12-2017

Analog Axon Hillock Neuron Design for Memristive Neuromorphic Systems

Ryan John Weiss
University of Tennessee

To the Graduate Council:

I am submitting herewith a thesis written by Ryan John Weiss entitled "Analog Axon Hillock Neuron Design for Memristive Neuromorphic Systems." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

Garrett S. Rose, Major Professor

We have read this thesis and recommend its acceptance:

Benjamin J. Blalock, Mark E. Dean

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Analog Axon Hillock Neuron Design for Memristive Neuromorphic Systems

A Thesis Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Ryan John Weiss

December 2017

Abstract

Neuromorphic electronics studies the physical realization of neural networks in discrete circuit components. Hardware implementations of neural networks take advantage of highly parallelized computing power with low energy systems. The hardware designed for these systems functions as a low power, low area alternative to computer simulations. With on-line learning in the system, hardware implementations of neural networks can further improve their solution to a given task.

In this work, the analog computational system presented is the computational core for running a spiking neural network model. This component of a neural network, the neuron, is one of the building blocks used to create neural networks. The neuron takes inputs from the connected synapses, which each store a weight value. The inputs are stored in the neuron and checked against a threshold. The neuron activates, causing a firing event, when the neuron's internal storage crosses its threshold. The neuron designed is an Axon-Hillock neuron utilizing memristive synapses for low area and energy operation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this thesis, the design and verification strategy for one componenet for a neuromorphic computing system architecture is presented. In order to understand the reasoning behind the circuit level component's design choices, it is imparitve to understand the background information that led to the developement of the circuit. The circuit designed is at its core an analog neuron that is designed for a syncronous neuromorphic system. The circuit component fits into a larger system that which drove the component's requirements. The requirements of the whole architecture combined together at different levels, high level simulation model and low level circuitry, to give the circuit design for this component stringent goals. The main goal for this work is developing a useful circuit that functions properly for our neuromorphic computing architecture by exicuting all required steps in the opreation time. Subsequent design implementation goals are a low-power, low-area circuit that meets the functional and timing requirements.

1.1 Neuromorphic System

The first step in building this circuit is to look at the neuromorphic computing system architecture that requires it as a component. A neuromorphic computing system is an alternative computational architecture to a von Neumann architecture. Neuromorphic systems are inspired by our own brains. To compete with the low power consumption of our brains, which is twenty percent of our total bodies power consumption, they would

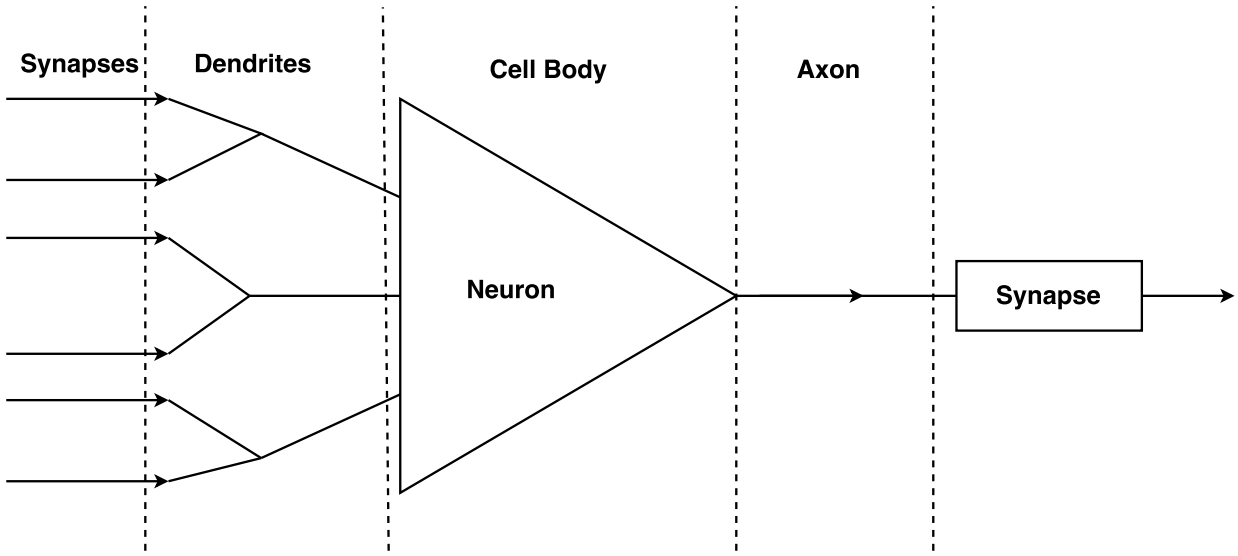

Figure 1.1: Labeled neuron cell

require 20 W or less power [10][6]. Neuromporphic systems also take advantage of our brains parrallel processing capabilities, which is a result of the brain's highly interconnected cells. The architecure the circuit in this work is designed to implement is Neuroscience-Inspired Dynamic Architecture, NIDA, which is a spiking neural network [16]. The circuit is designed to implement a limited physically realizable version of NIDA.

1.2 Biology

Due to the biological inspiration it is important we understand how the circuit components biological conterpart functions. The neocortex in the human brain consists of approximately twenty billion neurons [13]. Each neuron is connected to other neurons via synapses, with the neuron's in the neocortex having roughly ten thousand connections each. These cells create the electrical impulses in the brain that functions give the ability to think.

The neuron has three main components, input, storage, and output. The neuron is shown in a block diagram depiction in figure 1.1. The inputs into the neuron are its dendrites. Dendrites branch out to connect to the outputs of other neurons. The dendrites take the outputs of other neurons and bring them to the cell body of the neuron. The cell body is the summation center of all the inputs of a neuron. In the cell body, the neurons inputs are

stored and processed. Upon a neuron reaching a set electrical potential in the cell body, the neuron will fire an output signal. During the firing process, neuron resets its cell body and does not accumulate inputs. The reset time is known as a neuron's refractory period. The output signal is a voltage spike that is propogated out the neuron's axon. The axon connects to synapses that create a bridge to the next neuron's dendrites.

Synapses are a modulation of a neuron's ability to cause another neuron to fire. Synapses hold a relative strength between two neurons. Synapses grow to create stronger or weaker conncetions to neurons based on their use case. Hebb's postulate states that correlated neuron activety adjusts the weight of the synapse as seen in equation 1.1 [7]. The effective weight of the synapse, w, is changed based on the events of the neurons connected to it. The neuron whos output flows through the synapse, x_i, and the neuron who takes that synapse output as input, x_j, are the pre and post neuron, respectively.

$$\Delta w_i = x_i \times x_j \tag{1.1}$$

This is expanded upon in spike timing dependent plasticity. In STDP, the synapse increments or decriments its weight based on the time difference between a pre and post neuron fire [3]. The synapses ability to change its strength, which is its ability to cause a neuron to fire, is known as synaptic plasticity. The function of the changing effectiveness of one neuron to cause another neuron to fire gives the ability to learn. Spike timing dependent plasticity changes the weight of a synapse during its operation. This process is a type of unsupervised online learning because the updated behavior happens as the system runs, or is online, and has no feedback about whether or not the output is correct, or is unsupervised. The neuromorphic system implemented takes advantage of different combinations of connected neurons and synapses to solve problems and uses online learning to improve its results.

1.3 Neuron Model

The neuron can be expressed in functional electrical terms, the first of which is the integrate and fire model by Louis Lapicque in 1907 [1]. Equation 1.2 is a the basic integrate and fire

model for the function of a neuron. The inputs into the neuron are currents I(t) which are integrated on the capacitor C_m and create the voltage V_m. From equation 1.3 the fourier transform of a constant current input can be used to describe the refractory period. Upon the voltage, V_m, crossing a set threshold voltage, V_{th}, the neuron fires and resets its output voltage in time t_{ref}, from equation 1.3.

$$\mathrm{I(t)} = C_m \times \frac{dV_m(\mathrm{t})}{\mathrm{dt}} \tag{1.2}$$

$$\mathrm{f(I)} = \frac{\mathrm{I}}{C_m \times V_{th} + t_{ref} \times \mathrm{I}} \tag{1.3}$$

An improvement upon the integrate and fire model, is the leaky integrate and fire model which includes a leakage factor of the neuron shown in equation 1.4 [12]. The leakage factor gives a threshold input currents must cross to cause a neuron to fire. The threshold current is set by the threshold voltage, V_{th}, divided by the membrane resistance factor, R_m.

$$\mathrm{I(t)} - \frac{V_m\ (\mathrm{t})}{R_m} = C_m \times \frac{dV_m(\mathrm{t})}{\mathrm{dt}} \tag{1.4}$$

Another neuron model, the Hodkin-Huxley model takes into account multiple voltage-dependent currents with there own conductance equations [8]. This model better represents the bio-physical system, and consiquently, it is more complicated to reproduce and its intent is to realistic modeling the ion channels during the activity of a neuron. The detail of explaining the biological system goes beyond the basic functional application used in this neuromorphic system. In this neuromorphic system, the flow of current in and out of the neuron is controlled by the synapse which retains a set weight value, unless online learning occurs, and a neuron leakage parameter. The neuron does not try to match the biological signals in the neuron, but rather emulates the functionality to recreate the adaptive problem solving behavior of the brain.

