



Low-power hybrid memristor-CMOS spiking neuromorphic STDP learning system

Gabriel Maranhão¹  | Janaina Gonçalves Guimarães² 

¹ENE, Federal University of Brasília (UnB), Brasília, Brazil

²CAC, Federal University of Santa Catarina (UFSC), Blumenau, Brazil

Correspondence

Janaina Gonçalves Guimarães, CAC, Federal University of Santa Catarina (UFSC), Blumenau, Brazil.

Email: janaina.guimaraes@ufsc.br

Funding information

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: 000000/2018; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 311997/2017-1

Abstract

An electronic circuit that implements a neural network architecture with spike neurons was studied, proposed, and evaluated, primarily considering energy consumption. In this way, CMOS transistors were used to implement neurons, memristors were used to work as synapses, and the proposed network has a *spike-timing-dependent plasticity* (STDP) learning aspect. The validation of the circuit modules and the complete network architecture was performed using SPICE models. Since most data of company technologies is restricted, some universities provide predictive models to reproduce the real ones. In this paper, two types of *Integrate and Fire Neuron (I&F)* using 32 nm CMOS technology simulated in LTspice with BSIM4v4 model designed by Berkley University and applying predictive parameters provided by *Predictive Technology Model* (PTM) are presented. The simulation results obtained here reduces the bias voltage and the chip size to the most recent designs implemented. Finally, communication between neurons and synapses with STDP learning has been successfully simulated.

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and more powerful computational systems appear all the time, resulting in a cell phone from years ago being considered outdated.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 arrive at a physical barrier, and the size of the transistor. Different approaches are studied every year to increase the computational power of the systems without necessarily changing the transistors. Many 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. 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. Computational systems that use the CPU with von Neumann architecture are not optimized for these functions. A functional

processing unit capable of quickly performing these functions with high energy efficiency is the human brain [2]. 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 simultaneously. 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 to perform the same functions as a human brain, with low power consumption in such a small space. The neuron cell is responsible for numerous functions within the animal's biological system. It can fire nerve impulses; in other words, the action potential [3]. The synapses are the name given to the communication channel between two neurons and, each neuron makes 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 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

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *IET Circuits, Devices & Systems* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

potentials (or spikes). The action potential leaves the axon from the cell body and travels through the myelin sheath arriving at the gap (synapse) [4]. At this point, the synaptic strength is adjusted and, the signal crosses it via neurotransmitters and activates or not the postsynaptic signal. The novel microelectronic systems that try to reproduce the human brain's functionality are commonly called neuromorphic systems. 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 Neumann 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 many works [5–7]. 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. Also, some projects already use memristors to implement synapses and show excellent results [7–9]. A large part of this novel brain-inspired architecture is implemented, simulated, and even manufactured using MOSFET transistors. It is because MOSFET is the most commercially used device. Consequently, there is a vast collection of tools that work with MOSFET transistors, and also it 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 [10]. Due to its ability to ‘remember’, its resistive memory makes it useful to function as a device to represent synaptic weights.

Considering that one of the ideas of neuromorphic systems implemented in hardware is to seek the integration density of biological systems, to perform functions similar to those of the brain, this work aims to study proposals for low-power and small area neurons quantitatively. For this purpose, it aims to establish and evaluate an electronic circuit that implements a neural network architecture with spike neurons using analogue MOSFET neuron design. In its turn, spike neural networks (SNN), besides mimicking the neural dynamics of the brain, can usually achieve lower power consumption compared to artificial neural network (ANN) implementations. This feature is quite impressive when it comes to hardware realization [11].

The proposed network has an STDP learning aspect as it applies memristors to work as synapses. We evaluate two types of *Integrate and Fire Neuron* (I&F) [5, 12] using 32 nm CMOS technology, changing the technology in which they were originally proposed. Both circuits were simulated in LTspice with predictive MOS transistor models. The simulation results

obtained here reduce total power consumption in more than 85 % and reduce chip size to the most recent designs implemented using smaller CMOS technology. In the end, communication between neurons and synapses with STDP learning executed by the memristors has been successfully simulated.

2 | NEURAL NETWORKS AND ARCHITECTURES

The brain inspires artificial neural networks (ANNs). They are massively parallel architectures, composed of simple processors, called neurons, highly interconnected by the synapses. ANNs are adapted explicitly to solving visual perception and dynamic control problems. Specifically, Spiking Neural Networks (SNNs) have almost the same properties. However, the main difference between ANN and SNN is the time operation. SNNs work based on timing spike inputs while ANN are static. So, SNNs are closer to the brain's neural dynamics. Besides that, SNN based systems have lower power dissipation if compared to ANNs in the same application.

Energy-efficient hardware implementation of learning systems is a challenge when it comes to current von Neumann architecture-based computers. On early 1990's, Carver Mead first described the *Neuromorphic Electronic Systems*, ‘biology-inspired’ microelectronics, using the idea of an analogue electronic architecture, applying very-large-scale integration (VLSI) process, that can perform activities equivalent to biological neurons presented in mammals [13]. Those systems are intended to compete with von Neumann systems, improve energy efficiency, and cost of computational systems.

According to Indiveri [12], analogue 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 analogue VLSI neuromorphic systems such as conservancy of charge, integration, amplification, and size reduction [14]. Collective computation in a densely parallel analogue VLSI circuit is probably one of the most advantageous approaches to process the information on pattern recognition or sound recognition.

In neuromorphic systems, all the information processed is assigned and stored in each neuron and their synapses, responsible for the learning mechanism [15]. So, 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 [16]. All of them are trying to approximate the biological neuron behaviour.

The need for faster, efficient, reliable, and smaller systems leads the studies to choose digital or analogue approaches regarding design hardware neurons. Both paths have pros and cons when comparing the following characteristics, ease-of-design, robustness, scaling, and storage [17].

In terms of reliability and robustness, from the analogue point of view, it demands substantial care to create the design

to minimize effects such as variations on the power supply, leakage, and temperature variations, unlike binary digital circuits considering it has only two states. Despite being harder to design, analogue 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 [17].

In addition to the discussion about analogue 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 [16].

Initially, the simpler I&F model has a straightforward implementation and non-realistic when comparing to the biological system but successfully used to create dense networks and one of the most utilized models for analyzing neural network properties [18]. It defines the neuron using the membrane potential $v(t)$ and applying Kirchhoff's current law for the conservation of charge:

$$C_{mem} \cdot \frac{dv(t)}{dt} = I_{leak}(t) + I_S(t) + I_{inj}(t), \quad (1)$$

The membrane is stimulated by postsynaptic input deriving from other neurons and their synapses, $I_S(t)$. A current $I_{leak}(t)$ is responsible for modelling the leak conductance and potential that decays over a constant time and lastly, $I_{inj}(t)$, representing an external current stimulation. When the membrane potential hits a fixed threshold mark, the spike is generated [19].

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 bilayer by a capacitor, in other words, an equivalent circuit representation of a cell membrane [20].

