems used simple curve fitting methods that were able to reproduce some types of STDP but still failed to shed light on the biophysical basis of STDP. Furthermore, none of these neuromorphic systems were able to reproduce the various forms of STDP and relate them to the BCM rule. However, a recent discovery resulted in a new unified model that explains the general biophysical process governing synaptic plasticity using fundamental ideas regarding the biochemical reactions and kinetics within the synapse. This brilliant model considers all types of STDP and relates them to the BCM rule, giving us a fresh new approach to construct a unique system that overcomes all the challenges that existing neuromorphic systems faced. Here, we propose a novel analog very-large-scale-integration (aVLSI) circuit that successfully and accurately captures the whole picture of synaptic plasticity based from the results of this latest unified model. Our circuit was tested for all types of STDP and for each of these tests, our design was able to reproduce the results predicted by the new-found model. Two inputs are required by the system, a glutamate signal that carries information about the presynaptic stimuli and a dendritic action potential signal that contains information about the postsynaptic stimuli. These two inputs give rise to changes in the excitatory postsynaptic current which represents the modifiable synaptic efficacy output. Finally, we also present several techniques and alternative circuit designs that will further improve the performance of our neuromorphic system.

## Table of contents

Abstract .....	3
Table of contents.....	4
List of figures and tables.....	5
Acknowledgements.....	7
1. Introduction .....	9
2. Unified model of spike-timing-dependent synaptic plasticity .....	11
2.1. The neuron .....	11
2.2. The Synapse .....	12
2.2.1 Signaling mechanism through the synapse.....	12
2.2.2. Synaptic strength.....	13
2.2.3. Integration of synaptic inputs.....	14
2.3. Synaptic plasticity.....	14
2.4. Previous models of synaptic plasticity.....	15
2.5. Unified model of synaptic plasticity.....	16
2.5.1. Backpropagation of action potentials.....	17
2.5.2. NMDA receptor activation.....	17
2.5.2. Calcium transduction pathway.....	19
3. Fundamentals of low power neuromorphic circuit design.....	21
3.1. Neuromorphic circuit design overview.....	21
3.2. Metal oxide semiconductor field effect transistor (MOSFET) .....	23
3.2.1. MOSFET fundamentals.....	23
3.2.1. MOSFET in subthreshold regime .....	24
3.3. Basic subthreshold circuit blocks.....	25
3.3.1. Current mirrors.....	25
3.3.2. Current multipliers and dividers.....	26
3.3.3. Transconductance amplifiers.....	29
3.3.4. Differential pair.....	32
3.4. Summary .....	33
4. Development of a low power aVLSI synapse .....	35
4.1. Inputs .....	35
4.2. Outputs.....	37
4.3. Synaptic plasticity system design.....	37
4.3.1. Glutamate receptor dynamics.....	37
4.3.2. NMDA channel dynamics .....	39
4.3.3. Calcium integrator dynamics .....	40
4.3.4. Mechanistic design of the calcium to EPSC transduction circuit.....	41
4.3.5. Empirical design of the calcium to EPSC transduction circuit .....	43
4.4. Test and results.....	46
4.5. Relating rate-based to spike timing dependent plasticity .....	53
4.6. Discussion .....	56
5. Conclusion and future work .....	59
6. References.....	61

## List of figures and tables

Figure 2-1. A cartoon of a neuron showing its basic parts. ....	12
Figure 2-2. A cartoon of a synapse showing its basic parts. ....	13
Figure 2-3. Hebbian STDP .....	15
Figure 2-4. Block diagram of the unified STDP model. ....	16
Figure 2-5. Profiles of action potential and glutamate concentration. ....	19
Figure 3-1. The synapse and the physical gap between two neurons. ....	21
Figure 3-2. Schematics of the NMOS and PMOS. ....	24
Figure 3-3. Schematic of a current mirror. ....	26
Figure 3-4. Schematic of a low power one-quadrant multiplier. ....	27
Figure 3-5. Schematic of a low power four-quadrant multiplier. ....	28
Figure 3-6. Schematic of an ordinary transconductance amplifier. ....	29
Figure 3-7. Schematic of a wide linear range transconductance amplifier. ....	31
Figure 3-8. I-V relationship of the ordinary and wide linear range transconductance amplifier. ....	32
Figure 3-9. Schematic of a wide linear range differential pair. ....	33
Figure 4-1. The inputs of the synapse, glutamate and dendritic action potential. ....	35
Figure 4-2. Schematic of a basic spiking circuit. ....	36
Figure 4-3. The output spike generated by the spiking circuit. ....	36
Figure 4-4. Block diagram for the aVLSI circuit design of a synapse. ....	37
Figure 4-5. Schematic of the glutamate receptor circuit. ....	38
Figure 4-6 Schematic of the full NMDA channel dynamics circuit. ....	40
Figure 4-7. Schematic of the calcium integrator circuit. ....	41
Figure 4-8. Proposed schematic for the calcium to EPSC transducer. ....	43
Figure 4-9. The BCM rule curve. ....	44
Figure 4-10. Sigmoidal decomposition of the BCM curve. Resulting BCM curve from the sigmoidal decomposition. ....	45
Figure 4-11. Schematic of the alternative design for the EPSC transduction circuit. ....	45
Figure 4-12. BCM curve outputs from manipulations of the empirical EPSC transduction circuit. ....	47
Figure 4-13. Asymmetric STDP following Hebbian rule. ....	48
Figure 4-14. Asymmetric STDP following Anti-hebbian rule. ....	49
Figure 4-15. Symmetric STDP showing mixed STD and STP. ....	50
Figure 4-16. Symmetric STDP showing STP. ....	51
Figure 4-17. Symmetric STDP showing STD. ....	52
Figure 4-18. Rate-based plasticity curve. ....	53
Figure 4-19. Feedback diagram between the AMPA receptor output to the NMDA receptor. The blue lines correspond to the current signal path of our circuit while the red lines correspond to the proposed feedback path. ....	54
Figure 4-20. Calcium levels within a synapse increases with the presynaptic stimulation rate. ....	54
Figure 4-21. Modified block diagram showing proposed feedback path between the EPSC transduction stage and NMDA channel circuit. ....	55
Table 4-1. Input constants for the EPSC transductor tests. ....	46
Table 4-2. Power and area estimates of the STDP system. ....	47



## Acknowledgements

This project is a challenging undertaking for a lone student to hurdle by himself. For the past year and a half, numerous people had sacrificed their time to aid me in one way or another to complete this project. You know who you are, and with these few sentences, I wish to express my extreme gratitude to these people.

First of all, I would like to acknowledge my family, especially my parents and my brother who have supported me in my studies over the past four years. Their support, love and encouragement kept me hanging on even during the direst of times.

At this point, I wish to acknowledge Dr. Chi-sang Poon, a very brilliant mind who helped me formulate the project and educated me totally in the field of neuroscience. I wish to thank him for allowing me to work in his lab and do stuff that I really enjoyed. His guidance made me an improved researcher and helped me approach the challenges that I faced. He gave me the opportunity to write research papers which improved my communication and salesmanship skill. Lastly, he also taught me real life lessons.

Furthermore, I would like to acknowledge Dr. Kuan Zhou's mentorship and the expertise that he shared regarding circuit design while working on the project. His ideas were very useful and helped me solve some problems in my design. Even though he teaches at University of New Hampshire, he spends hours with me in the lab at MIT to share his insights regarding the project.

The next person I wish to thank is soon to be Dr. Chuang-Chung Lee who I collaborated closely with in developing the STDP circuit. He is the brilliant mind who proposed and discovered the unified model that explains the whole STDP story. This thesis is strongly based of his work and I am glad that his work can be implemented in silicon. Congratulations for recently tying the knot, as well as finishing your doctorate degree.

I would also like to acknowledge the immense help given to me by Dr. Guy Rachmuth, the developer of most of the chips that I used to complete this project. In the infancy stages of my research, he sacrificed a lot of his time coming to MIT during nights and weekends to pass on his research to me. His clear and concise explanations made me understand his thesis which 18 months ago, I had no clue what is all about. Now, I realized how powerful his research is and I am happy to continue and extend what he was doing when he was a graduate student at MIT. I hope you enjoy your married life as well.

Of course, I would also like to acknowledge the help of my friend, Steve Zhou, former graduate student at MIT and now a chip designer at Linear Technology. He spent some of his free time to help me design some of the circuits in the beginning of the project. His expertise in circuit design certainly shed some light on some of the confusion I had in the course of the project.

I would also like to thank all my other friends who kept supporting me in the course of my stay at MIT. They were my distractions, my conduits to the world outside of MIT. Without them, I might be a wreck right now.

To the rest of the Poon Lab, Paula, Yunguo, Chung, Zhen, Gang, and Hui. Thanks a lot for your friendship and for keeping me company as I do research. I enjoyed the lunches and dinners we all had together.

Lastly, I wish to thank God for being around when I need Him the most, especially in times of tribulations. Thank You so much for allowing me to complete my education.



## 1. Introduction

Artificial intelligence is a dream that we can only hope to achieve in the future. On average, scientists believe that we will be able to create an independent learning machine in the next few decades. For years, scientists have studied, developed, and tested multiple types of intelligent systems in the hopes of achieving the futuristic view of having artificial intelligence dominate the planet, but so far, although some sort of progress have been achieved, nobody was able to build a fully functional superior, intelligent and independent learning machine. The biggest reason why this is the case is because no one has developed an effective way to implement memory and learning in an artificial being.

If we think about it, why are we working so hard to create an entirely new system that mimics the behavior of the human mind? Why not just emulate how the mind works and behaves? Scientists around the world are beginning to realize this fact and slowly but surely, we see researchers faced by a seemingly impossible task find a similar and parallel challenge in the natural world that is solved by evolution. For instance, to create sleeker and better underwater vehicle designs, scientists studied how fishes maneuver [1]. Thus, instead of figuring out the most optimal aerodynamic design from scratch, scientists based their ideas using a proven design from mother nature.

In developing an efficient and effective learning machine with memory, we have the alternative approach of not starting from scratch. Instead, we can focus on understanding how the brain, a natural and perhaps one of the most complicated machines ever, actually works. To do so, we have to analyze basic structures of the brain because the brain's behavior is dependent on how these structures behave. Realistically, the most basic functional unit that we can easily replicate is the neuron. Once we successfully capture the full behavior of neurons and subsequently, networks of neurons in silicon, we can implement subsystems of the brain in an electronic chip. The caveat though is that these neurons should be so accurately modeled otherwise our system design might not work.

Neurons are simple minute structures. Simply put, they are transmission lines through which signals in the central nervous system flow. A much more important and interesting structure to study is the connection between any two neurons, or the synapse. The synapse is a very complex structure because it encapsulates the dynamics that describes the relationship between any two neurons. This is why scientists believe that learning and memory are encoded in the combination of synapses in the nervous system. The average person probably does not realize this but each of the quadrillion synapses in the human body has a property called synaptic strength and modifying this property leads to adaptation, habituation, and eventually learning and memory.

Our goal for this project is to emulate how the brain modulates this synaptic strength and implement an electronic synapse that behaves like its natural counterpart. We shall present to you a very robust model of synaptic strength modification, more commonly known as synaptic plasticity and we will propose a novel, compact, and low power analog circuit designed to fit in a chip that simulates the functions of a real synapse.



## 2. Unified model of spike-timing-dependent synaptic plasticity

Information storage in the brain depends on modifications in neuronal networks, is brought upon by inputs from the environment. More specifically, the connection between two neurons, which is called a synapse, can vary in strength depending on the stimuli acting upon the neurons. These stimuli originate from sensory inputs and go through the signal processing pathways of our brain and ultimately modulate the connection strength of synapses. This phenomenon of synaptic strength modification is called *synaptic plasticity*.

In the last decade, it was discovered that the strength of synaptic connection is influenced by the relative timing between presynaptic and postsynaptic spikes [2]. Hence, the term spike-timing-dependent synaptic plasticity or STDP was coined. This discovery is very crucial because it supports and expands the Hebbian theory, the most widely regarded neurological learning rule. It is such an important theory that the principal experimental protocol for inducing change in synaptic plasticity was based from it. Since the discovery of STDP, multiple experiments have been done to prove the Hebbian theory, however, the theoretical basis for STDP is not a popular study for research because very few biophysical models can account for the temporal asymmetry properties of STDP, as well as explain five different subtypes of STDP that were discovered [3].

Quite recently, a unified theoretical model that explains the multiple subtypes of STDP was developed [3]. The model consists of three essential components: backpropagation of action potentials through the axon and the dendrite, dual input activation of the N-methyl-D-aspartate (NMDA) receptors, and calcium dependent plasticity. First, the model carefully derives an accurate representation of the morphology of the backpropagating action potentials using Hodgkin-Huxley equations. Secondly, the model calculates the dual requirement of NMDA receptors for increasing the conductivity of calcium ionic channels. Lastly, calcium dependent plasticity was explained using a two-component model involving calcium dependent protein kinase and phosphatase [3].

Before we discuss the details of the unified STDP model, let us first summarize the important components of neuronal networks.

### 2.1. The neuron

Neurons are highly specialized single cells and are the most basic components of the nervous system. They are electrically excitable cells that process and transmit information around the nervous system. Neurons are the core components of the brain, and spinal cord in vertebrates and ventral nerve cord in invertebrates, and peripheral nerves (Lopez-Munoz, et al, 2006). Figure 2-1 shows a cartoon of a neuron and its different parts.

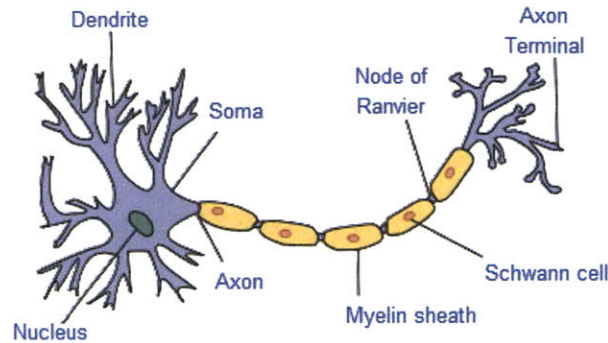


Figure 2-1. A cartoon of a neuron showing its basic parts.

The main parts of the nucleus are the *soma*, *axon*, and the *dendrite*. The soma acts as the processing unit of the neuron and is responsible for generating action potentials. These action potentials are propagated from the soma to the end of the neuron, also called *axon terminal*, by the axon. In the axon terminal, chemical *neurotransmitters* that encode the electrical signal are produced to cross the gap between the axon terminal and the next neuron. This gap is part of the connection system of the two neurons which is collectively referred to as the *synapse*. The neurotransmitters that cross the synapse bind to the input receptors on the dendrites of the postsynaptic neuron. Every neuron has multiple dendrites that are all connected to other neurons. Current signals that are coming from the dendrites, also referred to as *postsynaptic currents*, form *postsynaptic potentials* that are summed at the soma to produce new action potentials.

## 2.2. The Synapse

Synapses are special junctions that enable two neurons to communicate with each other. These structures allow neurons to form interconnected circuit networks within the nervous system and are very crucial to the biological computations that underlie perception, thought, and memory. They also provide the means for the nervous system to connect and control other systems of the body. For instance, the specialized synapse between a motor neuron and a muscle cell is called a neuromuscular junction [4]. A cartoon of the synapse is shown in Figure 2-2. The synapse has three main parts, the axon terminal that contains the neurotransmitters, the *synaptic cleft*, and the *dendrite spine*.

### 2.2.1 Signaling mechanism through the synapse

#### Neurotransmitter release

The first stage of signal propagation through the synapse is signified by the release of neurotransmitters. In the axon terminal, vesicles containing neurotransmitters are docked, ready to release their contents. The arrival of the action potential results in an influx of calcium  $[Ca]$  ions [5] whose presence triggers a biochemical process that results to the release of neurotransmitters to the synaptic cleft about 180 microseconds after  $[Ca]$  influx [6].

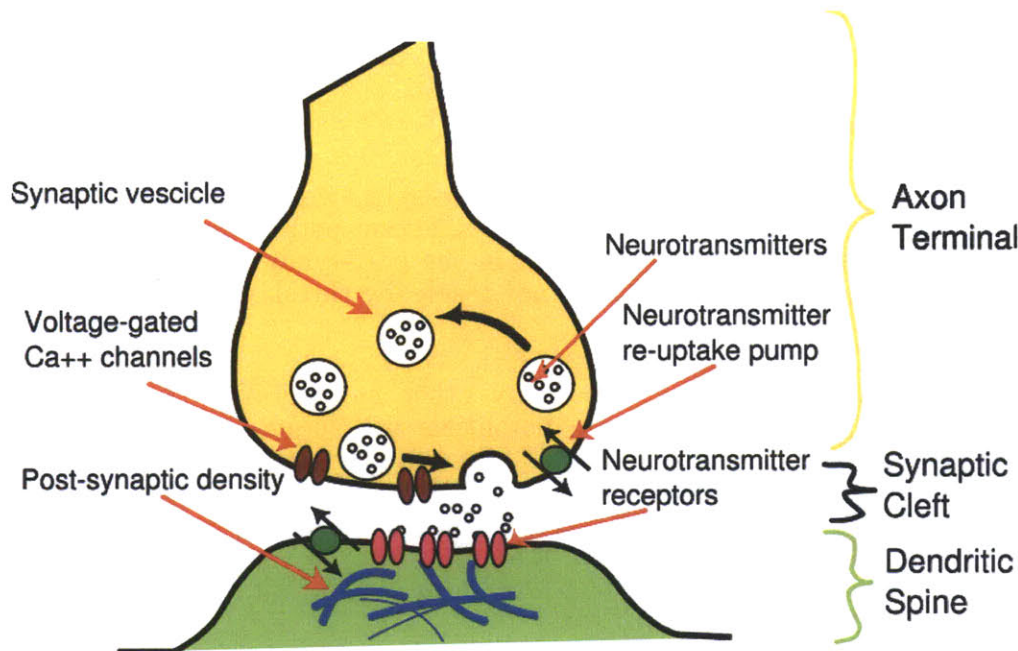


Figure 2-2. A cartoon of a synapse showing its basic parts.

## Receptor binding

Receptors on the dendrite spine bind to neurotransmitter molecules and respond by opening nearby ion channels in the post-synaptic cell membrane, causing ions to rush in or out forming postsynaptic currents that change the local membrane potential of the cell. The resulting change in voltage is called postsynaptic potential. In general, the result is an *excitatory postsynaptic potential* (EPSP), in the case of depolarizing *excitatory postsynaptic currents* (EPSC), or *inhibitory postsynaptic potential* (IPSP), in the case of hyperpolarizing *inhibitory postsynaptic currents* (IPSC). Whether a synapse is excitatory or inhibitory depends on what type of ion channel that conduct the postsynaptic current which in turn is a function of the type of receptors and neurotransmitter employed at the synapse.

## Termination

The last stage of signaling is termination. Presynaptic signals are terminated via the breakdown or reuptake of existing neurotransmitters. Reuptake is mainly localized in the presynaptic neuron and serves to recycle transported neurotransmitters.

### 2.2.2. Synaptic strength

The strength of a synapse is defined by the change in postsynaptic current as a result from the activation of postsynaptic neurotransmitter receptors. Changes in synaptic strength can be short term (short term potentiation/depression, or STP/STD) and causes no permanent structural changes in the neuron. Typically, this change lasts a few seconds to minutes. Sometimes, strength changes are long term (long term potentiation/depression, or LTP/LTD) [7]. For these types of changes, repeated or continuous synaptic activation results in an alteration of the

structure of the synapse itself. Learning and memory are believed to result from long term changes in synaptic strength, via the synaptic plasticity mechanism.

### 2.2.3. Integration of synaptic inputs

Having discussed the notion of synaptic strength, we can then differentiate a strong from a weak synapse. In strong synapses, action potentials in the presynaptic neuron will trigger another action potential in the post-synaptic neuron, whereas, in a weak synapse, the sum of EPSPs may not reach the threshold for action potential initiation.

Each neuron forms synapses with many other neurons and therefore receives multiple synaptic inputs. When action potentials fire simultaneously in several of these neurons, multiple EPSCs are created which all generate EPSPs that sums up in the soma. Hence, the output of a neuron may depend on the input of many others, each of which may have a different degree of influence, depending on the strength of its synapse with a specific neuron [8]. John Carew Eccles performed some of the important early experiments on synaptic integration, for which he received the Nobel Prize for Physiology or Medicine in 1963.

## 2.3. Synaptic plasticity

Synaptic plasticity, the phenomenon of synaptic strength modification, has been intensively studied because it is believed to be the underlying mechanism of learning and memory. STDP refers to the long term potentiation (LTP) or depression (LTD) of synapses as a function of the relative timing between the external stimuli to presynaptic and postsynaptic neuron [9]. Studies on STDP has boomed since STDP's discovery and it was soon discovered that a Hebbian synapse becomes potentiated (strengthened) if the presynaptic neuron is stimulated before the activation of postsynaptic neuron and becomes depressed (weakened) if the presynaptic neuron is stimulated after the activation of the postsynaptic neuron [3].