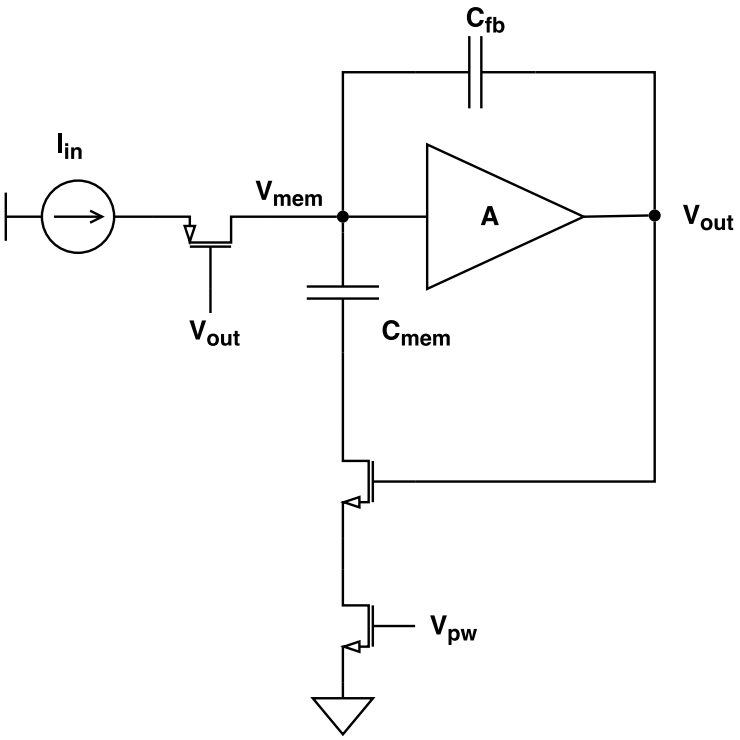

Figure 1.2: Axon hillock circuit schematic

1.4 Axon Hillock Circuit

In this system, an on chip circuit used to emulate the neuron behavior is required. The neuron circuit needs to accumulate and store inputs which can cause a firing event upon crossing a threshold and then reset itself. The axon hillock circuit proposed by Carver Meade in 1989 preforms these functions [11]. The circuit represents the functionality of the axon hillock in the neuron. As mentioned in section 1.2, the axon functions as the output of the neuron. The axon hillock is the part of the axon that is connected to the cell body. The axon hillock is where the output spike is generated upon the cell body crossing the threshold potential. The axon hillock circuit represents the cell body as the capacitor C_{mem}. The capacitor stores inputs and upon the voltage on the capacitor crossing the amplifier's threshold, generates an output spike, V_{out}. The refractory period of the neuron is set by the feedback transistors and the voltage V_{pw}. The inputs are blocked during the refractory period by a p-type transistor that is controlled by the output voltage. The feedback capacitor, C_{fb}, is positive feedback used to drive the voltage on the input capacitor to give the spike duration

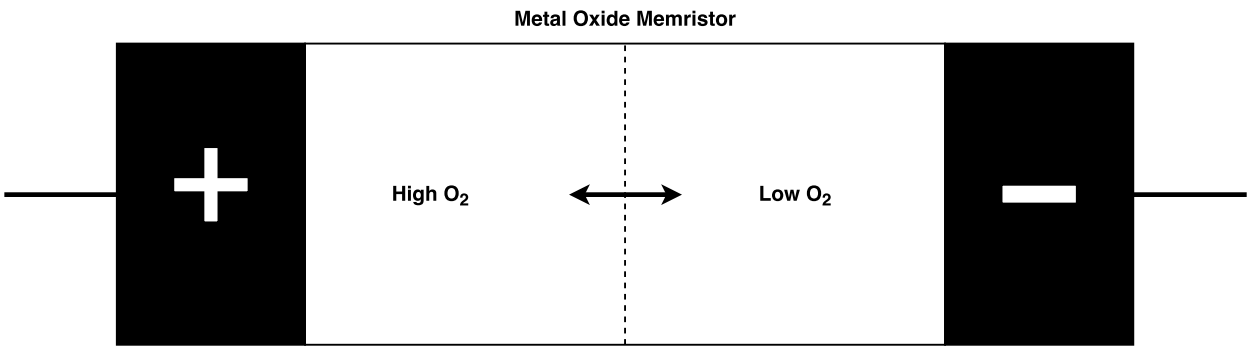

Figure 1.3: memristor

and set the reset voltage of the neuron. This circuit implements an analog process of spike creation and reseting that fits the functional criteria for our system.

1.5 Memristor

The memristor is a two terminal electrical device first proposed by Leon Chua in 1971 [5]. The theoretical device links change in flux with change in charge. This connection gives the device the relation between its current resistance state and its previous voltages applied. As seen in equations 1.5 and 1.6 the conductance value of the memristor is a function of the voltages that have been applied on the memristor.The change in state equation, equation 1.6, determines the change in resistance of the memristor based on voltage, time, and the devices current state. Simply put, the memristor is a two terminal resistive memory device. The physical realization of a resistive memory device happened in 2007 at HP [18]. The device was a titanium oxide, which switches its doping based on the voltage applied which in turn changed its resistance. The highly doped section gives the device low resistance and the less doped section gives the device high resistance. The size of the two regions changes with voltage applied on the device. The memristor is appropriate to serve as the synapse connection in the system [9]. It relies on the ability to switch resistance states to allow online learning in the system.

$$I(t) = G(x,V,t) \times V(t) \tag{1.5}$$

$$\Delta x(t) = f(x,V,t) \tag{1.6}$$

Chapter 2

System Design

The neuromorphic system will be implemented entirely on chip. The system design is for a 65 nm process which will be fabricated with our collaborators at SUNY. The process has on chip memristors, which will be fabraicated between the first two metal layers. The model used is for circuit simulation is from [2]. The model captures the resistive switching behavior of the memrsitor. As seen in figure 2.1, the memristor model has a switching voltage, V_{reset} and V_{set} which are the switching voltages from from a high resistance to a low resistance and low to high resistance, respectively. Since the thresholds are opposite polarities, the device is a bipolar, and its resistance states should cross the origin on a current verses voltage graph. The device will be able to switch to intermediate states, which means it will achieve more than the two maximum resistance values, which are denoted HRS and LRS for high and low resistance state. The onchip memristor will be a nano scale analog memory device used as the synapse's weight.

2.1 MrDANNA

The architecture the neuromorphic system follows to solve problems is named memristive dynammic adaptive neural network architecture, MrDANNA [4]. Figure 2.2 shows the block diagram of the MrDANNA physical design. The system consists of cores which contain multiple synapses inputing into a single neuron. The neuron provides the output to synapses in other cores. The dimensions of the array are predetermined by the chip size. On the

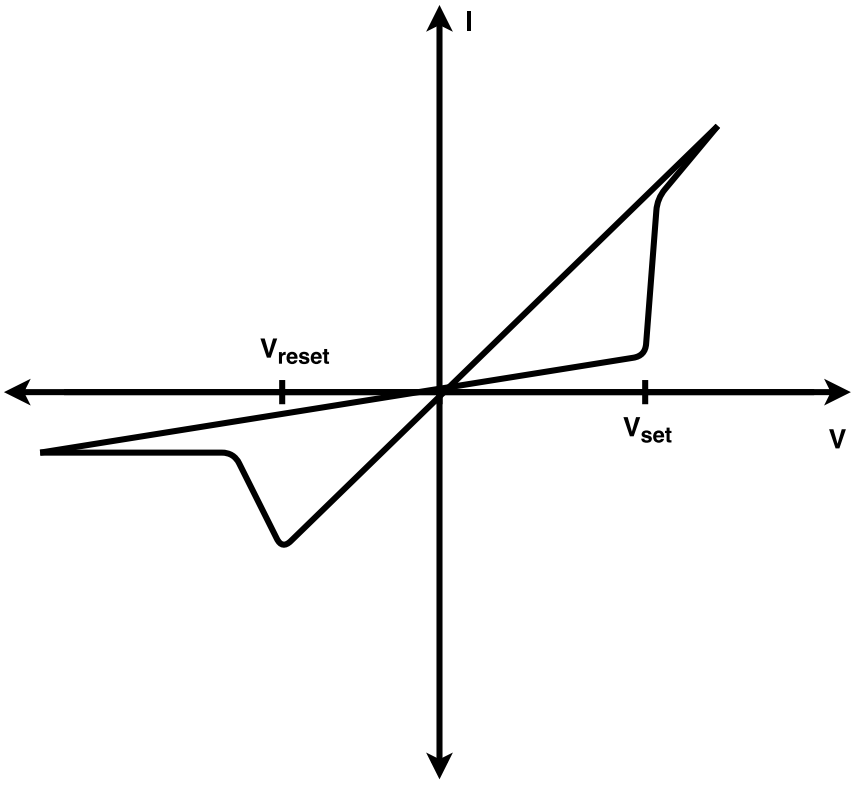

Figure 2.1: Memristor model

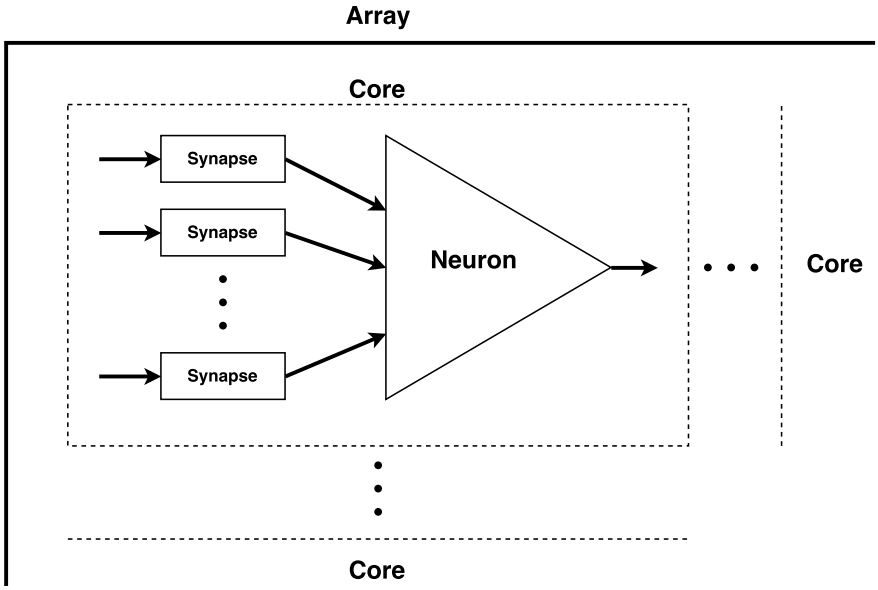

Figure 2.2: MrDANNA block diagram

chip, the possible connections between synapses and neurons are programmable. Finding how to program the chip to solve a specific problem uses a high level model simulator. The neuron circuit was modeled in this architecture to verify its profiecency at solving problems. MrDANNA, use evolutionary optimization to build networks of synapse and neuron connections that solve a set problem [15]. This is a form of offline learning, since it is preprocessing before the implementation is set. The evolutionary optimization takes a set number of inputs and outputs and builds conections of neurons and synapses with definable weights, delays and connections. Sets of networks are tested and networks that have the highest accuracy for solving the problem are kept and mutated. This happens every epoch, and each new epoch has some new random networks, the best networks from the previous epoch and some mutations of the best networks from the previous epoch. The desired result at the end of the evolutionary optimization is a network with the best chance at solving the problem. This preprocessing needs to accurately depict the hardware implementation such that results from the offline learning will reflect the results of the circuit. The axon hillock neuron used in this work was added as a model to the simulation and verifed that it can be used to solve problems.