First proposed by Hodgkin and Huxley, the Hodgkin–Huxley model [21] 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

$$I_m = I_C + I_{ion}, \text{ where} \quad (2)$$

$$I_C = C_{mem} \cdot \frac{dv(t)}{dt}, \quad I_{ion} = I_{Na} + I_K + I_{leaky} \quad (3)$$

2.1 | Analogue spike neuron

As part of neuromorphic systems, the analogue spiking neuron represents a fundamental piece. Several methods to reproduce a spike (the main activity for communication between cells also known as nerve impulses) of a biological neuron using VLSI have been described over the years. The action potential (spike) is the communication of neuronal cells, quickly carrying

information between and within tissues, mainly inside the brain. When the excitation on the neuron membrane reaches a specific threshold voltage, the *fire* occurs. The typical shape of an action potential is presented in Figure 1.

It is vital to take into account, as stated above, that analogue spiking neuron applying the I&F model has most details about the electrophysiology of real ones not included in the design. Instead, its simplicity is a significant advantage adequately describing the network mechanisms [22]. That simplicity resulted in the focus of theoretical and computational studies over decades and was still extensively used and improved on large dense neural networks.

An example related to this kind of hardware neuron is the Axon–Hillock (A-H), first proposed by Indiveri [12], presented in Figure 2, a simple implementation that can be useful to create dense networks and study neural network properties [18]. It usually has two additional inverters acting as a buffer for the output signal.

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. Thus, 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

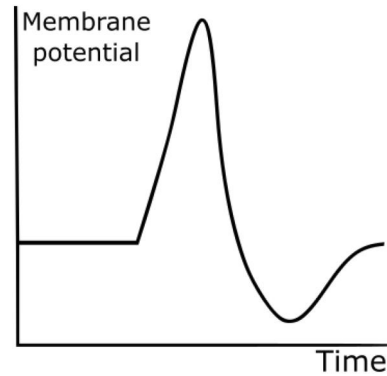


FIGURE 1 The pattern of a neuron action potential

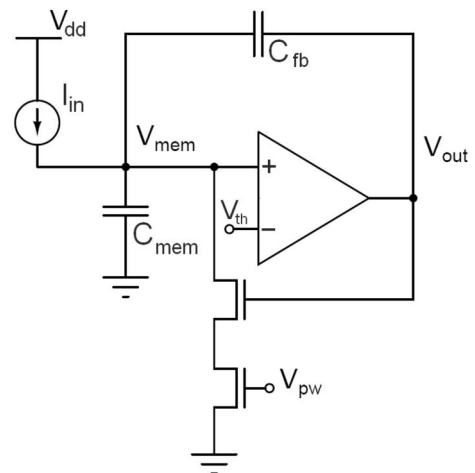


FIGURE 2 The Axon–Hillock circuit

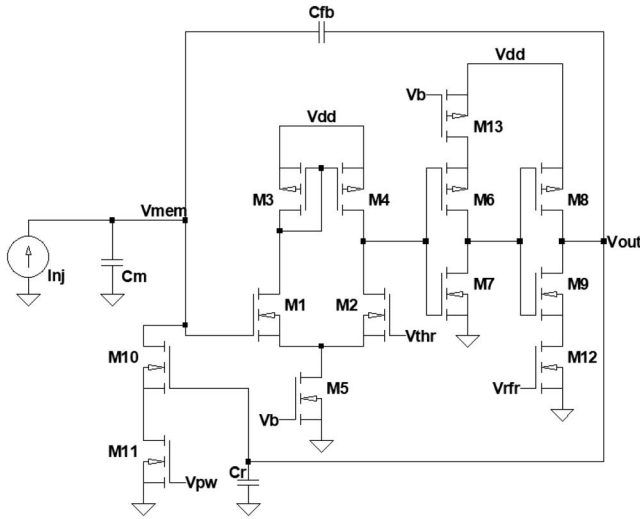


FIGURE 3 Regular Axon-Hillock integrate and fire neuron

neural learning mechanisms and requires fewer transistors to implement than cortical ones. In this section, two existing models of spike neurons are presented, and their modes of operation are demonstrated.

2.1.1 | Regular Axon-Hillock

Originally schemed by Indiveri [12], the integrate and fire neuron in Figure 3 contains 13 transistors, separated in four blocks, a differential pair (M1–M5), two inverters (M6–M9), and a reset circuit (M10–M11).

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 maintains 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 positive feedback created by the capacitor divider C_m and a feedback capacitor C_{fb} is activated because V_{out} turns on the reset transistor M10, allowing V_{mem} to increase with a rate given by

$$\Delta V_{mem} = \frac{C_{fb}}{C_m + C_{fb}} \cdot V_{dd} \quad (4)$$

On the other side, while V_{out} stays HIGH, C_m discharges in a rate determined by V_{pw} applied on the 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 is 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_{thr} .

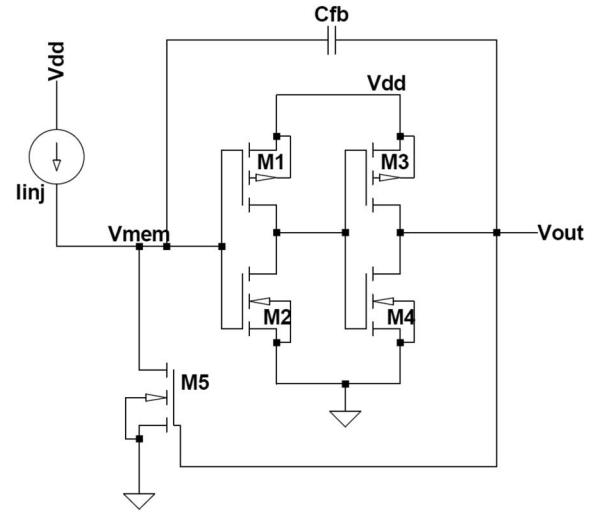


FIGURE 4 Low power A-H integrate and fire neuron

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 on I_r , as shown below:

$$t_{low} = \frac{C_{fb}}{I_{nj}} \cdot V_{dd}, \quad t_{high} = \frac{C_{fb}}{I_r - I_{nj}} \cdot V_{dd} \quad (5)$$

2.1.2 | Low-power Axon-Hillock

Analysing the low-power A-H implementation proposed in Ref. [5], it has fewer transistors on the 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 reduce power consumption drastically and achieve higher energy efficiency for each spike.

As presented in Figure 4, for this design C_m , reset circuit, and threshold differential pair were fully removed. Each part has its substitutes. The replacement for the membrane capacitance is the parasitic component of the 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 the increase or decrease I_r . Finally, the differential pair that evaluates whether or not the spike should occur is now replaced by the first inverter's switching voltage.

The behaviour is very similar to the regular neuron, the injection current I_{nj} increase charge and V_{mem} value. Once V_{mem} is sufficiently high for the switching voltage of the 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 begins again.

2.2 | Memristor synapse

The generic memristor SPICE model can reproduce several kinds of memristors [23]. Through its parameterization, different operating characteristics can be reached for different applications.

All the fitting parameters, and their initial values, for this particular model, are presented in Table 1. Using these default variables, we are able to plot the famous memristor pinched hysteresis loop represented in Figure 5. These variables are responsible for the three main characteristics: electron tunnelling, non-linear drift, and a voltage threshold [23].

