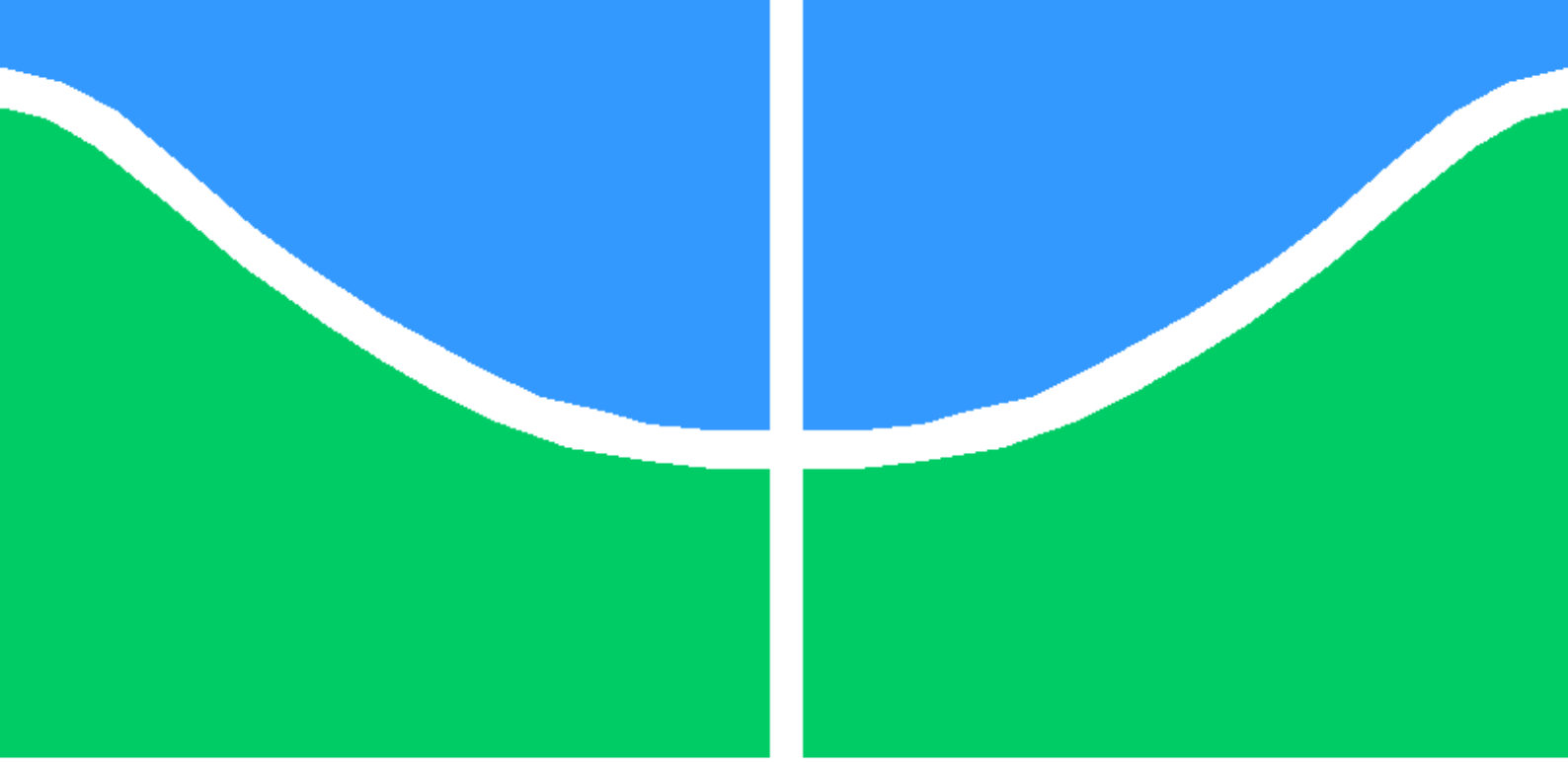


SPIKE NEURAL NETWORK ARCHITECTURE WITH MEMRISTIVE SYNAPSES USING PREDICTIVE CMOS MODEL

GABRIEL MARANHÃO SOARES

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA DE SISTEMAS ELETRÔNICOS E DE AUTOMAÇÃO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

SPIKE NEURAL NETWORK ARCHITECTURE WITH MEMRISTIVE SYNAPSES USING PREDICTIVE CMOS MODEL

GABRIEL MARANHÃO SOARES

Orientador: PROFª. DRA. JANAINA GONÇALVES GUIMARÃES, UFSC

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA DE SISTEMAS ELETRÔNICOS E DE AUTOMAÇÃO

PUBLICAÇÃO PPGENE.DM - 742/20
BRASÍLIA-DF, 28 DE FEVEREIRO DE 2020.

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

SPIKE NEURAL NETWORK ARCHITECTURE WITH MEMRISTIVE SYNAPSES USING PREDICTIVE CMOS MODEL

GABRIEL MARANHÃO SOARES

DISSERTAÇÃO DE MESTRADO ACADÊMICO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA DE SISTEMAS ELETRÔNICOS E DE AUTOMAÇÃO.

APROVADA POR:

Profª. Dra. Janaina Gonçalves Guimarães, UFSC
Orientador

Prof. Dr. Sandro Haddad, PGEA/UnB
Examinador interno

Prof. Dr. Daniel Guerreiro e Silva, PGEE/UnB
Examinador externo

BRASÍLIA, 28 DE FEVEREIRO DE 2020.

FICHA CATALOGRÁFICA
GABRIEL MARANHÃO SOARES
Spike Neural Network Architecture with Memristive Synapses using Predictive CMOS Model [Distrito Federal] 2020.
xxviii, 64p., 210x297 mm
(PPGEA/ENE/FT/UnB, Mestre, Engenharia de Sistemas Eletrônicos e de Automação, 2020)
Dissertação de Mestrado - Universidade de Brasília
Faculdade de Tecnologia - Departamento de Engenharia Elétrica
1. Neuromorphic systems 2. Spike neuron network
3. Memristor 4. STDP learning system
I. ENE/FT/UnB II. Spike Neural Network Architecture with Memristive Synapses using Predictive CMOS Model

REFERÊNCIA BIBLIOGRÁFICA

MARANHÃO, G. (2020) Spike Neural Network Architecture with Memristive Synapses using Predictive CMOS Model. Dissertação de Mestrado em Engenharia de Sistemas Eletrônicos e de Automação, Publicação PPGENE.DM-742/20, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 64p.

CESSÃO DE DIREITOS

AUTOR: Gabriel Maranhão Soares
TÍTULO: Spike Neural Network Architecture with Memristive Synapses using Predictive CMOS Model.
GRAU: Mestre ANO: 2020

_____

Gabriel Maranhão Soares
Rua dos Eucaliptos, Qd - 21, Lt - 9/20, Setor Goiânia 2
74.663-270 Goiânia - GO - Brasil.

Agradecimentos

Os agradecimentos principais são direcionados aos meus pais e irmão, pelo amor, incentivo e apoio incondicional, contribuindo diretamente para que eu pudesse ter um caminho mais fácil e prazeroso durantes todos esses anos.

Agradeço aos meus professores, que contribuíram para um melhor aprendizado, em especial a minha orientadora Profª. Janaína Guimarães, que tornou a realização desse projeto possível e me auxiliou em todas as etapas. Agradeço também a Universidade de Brasília e ao Departamento de Engenharia Elétrica a Faculdade de Tecnologia por terem me dado a chance e todas as ferramentas que me permitiram chegar hoje ao final desse ciclo de maneira satisfatória.

Por fim e não menos importante agradeço a minha namorada por todo o suporte e incentivo me dado nesses últimos anos, presente nos momentos mais felizes e nos mais difíceis sempre ao meu lado me ajudando. Agradeço também aos meus amigos que de uma forma ou de outra me ajudaram a chegar até aqui.

Resumo

O seguinte trabalho tem como objetivo propor e avaliar um circuito eletrônico que implemente uma arquitetura de rede neural com neurônios *spike*. Transistores do tipo MOSFET foram utilizados para implementar os neurônios. A rede proposta possui aprendizado do tipo *spike timing dependent plasticity* (STDP). Memristors foram utilizados como sinapses. A validação dos módulos de circuitos e do circuito da arquitetura completa foram realizados utilizando modelos SPICE dos dispositivos. Como a maioria dos dados sobre tecnologias de empresas possuem acesso restrito, algumas universidades fornecem modelos preditivos dos dispositivos com o intuito de reproduzir o comportamento real em tecnologias futuras, que foram empregados neste trabalho. Nesse projeto apresentamos dois tipos de *Integrate and Fire Neuron*(I&F) usando tecnologia CMOS de 32nm simulada no LTspice empregando o modelo BSIM4v4 concebido pela Universidade de Berkley e aplicando parâmetros preditivos fornecidos pelo *Predictive Technology Model* (PTM). Os resultados da simulação obtidos aqui, reduzem a tensão da fonte e o tamanho do chip em relação aos *designs* semelhantes mais recentemente implementado. Além disso, a comunicação entre neurônios e sinapses com um aprendizado STDP foi simulada com êxito.

Abstract

The following work aims to establish and evaluate an electronic circuit that implements a neural network architecture with *spike* neurons. This project uses MOSFET-type transistors to achieve the neurons, and the proposed network has a *spike-timing-dependent plasticity* (STDP) learning aspect. It applies Memristors to function as synapses. The validation of the circuit modules and the circuit for complete architecture were performed using SPICE models of the devices. Since most data of company technologies is restricted, some universities provide predictive models to reproduce the real ones. In this dissertation, we present two types of *Integrate and Fire Neuron (IF)* (IF) using 32nm CMOS technology simulated in LTspice with BSIM4v4 model designed by Berkley University and applying predictive parameters provided by *Predictive Technology Model* (PTM). The simulation results obtained here reduces the font tension and the chip size to the most recent designs implemented. Communication between neurons and synapses with STDP learning has been successfully simulated.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

$A-H$	Axon-Hillock
AFM	Atomic Force Microscopy
ANN	Artificial Neural Networks
$ASAP7$	Arizona State Predictive PDK 7nm
ASU	Arizona State University
$BSIM$	Berkeley Short-channel IGFET Model
$CMOS$	Complementary Metal-Oxide-Semiconductor
$DRAM$	Dynamic Random-Access Memory
$HKMG$	High Dielectric constant - Metal Gate
HP	High Performance
$I\&F$	Integrate and Fire
$IEDM$	International Electron Devices Meeting
$ITRS$	International Technology Roadmap for Semiconductors
LIF	Leaky Integrate and Fire
LP	Low Power
$LSTP$	Low Standby Power
MG	Multi-gate
$MOSFET$	Metal-Oxide-Semiconductor Field-Effect Transistor
PDK	Process Design Kit
PTM	Predictive Technology Models
$SRAM$	Static Random-Access Memory

$STDP$	Spike Timing Dependent Plasticity
$TiO2$	Titanium Dioxide
$VLSI$	Very-Large-Scale Integration

Chapter 1

INTRODUCTION

The main reason for a physics noble prize to John Bardeen and Walter Brattain in 1956, the creation of the transistor was a foundation for all existing electronic/semiconductor technology today. With the passing years, new and more powerful computational systems are appearing all the time, resulting in a cell phone from years ago being considered outdated. This exponential growth was theorized by Moore [1] and proved over the years. Although this growth is achieving its limits, computer systems are arriving at a physical barrier, the size of the transistor [2]. Different approaches are studied every year to increase the computational power of the systems without necessarily changing the transistors.

A lot of artificial intelligence systems are emerging over the years, demanding smaller and more energy-efficient computer systems. Many companies and universities invest money to create machine learning systems to solve unimaginable problems [3]. The leading problem is that linear processing units used in computers today were not initially developed to execute activities such as pattern recognition and learning mechanisms; neither are energetically efficient [4]. Computational systems that use the CPU with Von Neumann architecture are not optimized for these functions.

A functional processing unit capable of performing these functions quickly and with high energy efficiency is the human brain [5]. So why not study and create architectures capable of mimicking the learning aspects of a brain, using its most basic functions, neurons, and synapses. The human brain is a highly performing processing unit capable of storing, organizing, calculating, and executing various types of information at the same time [6]. It can recognize or recall data that has been "stored" for over 70 years. Although we use high-performance computing tools to facilitate the processing of digital information, none of them is capable of performing the same functions as a human brain, using so little energy in such a small space.

The neuron cell is responsible for numerous functions within the animals biological system. It can fire nerve impulses; in other words, the action potential [7, 8]. The Synapses are

the name given to the communication channel between two neurons and, each neurons make connections to a target neuron and excite or suppress their activity, forming circuits that can process information and carry out a response to adjacent courses. The synapses are strengthened by a a biological process called Spike-timing-dependent plasticity (STDP), it adjusts the connection strengths, between neurons in the brain, based on the relative timing of a particular neuron's output and input action potentials (or spikes) [9]. Fig. 1.1 shows how two neuron cells are connected through the synaptic gap. The action potential leaves the Axon from the cell body and travels through the Myelin Sheath arriving at the gap (synapse) [10]. At this point, the synaptic strength is adjusted and, the signal crosses it via neurotransmitters and activate or not the posynaptic signal.

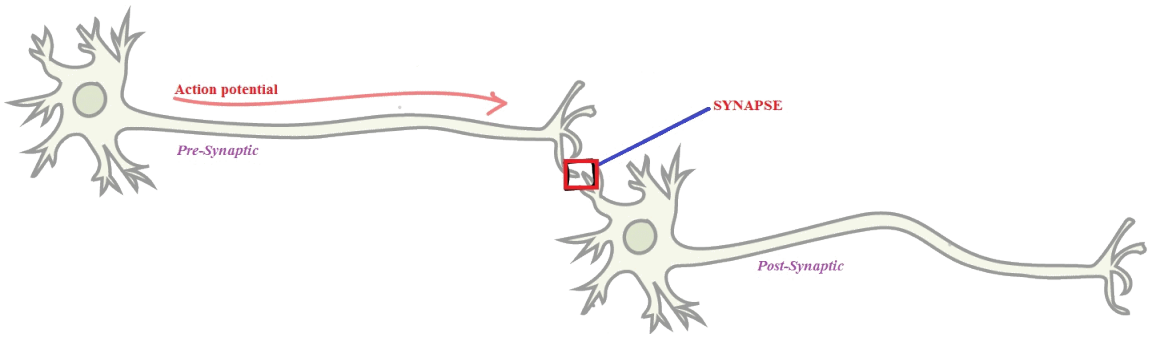


Figure 1.1: Simplified diagram of biological synapse interconnecting two neurons [8].

The novel microelectronic systems that try to reproduce that functionality of the human brain are commonly called neuromorphic systems [11]. There are numerous forms, already developed and detailed, of different architectures capable of reproducing the functioning of neurons and synapses. In the scientific literature, it is possible to find highly complex integrated circuits that try to simulate practically all biological characteristics. In contrast, other circuits are simpler and reproduce only the most basic functions, such as the action potential.

The need for computational systems with high processing capacity and, as always, greater energy efficiency led the studies of neuromorphic circuits to be, replaced or used alongside, the von Neuman architecture on conventional processors, where tasks that require neural networks would be processed. The development and implementation of microelectronic circuits with similar functions to biological neurons are presented in [12–20]. These works propose different approaches used in several practical applications, some more focused on the electrophysiology of neurons and others on the ability to create highly dense neural networks. One of these projects already use Memristors to implement synapses and show excellent results [17].

A large part of this novel brain-inspired architecture is implemented, simulated, and even manufactured using MOSFET transistors. The reason for this is because it is the most used device commercially. The availability of data and other works is enormous. There is a vast

collection of tools that work with MOSFET transistors, and also its has physical characteristics that benefit energy consumption, speed, and reliability of integrated circuits. Another device that works very well in neuromorphic systems is the novel memristors, also known as the fourth fundamental circuit element. The scientific bibliography shows its high capacity to simulate biological synapses between neurons [21]. Due to its ability to "remember," its resistive memory, making it useful to function as a device to represent synaptic weights.

The following work aims to establish and evaluate an electronic circuit that implements a neural network architecture with *spike* neurons using analog MOSFET neuron design. The proposed network has a *spike-timing-dependent plasticity* (STDP) learning aspect as it applies Memristors to function as synapses. In this dissertation, we present two types of *Integrate and Fire Neuron* (IF) [13–15] using 32nm CMOS technology, unlike the works in which they were proposed, simulated in LTspice with predictive MOS transistor models. The simulation results obtained here reduces power consumption, from $900pJ/Spike$ to $0.0485fJ/Spike$ and from $2fJ/Spike$ to $0.232fJ/Spike$, and also the chip size to the most recent designs implemented by using a smaller CMOS technology. Communication between neurons and synapses with STDP learning executed by the Memristors has been successfully simulated.

The subsequent dissertation was organized as follows: Chapter 2 introduces a brief revision of general concepts related to the devices used, MOSFET transistors and state-of-art memristors, and their application. Chapter 3 presents the architecture's designs that compose a neural network. Chapter 4 shows the methods used to simulate each block part of the system. Chapter 5 reveals the simulation result part by part comparing and evaluating it to other projects. Finally, the conclusion presents a quick summary of the work's achievements and what to expect for future projects.

