International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 3
***Article Received:** 18 January 2020 **Revised:** 25 February 2020 **Accepted:** 10 March 2020 **Publication:** 31 March 2020*

LIF CMOS Neuron for Neuromorphic Computing

Dr. Mahesh Koti
Assistant Professor, Department of Electronics and Communication Engineering
PES College of Engineering, Mandya -01

Abstract – As Artificial Intelligence is paving way for advancement of computing age, the need for efficient computation is coming to the forefront as most pivotal factor. Analog neurons and neuromorphic computing have emerged as promising candidates to address the need of emerging computational demand. This paper presents a LIF (Leaky Integrate and Fire) CMOS spiking neuron design, which is implemented and simulated in GPDK 180nm technology. The functionality, power efficiency and configuration viability of the neuron is tested through meticulously planned test cases. The neuron layout fits in $0.02mm^2$ and has approximate energy efficiency of 10pJ/Spike.

Keywords: *LIF Neuron, Memristor, STDP, CMOS Neuron, Spiking Neuron.*

1. Introduction

Three components make up the Von-Neumann architecture: memory, a central processing unit (CPU), and a connection between the two. Because data must be transmitted from memory to the CPU via a parallel data bus, the gap between the memory and CPU causes a bottleneck in processing. The Von Neumann architecture-based CPUs can process logic and compute at very high speeds, but they struggle with many tasks, like video motion detection and image recognition, and they are not designed to handle big data processing workloads. This architecture is not apt for implementing neural networks, which are derived from functioning of the brain [1]. Human brain is energy efficient computing system: functions such as vision, object recognition, speech recognition, and language translation are processes with ease by the massive parallel neuron network of the brain [4]. Brain-inspired architectures perform computing tasks by communicating spikes between large network of neurons, which are connected through synapses between each other and locally store memory in form of synaptic strength [2]. It mimics the biological neural cell where synapses receive the synaptic spikes from the other connected neurons. Designing a neural network on par with the brain requires understanding of the brains elementary function model and its network along with its essential synaptic communication protocols. A number of neuron models and networking structure have been developed over years, concepts like winner -take- all (WTA) and Spike timing dependent plasticity (STDP) [3] are an outcome of these efforts. Before the paper describes the design of CMOS neuron with its integral sub blocks, an outlook of neuron models and its relevant developments are presented. This is followed with discussion of LIF neuron structure and its sub blocks Operational amplifier, comparator, spike generator and finally the simulation results of testing the neuron.

2. Related works

The neural network comprises of neurons and synapses that make up the network, it is important to know how those components operate, and how they interact. This section discusses the different neuron models and synapse models.

2.1 Neuron models:

A biological neuron is made of cell, axon, and dendrite. The axon transmits information out of the neuron. Dendrites transmit information to the cell body and acts as receiver. Neurons interact by exchanging chemical molecules that gives signaling mechanism referred as synaptic spikes. The neuron works by charging and discharging resulting in exchange of spikes. This spiking consumes electrical energy in a neuron that can reach a certain threshold, which will cause the neuron to “fire” or, in biological terms, produce an action force that will travel through the axon of the neuron to affect the charging of other neurons through synapses. A number of neuron models have appeared over years which have been formulated to mimic the electrophysiological phenomenon of the biological neurons. Following are the most celebrated models of the neurons.

Hodgkin-Huxley model: The most popular biologically plausible neuron model is the Hodgkin-Huxley model [5]. The Hodgkin-Huxley model was first introduced in 1952 and is a relatively complex neuron model, consisting of different nonlinear equations that define neuron behavior about ion transfer within and outside the neuron.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 3
***Article Received:** 18 January 2020 **Revised:** 25 February 2020 **Accepted:** 10 March 2020 **Publication:** 31 March 2020*

Morris Lecar model: The simplest, but most biologically plausible neuron model is the Morris Lecar model, which reduces the model to an indirect two-dimensional equation [6]. It is a commonly used model in neuroscience and neuromorphic systems. Among these models, CMOS integrate and fire model and Conductance based silicon neuron model are the most adopted models [7], [4]. These models are explicitly considered for discussion in this work.

Conductance based silicon neuron: Conductance-based silicon neuron was reported by R. J. Wang, T. et.al. [7], where in synapse and soma are implemented as first-order low-pass filter-based based on "Tau-cell". Conductance-based synapse was adopted instead of using a separate low pass filter for excitatory and inhibitory inputs. This reduces the area and makes the circuit more compact with the penalty of identical time constants.

