

The Development and Application of Neuromorphic Silicon Neural Circuits

Yiheng Zheng^{1,a,*}

¹Maynooth International college of Engineering, Fuzhou University, Fuzhou, 350001, China
a. YIHENG.ZHENG.2023@mumail.ie

*corresponding author

Abstract: Silicon-based neuronal circuits, inspired by biomimetic principles, have garnered increasing scholarly interest and investigation. Building on the field of neuromorphic engineering, which focuses on designing and emulating circuits and systems that mimic the neural architecture of biological nervous systems, this study addresses high-speed modeling of large-scale neural systems and real-time applications of bio-inspired neural systems. The manuscript delineates the prevalent architectural components and methodologies for the fabrication of such circuits, offering a comprehensive survey of various neuromorphic silicon neurons (SiNs) that instantiate diverse computational paradigms. Additionally, the paper delves into the fundamental principles of competitive circuits, discussing and explaining their operational mechanisms and derivative applications. By comparing different design methodologies for basic circuits, it demonstrates how different fundamental circuits can be applied to implement spiking silicon neurons in various research fields. The study ultimately focuses on the sub-threshold operating region of these circuits and showcases the practical value of silicon neuron circuits through integrated circuits and vary-large-scale neural networks (VLSI) chip implementations. This approach highlights the potential of neuromorphic systems in achieving efficient and biologically plausible neural emulation.

Keywords: Neuromorphic silicon, Spiking behavior, Biomimetic neuron, Competing circuit.

1. Introduction

Neuromorphic engineering has emerged as a pivotal field in recent decades, aiming to replicate the architecture and functionality of biological neural systems through silicon-based circuits [1]. This interdisciplinary domain integrates principles from neuroscience, electrical engineering, and computer science to create systems capable of performing brain-like computations.

With rapid advancements in artificial intelligence (AI) and machine learning, neuromorphic systems hold significant potential to revolutionize real-time computing and low-power applications, both of which are essential for the next generation of AI systems and robotics. A key development within this field is the spiking neural network (SNN), which closely mirrors the spiking behavior of biological neurons, making it highly valuable for high-performance and energy-efficient computing tasks [2]. Such systems have been successfully deployed in the domains of pattern recognition, neurocomputational tasks, and sensory processing[3].

Despite these advancements, several challenges persist in the design and implementation of neuromorphic circuits, particularly concerning scalability, energy efficiency, and real-time performance [4]. Conventional computational architectures struggle to achieve efficient large-scale neural simulations while maintaining biological fidelity. Consequently, silicon-based neurons that operate in subthreshold regions have gained considerable attention due to their ability to emulate the spiking behavior of biological neurons with minimal power consumption [5]. These neurons leverage Very-Large-Scale Integration technology to overcome limitations related to scaling and energy efficiency, making them suitable for large-scale applications [6].

This paper investigates the design and implementation of spiking silicon neurons, with a particular focus on their subthreshold operation and competitive circuits, such as winner-takes-all (WTA) circuits. These circuits facilitate effective signal selection and processing within neuromorphic networks. By demonstrating their potential for real-time neural emulation, this study contributes to the advancement of neuromorphic computing systems, with promising applications in artificial intelligence, bio-inspired networks, and neuromorphic sensory systems.

2. Neuromorphic silicon neurons

Neuromorphic silicon neurons (SiNs) are designed to imitate the functions of biological neurons by utilizing analog VLSI (Very-Large-Scale Integration) technology. These neurons take advantage of CMOS transistors functioning in a subthreshold mode, thereby enabling a close approximation of the current-voltage characteristics exhibited by neuronal ion channels, while concurrently ensuring minimal power dissipation. Neuromorphic silicon neurons (SiNs) exhibit two key characteristics that distinguish them from other neural models.

Firstly, they operate using analog circuits, allowing them to directly replicate the spiking dynamics of biological neurons without relying on digital simulations. analog neuromorphic circuits share many common physical properties with protein channels in neurons. As a consequence, these types of circuits require far fewer transistors than digital approaches to emulating neural systems. Scholars tend to try to utilize the analog characteristics of transistors rather than simply using them as digital switches. [7]. This analog behavior is crucial as it aligns more closely with the inherently analog and variable nature of neural processing in the brain.