The profiles of STDP as functions of spike timing between presynaptic and postsynaptic spiking can vary significantly depending on which part of the nervous system is studied. Two major types of STDP responses exist: symmetric and asymmetric STDP. Symmetric STDP demonstrates no significant difference in the nature of synaptic strength change regardless of the temporal order between the presynaptic and postsynaptic stimuli. On the contrary, asymmetric STDP shows the opposite effect of long-term plasticity under opposite temporal order. Asymmetric STDP is further divided into two subtypes: Hebbian and anti-Hebbian. In the former subtype, the synaptic strength gets potentiated when EPSPs induced by presynaptic stimuli precede postsynaptic action potentials. Rat hippocampal synapses belong to this class. Figure 2-3 shows an example of a hebbian STDP. The latter subtype demonstrates depression under the same pre-post temporal stimulation. Synapses linking parallel fiber to Purkinje-like cells in electric fish are good examples of anti-Hebbian subtype of STDP [3].

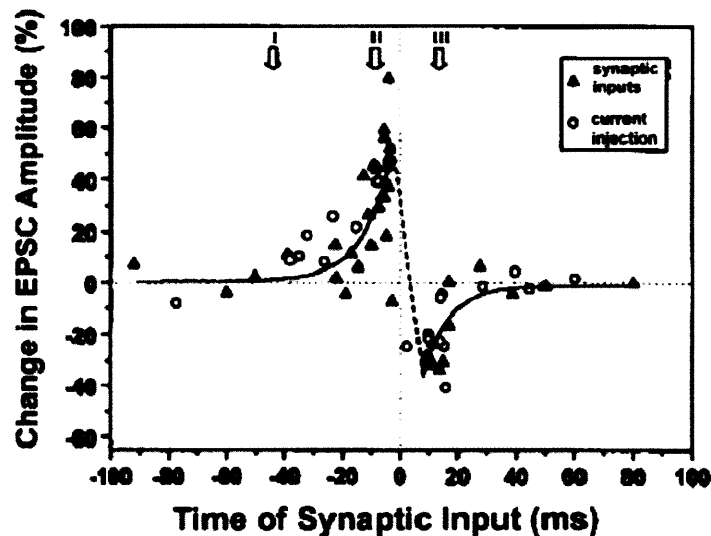


Figure 2-3. Hebbian STDP. *Figure obtained from [2].*

Knowing that there are various subtypes of STDP, we might wonder whether we can elucidate a unified theory about the mechanism of STDP that could potentially explain its different subtypes. Lee hypothesized that such a theory “exists and can be built as long as we accurately model essential molecular mechanisms that contribute to the temporal dependence of synaptic plasticity”. Such mechanisms include the back propagation of action potential to dendrites which can be seen as an essential process in transmitting postsynaptic signals. On the other hand, glutamate neurotransmitter signals could symbolize the presynaptic stimuli. The dynamics of calcium current that are gated by NMDA receptors could be responsible for temporally linking the presynaptic and postsynaptic stimuli [3].

## 2.4. Previous models of synaptic plasticity

Since the discovery of STDP, several models were proposed and developed to explain it. Some of these models aim to reproduce the experimental data with a mathematical expression to model the results from a macroscopic perspective. This type of modeling is called empirical. Perhaps, the simplest empirical model describes the STDP responses as two separate levels of synaptic plasticity subject that exponential decays. This type of model successfully demonstrates the time-dependent bidirectionality of STDP due to interacting depression and potentiation processes. However, the weakness of empirical models is the fact that we do not gain any significant knowledge about the biomolecular reactions at a cellular level which is responsible for STDP [3].

Other types of models focus instead on deriving physiological mechanisms at the biomolecular level to interpret experimental findings. These models are classified as mechanistic models. The calcium control hypothesis for instance [10], suggests that synaptic plasticity is dependent on the calcium level. This hypothesis clearly supports the well known and widely accepted BCM rule that was proposed by Bienenstock, Cooper, and Munro [11]. Eventually, scientists found out that NMDA receptors through which calcium ions flow into the postsynaptic membrane are the focal point of several mechanistic models [10, 12, 13]. These receptors serve as

temporal coincidence detectors that relate the timing of glutamate release from the presynaptic neuron and the back propagation of postsynaptic action potentials for both these stimuli are required to activate NMDA receptors which are both voltage and ligand gated ion channels. However, these mechanistic models can not explain asymmetric STDP responses. To overcome this issue, models that have two coincidence detectors were developed as well [14] but they did not quite have a strong biophysical reasoning.

These previous efforts are very important in shedding light in the mystery that surrounds STDP. However, none of them can completely explain the whole picture of STDP. No one admits this, but truth be told, the scientific community is craving for a fresh ideas for modeling STDP thereby causing more and more people to feverishly work to unlock this mystery. In the next section, we will present a novel and groundbreaking theory that we believe captures the whole picture of synaptic plasticity.

## 2.5. Unified model of synaptic plasticity

We conjectured that STDP requires a series of subsystems that interact with one another in a timely fashion. Unlike simple models that were developed in the past, we constructed a rather composite model consisting of several subsystems that are based on actual biophysical reactions in the synapse. We can summarize this model by these few sentences. Presynaptic stimuli lead to the release of glutamate neurotransmitter while postsynaptic activation leads to the propagation of the dendritic action potential. Calcium signaling reaction begins when NMDA receptors promote calcium influx which occurs when there is a glutamate influx and prolonged membrane depolarization due to the backpropagating dendritic action potential. Calcium kinase ( $Ca_4K$ ) then catalyzes the activation of Alpha - Amino - 3 - hydroxyl - 5 - Methyl - 4 - Isoxazole Propionic Acid (AMPA) receptors while calcium phosphatase ( $Ca_4P$ ) promotes the breakdown of these AMPA receptors [3]. The balance between the kinase and the phosphatase is determined by the total amount of calcium that flows to the postsynaptic membrane. Figure 2-4 summarizes this model.

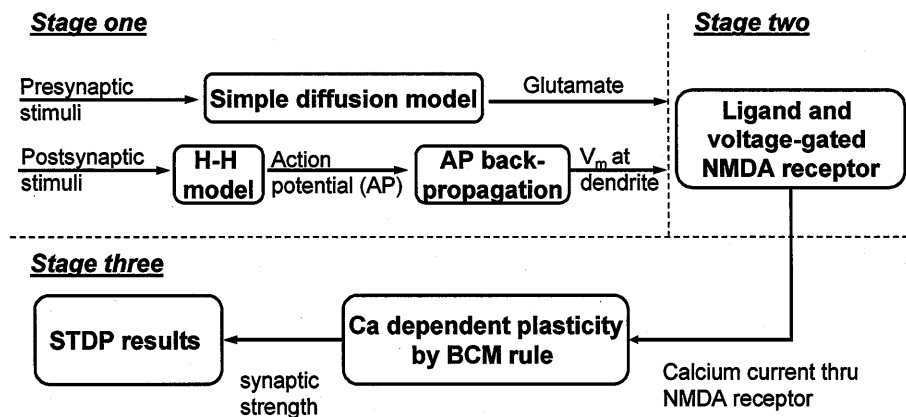


Figure 2-4. Block diagram of the unified STDP model. Figure is obtained from [3].

There are three stages in the STDP model. First is the backpropagation of action potential, second is the NMDA receptor activation, and third is the calcium



transduction pathway. We shall therefore organize our discussion based on these three stages. In the first stage, presynaptic stimuli get transduced to glutamate flux while the postsynaptic stimuli get back propagated as a dendritic action potential before they act on the NMDA receptors together. In the second stage, the influx of calcium enters the postsynaptic neuron through NMDA receptors that are both ligand and voltage gated. Finally in the third stage, activated NMDA receptors allow the accumulated calcium level to determine the synaptic strength change using the BCM rule. A two-component model was developed to model the calcium dependent plasticity and validated by experimental data of frequency-dependent plasticity [3].

### 2.5.1. Backpropagation of action potentials

To initiate an action potential, we follow the standard experimental protocol of spike timing dependent plasticity. Here, the postsynaptic neuron is stimulated milliseconds before or after the presynaptic neuron is stimulated. The action potential due to the stimulation of the postsynaptic neuron gets back propagated to the dendrites. We noted that most of the previous theoretical undertakings did not adopt rigorous biophysical models for plotting action potentials and did not consider detailed dynamics of backpropagation [3]. For example, the morphology of action potentials in the NMDA receptor-dependent bidirectional model was simply modeled as two monotonously decaying exponential functions [10]. In reality though, experimental recordings of action potentials typically illustrate depolarization, repolarization, and refractory phases. This is a very important detail that researchers failed to incorporate in their experiments and is the main reason why none of the past researches were able to explain the different forms of STDP.

To accurately represent the morphology action potential, we employed Hodgkin-Huxley equations that represent the dynamics of the ion channels of the neuron. We then decided to model the postsynaptic neuron by an equivalent circuit consisting of axon and dendrite compartments [3]. This is justifiable because the action potential back propagates through these two pathways in order for it to reach the dendrite.

We carefully examined the shape of the action potentials that pertains to each experiment to ensure a precise theoretical model. Ordinary Hodgkin-Huxley equations are adequate to model Hebbian asymmetric STDP. However, the shapes of the action potentials in anti-Hebbian asymmetric STDP and symmetric STDP shows *depolarizing after potential* (DAP) behavior. Hence, for these two cases we introduced DAP into H-H model by inverting the hyperpolarization part of action potentials.

### 2.5.2. NMDA receptor activation

It has been shown by recent studies that postsynaptic calcium transduction plays an essential role in regulating long term synaptic plasticity and it turns out that NMDA receptor-gated ion channels are the main pathways for calcium flux to the postsynaptic neuron [15, 16]. These NMDA-receptor channels are very sensitive channels. To open, they require both the binding of glutamate and a substantial degree of depolarization. This led us to define a property of these NMDA channels that represents how well they allow ions to pass through them. This property is called the conductance. Clearly, as the conductance increases, ions pass through the channel more easily.

We defined the conductance of NMDA receptors as  $g_{\text{NMDAR}}$  and as it turned out, it is actually a product of two terms: the dependence function of glutamate concentration (Glu) and that of membrane voltage ( $V_{\text{dent}}$ ). Note that glutamate flux mainly results from presynaptic stimuli while depolarization is caused by the back propagation of postsynaptic action potential [3].

Glutamate dependence can be described by the ligand-receptor model whose association constant is  $K_{\text{Glu}}$ . Glutamate concentration can be related back to the presynaptic stimuli by a simple diffusion model whose details are found in Lee's thesis. The voltage dependence of the NMDA channel has been modeled with the logistic function below:

$$g_{\text{NMDAR}}(\text{Glu}, V_{\text{dent}}) = \frac{\text{Glu}}{\text{Glu} + K_{\text{Glu}}} \times \frac{g_{\text{NMDARmax}}}{1 + e^{-k_{\text{NMDAR}}(V_{\text{dent}} - V_{1/2})}}$$

The exponential term of this equation is derived from the Arrhenius equation [3]. In qualitative terms, the voltage dependent activity function rises to a maximum  $g_{\text{NMDARmax}}$  following a sigmoidal curve whose half voltage is equal to  $V_{1/2}$  and whose slope is  $0.5k_{\text{NMDAR}}$  at the half voltage point [3].

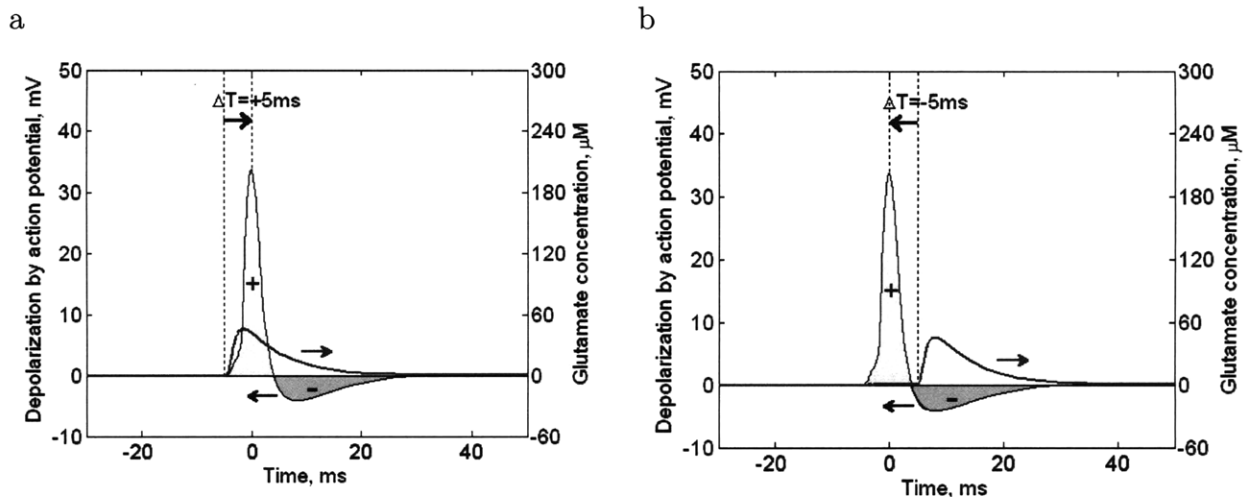
Once we found out the conductance of NMDA channels, we determined the amount of calcium flowing to the postsynaptic neuron. We reasonably assumed that the NMDA receptors are the major gateway of calcium inflow, thus, calcium concentration would be proportional to the integration of current flow through NMDA receptors over the time span of stimuli:

$$\text{Ca} \propto \int_{t_{\text{stimuli}}} I_{\text{NMDAR}} dt = \int_{t_{\text{stimuli}}} g_{\text{NMDAR}} V_{\text{dent}} dt$$

Where  $I_{\text{NMDAR}}$  is expressed as the product of the NMDA conductance  $g_{\text{NMDAR}}$  and dendrite membrane potential  $V_{\text{dent}}$  [3].

The lag time between the glutamate flux and arrival of dendritic action potential influences the level of NMDA receptor activation which in turn modulates the amount of calcium influx. Since both transient signals only last for a few tens of milliseconds, such temporal proximity between presynaptic and postsynaptic stimuli is necessary to trigger the interaction between the two signals. Figure 2-5 shows the transient profiles of glutamate from the presynaptic neuron and dendritic action potential at the postsynaptic neuron with lag time equal to  $\pm$  five milliseconds [3].

Intuitively, we can see that if the glutamate profile overlaps mostly with the positive portion of action potential (Figure 2-5.a), the NMDA receptor is activated causing more calcium to flow into postsynaptic neuron. On the other hand, if glutamate influx occurs at the same time as the negative portion of the action potential (Figure 2-5.b), the NMDA receptor is not fully opened, which ultimately results into less calcium inflow [3]. Now that we can determine the calcium level, we can use this to determine the direction of synaptic plasticity.



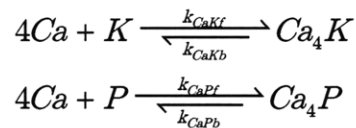
**Figure 2-5. Profiles of action potential and glutamate concentration.** (a) Onset of glutamate input precedes the peak of action potential by 5 ms. (b) Onset of glutamate input lags the peak of action potential by 5 ms. *Figure obtained from [3].*

## 2.5.2. Calcium transduction pathway

Accumulated calcium in the postsynaptic compartment through the action of the NMDA receptors strongly influences the action of synaptic plasticity. Depending on the total amount of calcium inflow, postsynaptic calcium can either upregulate or downregulate the signal transduction pathways that lead to synaptic plasticity change. Synaptic strength tends to exhibit LTD at low levels of calcium and LTP at high levels [17, 18]. This forms the basis of the BCM rule. Basically, this rule states that an omega-shaped function describes the dependence of synaptic strength on postsynaptic activity [11]. Eventually, this rule was used in STDP models to relate synaptic strength to intracellular calcium concentration [10]. Detailed mechanistic models involving complex multi-step calcium dependent enzyme activation are quite challenging to reproduce [3], but after extracting essential reaction mechanisms, a bidirectional calcium binding model was developed with sufficient biophysical details to describe the phenomena.

Calcium in the postsynaptic neuron influences synaptic plasticity through two steps. To begin, the calcium either binds with calcium-dependent protein kinase or phosphatase. Afterwards, these enzyme-calcium complexes either insert active AMPA receptors [19] to the postsynaptic membrane if they are kinases or remove AMPA receptors from the membrane if they are phosphatases [3, 20]. The total number of active AMPA receptors remaining in the postsynaptic membrane determines the magnitude of the EPSC.

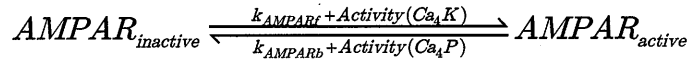
We can describe the binding between the calcium and two different types of enzymes using the reaction below:



Trapped calcium can either bind with calmodulin-dependent protein kinase II (K for short) or protein phosphatase 1 (P for short). The binding between enzymes and calcium is assumed to be cooperative and four calcium ions are required to fully activate the kinase ( $\text{Ca}_4\text{K}$ ) or protein phosphatase ( $\text{Ca}_4\text{P}$ ). To simplify things, we define the binding ratio of  $\text{Ca}_4\text{K}$  as  $R_{\text{CaK}}$  and that of  $\text{Ca}_4\text{P}$  as  $R_{\text{CaP}}$ . In equilibrium the values of  $R_{\text{CaK}}$  and  $R_{\text{CaP}}$  are:

$$R_{\text{CaK}} = \frac{\text{Ca}^4}{\text{Ca}^4 + K_{\text{CaK}}^4} ; \quad R_{\text{CaP}} = \frac{\text{Ca}^4}{\text{Ca}^4 + K_{\text{CaP}}^4}$$

Afterwards, calmodulin-dependent protein kinase II bound by calcium ( $\text{Ca}_4\text{K}$ ) can insert and activate AMPA receptors whereas calcium-bound protein phosphatase 1 ( $\text{Ca}_4\text{P}$ ) may internalize and deactivate AMPA receptors. This reaction is shown in this reaction:



Where  $k_{\text{AMPA}Rf}$  and  $k_{\text{AMPA}Rb}$  denote the forward and backward reaction rate constants of AMPA receptor activation.

$\text{Ca}_4\text{K}$  activity can be computed as the product of  $V_{\text{CaK}}$  and  $R_{\text{CaK}}$ . Similarly, the same relationship holds to the activity of  $\text{Ca}_4\text{P}$  too. Intuitively,  $V_{\text{CaK}}$  and  $V_{\text{CaP}}$  are proportional constants that quantify the contributions of  $\text{Ca}_4\text{K}$  and  $\text{Ca}_4\text{P}$  to the rate constants, respectively. Changes in the activity of  $\text{Ca}_4\text{K}$  and  $\text{Ca}_4\text{P}$  caused by calcium inflow would modify the number of active and inactive AMPA receptors. If we assume that the total number of AMPA receptors remains constant, the difference between synaptic strength before and after stimuli ( $\Delta W$ ) is proportional to the change in normalized forward rate constants [3].

$$\Delta W \propto \frac{k_{\text{AMPA}Rf} + V_{\text{CaK}} R_{\text{Ca}_4\text{K}}}{k_{\text{AMPA}Rb} + V_{\text{CaP}} R_{\text{Ca}_4\text{P}} + k_{\text{AMPA}Rf} + V_{\text{CaK}} R_{\text{Ca}_4\text{K}}} - \frac{k_{\text{AMPA}Rf}}{k_{\text{AMPA}Rb} + k_{\text{AMPA}Rf}}$$

After determining the relationship of  $\Delta W$  with the total calcium in the postsynaptic neuron, we now have a working model that predicts how  $\Delta W$  changes as we stimulate the presynaptic and postsynaptic neuron with varying lag times. This is the unified model of synaptic plasticity which we use to construct our circuits.

### 3. Fundamentals of low power neuromorphic circuit design

#### 3.1. Neuromorphic circuit design overview

Neuromorphic circuit design was established by Carver Mead in the late 1980's to describe analog very large scale integration (aVLSI) circuits that emulate neuronal networks present in the nervous system [21]. Recently, the term neuromorphic circuit design has been used to refer to analog, digital or mixed signal VLSI systems that implement circuit models and real time simulation of different neuronal systems. Such systems include perception, motion control, and signal processing pathways in the brain [22].

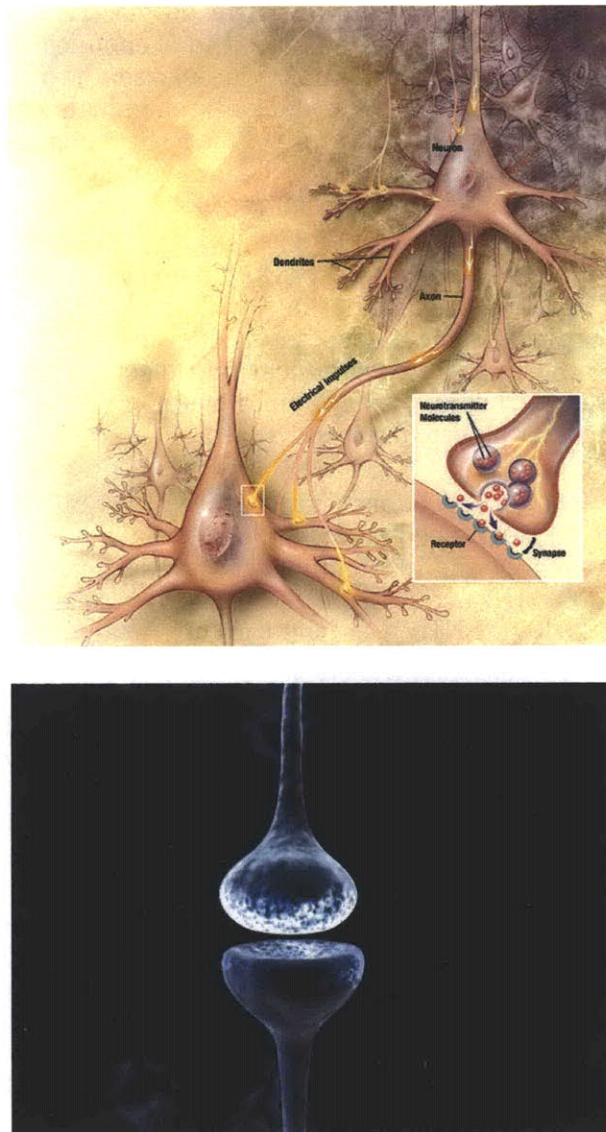


Figure 3-1. The synapse and the physical gap between two neurons. Chemicals need to travel between the gap to send information between the two neurons.