A CMOS leaky integrate and fire model: Xinyu Wu, Vishal Saxena et.al [4], have presented a CMOS spiking neuron design, which was implemented in a 0.18μm CMOS technology. The CMOS neuron design is based on leaky integrate and fire model [1] and consists of an operational amplifier, phase controller, comparator, and spike generator sub-blocks along with membrane capacitance C mem, a leaky resistor R leaky and switches SW1 to SW3. The CMOS neuron was designed to operate in integration and firing modes, i.e., dual-mode. The mode of operation is controlled by the phase signal Φ fire and Φ int.

2.2 Synapse models:

Just as some neuromorphic work focuses primarily on neuron models, there was also a focus on developing synapse models. synapse models are often simpler unless they try to explicitly model biological behavior. One popular insertion of complex synapse models is the plasticity method, which causes the strength of the neuron or the amount of weight to change over time. Plastic methods are related to learning in the biological brain. The artificial neural network models are closer to biological neurons in the brain. If the neuron receives the current from other neurons and membrane potential exceeds a threshold voltage, then the output spike will be generated and delivered to other neurons. Therefore, spike timing is considered in the neuron model. Learning is an essential criterion for the working of neuromorphic computing networks, the synapses adopt their weight in accordance with the learning. Different learning mechanisms have emerged over years. Here a brief account of the winner-take-all (WTA) and Spike timing-dependent plasticity (STDP) learning mechanism is given.

Winner -take- all and Spike timing dependent plasticity: Idongesit E. Ebong, et.al, highlighted the two basic learning rules: They are winner -take- all (WTA) and Spike timing dependent plasticity (STDP).

Winner-Take-All: WTA is an algorithm in which one neuron clearly inhibits its neighbors in order to take the prize. There exists a WTA variation that is designed for design flexibility called the kWTA. In kWTA, two or more neurons might have ended up winning the prize, but the concept remains the same [10].

Spike-timing-dependent plasticity: STDP is a form of long-term synaptic plasticity in which the precise order and timing of pre-synaptic and post-synaptic action potentials trigger long-term synaptic potentiation or depression [8], [3]. There are two forms of STDP: Asymmetric and symmetric. Symmetric STDP performs the same weight adjustment independent of the spike order between the pre-neuron and post-neuron. Asymmetric STDP reverses weight adjustment based on the spike time difference between the pre-neuron and the post-neuron. In the asymmetric STDP case, if the pre neuron spikes before the post neuron, the synaptic weight is increased. If the order of spikes is reversed, the synaptic weight is decreased. In both cases, the larger is the duration between the pre-neuron and the post-neuron spikes, the lesser is the magnitude of the synaptic change. Most circuit implementations take advantage of the asymmetric implementation [3]. Idongesit E. Ebong, et.al have implemented asymmetric STDP based on the equation.

$$\Delta w(t2-t1) = \begin{cases} Ae^{-(t2-t1)\tau+}, & \text{t2-t1>} \\ -Ae^{i(t2-t1)\tau-}, & \text{t2-t1<0} \end{cases} \quad (2.2.1)$$

Where, Δw = Change in synaptic weight.

t2-t1 = Time difference between preneuron and the post neuron.

A+ = Maximum change in the positive direction.

A-= Maximum change in the negative direction.

τ = Time constants.

3. LIF Neuron Structure

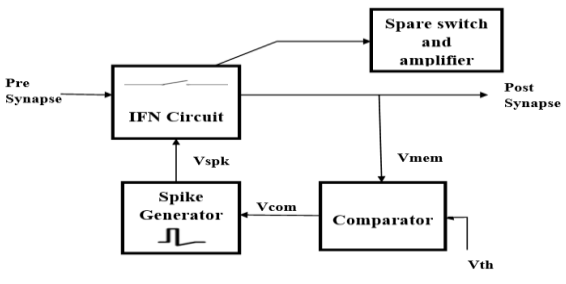

Fig.3.1 Block diagram of the CMOS Neuron Circuit.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 3
Article Received: *18 January 2020* ***Revised:*** *25 February 2020* ***Accepted:*** *10 March 2020* ***Publication:*** *31 March 2020*

The Fig.3.1 shows the block diagram of CMOS neuron, inspired by the model [2]. It consists of IFN circuit, comparator and STDP-compatible spike generator. The IFN circuits are designed to generate spikes to match spiking behaviors of biological neurons. The IFN circuit integrates the input synaptic currents and generates Vmem signal as output which is compared with the threshold voltage Vth. Whenever Vmem exceeds threshold, comparator generates Vcom signal which drives the spike generator. As shown in Fig.3.2, the spare block consists of matrix of switches and operational amplifier that provides the reconfigurability for the neurons with respect to the driving capability and noise reduction. It also makes the implementation of area and power efficient.