A secondary, yet pivotal, characteristic is the energy efficiency of SiNs. By studying the physics of computation of neural systems, and reproducing it through the physics of transistors biased in the subthreshold regime, neuromorphic engineering seeks to emulate biological neural computing systems efficiently, using the least amount of power and silicon real-estate possible[8].

In the “subthreshold” region of operation, the transistor current is exponentially dependent on its terminal voltages analogous to the exponential dependence of active populations of voltage-sensitive ionic channels as a function of the potential across the membrane of a neuron. This congruence permits the construction of compact circuits, utilizing a limited number of transistors, to implement electronic models of voltage-controlled conductance-based neurons and conductance-based synapses (as detailed in Reference [9]).

SiNs leverage CMOS transistors operating in the subthreshold region, consuming considerably less power than their digital counterparts. This low power consumption makes SiNs ideal for large-scale neural network implementations and real-time applications, especially in scenarios where energy efficiency is a priority.

3. Spiking model of neuromorphic silicon neural circuits

3.1. Subthreshold operation

Silicon neurons operate in the subthreshold region of MOSFETs, wherein the correlation between current and gate voltage assumes an exponential form. This characteristic facilitates the precise simulation of the membrane potential fluctuations observed in biological neurons, particularly below the action potential threshold. The key advantage of this subthreshold operation is its extremely low energy consumption, making it ideal for large-scale neuromorphic systems. Additionally, the exponential current-voltage characteristic enables accurate simulation of neuronal behaviors, such as the accumulation of inputs, threshold-triggered firing, and voltage resetting, which are essential for bio-inspired neural computation.

3.2. Basic competitive circuit - winner-takes-all circuit (WTA)

Focus on a particularly simple yet powerful model that describes a population of N homogeneous excitatory units that excites a single global inhibitory unit which feedback to inhibit all the excitatory units (Fig. 1). Which provides feedback inhibition. Small filled circles indicate inhibitory synapses and small empty circles indicate excitatory synapses.

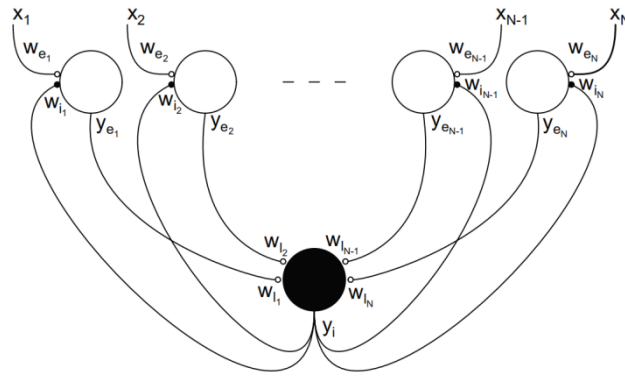


Figure 1: Network of N excitatory neurons (empty circles) projecting to one common inhibitory neuron (filled circle) [10].

To construct a stratified architectural framework, an analog competitive circuit model, drawing inspiration from biological systems, was conceptualized and designed.

For sake of simplicity, the study neglect the dynamics of the system and examine only the steady-state solutions. Dynamic properties of these networks and of other physiological models of competitive mechanisms are described in detail.

A winner-take-all (WTA) circuit (Fig.2) is a network of competing cells (neural, software, or hardware) that reports only the response of the cell that has the strongest activation while suppressing the responses of all other cells.

The circuit essentially implements a $\max()$ function. These circuits are typically used to implement and model competitive mechanisms among populations of neurons.

Assuming that the circuit has only two input signals, we divide it into three cases: input 1 is equal to input 2, input 1 is much larger than input 2, and input 1 is slightly larger than input 2.