The first thing to consider when designing a neuromorphic circuit is to understand how the biochemical processes of neurons compare with available circuit techniques and overall architectures. Circuits are an excellent way to represent neuronal networks because neuronal processes are governed by ion diffusion and drift in neuronal ion channels [21, 23]. Science has evolved to such an extent that accurate mathematical models have been developed to characterize the physics of neuronal biological processes. What is amazing though is that circuits also follow the same physical laws which allow us to reproduce biological computations using analog circuits.

A basic neuronal network is usually comprised of a few interconnected neurons that have specific behavior. Information is usually encoded in terms of action potentials or digital spikes via the spiking frequency, spike amplitude, or in some cases, spike morphology. These action potentials propagate through the axon which can be modeled as the information carrier medium of the neuronal network. To travel to another neuron, the information stored in an electrical action potential has to be transduced so this information can be stored in a chemical format. This change is essential because there is no physical connection between two neurons for electrical signals to travel to. Literally, a gap exists and only chemicals can pass through this gap, which is also referred to as the synapse. Once the chemical reaches the next neuron, the information is transduced back to an electrical action potential [24-27] Figure 3-1 shows the synapse and the physical gap between two neurons.

In constructing a complete neuronal network, we need robust and accurate circuit models of action potential generators that will be used to encode electrical information and chemical signal generators to encode chemical information. We also need a synapse that is able to analyze chemical and electrical inputs and follow the observed synaptic plasticity behavior. To build such circuits we then face many design considerations such as power, area and speed.

Fortunately, the neuron is not a fast signal processing unit. Hence, circuit speed is not a big design issue that we need to face. Now, we are not saying that the brain is a slow machine. However, compared to modern circuits which typically operate in high frequencies (1 - 1000 MHz), the brain is fairly slow as neurons typically reside in low frequency bands (10- 100 KHz). So the two important design issues are power and area. Area is typically solved by building circuits using the minimum amount of transistors while avoiding large capacitors and resistors. On the other hand, power issue is solved by designing circuits that operate in subthreshold or weak inversion regime. These circuits typically have transistors that draw currents in the picoampere (pA) and nanoampere (nA) range which is coincidentally similar to the order of magnitude of currents in neuronal circuits. This fact makes it even more appealing and easier to emulate neuronal networks using low power aVLSI circuits.

The biggest drawback of analog design, thermal noise, is not a big problem when capturing the behavior of neuronal networks because these networks are inherently noisy. Now we might wonder, why not use the alternative design technique, digital design? The main reason why we do this is because analog computation is almost instant, as computation occurs in real time and is limited only by the motion of dopant ions in semiconductors. Digital design on the contrary, needs more complicated circuitry to implement biological computation, as every bit of precision needs to be encoded. Thus, we burn more power and more importantly, it takes longer time to get results.

One intuitive example to see the beauty of analog versus digital design is to consider the simple process of addition. Efficient and fast digital adders takes an order of  $n \log(n)$  time and circuitry to perform  $n$  bits of addition [28]. On the other hand, to perform addition in analog domain, we simply need to combine two current sources in one node and observe the current exiting that node in real time! Kirchoff current laws naturally dictate that the sum of those currents must be the magnitude of the current observed at the output node.

For this chapter, we will discuss the operation of weak inversion transistor circuits. Furthermore, low power and area saving techniques will be also covered in this section. Finally, we shall also present to you basic circuit building blocks that will enable us to build useful aVLSI circuits that emulate neuronal networks.

## 3.2. Metal oxide semiconductor field effect transistor (MOSFET)

Every aspiring electrical engineering student probably knows a lot about resistors, capacitors, inductors, and voltage sources. However, these elementary devices, while useful for composing basic circuits and understanding circuit theory in general, are not the main components of VLSI circuit design. Truth be told, the main device used in circuit design is actually the **transistor**.

As an engineering student myself, I have seen a lot of students understand and appreciate resistor and capacitor networks. However, the same students suddenly become overwhelmed when they get introduced to the transistor. How come? This device is just like a resistor! Like a resistor, it has current-voltage (I-V) characteristics, albeit not as linear as the resistor. Nevertheless, it has fully characterized I-V curves that have been modeled mathematically. Upon realizing this fact, I began to appreciate how fascinating this little device is in creating useful circuits.

So why do people use transistors instead of resistors and capacitors? The reason is quite simple: transistors are small, flexible, and easy to produce. They can also be operated in very low power, i.e. they use up very low amounts of currents (pA levels) to perform interesting operations. Resistors on the other hand, need to be huge in order use up low power. A resistor that operates in similar power levels as a transistor would be at least thousand times larger than the transistor. This is very detrimental to chip design as we do not want chips are bulky. Because of this, we try to limit the amount of resistors that we put in our design. Same thing goes for capacitors.

However we have to be careful in working with transistors for they are strange gizmos whose behavior varies depending on the amount of current flowing through them. Since most of our circuits are low power, we will focus on understanding the basics of subthreshold transistor operation.

### 3.2.1. MOSFET fundamentals

The MOSFET is a four terminal device that is used for most integrated circuits we can find in the market nowadays. Its four terminals are called Gate (G), Source (S), Drain (D), and Bulk (B). Without going into all details how these four terminals interact, let me just say that the voltages between any of these two terminals react

to current that flows through them. Standard schematics of transistors are shown in Figure 3-2.

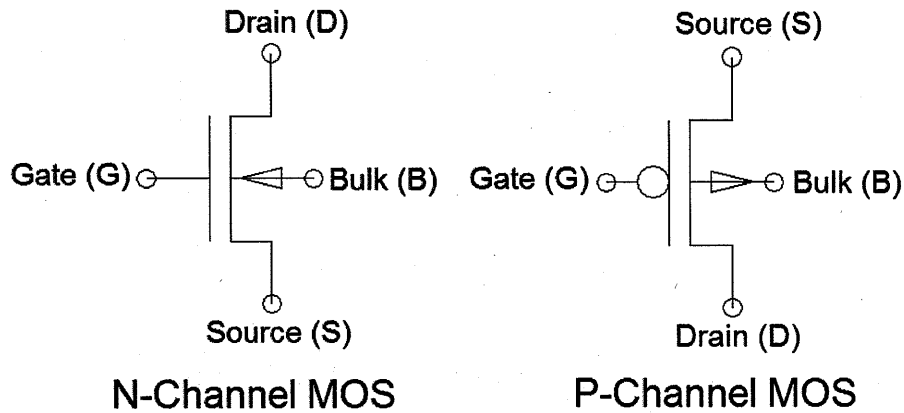


Figure 3-2. Schematics of the NMOS and PMOS.

There are two types of MOSFETS, the N-channel MOSFET (NMOS) and the P-channel MOSFET (PMOS). NMOS is primarily used to sink current while PMOS are used to source currents. Currents flow opposite across these two types of transistors because the dopant ions for these two are of opposite polarity. The theory behind these transistors can be found in any electronics textbook and I will not waste time explaining this. Instead, I will focus on explaining the essential I-V relationship of the MOSFET in subthreshold or low power regime.

### 3.2.1. MOSFET in subthreshold regime

We can apply and measure the voltage between the MOSFET's gate and source. This voltage is referred to as  $V_{GS}$ . This affects the current flowing through the drain  $I_D$ .  $V_{GS}$  can be modified by two ways – by setting a fixed current through the drain of the transistor or by applying a fixed voltage across the gate and source of the transistor. Now, all MOSFETs have a certain threshold  $V_{GS}$  where the  $I_D$  to  $V_{GS}$  relationship changes dramatically. This threshold voltage is referred to as  $V_{th}$ . We call the MOSFET operation as above threshold when  $V_{GS} > V_{th}$ . If  $V_{GS} < V_{th}$ , the MOSFET is operating in the subthreshold regime.  $I_D$  is much more larger in above threshold than in subthreshold (about 1,000 - 1,000,000 times larger).

In subthreshold regime, the Boltzmann distribution of electron energies allows some of the more energetic electrons at the source to enter the channel and flow to the drain, resulting in a subthreshold current,  $I_D$  that is an exponential function of  $V_{GS}$ . In weak inversion, the current varies exponentially with gate-to-source bias  $V_{GS}$  as given approximately by [29, 30]

$$I_D = I_o e^{\frac{\kappa V_{GS}}{\phi_t}} e^{\frac{(1-\kappa)V_{BS}}{\phi_t}} \left( 1 - e^{-\frac{V_{DS}}{\phi_t}} \right) \quad (1)$$

$I_o$  is the intrinsic current that flows when  $V_{GS}$  is equal to zero. This term is process dependent or in other words, it depends on how the chip and silicon was



manufactured.  $\phi_t$  is the thermal voltage and is defined as  $\phi_t = kT/q$ , where  $k$  is the Boltzmann constant,  $q$  is the charge of a single electron and  $T$  is the absolute temperature of the room. The value of  $\phi_t$  is typically around 25 mV in room temperature. On the other hand,  $\kappa$  is also process dependent constant and typically its value is around 0.7.

(1) could be simplified if we short circuit the bulk to the source. This will set  $V_{BS}$  to be equal to zero and the second exponential term becomes equal to one. Also, if we bias the transistor so that it is saturated which is setting  $V_{DS}$  to be 100 mV or bigger, the third exponential term would simplify to about zero because  $V_{DS} \geq 4\phi_t$  which would put make the last term equal to above 0.99. After these series of simplifications, our new expression for  $I_D$  becomes:

$$I_D \approx I_o e^{\frac{\kappa V_{GS}}{\phi_t}} \quad (2)$$

We then get a purely exponential relationship between  $V_{GS}$  and  $I_D$ .

### 3.3. Basic subthreshold circuit blocks

Now that we have an idea how the MOSFET in subthreshold regime works, we can start discussing basic building blocks that we will use to build neuromorphic circuits. These circuits are comprised of several transistors and in some cases resistors and capacitors that work together to produce a desirable circuit behavior.

#### 3.3.1. Current mirrors

Current mirrors allow us to create multiple scalable copies of a current from a control circuit to another active circuit by modulating the current drive of the control circuit. The mirrored current is held equal to the control current regardless of the load in the active circuit. Mirrors are essential if we want a certain signal carrying current to be used in several active parts of our circuit. Properly biasing circuits also require current mirrors for creating accurate, non-fluctuating current sources are quite challenging. If all our circuits would require accurate biasing current sources, it is probably better to create one accurate current source and create scaled mirrors of it for biasing our circuits. This is much better than creating multiple accurate current sources which would consume a lot of space in the chip.

The basic current mirror is shown below in Figure 3-3. The input current  $I_{in}$  is mirrored to the output current  $I_{out}$ . M1 and M2 represent the  $W/L$  ratios of the left and right transistor respectively, and  $W$  and  $L$  are the widths and lengths of the transistor respectively. So by varying the size of the transistor we can create scale copies of the  $I_{in}$ . So how exactly does a current mirror work?

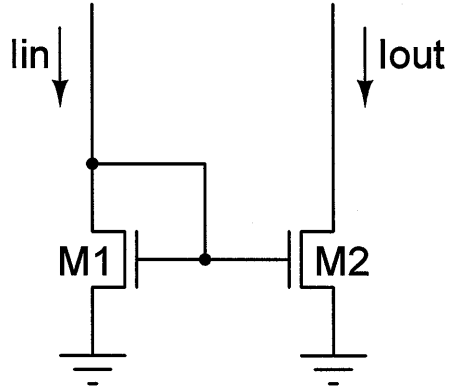


Figure 3-3. Schematic of a current mirror.

Clearly,

$$I_{in} = I_{o1} e^{\frac{\kappa V_{GS1}}{\phi_t}} \quad (3)$$

Since we are controlling  $I_{in}$ ,  $V_{GS1}$  will be set to a value depending on  $I_{in}$  such as:

$$V_{GS1} = \frac{\phi_t}{\kappa} \ln \left( \frac{I_{in}}{I_o} \right)$$

Because the gates of the two NMOS are tied together and they share a common source :

$$V_{GS1} = V_{GS2} \quad (4)$$

Now the main difference between the two transistors is M1 and M2. However,  $I_o$  is proportional to M. Given that:

$$I_{out} = I_{o2} e^{\frac{\kappa V_{GS2}}{\phi_t}} \quad (5)$$

Combining (3), (4), (5), gives us:

$$I_{out} = \frac{M2}{M1} I_{in} \quad (6)$$

So, to create equal copies of  $I_{in}$ , we just need to set  $M1 = M2$ .

### 3.3.2. Current multipliers and dividers

Current mode circuits employ current signals to perform a certain task. For arithmetic operations, currents are so easy to work with because addition and subtraction can be easily done by connecting multiple current sources together. Kirchoff's current law simply mandates that the total current entering at a junction must equal the current exiting it. The same thing can be said for current

multiplication and division. Because of the beauty of transistor current mode design, it is very easy to create compact low-power current multipliers and dividers.

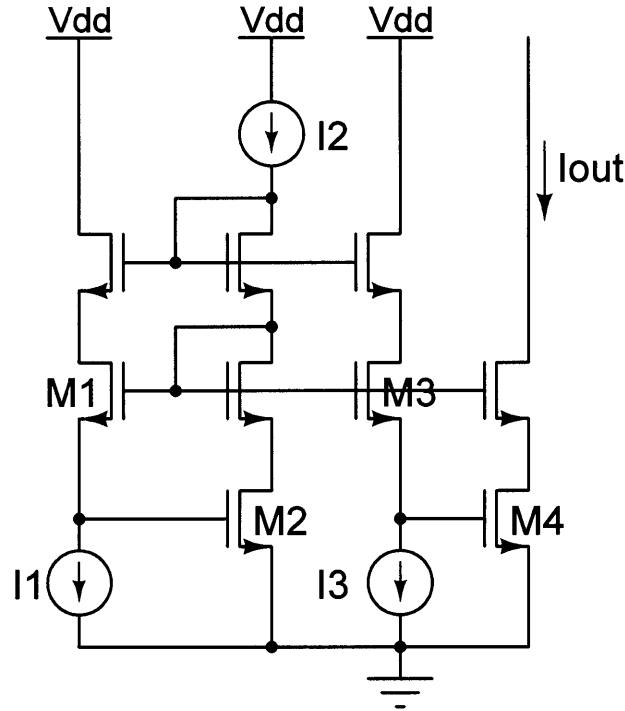


Figure 3-4. Schematic of a low power one-quadrant multiplier.

Consider the circuit in Figure 3-4. It is a circuit that performs low power division single quadrant multiplication and division [31]. This circuit works because of the translinear principle [32]. The translinear principle states that in a closed loop containing an even number of translinear elements, in our case, MOSFETS, with an equal number of clockwise and counter-clockwise components, the product of the currents through the clockwise MOSFETS equals the product of the currents through the counter-clockwise MOSFETS. Hence:

$$\prod_{n \in CW} I_n = \prod_{n \in CCW} I_n \quad (7)$$

In our circuit, M1, M2, M3, M4 form a translinear loop with M1, M2 forming the clockwise branch and M3, M4 forming the counter-clockwise branch. Combining (7), and these facts: a.  $I_{out}$  flows through M1, b.  $I_3$  flows through M2, c.  $I_2$  flows through M4, and d.  $I_1$  flows through M3, we can solve:

$$\begin{aligned} I_{out} I_3 &= I_2 I_1 \\ I_{out} &= \frac{I_2 I_1}{I_3} \end{aligned} \quad (8)$$

Thus, to multiply two currents we simply need two current sinks and replace  $I_2$  and  $I_1$  with these sinks.  $I_3$  could just be a constant current with a magnitude of 1. Similarly, we can divide two currents by using  $I_1$  and  $I_3$  has current signals and use  $I_2$  as a constant current. We can then obtain the monophasic output from the  $I_{out}$  node.

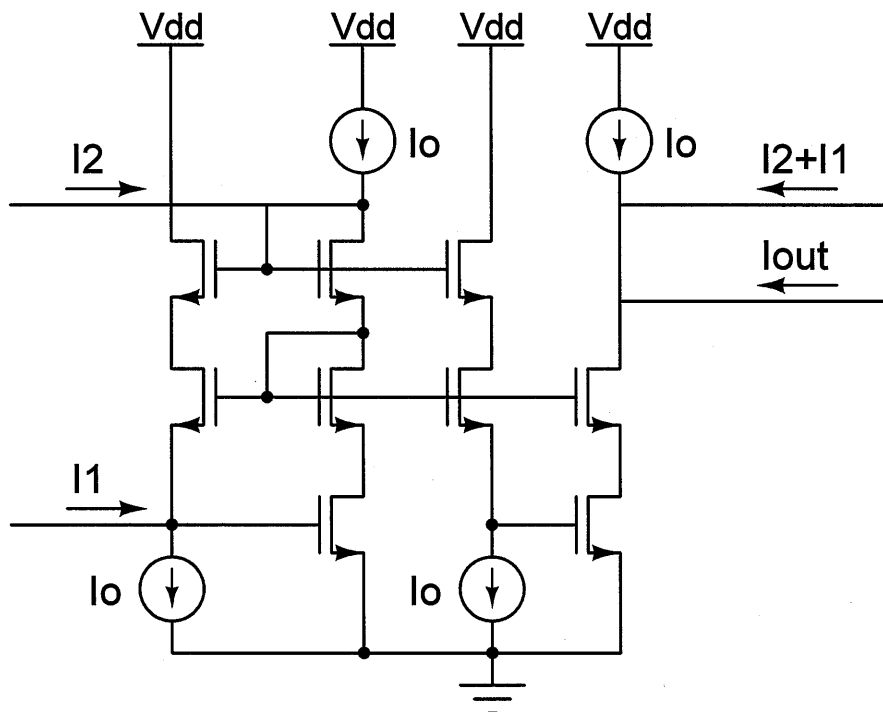


Figure 3-5. Schematic of a low power four-quadrant multiplier.

We can extend the previous schematic to obtain a four quadrant multiplier/divider design. The disadvantage of single quadrant design is that our inputs and outputs can only be monophasic. Four quadrant design allows us to have biphasic inputs and outputs which makes our arithmetic more flexible. The schematic for such a four quadrant design is shown in Figure 3-5. Here, we use the same old divider but we make slight modifications to the bias currents. Now, we use eight current inputs 4 of which are instances of a constant reference current  $I_o$ , 2 of them are instances of  $I_1$ , and the last 2 are instances of  $I_2$ . So now let us derive  $I_{out}$  based from our schematic.

Applying our new schematic to equation (8) gives us:

$$\frac{(I_1 + I_o)(I_2 + I_o)}{I_o} = I_1 + I_2 + I_o + I_{out}$$

$$\frac{I_1 I_2 + I_1 I_o + I_2 I_o + I_o^2}{I_o} = I_1 + I_2 + I_o + I_{out}$$

$$\frac{I_1 I_2}{I_o} + I_1 + I_2 + I_o = I_1 + I_2 + I_o + I_{out}$$

We can then get a biphasic output:

$$\frac{I_1 I_2}{I_o} = I_{out}$$

### 3.3.3. Transconductance amplifiers

We have been talking a lot about current mode circuits and current mode arithmetic but most of the time signals are usually in the form of voltages not currents. The question how can we perform current mode arithmetic to voltage signals? Well the obvious answer is to convert voltage to current. However, how do we do this efficiently without worrying about loading problems? Of course, a rather simplistic approach would be to apply the voltage to a resistor which would give a corresponding current. However, if we apply this current to a load circuit which has a non-zero input impedance, our current will decrease and our arithmetic will have errors in it. To solve this problem of converting voltage to current, we use one of the most widely-used building block of analog circuits, the transconductance amplifier as shown in the Figure 3-6 below.

Transconductance amplifiers are usually implemented as a voltage-controlled linear conductor. In reality though, the transconductance amplifier is a non linear device that implements an output sigmoidal hyperbolic tangent ( $\tanh$ ) current,  $I_{out}$  from a differential input voltage  $V_+ - V_-$ . However, in the middle of the function lies a linear range where we can approximate  $I_{out}$  to be linearly related to  $V_+ - V_-$ . To understand the function of the transconductance amplifier, let us solve for its current-voltage characteristic function.

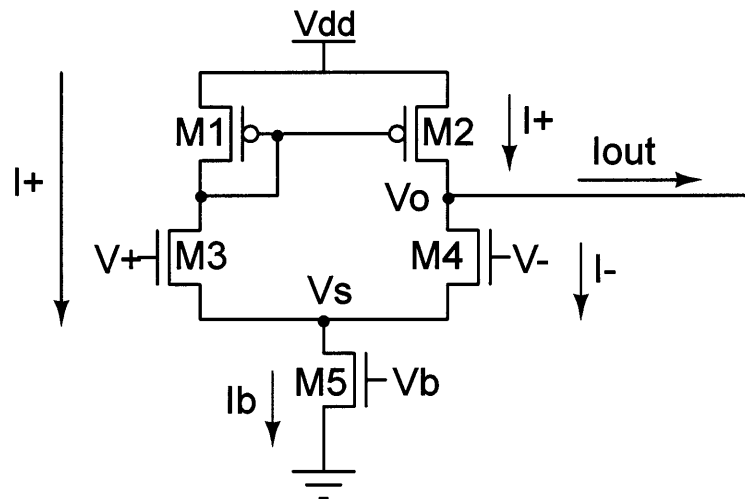


Figure 3-6. Schematic of an ordinary transconductance amplifier.