It was considered the main reason for using memristors during this work, to have similar behaviour to neurological synapses and be able to perform the 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. Each pre- and post-synaptic spike works together to modify the synaptic weight, in this case, the state variable.

2.3 | Spike neural network

One of the main characteristics of a neural network is how 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, 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 of an I&F neuron–memristor–I&F neuron, as shown in Figure 6.

Figure 6 shows a pre-synaptic green pulse and a post-synaptic blue pulse at different times, making the existence of an STDP curve

TABLE 1 Fitting parameters to model different memristor devices

Parameters	Description	Default values
a_1, a_2, b	Control the I-V relationship	0.17, 0.17, 0.05
V_p, V_n	Positive and negative voltage thresholds	0.16 V, 0.15 V
A_p, A_n	The magnitude of the exponential represents how quickly the state changes once the threshold is surpassed	4000, 4000
x_p, x_n	Points where the state variable motion becomes limited	0.3, 0.5
α_p, α_n	Rate at which the SV motion decays after passing the points x_n or x_p	1.0, 5.0
x_0	Initial value for state variable	0.11
η	If 1 a positive voltage will increase SV value ^a If 0 a positive voltage will decrease SV value ^a	1

^aSV: State-variable, normalized value for conductivity, between 0 and 1.

during $\Delta t = t_{\text{post}} - t_{\text{pre}}$. The shape of action potential will profoundly influence the resulting STDP-learning function; it must have a narrow positive spike with a broad 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 comprehensive spike neuron network. Being also scalable, so large networks could be achieved using that set-up. Figure 7 represent 4×4 crossbar array composed of the primary cell described above.

3 | METHODOLOGY

All the steps and choices made for the development of this work, from the models to the simulation of the spike neural network will be described below.

3.1 | Simulation models

3.1.1 | MOSFET

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 a transistor's physical characteristics. Composed of more than 220 variables for modelling the BSIM4 [24], and ASU Predictive Technology Model (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

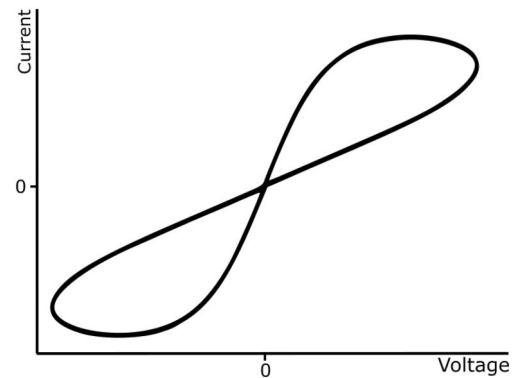


FIGURE 5 I-V Characteristics of a regular memristor

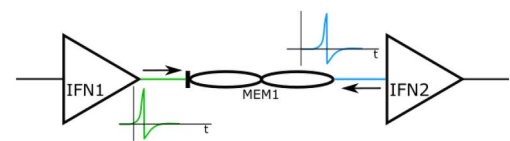


FIGURE 6 Pre- and post-synaptic pulse between spike neurons with memristor synapse

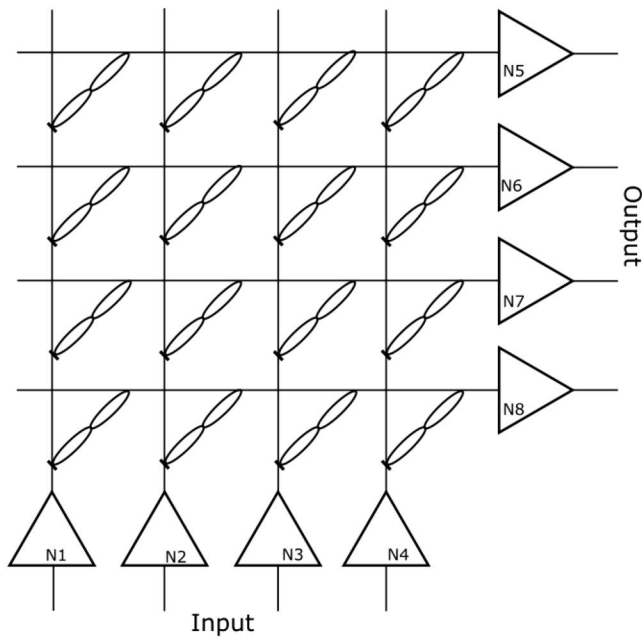


FIGURE 7 4 × 4 Crossbar array network

suggested by the compact model coalition (CMC), a group responsible for standardizing 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 no longer SPICE; it uses Verilog-A. LTspice XVII is not prepared to compile such a type of file. The parameter characterization file provided by ASU PTM was chosen based on analysis developed by previous works [26, 27]. So, ASU PTM HP HKMG 32 nm made available by the Arizona State University [28] was used in this work.

3.1.2 | Memristor

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

The memristor model proposed by Yakopic et al. [23] was chosen to be used as the synaptic circuit, as an alternative for those physical models. This specific design is responsible for translating the equations of scientific literature from different configurations and grouping them all in the so-called Generic Model, capable of representing with equivalence different types of memristors and 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 [10],

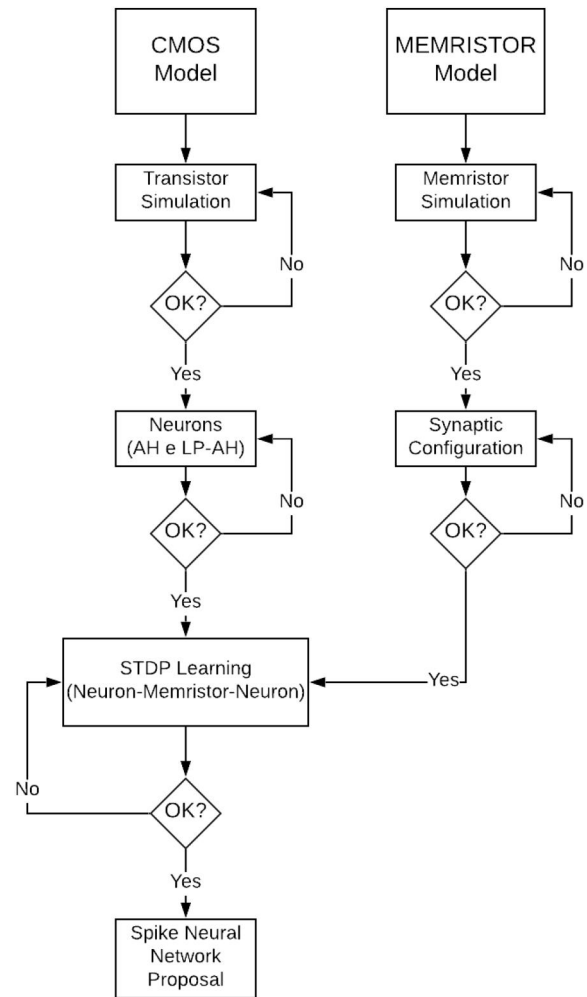


FIGURE 8 Simulations flow chart

ideals for sweep behaviour, and power-efficient when driving STDP learning operation. Using these real memristor models as exemplars and fitting the SPICE model presented in this session to have the same characteristics, achieving the ideal hysteresis is possible.