Chapter 2

DEVICES AND MODELS

2.1 CMOS

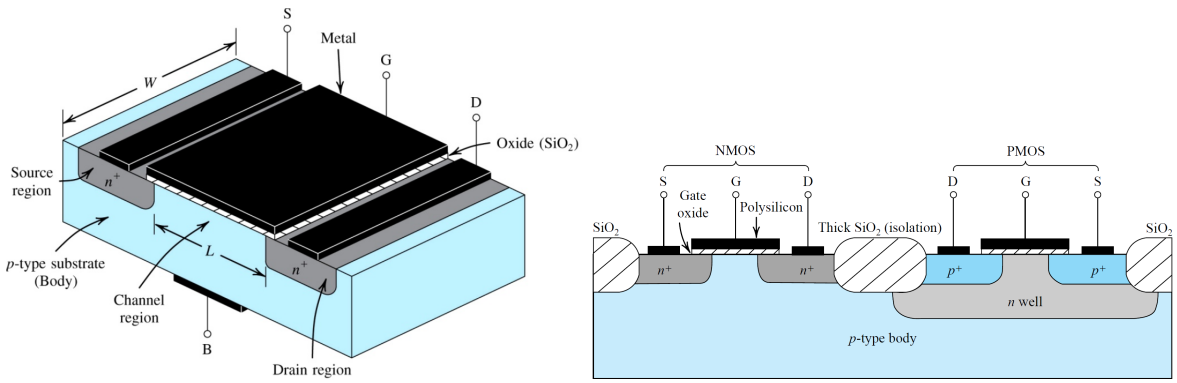
The technological electronic development that exists today has its main reason for the invention of the CMOS (Complementary metal-oxide-semiconductor) manufacturing process of analog and digital circuits that uses NMOS and PMOS transistors on a complementary and symmetrical way. That lead to the creation of the personal computer, and all side technologies came with it. The MOS transistor has become the most manufactured device in human history [22], with over 95% of all integrated circuits applying CMOS technology [23].

Over the years, the process of manufacturing CMOS integrated circuits has changed dramatically and continues to evolve. This development is supported by Moore's Law [1], which predicted the evolution of these systems through the number of transistors per chip and also their dimension, thus making these characteristics the main feature of the manufacturing process. Measured on metric scale, the so-called semiconductor manufacturing process nodes, in general, refer to the gate length of a transistor.

As the Fig. 2.1 shows, MOSFET transistors have four terminals, source (S), gate (G), drain (D) and body (B). The N-channel transistor is manufactured on a silicon p-type substrate ensuring support for the device. Two highly doped n-type regions represent the source and drain terminals, and there is a super-thin layer of silicon dioxide (insulator) between the p-type substrate and the metal layer of the gate. Simplifying its operation, when applying a voltage to the gate terminal, a depletion layer is created between the two doped type-n regions allowing the current flow between these terminals. The voltage value V_{GS} will control the current I_{DS}.

As the name says, complementary, the CMOS manufacturing process uses p-type and n-type transistors using a complementary approach. Fig. 2.1a presents the physical configuration of an NMOS transistor and its components, making it possible to note the characteristics such as gate width, transistor length, doping regions, and the insulating layer. Fig. 2.1b

exhibits the Cross-section of a CMOS integrated circuit. It is a general procedure of how a CMOS system is made on a single p-type substrate, using the region called n-well to include the p-type transistor [24].



a) NMOS transistor. b) CMOS Integrated circuit.

Figure 2.1: Physical structure of MOS transistors [24].

2.1.1 CMOS MODELS

The vast majority of integrated circuits use MOSFET (metal-oxide-semiconductor field-effect transistor) as the primary device, applying CMOS technology to construct the integrated circuits. Most of the commercial CMOS processes are kept secretly by companies. Thus many researchers dedicate their studies to recreate predictive models to be used by the academic community. The Arizona State University Predictive TechnologyModels (ASU PTM) [25] and the BSIM (Berkeley Short-channel IGFET Model) Group [26] are two of the best and better-studied models [27].

The difference between both BSIM Group and ASU PTM is that the first one provides a complete mathematical model, already implemented inside most SPICE based platforms, of the MOSFET technology, and the second uses this model as a base to their prediction of real physical parameters. Using BSIM and ASU PTM models, it simulates different types of transistors for different applications. These groups are continually evolving, introducing new models and technologies year by year, maintaining the progress projected by Moore's law [1].

In an attempt to continue using MOSFETS over time in addition to the size reduction of integrated circuits, researchers introduced new materials and technologies over the years. The establishment of a high-k insulator (dielectric constant) was added to the gate, with a dielectric constant higher than silicon dioxide, generating larger gate capacitance, suppressing the tunneling of carriers by the gate, thus reducing the current in the gate drastically [28].

A MOSFET ideals region of operation depends not unique but mostly by its threshold voltage (V_{th}) as it may be saturation, triode, and cut-off. Fig. 2.2 presents how these three

regions are achieved. In the Saturation regime, the transistor is saturated, itr will be biased so that the maximum amount of gate voltage is applied to the device, causing the depletion layer and electrical current through the switch to be maximum. The triode region is the operating region where the inversion region exists and current flows, but this region has begun to taper near the source. The potential requirement here is $V_{DS} < V_{GS} - V_{th}$. Finally, the cut-off is the moment where $V_{GS} > V_{th}$, that is, the gate to source voltage is not high enough to create a depletion layer so no current flows.

There is a special region of operation that can be fount are real transistors models. Differently from Fig. 2.2, a V_{GS} slightly less than V_{th} cause the transistor to work on sub-threshold region or weak inversion regime. On the weak inversion region, gate-to-source voltage is less than the threshold voltage. Ignored by digital implementations, the sub-threshold region of operation presents non-linear effects. An analog view of this domain holds valuable importance since transistors will work with small power consumption in addition to having an exponential gain [29].

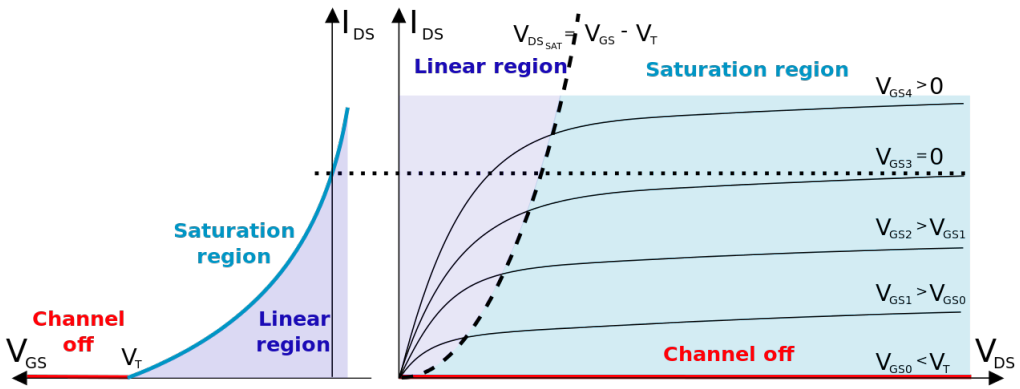


Figure 2.2: I–V characteristics and output plot of a n-type MOSFET.

Recently, data from industrial processes described at IEDM and the International Technology Roadmap for Semiconductors (ITRS) became available for many nodes. A study evaluating a wide array of prediction models was published, and some conclusions of the results comparing different planar and multigate processes to the industrial ones are presented in Table 2.1 [27].

Table 2.1: Discrepancies between PTMs and commercial processes.

Model	Divergence
ASU 1.0	I_{off} high below 45nm, P/N ratio high below 90nm PMOS gate leakage 100x lower than NMOS
ASU 2.0	I_{on}, I_{off}, Igate, P/N ratio, subthreshold slope high C_{diff} low
ASU 2.1 HKMG HP	Reasonably consistent with industry trends Bug in 45nm PMOS gate leakage model PMOS gate leakage 100x lower than NMOS
ASU 2.1 HKMG LP	FO4 delay slow at high temperature and does not scale with feature size
ASU MG HP 2012, ASU MG LSTP 2012	Ion higher than ITRS and commercial processes PMOS gate leakage 1000x lower than NMOS
ASAP 7	Reasonably consistent with industry trends

The Berkeley Group model is a detailed analytical model for SPICE simulations, physics-based, accurate, scalable, robust, and predictive MOSFET. It is already implemented in most SPICE based software using the last updated version from February 2017. This model establishes a group of parameters that define all the physical characteristics of the devices, such as effective oxide thickness, channel length, and width, threshold voltage, channel charge, sub-threshold swing, body current, capacitance's and so many more. It is not the goal of this design to give real values to these variables but to create the mathematical system that describes the action of each of these constants on the transistor operation.

Responsible for parameter characterization, the Arizona predictive model is the one to predict, based on calculations and comparison, all the system variables described by the BSIM Group. ASU PTM has an extensive data regarding different types of transistor modeling. It includes high performance and low power MOSFET of different fabrication processes along with CNT-FET and FinFET's. Regarding single-channel MOSFET's PTM ASU has devices available from 180nm to 16nm fabrication process. As shown on Table 2.1 of section 2.1.1 some configurations are not reliable comparing to commercial/industrial devices. Only two demonstrated characteristics similar to real ones, reported on [27], for planar processes, the chosen one is ASU PTM 2.1 HP HKMG 32nm. For not planar processes such as FinFET multi-gate, ASAP7 exhibits the best results.

2.2 MEMRISTORS

Until the 1970's, there were only three passive and fundamental circuit elements, resistors, capacitors, and inductors. Using all the mathematical equations that describe these three devices, Dr. Chua deduced and proved the existence of a fourth missing fundamental

component defined by the relationship between charge and flow. The Memristor (*memory resistor)* is a passive non-linear circuit element capable of retaining value after cutoff power supply [30]. Fig. 2.3 indicate the customarily used symbols for this device.

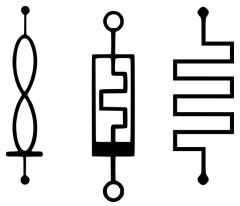


Figure 2.3: Memristor possible symbols.

The memory resistor mentioned above is not an energy storage device such as capacitors and inductors. At the moment the current goes to zero, the voltage also must be zero. Causing the I-V characteristic, shown in Fig. 2.4, of this device to be a pinched hysteresis loop, switching between different resistance states, demonstrating its memory effect [31]. Although the Memristor is mathematically proven, until 2008, memristance as a property of a known material was nearly nonexistent.

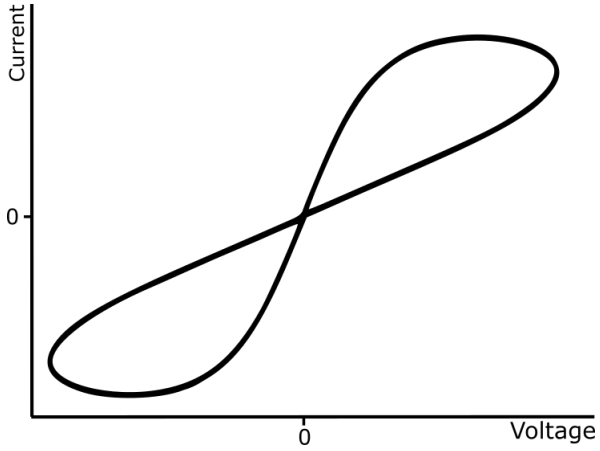


Figure 2.4: I-V Characteristics of Memristor.

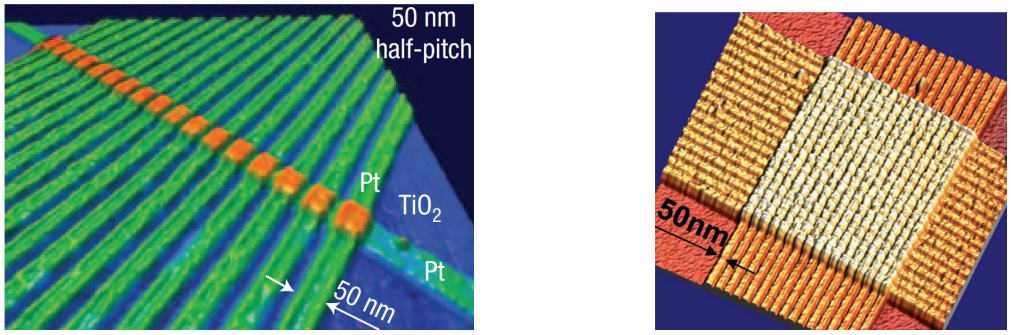
HP Labs introduced the first fabricated device with an analogous memristor mode of operation in 2008 [32], a thin-film titanium oxide device. They are composed of a stoichiometric titanium dioxide (TiO2) and an oxygen-deficient titanium dioxide (TiO2-x) layer compressed between two platinum electrodes. Since oxygen *holes* have low mobility, they tend to stay in the same position after any alteration on the arrangement provoked by a previously applied voltage, similar properties of a non-volatile memory [33].

2.2.1 Generalized model

Since the memristor model fabricated by HP labs [32], SPICE modeling of this device started to emerge [34–36]. It is making possible more accurate circuit simulations. Re-

searchers developed various models using equations that approximate the functionality of published memristor devices shown in Fig. 2.5.

An essential feature of a memristor and mainly used for characterization is the *state variable*. It provides a change in resistance based on the dynamics in each device, variations on Memristor physical proprieties. For example, the device produced by HP Labs [32] has its state variable represented by the thickness of the oxygen-deficient titanium dioxide layer. As the value increases, the overall resistance lowers.

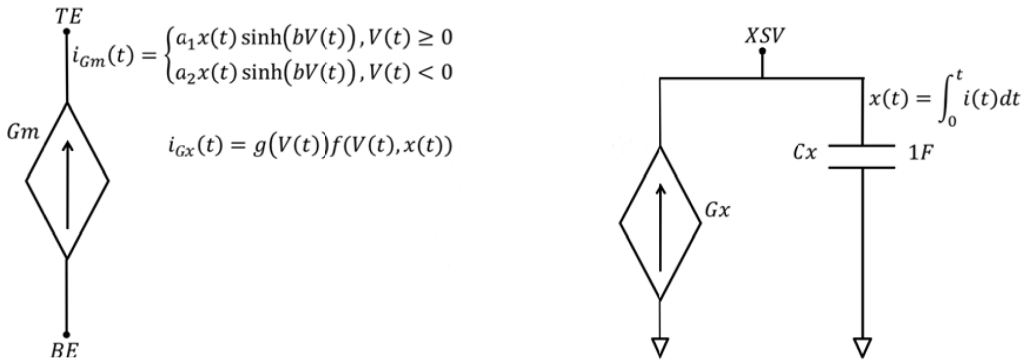


a) 1 x 17 Nano-crosspoint [37]. b) 17 x 17 Cross-bar [38].

Figure 2.5: Atomic force microscopy image of a Memristor devices with 50 nm half-pitch.

After studying several different types and distinct memristor applications, Dr. Chris Yakopcic developed a generic SPICE model correlated to multiple devices, where all the functions observed on the previous works were able to be re-created from this single model [39]. Fig. 2.6 present a two circuit model of Memristor proposed by Yakopcic.

Circuit on Fig. 2.6a is responsible for calculating the generalized I–V relationship for the proposed memristor SPICE model. The equation on this image is based on a hyperbolic sinusoid present on memristor operation, causing the device to have a change in conductivity beyond a certain voltage threshold. The parameters on this equation (a and b) are used to fit different device structures for different proposes. The state variable, $x(t)$, provides a change in resistance based on the dynamics off each device, a value between 0 and 1. The circuit on Fig. 2.6b presents the port XSV, it provide a convenient method for plotting the internal state variable.



a) Circuit to determine I-V relationship. b) Determine value of state variable.

Figure 2.6: Memristor SPICE model proposed by C.Yakopcic [39].

C. Yakopcic work [40] results in a mathematical model translated to SPICE configuration of a memristor. By modifying the parameters, such as, the thickness of the dielectric layer, conduction in the device, speed of ion motion, threshold voltages, and state variable motion, it is possible to reproduce different types of memristors.

2.2.2 Applications

Despite yet to initiate industrial process and industrial fabrication, the memristor market already values hundreds of millions of dollars [41]. That is because of the vast number of systems and applications this device could enhance, improving energy-efficient, area consumption, and versatility of some electronic systems. A few crucial applications are ROIC circuits(Read-Out Integrated Circuit), memory systems, and also state-of-the-art neuromorphic systems [42].

A ROIC circuit is used especially for reading photodetectors and commonly are placed both on the same chip. It integrates and amplifies the weak detector light signal into a analog voltage level. Normally these integrated circuits usually contain a capacitor responsible for integration and some additional control circuits. Consequently occupying a large area and limiting the space for the main area (sensitive detectors). A memristor could be used to replace these spacious capacitors, significantly reducing the unit cell size, making available larger area for sensitive detectors [43].