2.2 High Level Model

The initial high level model used in the simulator models the dynamics of the synapse and neurons in an abstract ideal form. The neuron takes inputs from the synapses that are connected that fired at a given time. The neuron adds each weight value of a synapse that fires together with the current accumulated value to update the neurons stored value. If the stored value reaches the neuron's threshold value, the neuron fires. When neuron's fire they enter a refractory period which resets the stored value and blocks inputs. A neuron fire updates the weight of the synapses it is conncected too. A preliminary representation of the neuron in this work was added to the simulator which more closely represents the physically implemented system. These preliminary results of the neuron were found to be a possible candidate for solutions in the simulator. These results led to some design choices, which will

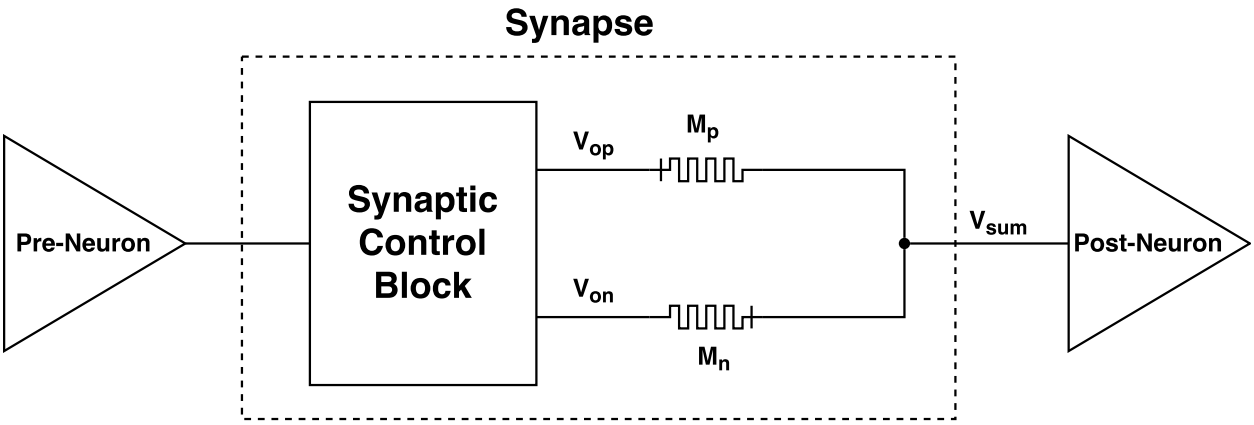

Figure 2.3: Bi-memristor synapse block diagram

be expanded upon later. The design improvements revolved around increasing a synapse's ability to cause a neurons to fire.

2.3 Synapse

The synapse used in this system is the bi-memristor synapse from [14]. The synapse stores an effective weight range of positive and negative values. The weight value is determined by the difference in resistance values of the two memristors. When M_p is a higher resistance than M_n, the weight is negative since more current is flowing out of the synapse than into the synapse. If the resistance values are reversed, the synapse has a positive weight, which adds voltage to the neuron's stored voltage. When the resistance values are equal a zero weight is applied which would not cause a neuron to fire. However, this functionality requires the summing node, V_{sum}, to be held at a constant voltage. In the neuron implemented, the summing node is floating, which changes the weight relationship. The difference in the weight relationship for this neuron's input implementation will be explained in section 3.1.

Besides holding a weight value, the synapse also changes its weight value based on the inputs. The synapse in this system implements a single cycle of learning. The condition for positive and negative weight change, from [19], are a pre and post neuron firing within once clock cycle of each other. The synapse is potentiated, which means the weight value is increased, when the neuron who receives the synapses weight value, the post neuron, fires directly after the neuron that feeds into the synapse, the pre neuron. If the pre neuron fires

directly after the post neuron, the synapse weight depressed, or decreases. These learning rules are termed long term potentiation and depression respectively, because the weight change is stored indefinetly. Only synapses that fired the clock cycle before and after the post neuron fired will have their weights updated. All other synapses are left at their current weight value. To perform the weight update the memristors must be driven above their threshold voltages, V_{reset} and V_{set}. To accomplish this the memristors are driven to the supply rails on either side. The synapse control block drives one node of the memristor, while the other node is driven by the neuron.

Chapter 3

Neuron

The neuron implemented in this system is the synchronous axon hillock from [19]. This circuit is chosen because of its relatively small area and energy consumption. In designing this neuron for this system, the proper functionality for the neuromorphic system is the main priority. The neuron accumulates inputs from its connected synapses, fires an output voltage spike upon crossing a voltage threshold and then resets itself. In figure 3.1, we see the schematic for the neuron sent to fabrication, which has an additional p-type transistor added to the circuit from [19]. This p-type transistor is a keeper transistor which helps in the proper operation of the domino logic circuit. As described in [17], this transistor helps mitigate leakage and keeps the output of the domino logic set during an evaluate phase of a logical low input. This is imperative for this circuit, because the domino logic amplifier serves as a single bit digital to analog converter, and the output, F_{post}, should remain logical

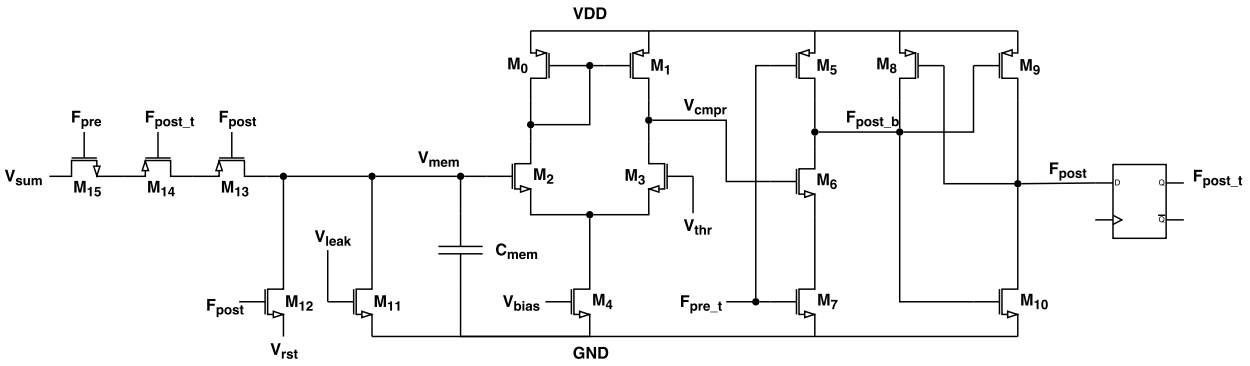

Figure 3.1: Synchronous axon hillock schematic

Table 3.1: Device sizes used for neuron

Parameter	Value
C_{mem}	150 pF
M_0, M_1	$\frac{1.5\,\mu m}{180\,nm}$
M_2, M_3, M_4	$\frac{1.2\,\mu m}{120\,nm}$
M_5, M_7, M_8	$\frac{150\,nm}{60\,nm}$
M_6	$\frac{300\,nm}{60\,nm}$
M_9	$\frac{1.8\,\mu m}{60\,nm}$
M_{10}	$\frac{300\,nm}{60\,nm}$
M_{11}, M_{12}	$\frac{150\,nm}{60\,nm}$
M_{13}, M_{14}	$\frac{300\,nm}{60\,nm}$
M_{15}	$\frac{300\,nm}{60\,nm}$

low until the input into the domino logic is a high voltage. Since the input into the domino logic gate is analog, there is significantly more leakage than if it was digital. The sizing and requirements of the domino logic gate are further explained in section 3.2.