3.2 | Circuits

Two types of I&F neurons were chosen to be implemented due to their simple operation and easy implementation, thus enabling them to function together with synapses and higher neuron density. The regular A-H circuit on Figure 3 and the low-power A-H circuit in Figure 4 were used to develop the neural network.

3.3 | Simulations

The most common and first approach used by companies and universities to simulate the behaviour 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 was used to perform the simulations, starting with minor blocks and ending with the major project.

A brief diagram is shown in Figure 8, showing how the simulation hierarchy went step by step. Starting with the most basic simulations of the NMOS and PMOS transistors and ending with the neuron–synapse–neuron triad. First, a MOSFET evaluation to ensure the expected operation must have a high gain in weak inversion regimes and use a low supply voltage. Second, the neuron circuits are tested with the selected transistor model, observing parameters for better operation and making adjustments. 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.

4 | RESULTS AND ANALYSIS

This work used the 32-nm MOSFETs for designing the neurons. 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 neuron's spike frequency and will be discussed later [29].

FIGURE 9 Waveforms of V_{mem} , V_{out} and I_{or}

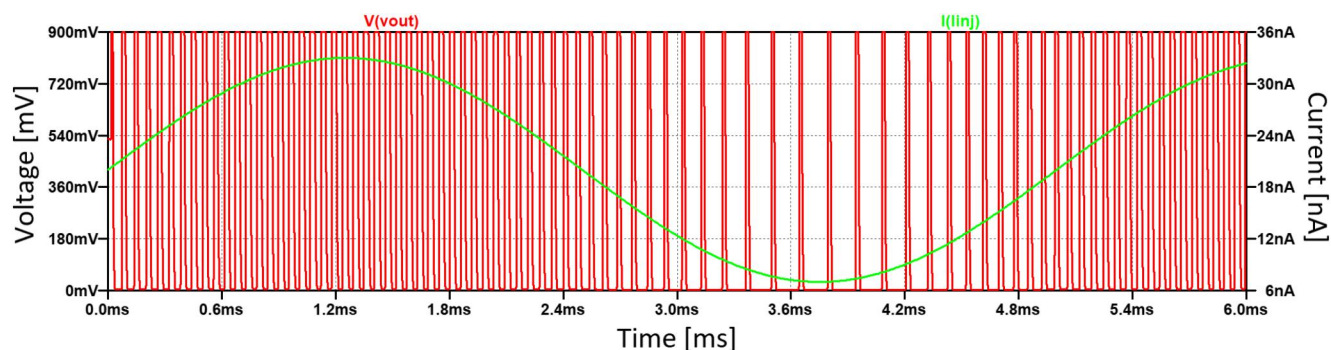
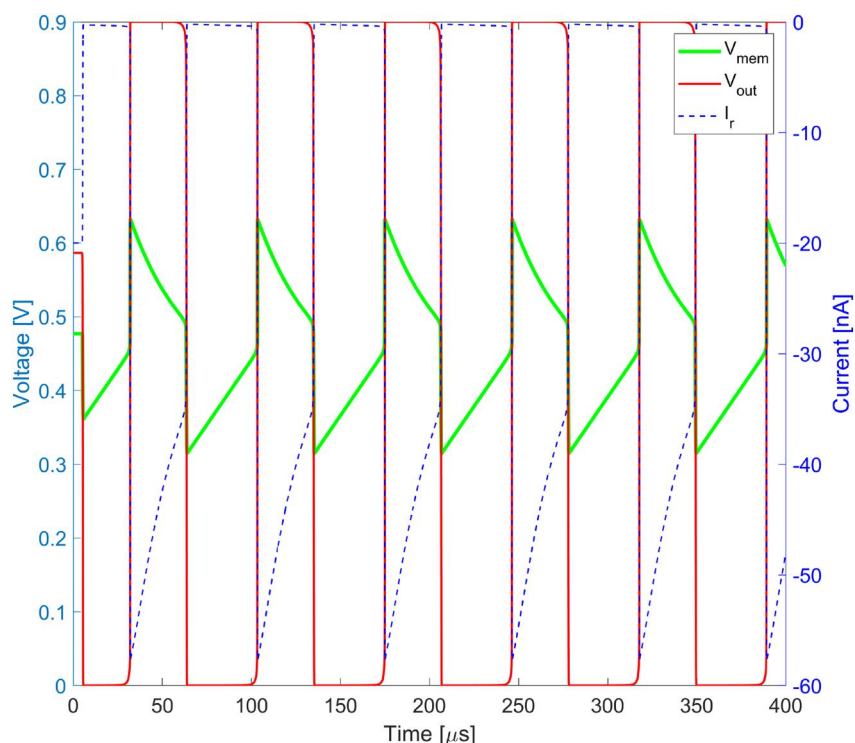


FIGURE 10 Analogue spike neuron 1 encoding an input signal

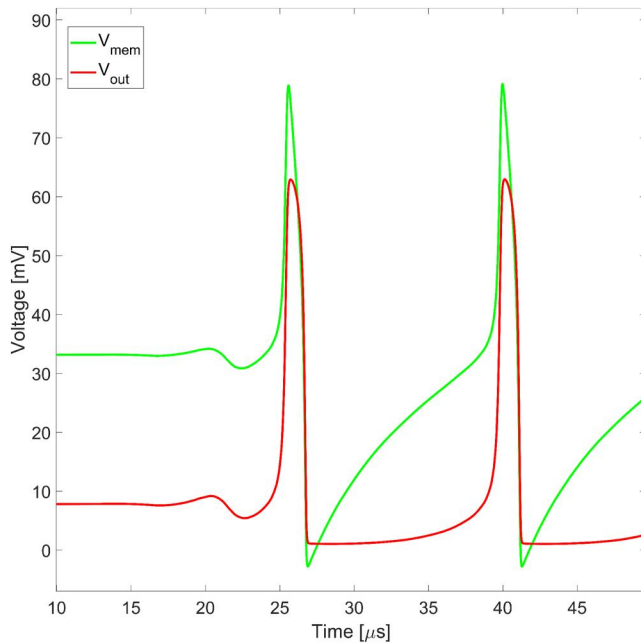


FIGURE 11 Waveforms of V_{mem} and V_{out}

4.1 | Circuits

4.1.1 | Analogue spike neuron 1

The first neuron, a regular A-H, was built using the circuits mentioned previously. Using a power supply voltage of 0.9 V. Besides, the capacitive divider was calculated to have a value of 0.2, that is $\Delta V_{\text{mem}} = 0.2 \times V_{\text{dd}}$, so $\Delta V_{\text{mem}} = 0.18$ V. This variation is how much V_{mem} will increase and decrease when V_{out} goes from ground to 0.9 V, the exact spike pulse. Using $I_{\text{nj}} = 20$ nA, $V_{\text{pw}} = 0.22$ V and $V_{\text{rfr}} = 0.53$ V.

Two other important components to show are V_{out} (spike train), and I_r . Equation (5) presented that I_r has primary importance determining the t_{high} time, how long the output spike will stay at 0.9 V. Figure 9 presents a complete graph with all main tree 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 system's refractory period and a new cycle.