Consequently, the input signal of greater magnitude invariably prevails, with its corresponding output signal predominating the allocated bias current. Competitive inhibition is fundamental, where nodes inhibit each other's activity, allowing only the strongest input to become the "winner." Non-linear activation functions, like ReLU, ensure that only inputs surpassing a certain threshold

produce significant output, suppressing weaker signals. Positive feedback mechanisms can further amplify the activity of the leading node, reinforcing its dominance.

Normalization processes limit overall input energy, ensuring only the strongest signal is amplified. These mechanisms collectively enable WTA circuits to effectively focus on the most prominent input, useful in applications like feature selection and signal enhancement.

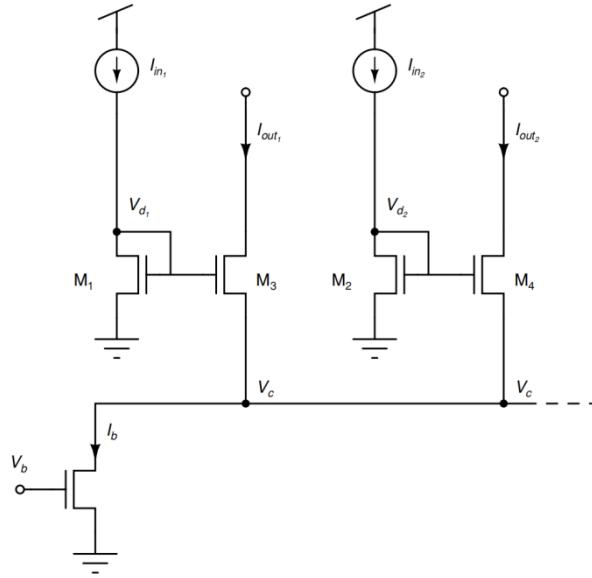


Figure 2: Winner take all circuit illustration.

3.3. Spiking behavior of silicon-based neural networks

The spiking behavior of silicon-based neural circuits primarily refers to the emulation of the firing (spiking) activities of biological neurons[11]. This behavioral attribute holds considerable significance within the disciplines of neuroscience and neuromorphic computing, as it aids researchers in understanding and modeling the functioning of the brain. The subsequent elements delineate pivotal characteristics of spiking behavior in silicon-based neural circuits: In silicon-based neural circuits, spiking behavior can reflect the intensity and characteristics of input signals through spiking patterns. These circuits encode complex information via the frequency, timing, and patterns of spikes. Typically, they employ models such as the Hodgkin-Huxley model, the Integrate-and-Fire model, and the Ishikari model to simulate the dynamic properties of neurons.

Additionally, silicon-based neural circuits can implement spike-timing-dependent plasticity (STDP), a synaptic plasticity mechanism that adjusts synaptic strength based on the temporal relationship of spikes, which is crucial for simulating learning and memory processes. Due to their low power consumption and high parallel processing capabilities, silicon-based neural circuits have significant potential applications in artificial intelligence and machine learning, facilitating the development of efficient neuromorphic computing systems.

3.4. Peak threshold and adaptivity of spiking behavior

Delves into the peak threshold and adaptive mechanisms of spiking behavior in silicon-based neural circuits. It investigates the membrane potential dynamics that precipitate the generation of spikes upon a neuron's receipt of sufficient input to achieve the peak threshold. The section also examines how conductance regulation and synaptic weight adjustments contribute to the network's learning and memory capabilities. Furthermore, the spiking behavior is influenced by the design of ion

channels and circuit parameters, with adaptive feedback mechanisms enabling neurons to adjust based on their firing history. These features collectively enhance the energy efficiency and flexibility of silicon neural networks in processing complex signals.

Membrane Potential Dynamics: Empirical evidence suggests that spiking behavior is not exclusively governed by a static threshold but is also influenced by the historical trajectory of the membrane potential [12]. Silicon-based neuron models simulate the membrane potential changes of biological neurons. When a neuron receives sufficient input signals, the membrane potential reaches a threshold, triggering a spike. The peak threshold is the critical point at which this spike is initiated, determining when the neuron fires an action potential.

Conductance Regulation: The adaptivity of silicon-based neurons is often achieved through the regulation of conductance. The strength and frequency of input signals can affect the neuron's conductance, allowing the peak threshold to adjust dynamically based on input conditions. This mechanism is analogous to the adaptation phenomenon observed in biological neurons, ensuring that the neural network remains sensitive to continuous changes in input.