The main application and one that represents large part of memristor market value are the memory systems. After decades improving operation and reducing transistor size, they are reaching their limits. It is also known that a considerable percentage of chip area is dedicated to memory components made of transistor arrays, limiting chip area for processing units. Random-Access Memory are the ones circuits with memristor device could improve their performances.

The volatile nature of SRAM/DRAM leads to a problem, as soon the power supply goes

off the storage data are lost [44]. Nowadays the processing units require high-performance, power-efficient and high-density nonvolatile memory [45] . Because of their higher densities and lower leakage potential, memory resistor may be the solution to power, area consumption and volatility of some memory systems such as SRAM/DRAM [46].

As mentioned above the high density of a memristor is due to two facts, first their nanoscale dimensions, utilizing less area than flip-flops on SRAM/DRAM. Secondly it is an analog circuit element allowing more than one bit to be stored per unite cell such as "00", "01", "10" and "11" instead of only "1" and "0" [47]. Three different thresholds potentials would be used, making possible a 2-bit device showing on Fig. 2.7.

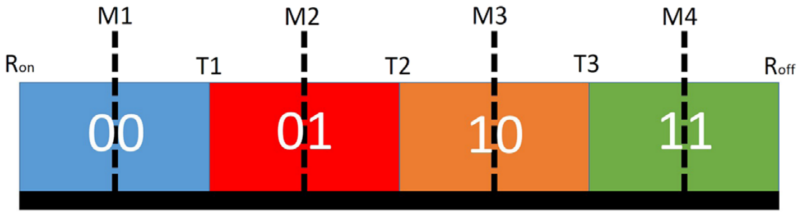


Figure 2.7: 2-bit memristor-based memory cell [47].

Finally, the newest operation mode related to a memristor is a building block for neuromorphic systems. Like electrochemical impulses that change synaptic weights on the brain tissue, voltage pulses applied to memristor modify its conductance, then it has the ability to continuously modulate the resistance over time. Fig. 2.8 shows a almost static sweeps where the resistance changes when a large voltage is applied. Essentially it operates similarly to a synapse, remembering the charge/flux ever to flow through it [48]. Making these devices favorable for massively parallel, large-scale neuromorphic systems [21].

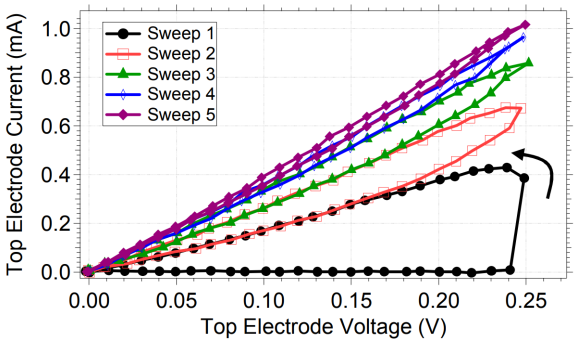


Figure 2.8: Successive voltages sweeps results on reduction of the device resistance [49].

The synaptic property of a memristor began to be widely studied by researchers using it together with CMOS neurons, mainly spike neurons, making viable the creation of state-of-the-art neuromorphic systems similar to biological ones [50]. A high density memristor/CMOS-spike-neuron crossbars, Fig. 2.9, may be used to perform parallel analog computations between a set of highly connected inputs and outputs [51].

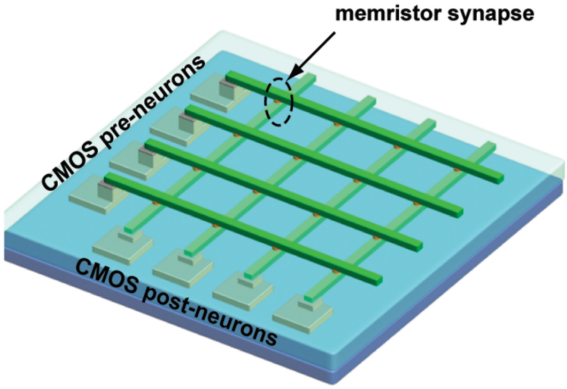


Figure 2.9: Crossbar structure of CMOS neurons and memristor synapses [21].

Memristors allow implementations that simulate the spike-timing-dependent-plasticity (STDP) ability of real synapses, providing an ideal architecture to machine learning operations direct on hardware. STDP describes the neuronal synaptic learning mechanisms, which were originally postulated as a family of computer learning algorithms, refining the Hebbian correlation-based plasticity proposed in 1949 [52]. It is responsible for adjust strengths between neurons based on time difference of post-synaptic and pre-synaptic spike, as shown in Fig. 2.10, the ideal STDP function of a synapse.. If *post* fires earlier than *pre*, the weights get stronger, otherwise get weaker [53–55]. This variation on weights, ΔW, can be expressed by the following equation:

$$\Delta W(\Delta t) = \begin{cases} W_+ e^{\frac{\Delta t}{\tau_+}} & (\Delta t > 0) \\ -W_- e^{\frac{\Delta t}{\tau_-}} & (\Delta t < 0) \end{cases} \tag{2.1}$$

Where W_+ W_- are the limit values of ΔW, and τ is the time constant that determines the weight update rate [56].

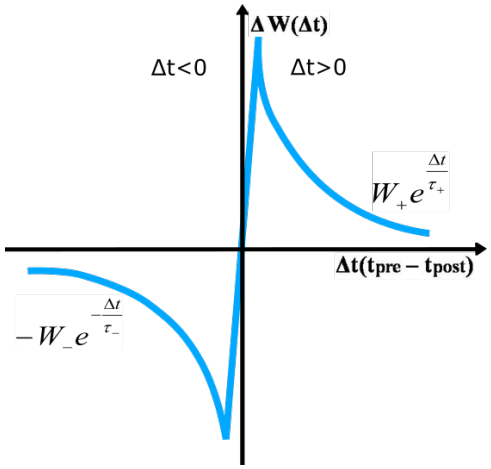


Figure 2.10: Synaptic weight modification in STDP on pre/post spike interval.

Chapter 3

NEURAL NETWORKS AND ARCHITETURES

Artificial Neural Networks (ANNs) are inspired by the brain. They are massively parallel architectures, composed of simple processors, called neurons, highly interconnected by the synapses [57, 58]. ANNs are specifically adapted to solving visual perception and dynamic control problems [58].

Energy-efficient hardware implementation of learning systems is a challenge when it comes to current von Neumann architecture-based computers. In order to substitute von Neumam systems, improve energy efficiency and also cost of computational systems Carver Mead first described, on early 90's, the *Neuromorphic Electronic Systems*, "biology-inspired" microelectronics, using the idea of an analog electronic architecture, applying very-large-scale integration (VLSI) process, that are able to perform activities equivalent to biological neurons presented in mammals [59].

According to G. Indiveri [60], analog VLSI technology is an appropriate process to create neurons and neuromorphic systems. There is a correlation among the biological neurons presented on animals and the analog VLSI neuromorphic systems such as conservancy of charge, integration, amplification and size reduction [61]. Collective computation in a densely parallel analog VLSI circuit probably is one of the most advantageously approach to process information on patter recognition or sound recognition [62].

During the last decade, called by IBM the era of cognitive computing [63], novel neuromorphic architecture start to emerge. DARPA, IBM and Intel are an example of an agency and companies that are committed to research and develop of extremely low power and small chip area neurocomputing integrated circuits. The SpiNNaker [64], TrueNorth [65] and Loihi [66] are a good example. Other smaller studies on this subject but with high performance and admirable result already using state-of-the-art memristor, described in Section 2.2, are also being developed e.g. [67, 68].

3.1 HARDWARE BASED NEURONS

In neuromorphic systems, all the information processed are assigned and stored in each neuron and their synapses, being responsible for the learning mechanism [69]. In order to start understanding the microelectronic neuron itself, it is necessary to distinguish the types of neuron models. Many different approaches of neuromorphic implementations were studied in the last two decades, using various kinds of circuits, neural functionality and styles to transfer data from neuron to neuron [70]. All of them trying to approximate the biological neuron behavior.

The need for faster, efficient, reliable and smaller systems lead the studies to choose for digital or analog approaches regarding design hardware neurons. Both paths/ has some pros and cons when comparing the following characteristics, ease-of-design, robustness, scaling and storage [12, 71, 72].

In terms of reliability and robustness, from the analog point of view, it demands a substantial care creating the design in order to minimize effects such as variations on power supply, leakage, temperature variations, unlike binary digital circuits considering it has only two states. [72]. Despite being harder to design, analog neuron circuits still require far less chip area than digital ones, as well as they are approximately 20 times more energy efficient considering that digital implementations demand high signal-to-noise ratio [12].

In addition to the discussion about analog or digital design, one more method of classifying microelectronic neurons operation refers to less or more biologically realistic. The first one is the Integrate and Fire (I&F) model and the other is the more complex cortical or conductance based model [70].

Initially the simpler, I&F model, has a straightforward implementation and non-realistic when comparing to biological system but successfully used to create dense networks and one of the most utilized model for analyzing neural network properties [73, 74]. It defines the neuron using the membrane potential $v(t)$ and applying Kirchhoff's current law for the conservation of charge:

$$C_{mem} \cdot \frac{dv(t)}{dt} = I_{leak}(t) + I_S(t) + I_{inj}(t), \tag{3.1}$$

The membrane is stimulated by post synaptic input deriving out of other neurons and their synapses, $I_S(t)$. A current $I_{leak}(t)$ is responsible for modeling the leak conductance and potential that decays over a constant time and lastly $I_{inj}(t)$, representing a external current stimulation. At the moment the membrane potential hit a fixed threshold mark, the spike is generated [74–76].

Alternatively, the conductance based neurons are more realistic because they have properties remarkably similar to those cortical neurons, by having protein molecule ion channels

represented through conductance and their lipid bi-layer by a capacitor, in other words, an equivalent circuit representation of a cell membrane [77, 78].

First proposed by Hodgkin and Huxley, the Hodgkin-Huxley model [79] represents a simple biophysical perception of an excitable cell in which current flows through the membrane due to charging of the membrane capacitance, I_C, and the action of ions across ion channels. Two ion channels, sodium and potassium, are characterized by I_{Na} and I_K respectively. Thus the membrane current I_m is given by:

$$I_m = I_C + I_{ion},\ where \tag{3.2}$$

$$I_C = C_{mem} \cdot \frac{dv(t)}{dt}\ ,\ I_{ion} = I_{Na} + I_K + I_{leaky} \tag{3.3}$$

3.2 ANALOG SPIKE NEURON

As part of neuromorphic systems the Analog Spiking Neuron represents a fundamental piece. Several methods to reproduce a spike (the main activity for communication between cells also knows as nerve impulse) of a biological neuron using VLSI have been described over the years. The action potential (spike) is the method of communication of neuronal cells, quickly carrying information between and within tissues, mainly inside the brain. When the excitation on neuron membrane reaches a specific threshold voltage, the *fire* occurs. The typical shape of an action potential is presented in Fig. 3.1.

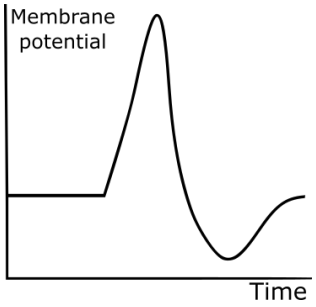


Figure 3.1: The pattern of a neuron action potential.

It is important to take into account, as stated above, that analog spiking neuron applying the I&F model has most details about the electrophysiology of real ones not included on the design, instead, its simplicity is a major advantage effectively describing the network mechanisms [80]. That simplicity resulted on it been the focus of theoretical and computational studies over decades and still extensively used and improved on large dense neural networks.

An example related to this kind of hardware neuron is the Axon-Hillock, first proposed by G. Indiveri [60], presented in Fig. 3.2, a simple implementation that can be useful to

create dense networks and study neural network properties [73]. It usually has two additional inverters acting as buffer for the output signal.

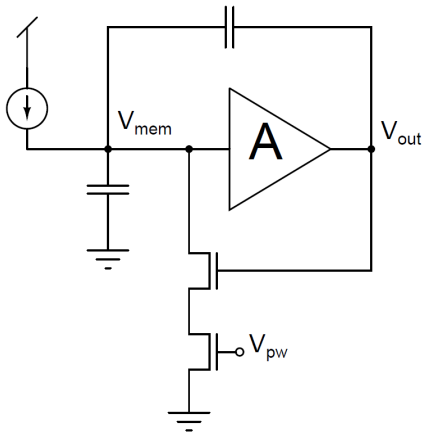


Figure 3.2: The Axon-Hillock Circuit.

The central part of a hardware neural network is the neuron itself. Responsible for adding multiple synaptic signals, and once exceed a triggering threshold, it executes the spike or action potential. The Compact leaky integrate-and-fire neuron circuits have a reasonable accuracy to the biological neuron and are a reliable abstraction of neurons. It is useful for neural learning mechanisms and requires a lower number of transistors to implement than cortical ones. In this section, two existing models of spike neurons are presented, and their modes of operation are demonstrated.

3.2.1 Regular Axon-Hillock

Originally schemed by G. Indiveri [60], the integrate and Fire neuron in Fig. 3.3 contains thirteen transistors, separated in four blocks, a differential pair (M1-M5), two inverters (M6-M9) and a reset circuit (M10-M11).

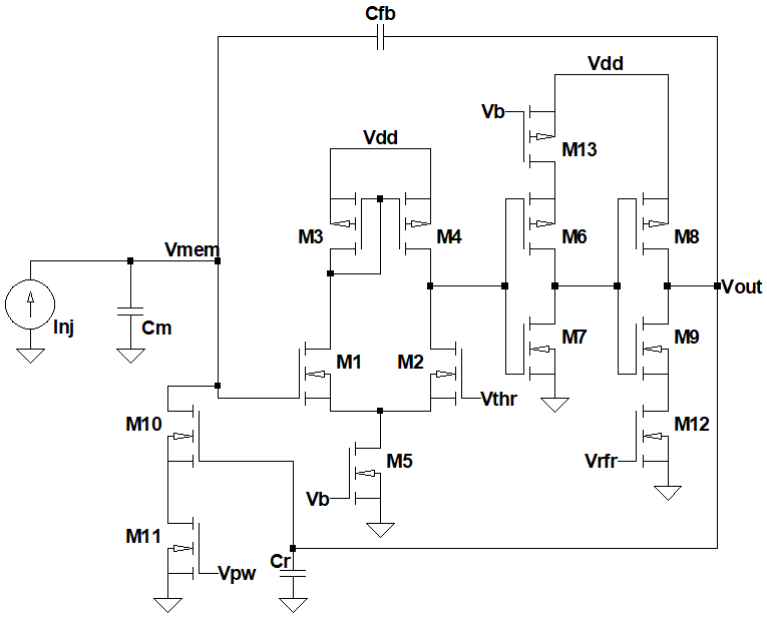


Figure 3.3: Regular A-H Integrate and Fire Neuron.

The circuit works as follows, injection DC current I_{nj} charges the membrane capacitor C_m causing V_{mem} to increase linearly with time. The differential pair works as a comparator, comparing the V_{mem} value with a fixed threshold value V_{thr}. While $V_{mem} < V_{thr}$ the output maintain a LOW state. After a time V_{mem} reaches a considerable value bigger than V_{thr}, the output goes to HIGH and through the two inverters working as buffers V_{out} also goes HIGH. At this moment a positive feedback created by the capacitor divider C_m and a feedback capacitor C_{fb} is activated because V_{out} turn on the reset transistor M10, allowing V_{mem} to increase with a rate given by:

$$\Delta V_{mem} = \frac{C_{fb}}{C_m + C_{fb}} \cdot V_{dd} \tag{3.4}$$

In the other side, while V_{out} stays HIGH, C_m discharges in a rate determined by V_{pw} applied on M11 gate. A reset current I_r goes in the direction of M10 drain terminal, if higher than I_{nj} the membrane capacitor discharging take V_{mem} down until became less than V_{thr} thus V_{out} returns to LOW. At this time the neuron enters in the refractory period, a moment where none spike are possible. With V_{out} at a LOW state the output of the first inverter is HIGH driving M9 to turn on and C_r to discharge through M12 at a rate set by V_{rfr}.

The V_{pw} is responsible to the pulse width of the output spike and the injection current will control both t_{low} (interval inter spikes) and t_{high} (pulse duration) that also depends of I_r, as shown below:

$$t_{low} = \frac{C_{fb}}{I_{nj}} \cdot V_{dd}\ ,\ t_{high} = \frac{C_{fb}}{I_r - I_{nj}} \cdot V_{dd} \tag{3.5}$$

3.2.2 Low Power Axon-Hillock

Analyzing the low power Axon-Hillock implementation proposed on [13], it has less transistors on circuit, composed by just two inverters (M1-M4) with the same feedback capacitor C_{fb} from regular A-H. Some modifications were applied to the original circuit trying to drastically reduce power consumption and achieve higher energy efficiency for each spike.

As presented on Fig. 3.4, for this design C_m, reset circuit, and threshold differential pair were fully removed. Each part has its own substitutes. The replacement for the membrane capacitance is the parasitic component of first inverter capacitance. The reset circuit that controlled I_r through V_{pw} was replaced by only one transistor M5, and the W/L ratio of M3 and M5 are now responsible for increase or decrease I_r. Finally the differential pair that evaluates whether or not the spike should occur are now replaced by the switching voltage of the first inverter.