In this circuit, the spike frequency is influenced directly by the refractory period; the shorter its value, the higher the pulse frequency. I_{nj} and V_{rfr} mainly determine this factor.

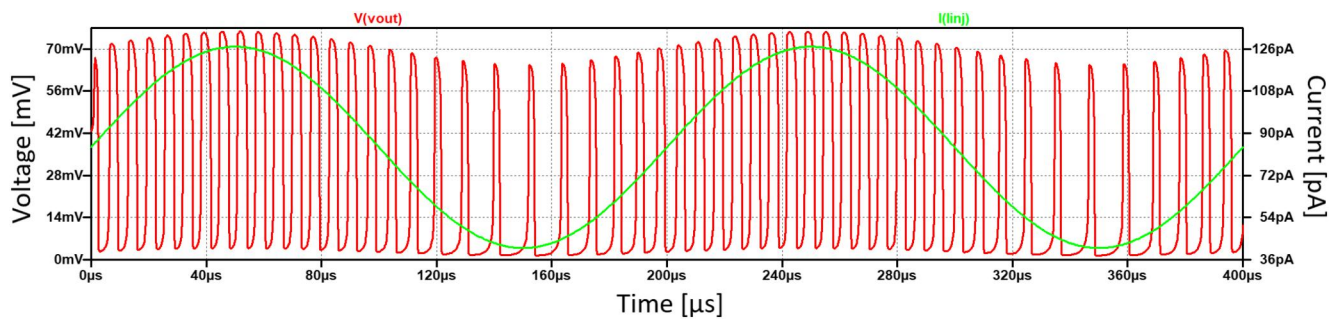
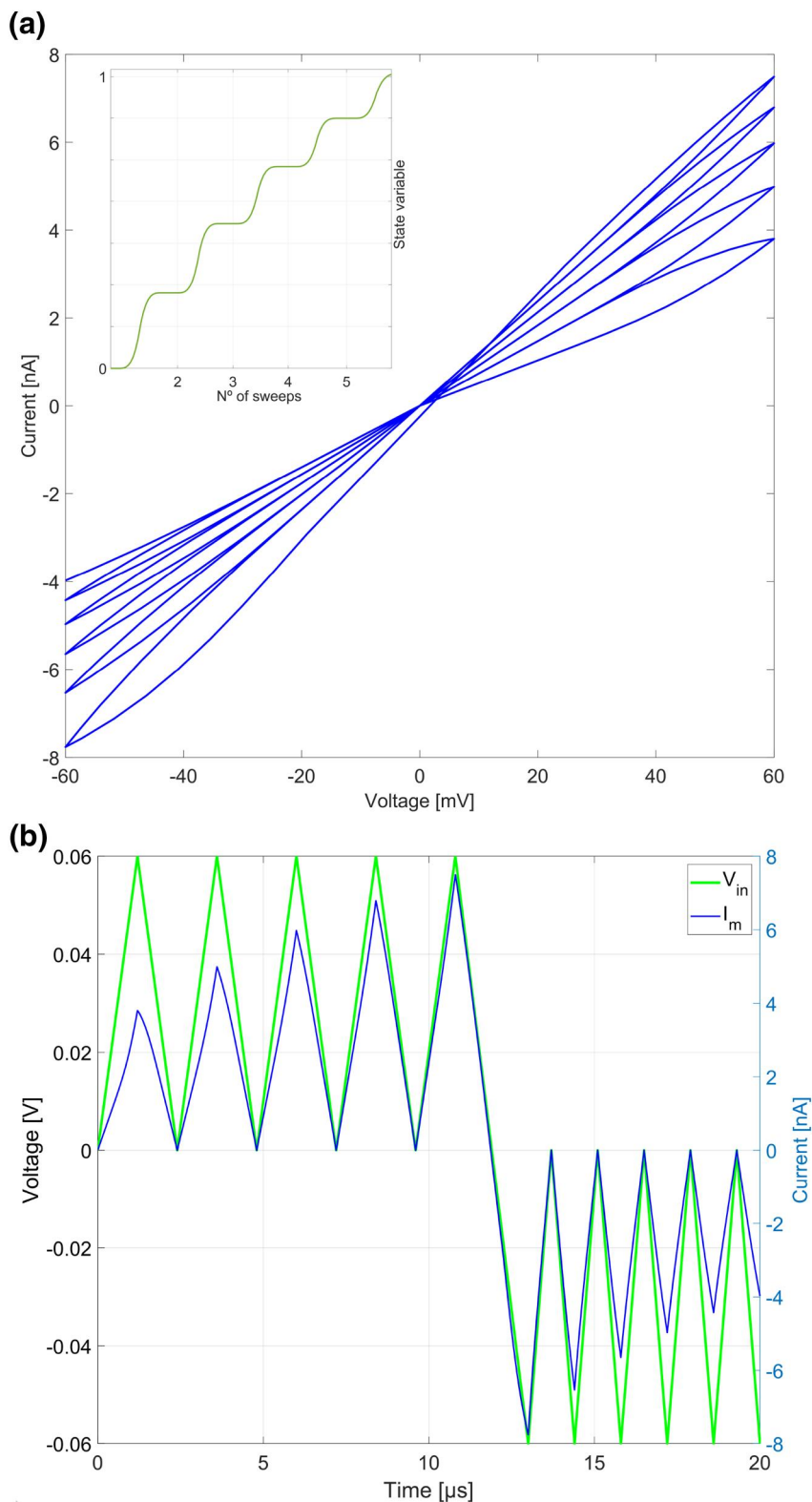


FIGURE 12 Analogue spike neuron 2 encoding an input signal

TABLE 2 Performance comparison of this work neurons and other VLSI neuron models

References	Process (nm)	Spike frequency (Hz)	Area (μm)	Power (W)	Energy efficiency (pJ/spike)
[30]	350	200	244	0.3–1.5 μ	900
[17]	65	1.9×10^6	120	78 μ	41
[31]	350	10^6	-	8–40 μ	8.5–9
[7]	180	33×10^6	-	—	9.3
[32]	180	$100\text{--}500 \times 10^3$	-	35p	2
[33]	90	100	442	40.2p	0.4
[34]	65	26×10^3	35	105p	0.004
[5]	65	15.7×10^3	31	30p	0.002
Analogue spike neuron 1	32	$10\text{--}20 \times 10^3$	≈ 21	4.75n	0.0485
Analogue spike neuron 2	32	$70\text{--}220 \times 10^3$	≈ 8	23.73p	0.00023

FIGURE 13 I-V characteristics of the memristor with synaptic behaviour using triangular pulses after adjust voltage parameters: (a) voltage sweeps of the adjusted synaptic memristor and (b) applied pulse train to synaptic memristor fitted



When evaluating only I_{nj} , the I&F neuron can also behave like an encoder, encoding the input current to a spike train, Figure 10. The number of spikes is proportional to the input current.