Synaptic Weight Adjustment: In spiking neural networks, changes in synaptic weights impact the total input received by the neuron, thereby influencing the accumulation of membrane potential and the triggering of the threshold. By adaptively adjusting synaptic weights, the network can learn and memorize input patterns, enhancing its responsiveness to specific signals.

Ion Channel and Circuit Design: The spiking behavior of silicon-based neurons is also constrained by the design of ion channels and circuit parameters. Factors such as the gating mechanism of ion channels, conduction time, and recovery time can all affect the peak threshold and adaptivity.

Feedback Mechanisms: Silicon-based neural networks may include adaptive feedback mechanisms to adjust the threshold based on the neuron's firing history. This mechanism is similar to the "afterhyperpolarization" observed in biological neurons, preventing neurons from firing too frequently within a short period.

3.5. Significance and application of peak behavior to neuronal circuits

Based on the spiking behavior of silicon neurons, owing to their spike threshold properties and feedback mechanisms, further advancements have been achieved in the fields of neuromorphic computing, multithreaded signal processing, deep learning, neural plasticity and learning capability, real-time signal processing, as well as bio-simulated neural interfaces. The spiking behavior of silicon-based neurons forms the basis of neuromorphic computing.

By mimicking the brain's pulse-based information transmission and parallel processing, neuromorphic computing offers more efficient processing than traditional von Neumann architectures, especially in tasks like pattern recognition, image and speech processing. This spiking behavior supports the development of brain-like neural networks, with the potential for more efficient AI systems.

The spiking behavior of silicon-based neurons is particularly apt for neuromorphic computing for several reasons: These neurons are event-driven, only processing information when a spike occurs, which significantly reduces energy consumption compared to traditional computing systems.

The temporal encoding of spikes facilitates dynamic information encoding, augmenting tasks such as pattern recognition and image processing by closely paralleling the functionalities of biological neurons. The independent spiking of neurons allows for large-scale parallelism, crucial for handling complex, real-time tasks in neuromorphic systems.

Neuromorphic systems leverage mechanisms like Spike-Timing Dependent Plasticity (STDP) to learn and adapt, thereby exhibiting improvement over time. Silicon neurons communicate sparsely

and asynchronously, conserving energy and enabling scalability even in complex, large networks. Intel's Loihi chip implements spike-timing-dependent plasticity (STDP) with silicon neurons, enabling adaptive learning similar to biological brains. This model aims to capture the dynamic and lifelong learning abilities seen in neuromorphic systems, significantly contributing to brain-like computing and improving energy efficiency in AI applications [13].

Research on BrainScaleS chips also emphasizes the role of silicon neurons in advancing neuromorphic computing by simulating spiking behavior at an accelerated rate, enabling more complex brain-inspired models. These neurons help researchers simulate up to 200,000 neurons and millions of synapses, forming the basis for large-scale neuromorphic systems [14].

The spiking behavior of silicon-based neurons enables the efficient handling of diverse signals through mechanisms of energy efficiency, temporal and asynchronous encoding, sparse coding, and robustness. Additionally, their dynamic nature and resemblance to biological neural systems make them well-suited for complex signal environments, offering inherent advantages in diverse signal processing tasks. Spiking neurons use time to encode information, where the information is not only contained in the amplitude but also in the timing and sequence of spikes. This enables parallel processing of diverse signals with varying intervals. Moreover, asynchronous signal processing allows neurons to handle inputs without global synchronization, ideal for handling complex and rapidly changing signals. Spiking neurons operate with sparse coding, meaning fewer spikes represent a large amount of information.

This sparsity enables the processing of high-dimensional, complex data while reducing redundancy, which is critical for extracting key information from diverse signals. "Spiking neurons are highly suited for temporal encoding because they transmit information based on the timing of spikes rather than continuous signal amplitudes.