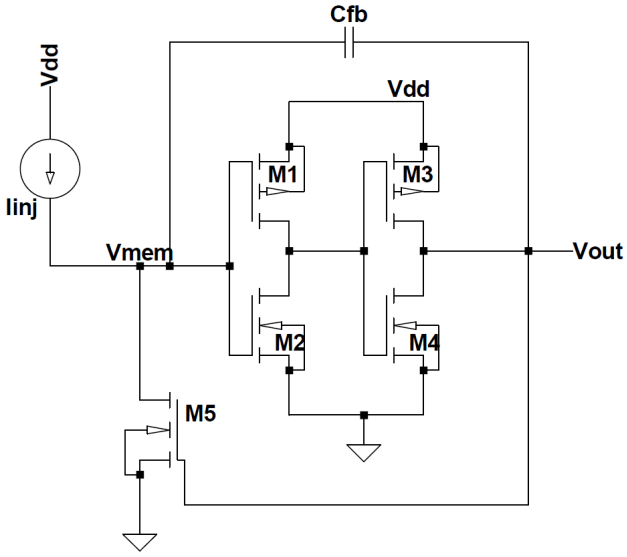


Figure 3.4: Low Power A-H Integrate and Fire Neuron.

The comportment is very similar to the regular neuron, the injection current I_{nj} increase charge and V_{mem} value. Once V_{mem} is sufficient high for the switching voltage of first inverter (V_{thr}) V_{out} goes high through the inverters. The C_{fb} works again as a feedback capacitor increasing V_{mem} and turning ON M5. At this point I_r is now controlled by the pull up (M3) and pull down (M5) transistors, causing V_{mem} to decrease over time and the cycle begin again.

3.3 MEMRISTOR SYNAPSE

The generic memristor SPICE model introduced on Session 2.2.1 is able to reproduce several kind of memristors. Through its parameterization, different operating characteristics

can be reached for different applications. The subcircuit code seen in Fig. 3.5 was developed mainly for LTspice using the circuit description already shown on Fig. 2.6. All the fitting parameters for this model are presented on Table 3.1. These variables are responsible for the three main characteristics which are electron tunneling, non-linear drift, and a voltage threshold.

The Subcircuit code on Fig. 3.5 is composed by four main groups, *Memristor Boundaries*, function to limit state variable motion. *Device Threshold* to ensure a point where the state variable starts to change. *State Variable motion* as the name says, describes how the SV motion will occur. Lastly the *I-V Response* a hyperbolic sine shape that leads the device to have a gain in conductivity above voltage threshold. A optional circuit at the end of the subcircuit establishes a third terminal only to compute the state variable outline.

```
.subckt mem_yako TE BE XSV

.params a1=0.17 a2=0.17 b=0.05 Vp=0.16
        +Vn=0.15 Ap=4000 An=4000 xp=0.3
        +xn=0.5 alphap=1 alphan=5 xo=0.11
        +eta=1

** Memristor Boundaries - functions to ensure
**  zero state variable motiom
    .func wp(V) = xp/(1-xp) - V/(1-xp) + 1
    .func wn(V) = V/(1-xn)

** Device Threshold - Function G(V(t))
    .func G(V) = IF(V <= Vp, IF(V >= -Vn, 0, -An*
    +(exp(-V)-exp(Vn))), Ap*(exp(V)-exp(Vp)))

** SV motion - Function F(V(t),x(t))
    .func F(V1,V2) = IF(eta*V1 >= 0, IF(V2 >= xp,
    +exp(-alphap*(V2-xp))*wp(V2) ,1), IF(V2 <=
    +(1-xn),exp(alphan*(V2+xn-1))*wn(V2) ,1))

** IV Response - Hyperbolic sine due to
** MIM structure
    .func IVRel(V1,V2) = IF(V1 >= 0, a1*V2*sinh(b
       *V1),
    +a2*V2*sinh(b*V1) )

** Circuit to determine state variable
** dx/dt = F(V(t),x(t))*G(V(t))

    Cx XSV 0 {1}
    .ic V(XSV) = xo
    Gx 0 XSV value={eta*F(V(TE,BE),V(XSV,0))
    +*G(V(TE,BE))}

** Current source for memristor IV response
    Gm TE BE value = {IVRel(V(TE,BE),V(XSV,0))}
.ends mem_yako
```

Figure 3.5: C.Yakopic Memristor SPICE subcircuit [39].

Table 3.1: Fitting parameters to model different Memristor devices.

Parameters	Description
a_1, a_2, b	Controls the I-V relationship
V_p, V_n	Positive and negative voltage thresholds
A_p, A_n	The magnitude of the exponential represents how quickly the state changes once the threshold is surpassed.
x_p, x_n	Points where the state variable motion becomes limited.
α_p, α_n	Rate at which the SV motion decays after passing the points xn or xp
x_o	Initial value for state variable
η	If 1 a positive voltage will increase SV value If 0 a positive voltage will decrese SV value

To demonstrate the operation of the system a simple circuit, Fig. 3.6, was set up to plot distinct I-V curves using different parameters. Fig. 3.7 presents two different curves for the same subcircuit where the famous hysteresis expected for a memristor are shown.

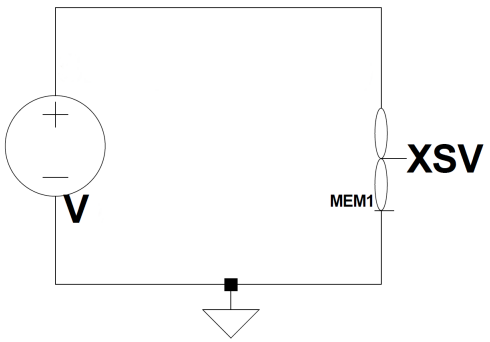
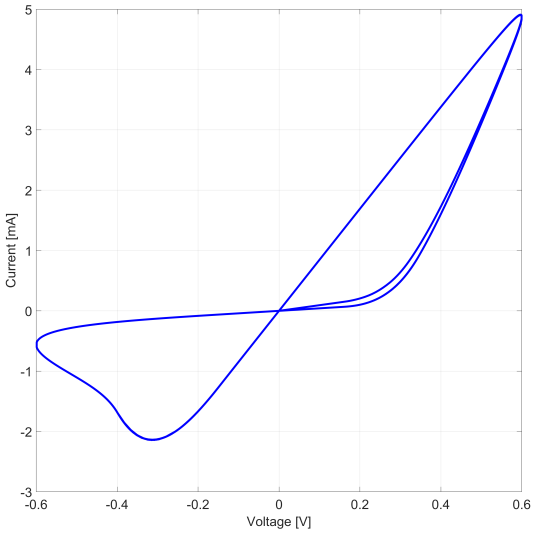
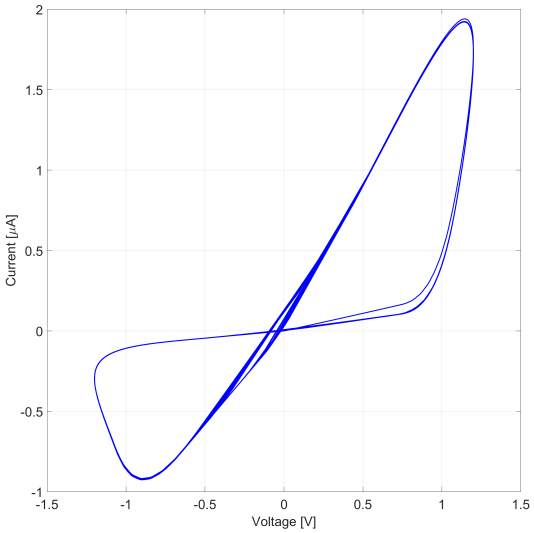


Figure 3.6: Basic circuit to extract Memristor I-V characteristics.



a) Parameters - $a_1 = 0.17, a_2 = 0.1b = 0.05, V_p = 0.16, V_n = 0.15, A_p = 4000, A_n = 4000, x_p = 0.3, x_n = 0.5, \alpha_p = 1, \alpha_n = 5, x_o = 0.11, \eta = 1.$



b) Parameters - $a_1 = 5.5 * 10^{-5}, a_2 = 3.5 * 10^{-5}, b = 0.04, V_p = 0.75, V_n = 0.65, A_p = 8 * 10^4, A_n = 8 * 10^4, x_p = 0.3, x_n = 0.5, \alpha_p = 1.2, \alpha_n = 2, x_o = 0.1, \eta = 1.$

Figure 3.7: Simulation of Memristor IV Characteristics.

Considering the main reason for using memristors during this work, to have equivalent behavior to neurological synapses (Fig. 2.8 and be able to perform STDP Learning mechanism. It is necessary to configure the presented model to a kind of device that acts the same way. By using a minimal threshold voltage from the spike to activate a change through multiple resistance states. Where each pre and post synaptic spike, works together to modify the synaptic weight, on this case, the state variable.

3.4 STRUCTURE OF A SPIKE NEURAL NETWORK

One of the main characteristics of a neural network is the way in which synaptic weights are updated. Each pre and post synaptic spike originated from a biological neuron starts the process of adjusting the weights based on the time variation between the pulses. The learning process previously explained, Spike-timing-dependent plasticity (STDP), requires at least one triad of neuron-synapse-neuron, since the synaptic weight will depend on the inputs and outputs. On the current work, the triad will be composed by a I&F Neuron-Memristor-I&F Neuron, as shown on Fig. 3.8.

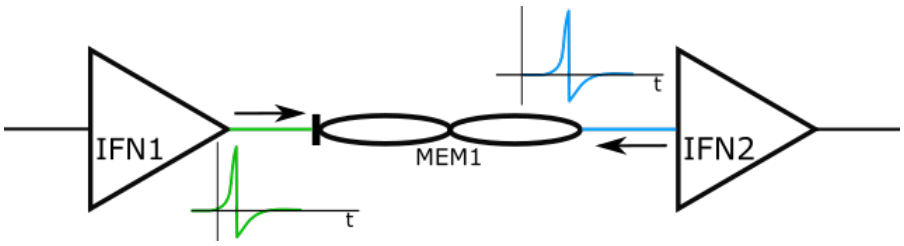


Figure 3.8: Pre- and post-synaptic pulse between spike neurons with memristor synapse.

The image shows a pre-synaptic green pulse and a post-synaptic blue pulse at different time, making possible the existence of a STDP curve (Fig. 2.10) during $\Delta t = t_{post} - t_{pre}$. The shape of action potential will highly influence the resulting STDP-Iearning function, it must have a narrow positive spike with a large voltage peak and a soften negative extremity.

After configuring the circuit block proposed by this work, the system could be increased to a crossbar array, the fully complete spike neuron network. Being also scalable, so large networks could be achieve using that setup. Fig. 3.9 represent four by four crossbar array composed of the main cell described above.

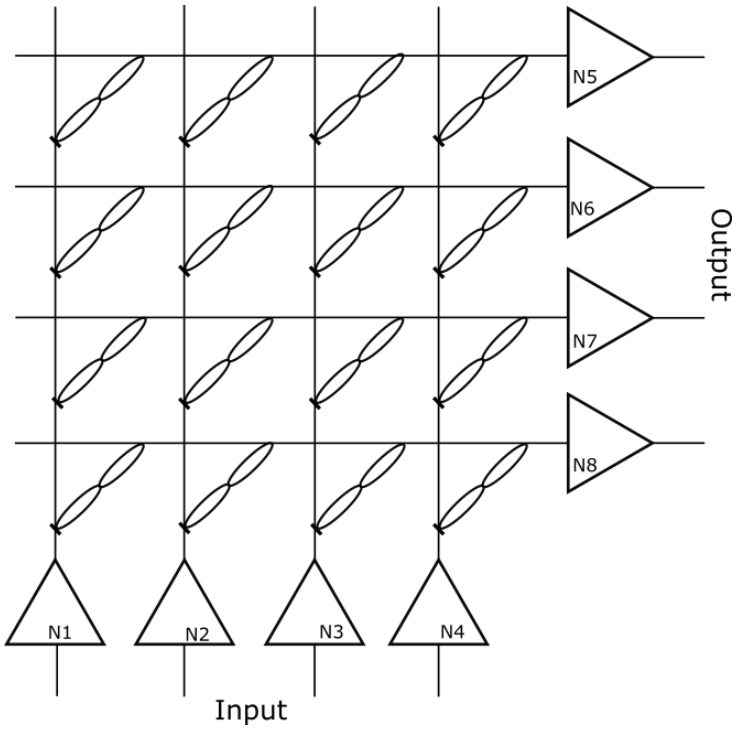


Figure 3.9: 4x4 Crossbar array network.

Chapter 4

METHODOLOGY

4.1 MODELS

To achieve the objective of developing a functional neuron and a synapse, an investigation was carried out in the scientific literature to evaluate the devices to be used during the work, the aim was to find systems that represent the same characteristics of real devices with reliability. For this, some types of transistors and memristors were studied to arrive at the optimal ones.

4.1.1 MOSFET

Not all models of MOSFET transistors used by circuit simulation software are real representations. Devices that have similar characteristics are used, but in extreme situations they may not work the same way, such as weak inversion, low or high temperatures, low power supply. Other models are so complex that they require software capable of 3D simulations of all transistor parts in order to execute all functions, being impossible to simulate the circuit in the LTspice XVII software that was used in this work.

The choice of transistors to be used in this work was based on proximity with commercial ones, availability and adaptability to the LTSPICE platform bringing the whole circuit closer to reality. As these industrial models are hidden by companies, it was necessary to select mathematical and predictive systems that simulate all the physical characteristics of a transistor. Composed of more than 220 variables for modeling the BSIM4 [26] and ASU PTM [25] form a refined model for simulations.

The BSIM4 model is more modern than BSIM3, considering that the last one stopped being updated in 2005 and the most recent update for BSIM4 was in 2017 adding changes suggested by the Compact Model Coalition (CMC), a group responsible to standardize the design of semiconductor device models. However, there are still more contemporary models,

the multi-gate devices, such as FinFET also available at BSIM Group site as BSIM-CMG. The problem is the language in which this model was implemented is no longer SPICE, it uses Verilog-A. LTspice XVII is not prepared to compile such type of file. A few tests using FinFETs were made using ngSPICE, also a freeware, but they are very limited and unreliable differently from BSIM4 on LTspice.

The parameter characterization file provided by ASU PTM was chosen based on two works [27, 81] that evaluated different files from this group. As already shown on Table 2.1 some predicted models are more accurate than others, remaining to be evaluated these two models, ASU PTM 2.1 HP HKMG 32nm and ASAP7. The first one is refers to a device applied on planar CMOS process. The ASAP7 is a Process design kit (PDK) with a set of files describing FinFET for multigate process and as already said, do not work with LTSPICE.

4.1.2 MEMRISTOR

As an extremely new circuit element, commercial manufacturing models do not yet exist. Some prototypes already presented were created to physically represent the functioning of a memristor. These physical models produced are usually adapted for only a single function as a memory, ROIC circuits or synapses. Although they are functional and reliable, there is a certain complexity in translating these devices for SPICE simulations.

As an alternative for those physical models and the availability to implement with SPICE an already presented on Section 3.3 the memristor by C.Yakopic [39] was chosen to be used as the synaptic circuit part. This specific design is responsible for translating the equations of scientific literature from different configuration and grouping them all in the so-called Generic Model, capable of representing with equivalence different types of memristors, in addition to being all developed for the LTSPICE platform. With a simple adjustment of parameters a Generic memristor functioning as a synapse capable of STDP learning can be implemented.

Over the years some researchers were able to create real memristors with the same ability previously mentioned [21, 49], ideals for sweep behavior and power efficient when driving STDP learning operation. Using these real memristor models as exemplars and fitting the SPICE model presented on this session to have the same characteristics, it is possible to achieve the ideal hysteresis with sweeps as shown early on section 2.2.2 Fig 2.8.

4.2 CIRCUITS

During the bibliographic review phase several neuron representation circuits were evaluated to be used. An initial study was made about implementations of cortical neurons, those

that try to represent the electrophysiology of a real neuron [82,83] they have a greater implementation complexity and require more computational power for simulations. The objective of the work is to create a neural network that uses several neurons, the need to apply highly complex models was not necessary, since the intention is not only to represent the physical characteristics of a neuron.

Finally, two types of I&F neurons were chosen to be implemented due to their simple functioning and easy implementation, thus enabling them to function together with synapses and in a higher neuron density as well. The circuits shown in section 3.2, the Regular A-H on Fig. 3.3 and the Low Power A-H on Fig. 3.4 were used as shown on the images. In an attempt to improve these circuits, all of their transistors found in them were replaced by the model presented on section 4.1.1.

4.3 SIMULATIONS

The most common and first approach used by companies and universities to simulate the behavior of integrated circuits is SPICE, capable of simulating the circuit and all the characteristics of a transistor before committing to manufacture an integrated circuit. The simulations performed in this work use the LTspice XVII software and the collected data were manipulated with MATLAB. A hierarchical methodology were used to perform the simulations, starting with minor blocks and ending with the major project.