Finally, using default parameters such as $V_{DD} = 0.9$ V, $I_{nj} = 10$ nA, $V_{rfr} = 0.53$ V, $V_{pw} = 0.20$ V, and

$W_{NMOS} = 52$ nm, the power consumption of one neuron turns out to be 4.75 nW, and it is directly proportional of the spike frequency since more spikes will dissipate more energy. Thus all significant parameters that influence pulse frequency will direct impact on power consumption. More important to analyze than energy dissipation is energy efficiency (Joules/

TABLE 3 Fitting parameters for the memristor synapses

Parameters	Yakopcic characterization [23]	This work values
a_1	0.37×10^{-7}	0.37×10^{-7}
a_2	0.435×10^{-7}	0.435×10^{-7}
b	0.7	0.7
V_p	1.5 V	0.055 V
V_n	0.5 V	0.050 V
A_p	0.005	5.0×10^6
A_n	0.08	8.0×10^7
x_p	0.2	0.2
x_n	0.5	0.5
α_p	1.2	1.2
α_n	3.0	3.0
x_o	0.1	0.2
η	1	1

spike); for the regular neuron, a 48.5 fJ/spike was achieved, being directly dependent on variables that considerably change t_{high} and t_{low} characteristics. At the end of the next session, a table presents this work and other power dissipation and energy efficiency values.

4.1.2 | Analogue spike neuron 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, 0.15 V was selected to be the power supply and current source of $I_{nj} = 35$ pA. Using $C_{fb} = 5$ fF, the capacitor divider works the same way of the first neuron. The difference is that the parasitic capacitance of the first inverter has significant influence now. Figure 11 presents the spike pulse for the low-power Spike Neuron with V_{mem} and V_{out} . Comparing with Figure 9, 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 Figure 1 with a voltage peak near 70 mV.

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.

The low-power neuron can also behave like an encoder, as shown in Figure 10 presented in the last section. Applying a sine wave for injection current, the spikes fire proportional to the current as shown in Figure 12. I_{nj} will alter spike pulse time

and the frequency in which it occurs as well as the regular neuron.

Lastly, to extract the power dissipation and energy efficiency a standard model was used, without significant changes to variables, $V_{\text{DD}} = 150$ mV, $I_{nj} = 38$ pA, and $W_{\text{NMOS}}/P_{\text{MOS}} = 52$ nm. The power consumption for a single neuron was 23.73 pW and the energy efficiency of a single spike 0.232 fJ/Spike. As with the first neuron presented, the parameters that modify pulse frequency and pulse size directly influence these values.

Concluding the simulation results for neurons section, Table 2 presents the results of this work and other 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 concerning the processing capacity of neurons. So, the two proposed circuits have achieved good results among these projects.

There are two that deserve more attention, from all eight referenced models, Indiveri [30] and Danneville [5]. 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 900 pJ/Spike to 0.0485 fJ/Spike and the low power I&F from 2 to 0.232 fJ/Spike. An estimate of the areas occupied by analogue neurons 1 and 2 was obtained scaling down previous implementations [5] and were also included in the table.

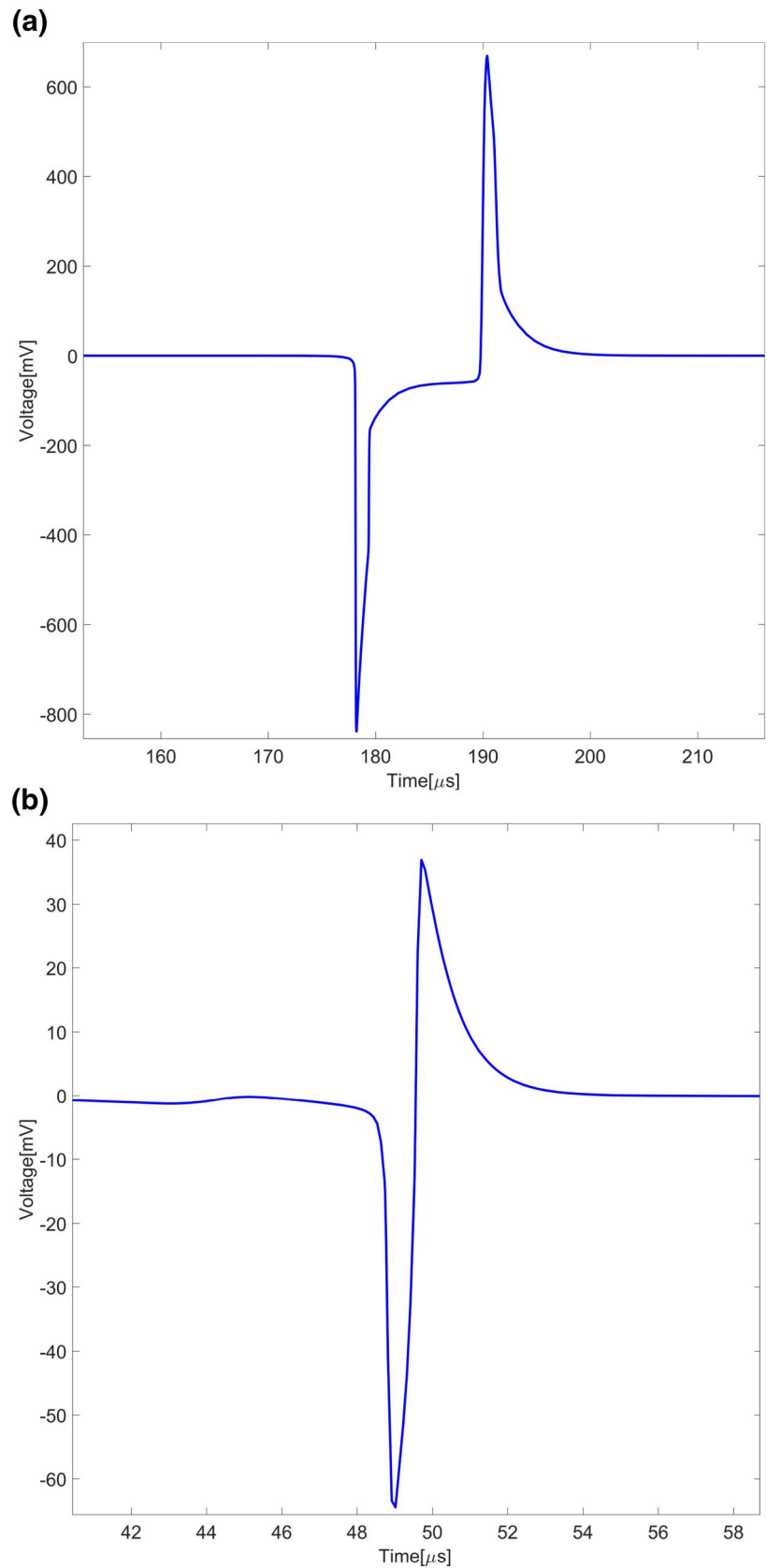
4.1.3 | Memristor synapse

Given the generic model of memristor presented, Yakopcic also demonstrates a synaptic representation of his memristor modelling the device-related on [10]. It was possible to extract the necessary parameters for our composition of the final memristor used through these data. A model to operate with low voltages, a higher frequency than the one presented by him and to perform as the characteristic I-V curve for the synaptic memristor.