Assuming the transistors are all saturated and all the bulk terminals are tied to the source, equation (2) should hold for all transistors. Considering M3, the current that flows through it is  $I_+$  and its  $V_{GS}$  is  $V_+ - V_s$ . Applying (2) gives us:

$$I_+ = I_o e^{\frac{\kappa(V_+ - V_s)}{\phi_t}} \quad (9)$$

Similarly, if we consider M4:

$$I_- = I_o e^{\frac{\kappa(V_- - V_s)}{\varphi_t}} \quad (10)$$

By using KCL:

$$\begin{aligned} I_+ + I_- &= I_b \\ I_+ - I_- &= I_{out} \end{aligned} \quad (11)$$

Combining (9), (10), and (11):

$$\frac{I_{out}}{I_b} = \frac{I_o e^{\frac{\kappa(V_+ - V_s)}{\varphi_t}} - I_o e^{\frac{\kappa(V_- - V_s)}{\varphi_t}}}{I_o e^{\frac{\kappa(V_+ - V_s)}{\varphi_t}} + I_o e^{\frac{\kappa(V_- - V_s)}{\varphi_t}}}$$

$$\frac{I_{out}}{I_b} = \frac{e^{\frac{\kappa(V_+ - V_s)}{\varphi_t}} - e^{\frac{\kappa(V_- - V_s)}{\varphi_t}}}{e^{\frac{\kappa(V_+ - V_s)}{\varphi_t}} + e^{\frac{\kappa(V_- - V_s)}{\varphi_t}}}$$

$$I_{out} = I_b \frac{e^{\frac{-\kappa V_s}{\varphi_t}} \left( e^{\frac{\kappa V_+}{\varphi_t}} - e^{\frac{\kappa V_-}{\varphi_t}} \right)}{e^{\frac{-\kappa V_s}{\varphi_t}} \left( e^{\frac{\kappa V_+}{\varphi_t}} + e^{\frac{\kappa V_-}{\varphi_t}} \right)}$$

$$I_{out} = I_b \frac{e^{\frac{\kappa(2V_+ - V_- + V_-)}{2\varphi_t}} - e^{\frac{\kappa(2V_- + V_+ - V_+)}{2\varphi_t}}}{e^{\frac{\kappa(2V_+ - V_- + V_-)}{2\varphi_t}} + e^{\frac{\kappa(2V_- + V_+ - V_+)}{2\varphi_t}}}$$

$$I_{out} = I_b \frac{e^{\frac{\kappa(V_+ + V_-)}{2\varphi_t}} \left( e^{\frac{\kappa(V_+ - V_-)}{2\varphi_t}} - e^{\frac{\kappa(V_- - V_+)}{2\varphi_t}} \right)}{e^{\frac{\kappa(V_+ + V_-)}{2\varphi_t}} \left( e^{\frac{\kappa(V_+ - V_-)}{2\varphi_t}} + e^{\frac{\kappa(V_- - V_+)}{2\varphi_t}} \right)}$$

$$I_{out} = I_b \frac{e^{\frac{\kappa(V_+ - V_-)}{2\phi_t}} - e^{-\frac{\kappa(V_- - V_+)}{2\phi_t}}}{e^{\frac{\kappa(V_+ - V_-)}{2\phi_t}} + e^{\frac{\kappa(V_- - V_+)}{2\phi_t}}}$$

$$I_{out} = I_b \tanh\left(\frac{\kappa(V_+ - V_-)}{2\phi_t}\right) \quad (12)$$

Notice that even though the overall I-V relationship is nonlinear, for a certain voltage range, the I-V relationship is quite linear. This voltage range is called the linear range of the transconductance amplifier. The linear range  $V_L$  is defined as:

$$V_L = \frac{2\phi_t}{\kappa}$$

To obtain an effective linear V-I converter, we need a wide linear range design. Past research has led to the development of such wide linear range transconductance amplifiers [33, 34]. The I-V relationship is still the same but the linear range has increased due to changes in the effective value of  $\kappa$ . The schematic for such an amplifier is shown below in Figure 3-7. You can clearly see the effects of having such a wide linear range design. For an ordinary, narrow range design the linear range is limited to 75 mV, hence we can only have voltage inputs of that magnitude. On the other hand, for wide range designs, we can have voltage inputs of 2 volts or more, which gives us more flexibility in circuit design. This difference is illustrated in Figure 3-8.

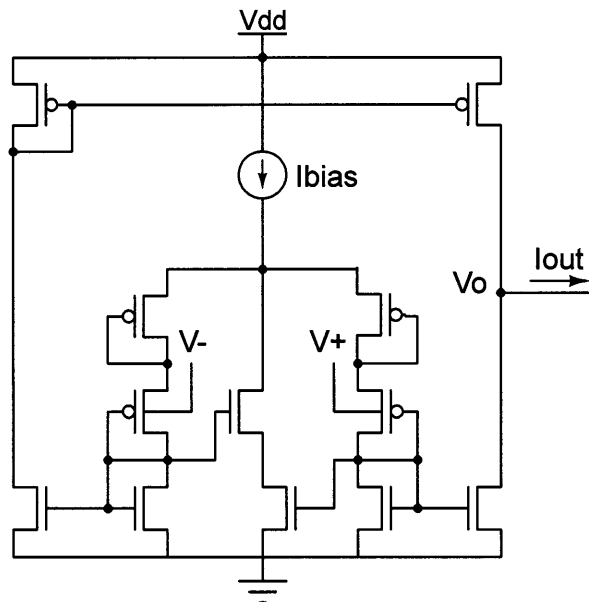


Figure 3-7. Schematic of a wide linear range transconductance amplifier.

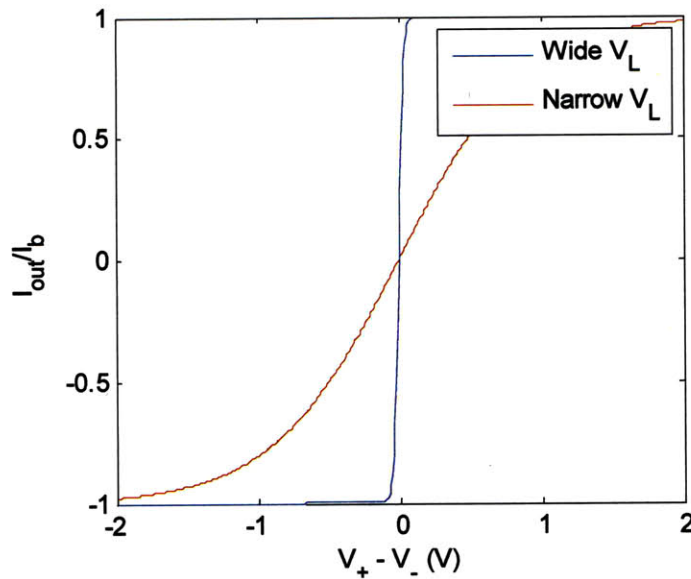


Figure 3-8. I-V relationship of the ordinary and wide linear range transconductance amplifier.

Transconductance amplifiers can be used as comparators as well. If  $V_+ > V_-$ ,  $V_o$  becomes low. On the other hand if  $V_- > V_+$ ,  $V_o$  becomes high. So it is quite common to see transconductance amplifiers used as cheap comparators because of its simplicity especially if switching voltage accuracy is not a big design issue.

One last thing to note,  $V_+$  and  $V_-$  must have values of between roughly 1.5V and  $V_{dd} - 0.8V$  to ensure that all our transistors remain saturated. Hence, it is traditional practice to append a common mode voltage  $V_{cm}$  on top of both  $V_+$  and  $V_-$ .

### 3.3.4. Differential pair

The use of differential pairs in neuromorphic circuit design is very common because of its intrinsic sigmoidal I-V relationship which is present in most neuronal networks. For instance, NMDA receptors in the synapse have a sigmoidal I-V dependence relationship such as for increasingly negative differential voltage, the current output approaches zero while for increasingly positive differential voltage, the current output approaches the bias current. If the differential voltage is zero, the current is half the bias current. Again to obtain a wide sigmoidal range, we want to use a differential pair that has a wide linear range [22]. The design shown in Figure 3-9 is one example of a wide linear range circuit.



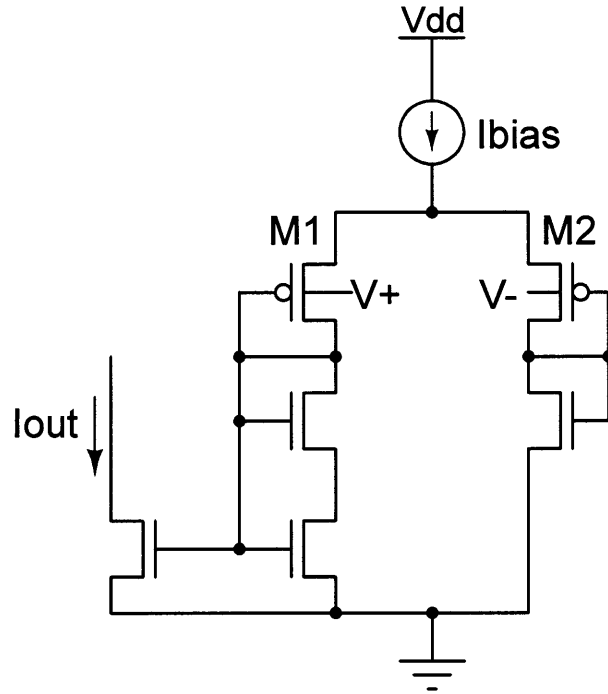


Figure 3-9. Schematic of a wide linear range differential pair.

The exact transfer function between  $I_{out}$ ,  $V_+$ , and  $V_-$  is:

$$I_{out} = I_{bias} \frac{1}{1 + \frac{\frac{W_1}{L_1}}{\frac{W_2}{L_2}} e^{\frac{\kappa(V_+ - V_-)}{\phi_t}}} \quad (13)$$

Where,  $W_1/L_1$ , and  $W_2/L_2$  are the widths to length ratios of M1 and M2 respectively.

### 3.4. Summary

The circuits that we described above are simple building blocks to construct a complete synapse that exhibit plasticity. We would exploit the laws of physics to mimic real life neurons which are also governed by similar principles. As you notice some designs have narrow linear range while some have wide linear range. We would decide depending on the application of such a device whether we need a narrow or wide range design. Chapter 4 will illustrate our synapse system that takes advantage of all these MOSFET building blocks.



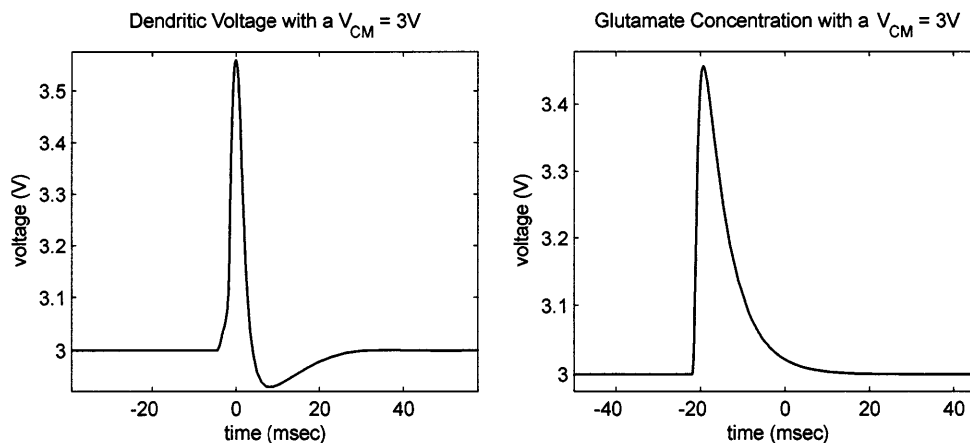
## 4. Development of a low power aVLSI synapse

For almost a decade, since the phenomenon of synaptic plasticity was discovered, neuromorphic engineers have attempted to create circuit network models of synapses. Every year, new neuromorphic circuits have popped up claiming to be better than previous models in several aspects: power, stability, memory retention, and network scalability [35-37]. Some of their claims are true but they all fail to capture the most essential aspect of synaptic plasticity – its behavior.

Some of these models are inaccurate and does not reproduce experimental results. As an example, some do not even reproduce the correct exponential-like shape of the synaptic change function in hebbian STDP [37]. Others are just too simplistic and merely employ curve fitting methods to reproduce the STDP curves [38]. None of the previous neuromorphic implementations capture all known types of STDP which is a very important issue since their technology can only be used in special circumstances and cannot be used as a general way to form intelligent machines. Furthermore, these circuits were not able to relate calcium based synaptic plasticity to STDP – in other words, no other system was able to show both the BCM learning rule and STDP. In short, the existing synaptic implementations are neither accurate nor complete. Our research will aim to ameliorate these issues and here we shall present our approach to develop a robust and versatile circuit model that will unify the various forms of STDP as well as the BCM rule.

### 4.1. Inputs

Synaptic plasticity is influenced by two essential inputs – the glutamate ligand concentration and the back propagating dendritic action potential. Lee's unified theory suggests that the timing of these two inputs is essential in capturing STDP [3], thus, in this regard, we have to ensure that our implementation reflects how the EPSC level is modulated by the timing of these two inputs. Hence, we would need two external circuit blocks that would emulate the behavior and timing of these two inputs.



**Figure 4-1. The inputs of the synapse, glutamate and dendritic action potential.**

Both the glutamate and dendritic signals are time varying continuous signals that have the shapes shown in Figure 4-1. To reproduce these types of shapes, we employ

spiking circuits such as the one shown in the schematic in Figure 4-2, which are similar to existing spiking systems [22, 39]. The circuit in Figure 4-2 emits a spike or a series of spikes in the voltage node,  $V_{mem}$ , based on the current stimuli  $I_{input}$  that we inject. The heart of the circuit is the transconductance amplifier formed by transistors M1-M4. This transconductance amplifier is used as a comparator such that when the voltage  $V_{mem}$  rises above  $V_{thresh}$ ,  $V_x$  decreases, turning off transistor M5 which causes the current source  $I_{Na}$  to charge  $C_{mem}$  to a higher voltage  $V_{mem}$ .  $V_x$  also charges the feedback capacitor,  $C_{feedback}$  such that when its voltage  $V_{feedback}$  rises above a certain threshold, transistor M6 turns on causing the current source  $I_K$  to discharge  $C_{mem}$  back to its initial resting voltage  $V_{rest}$ . This last event forces  $V_x$  to return to its resting value of around  $V_{dd}$  which turns on M8, discharging  $C_{feedback}$  and thereby turn off M6. Overall, this results in a rapid up-down spike shown below in Figure 4-3.

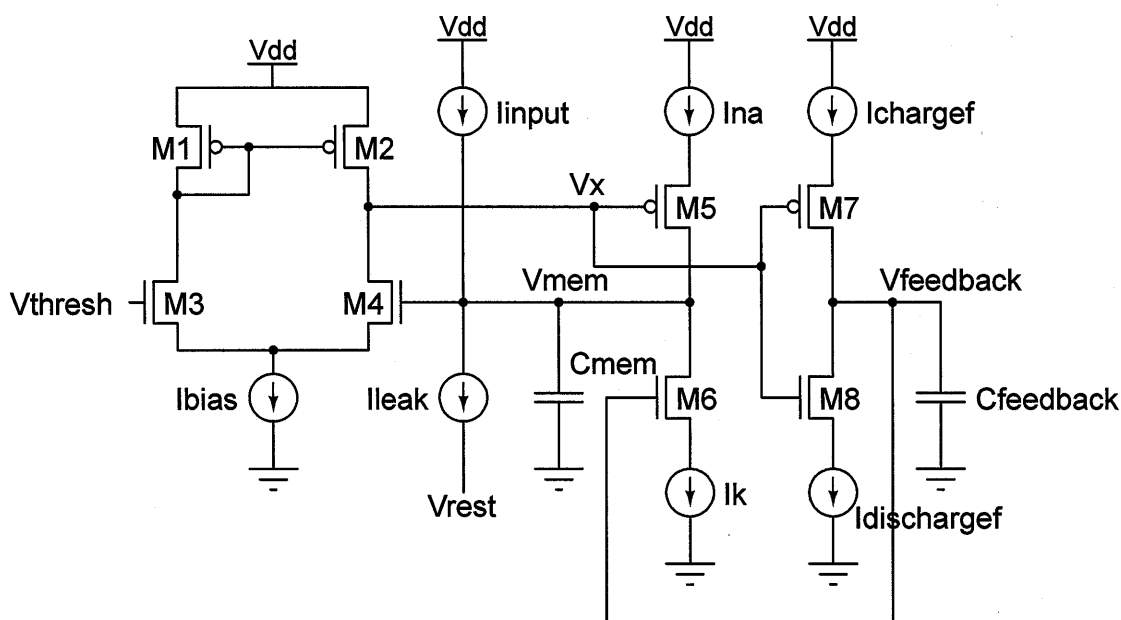


Figure 4-2. Schematic of a basic spiking circuit.

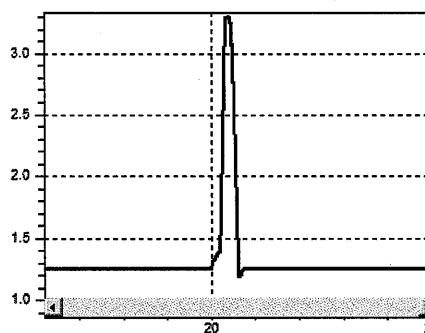


Figure 4-3. The output spike generated by the spiking circuit.

## 4.2. Outputs

Our synapse has two measurable outputs. The main output of our system is the synaptic weight which is represented by the excitatory post synaptic current (EPSC). The secondary output used to modulate the EPSC is the total calcium level ( $Ca_{total}$ ) of the synapse. Our system determines the  $Ca_{total}$  from our two signal inputs and maps this value to the proper EPSC as determined by the BCM learning rule.

## 4.3. Synaptic plasticity system design

Two options were considered in developing our system. One option is to use voltage centered design while the other option is to use current mode design. Existing circuit technology makes it much easier and simpler to design the system using a current mode approach. A lot of neuromorphic arithmetic favors transfer functions that are inherent in several current mode design systems. This ultimately led us to reject voltage mode design and stick to mostly current mode implementation. Figure 4-4 illustrates the overall block diagram for our system.

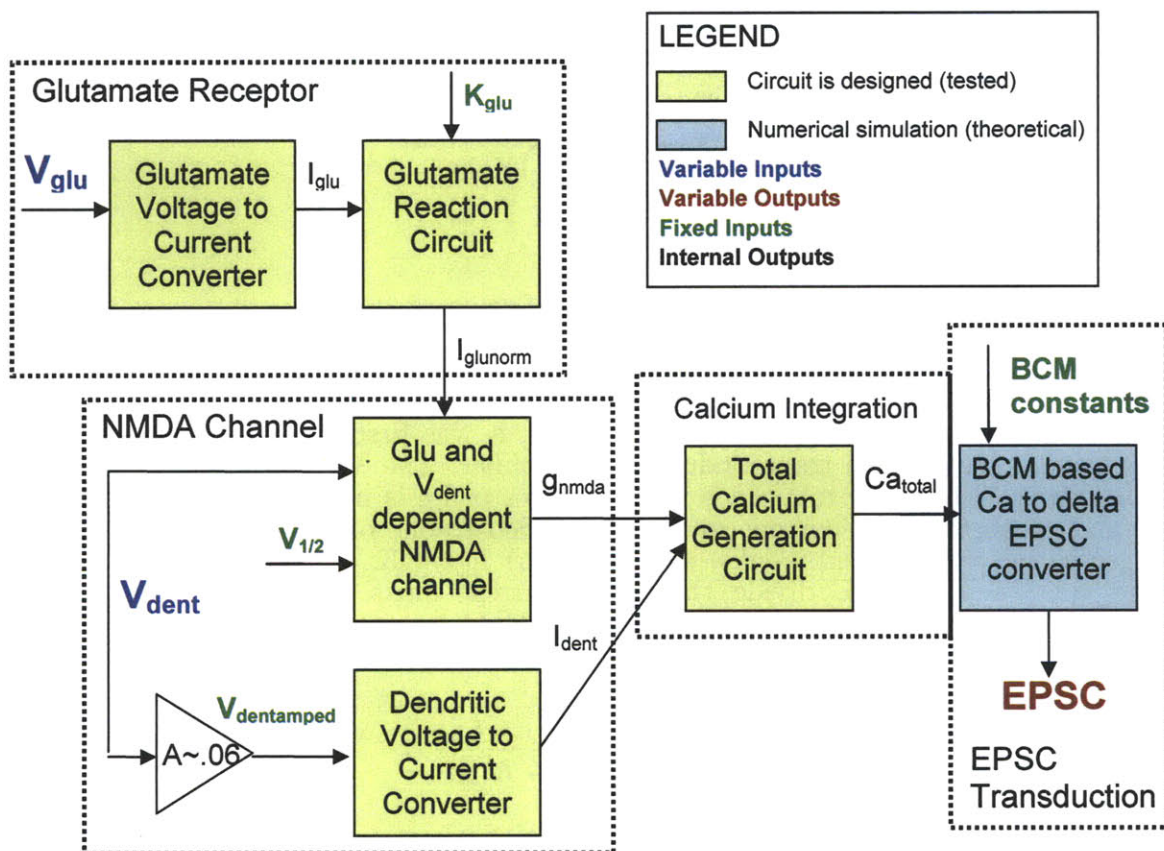


Figure 4-4. Block diagram for the aVLSI circuit design of a synapse.