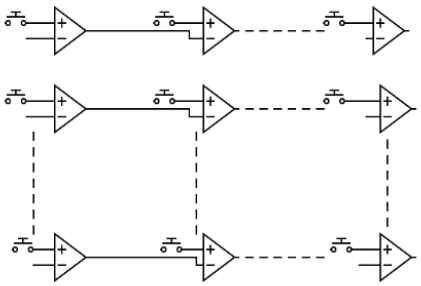

Fig.3.2. Spare switch and amplifier block

4. The Design of CMOS Neuron

This section discusses the design of CMOS neuron circuit blocks.

4.1 Differential Amplifier:

To integrate the spikes arising from pre synapses, an integrator was required to be designed [2]. Differential amplifier founds an important integral part of this integrator. Here the design of a differential amplifier with a cascode load is discussed.

The Differential amplifier forms the first block of the LIF neuron. As shown in Fig.4.1.1, comprising of cascode load is designed with using NMOS input differential pairs NM3, NM4 because these input type can perform larger output gain compared to PMOS input type. the cascode topology is employed for achieving a high gain. PM1 and PM2 form a current mirror. NM5 and NM6 are used to provide biasing to this amplifier. The body of all the transistors is connected with their source. Current from NM5 will divide equally into two arms of PM1, NM1, NM3, and PM2, NM2, NM4.

The basic differential amplifier circuit converts the difference between its two input voltages to a corresponding current as of its output. This circuit is used in the implementation of CMOS neurons. Table 1 shows the design specifications and Table 2 shows the dimensions for each transistor according to the W/L ratio.

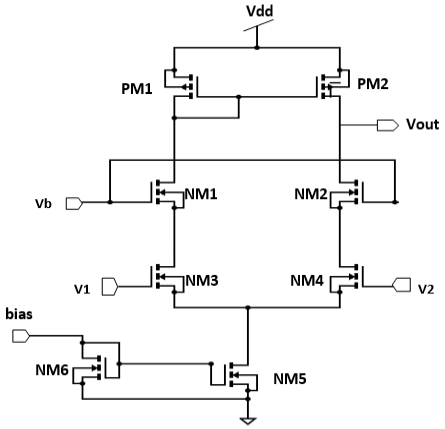

Fig.4.1.1: Cascode Differential Amplifier

Table I Design Specification

Open loop Gain	100 V/V
Power Supply	2.5v
Maximum Power Dissipation	≤ 2mw
Slew Rate	10 V/us
Gain B/W at -3 db gain	100Khz
ICMR	≤2v

The Differential amplifier is designed for 180 nm technology for the following specifications: -

STEP 1 — To find DC current from Slew Rate

$$SR = \frac{I}{c} \qquad (4.1.1)$$

STEP 2 — To find aspect ratios of MOSFETs PM1 and PM2 from Input Common-Mode Range. Let the input common-mode range be v_{cmax} to v_{cmin}Let Vin2 be grounded and v_{in1} be equal to v_{cmax}.

$$\left.\begin{aligned} v_x &\geq v_{in1} - v_{th} \\ v_x &\geq v_{cmax} - v_{th} \end{aligned}\right\} \qquad (4.1.2)$$

$$i_{pm1} = \frac{i}{2} \qquad (4.1.3)$$

$$i_{pm1} = \mu P \cdot \frac{c_{ox}}{2}\left(\frac{\omega}{L}\right)[v_{ds} - v_{th}]^2 \qquad (4.1.4)$$

$$\left(\frac{w}{L}\right)_{pm1} = \frac{2 \cdot i_{pm1}}{\mu p \cdot c_{0x} \cdot [vx - v_{th}]^2} \qquad (4.1.5)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 3
Article Received: *18 January 2020* ***Revised:*** *25 February 2020* ***Accepted:*** *10 March 2020* ***Publication:*** *31 March 2020*

STEP 3 — To find aspect ratios of MOSFETs 5 and 6 from Input Common Mode Range.