To configure each of the fitting parameters presented, Yakopcic prepared a detailed study [35] showing the paths to design a wide range of memristors by changing each variable, creating a device that operates as desired. With this, it was possible to verify which parameters required adjustment. First, the threshold voltages, V_n and V_p , setting the points where the conductivity is biggest for both positive and negative slope. Also, adjusting A_n and A_p , which represents the variation on conductivity over time, reaching the neurons' frequencies of operation.

In order to adjust the model to perform using triangular peaks of 60 mV (an approximation for the neurons action potential), as indicated in Figure 13b, it is necessary to make, $V_p = 55$ and $V_n = 50$ mV to set threshold voltages. The variable responsible for fitting how quickly the state variable

FIGURE 14 Spike-timing-dependent plasticity curve: (a) regular neuron, and (b) low power neuron



changes once the threshold is surpassed is A_p and A_n , changing it to 10^9 scales, will adjust the time response to microseconds, resulting in Figure 13a.

Table 3 presents a list of all fitting parameters for the memristor synapse model proposed by Yakopcic and the ones adjusted to perform alongside this work.

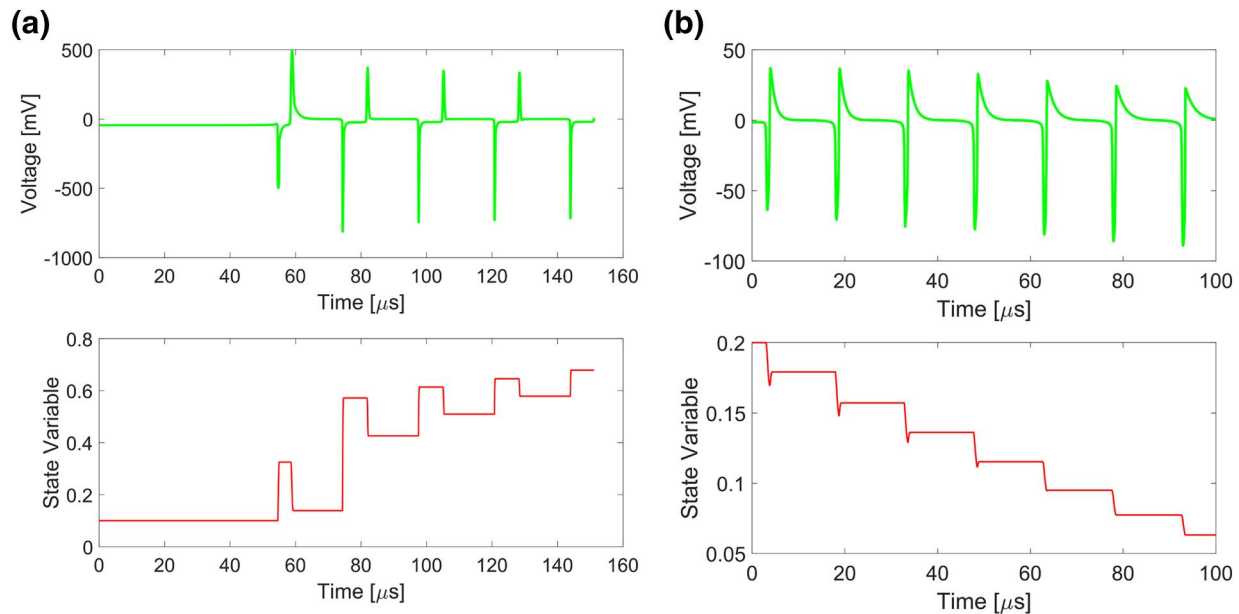


FIGURE 15 Synaptic weights updated by spike-timing-dependent plasticity: (a) regular neuron, and (b) low power neuron

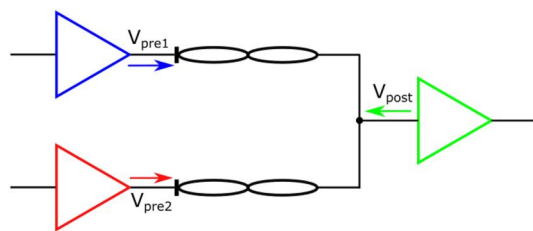


FIGURE 16 The simple spike neuron network

4.1.4 | STDP learning

After presenting the results for each circuit block of this project, it is time to show these pieces working together.

From the configuration neuron-memristor-neuron in Figure 6, the STDP curve, responsible for adjustment on the state variable or synaptic weights generated at the memristor between the implemented neurons are revealed in Figure 14 for each neuron circuit.

To show how these two curves are capable of modifying the synaptic weights, Figure 15 shows both neurons the gradual alteration of the memristor's state variable, which is precisely the change in the synaptic weights.

The increase and decrease of synaptic weight depend on how the STDP is acting. It is also possible to notice that in Figure 15b, 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 the state variable keeps its value constant if nothing is applied in both cases.

Finally, a system was created to show how different pre-synaptic pulses from two different neurons in a different time change the synaptic weights of their respective synapses using

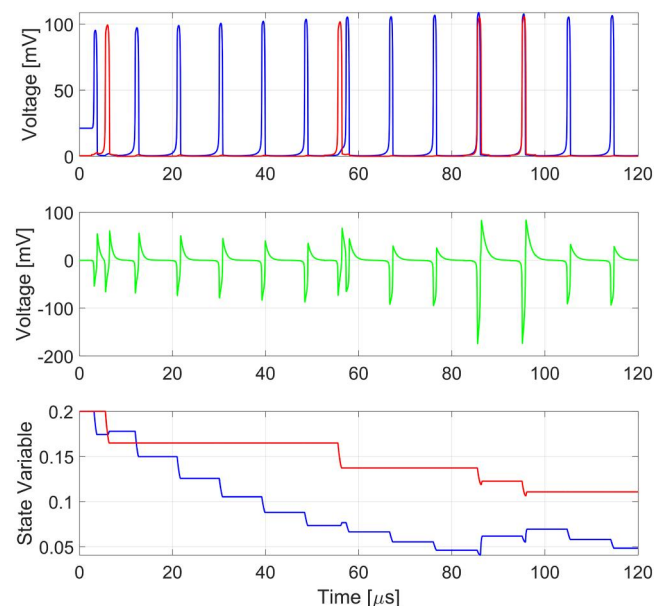


FIGURE 17 Neurons operation and spike-timing-dependent plasticity learning with two input neurons and one output neuron. Each colour corresponds to a specific neuron in Figure 16

the same postsynaptic signal from a third neuron. In other words, Figure 16 presents the beginning of a spike neural network and how the circuits presented manage to work together considering the blue neuron, a regular frequency of spikes and the red one a random spike generator.

In Figure 17, 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 postsynaptic signal. This phenomenon will

occur at all times, even in more extensive neural networks. So the synapses of the existing neural network are updated. In other words, they are learning patterns.

5 | CONCLUSION

This work proposed and validated the basis of a spike neural network architecture using two different types of analogue CMOS neurons and configuring a memristor to function as a synapse enabling the STDP learning process aiming to achieve a low-power neuromorphic configuration. Through a bibliographic review covering all aspects of other studies, and evaluating methods for simulations and data acquisition, two existing neuron designs were implemented and simulated using a different CMOS process technology from previous works [5, 6, 30]. 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 the less effective area. In the second part, a generic memristor model was configured to work together with these neurons making the creation of a neuromorphic system able to update synaptic weights, in other words, to perform an online training of the neural network.