A brief diagram is shown in Fig. 4.1, showing how the simulation hierarchy went step by step. Starting with the most basic simulations of the NMOS and PMOS transistors and ending with neuron-synapse-neuron triad. First a MOSFET evaluation to ensure the expected operation, it must have a high gain in weak inversion regimes and use a low supply voltage. Moving on to testing the neuron circuits with the selected transistor model, observing parameters for better operation and making adjustments if necessary. After that, neuron simulations were paused to find and test the best way to recreate synapses. The first tests with the memristor were to guarantee its functioning on the LTSPICE platform, following by its synaptic configuration, using several parameters to find the ideal curve. Finally, the two main blocks were placed to work together, testing final adjustments to ensure the best STDP curve to update the synaptic weights.

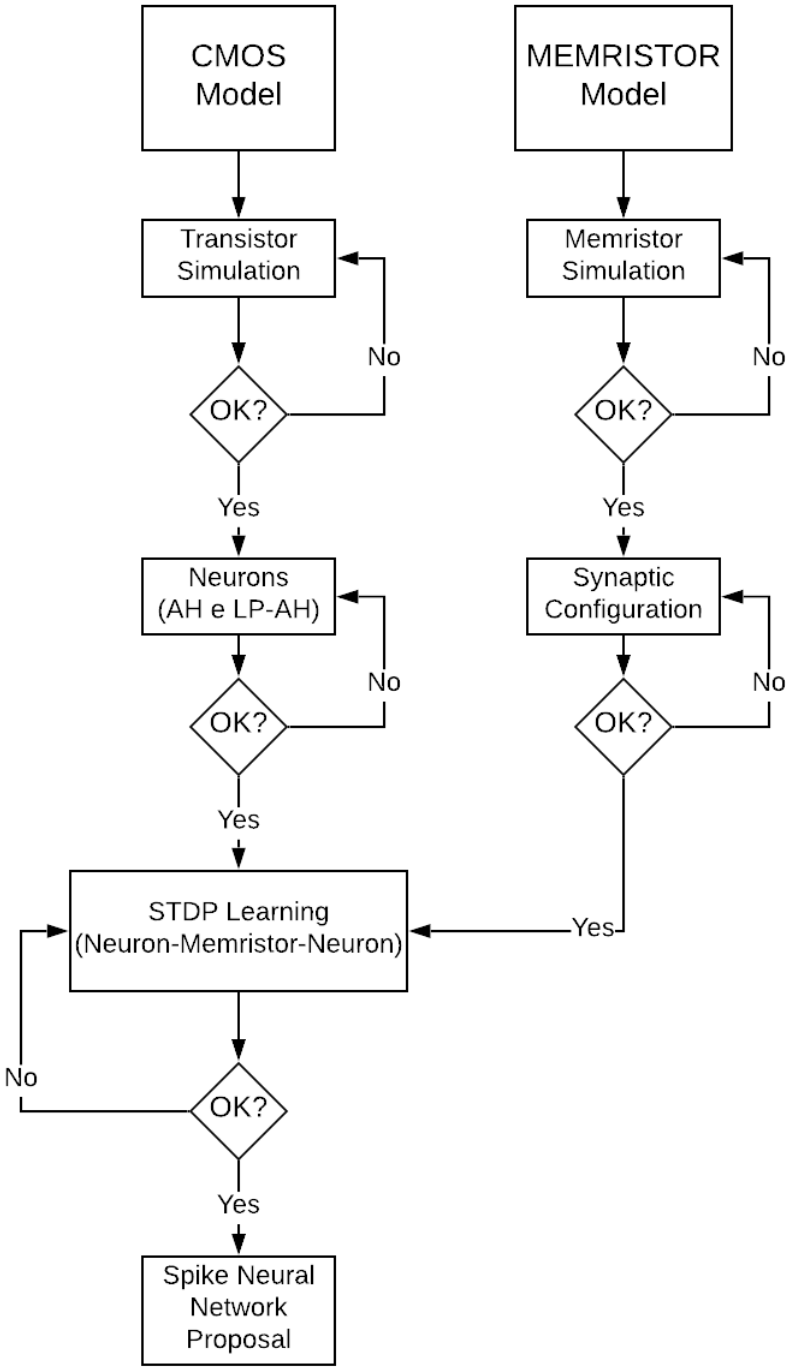


Figure 4.1: Simulations flow chart.

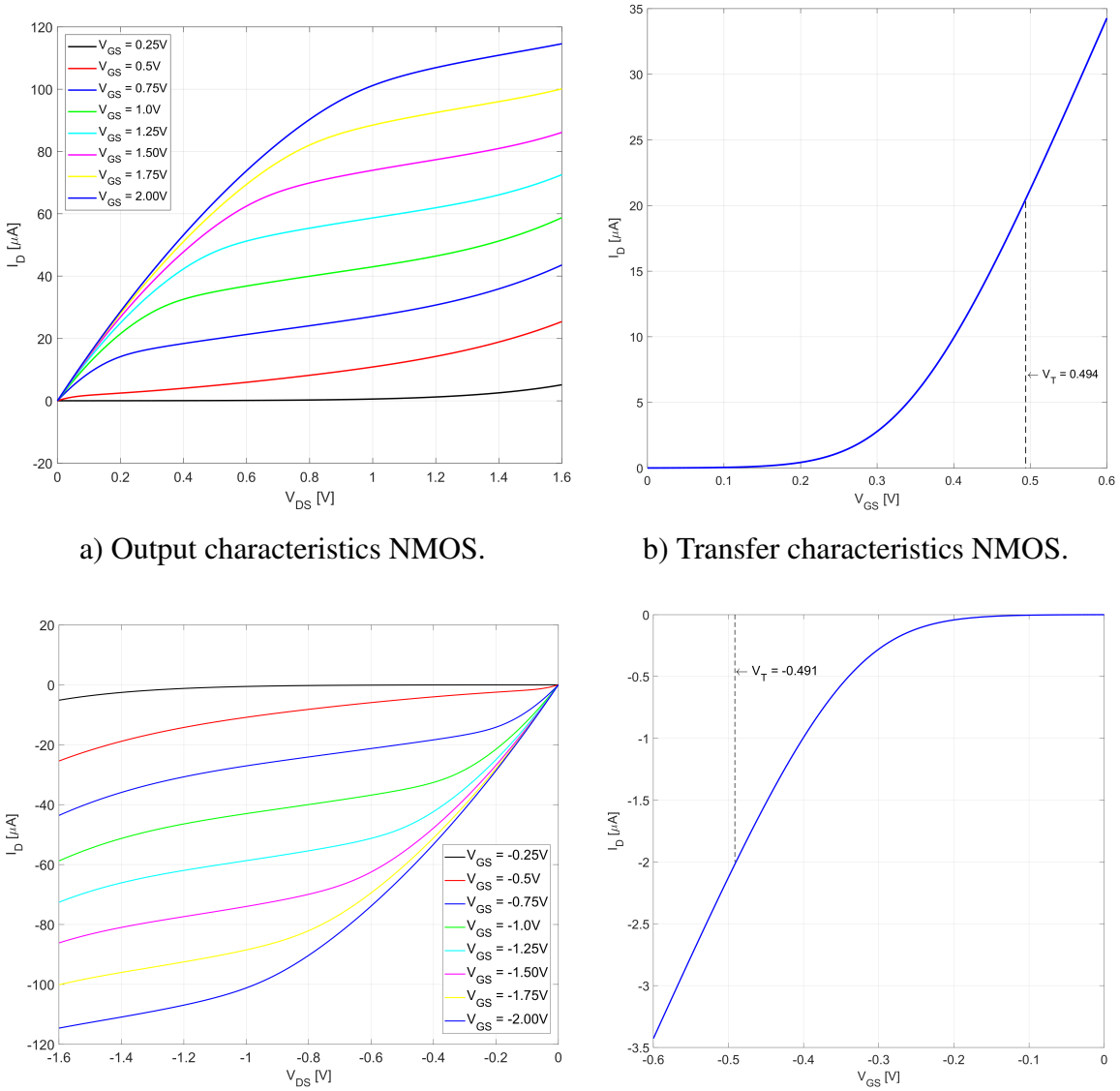
Chapter 5

RESULTS AND ANALYSIS

5.1 MODELS

Going further with the selected devices, ASU 2.1 HP 32nm process and using BSIM4 model, it is necessary to describe and analyze the selected model of MOSFET by itself. First, threshold voltage are $0.49396V$ for NMOS and $-0.49155V$ for PMOS. Output characteristics of a single NMOS and PMOS are shown in Fig. 5.1a and Fig. 5.1c respectively, for different values of gate-source voltage. It is good to indicate that as a near-real model, even for $V_{gs} < V_{th}$, sub-threshold voltages, the transistor still present a similar curve to other tensions. Also analyzing transfer characteristics exhibited in Fig. 5.1b and Fig. 5.1d, from $V_{gs} = 0V$ to $V_{gs} \approx \pm 0.49V$, called sub-threshold region, the transistor yet has the ability to work with exponential gain. The weak inversion region is the most important for this work as it implies huge gain and low power consumption.

From the 32 nanometer (32 nm) lithography process, this work use the MOSFET's of minimum gate length 32nm. For the gate width, different values will imply mainly on altered gain and power consumption, both been directly proportional to gate width. Smaller gate widths are expected, presuppose a smaller area and power consumption, although the gain will drop too. Another characteristic influenced by the gate width is the spike frequency of the neuron, and will be discussed further on section 5.2.1.



a) Output characteristics NMOS.

b) Transfer characteristics NMOS.

c) Output characteristics PMOS.

d) Transfer characteristics PMOS.

Figure 5.1: ASU PTM 32nm HKMG CMOS.

5.2 CIRCUITS

5.2.1 ANALOG SPIKE NEURON 1

The first neuron, a regular Axon Hillock, was built using the circuits mentioned previously on section 3.2.1. Using a power supply voltage of $0.9V$. In addition, the capacitive divider was calculated to have a value of 0.2, that is, $\Delta V_{mem} = 0.2 \cdot V_{dd}$ so $\Delta V_{mem} = 0.18V$. This variation is how much V_{mem} will increase and decrease when V_{out} goes from ground to $0.9V$, the exact spike pulse. Using $I_{nj} = 20nA$, $V_{pw} = 0.22V$ and $V_{rfr} = 0.53V$ the wave form of V_{mem} over time is represented in Fig. 5.2.

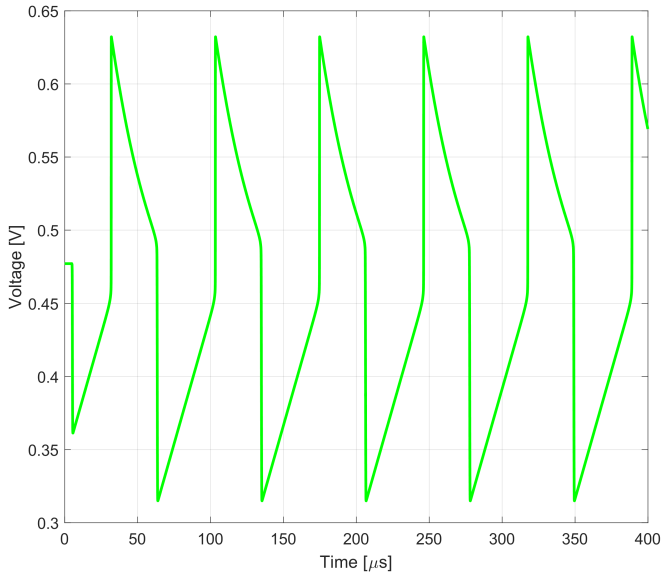


Figure 5.2: V_{mem} waveform.

Two other important components to plot are V_{out} (spike train) and I_r. Equation (3.5) presented that I_r has a major importance determining the t_{high} time, how long the output spike will stay at 0.9V. Fig. 5.3 presents a complete graph with all tree main characteristics alongside. As expected, the membrane voltage reaches a predetermined value and the spike happens, it is represented by the red output signal. At this point I_r increases its value by discharging C_m starting the refractory period of the system and a new cycle.

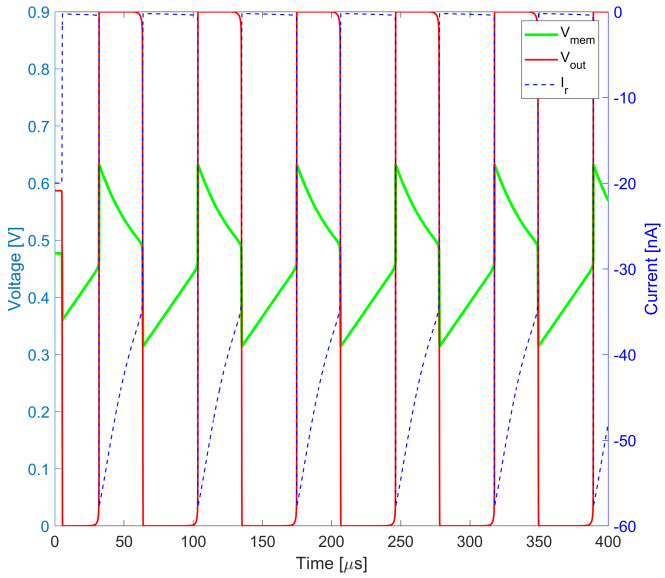


Figure 5.3: Waveforms of V_{mem}, V_{out} and I_{or}.

Some circuit parameters has a bigger influence on the final results than others, because of that it is interesting to investigate what each constant modifies on the circuit. It will be evaluated time between spikes, output pulse duration and frequency of spikes.

Decreasing V_{pw} at M11 gate is also expected to decrease I_r thus increasing t_{high}. On

the other hand, t_{low} depends only by injection current I_{nj} thus remains constant for different values of V_{pw}. Differently, decreasing I_{nj} and maintaining V_{pw} will affect a lot t_{low}, as it has a major dependence by the injection current. Tables 5.1 and 5.2 shows I_{nj} and V_{pw} dependence of t_{high} and t_{low} respectively. Estimated values were calculated using Eq. 3.5, and it represents an estimated value. By simulating real circuits, other variables influence their values. The propose of these tables is not to compare the values quantitatively. Instead, they indicate an increase and decrease of the times by changing these variables.

Table 5.1: Influence of I_{nj} in time variables with V_{pw} fixed

$V_{pw} = 0.22$		Value of I_{nj}		
		$7nA$	$17nA$	$27nA$
$t_{high}(\mu s)$	Estimated	22.50	28.12	37.50
	Obtained	20.56	27.96	45.58
$t_{low}(\mu s)$	Estimated	160.71	61.170	41.16
	Obtained	117.47	46.89	30.41

Table 5.2: Influence of V_{pw} in time variables with I_{nj} fixed

$I_{nj} = 20nA$		Value of V_{pw}		
		$0.20V$	$0.22V$	$0.24V$
$t_{high}(\mu s)$	Estimated	87.36	30.26	14.76
	Obtained	107.85	31.60	15.67
$t_{low}(\mu s)$	Estimated	50.62	50.62	50.62
	Obtained	40.02	40.07	40.16

From the tables 5.1 and 5.2 we can get some information. The spike frequency is influenced directly by the refractory period, if it is shorter, the pulse frequency will be higher. This factor is mainly determined by I_{nj} and V_{rfr}. When evaluating only I_{nj}, the I&F neuron can also behave like an encoder, encoding the input current to a spike train, Fig. 5.4. The amount of spikes are proportional to the input current.

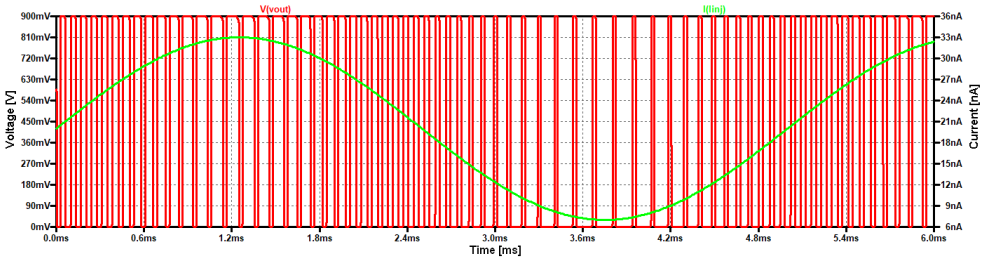


Figure 5.4: Neuron encoding an input signal.

To indicate the consequence by altering V_{rfr} a graph was created, Fig 5.5, showing the frequency of the spike train when applying different DC input current values for different

V_{rfr}. The results of the graph show that a higher V_{rfr} will cause the system to have a smaller spike frequency when maintaining the injection current constant.