The sizing of all the components that are to be fabricated are listed in table 3.1. The p-type transistors, M_{13} and M_{14}, n-type transistor, M_{15}, and the capacitor, C_{mem} define the input accumulation functionality. Transistor M_{11} is a leakage transistor that sets a coninuous loss of the accumulated voltage based on the voltage V_{leak}. The neuron's accumulation is explained in greater detail in section 3.1. The voltages, V_{rst} and V_{thr}, are analog bias voltages that are chosen based on results from high level simulations. V_{thr} is the threshold voltage that the neuron must accumulate to fire, and V_{rst} is the voltage the neuron resets to after it fires. From the preliminary results high level model, mentioned in section 2.2, the reset voltage, V_{rst}, is made a controlled parameter and not set to the lower rail voltage as seen in [19]. A higher reset voltage reduces the number of synapse fires into the neuron it takes to cause it to fire. The transistors, M_0 through M_4, form a differential amplifier that is used to apply the adjustable threshold voltage, V_{thr}. Without the differential amplifier, the threshold would be set by the switching voltage of the domino logic inverter. Transistors M_5 through M_{10} are a domino logic amplifier that creates the output voltage spike. Transistors M_{12}

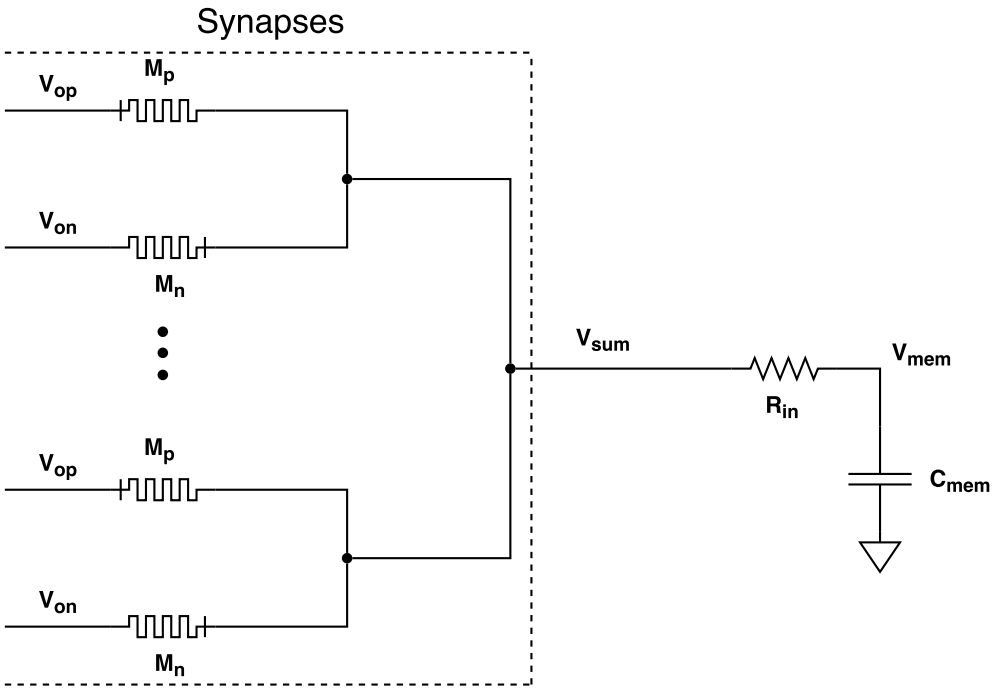

Figure 3.2: Synchronous axon hillock input schematic

through M_{14} are for reseting the neuron and implementing the refractory period. M_{12} drives the accumulated voltage, V_{mem}, to the reset voltage, V_{rst}, when the neuron has fired. M_{13} and M_{14} block inputs into the neuron defining its refractory period's length. The operation and sizing of the differential amplifier and domino logic inverter, and reset transistors are futher explain in section 3.2.

3.1 Neuron Input

As mentioned in the section 2.3, due to the floating input of the synchoronous axon hillock neuron, the input does not add in the weight value of the connected synapse directly. Assuming the weight is proportional to the difference in the resistance of the two memristors, the change in accumulated voltage is a function of the weight of all synapses that fired at the same time and the current accumulated voltage of the neuron that received the synapse fires. When a synapse fires into a neuron, the charge on the neuron, V_{mem}, is driven to the voltage on the summing node V_{sum}, from figure 3.1. The voltage on the summing node is the voltage created by resistive voltage division by the two memristors shown in equation 3.1. This happens

because the resistance of the transistors in series at the input, M_{13} through M_{15} have higher resistance than the memristors in the synapse. To accomplish this the transistors are placed in series and are intentionally given a low width to length ratio. The ability of the synapse to charge or discharge the capacitor in the neuron is proportional to the input resistance times the capacitance from equation 3.2. This relationship means the high input resistance allows the size of the capacitor, C_{mem}, to be reduced for the same charging time constant. If more than one synapse fires at a time, the voltage on the summing node is the average of the voltage that would be created by each synapse individually. Leakage through synapses that are not firing is present but is negligible because those synapses are set to a high resistance.

Two factors effect the possible voltages that the summing node can achieve. For this setup the maximum voltage the summing node can achieve happens when the memristor driven by V_{op}, M_p, is at the minimum possible resistance state and the memristor driven by V_{on}, M_n, is at the maximum possible resistance state. The minimum negative weight happens when M_p is at the maximum resistance state and M_n is at its minimum. These two resistance states set the maximum and minimum possible voltage on the summing node. The voltage these minimum and maximum synaptic weights create is also a factor of the voltages used to drive the memristors. The driving voltages are set based on the switching voltage of the memristors. The difference between the voltages V_{op} and V_{on} used for this system must be below the switching voltage of the memristors. This limit is set so the memristors will not change states while trying to read the synapses weight value.

The effect of this voltage on the change in voltage is approximated in equation 3.2. The voltage change is proportional to the difference between the accumulated voltage and the summing node voltage. If the summing node voltage is a higher voltage than V_{mem}, ΔV_{mem} is positive. If the summing node is a lower voltage then V_{mem} the accumulated voltage decreases. The alteration of the neuron's accumulated voltage, V_{mem}, is effected in both direction and magnitutdeby the synapse weight and the current accumulated voltage. Equation 3.2 does not consider the changing resistance of the input resistance into the neuron, R_{in}, which is also dependent on V_{sum} and V_{mem}. The T in equation 3.2 is the clock period of the system, which is 50 ns, since the synapse is activated for one clock cycle. Simpler approximations were used in the high level model discussed in section 2.2, and the future high level model

of the system will use these or more detailed equations to describe the synapse weight to neuron accumulated voltage relationship.

The voltage that the capacitor can be driven to is limited by the input transistors. Since the input transistors are p-type and n-type in series, the maximum and minimum voltage that V_{mem} can be driven to is based on the threshold voltages for the two types of transistors. This means if V_{mem} is at the threshold of the p-type transistors, approximately 450 mV, and a synapse fires that creates a voltage below that on the summing node, the resulting voltage of V_{mem} will remain approximately the same, ignoring sub-threshold operation. Ultimately, the neuron has a lower limit on its accumulation of negative weight synapses, which is the threshold of p-type transistors. The lower limit will be a factor in the high level simulator. The lower limit can be a lower voltage if low threshold voltage transistors are used and can be the lower rail voltage if transmission gates are used instead of single transistors. The upper limit of accumulating voltage should not be effected by the threshold voltage of the n-type transistor. The threshold voltage needs to be the upper limit that the neuron can accumulate and for this circuit it has to be below the upper rail voltage minus the threshold voltage of the n-type transistor. The lower limit only effects decreasing the accumulated voltage. If the accumulated voltage is below the lower limit, it will not be effected by summing node voltages that are below it, but any summing node voltage above it will increase the accumulated voltage as intended.

The neuron's new accumulated voltage is based on the events of its synapstic inputs but is not determined solely by looking at the resistance of the memristors and adding in a value based on the difference of those resistance values. The output of the synapse creates a voltage via voltage division, and the change in the accumulated voltage is dependent in magnitude and direction on the current accumulated voltage. As the accumulated voltage aproaches the voltage on the summing node it decreases its change in voltage. The ability for the synapse to drive the neuron is also limited by the input transistors. Transistor M_{13} and M_{14} are used by the neuron to create the refractory period, but play a vital role in the charge accumulation. Transistor M_{15} is used to only pass in voltages when a synapse has activated. These transistors work together with the capacitor, C_{mem}, and the synapse to accumulate inputs.

$$V_{sum} = V_{on} + (V_{op} - V_{on})\frac{M_p}{M_p + M_n} \tag{3.1}$$

$$\Delta V_{mem} = \frac{1}{C_{mem}}\frac{V_{sum} - V_{mem}}{R_{in}} \times \mathrm{T} \tag{3.2}$$

3.2 Neuron Fire and Reset

Aside from accumulating inputs, the neuron must output a voltage spike and reset itself. The neuron should output a voltage spike and reset itself when a synapse fires into the neuron and causes the neurons accumulated voltage to cross a threshold. The threshold of this neuron is set by a differential amplifier that acts as a comparator. The differential amplifier, M_0 through M_4, outputs a low or high voltage, V_{cmpr}, based on the difference between its input voltages. When the voltage V_{mem} is less than the V_{thr}, the differential amplifier's output voltage, V_{cmpr}, is low. When V_{mem} goes above V_{thr}, V_{cmpr} goes to a high voltage. The differential amplifier gives the neuron an adjustable threshold based on the voltage applied to V_{thr}. Due to the nature of the input inito the neuron as described in section 3.1, the adjustable threshold acts as a weight shifting system. As the threshold is moved to lower voltages more synaptic weights have the ability cause the neuron to fire, and as the threshold is moved to higher voltages less synaptic weights can cause the neuron to fire. The highest appropriate threshold voltate is below the rail voltage minus the threshold voltage of the n-type transistor or below the maximum V_{sum} that can be produced. The differential amplifier's goals in the design are to output a high or low voltage depending on the accumulated voltage and threshold and to operate at the clock speed of the system. The system is designed to operate at a frequency of 20 MHz. Delay from the voltage V_{cmpr} delays the possibility of the output voltage spike. The differentail amplifier design transitions from a low to high output quickly to create the output voltage as fast as possible and allow the width of the output voltage spike to be close to the full clock cycle. In order to make the comparason, the differential amplifer needs to have enough gain to accurately portray the different outputs to the next stage. Because of these requirements, the widths of the p-type transistors M_0 and M_1, are increased to give higher small signal resistance and thus higher gain. The current provided by V_{bias} on M_4 is

high by having a a large voltage and aspect ratio, respectively. The transistors M_2 and M_3 are set to match and operate within the confines of the requirements. When a synapse causes a neuron's accumulated voltage to cross the threshold, the neuron creates a final internal analog voltage that signifies the digital output should high.