Studies on the performance of networks are also foreseen, going beyond 32 nm technology, to assess how power dissipation would be modified with the reduction of the technological process. In the future, the SNN proposed in this work will be enlarged for performing a specific application concerning an image recognition task by simulation. Also, a training methodology will be proposed. If necessary, the network circuit will be adjusted to fit in the application. Depending on the performance results provided, in the future, the physical design and fabrication are planned.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. Janaina is grateful to PQ/CNPq for financial support. The authors also kindly acknowledge the site <http://ptm.asu.edu> for the models.

ORCID

Gabriel Maranhão  <https://orcid.org/0000-0001-5651-6661>
Janaina Gonçalves Guimarães  <https://orcid.org/0000-0002-6551-5094>

REFERENCES

- Moore, G.E.: Cramming more components onto integrated circuits, reprinted from electronics, volume 38, number 8, april 19, 1965 114 ff. IEEE Solid-State Circuits Society Newsletter. 11(3), 33–35 (2006)
- 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, 110(33), 13642–13647 (2013)
- Marani, E., Lakke, E.A.J.F.: Peripheral nervous system topics. In: Mai, J. K., Paxinos, G. (Eds.) The human nervous system. pp. 82–140. Academic Press, Elsevier (2012). <https://doi.org/10.1016/B978-0-12-374236-0.10004-5>
- Brady, S., et al.: Basic neurochemistry: principles of molecular, cellular, and medical neurobiology. Academic Press (Elsevier Science) (2012)
- Danneville, F., et al.: A Sub-35 pW Axon-Hillock artificial neuron circuit. Solid State Electron. 153, 88–92 (2019)
- Srivastava, S., Rathod, S.S.: Silicon neuron-analog cmos vlsi implementation and analysis at 180 nm. In: 2016 3rd International Conference on Devices, Circuits and Systems (ICDCS), pp. 28–32 (2016)
- 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. 62(11), 1088–1092 (2015)
- Yan, X., et al.: Robust Ag/ZrO₂/WS₂/Pt memristor for neuromorphic computing. ACS Appl. Mater. Interfaces. 11(51), 48029–48038 (2019)
- Zhou, Z., et al.: Synapse behavior characterization and physical mechanism of a TiN/SiO_x/p-Si tunnelling memristor device. J. Mater. Chem. C. 7(6), 1561–1567 (2019)
- Jo, S.H., et al.: Nanoscale memristor device as synapse in neuromorphic systems. Nano Lett. 10(4), 1297–1301 (2010)
- Deng, L., et al.: Rethinking the performance comparison between SNNs and ANNs. Neural Networks. 121, 294–307 (2020)
- Indiveri, G.: Computation in neuromorphic analog vlsi systems. In: Tagliaferri, R., Marinaro, M. (eds.) Neural Nets WIRN Vietri-01, pp. 3–20. Springer London, London (2002)
- Mead, C.: Neuromorphic electronic systems. Proceedings of the IEEE, 78(10), 1629–1636 (1990)
- Poon, C.S., Zhou, K.: Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities. Front Neurosci. 5, 108 (2011)
- Bartolozzi, C., Indiveri, G.: Synaptic dynamics in analog VLSI. Neural Comput. 19(10), 2581–2603 (2007)
- Nawrocki, R.A., Voyles, R.M., Shaheen, S.E.: A mini review of neuromorphic architectures and implementations. IEEE Trans. Electron. Dev. 63(10), 3819–3829 (2016)
- Joubert, A., et al.: Hardware spiking neurons design: analog or digital? The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–5 (2012)
- 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. 65(12), 4299–4312 (2018)
- Aamir, S.A., et al.: A highly tunable 65-nm CMOS LIF neuron for a large scale neuromorphic system. European Solid-State Circuits Conference, pp. 71–74 (2016)
- Livi, P., Indiveri, G.: A current-mode conductance-based silicon neuron for address-event neuromorphic systems. In: IEEE International Symposium on Circuits and Systems, IEEE, pp. 2898–2901 (2009)
- Hodgkin, A.L., Huxley, A.F.: A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. 117(4), 500–544 (1952)
- Newhall, K.A., et al.: Dynamics of current-based, Poisson driven, integrate-and-fire neuronal networks. Commun. Math. Sci. 8(2), 541–600 (2010)
- Yakopcic, C., et al.: Generalized memristive device SPICE model and its application in circuit design. IEEE Trans. Comput. Aided Des. Integrated Circ. Syst. 32(8), 1201–1214 (2013)
- Sheu, B.J., et al.: BSIM: Berkeley short-channel igfet model for MOS transistors. IEEE J. Solid State Circ. 22(4), 558–566 (1987)
- Zhao, W., Cao, Y.: New generation of predictive technology model for sub-45 nm early design exploration. IEEE Trans. Electron. Dev. 53(11), 2816–2823 (2006)
- Harris, D.M., et al.: Evaluation of predictive technology models. Microelectron. J. 80, 7–17 (2018)
- Zhao, W., et al.: Predictive technology modeling for 32nm low power design. In: 2007 International Semiconductor Device Research Symposium, ISDRS. 2007 International Semiconductor Device Research Symposium, ISDRS, pp. 1–2 (2007)
- Zhao W., Cao Y. New Generation of Predictive Technology Model for Sub-45 nm Early Design Exploration. IEEE Transactions on Electron Devices 53, (11), 2816–2823 (2006). <https://doi.org/10.1109/ted.2006.884077>

29. Maranhão, G., Guimarães, J.G.: Integrate and fire neuron implementation using cmos predictive technology model for 32nm. In: *2019 34th Symposium on Microelectronics Technology and Devices (SBMicro)*, IEEE, vol. 1, pp. 1–4 (2019)
30. Indiveri, G., Chicca, E., Douglas, R.: A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity. *IEEE Trans Neural Network.* 17(1), 211–221 (2006)
31. Wijekoon, J.H.B., Dudek, P.: Compact silicon neuron circuit with spiking and bursting behaviour. *Neural Networks.* 21(2-3), 524–534 (2008)
32. 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*, IEEE, pp. 366–369.(2017)
33. Cruz-Albrecht, J.M., Yung, M.W., Srinivasa, N.: Energy-efficient neuron, synapse and STDP integrated circuits. *IEEE Transactions on Biomedical Circuits and Systems.* 6(3), 246–256 (2012)
34. Sourikopoulos, I., et al.: A 4-fj/spike artificial neuron in 65 nm cmos technology. *Front. Neurosci.* 11, 123 (2017)
35. Yakopcic, C., et al.: Memristor model optimization based on parameter extraction from device characterization data. *IEEE Trans. Comput. Aided Des. Integrated Circ. Syst.* 39(5), 1084–1095 (2019)

How to cite this article: Maranhão G, Guimarães JG. Low-power hybrid memristor-CMOS spiking neuromorphic STDP learning system. *IET Circuits Devices Syst.* 2021;15:237–250. <https://doi.org/10.1049/cds2.12018>