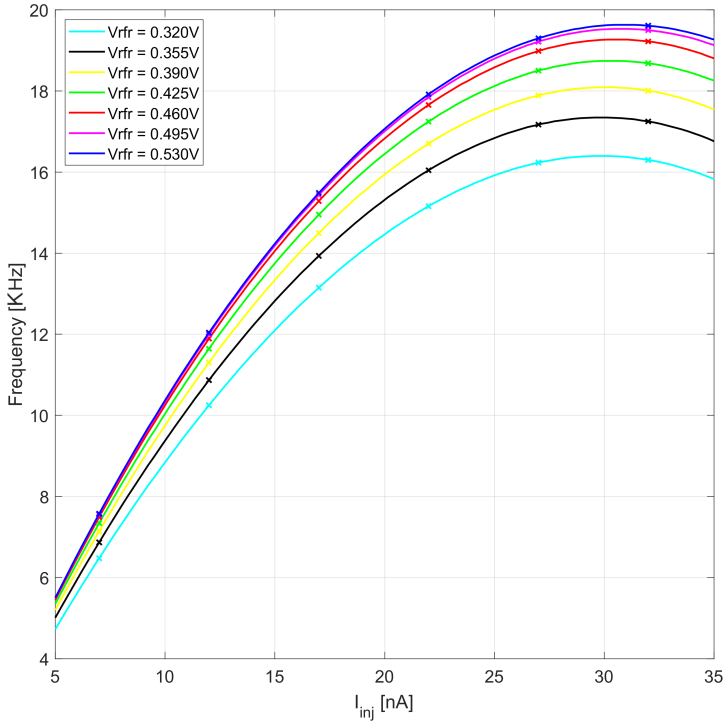


Figure 5.5: Spiking frequency characteristic's.

The MOSFET gate width will also affect the spike frequency, not as much as V_{rfr} and I_{nj}, however it still has a significant change. We indicate such differences on spike frequency by using V_{out} signal resulted from sequential simulations varying only the gate width of all transistors for the entire neuron. Applying the FFT (Fast Fourier transform) on the output spike train signal from each iteration, resulted on plot of the frequency domain, Fig. 5.6. Thereby, it was capable to identify that gate widths with values very close to the gate length, $32nm \leq W \leq 50nm$ with $W/L \approx 1$, resulted in pulse train with lower frequency. The green and red signal were obtained using a gate width near gate length, $W_{green} = 36nm$ and $W_{red} = 42nm$.

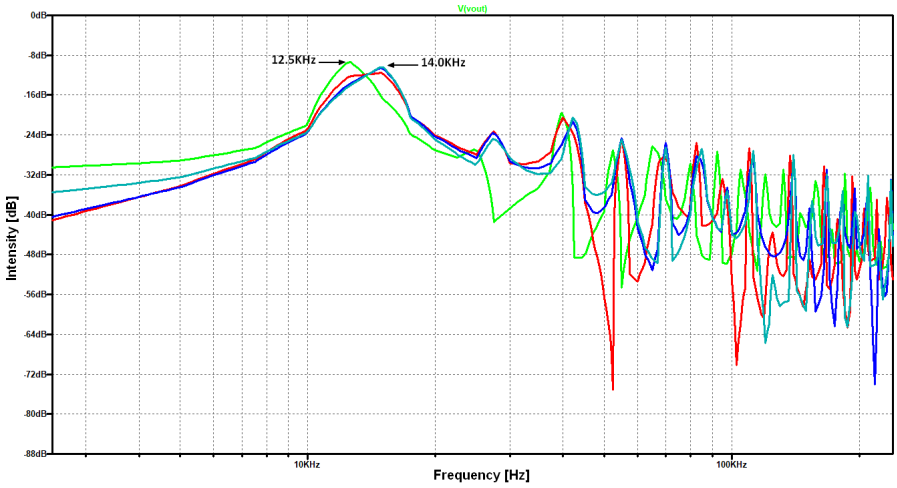


Figure 5.6: Influence of MOSFET gate width on spike frequency.

Finally, using default parameters such as $V_{DD} = 0.9V, I_{nj} = 10nA, V_{rfr} = 0.53V, V_{pw} = 0.20V$ and $W_{NMOS} = 52nm$, power consumption of one neuron turns out to be $4.75nW$ and it is directly proportional of the spike frequency since more spikes will dissipate more energy, thus all major parameters that influence pulse frequency will direct impact on power consumption. More important to analyze than energy dissipation is energy efficiency ($Joules/spike$), for the regular neuron a $48.5fJ/spike$ was achieved, being directly dependent on variables that considerable change t_{high} and t_{low} characteristics. At the end of next session a table is presented with this work and other values of power dissipation and energy efficiency.

5.2.2 ANALOG SPIKE NEURON 2

The low power Axon Hillock neuron was created using the circuit on section 3.2.2. A different approach was used to extract the results of the simulation of this neuron, taking into account that there are fewer parameters to be configured. Respecting the low power consumption, $0.15V$ was selected to be the power supply and current source of $I_{nj} = 35pA$. Using $C_{fb} = 5fF$ the capacitor divider works the same way of first neuron, the difference is that the parasitic capacitance of first inverter has major influence now. Fig. 5.7 presents the spike pulse for the Low Power Spike Neuron with V_{mem} and V_{out}. Comparing with Fig. 5.3, the results shown here are different when evaluating the shape of the curves and also peak voltage values. There is no longer an output signal with a square wave shape since the differential pair was removed, resulting in a spike much more like the ideal waveform of the action potential in Fig. 3.1 with a voltage peak near $70mV$.

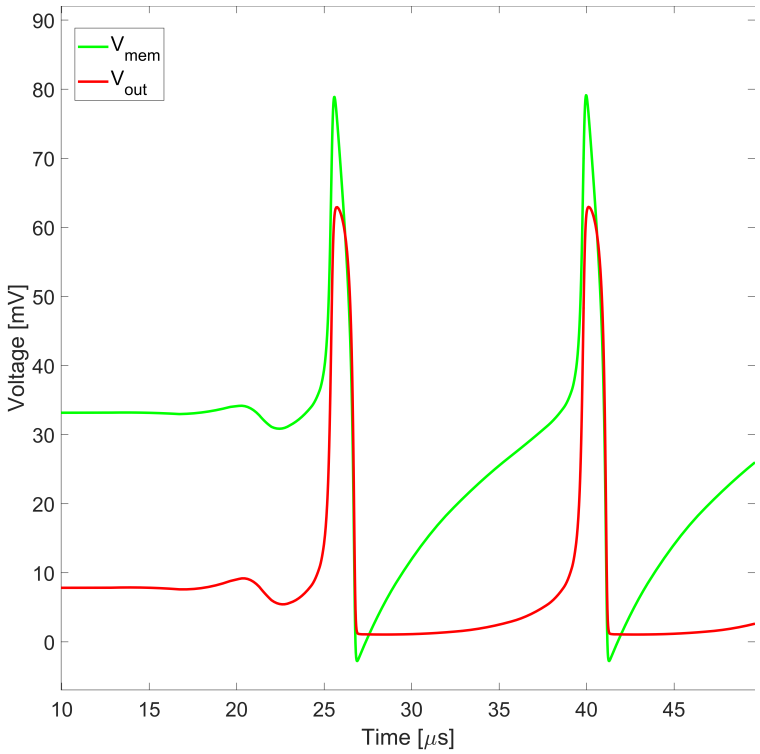


Figure 5.7: Waveforms of V_{mem} and V_{out}.

Like the previous neuron, the injection current I_{nj} remains one of the main factors responsible for adjusting pulse period and frequency. However, as mentioned earlier, the pulse width control is no longer adjusted by V_{pw}, the withdrawal of the second transistor next to $M5$ turns the W/L ratio of $M5$ and $M3$ now responsible for controlling I_r and consequently the width of the spike, in other words, the period in which pulse remains HIGH. Table 5.3 and Table 5.4 show how I_{nj} and the gate width of $M5$ and $M3$ affects t_{high} and t_{low} for the low power neuron.

Results on Table 5.3 indicate that I_{nj} continues to have major influence on t_{low} but not a substantial importance on t_{high} as expected, being the same for the regular neuron. In distinction to the regular neuron where V_{pw} modulate pulse characteristics, Table 5.4 shows results for each transistor gate width while maintaining the other transistor aspect ratio. It reveals that W_{M5} has a larger influence over t_{high} and for t_{low} both transistors do not have a big impact, this was expected as I_r is not responsible for t_{low}.

Table 5.3: Influence of I_{nj} in time variables with W_{M3} and W_{M5} fixed.

$W_{M5} = W_{M3} = 52nm$	Value of I_{nj}		
	$39pA$	$51pA$	$63pA$
$t_{high}(\mu s)$	2.05	2.23	2.40
$t_{low}(\mu s)$	10.91	7.42	5.76

Table 5.4: Influence of W_{M3} and W_{M5} in time variables with I_{nj} fixed.

$I_{nj} = 39pA$		Value of $W_{M3,M5}$		
		$32nm$	$52nm$	$64nm$
$t_{high}(\mu s)$	W_{M3}	2.17	2.05	1.96
	W_{M5}	3.41	2.05	1.17
$t_{low}(\mu s)$	W_{M3}	7.69	10.91	11.70
	W_{M5}	10.67	10.91	11.05

As shown in Fig. 5.4 presented on last section, the low power neuron can also behave like an encoder. Applying a sine wave for injection current the spikes fire proportional to the current as shown on Fig. 5.8. I_{nj} will not only alter spike pulse time but also the frequency in witch it occurs as well as the regular neuron. Fig. 5.9 exhibit the influences of input current on spike frequency and shows that it has the same pattern from previous neuron, as expected.

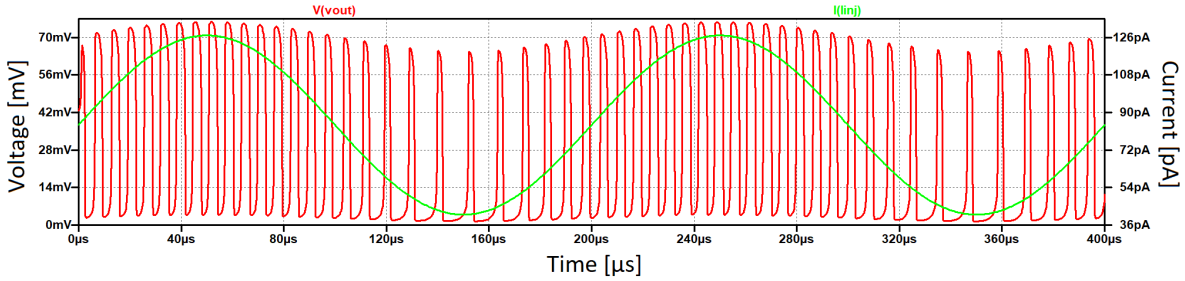


Figure 5.8: Low power neuron encoding an input signal.

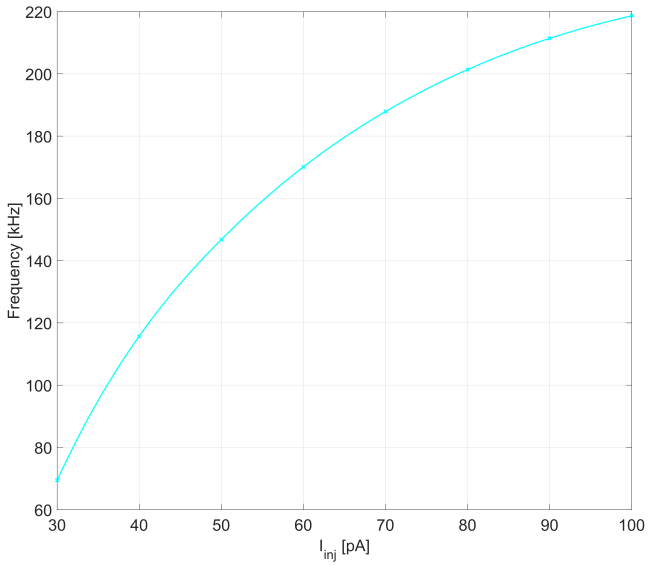


Figure 5.9: Spiking frequency characteristic's for low power Neuron.

Because this neuron model uses very low voltages for the power source, it is interesting to make an evaluation of how a small variation on V_{DD} will affect neuron functionalities.

Fig. 5.10 shows one of the possible effects if the voltage source has noise altering it voltage value. Only a small difference of $20mv$ are enough to reduce pulse frequency from $\approx 125kHz$ to $\approx 70kHz$. This could lead to some problems during fabrication, requiring a fine industrial method to reduce noise perturbation.

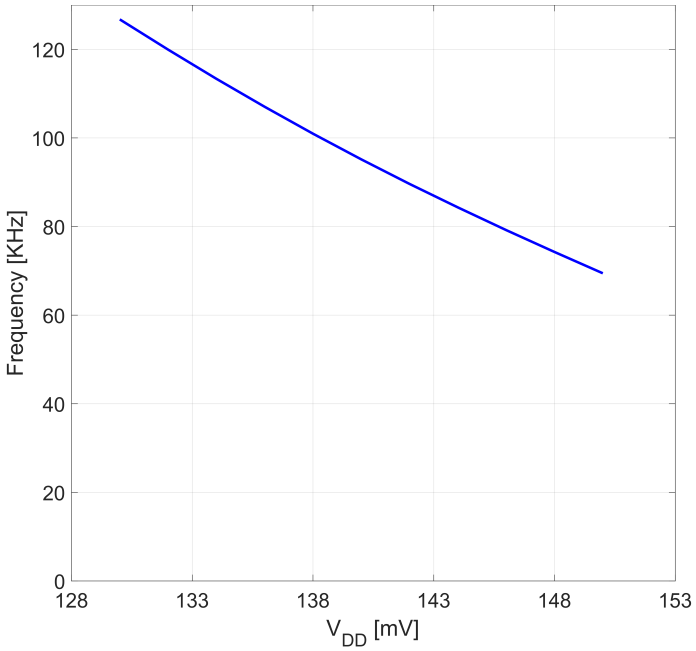


Figure 5.10: Spiking frequency as function of V_{DD}.

Lastly, the reason for the name given to this neuron, the energy consumption. To extract the power dissipation and energy efficiency a standard model was used, without major changes to variables, $V_{DD} = 150mV, I_{nj} = 38pA$ and $W_{NMOS/PMOS} = 52nm$. The power consumption for a single neuron was $23.73pW$ and the energy efficiency of a single spike $0.232fJ/Spike$. As with the first neuron presented, the parameters that modify pulse frequency and pulse size has a direct influence on these values.

Concluding the results for neurons section, Table 5.5 present the results of this work and others main differences when evaluating power consumption, energy per-spike and fire frequency. Energy per spike should be used as a figure of merit as it provides a fair comparison of power consumption with regard to the processing capacity of neurons. Considering this, the two proposed circuits have achieved good results among these projects.

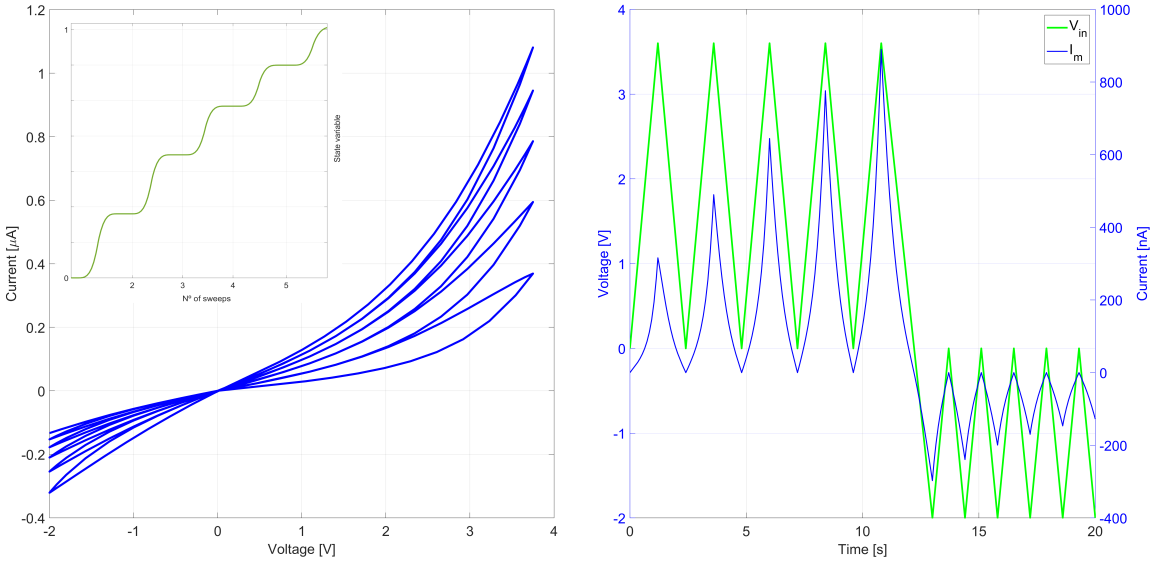
From all eight referenced models there are two that deserve more attention, Indiveri [15] and Danneville [13]. As already said, the two neuron circuits used here are almost the same proposed on these works. With considerable changes and adjustments to the MOSFET model and transistor W/L ratio. On the table it can be seen that both works presented here were capable of obtaining better energy efficiency for each spike. The regular I&F neuron reduced from $900pJ/Spike$ to $0.0485fJ/Spike$ and the low power I&F from $2fJ/Spike$ to $0.232fJ/Spike$.