The neuron output is a spike voltage which is a digital logic high that lasts one clock cycle. This is acheived through the domino logic gate driving the output flip flop. The domino logic inverter works as a timing driven analog to digital one bit converter that only checks to see if the digital voltage should go high. As mentioned in 3, the domino logic amplifier, which is the domino logic inverter followed by a regular inverter, uses a keeper p-typer transistor, M_8. The leakage through M_6 and M_7 increases as the accumulated voltage approaches the threshold voltage, since the output voltage of the differential ampliefier increases. For the situation where the accumulated voltage is close to the threshold voltage the neuron should not fire. In order to hold the output low when there is a synapse fire that did not cause the neuron to cross the threshold, the keeper transistor M_8 must offset the leakage through M_6 and M_7. When the neuron does cross the threshold, the domino logic gate should switch. Because of the keeper transistor, the current drive of the transistors M_6 and M_7 need to produce a higher on current to drive the node F_{post_b} to a low voltage. This is accomplished by sizing M_6 to a larger aspect ratio.

While the keeper transistor helps mitigate a fire when the neuron has not crossed the threshold, it negatively impacts the required energy for the neuron to fire at a set delay. Because of this tradeoff the keeper transistor, M_8, is intentionally sized to a drive a small current. The small size allows the transistor to help offset the leakage, while not driving the current that M_9 must produce to create the output voltage signal up significantly. The p-type transistor, M_9, that pulls up the output voltage to rail and creates the spike needs to be much larger to overcome the keeper transistor and drive the output load. The input transistor into the domino logic, M_6, also must be able to drive more current than the keeper transistor, M_8, to generate the spike. The sizing of these two transistors directly effect the delay and energy of the output voltage spike of the neuron. The delay of the neuron is the time from the possibility of a fire, which starts at the clock cycle after a synapse fire, and when the output voltage F_{post} goes high. As with the differential amplifier, the domino logic

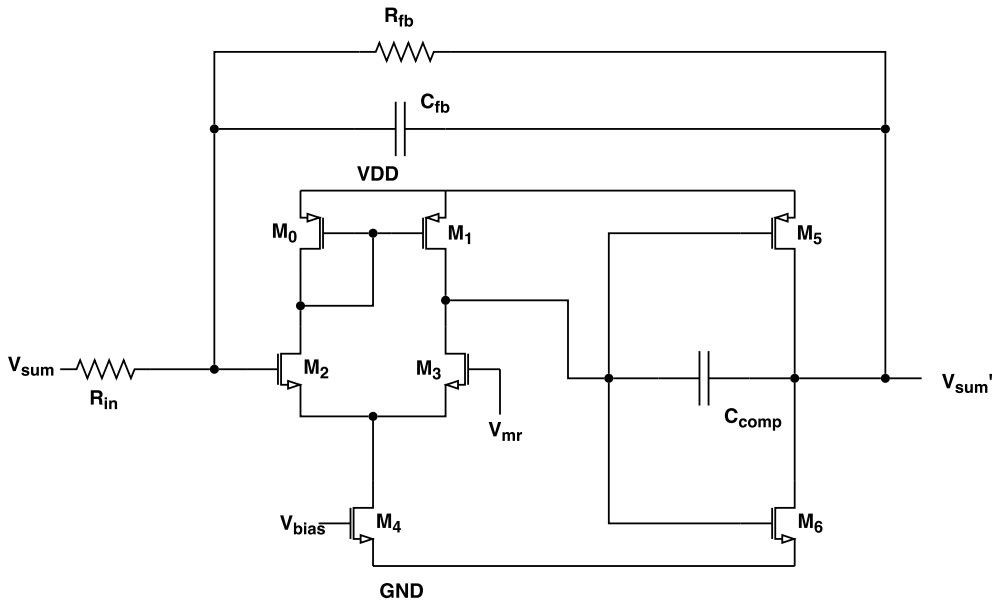

Figure 3.3: Input gain stage schematic for the neuron

amplifier is designed to be much faster than the frequency requirement so the pulse width of the output voltage spike is close to the entire clock period.

The output spike, F_{post}, drives the gates of the reset and refractory period transistors, M_{12} through M_{14}. When the output voltage goes high, the input transistors, M_{13} and M_{14} are turned off and block inputs into the neuron. At the same time, M_{12} discharges the capacitor, C_{mem}, to the reset voltage, V_{rst}. The current M_{12} can sink must be large enough to discharge the capacitor to the reset voltage within the refractory period. M_{12} is sized with a small aspect ratio to reduce its loading capacitance on the output of the domino logic because a larger transistor is unnecessary. As discussed in section 3.1, the aspect ratio of the input transistors is small to increase their resistance, but this also serves to reduce the load capacitance on the output.

3.3 Neuron Input Stage

As discussed in section 3.1, the input into the neuron is limited in part by the voltages V_{op} and V_{on} and the memristance states of the memristors, M_p and M_n. The voltages the summing node, V_{sum}, can reach, and the neuron's accumulated voltage, V_{mem}, are dependent on the

Table 3.2: Device sizes used for input stage

Parameter	Value
C_{fb}	44 fF
C_{comp}	18 fF
R_{fb}	40 kΩ
R_{in}	5 kΩ
M_0, M_1	$\frac{1.5\,\mu m}{180\,nm}$
M_2, M_3	$\frac{1.2\,\mu m}{120\,nm}$
M_4	$\frac{1.5\,\mu m}{120\,nm}$
M_5	$\frac{3\,\mu m}{120\,nm}$
M_6	$\frac{1.4\,\mu m}{120\,nm}$

possible resistance values of the memristors and their switching threshold. The difference between voltages V_{op} and V_{on} when the synapse is firing must be below the switching threshold of the memristors to not change their resistance value at an unintended time. Because of these relationships between switching threshold and memristance states, the neuron has a condensed range of applicable inputs. Another adjustment made due to the high level simulations is an input gain stage that increases the range of the summing node voltage. This is accomplished by a voltage gain amplifier as seen in figure 3.3. The amplifier circuit consists of a differential pair and a push-pull gain stage. M_2 is reference to be the mid-rail of the system, which due to the negative feedback drives the summing node close to mid-rail. The input stage is placed at the summing node and generates a new voltage, V'_{sum} to be applied on the capacitor in the neuron.

The summing node is now held close to the mid-rail becuase the positive voltage input into the differential amplifier, V_{mr}, is mid-rail. The voltages applied on the memristor, V_{op} and V_{on}, can now be as high as the switching threshold of the memristors. The output of the synapse into the gain stage is the sum of all the current differences between the two memristors. The accumulation of input fires now takes the sum of the synapse fires and creates a voltage at the output of the gain stage. The relationship of the resistance values of the memristors to their impact on the neuron is negated and amplified. This switches memristors relationship

to whether the synapse holds a positive and negative weight. The system is designed such that the output voltage of the gain stage can reach close to rail voltage when the maximum number of inputs fire at the maximum weight. The effect of the voltage created by the gain stage is still equation 3.2, but now V_{sum} is V'_{sum} which can reach closer to supply rail voltages because it does not depend on the maximum ratio memristor resistances but the gain of the input stage. The functionality of multiple fires changes from the average of the weight of the synapses to the addition of all synapse weights fired. The input resistor, R_{in} holds the place of an n-type transistor that would diconnect the neuron input stage during the learning event. Since the weight relationship to the memristor states are inversed, the voltages driving the memristors would have to be inverted in the synapse. This is accomplished by switching V_{op} and V_{on}. Ultimately, this circuit will be tested as another method along with the higher reset voltage to cause neurons to fire with fewer synaptic inputs.

Chapter 4

Results

This chapter covers the results from the circuit simulations of the neuron. Figures 4.1 shows the layout of the neuron, and table 4.1 shows the layout area for the neuron and some of the major components. The simulation results discussed in this chapter show the functionality of the neuron. The intended results of the functionality of the neuron as a whole system and the performance results of the neuron operation are described in section 4.1. The wave forms in the figures in this chapter are from simulations of the parasitic extracted layouts of the circuit. Aside from the proper functionality of the circuit shown in this chapter, the resulting energy, power and delay are tabulated. Section 4.2 describes the methodology for determining the energy and power usage for the neuron. Section 4.3 explains how the delay of the neuron is calculated. These results show the neuron fits the requirements of the system. Lastly, the effect and cost of the neuron input stage are examined.