### 4.3.1. Glutamate receptor dynamics

Glutamate ligand concentration as a function of time,  $V_{glu}$ , combines with glutamate receptors in the postsynaptic membrane with an association constant of  $K_{glu}$ . Ultimately, in equilibrium, the ratio of bound glutamate to the total glutamate concentration is given by:

$$\frac{Glu}{Glu + K_{glu}}$$

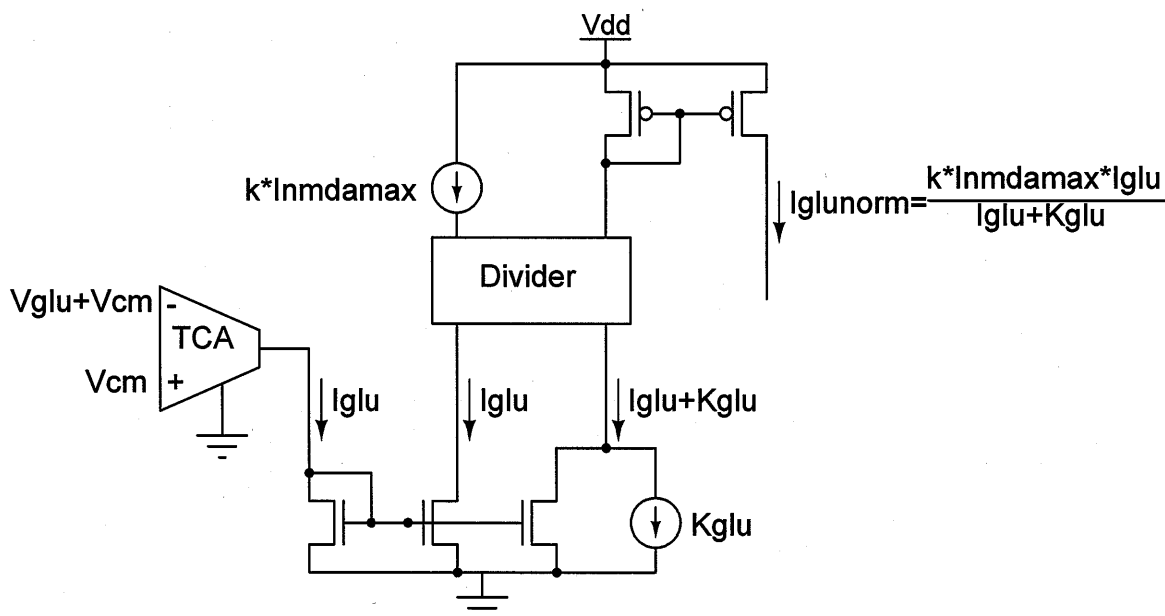


Figure 4-5. Schematic of the glutamate receptor circuit.

The equation above is a simple arithmetic equation that involves addition and division. Using aVLSI, we could implement this using a current adder and a divider. However, since our input  $V_{glu}$  is a voltage signal, it was first converted to a current signal using a wide range transconductance amplifier. The V-I conversion rate that was used was  $10 \text{ mV} = 1 \text{ nA}$ . Since two instances of  $I_{glu}$  is needed,  $I_{glu}$  was mirrored twice and a constant  $K_{glu}$  current source was added to one of the mirrored instances of  $I_{glu}$ , creating two current sinks: a)  $I_{glu}$  and b)  $I_{glu} + K_{glu}$ . A translinear current divider was then used to divide these two current sinks to generate an output current signal that represents the maximum NMDA receptor conductance. Hence, the overall transfer function of our glutamate receptor is given by:

$$I_{glunorm} = \frac{k I_{NMDAmax} I_{glu}}{I_{glu} + K_{glu}}$$

Where  $I_{NMDAmax} = 1 \text{ nA}$  and  $k = 10$ .  $I_{glunorm}$  is a current sink so it is mirrored to transform it to a current source to ensure that it segues seamlessly with the next circuit block. Figure 4-5 depicts the schematic of the circuit that we have developed for this first stage. The output current of this stage is qualitatively equal to what the model predicts except with a constant gain factor of  $k$  and  $I_{NMDAmax}$ . This is not an

issue because  $I_{\text{NMDAmax}}$  is used in the NMDA receptor circuit. The gain term is not a big deal as long as we keep track of the value of  $k$ .

### 4.3.2. NMDA channel dynamics

A wide-range differential pair is the main component used to represent the dynamics of the NMDA channel. If we recall, the model for NMDA channel dynamics states that the conductance of the NMDA channel is dependent to both the dendritic voltage and glutamate concentration and this dependence is given by the equation:

$$g_{\text{NMDA}}(V_{\text{dent}}, \text{Glu}) = \frac{g_{\text{NMDAmax}}}{1 + e^{k_{\text{NMDA}}(V_{1/2} - V_{\text{dent}})}} \frac{\text{Glu}}{\text{Glu} + K_{\text{glu}}}$$

This complicated looking sigmoidal equation seems challenging to implement in aVLSI at first glance. However, an analog circuit that naturally expresses this type of behavior already exists in the form of the differential pair whose I-V characteristic is quite similar to the NMDA channel conductance dynamics as shown below in this equation:

$$I_{\text{out}} = I_{\text{bias}} \frac{1}{1 + \frac{\frac{W_1}{L_1}}{\frac{W_2}{L_2}} e^{\frac{\kappa(V_+ - V_-)}{\phi_t}}}$$

Certainly, the differential pair can be used to obtain  $g_{\text{NMDA}}$  but caution was exercised to ensure that the NMDA was accurately mapped with the differential pair. First of all, the bias current  $I_{\text{bias}}$  was represented by  $I_{\text{glunorm}}$ , the output current of the glutamate receptor circuit. Secondly, the transistor geometry terms were normalized to 1 by making the transistor sizes equal. Thirdly, we mapped  $V_+$  to  $V_{1/2}$  and  $V_-$  to  $V_{\text{dent}}$ .

Mapping the  $\kappa/\phi_t$  factor to  $k_{\text{NMDA}}$  is very tricky because their values are quite different. Naively mapping the two directly without considering the scaling error between the two values would result in an incorrect emulation of the NMDA channel dynamics (for example, the dynamics might become too fast or too slow). We analyze this difference by solve the value of  $\kappa$  in a wide linear range differential pair using this equation [22]:

$$\kappa = \frac{1 - \kappa_p}{1 + \frac{\kappa_n}{\kappa_p}}$$

$\kappa_p$  and  $\kappa_n$  represent the coupling coefficients of the PMOS and NMOS in the circuit. These values are process dependent and typically fall between 0.5 and 1 and are usually approximated as 0.7, which gives us the approximate working value of  $\kappa$  of 0.15. Thus,  $\kappa/\phi_t$  has a value of about  $6 \text{ V}^{-1}$  at room temperature. In comparison,

$k_{NMDA}$  is about  $100 \text{ V}^{-1}$ , hence, to ensure that  $g_{NMDA}$  is calculated correctly,  $V_{dent}$  and  $V_{1/2}$  needs to be scaled up by a factor of about 16.7. Common amplifying techniques were used to implement a scaled up input version of  $V_{dent}$ , a signal we call  $V_{dentamped}$ .

Another main component of our NMDA circuit is the conversion of our scaled dendritic voltage,  $V_{dentamped}$  to the current,  $I_{dent}$ . We can implement this by simply using a transconductance amplifier with an I-V ratio of 1 mV is to 1 nA. This completes the necessary temporal inputs for our next circuit block, the calcium integrator. Figure 4-6 below shows the schematic of the NMDA channel circuit and the equation below summarizes the mathematical equivalent of  $g_{NMDA}$ .

$$kg_{NMDA} = \frac{kg_{NMDAmax}}{1 + e^{\frac{\kappa(V_{1/2} - V_{dent})}{\phi}}} \frac{I_{Glu}}{I_{Glu} + K_{glu}}$$

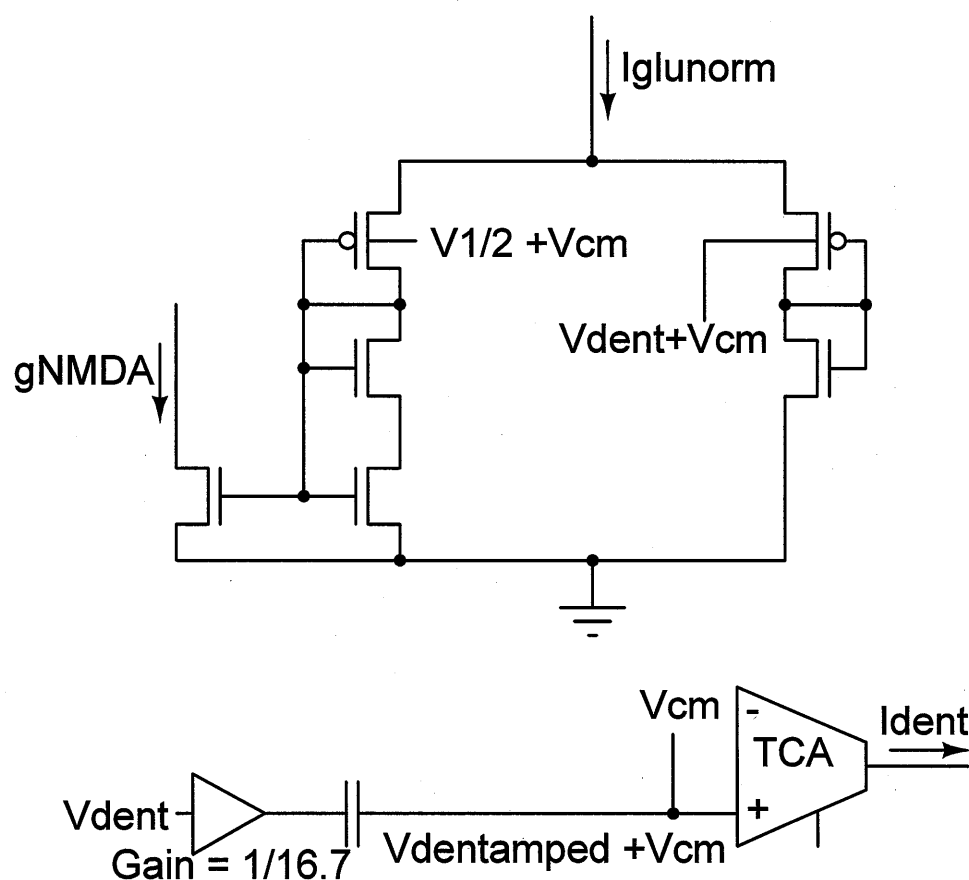


Figure 4-6 Schematic of the full NMDA channel dynamics circuit.

### 4.3.3. Calcium integrator dynamics

Calcium integration determines how much calcium is present in our synapse by integrating the product of our NMDA conductance and the incoming dendritic voltage. This relation is captured by this equation from Lee's model:



$$Ca_{total} = \int_{t_{start}}^{t_{end}} g_{NMDA} V_{dent} dt$$

$t_{start}$  refers to the time at which the first stimulus arrives and  $t_{end}$  refers to the time at which the final stimulus ends. Notice the the multiplication between the two temporal signals  $g_{NMDA}$  and  $V_{dent}$ . Again, similar to the glutamate receptor circuit, we implement a four quadrant multiplier because  $V_{dent}$  can be a biphasic signal ( $g_{NMDA}$  is monophasic since conductance cannot be negative).

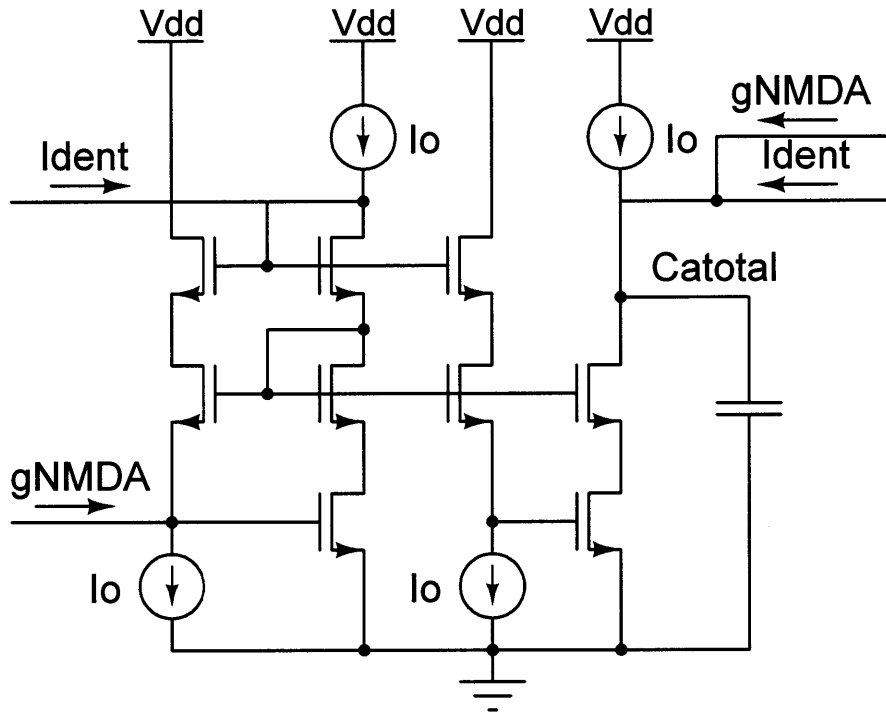


Figure 4-7. Schematic of the calcium integrator circuit.

Integration is the second operation in this system and is very trivial to implement for current and time integration can be easily achieved by using a capacitor. The total current,  $Ca_{total}$  is represented by the accumulated voltage in the capacitor within the stimuli time window. The schematic in Figure 4-7 summarizes what the circuit described above. The overall transfer function of the circuit implementation of  $Ca_{total}$  is given by:

$$Ca_{total} = \int_{t_{start}}^{t_{end}} \frac{kg_{NMDA} I_{dent}}{C_x} dt$$

#### 4.3.4. Mechanistic design of the calcium to EPSC transduction circuit

Here, we propose a mechanistic design for the calcium to EPSC transducer. First, we recall Lee's model for EPSC current generation. His model introduces two equations that govern the changes in EPSC:

$$R_{CaK} = \frac{Ca_{total}^4}{Ca_{total}^4 + K_{CaK}^4} \quad R_{CaP} = \frac{Ca_{total}^4}{Ca_{total}^4 + K_{CaP}^4}$$

$$\Delta W = \alpha \left( \frac{k_{AMPArf} + V_{CaK} R_{CaK}}{k_{AMPARb} + V_{CaP} R_{CaP} + k_{AMPArf} + V_{CaK} R_{CaK}} - \frac{k_{AMPArf}}{k_{AMPARb} + k_{AMPArf}} \right) \quad (14)$$

$R_{CaK}$  represents the binding ratio of  $Ca_{total}$  with kinase while  $R_{CaP}$  represents the binding ratio of  $Ca_{total}$  with phosphatase. The rest of the terms are simply constants that are specific to each type of STDP.  $\Delta W$  represents the amount at which the EPSC changes after each stimulus.

The equation above suggests that the we first need to create powers of the voltage signal  $Ca_{total}$ . Current mode multiplication easily achieves this goal, but to do this, voltage  $Ca_{total}$  must be first converted to a current  $I_{CaT}$ , which was done using a wide linear range transconductance amplifier. Two copies of  $I_{CaT}$  was produced using a current mirror and these copies served as inputs to a current multiplier to give us  $I_{CaT}^2$ . Again, mirroring and squaring this current again gives us  $I_{CaT}^4$ .

Four instances of this current signal,  $I_{CaT}^4$  were generated using current mirrors. These four signals were sent to two current dividers and the corresponding constant currents,  $K_{CaK}^4$  and  $K_{CaP}^4$  were added to two of these instances to obtain the  $R_{CaK}$  and  $R_{CaP}$  currents. Next,  $R_{CaK}$  and  $R_{CaP}$  were multiplied by two constants  $V_{CaK}$  and  $V_{CaP}$ , respectively, using the scaling capability of the current mirror (remember M2/M1 ratio from equation 6). Again analog arithmetic blocks were used to implement equation (14) to the several constant currents and our signal currents. Finally, using a current mirror, the resulting current was scaled by the factor  $\alpha$  which is dependent to the type of STDP that was being emulated resulting to a final output current of  $\Delta W$ .

The proposed circuit schematic that implements the mechanistic approach above is illustrated below in Figure 4-8. For the exact values of the constants used, please refer to Lee's manuscript [3].

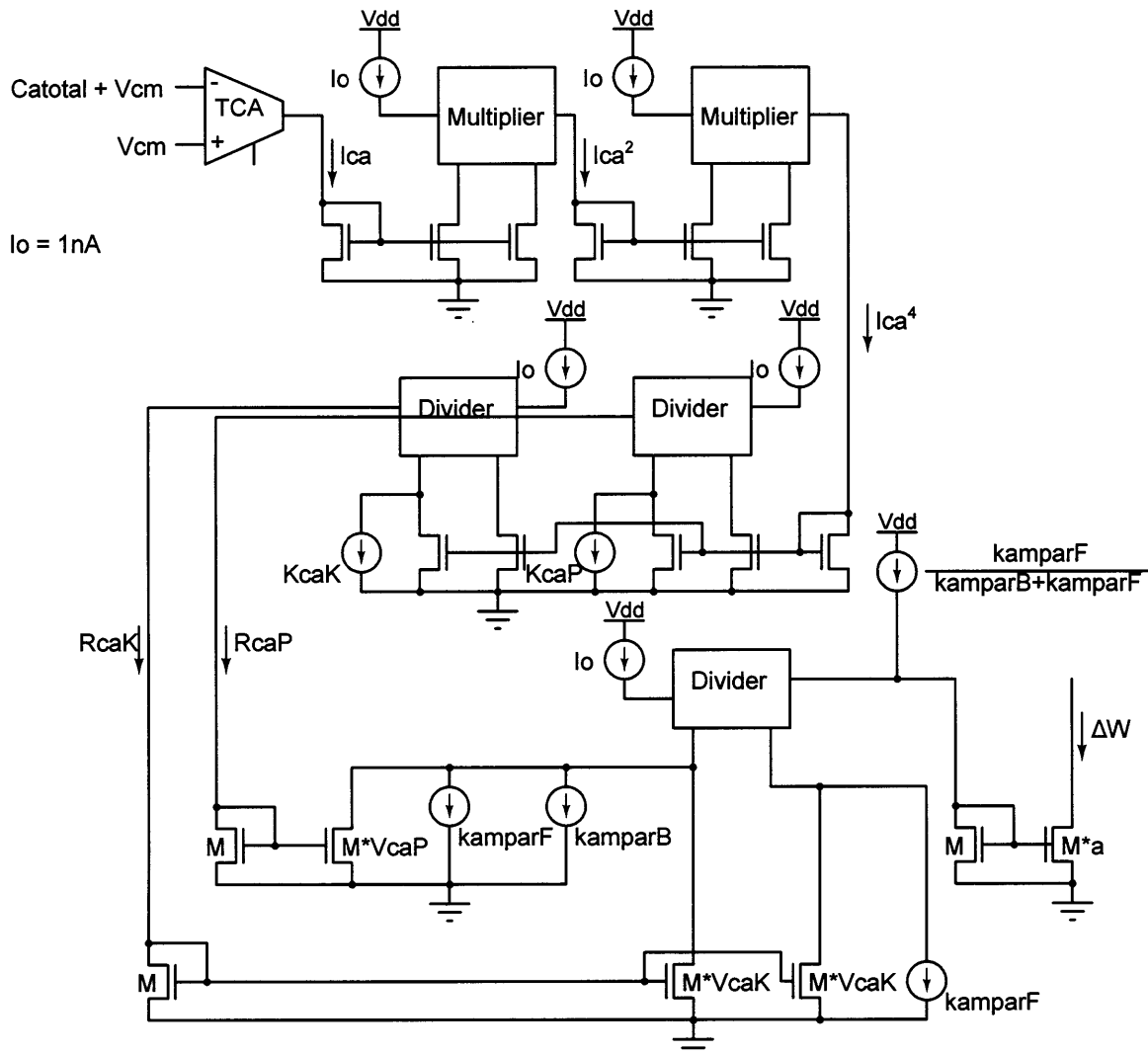


Figure 4-8. Proposed schematic for the calcium to EPSC transducer.

#### 4.3.5. Empirical design of the calcium to EPSC transduction circuit

The mechanistic implementation of the EPSC transduction circuit is the most accurate representation of the biophysical processes within the synapse. However, this accuracy compromises the circuit's size and complexity causing our implementation to become extremely bulky and prone to mismatch and leakage problems, making it more difficult to test and debug in a chip implementation. This leads to an alternative and much simpler empirical design which saves area, power, and complexity.