Table 5.5: Comparison of the performances of this work 32nm neuron circuit and other VLSI neuron models

References	Process (nm)	Spike Frequency (Hz)	Power (W)	Energy Efficiency (pJ/Spike)
[15]	350	200	$0.3 - 1.5\mu$	900
[12]	65	$1.9 \cdot 10^6$	78μ	41
[16]	350	10^6	$8 - 40\mu$	$8.5 - 9$
[17]	180	$33 \cdot 10^6$	–	9.3
[18]	180	$100 - 500 \cdot 10^3$	$35p$	2
[19]	90	100	$40.2p$	0.4
[20]	65	$26 \cdot 10^3$	$105p$	0.004
[13]	65	$15.7 \cdot 10^3$	$30p$	0.002
Neuron 1	32	$10 - 20 \cdot 10^3$	$4.75n$	0.0485
Neuron 2	32	$70 - 220 \cdot 10^3$	$23.73p$	0.00023

5.2.3 MEMRISTOR SYNAPSE

Given the generic model of memristor chosen to be used and adjusting its parameters, it is possible to obtain a characteristic IV curve for the synaptic memristor. Using a sequence of five positive and five negative triangular peaks, Fig. 5.11b, initially representing the pulses, the operation mode as a synapse can be achieved as shown in Fig. 5.11a. The triangular spikes here has peak of $3.6V$ and occurs every 2.4 seconds. Using the right parameters as $a_1 = 3.7{*}10^{-7}, a_2 = 4.35{*}10^{-7}, b = 0.7, V_p = 1.5, V_n = 0.5, A_p = 0.005, A_n = 0.08, x_p = 0.2, x_n = 0.5, \alpha_p = 1.2, \alpha_n = 3, x_o = 0.1$ and $\eta = 1$. The I-V plot shows how successive voltages sweeps results on reduction of the device resistance and after each spike the device have a major change on its conductivity.

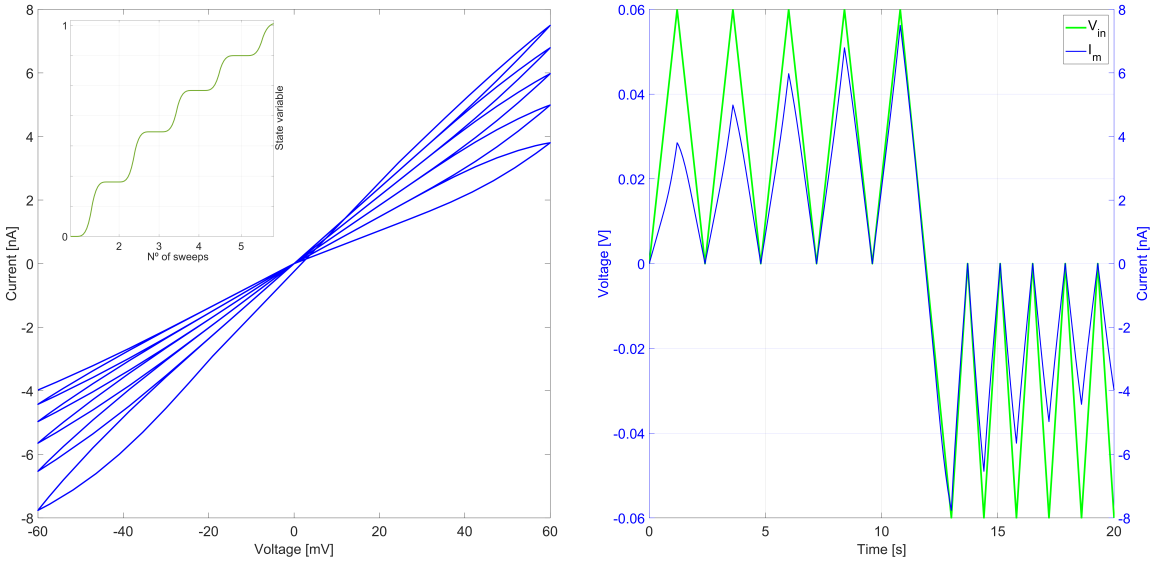


a) Voltages sweeps of synaptic memristor. b) Applied pulse train to synaptic memristor.

Figure 5.11: I-V characteristics of the memristor with synaptic behavior using triangular pulses.

This model was adjusted for high voltage and low frequency. To work as a synapse for the presented neurons, low voltages are required. To adjust the model in order to perform using triangular peaks of $60mV$, as indicated on Fig. 5.12b, are necessary, $V_p = 55mV$ and $V_n = 50mV$ to set threshold voltages, resulting on Fig. 5.12a.

Both configurations presented here work on a period of time of 20 seconds, this is too much for the neuron simulation, as it was presented on the previous session that each spike lasts micro-seconds. The variable responsible for fitting how quickly state variable changes once the threshold is surpassed is A_p and A_n, changing it to 10^6 scale adjusts the time problem and the graph after this adjustment is just the same of Fig. 5.12.



a) Voltages sweeps of adjusted synaptic memristor.

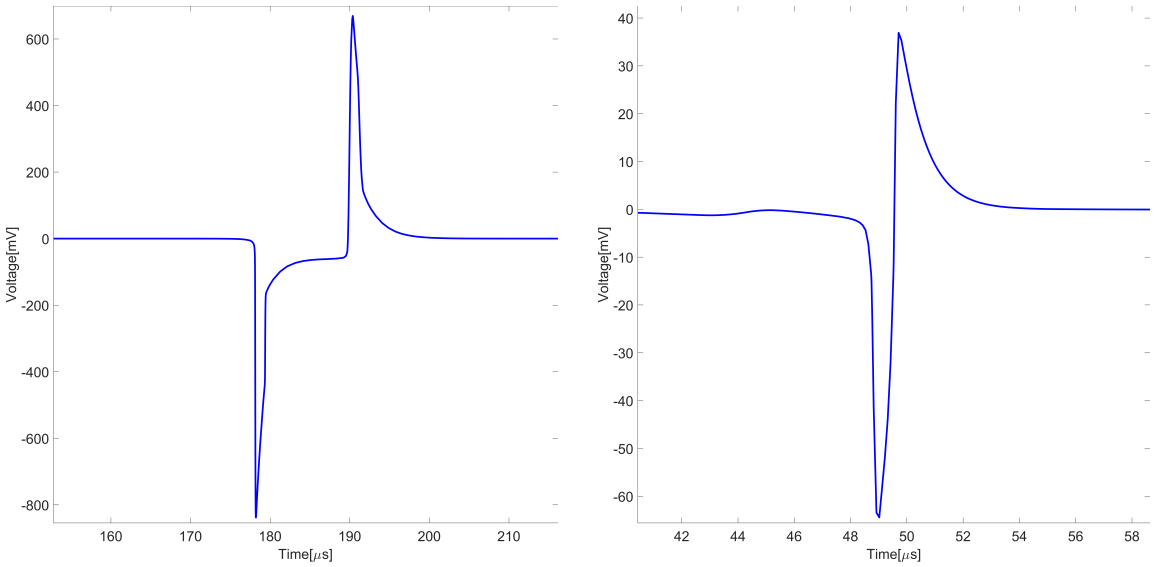
b) Applied pulse train to synaptic memristor fitted.

Figure 5.12: I-V characteristics of the memristor with synaptic behavior using triangular pulses after adjust voltage parameters.

5.2.4 STDP LEARNING

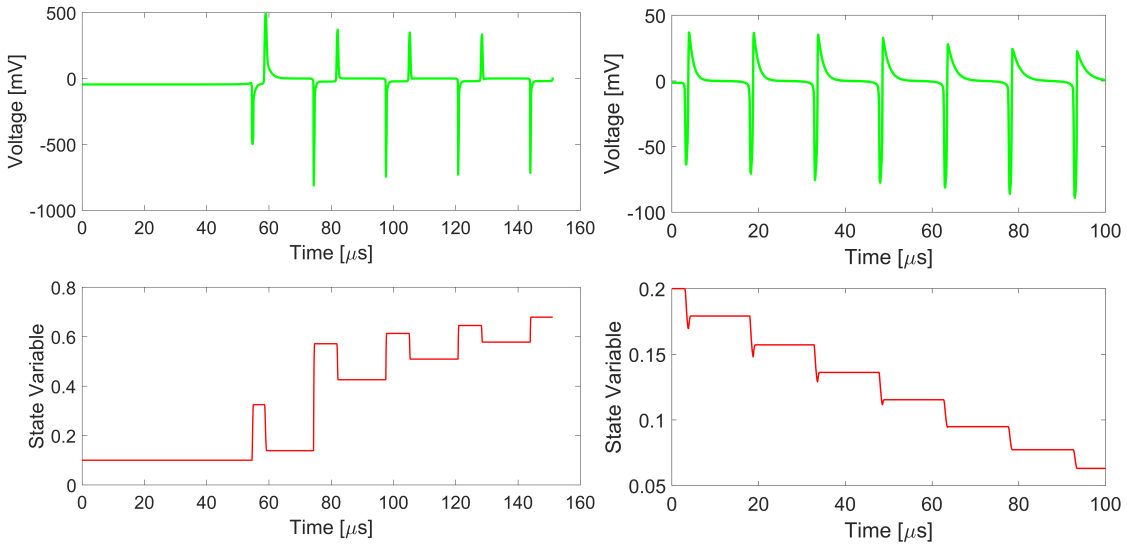
After presenting the results for each circuit block of this project, it is time to show these pieces working together. At the end of Section 3.4 it was shown, how pre- and post-synaptic pulse between spike neurons could update synaptic weight, the so called Spike-timing-dependent plasticity learning mechanism of real neurons and synapses. Using the configuration Neuron-Memristor-Neuron on Fig. 3.8, the STDP curve, responsible for adjustment on the state variable or synaptic weights, generated at the memristor between the implemented neurons are revealed on Fig. 5.13 for each neuron circuit.

To show how these two curves are capable of modifying the synaptic weights, Fig. 5.14. shows for both neurons the gradual alteration of the memristor's state variable, which in this case is precisely the change in the synaptic weights. Presented on equation 2.1, the increase and decrease of synaptic weight depends on how the STDP is acting. It is also possible to notice that in Fig. 5.14b, the positive peak begins to decrease over time until it becomes less than the threshold voltage of the memristor, thus the positive changes on the state variable stop occurring. It is also possible to note that in both cases the state variable keeps its value constant if nothing is applied.



a) Regular neuron. b) Low power neuron.

Figure 5.13: Spike-timing-dependent plasticity curve.



a) Regular neuron. b) Low power neuron.

Figure 5.14: Synaptic weights updated by STDP.

To finalize the results of simulations, a system was created to show how different pre-synaptic pulses from two different neurons in a different period of time change the synaptic weights of their respective synapses using the same post-synaptic signal from a third neuron. In other words, Fig. 5.15 presents the beginning of a spike neural network and how the circuits presented manage to work together. Using, for the blue neuron, a regular frequency of spikes and for the red one a random spike generator.

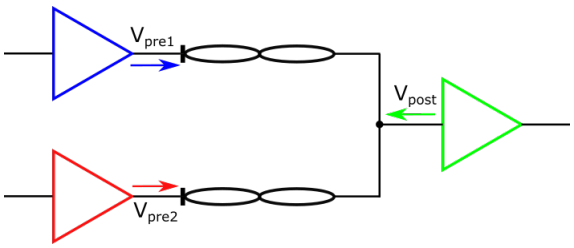


Figure 5.15: The small Spike Neuron Network.

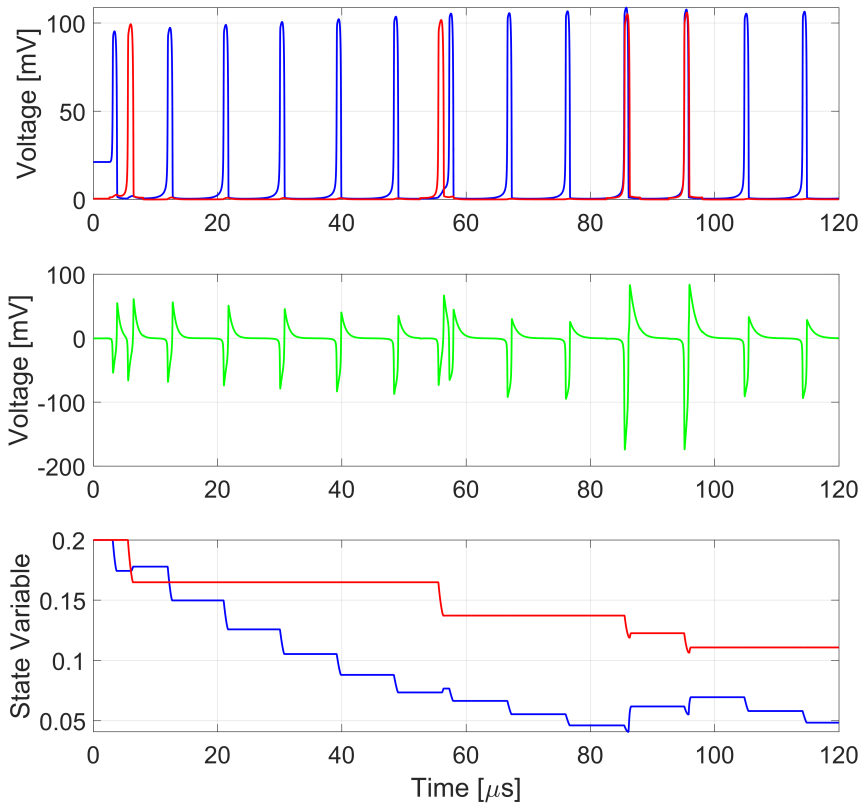


Figure 5.16: Neurons operation and STDP learning with two input neurons and one output neuron.

On the figure it is possible to observe the performance of the second neuron being responsible for updating the synaptic weights not only for the second memristor but also being part of the update for the first memristor, this is because both pulses influence the post-synaptic signal. This phenomenon will occur at all times, even in larger neural networks. So the synapses of the present neural network are updated, in other words they are learning patterns.

Chapter 6

Conclusion

Despite Moore's law for many years predicting miniaturization of transistors and increasing transistor density on integrated circuit, this growth pattern may be arriving at the biggest obstacle, the physical barrier. The need to continue the development of technologies has brought to light new architectures that optimize information processing without the main need to reduce the size of the transistors. A great example of this are the processing units used for machine learning operations, that increasingly need computer systems with more processing power. One way to improve the processing capacity is using bio-inspired systems, such as neuromorphic systems, that try to simulate the basic functioning of a human brain.

During the last years CMOS has been widely used in electronics systems and analog neuromorphic systems are not that different. Reducing power supply and transistor aspect ratio, by changing length gate are main priority when refers to new technologies smaller and with better energy efficiency. Although it became harder and harder to simulate real physical transistor because of the lack of real parameters, some predictive models shows good results when comparing with real transistors. Novel CMOS process arises all time and the predictive models are the base to create it. Therefore, it is necessary that these predictions are as close as possible to reality, such that simulated circuits can accurately represent real ones.

In addition to the MOSFET devices, the idealization and manufacture of the memristor brings many innovations to microelectronics in general. Its memory capacity on a passive device makes it a great synapse solution for neuromorphic systems. However, as it is still a very new device, commercial processes are yet to come, and most simulations and tests are done with mathematical models that describe their behavior.

This work proposed and validated the basis of a spike neural network architecture using two different types of analog CMOS neurons and configuring a memristor to function as a synapse enabling the STDP learning process, all the proposed objectives have been achieved. Through a bibliographic review covering all aspects of other studies, and evaluating meth-

ods for simulations and data acquisition, two existing neuron designs were implemented and simulated using a different CMOS process technology from previous works [13–15]. We were able to reproduce the same results even with reduced power supply (lower power consumption) and the transistor length almost six times shorter leading to less effective area. On the second part a generic memristor model was configured to work together with these neurons making possible the creation of a neuromorphic system able to update synaptic weights, in other words, to perform an online training of the neural network.

Some aspects of the system were left out of the work to be studied and designed in future projects. The next step, after a SPICE simulation, is the development of the circuit layout design to estimate the real size of each neuron and mismatch evaluation to identify the limiting value of gate width. To complete the spike neural network, it needs some adjacent circuits working together. Adding a method to select network training mode or control of input and output in the neural network, in other words, determine when the system is using training data or performing a data recognition. Finally, use the whole system to work with real data by training and accomplish pattern recognition.

REFERENCES

[1] MOORE, G. E. *Cramming More Components onto Integrated Circuits*. [S.l.].

[2] CAVIN, R. K.; LUGLI, P.; ZHIRNOV, V. V. Science and engineering beyond moore's law. In: *Proceedings of the IEEE*. [S.l.: s.n.], 2012. v. 100, n. SPL CONTENT, p. 1720–1749. ISSN 00189219.

[3] BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 35, n. 8, p. 1798–1828, 2013. ISSN 01628828.

[4] GARCÍA-MARTÍN, E. et al. Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, Academic Press Inc., v. 134, p. 75–88, 12 2019. ISSN 07437315.

[5] TOMASI, D.; WANG, G. J.; VOLKOW, N. D. Energetic cost of brain functional connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, National Academy of Sciences, v. 110, n. 33, p. 13642–13647, 8 2013. ISSN 00278424.