Table 4.1: Layout area

Component	Area
C_{mem}	$84\,\mu m^2$
Differential Amplifier	$24\,\mu m^2$
Domino logic Amplifier	$17\,\mu m^2$
Total	$229\,\mu m^2$

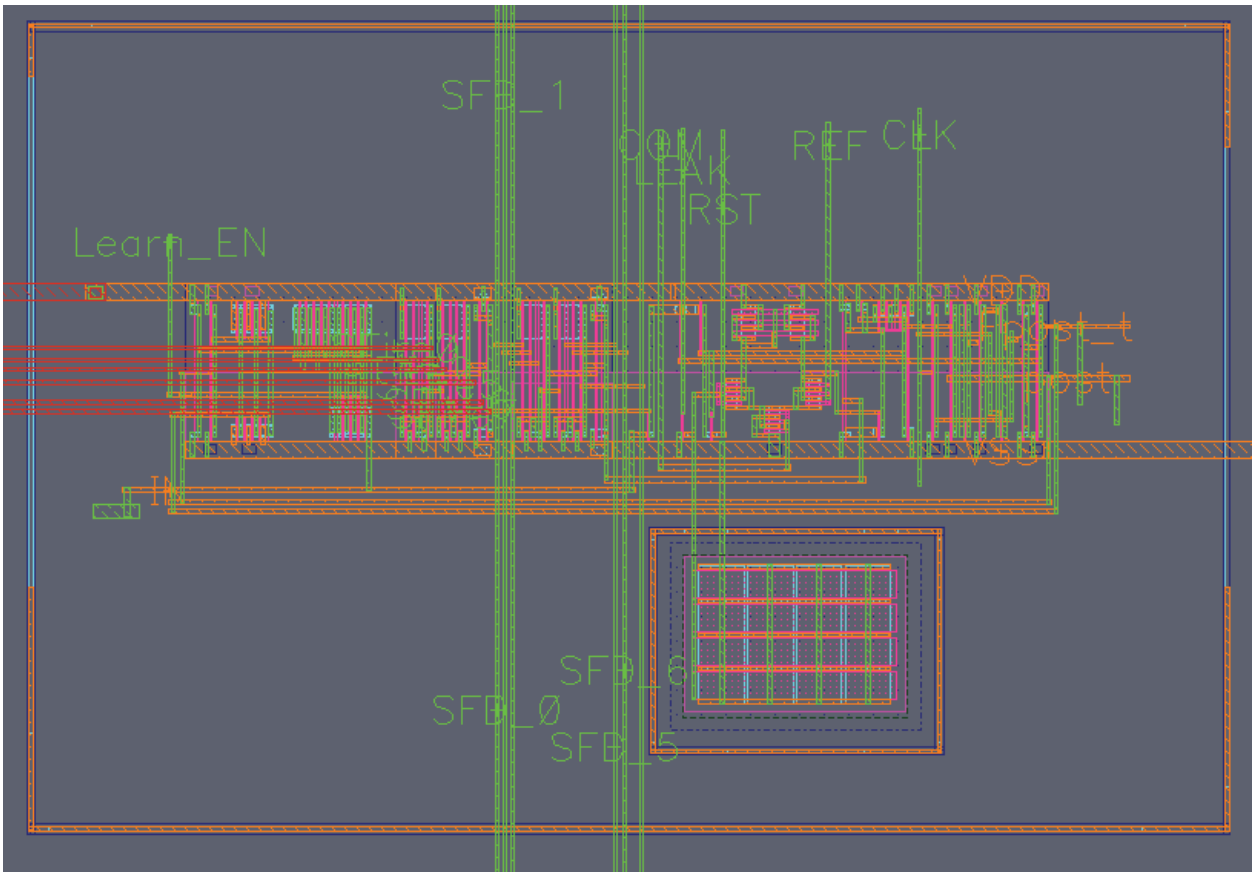

Figure 4.1: Layout of the neuron that is being fabricated

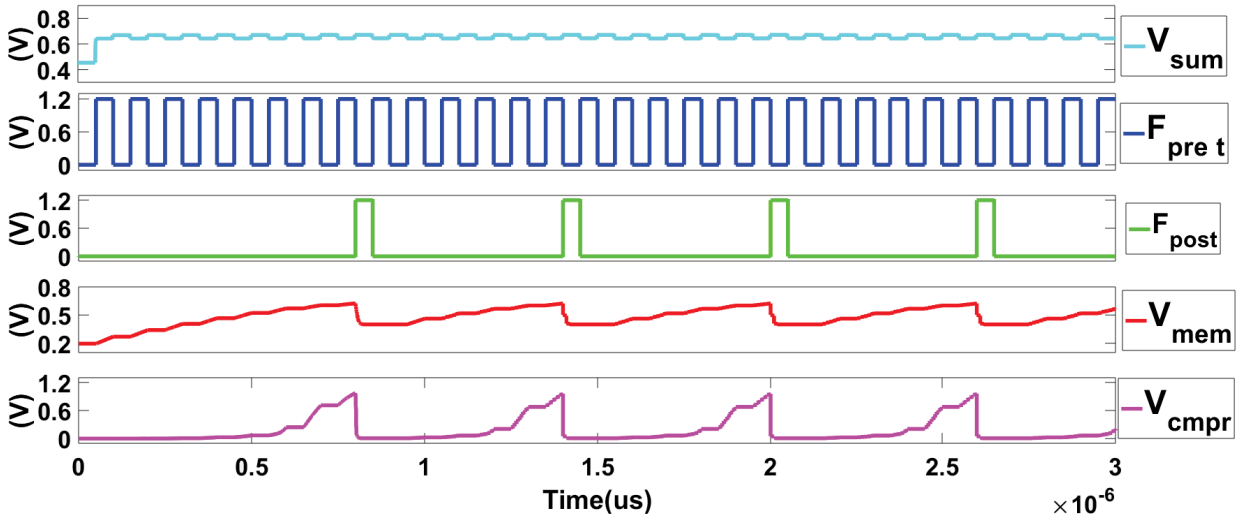

Figure 4.2: Waveform showing functionality of the neuron in the system with only positive weight synapse fires

4.1 Functionality

The results of the simulations of the extracted layout show proper functionality of the circuit. Figures 4.2 and 4.3 show waveforms depicting the operation of the neuron. The top line is the summing node voltage V_{sum}, which is driven to a voltage based on the resistance values of the memristors in the synapse. These simulations use resistors that are hardcoded to possible resistance values of the memristors. The resistors are set to 10 kΩ or 15 kΩ. Figure 4.2 only uses a positive weight synapse. The applied voltages on the resistors, V_{op} and V_{on}, are 800 mV and 400 mV. Solving equation 3.1, the resulting summing node should be 640 mV and the voltage at the summing node from the simulation is 641 mV. From figure 4.3, two other hardcoded synapses are used. One is set to a zero weight, which means the resistances are equal, and the other is set to a negative weight. The zero weight synapse should create a voltage of 600 mV, and the simulation shows V_{sum} is 597 mV. For the negative weight, the resistance values are reversed from the previous positive weight, and after solving the equation 3.1 the V_{sum} should be 560 mV. The voltage produced at V_{sum} by the negative weight synapse in the simulation is 553 mV. The second line is the delayed synaptic fire signal. This singal is a digital high for a clock period, 50 ns, the clock cycle after there is a synapse fire. This signal allows the neuron to output its fire. The third line is the output fire signal, F_{post}. This signal

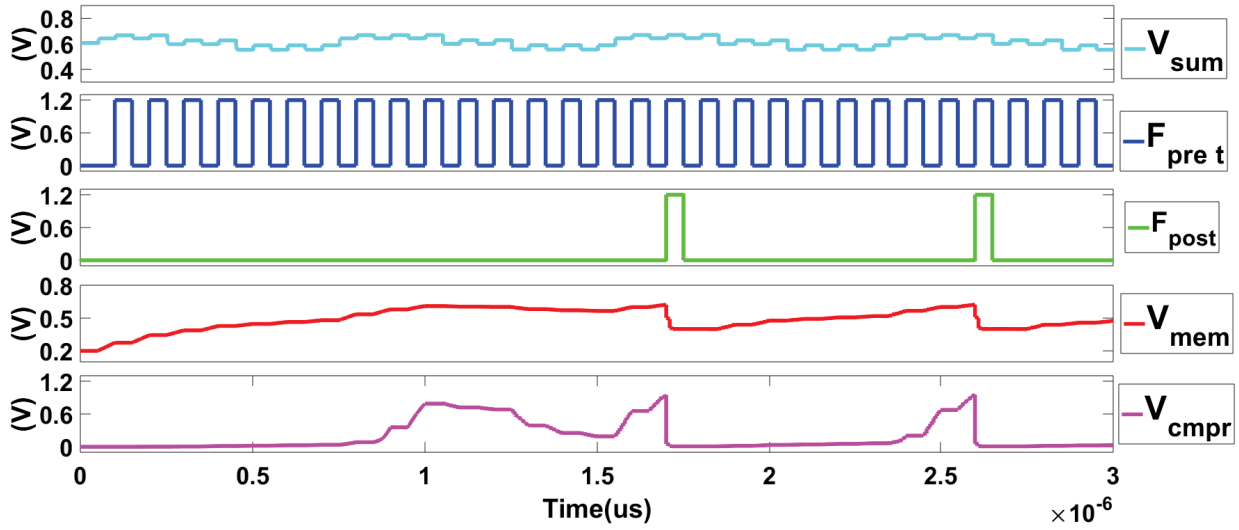

Figure 4.3: Waveform showing functionality of the neuron in the system with positive, zero, and negative weight synapse fires

goes to a digital high when a synapse fire has caused the neuron to accumulate a voltage above its threshold. From the figures 4.2 and 4.3, the output goes high only when the delayed synaptic input is high. The accumulated voltage, V_{mem}, is the fourth line. The voltage changes during a synapse input based on equation 3.2. To solve the equation, an estimated input resistance of 500 kΩ is used. At the reset voltage, 400 mV in these simulations, the accumulated voltage after a synapse for of positive weight should be 467 mV. From figure 4.2, the first synapse causes an accumulation of 61 mV resulting in a V_{mem} of 461 mV after the neuron has reset. The second change in voltage for figure 4.2 should be 50 mV, given V_{mem} is now 461 mV and V_{sum} is still 641 mV. The accumulated voltage after the second synapse fire in figure 4.2 is 517 mV, which is a change of 56 mV. Negative accumulations happen when V_{sum} is a lower voltage than V_{mem}. This situation occurs in figure 4.3, since zero and negative weight synapses are used. Solving equation 3.2 gives a decrease in accumultaed voltage when the accumulated voltage is greater than 560 mV. The negative weight synapses fire at 1.25 µs in figure 4.3 should decrease the accumulated voltage by 13 mV from solving the equation 3.2 given V_{mem} begins at 602 mV. The decrease in voltage in the simulation is 21 mV. The magnitude and direction of the change in accumulated voltage based on the synapse inputs from simulation show that the equation used to estimate the change in accumulated voltages properly approximates the input behavior. The main source of error in this approximation