Additionally, they can operate asynchronously, which allows them to handle dynamic, real-time systems efficiently" [15]. "Spiking neural networks utilize sparse coding by using fewer spikes to represent information, thereby achieving energy efficiency and suitability for the processing of high-dimensional data with minimal redundancy" [16].

The spiking behavior of silicon-based neurons significantly influences the development of bioinspired neural interfaces and neuroplasticity, primarily due to its adaptive properties that enable effective regulation of signal feedback timing. An increasing number of machine learning, deep learning, and bioinspired modules are recognizing the importance of spiking behavior and integrating it into innovative connections. The utility and recognition of trajectory neurons are anticipated to be further augmented within these fields.

4. Implementation of silicon-based neurons

4.1. Very large scale integration

The implementation of silicon-based neurons relies on the mathematical modeling of neurons, with the most common models being the Hodgkin-Huxley model and the integrate-and-fire model. This process is combined with insights from bionics regarding synaptic activity and further integrated with Very Large Scale Integration (VLSI) hardware design to simulate the plasticity of biological synapses.

VLSI technology enables the integration of millions of neurons and synapses onto a single chip, providing a highly scalable hardware architecture that can expand the size of neural networks as computational demands grow. For less complex tasks, a limited number of neurons may be adequate, whereas the execution of more intricate tasks, such as pattern recognition or deep learning, necessitates the coordinated activation of a multitude of neurons.

VLSI facilitates flexible scaling of neuron numbers to meet varying application requirements. Nevertheless, the escalation of neuron counts gives rise to constraints pertaining to chip real estate and power consumption. To address this, silicon-based neurons can leverage inter-chip connectivity to build larger neural networks. This interconnectivity allows multiple VLSI chips to function in tandem, enabling more sophisticated tasks, such as large-scale parallel processing for image recognition and complex data analysis.

The design of silicon-based neurons involves several key steps, utilizing both hardware and software tools, with VLSI technology as the foundation. Initially, neuronal behavior is modeled based on biological neuron models, describing how neurons process inputs and generate action potentials. After modeling, electronic components are used to implement these behaviors, often through hardware description languages like Verilog or VHDL. Key elements such as transistors, operational amplifiers, and memristors are configured to simulate neuron firing and synaptic plasticity.

Once designed, VLSI technology integrates these circuits onto silicon chips. After fabrication, functional validation and debugging ensure the neurons perform as intended, using tools like probe stations and logic analyzers.

At the network level, the coordination, learning, processing velocity, and power efficiency of multiple neurons are evaluated. Using CMOS processes, millions of transistors are integrated into a single chip. Post-validation, the neuron chips are deployed into larger neural networks, where neural algorithms are implemented for applications like image recognition or AI. Ongoing training further refines network performance.

4.2. Driving principle and model

Conductance dynamics and thermodynamically equivalent models explain two critical approaches to simulating neural activity in silicon-based neurons.

Conductance-Based Models simulate neuron behavior by representing changes in membrane conductance using first-order differential equations. In hardware, this is actualized through transconductance amplifiers and log-domain circuits, which mimic the conduct of ion channels as delineated by the Hodgkin-Huxley model. These circuits allow the neuron to simulate the passive conductance of real neurons. Thermodynamically Equivalent Models focus on the state transitions of voltage-dependent gating particles, which regulate ion channels. Transistor circuits replicate the dynamics of these gating particles by mimicking the movement of charged particles in a neuron. The source and drain currents in transistors represent the rates of channel opening and closing, providing a biophysically faithful simulation of neuron conductance.

5. Conclusion

This paper provides an in-depth analysis of neuromorphic silicon neurons (SiNs) and their use of competitive circuits to replicate biological neural functions. By leveraging analog Very-Large-Scale Integration (VLSI) technology, these SiNs accurately model neural dynamics while maintaining energy efficiency through subthreshold operation. Such advancements have significant potential for real-time computational applications.

However, the scalability of these systems to more extensive networks while sustaining both efficacy and energy conservation remains an impediment. Issues like limited chip space and high power consumption must be overcome to fully realize the potential of SiNs. Future work should focus on improving design methodologies and developing more advanced learning algorithms inspired by biological systems.