[6] BARON, R. *The cerebral computer: An introduction to the computational structure of the human brain*. [S.l.: s.n.], 2013.

[7] MAI, J.; PAXINOS, G. *The human nervous system*. [S.l.: s.n.], 2011.

[8] Khan Academy. *The synapse, Human biology*. Disponível em: <https://www.khanacademy.org/science/biology/human-biology/neuron-nervous-system/a/the-synapse>.

[9] MARKRAM, H. et al. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, American Association for the Advancement of Science, v. 275, n. 5297, p. 213–215, 1 1997. ISSN 00368075.

[10] MORELL, P.; QUARLES, R. H. The Myelin Sheath. Lippincott-Raven, 1999. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK27954/>.

[11] SMITH, L. S.; HAMILTON, A. *Neuromorphic Systems*. WORLD SCIENTIFIC, 1998. (Progress in Neural Processing, v. 10). ISBN 978-981-02-3377-8. Disponível em: <https://www.worldscientific.com/worldscibooks/10.1142/3702>.

[12] JOUBERT, A. et al. Hardware spiking neurons design: Analog or digital? In: *Proceedings of the International Joint Conference on Neural Networks*. [S.l.: s.n.], 2012. ISBN 9781467314909.

[13] DANNEVILLE, F. et al. A Sub-35pW Axon-Hillock artificial neuron circuit. *Solid-State Electronics*, Elsevier Ltd, v. 153, p. 88–92, 3 2019. ISSN 00381101.

[14] SRIVASTAVA, S.; RATHOD, S. S. Silicon neuron-analog CMOS VLSI implementation and analysis at 180nm. In: *Proceedings of the 3rd International Conference on Devices, Circuits and Systems, ICDCS 2016*. [S.l.: s.n.], 2016. ISBN 9781509023097. ISSN 1099-1263.

[15] INDIVERI, G.; CHICCA, E.; DOUGLAS, R. A VLSI Array of Low-Power Spiking Neurons and Bistable Synapses With Spike-Timing Dependent Plasticity. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, v. 17, n. 1, 2006.

[16] WIJEKOON, J. H.; DUDEK, P. Compact silicon neuron circuit with spiking and bursting behaviour. *Neural Networks*, v. 21, n. 2-3, p. 524–534, 3 2008. ISSN 08936080. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0893608007002705>.

[17] WU, X. et al. A CMOS Spiking Neuron for Brain-Inspired Neural Networks with Resistive Synapses and in Situ Learning. *IEEE Transactions on Circuits and Systems II: Express Briefs*, Institute of Electrical and Electronics Engineers Inc., v. 62, n. 11, p. 1088–1092, 11 2015. ISSN 15497747.

[18] SHAMSI, J.; MOHAMMADI, K.; SHOKOUHI, S. B. A low power circuit of a leaky integrate and fire neuron with global reset. In: *2017 25th Iranian Conference on Electrical Engineering, ICEE 2017*. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2017. p. 366–369. ISBN 9781509059638.

[19] CRUZ-ALBRECHT, J. M.; YUNG, M. W.; SRINIVASA, N. Energy-efficient neuron, synapse and STDP integrated circuits. *IEEE Transactions on Biomedical Circuits and Systems*, v. 6, n. 3, p. 246–256, 2012. ISSN 19324545.

[20] SOURIKOPOULOS, I. et al. A 4-fJ/Spike Artificial Neuron in 65 nm CMOS Technology. *Frontiers in Neuroscience*, Frontiers Research Foundation, v. 11, n. MAR, p. 123, 3 2017. ISSN 1662-453X. Disponível em: <http://journal.frontiersin.org/article/10.3389/fnins.2017.00123/full>.

[21] JO, S. H. et al. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Letters*, v. 10, n. 4, p. 1297–1301, 4 2010. ISSN 15306984.

[22] David Laws. 13 Sextillion and Counting: The Long and Winding Road to the Most Frequently Manufactured Human Artifact in History. *Computer History Museum*, 4 2018.

[23] BAKER, R. J. *CMOS: circuit design, layout, and simulation*. Third edition. Piscataway, NY: [s.n.], 2010. ISBN 978-0-470-88132-3.

[24] SEDRA, A.; SMITH, K. Microelectronic circuits: theory and applications. 2013.

[25] Arizona State University. *Predictive Technology Model*. Disponível em: <http://ptm.asu.edu/>.

[26] University of California - Berkeley. *BSIM4 – BSIM Group*. Disponível em: <http://bsim.berkeley.edu/models/bsim4/>.

[27] HARRIS, D. M. et al. Evaluation of predictive technology models. *Microelectronics Journal*, Elsevier Ltd, v. 80, p. 7–17, 10 2018. ISSN 00262692.

[28] DUNGA, M. V. *Nanoscale CMOS Modeling*. Tese (Doutorado) — EECS Department, University of California, Berkeley, 3 2008. Disponível em: <http://www2.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-20.html>.

[29] AIYAPPA, B. N. et al. Amplifier design in weak inversion and strong inversion — A case study. In: *2017 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 2017. p. 1227–1231. ISBN 978-1-5090-3800-8. Disponível em: <http://ieeexplore.ieee.org/document/8286575/>.

[30] CHUA, L. . *Memristor-The Missing Circuit Element*. [S.l.], 1971. v. 18, n. 5, 507 p.

[31] MUTHUSWAMY, B. et al. Memristor modelling. In: *Proceedings - IEEE International Symposium on Circuits and Systems*. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2014. p. 490–493. ISBN 9781479934324. ISSN 02714310.

[32] STRUKOV, D. B. et al. The missing memristor found. *Nature*, v. 453, n. 7191, p. 80–83, 5 2008. ISSN 00280836.

[33] MILLER, K. et al. Memristive behavior in thin anodic Titania. *IEEE Electron Device Letters*, v. 31, n. 7, p. 737–739, 7 2010. ISSN 07413106.

[34] JOGLEKAR, Y. N.; WOLF, S. J. The elusive memristor: properties of basic electrical circuits. 7 2008. Disponível em: <http://arxiv.org/abs/0807.3994 http://dx.doi.org/10.1088/0143-0807/30/4/001>.

[35] BIOLEK, Z.; BIOLEK, D.; BIOLKOVÁ, V. *SPICE Model of Memristor with Nonlinear Dopant Drift*. [S.l.].

[36] BATAS, D.; FIEDLER, H. A memristor SPICE implementation and a new approach for magnetic flux-controlled memristor modeling. *IEEE Transactions on Nanotechnology*, v. 10, n. 2, p. 250–255, 3 2011. ISSN 1536125X.

[37] YANG, J. J. et al. Memristive switching mechanism for metal/oxide/metal nanodevices. *Nature Nanotechnology*, Nature Publishing Group, v. 3, n. 7, p. 429–433, 2008. ISSN 17483395.

[38] MIAO, F. et al. Anatomy of a nanoscale conduction channel reveals the mechanism of a high-performance memristor. *Advanced Materials*, v. 23, n. 47, p. 5633–5640, 12 2011. ISSN 09359648.

[39] YAKOPCIC, C. et al. Generalized memristive device SPICE model and its application in circuit design. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, v. 32, n. 8, p. 1201–1214, 2013. ISSN 02780070.

[40] YAKOPCIC, C. et al. Memristor SPICE Modeling. Disponível em: <https://pdfs.semanticscholar.org/3054/267ed6237ce1c5b33311ab985020a7fd99a6.pdf>.

[41] Research ltd. *Memristors Market - Growth, Trends, and Forecast (2020 - 2025)*. Disponível em: <https://www.researchandmarkets.com/reports/4771724/memristors-market-growth-trends-and-forecastrela1-4618275>.

[42] MARANI, R.; GELAO, G.; PERRI, A. G. A review on memristor applications. 6 2015. Disponível em: <http://arxiv.org/abs/1506.06899>.

[43] YAKOPCIC, C. et al. Memristor-based unit cell for a detector readout circuit. In: *Unconventional Imaging, Wavefront Sensing, and Adaptive Coded Aperture Imaging and Non-Imaging Sensor Systems*. [S.l.]: SPIE, 2011. v. 8165, p. 81651F. ISBN 9780819487759. ISSN 0277786X.

[44] HAMDIOUI, S.; TAOUIL, M.; HARON, N. Z. Testing open defects in memristor-based memories. *IEEE Transactions on Computers*, IEEE Computer Society, v. 64, n. 1, p. 247–259, 1 2015. ISSN 00189340.

[45] MEENA, J. S. et al. Overview of emerging nonvolatile memory technologies. *Nanoscale Research Letters*, Springer New York LLC, v. 9, n. 1, p. 1–33, 9 2014. ISSN 1556276X. Disponível em: <http://nanoscalereslett.springeropen.com/articles/10.1186/1556-276X-9-526>.

[46] PAL, S. et al. Design of power-And variability-Aware nonvolatile rram cell using memristor as a memory element. *IEEE Journal of the Electron Devices Society*, Institute of Electrical and Electronics Engineers Inc., v. 7, p. 701–709, 2019. ISSN 21686734.

[47] MOSTAFA, H.; ISMAIL, Y. Statistical yield improvement under process variations of multi-valued memristor-based memories. *Microelectronics Journal*, Elsevier Ltd, v. 51, p. 46–57, 5 2016. ISSN 00262692.

[48] HU, M. et al. Memristor crossbar-based neuromorphic computing system: A case study. *IEEE Transactions on Neural Networks and Learning Systems*, Institute of Electrical and Electronics Engineers Inc., v. 25, n. 10, p. 1864–1878, 10 2014. ISSN 21622388.

[49] OBLEA, A. S. et al. Silver chalcogenide based memristor devices. In: *Proceedings of the International Joint Conference on Neural Networks*. [S.l.: s.n.], 2010. ISBN 9781424469178.

[50] WU, X.; SAXENA, V.; ZHU, K. A CMOS spiking neuron for dense memristor-synapse connectivity for brain-inspired computing. *Proceedings of the International Joint Conference on Neural Networks*, IEEE, v. 2015-Septe, p. 1–6, 2015.

[51] YAKOPCIC, C. et al. Memristor-based neuron circuit and method for applying learning algorithm in SPICE. *Electronics Letters*, Institution of Engineering and Technology, v. 50, n. 7, p. 492–494, 3 2014. ISSN 00135194.

[52] HEBB, D. O. *The organization of behavior, A neuropsychological study*. New York: Wiley, 1949.

[53] LINARES-BARRANCO, B.; SERRANO-GOTARREDONA, T. Memristance can explain Spike-Time-Dependent-Plasticity in Neural Synapses. *Nature Precedings*, Springer Science and Business Media LLC, 3 2009.

[54] ZAMARREÑO-RAMOS, C. et al. On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex. *Frontiers in Neuroscience*, n. MAR, 2011. ISSN 16624548.

[55] CHEN, L. et al. STDP learning rule based on memristor with STDP property. In: *Proceedings of the International Joint Conference on Neural Networks*. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2014. p. 1–6. ISBN 9781479914845.

[56] FROEMKE, R. C.; DAN, Y. Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature*, v. 416, n. 6879, p. 433–438, 3 2002. ISSN 00280836.

[57] ALPAYDIN, E. *Introduction to machine learning*. [S.l.: s.n.], 2014. 613 p. ISBN 0262028182.

[58] BOUT, D. E. van den; MILLER, T. K. A Digital Architecture Employing Stochasticism for the Simulation of Hopfield Neural Nets. *IEEE Transactions on Circuits and Systems*, v. 36, n. 5, p. 732–738, 1989. ISSN 00984094.

[59] MEAD, C. Neuromorphic electronic systems. *Proceedings of the IEEE*, v. 78, n. 10, p. 1629–1636, 1990. ISSN 00189219. Disponível em: <http://ieeexplore.ieee.org/document/58356/>.

[60] INDIVERI, G. Computation in Neuromorphic Analog VLSI Systems. In: *12th Italian Workshop on Neural Networks. WIRN*. [s.n.], 2001. p. 3–19. Disponível em: <http://ncs.ethz.ch/pubs/pdf/Indiveri01c.pdf>.

[61] POON, C. S.; ZHOU, K. *Neuromorphic silicon neurons and large-scale neural networks: Challenges and opportunities*. 2011.

[62] LANDE, T. S. *Neuromorphic systems engineering : neural networks in silicon*. [S.l.]: Kluwer Academic, 1998. 462 p. ISBN 9780792381587.

[63] Nora AlNashwan. *The Age of Cognitive Computing - IBM Code*. 2018. Disponível em: <https://developer.ibm.com/code/2018/01/21/the-age-of-cognitive-computing/>.

[64] PAINKRAS, E. et al. SpiNNaker: A 1-W 18-Core System-on-Chip for Massively-Parallel Neural Network Simulation. *IEEE Journal of Solid-State Circuits*, v. 48, n. 8, p. 1943–1953, 8 2013. ISSN 0018-9200. Disponível em: <http://ieeexplore.ieee.org/document/6515159/>.

[65] AKOPYAN, F. et al. TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Institute of Electrical and Electronics Engineers Inc., v. 34, n. 10, p. 1537–1557, 10 2015. ISSN 02780070.

[66] DAVIES, M. et al. Loihi: A Neuromorphic Manycore Processor with On-Chip Learning. *IEEE Micro*, IEEE Computer Society, v. 38, n. 1, p. 82–99, 1 2018. ISSN 02721732.

[67] WANGCHETAN, R. A neuromorphic hardware architecture using the Neural Engineering Framework for pattern recognition. p. 1–12, 2015.

[68] KIM, Y.; ZHANG, Y.; LI, P. A digital neuromorphic VLSI architecture with memristor crossbar synaptic array for machine learning. In: *International System on Chip Conference*. [S.l.: s.n.], 2012. p. 328–333. ISBN 9781467312950. ISSN 21641676.

[69] BARTOLOZZI, C.; INDIVERI, G. Synaptic dynamics in analog VLSI. *Neural Computation*, v. 19, n. 10, p. 2581–2603, 10 2007. ISSN 08997667.

[70] NAWROCKI, R. A.; VOYLES, R. M.; SHAHEEN, S. E. *A Mini Review of Neuromorphic Architectures and Implementations*. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 10 2016. 3819–3829 p.

[71] SARPESHKAR, R. *Analog Versus Digital: Extrapolating from Electronics to Neurobiology*. [S.l.]: MIT Press Journals, 10 1998. 1601–1638 p.

[72] KAKKAR, V. *Comparative Study on Analog and Digital Neural Networks*. [S.l.], 2009. v. 9, n. 7, 14 p.

[73] AAMIR, S. A. et al. An accelerated LIF neuronal network array for a large-scale mixed-signal neuromorphic architecture. *IEEE Transactions on Circuits and Systems I: Regular Papers*, Institute of Electrical and Electronics Engineers Inc., v. 65, n. 12, p. 4299–4312, 12 2018. ISSN 15498328.

[74] BURKITT, A. N. A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biological Cybernetics*, v. 95, n. 1, p. 1–19, 7 2006. ISSN 03401200.

[75] AAMIR, S. A. et al. A highly tunable 65-nm CMOS LIF neuron for a large scale neuromorphic system. *European Solid-State Circuits Conference*, v. 2016-Octob, p. 71–74, 2016. ISSN 19308833.

[76] DUTTA, S. et al. Leaky Integrate and Fire Neuron by Charge-Discharge Dynamics in Floating-Body MOSFET. *Scientific Reports*, Nature Publishing Group, v. 7, n. 1, 12 2017. ISSN 20452322.

[77] LIVI, P.; INDIVERI, G. A current-mode conductance-based silicon neuron for Address-Event neuromorphic systems. In: *Proceedings - IEEE International Symposium on Circuits and Systems*. [S.l.: s.n.], 2009. p. 2898–2901. ISBN 9781424438280. ISSN 02714310.

[78] ALVADO, L. et al. Hardware computation of conductance-based neuron models. *Neurocomputing*, Elsevier, v. 58-60, p. 109–115, 2004. ISSN 09252312.

[79] HODGKIN, A. L.; HUXLEY, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, v. 117, n. 4, p. 500–544, 8 1952. ISSN 0022-3751. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1952.sp004764>.

[80] NEWHALL, K. A. et al. Dynamics of current-based, Poisson driven, integrate-and-fire neuronal networks. *Communications in Mathematical Sciences*, International Press, v. 8, n. 2, p. 541–600, 2010. Disponível em: <https://projecteuclid.org/euclid.cms/1274816894>.

[81] ZHAO, W. et al. Predictive technology modeling for 32nm low power design. In: *2007 International Semiconductor Device Research Symposium, ISDRS*. [S.l.: s.n.], 2007. ISBN 1424418917.

[82] RADHIKA, E.; KUMAR, S.; KUMARI, A. Low power analog VLSI implementation of cortical neuron with threshold modulation. In: *2015 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2015*. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2015. p. 561–566. ISBN 9781479987917.

[83] HSIEH, H. Y.; TANG, K. T. VLSI implementation of a bio-inspired olfactory spiking neural network. *IEEE Transactions on Neural Networks and Learning Systems*, v. 23, n. 7, p. 1065–1073, 2012. ISSN 2162237X.