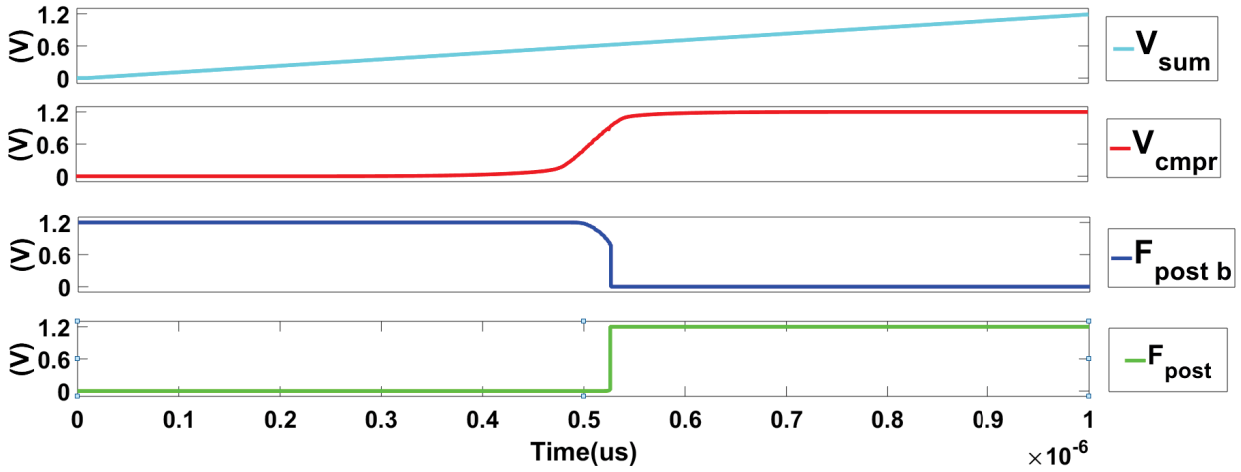

Figure 4.4: Waveform showing output fire event current

is the static estimated resistance value used. The last line in the figures 4.2 and 4.3 is the output voltage of the differential amplifier, V_{cmpr}. The threshold voltage used is 600 mV which sets a necessary accumulated voltage the neuron must acquire at V_{mem} in order to cause the output fire. The accumulated voltage the neuron must be above 620 mV which gives a V_{cmpr} of greater than 900 mV. When V_{cmpr} is above 900 mV, the current through M_6, M_7, and M_8 is large enough such that the voltage drop across M_8 causes the output to switch to a high voltage.

Another look at the firing mechanisim is shown in figure 4.4. This schematic simulation shows the the switching mechanism as the voltage V_{mem} is swept from a low voltage to a high voltage. A switching threshold voltage, V_{thr} of 600 mV is still used. The top line is the voltage V_{mem} which is swept up from ground to 1.2 V. The second line is the output of the differential amplifier, V_{cmpr}, which switches from 0 V to 1.2 V when V_{mem} goes above the V_{thr}. The third line is F_{post_b}, which switches from a high to low voltage. When V_{cmpr} has risen close to V_{thr}, the voltage at F_{post_b} decreases due to the current through M_6 and M_7 is pulled through M_8. When the voltage drop across M_8 reaches the switching threshold of the output inverter, M_9 and M_{10}, the output voltage, F_{post} goes low. The last line is the output voltage, F_{post}, which goes high. This occurs due to F_{post_b} going below the inverter threshold. Once the accumulated voltage is high enough above the threshold voltage to cause the switching, M_8 is turned off and F_{post_b} and F_{post} are supply rail voltages.

Table 4.2: Delay and energy for schematic and layout simulations

Parameter	Schematic	Layout
Energy per spike	130 fJ	138 fJ
Leakage Current	24 µA	26 µA
Average power	7.6 µW	7.9 µW
Delay	250 ps	500 ps

4.2 Energy and Power

The neuron design is intended to use low energy and power. The energy and power results from 4.2 show the resulting energy per spike and average power for the circuit. The energy per spike is the energy used to create the output voltage spike, F_{post}. From the simulation that produced 4.2, producing the output voltage on node F_{post} requires the highest energy consumption from the neuron. The maximum current pulled by the neuron is 100 µA, and occurs during the spike creation. The spike creation uses the most energy in the system because the domino logic amplifier must evaluate the voltage V_{cmpr} and drive the output load capacitance. The leakage current from 4.2 is the maximum leakage current caused by the neuron. The maximum leakage current occurs when the neuron has accumulated the highest voltage that does not produce a spike. In this situation, when the neuron would preduce a spike the leakage current is a combination of the leakage through the domino logic amplifier and the current through the differential amplifier.

The average power for the system has many factors. Since this is a mixed-signal circuit, static power and dynamic power are factors. The static power is primarily from the current through the differential amplifier, while dynamic power is from the domino logic inverter. Each synaptic event that does not cause a fire adds to the power, eventhough the output is not switching. This is due to the increased leakage through M_6 and M_7 as V_{cmpr} increases. The average power taken from the falling edge of a firing event to the falling edge of the next firing event from the simulation that produced figure 4.2. Improvements to power consumption can be made and are discussed in chapter 5.

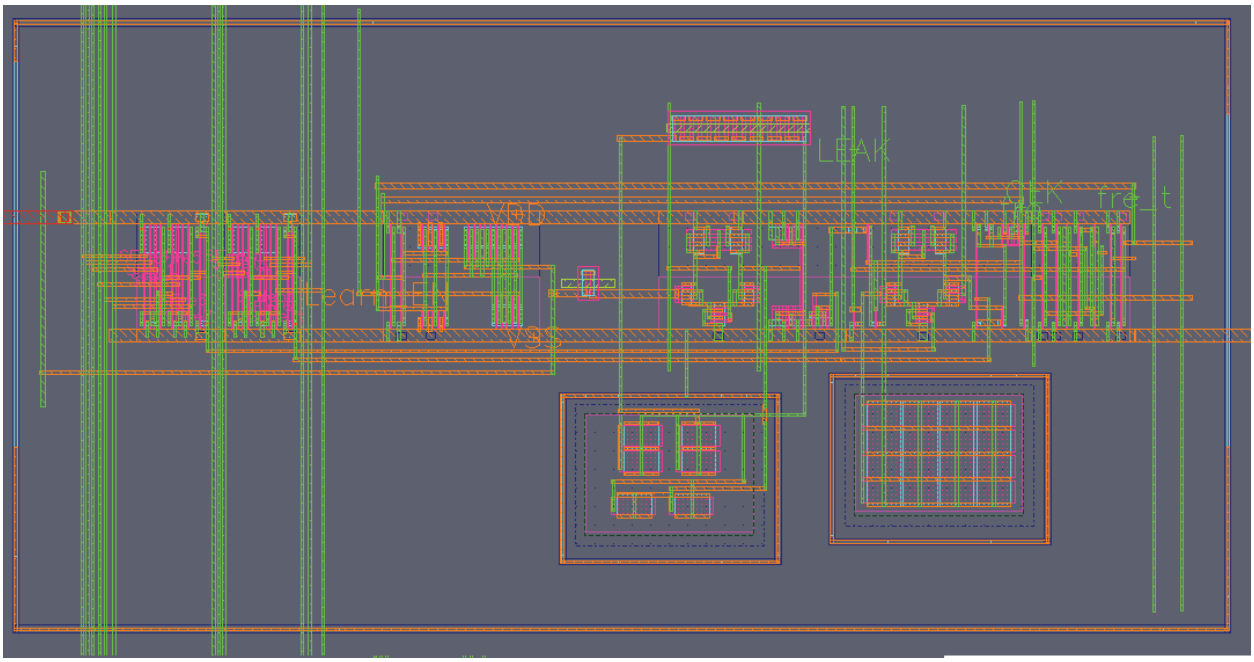

Figure 4.5: Layout of the neuron with the input stage that is being fabricated

4.3 Delay

The delay of the neuron is calculated as the latency of the output fire signal, F_{post}. The delay is shown in table 4.2. The delay is variable due to the analog voltage input into the domino logic amplifier, but is close to the values in 4.2. The lag from the rising edge of the clock and the output voltage spike effects the online learning operation of the system. Ideally the learning circuitry would evaluate the output voltage spike immediately, but this is unrealistic. The goal is to have a delay that is much smaller than the clock period, which allows the online learning circuitry to perfom the weight update. This goal is achieved because the neuron's simulated delay is 500 ps, which is two orders of magnitude less than the clock period.