Understanding the circuit's input-output transfer function is the most essential thing to do. This was achieved by plotting equation (14) as a function of its input,  $Ca_{total}$  which results to a famous curve known as the BCM rule curve [11] as illustrated in Figure 4-9. Looking at the curve, it seems that it could be easily decomposed as a sum of multiple sigmoidal functions, which is reminiscent to the Fourier series decomposition where a signal is broken down into multiple sine and cosine terms. Here, instead of using sine and cosine functions, sigmoidal functions

were employed because of the characteristic shape of the curve. We observed that the BCM curve is composed of a fast decreasing segment followed by a slow increasing segment that can be minimally represented by a sum of two sigmoidal curves.

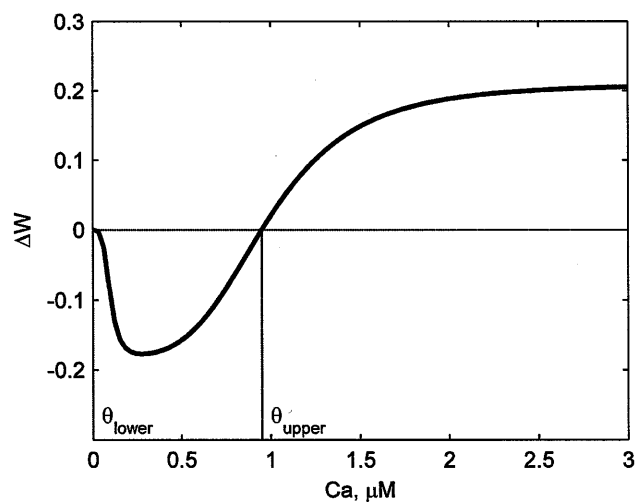


Figure 4-9. The BCM rule curve. *Figure is obtained from [3].*

### Biophysical relevance of sigmoidal decomposition

Sigmoidal decomposition of the BCM curve corresponds to the breakdown of the difference in the chemical kinetics of the binding between calcium to phosphatase and kinase. Phosphatase kinetics is faster and dominates in low  $Ca_{total}$  concentrations and corresponds to a decrease in EPSC current from the AMPA channels. On the other hand, kinase kinetics is slightly delayed and slower but has the direct opposite effect of increasing the EPSC current from the AMPA channels in higher  $Ca_{total}$  concentrations. Hence, the phosphatase and kinase kinetics then maps respectively to the decreasing and increasing sigmoids. To obtain the best fit shapes of the both individual sigmoidal curves, we used a least error method algorithm. Once these curves were obtained, we summed them up to produces the BCM output curve shown in Figure 4-10a and 4-10b.

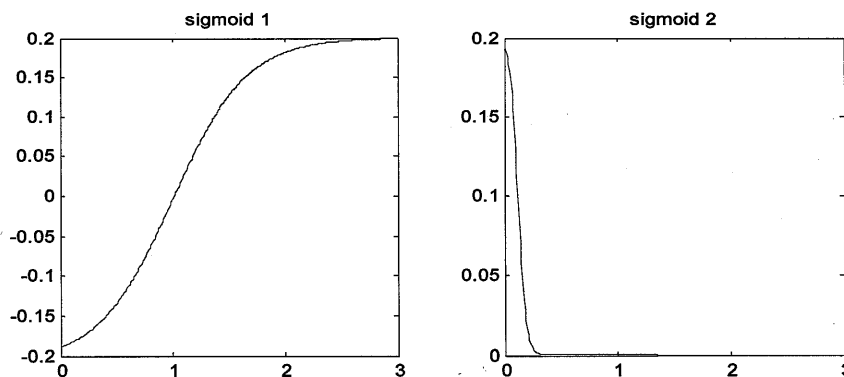


Figure 4-10a. Sigmoidal decomposition of the BCM curve. The two sigmoids that are used to generate the BCM curve.

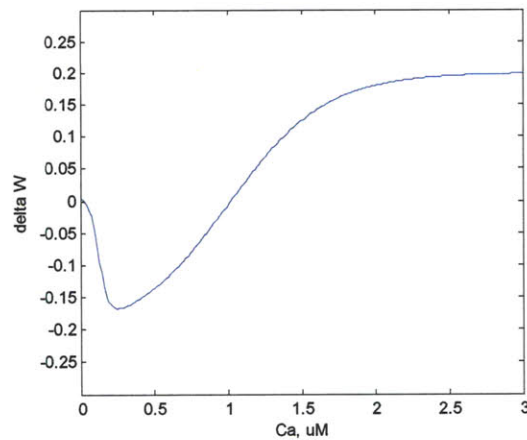


Figure 4-10b. Resulting BCM curve from the sigmoidal decomposition.

### Hardware implementation of sigmoidal decomposition

Sigmoidal decomposition simplifies the needed hardware for implementing the transduction circuit. Hyperbolic tangent functions like equation (12) accurately represents a sigmoidal curve implying that two transconductance amplifiers can be used to generate the two sigmoidal curves in Figure 4-10. The current outputs of these two amplifiers are added together by connecting them to a single node resulting to an output current,  $I_{out}$ , that represents the instantaneous value of the change in EPSC,  $\Delta W$ . The first amplifier has a narrow dynamic range to simulate the fast decreasing segment of the function. This amplifier is shown in the left arm of Figure 4-11. Notice that, the voltage inputs were obtained from the gate of the middle transistors eliminating the increase in linear range obtained from gate degeneration in standard wide linear range amplifiers [34]. To finely tune the linear range we used bump linearization transistors. On the other hand the second amplifier was designed to have a wider dynamic range to emulate the slow increasing segment of the BCM curve. To do this, we used our old wide linear range design (Figure 3-7), only this time we removed the bump linearization transistors to slightly tune down the linear range. A schematic of our alternative implementation of the EPSC transduction system is shown in Figure 4-11.

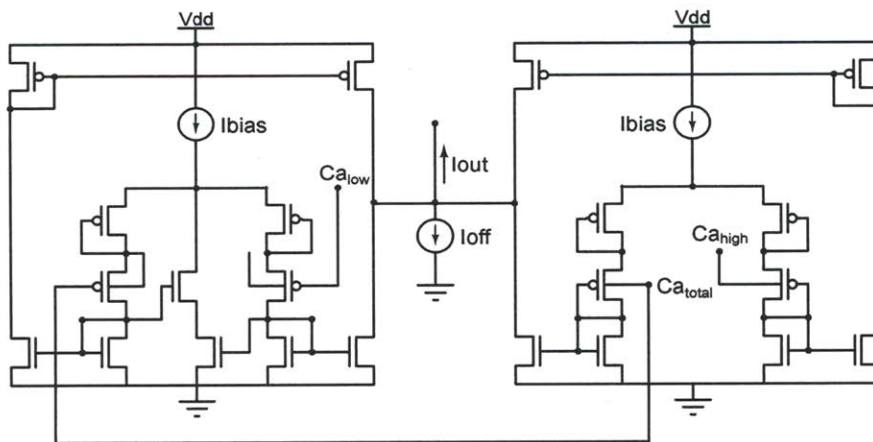


Figure 4-11. Schematic of the alternative design for the EPSC transduction circuit.

Our design includes 4 tunable constant inputs to facilitate for minor and common changes to the shape of the BCM curve. The first current input,  $I_{\text{bias}}$ , sets the maxima and minima of the BCM curve, which represents the maximum amount of potentiation and depression of the synapse.  $I_{\text{off}}$ , on the other hand, represents the current that sets the vertical offset of the BCM curve. Changing this current shifts the BCM curve **up or down**. The two voltage inputs,  $Ca_{\text{low}}$  and  $Ca_{\text{high}}$  set the zero crossings of the BCM curve. Modifying these two values shifts the BCM curve **left or right**. The last input is the variable voltage input,  $Ca_{\text{total}}$ , which represents the instantaneous calcium voltage that is transduced to an appropriate change in EPSC.

#### 4.4. Test and results

Our synaptic system was designed using Tanner tools' S-Edit and the system's performance was tested using T-Spice. To benchmark the performance of our circuit, its output synaptic plasticity behavior was compared with the behavior of Lee's model. Synaptic plasticity was induced by sweeping the two inputs,  $V_{\text{dent}}$  and  $V_{\text{glu}}$ , with varying time offsets,  $\Delta T$ . A positive  $\Delta T$  implies that a dendritic action potential that comes after the glutamate stimulus while a negative  $\Delta T$  signifies that a dendritic action potential arrived before the glutamate stimulus. Lee's model predicted five different types of STDP, hebbian, anti-hebbian, mixed LTP and LTD, STP and STD [3] and our design was tested for all types by modulating the shape of  $V_{\text{dent}}$  input for Lee's hypothesis suggests that the different shapes of  $V_{\text{dent}}$  lead to different types of STDP.

One of the two essential outputs of our system is the voltage  $Ca_{\text{total}}$  that has accumulated within the stimuli period. As expected, this value varies as a function of  $\Delta T$  like the model predicts.  $Ca_{\text{total}}$  was then used as an input to our mechanistic EPSC transducer circuit to obtain the change in EPSC amplitude,  $\Delta W$ .  $\Delta W$ , perhaps the most important output of circuit, represents the degree of potentiation and depression of the synaptic system. Both  $Ca_{\text{total}}$  and  $\Delta W$  were obtained and plotted for several values of  $\Delta T$  for each of the different types of STDP and their plots are summarized in Figures 4-13 to 4-17.

We also tested our empirical implementation of the EPSC transducer to ensure that it matches well with the BCM curve. Our first test involves sweeping the  $Ca_{\text{total}}$  input of the empirical EPSC transducer system within a range of physiological values while measuring the EPSC output generated keeping the other four constant inputs to their nominal values. For the second test, we sweep  $Ca_{\text{total}}$  again, but this time we scaled in half the values of  $I_{\text{bias}}$  and  $I_{\text{off}}$  to produce a vertically scaled down version of the BCM curve. Lastly, for the third test, we repeated the conditions of the second test, but we increased the values of  $Ca_{\text{low}}$  and  $Ca_{\text{high}}$  by one volt to shift the curve to the right. The inputs that we used are summarized in Table 4-1 and the output curves of the three tests are shown in Figure 4-12.

Table 4-1. Input constants for the EPSC transductor tests.

Inputs	Test 1 - Normal	Test 2 - Scale down	Test 3- Scale down and right shift
$Ca_{\text{low}}$	0 V	0 V	1 V
$Ca_{\text{high}}$	1 V	1 V	2 V
$I_{\text{bias}}$	40 nA	20 nA	20 nA
$I_{\text{off}}$	32 nA	16 nA	16 nA

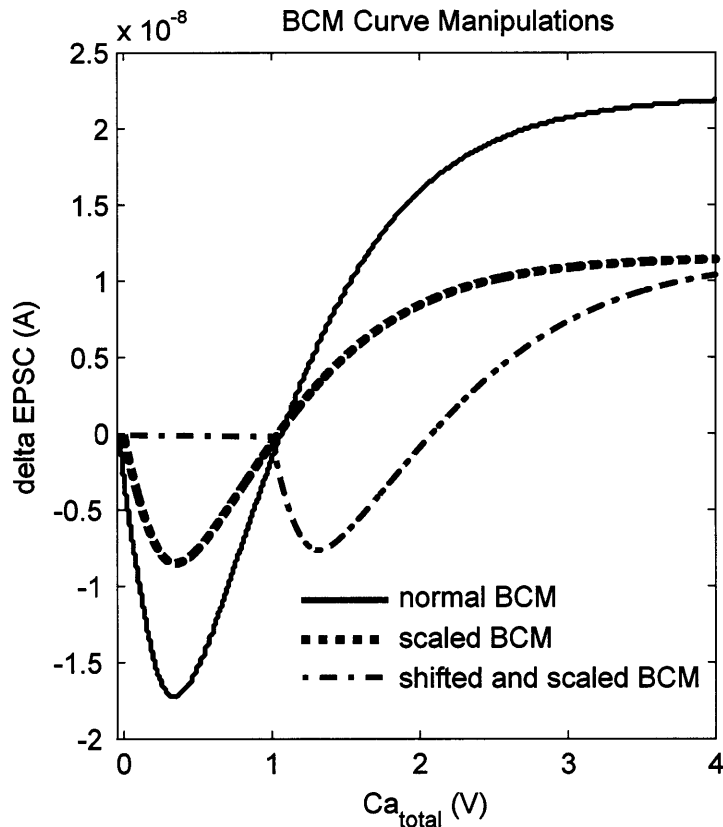


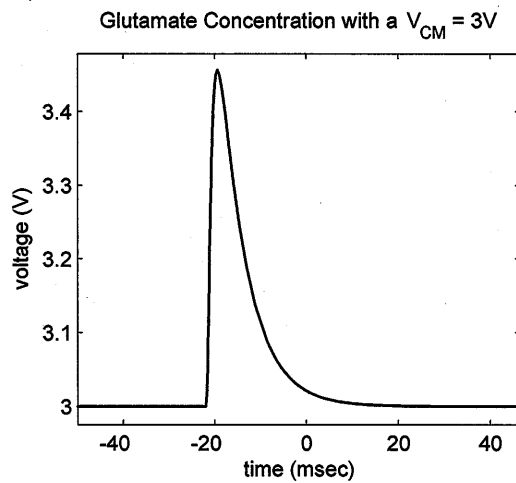
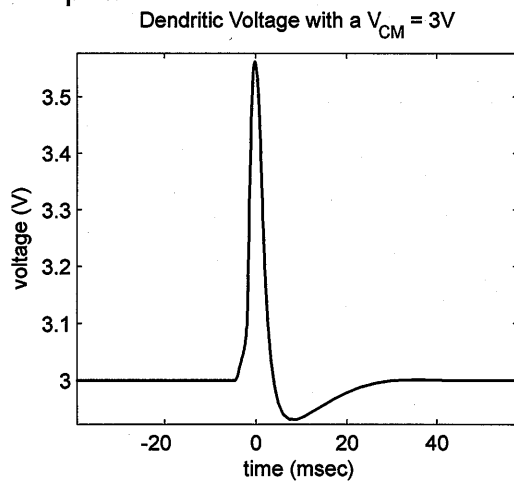
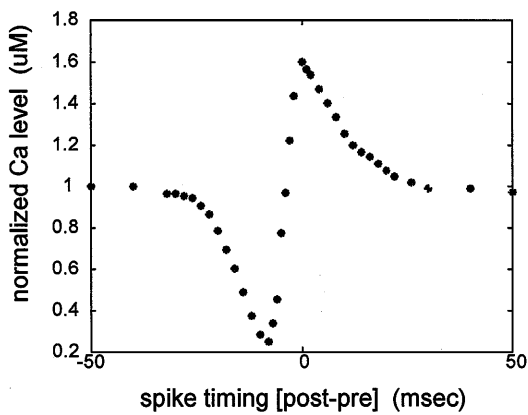
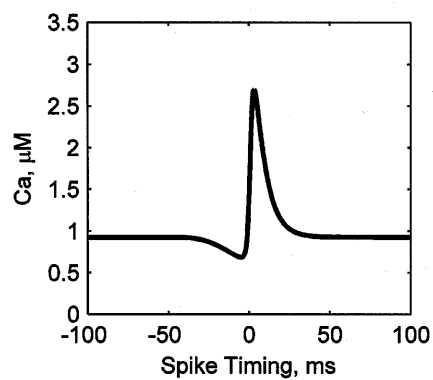
Figure 4-12. BCM curve outputs from manipulations of the empirical EPSC transduction circuit.

To further characterize the performance of our circuit, its power consumption was also determined and it measures about **1.98  $\mu$ W** when the circuit is idle and **2.48  $\mu$ W** when the circuit is active (processing incoming signals). The measurements above were done using a 5 volt power supply. Using a lower power supply voltage of 3.3 volts reduces the power consumption to around **1.3  $\mu$ W – 1.63  $\mu$ W**. To complete our benchmark tests, an estimation of the area of the chip was also obtained using layout design rules from Tanner Tool's layout editor software. We estimate the chip size to be about **1.2 mm x 0.8 mm** using a half micron MOSIS technology. These findings are summarized in Table 4-2 below.

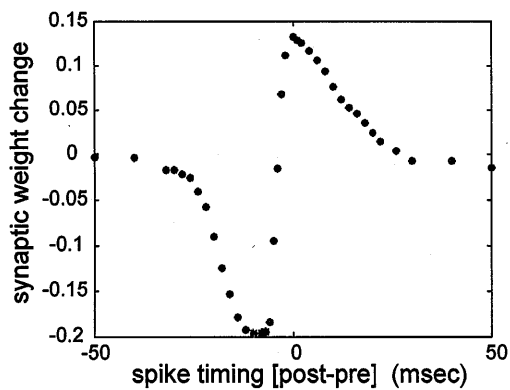
Table 4-2. Power and area estimates of the STDP system.

Power Consumption (Vdd = 5 volts)	1.98 – 2.48 $\mu$ W
Power Consumption (Vdd = 3.3 volts)	1.3 – 1.63 $\mu$ W
Chip Area	1.2 mm x 0.8 mm

## a. Inputs

b. Normalized  $Ca_{total}$ c. Theoretical Normalized  $Ca_{total}$ 

## d. EPSC Amplitude Change



## e. Theoretical EPSC Amplitude Change

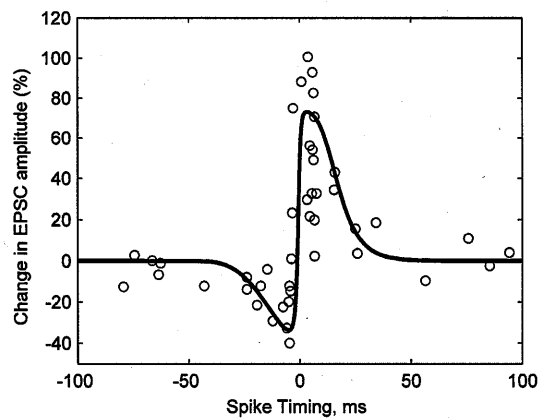
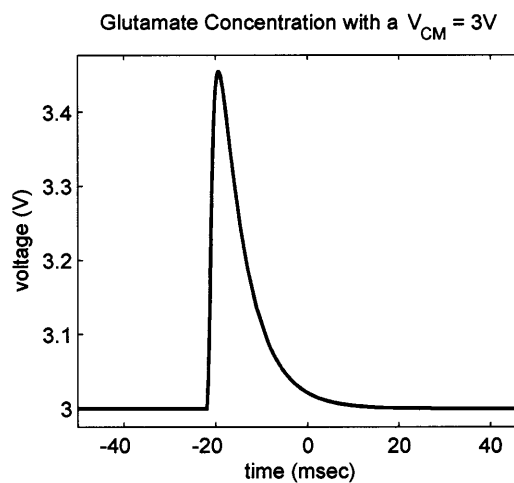
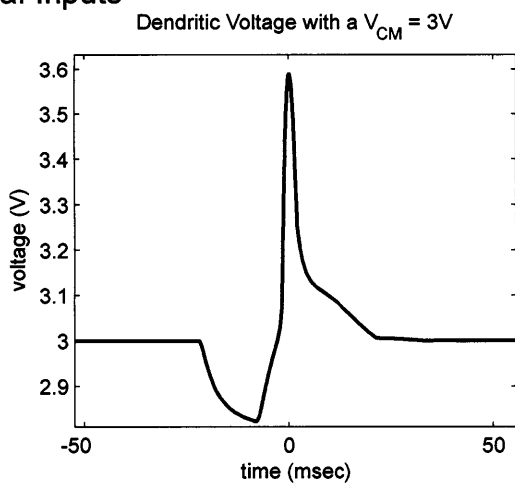
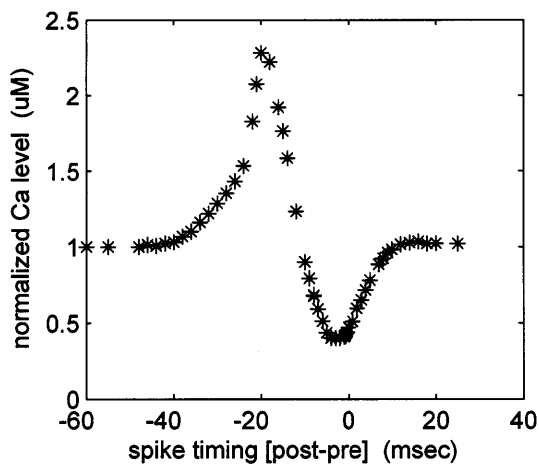
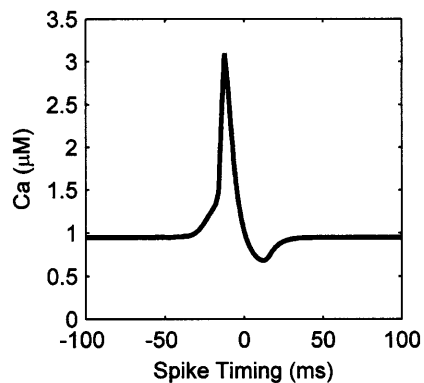


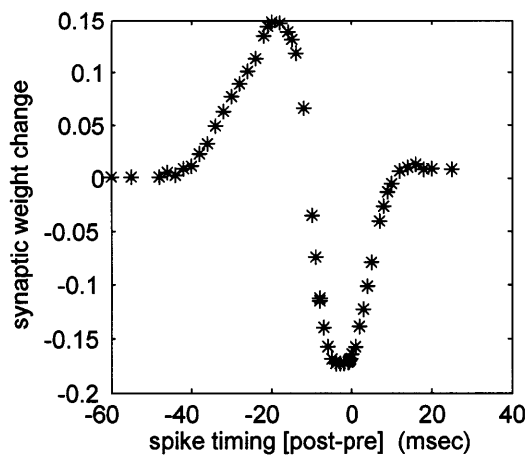
Figure 4-13. Asymmetric STDP following Hebbian rule. *Theoretical figures were obtained from [3].*