In looking to the future, the domain of neuromorphic computing is replete with uncharted territories ripe for investigation, with several auspicious research vectors. Domains such as the enhanced integration of spiking neural networks with deep learning methodologies, the advancement of hardware-software co-design paradigms, and the investigation of hybrid models that fuse conventional neural networks with neuromorphic concepts offer substantial opportunities for innovation.

Additionally, the development of more robust feedback mechanisms and the application of neuromorphic systems in dynamic environments could further enrich their utility. By addressing current limitations and focusing on these innovative avenues, researchers can propel the field of neuromorphic computing into new domains, paving the way for advanced AI systems and more effective bioinspired applications.

References

- [1] Mead, C.A. (1990) *Neuromorphic electronic systems. Proceedings of the IEEE*, 78(10), pp.1629-1636.
- [2] Maass, W. (1997) *Networks of spiking neurons: The third generation of neural network models. Neural Networks*, 10(9), pp.1659-1671.
- [3] Indiveri, G., Linares-Barranco, B., Hamilton, T.J., Van Schaik, A., Etienne-Cummings, R., Delbruck, T., ... & Liu, S.C. (2011) *Neuromorphic silicon neuron circuits. Frontiers in Neuroscience*, 5, p.73.
- [4] Ponulak, F. & Kasinski, A. (2011) *Introduction to spiking neural networks: Information processing, learning and applications. Acta Neurobiologiae Experimentalis*, 71(4), pp.409-433.
- [5] Chicca, E. & Indiveri, G. (2007) *A recipe for creating a faulty transistor in silicon spiking neurons. IEEE Transactions on Neural Networks*, 18(1), pp.241-244.
- [6] Yigit, M., Indiveri, G. & Chicca, E. (2016) *Modeling spike-based silicon neurons with subthreshold CMOS circuits. IEEE Transactions on Circuits and Systems I: Regular Papers*, 63(12), pp.2342-2351.
- [7] Indiveri, G., & Horiuchi, T.K. (2011). "Frontiers in neuromorphic engineering." *Frontiers in Neuroscience*, 5, pp. 118.
- [8] Qiao, N., Mostafa, H., Corradi, F., Osswald, M., Stefanini, F., Sumislawska, D., & Indiveri, G. (2015). "A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses." *Frontiers in Neuroscience*, 9, pp. 141.
- [9] Liu, S.-C., & Delbruck, T. (2010). "Neuromorphic sensory systems." *Current Opinion in Neurobiology*, 20(3), pp. 288-295.
- [10] Liu, S.-C., Kramer, J., Indiveri, G., Delbrück, T. & Douglas, R. (2002) *Analog VLSI: Circuits and Principles. A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England.*
- [11] Indiveri, G., Chicca, E. & Douglas, R.J. (2006). "A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity." *IEEE Transactions on Neural Networks*, 17(1), pp.211-221.
- [12] Gerstner, W. & Kistler, W.M. (2002) *Spiking Neuron Models: Single Neurons, Populations, Plasticity. Cambridge University Press, Cambridge.*
- [13] Davies, M., Srinivasa, N., Lin, T.H., Chinya, G., Cao, Y., Choday, S.H., & Suri, M. (2018) *Loihi: A neuromorphic manycore processor with on-chip learning. IEEE Micro*, 38(1), pp.82-99.
- [14] Schemmel, J., Grünbl, A., Meier, K. & Millner, S. (2010) *Implementing synaptic plasticity in a VLSI spiking neural network model. In: Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN)*, pp.1-8.
- [15] Ponulak, F., Kasinski, A. (2011) 'Deep Learning With Spiking Neurons: Opportunities and Challenges'. *Frontiers in Neuroscience*, 5, pp.1-16. Available at: <https://www.frontiersin.org/articles/10.3389/fnins.2011.00016/full> (Accessed: [add access date]).
- [16] Kasabov, N. (2010) 'To spike or not to spike: A probabilistic spiking neuron model'. *Neural Networks*, 23(1), pp.16-19. Available at: [ScienceDirect/IEEE Xplore].