We know that

$$i_1 = \mu_n \cdot \frac{c_{0x}}{2} \cdot \left(\frac{w}{L}\right)_1 \cdot \left[v_{gs} - v_{th}\right]^2 \quad (4.1.6)$$

Therefore,

$$v_{gs} = v_{th} + \sqrt{\frac{2 \cdot i_1}{u_n c_{03} \cdot \left(\frac{w}{L}\right)}} \quad (4.1.7)$$

$$\left.\begin{aligned} v_{cmin} &\geq v_{gs} + v_{dsat} \\ v_{dsat} &\geq v_{gs} + v_{cmin} \end{aligned}\right\} \quad (4.1.8)$$

We know, $i_{5=}i$

$$\left(\frac{w}{L}\right)_5 = \frac{2 \cdot i_5}{\mu n \cdot c_{ox} \cdot [v_{dsat}]^2} \quad (4.1.9)$$

Transistor	W/L
PM 1	500n/1u
PM 2	500n/1u
NM 1	8u/180n
NM 2	8u/180n
NM 3	8u/180n
NM 4	8u/180n
NM 5	2u/180n
NM 6	2u/180n

Table-2: Transistor Dimensions for the differential amplifier

The integrated signal is continuously monitored by the comparator for threshold crossing**.**

4.2 Comparator:

The Comparator is used to compare the neuron membrane potential (Vmem) against the firing threshold (Vth). As shown in Fig.4.2.1, the comparator circuit uses 11 transistors. The design is based on differential pair MC1 and MC2, the bias transistor MC0, load devices MC3 and MC4 and the cross-coupled pair MC5 and MC6. The purpose of current mirror and transistors MC7, MC8, MC9, MC10 is to provide additional gain and to covert the differential output to a single-ended output. The bias transistor MC0 sets the current flowing the circuit. The full derivation of the comparator, as well as stability conditions, are presented in Gregorian's work [8]. However, we mention the formulas as follows:

$$V_{trig+} = \sqrt{\frac{i_0}{K'(W/L)_1}} \cdot \frac{\sqrt{\alpha}-1}{\sqrt{1+\alpha}} \quad (4.2.1)$$

$$V_{trig-} = \sqrt{\frac{i_0}{K'(W/L)_1}} \cdot \frac{1-\sqrt{\alpha}}{\sqrt{1+\alpha}} \quad (4.2.2)$$

$$\alpha = \frac{(W/L)_5}{(W/L)_3} = \alpha = \frac{(W/L)_6}{(W/L)_4} \quad (4.2.3)$$

$$k' = \frac{1}{2} . \mu * c_{Ox} \quad (4.2.4)$$

Where, W/L= transistor width to length ratio

c_{ox} = oxide capacitance

μ = mobility

α = positive feedback factor

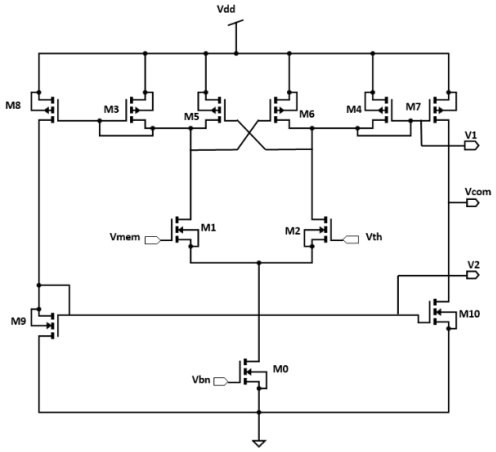

Fig.4.2.1 Circuit diagram of Comparator

In this circuit, there are two paths of feedback. The first feedback path is negative feedback through common-source node MC1 and MC2 and the second feedback path is the positive feedback, through the gate-drain connection of MC5 and MC6.

Transistor	W/L
MC8	400n/1u
MC3	800n/4u
MC5	400n/1u
MC6	800n/4u
MC4	2u/500n
MC7	400n/1u
MC1	12u/180n
MC2	12u/180n
NM9	2u/180n
NM10	2u/180n
MC0	1u/500n

Table 2: Transistor dimensions

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 3
Article Received: *18 January 2020* ***Revised:*** *25 February 2020* ***Accepted:*** *10 March 2020* ***Publication:*** *31 March 2020*

4.3 Spike Generator:

The Spike generator is designed, produced the customized spike with reconfigurable timing and amplitude parameter settings.