## a. Inputs

b. Normalized  $Ca_{total}$ c. Theoretical Normalized  $Ca_{total}$ 

## d. EPSC Amplitude Change



## e. Theoretical EPSC Amplitude Change

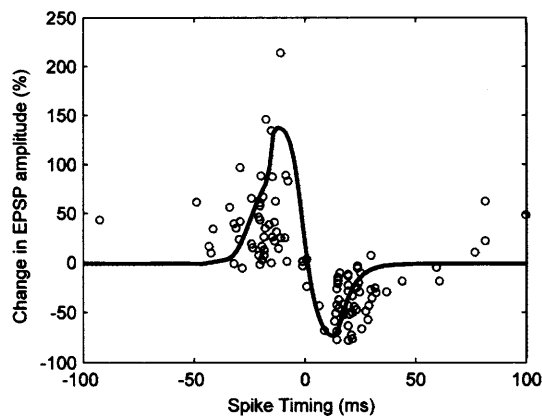
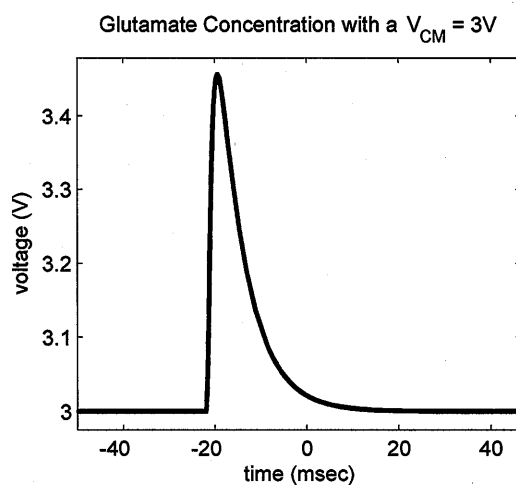
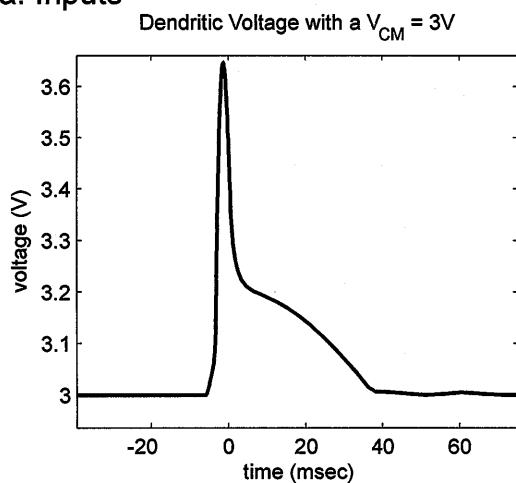
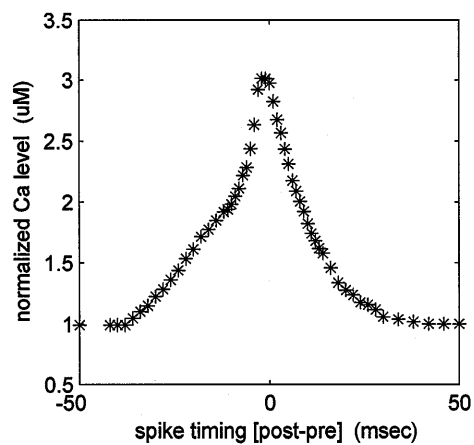
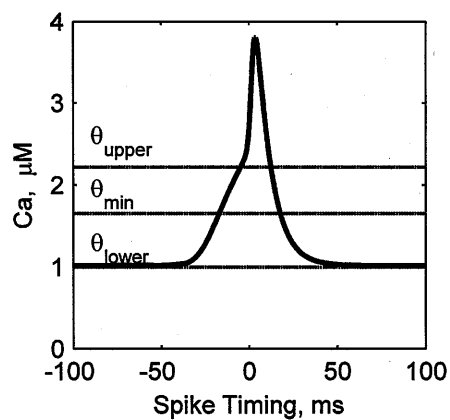
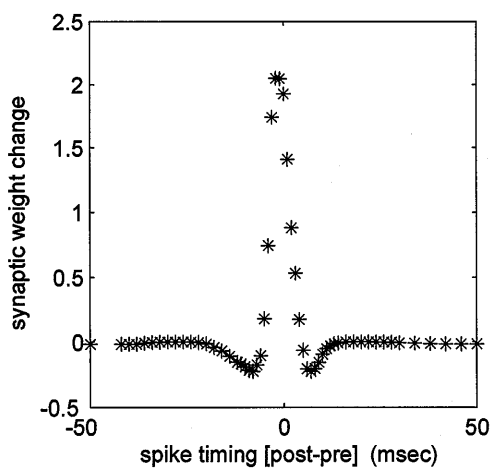


Figure 4-14. Asymmetric STDP following Anti-Hebbian rule. *Theoretical figures were obtained from [3].*

## a. Inputs

b. Normalized  $Ca_{total}$ c. Theoretical Normalized  $Ca_{total}$ 

## d. EPSC Amplitude Change



## e. Theoretical EPSC Amplitude Change

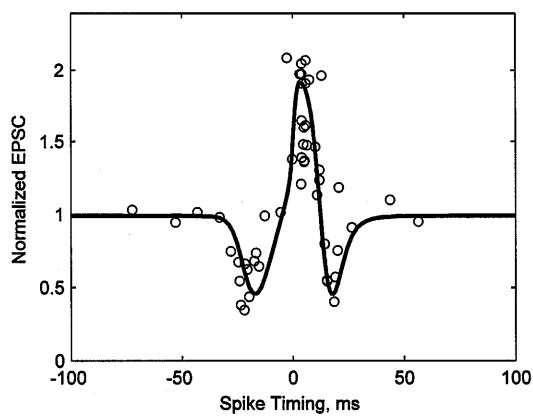
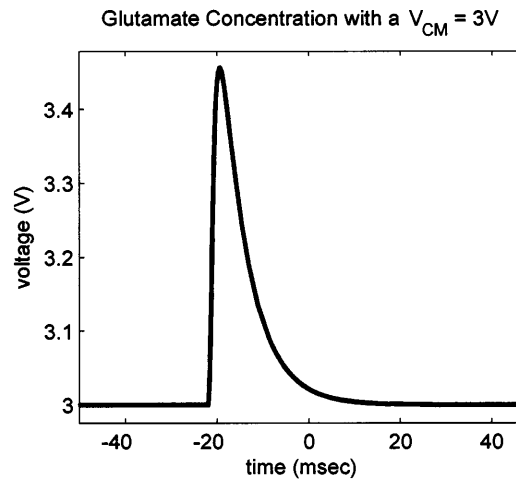
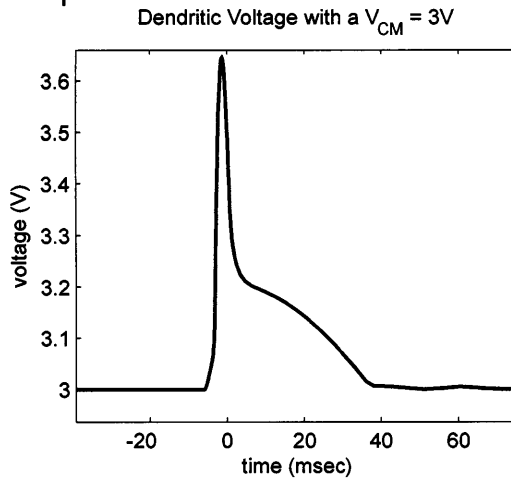
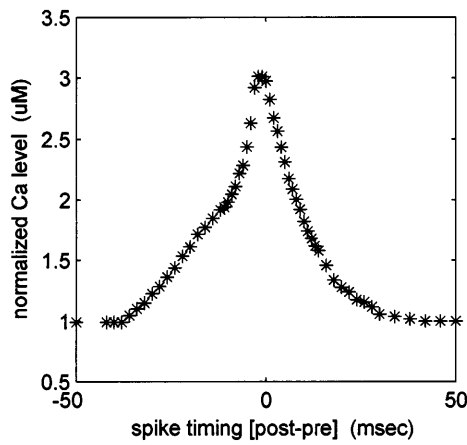
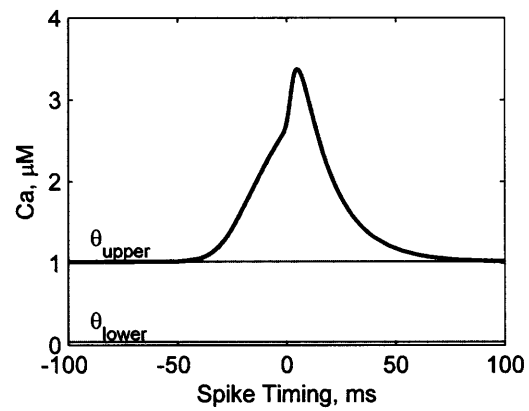
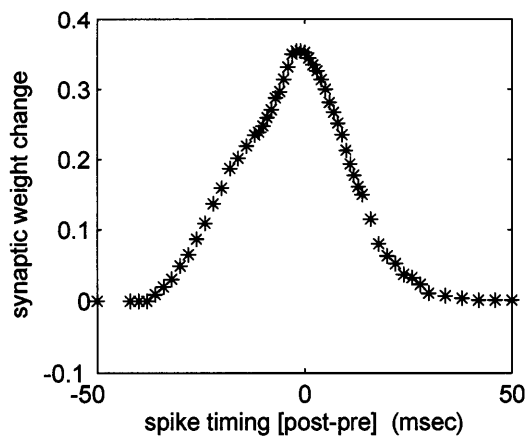


Figure 4-15. Symmetric STDP showing mixed STD and STP. *Theoretical figures were obtained from [3].*

## a. Inputs

b. Normalized  $Ca_{total}$ c. Theoretical Normalized  $Ca_{total}$ 

## d. EPSC Amplitude Change



## e. Theoretical EPSC Amplitude Change

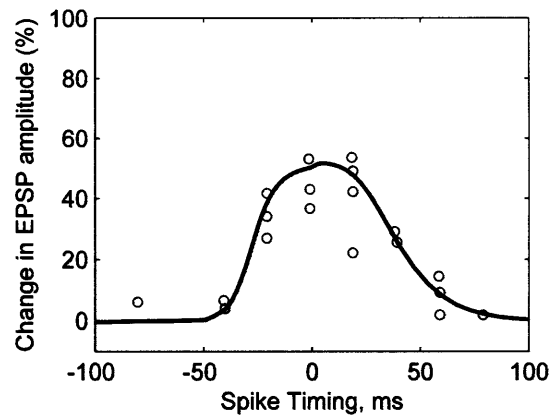
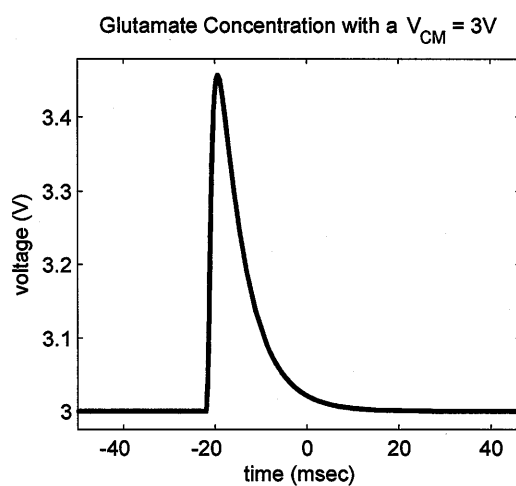
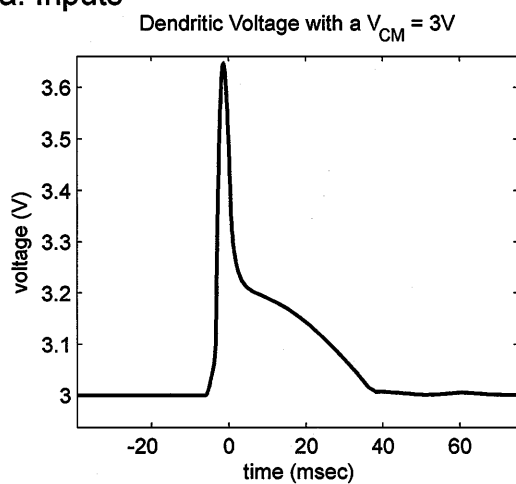
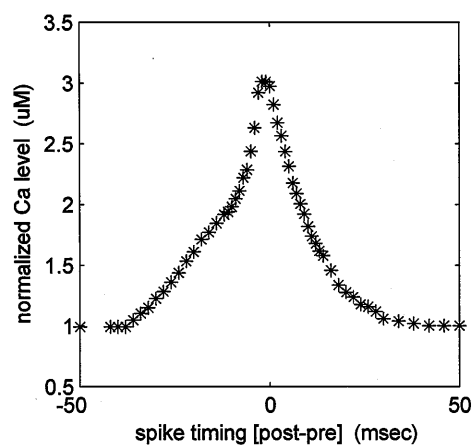
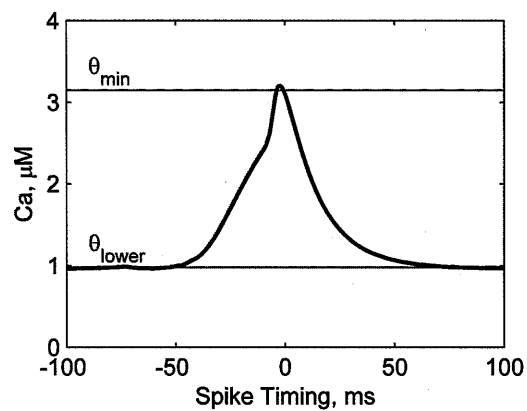
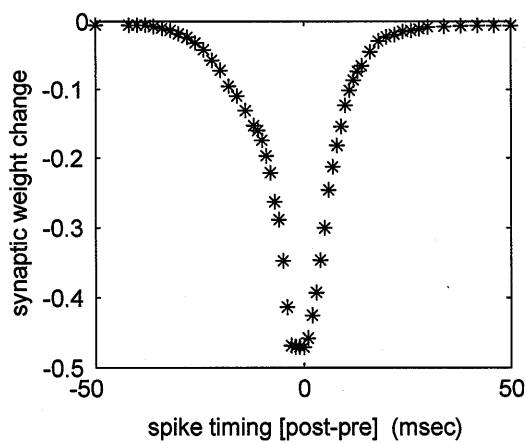


Figure 4-16. Symmetric STDP showing STP. *Theoretical figures were obtained from [3].*

## a. Inputs

b. Normalized  $Ca_{total}$ c. Theoretical Normalized  $Ca_{total}$ 

## d. EPSC Amplitude Change



## e. Theoretical EPSC Amplitude Change

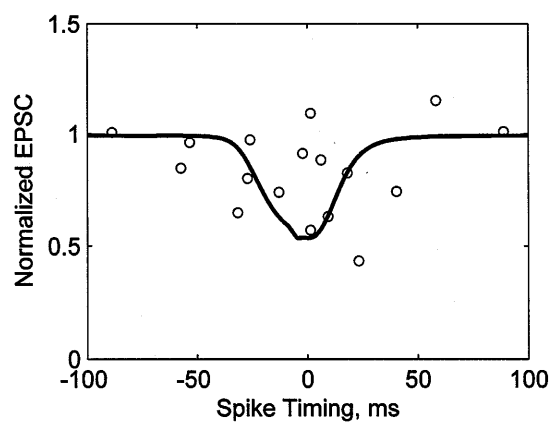


Figure 4-17. Symmetric STDP showing STD. *Theoretical figures were obtained from [3].*

## 4.5. Relating rate-based to spike timing dependent plasticity

Rate-based plasticity is the classic synaptic plasticity protocol that was characterized by Bienenstock, Cooper and Munro to form the BCM curve. In the actual experiment, synaptic strength was measured as a function of the presynaptic neuron's stimulation rate. The relationship between the stimulation rate and change in synaptic strength turns out to be similar to the BCM curve shown in Figure 4-9 [11]. Recently, studies made by Rachmuth [22, 40] resulted in a circuit that reproduces this relationship between rate and synaptic strength modification as illustrated in Figure 4-18.

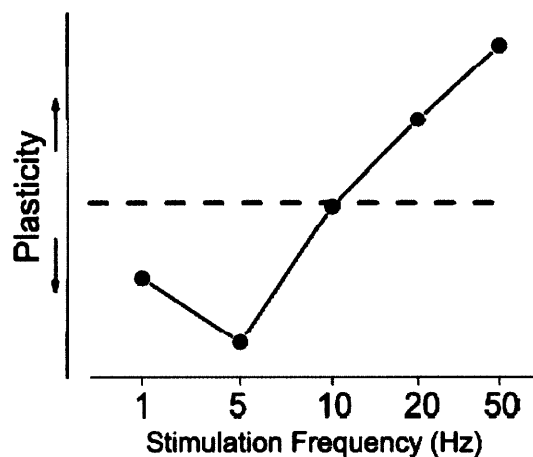


Figure 4-18. Rate-based plasticity curve. *Figure obtained from [40].*

The relationship between STDP and calcium-based plasticity was fully characterized by Lee's model [3]. However, it was not clear how rate-based plasticity relates to STDP. In rate-based plasticity, the postsynaptic stimulus is absent because we only stimulate the presynaptic neuron which presents a slight problem to the model. To amend this problem, we propose a slight extension to our implementation. In Figure 4-19, we present a block diagram that illustrates our extension. Our current circuit's signal path is symbolized by the blue lines and to connect STDP to rate-based plasticity, we append the feedback signal path depicted by the red lines. This feedback path corresponds to the coincidence detection by the NMDA channel of the EPSP signal that originates from the AMPA channels. In STDP, this feedback path was ignored because the amplitude of the EPSP signal is significantly smaller than the back propagating dendritic potential's magnitude. However, this is not the case in rate-based plasticity where only the EPSP signal is detected by the NMDA channel.

In the rate based protocol when only the presynaptic neuron is stimulated, we hypothesize that the glutamate signal is paired with the EPSP signal alone (in contrast, in STDP protocol glutamate is paired with the dendritic action potential) because the dendritic signal is absent. This results to a much lower calcium influx through the NMDA channels. However, since the protocol states that we have to stimulate multiple times, the calcium level within the synapse will eventually reach a steady state level,  $Ca_{ss}$ . If we only stimulated the synapse once, the increase in calcium would be so minute that it would be buffered almost instantaneously. This

must be the basis why we stimulate the presynaptic neuron multiple times at a higher rate. After calcium reaches its steady state value, it is then used to create and degrade AMPA receptors via the action of kinase and phosphatase respectively [3].

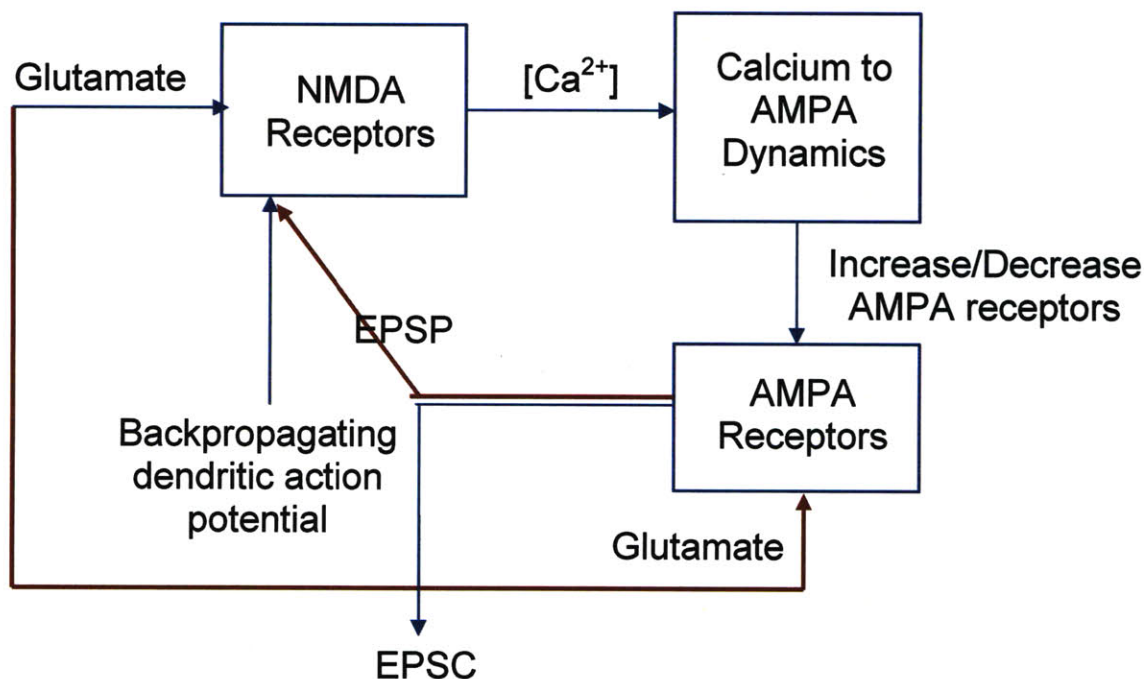


Figure 4-19. Feedback diagram between the AMPA receptor output to the NMDA receptor. The blue lines correspond to the current signal path of our circuit while the red lines correspond to the proposed feedback path.