4.4 Input Stage

The layout of the neuron with the input stage stage discussed in section 3.3 is shown in figure 4.5. The area used for the input stage and its major components are listed in table 4.3. The input stage adds a total area of 155 µm^2 to the 229 µm^2 for the neuron . Besides an increase in area cost, the neuron also adds additional power consumption. The input stage uses 36 µW

Table 4.3: Layout area of input stage

Component	Area
C_{fb} and C_{comp}	26 µm^2
R_{fb}	10.8 µm^2
Total	155 µm^2

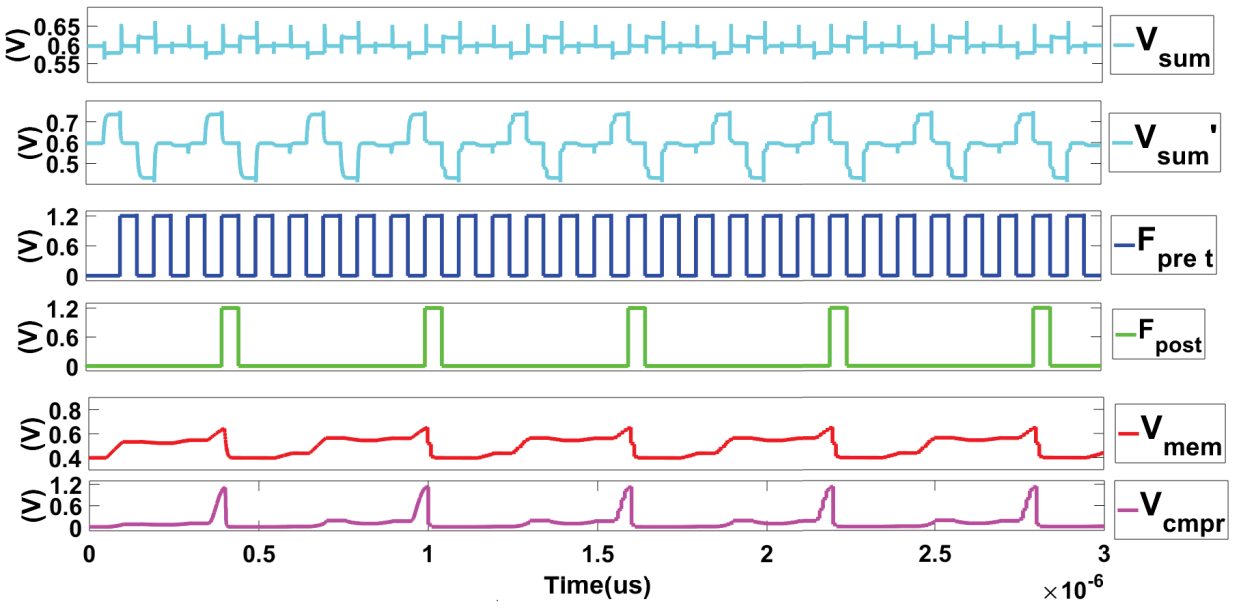

Figure 4.6: Waveform showing functionality of the neuron with the input stage in the system

of power nominally. This additional power cost is a tradeoff for the change in functionality which is the increases in the relative strength of the synaptic weights and the addition of the weights of all inputs firing at the same time. The same positive, zero, and negative weights for the simulations without the input stage are used with the input stage. The input stage neuron's waveforms can be seen in figure 4.6. V_{sum} is now a virtual ground node that is set close to mid-rail, and the new voltage applied onto the neuron is V'_{sum}. For the same weights larger voltages are produced because of the input stage and larger changes to the voltage V_{mem} can occur.

Chapter 5

Future Work and Conclusion

The continuation of this work consists of testing the fabricated chip and finalizing the high level model. The same circuit tests done in simulation need to be performed on the physical device. Testing the chip will be conducted in a probe station. The input voltages used in the simulations will be applied as stimulus and the responding output fires will be characterized. The tests will be repeated numerous times with different voltages for all the voltage inputs that are intended to be high level choices. Of these, the effects of the reset voltage, leakage voltage, and reference voltage need to be carefully characterized so that the high level model can accurately capture the different use cases. After characterizing the neuron for different stimuli, the final high level model can be implemented. The model should be parameterizable to match the possible configurations of the neuron's analog voltage stimuli. With the high level, model tests can be run to find the necessary size of the array to solve different problems. In characterizing the neuron, parameters like energy and power will be found and used to give estimates from high level simulations. A goal of the future high level model is to take the modularized neuron and generate networks that can prioritize energy consumption or minimizing area. Improvements for power will be considered from the physical results. One possible improvement to reduce power would be to turn of the differential amplifier when it is not in use. This would require the current mirror setting the differential amplifier's tail current to turn on and off at the appropriate times.

This axon hillock analog neuron circuit is developed for this neuromophic architecture on a process that can fabricate on-chip memristor for the synapses. The neuron shows proper

functionality with considerations made to reduce area and energy. While developing the circuit, possible improvements to the usage of this circuit are added to add flexibility to the possible network configurations. The neuron design takes into consideration the impactfulness of the memristors, while serving as a key component in a neuromorphic system.

Bibliography

[1] L.F Abbott. “Lapicque’s introduction of the integrate-and-fire model neuron (1907)”. In: *Brain Research Bulletin* 50.5-6 (Nov. 1999), pp. 303–304. DOI: 10.1016/s0361-9230(99)00161-6 (cit. on p. 3).

[2] S. Amer et al. “A Practical Hafnium-Oxide Memristor Model Suitable for Circuit Design and Simulation”. In: *ISCAS: International Symposium on Circuits and Systems*. Baltimore, MD, May 2017 (cit. on p. 7).

[3] Natalia Caporale and Yang Dan. “Spike Timing–Dependent Plasticity: A Hebbian Learning Rule”. In: *Annual Review of Neuroscience* 31.1 (July 2008), pp. 25–46. DOI: 10.1146/annurev.neuro.31.060407.125639 (cit. on p. 3).

[4] G. Chakma et al. “A Mixed-Signal Approach to Memristive Neuromorphic System Design”. In: *60th IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*. Boston, MA, Aug. 2017 (cit. on p. 7).

[5] L. Chua. “Memristor-The missing circuit element”. In: *IEEE Transactions on Circuit Theory* 18.5 (1971), pp. 507–519. DOI: 10.1109/tct.1971.1083337 (cit. on p. 6).

[6] Daniel Drubach. *The Brain Explained*. PRENTICE HALL, Sept. 11, 1999. 168 pp. ISBN: 0137961944. URL: http://www.ebook.de/de/product/3648593/daniel_drubach_the_brain_explained.html (cit. on p. 2).

[7] D. O. Hebb. *The Organization of Behavior*. Taylor & Francis Inc, June 12, 2002. 378 pp. ISBN: 0805843000. URL: http://www.ebook.de/de/product/3884835/d_o_hebb_the_organization_of_behavior.html (cit. on p. 3).

[8] A. L. Hodgkin and A. F. Huxley. “A quantitative description of membrane current and its application to conduction and excitation in nerve”. In: *The Journal of Physiology* 117.4 (Aug. 1952), pp. 500–544. DOI: 10.1113/jphysiol.1952.sp004764 (cit. on p. 4).

[9] Sung Hyun Jo et al. “Nanoscale Memristor Device as Synapse in Neuromorphic Systems”. In: *Nano Letters* 10.4 (Apr. 2010), pp. 1297–1301. DOI: 10.1021/nl904092h (cit. on p. 6).

[10] *Macmillan Encyclopedia of Physics*. Macmillan Library Reference, 1996. ISBN: 9780028645865. URL: https://www.amazon.com/Macmillan-Encyclopedia-Physics-John-Rigden/dp/0028645863?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0028645863 (cit. on p. 2).

[11] Carver Mead. *Analog VLSI and Neural Systems*. Addison Wesley Publishing Company, 1989. ISBN: 0201059924. URL: https://www.amazon.com/Analog-VLSI-Neural-Systems-Carver/dp/0201059924?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0201059924 (cit. on p. 5).

[12] *Methods in Neuronal Modeling - 2nd Edition: From Ions to Networks (Computational Neuroscience)*. A Bradford Book, 1998. ISBN: 0-262-11231-0. URL: https://www.amazon.com/Methods-Neuronal-Modeling-Computational-Neuroscience/dp/0262112310?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0262112310 (cit. on p. 4).

[13] Randall C. O'Reilly et al. *Computational Cognitive Neuroscience*. Wiki Book, 1st Edition, URL: http://ccnbook.colorado.edu, 2012. URL: http://ccnbook.colorado.edu (cit. on p. 2).

[14] S. Sayyaparaju et al. "Circuit Techniques for Online Learning of Memristive Synapses in CMOS-Memristor Neuromorphic Systems". In: *27th ACM Great Lakes Symposium on VLSI*. Banff, Alberta, Canada: ACM, May 2017, pp. 479–482. DOI: 10.1145/3060403.3060418 (cit. on p. 10).

[15] C. D. Schuman et al. "An Evolutionary Optimization Framework for Neural Networks and Neuromorphic Architectures". In: *International Joint Conference on Neural Networks*. Vancouver, July 2016 (cit. on p. 9).

[16] Catherine D. Schuman. "Neuroscience-Inspired Dynamic Architectures". PhD thesis. University of Tennessee, May 2015 (cit. on p. 2).

[17] P. Srivastava, A. Pua, and L. Welch. “Issues in the design of domino logic circuits”. In: *Proceedings of the 8th Great Lakes Symposium on VLSI (Cat. No.98TB100222)*. IEEE Comput. Soc. DOI: 10.1109/glsv.1998.665208 (cit. on p. 12).

[18] Dmitri B. Strukov et al. “The missing memristor found”. In: *Nature* 453.7191 (May 2008), pp. 80–83. DOI: 10.1038/nature06932 (cit. on p. 6).

[19] R. Weiss, G. Chakma, and G. S. Rose. “A Synchronized Axon Hillock Neuron for Memristive Neuromorphic Systems”. In: *60th IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*. Boston, MA, Aug. 2017 (cit. on pp. 10, 12, 13).

Appendix

A Abbreviations and Symbols

NIDA	Neuroscience-Inspired Dynamic Architecture
STDP	Spike Timing Dependent Plasticity
HRS	High Resistance State
LRS	Low Resistance State
MrDANNA	Memristive Dynamic Adaptive Neural Network Architecture

Vita

Ryan Weiss is from Nashville, Tennessee. He graduated from Father Ryan High School in 2012 and began his collegiate studies at the University of Tennessee, Knoxville, in pursuit of a Bachelor of Science degree in Electrical Engineering from the Department of Electrical Engineering and Computer Science in the College of Engineering. He has worked as a summer intern at Nissan. He has worked on graduate and undergraduate research at the University of Tennessee since 2015. HE graduated cum laude, with his Bachelor of Science degree in Electrical Engineering, in 2016. Ryan is pursuing a Doctor of Philosophy at the University of Tennessee.