As shown in Fig.4.3.1, the spike generator circuit is designed by selecting the voltage reference levels and an RC charging circuit for the positive pulse and negative tail respectively. A shape of the action potential Vspk influences the STDP learning function. A biological like STDP pulse with exponential rising edges is very difficult to realize in the circuit. However, a bio-inspired STDP pulse can be achieved with a simpler action potential shape: a short narrow positive pulse of large amplitude followed by a larger slowly decreasing negative tail [11].

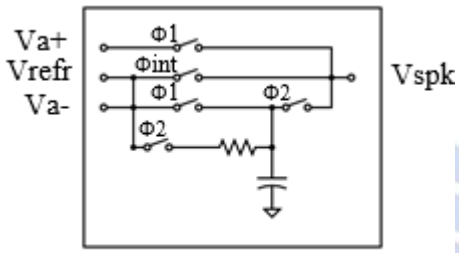

Fig.4.3.1: Spike generator

5 Simulation Result

The circuits are designed and simulation is done using cadence virtuoso in 0.18um CMOS technology. The layout simulation is carried out in a cadence virtuoso layout XL design environment. The design rule check (DRC) and layout versus schematic (LVS) were executed successfully without errors.

5.1 Schematic testbench and simulation of Differential amplifier:

In Figure 5.1.1, a test bench is used to check the transient response of the amplifier. In this case, a sine wave signal is applied to the two inputs, and the effective value of the output signal is the voltage difference between the two terminals. Fig.5.1.2 shows the transient response of the differential amplifier. Simulation is done by taking the parameters Vdd=1.8V, V1 and V2=1mv amplitude, frequency=10khz, offset=500mV, bias current Idc=232uA.

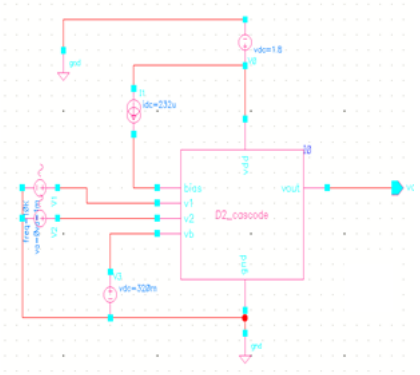

Fig.5.1.1 Testbench for Differential amplifier

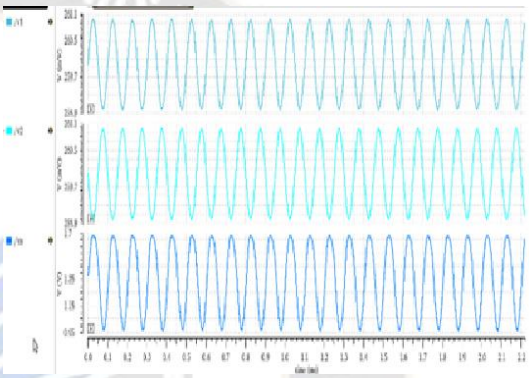

Fig.5.1.2. Transient Response of the differential amplifier

5.2 Simulated Outputs of the Comparator:

The transient behavior of the comparator is illustrated in Fig.5.2.1, The membrane potential Vmem is compared with a threshold voltage Vth, crossing which gives the output signals V1, V2, Vcom and the overall circuit performance is depending on the devices size and dimensions which are shown in Table 2. Here we have taken the parameters as VDD = 2.2 V, pulse width = 2ms, period = 4ms.

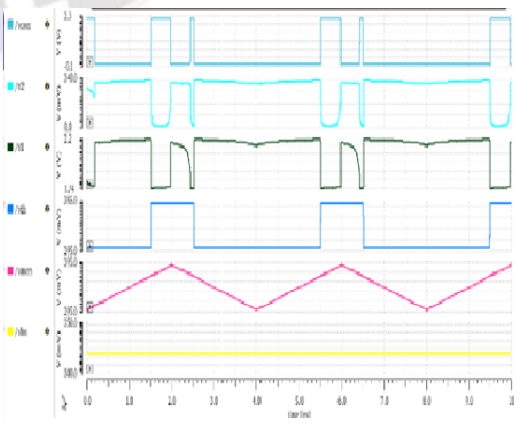

Fig.5.2.1: Transient simulation of the comparator

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 3
Article Received: *18 January 2020* ***Revised:*** *25 February 2020* ***Accepted:*** *10 March 2020* ***Publication:*** *31 March 2020*

5.3 Simulated Outputs of the Spike generator circuit:

The transient response of the spike generator is shown in Fig.5.3.1. For controlling the spike generation, a digital phase controller generates two non-overlapping control signals Φint and Φfire, together with another two signals implemented using pulse circuits, Φ1 for the pulse and Φ2 for the negative tail. For instance, spike parameters Va+ =3V, Va- = 1V and Vrefr = 2V were chosen for a device.