Rachmuth discovered the relationship between the frequency of stimulation and  $Ca_{ss}$  after he designed a simple rate-based synaptic plasticity system [22, 40]. As shown in Figure 4-20, we can see that the  $Ca_{ss}$  value **monotonically increases** as a function of frequency. One observation that we note is that the waveform of  $Ca_{ss}$  has a sawtooth pattern. The upstroke is explained by the inflow of calcium while the downstroke is explained by the buffering of calcium via biochemical pathways that include its binding with kinase and phosphatase.

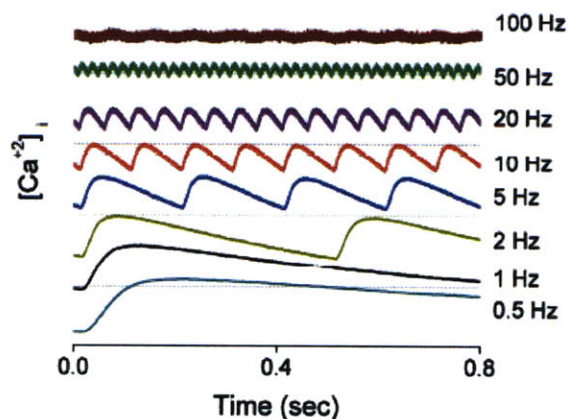


Figure 4-20. Calcium levels within a synapse increases with the presynaptic stimulation rate. Figure obtained from [40].

Lee's model suggests that at low levels of  $Ca_{ss}$ , calcium binds primarily with phosphatase, causing synaptic depression to become more expressed as  $Ca_{ss}$  increases. However after  $Ca_{ss}$  reaches a certain threshold value, it begins to bind more to kinase which slows the rate of synaptic depression and as  $Ca_{ss}$  further increases, more of it becomes bound to kinase which eventually leads to potentiation. This phenomenon explains why the BCM curve has its characteristic shape. Now, since  $Ca_{ss}$  monotonically increases with the presynaptic stimulation rate (i.e. as the rate increases,  $Ca_{ss}$  increases), the relationship between rate and changes in synaptic efficacy must exhibit a curve similar to the BCM rule function. Hence, this completes the loop between the presynaptic stimulation rate, calcium, and synaptic efficacy. To implement this relationship in silicon in the future, we suggest that the inclusion of an EPSC to EPSP conversion circuit similar to Rachmuth's design [22, 40] that serves as a feedback between the calcium to EPSC transduction circuit and the NMDA channel circuit as depicted in the modified block diagram in Figure 4-21 below. The feedback flow is indicated by the purple path.

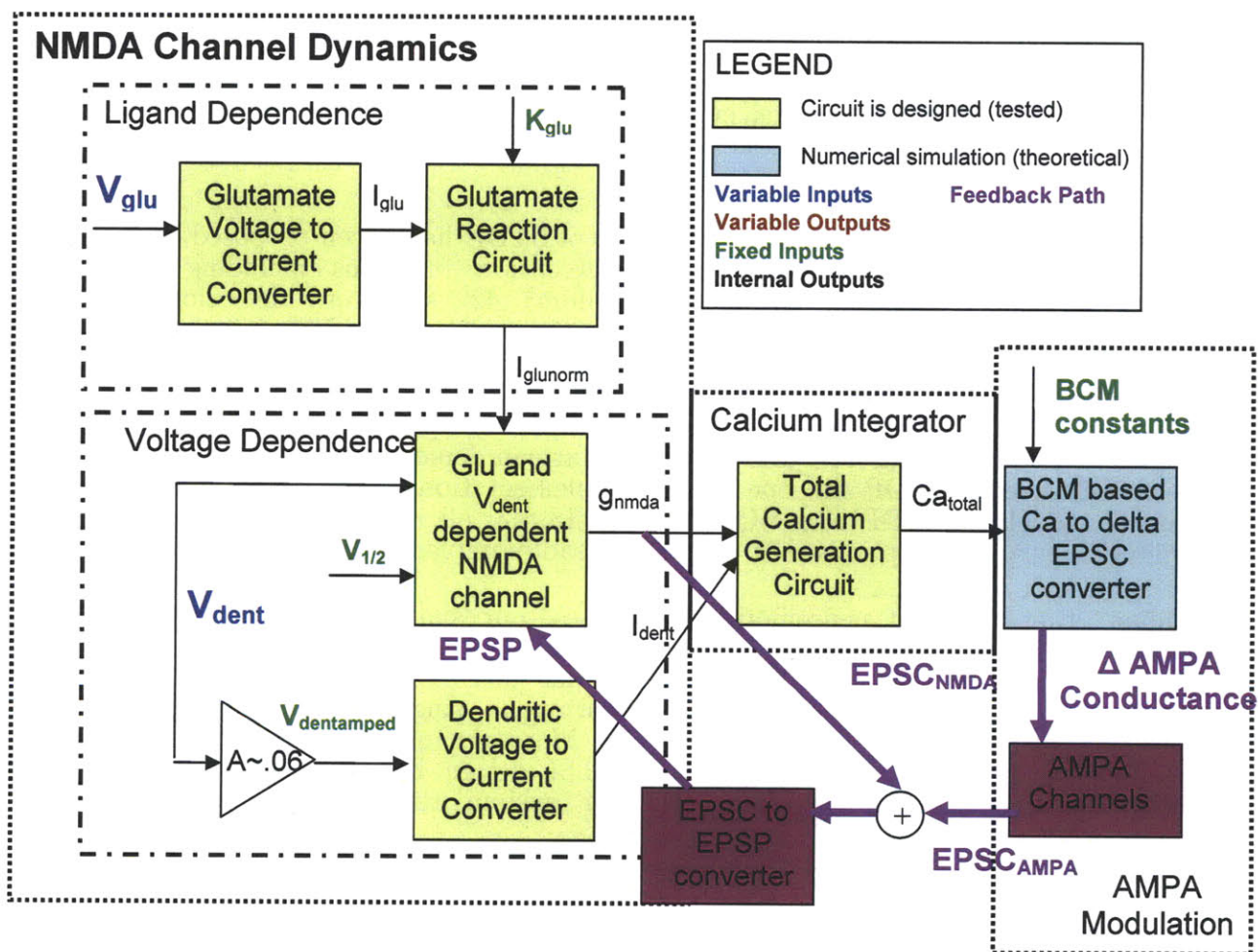


Figure 4-21. Modified block diagram showing proposed feedback path between the EPSC transduction stage and NMDA channel circuit.

## 4.6. Discussion

Multiple types of STDP were accurately captured by our neuromorphic synaptic circuit. The  $Ca_{total}$  and  $\Delta W$  curves qualitatively matches with what Lee's model predicts. Minor scaling non-idealities were observed in the magnitude of the two curves but this was expected because the traces generated by Lee's model represent the overall change in  $Ca_{total}$  and  $\Delta W$  after continuous stimulation of the presynaptic and postsynaptic neuron which follows the standard protocol used in generating the invitro data for STDP [2]. On the other hand, our data only measures the effect of a single stimulation.

Compared to current neuromorphic technologies (CMOS-nano or traditional CMOS neuromorphic circuits, our novel synaptic circuit design is the only system that provides superior functional versatility as well as robustness and holds the best promise for emulating emergent cognitive functions in large-scale neural networks. As a baseline, a neuromorphic synapse must be able to emulate multiple forms of STDP (Hebbian, anti-Hebbian and symmetrical mixed LTP and LTD, STP, and STD). Furthermore, it must be able to reproduce the classical rate-based BCM learning rule. To the best of our knowledge, currently, no other neuromorphic technologies are capable of reproducing all these forms of synaptic plasticity as well as our system.

As an example, several neuromorphic models of STDP have been proposed in the engineering literature for various sensory perception functions including odor detection [41], liquid state machine (for cerebellum) [42], synchrony detection and amplification [38]. All these implementations used only curve fitting instead of mechanistic methods, unlike our system. In other words, these implementations can only model a specific STDP curve (for example, Hebbian). None of them can actually demonstrate different learning curves based on different experimental conditions. In particular, none of the neuromorphic circuits so far can demonstrate the relationship between BCM and STDP in one single implementation of a synapse. Other theoretical models of STDP and BCM [10, 12, 13, 43] all rely on certain abstract (non-physiological) assumptions that cannot be readily implemented in CMOS.

Unlike other types of neuromorphic synapses, our synapse demonstrates and explains the significance of the singularity in the transition point between potentiation and depression in the multiple forms of STDP. None of the circuit models could emulate this smooth and continuous singularity. Some circuits demonstrate singularity but not continuity [37]. Moreover, our system exhibits pure STP and pure LTD behaviors. No other aVLSI design has even attempted to reproduce these behaviors, which makes our system the best design in the neuromorphic world.

Presently, our iono-neuromorphic STDP implementation has four circuit blocks as shown in Figure 4-4. Although the number of transistors that we used is about seven times as that of phenomenologic STDP circuits demonstrated by others [37], the circuit area is not significantly increased as the capacitor is the dominant device. This is one issue that we plan to address as we are currently exploring several ways to implement low area and high density designs for memory storage besides using a capacitor. Our power consumption in our implementation is slightly larger than conventional neuromorphic designs. Another one of our goals in the future is to reduce the complexity and power consumption of our neuromorphic STDP



implementation. We can easily scale down the bias current inputs of our circuit blocks and decrease our power supply voltage and we estimate that this should reduce our power consumption by roughly one or two orders of magnitude.

This first version of our circuit design can be further improved because several details in our model can be simplified. For instance, in our mechanistic model of the calcium to EPSC transducer, most computations in our model are divisions and multiplications. These can be converted to linear combinations of exponential functions that can be efficiently implemented in subthreshold analog transistor circuits. We have already made significant progress towards this end with our empirical sigmoidal decomposition circuit and our preliminary results look very promising. Figures 4-12 and 4-9 looks very similar to each other which shows that we are in the verge of perfecting the implementation of the BCM rule curve. Another benefit of our empirical implementation is its easiness to segue with the previous circuit block, the calcium integration circuit. This is because we can directly use the  $Ca_{total}$  output of the calcium integration circuit block without any need to convert it to a current signal. The transconductance amplifiers in our empirical design perform this conversion automatically. Our empirical transducer design is also very flexible and easy to use. With its multiple constant inputs, we can easily scale and shift the BCM curve to adapt to different types of synapses that exhibits the multiple forms of STDP. We can also make our circuit adapt to rate-based synaptic models by inserting a feedback path between our fourth and second stage circuit blocks as seen in Figure 4-21.

Before our final product is rolled out, we hope to discover and investigate new forms of STDP that we can test with our novel technology. Demonstrating that our system works with newly-discovered synaptic plasticity phenomena boosts the functionality and usability of our system to greater heights. We are very optimistic that our implementation can easily achieve this with ease and our vision is for our system to become a fixture in artificial intelligence applications in the future.



## 5. Conclusion and future work

We have designed and implemented a fully functional aVLSI synaptic circuit that unifies the different subtypes of synaptic plasticity and relate them to the BCM rule as well. Not only does our circuit emulate the empirical behavior of the synapse, it also captures the biophysical mechanisms that govern the process of synaptic plasticity – something no other neuromorphic circuit has yet to achieve. This robustness is the key feature of our design which makes it an ideal system to understand and simulate the behavior of more complex neuronal networks consisting of multiple neurons and synapses. Our superior design is very promising and should definitely be a blueprint for potential chip development and fabrication in the near future.

Since our implementation is the first of its kind, it has a lot potential for improvement. For instance, we proposed a dual transconductance circuit that decomposes the complex BCM function into two sigmoidal functions, making the BCM curve function easier to implement in silicon. This alternative approach will certainly make our circuit smaller and simpler enabling us to pack more synapses in a chip once our design is fabricated. Another aspect of our design that we can certainly improve on is the memory of our capacitor that holds the value of  $Ca_{total}$ . Currently our design causes this memory to fade fast but by using certain low leakage memory techniques such as floating gate technology [44], we can definitely improve the length of time that our memory is stored.

Our circuit is designed for low power applications, but our design was relatively conservative about the power consumption cuts that we made. We simply exploited subthreshold circuit design to enable us to cut down power by several orders of magnitude. However, we have not actually determined the lowest operating power limit of our circuit as we focused more in capturing an accurate representation of the synaptic model and not optimizing the circuit's power consumption. This leaves us with lots of room for aggressive power reductions. For instance, we can lower the power supply from 5 to 3.3 V which would reduce our power to about 60% of its consumption now. Moreover, we can also decrease the bias currents of our circuit until we begin to compromise our circuit's signal to noise ratio to further cut down our power by about two orders of magnitude.

We can also reduce the size of our system further by ensuring that we layout and fabricate our system using the smallest possible fabrication size technology to achieve a high density chip. After fabricating our chip, we could then begin implementing small neuronal networks in the cortex, cerebellum [45], and the brain stem. In an ideal world, if everything goes as planned, we can eventually integrate these networks together to form a more complex architecture that mimics physiological systems in an organism such as memory formation or respiratory control.

Once the chip becomes feasible for use, we can also begin replacing biological systems in the central nervous systems with artificial on-chip ones. A person who is suffering from a chronic nervous system disease could easily be implanted with a chip that allows him to replace the diseased “neuronal network”. Real machine learning will replace current supercomputer clusters that claims to implement so called neural networks, the software replicates of the process of learning. With a chip that

naturally allows learning and memory, we expect a wide scale improvement of technological infrastructures in the whole world.

Remember that every complex system in the world is nothing without the simple and miniscule components that make them work. The human brain, for instance, is perhaps one of the most complicated systems but let us face it, it is just composed of neurons and synapses which are interrelated to each other through synaptic plasticity. Once we uncover some of the mysteries of these neurons and synapse, which we have accomplished in our undertaking, we certainly took a huge step forward in field of neuroscience and engineering and now perhaps the dream of an artificial intelligence dominated future seems more realistically feasible to achieve.

## 6. References

1. Bandyopadhyay, P.R., *Maneuvering hydrodynamics of fish and small underwater vehicles*. Integrative and Comparative Biology, 2002.
2. Bi, G.Q. and M.M. Poo, *Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type*. J Neurosci, 1998. **18**(24): p. 10464-72.
3. Lee, C.-C., Poon C-S., and G.J. McRae, *A three-stage kinetic model unifying spike timing dependent plasticity*. 2008, Massachusetts Institute of Technology.
4. Kandel, E.R., J.H. Schwartz, and T.M. Jessell, *Principles of neural science*. 4th ed. 2000, New York: McGraw-Hill, Health Professions Division. xli, 1414 p.
5. Llinas, R., I.Z. Steinberg, and K. Walton, *Relationship between presynaptic calcium current and postsynaptic potential in squid giant synapse*. Biophys J, 1981. **33**(3): p. 323-51.
6. Lopez-Munoz, F., J. Boya, and C. Alamo, *Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramon y Cajal*. Brain Res Bull, 2006. **70**(4-6): p. 391-405.
7. Bliss, T.V. and T. Lomo, *Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path*. J Physiol, 1973. **232**(2): p. 331-56.
8. Alberts, B., *Molecular biology of the cell*. 4th ed. 2002, New York: Garland Science. xxxiv, 1463, [86] p.
9. Bliss, T.V. and G.L. Collingridge, *A synaptic model of memory: long-term potentiation in the hippocampus*. Nature, 1993. **361**(6407): p. 31-9.
10. Shouval, H.Z., M.F. Bear, and L.N. Cooper, *A unified model of NMDA receptor-dependent bidirectional synaptic plasticity*. Proc Natl Acad Sci U S A, 2002. **99**(16): p. 10831-6.
11. Bienenstock, E.L., L.N. Cooper, and P.W. Munro, *Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex*. J Neurosci, 1982. **2**(1): p. 32-48.
12. Abarbanel, H.D., et al., *Biophysical model of synaptic plasticity dynamics*. Biol Cybern, 2003. **89**(3): p. 214-26.
13. Abarbanel, H.D., R. Huerta, and M.I. Rabinovich, *Dynamical model of long-term synaptic plasticity*. Proc Natl Acad Sci U S A, 2002. **99**(15): p. 10132-7.
14. Karmarkar, U.R. and D.V. Buonomano, *A model of spike-timing dependent plasticity: one or two coincidence detectors?* J Neurophysiol, 2002. **88**(1): p. 507-13.

15. Cho, K., et al., *An experimental test of the role of postsynaptic calcium levels in determining synaptic strength using perirhinal cortex of rat*. J Physiol, 2001. **532**(Pt 2): p. 459-66.
16. Yang, S.N., Y.G. Tang, and R.S. Zucker, *Selective induction of LTP and LTD by postsynaptic  $[Ca^{2+}]_i$  elevation*. J Neurophysiol, 1999. **81**(2): p. 781-7.
17. Cowan, W.M., et al., *Synapses*. 2001, Baltimore: Johns Hopkins University Press. xiii, 767 p.
18. Dan, Y. and M.M. Poo, *Spike timing-dependent plasticity: from synapse to perception*. Physiol Rev, 2006. **86**(3): p. 1033-48.
19. Takahashi, T., K. Svoboda, and R. Malinow, *Experience strengthening transmission by driving AMPA receptors into synapses*. Science, 2003. **299**(5612): p. 1585-8.
20. Jorntell, H. and C. Hansel, *Synaptic memories upside down: bidirectional plasticity at cerebellar parallel fiber-Purkinje cell synapses*. Neuron, 2006. **52**(2): p. 227-38.
21. Mead, C., *Analog VLSI and neural systems*. 1989, Reading, Mass.: Addison-Wesley. xxii, 371 p., 8 p. of plates.
22. Rachmuth, G. and C. Poon, *Transistor analogs of emergent iono-neuronal dynamics*. HFSP Journal, 2008.
23. Mead, C., *Neuromorphic electronic systems*. Proceedings of the IEEE, 1990. **78**(10): p. 1629-1636.
24. Fitzhugh, R., *Motion picture of nerve impulse propagation using computer animation*. J Appl Physiol, 1968. **25**(5): p. 628-30.
25. Getting, P.A., *Emerging principles governing the operation of neural networks*. Annu Rev Neurosci, 1989. **12**: p. 185-204.
26. Guttman, R., L. Feldman, and E. Jakobsson, *Frequency entrainment of squid axon membrane*. J Membr Biol, 1980. **56**(1): p. 9-18.
27. Nagumo, J., S. Arimoto, and S. Yoshizawa. *An active pulse transmission line simulating nerve axon*. in *Proceedings of the IRE*. 1962.
28. Katz, R.H. and G. Borriello, *Contemporary logic design*. 2nd ed. 2005, Upper Saddle River, N.J.: Pearson Prentice Hall. xviii, 590 p.
29. Gray, P.R., *Analysis and design of analog integrated circuits*. 4th ed. 2001, New York: Wiley. xviii, 875 p.
30. van der Meer, P.R., A. van Staveren, and A.H.M.v. Roermund, *Low-power deep sub-micron CMOS logic: Subthreshold Current Reduction*. Dordrecht: Springer, 2004. 78.

31. Wilamowski, B.M. *VLSI analog multiplier/divider circuit*. in *Industrial Electronics*. 1998.
32. Gilbert, B., *Translinear circuits: a proposed classification*. Electronics Letters, 1975. 1(11): p. 14-16.
33. Ismail, A.M. and A.M. Soliman, *Novel CMOS wide-linear-range transconductance amplifier*. Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions, 2000. 47(8): p. 1248-1253.
34. Sarpeshkar, R., L. R.F., and C. Mead, *A low-power wide-linear-range transconductance amplifier*. Analog Integrated Circuits and Signal Processing, 1997.
35. Boegerhausen, M., P. Suter, and S.C. Liu, *Modeling short-term synaptic depression in silicon*. Neural Comput, 2003. 15(2): p. 331-48.
36. Chicca, E., et al., *A VLSI recurrent network of integrate-and-fire neurons connected by plastic synapses with long-term memory*. IEEE Trans Neural Netw, 2003. 14(5): p. 1297-307.
37. Indiveri, G., E. Chicca, and R. Douglas, *A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity*. IEEE Trans Neural Netw, 2006. 17(1): p. 211-21.
38. Bofill-i-petit, A. and A.F. Murray, *Synchrony detection and amplification by silicon neurons with STDP synapses*. IEEE Trans Neural Netw, 2004. 15(5): p. 1296-304.
39. Van Schaik, A., *Building blocks for electronic spiking neural networks*. Neural Networks, 2001(14): p. 617-628.
40. Rachmuth, G., *Analog VLSI circuit models of neuronal dynamics and synaptic plasticity*, in *Division of Engineering and Applied Sciences*. 2005, Harvard University: Cambridge.
41. Jacob, K.T., et al., *Analog VLSI circuit implementation of an adaptive neuromorphic olfaction chip*. IEEE Transactions on Circuits and Systems, 2007. 54: p. 60-73.
42. Oliveri, A., R. Rizzo, and A. Chella. *An application of spike-timing-dependent plasticity to readout circuit for liquid state machine*. in *Neural Networks, International Joint Conference*. 2007.
43. Izhikevich, E.M. and N.S. Desai, *Relating STDP to BCM*. Neural Comput, 2003. 15(7): p. 1511-23.
44. Kahng, D. and S.M. Sze, *A floating-gate and its application to memory devices*. The Bell System Technical Journal, 1967. 46(4): p. 1288-1295.
45. Raymond, J.L., S.G. Lisberger, and M.D. Mauk, *The cerebellum: a neuronal learning machine?* Science, 1996. 272(5265): p. 1126-31.