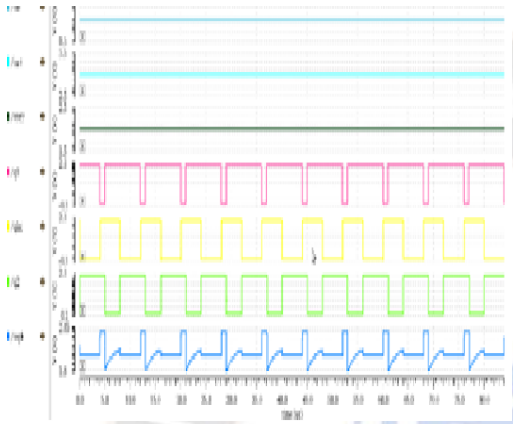

Fig.5.3.1. Transient response of the spike generator circuit

To optimize the neuron circuit for the area and power dissipation, the layouts are handcrafted. This resulted in effective neuron implementation with the following figure of merits. The schematic and layout of the comparator design are shown in Fig.5.3.1.

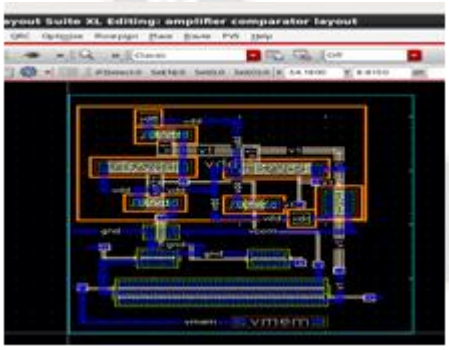

Fig.5.3.1. A layout design of the comparator

Conclusion

This paper presented a LIF neuron design for a brain-like neuromorphic computing system. It combines the CMOS design of the IFN circuit based on the operational amplifier with comparator and spike generator circuits. The design was implemented in a 180 nm CMOS process. Simulation results showed that the neuron is effective while consuming energy 12PJ/spike, which is on par in computation compared to the circuits that have been published till now.

References

1. P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, et al., “A million spiking-neuron integrated circuit with a scalable communication network and interface,” Science, vol. 345, no. 6197, pp. 668–673, Aug. 2014.
2. Xinyu Wu, Student Member, IEEE, Vishal Saxena, Member, IEEE, Kehan Zhu, Student Member, IEEE, and Sakkara ani Balagopal “A CMOS Spiking Neuron for Brain-Inspired Neural Networks with Resistive Synapses and In-Situ Learning”.
3. Ebong and P. Mazumder, “CMOS and memristor-based neural network design for position detection,” Proceedings of the IEEE, vol. 100, no. 6, pp. 2050 - 2060, 2012.
4. Xinyu Wu, Student Member, IEEE, Vishal Saxena, Member, IEEE, and Kehan Zhu, “Homogeneous Spiking Neuromorphic System for Real-World Pattern Recognition”.
5. Hodgkin and A. Huxley, “A quantitative description of membrane current and its application to conduction and excitation in nerve,” The Journal of physiology pp. 500-544, 1952.
6. Borisyuk, “Morris–lecar model,” in Encyclopedia of Computational Neuroscience. Springer, 2015, pp. 1758–1764.
7. R. Wang, T. J. Hamilton, et al., “A generalised conductance-based silicon neuron for large-scale spiking neural networks,” in International Symposium on Circuits and Systems, 2014, pp. 1564–1567.
8. R. Gregorian, Introduction to CMOS Op-Amps and comparators. John Wiley and sons, 1999.
9. M. R. Azghadi, “Spike-Based Synaptic Plasticity in Silicon: Design, Implementation, Application, and Challenges,” Proceedings of the IEEE, vol. 102, no. 5, 2014.
10. Giacomo Indiveri, A Current-Mode Hysteretic Winner-take-all Network, with Excitatory and Inhibitory Coupling, 2000.
11. S. Saïghi, C. G. Mayr, T. Serrano-Gotarredona, H. Schmidt, et al., “Plasticity in memristive devices for spiking neural networks,” Frontiers in Neuroscience, vol. 9, 2015.
12. G. Indiveri and S.-C. Liu, “Memory and information processing in neuromorphic systems,” Proceedings of the IEEE, vol. 103, no. 8, pp. 1379–1397, 2015.