



INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

SHORT ABSTRACT OF THESIS

Name of the Student	: Ashvinikumar Pruthviraj Dongre
Roll Number	: 186102005
Programme of Study	: Ph.D.
Thesis Title	: Design of RRAM-Based Integrate and Fire Neuron And Programmable Synapse for Neuromorphic Computing
Name of the Supervisor(s)	: Dr. Gaurav Trivedi
Thesis Submitted to the Department/Center	: EEE Department
Date of completion of the Viva-Voce Exam	: 26/3/24
Key words for description of Thesis Work	: Neuromorphic Computing, RRAM, Integrate and Fire Neuron, Synapse.

SHORT ABSTRACT

A human brain can perform compute-intensive tasks, such as multi-object recognition, reasoning, and decision-making, consuming only 20 W power. Whereas, to recognize 1000 different objects, a CPU consumes around 250 W power. Around 10^{11} neurons in the human brain are interconnected through approximately 10^{15} synapses responsible for the brain's exceptional computing capacity. The advancements in processing technology have reduced the technology nodes drastically, which further reduced the power consumption of the processors; still, they cannot match the low power consumption of the human brain. Even with the latest technological advancements, optimizing the processors with Von Neumann architectures for speed and power becomes challenging because of the memory Bottleneck.

The root cause of the memory bottleneck in a processor is the separation of memory and processing units. Even though the processors can be designed to be superfast, the applications that run on these processors, especially the artificial intelligence applications, need large amounts of data to be transferred from memory to the processing unit. A simple matrix multiplication involves multiple reading, processing, and writing operations. It worsens further with an increase in the size of the matrix.

This motivated researchers to explore other paradigms, such as in-memory and near-memory computing, where computations are performed in the system's memory. Such modifications at the architectural level have improved the performance of the processing units. However, the ever-increasing

demands of AI applications have forced researchers to look deep into the brain's functioning to optimize area, speed, and power. The human brain does not have separate memory and processing units. Therefore, it does not require any read or write cycles. This evolved a new neuromorphic computing era, in which brain's critical algorithmic and computational features are emulated in silicon-based hardware to improve performance at minimal power consumption. Since neurons and synapses form the basic elements of a neuromorphic architecture, it is anticipated that optimizing these elements would result in area and power-efficient large-scale neuromorphic computing architectures.

The scientific community prefers neuromorphic systems realization using digital logic. Since the number of neurons required for practical applications is large, this increases the overall power and area consumption. Implementation of neurons in the analog domain is also an attractive solution. However, the large-scale realization of such architectures becomes inefficient because of its high power consumption even when implemented on the lower technology nodes. Efforts have been made to design analog neurons using CMOS transistors, but the energy consumption remains in the range of pJs. Motivated by these facts, we propose a Resistive Random Access Memory (RRAM) based integrate and fire neuron. RRAM is employed as a voltage divider for integrate and fire (I&F) operation. The proposed neuron exhibits temporal integration, triggering threshold, and refractory period like a biological neuron, making it a suitable candidate for large-scale neuromorphic systems.

Another major component of neuromorphic systems is the synapse. RRAMs have been widely explored to design synapses. However, variability in RRAMs is a major hindrance while implementing large-scale neuromorphic architectures. Multilevel cells have been extensively explored to obtain multi-bit precision in a single cell. Although it provides considerable advantages regarding area utilization and power consumption, implementing circuits for precise programming of the resistive state is a significant challenge. Therefore, this thesis proposes an RRAM-based synaptic architecture with a continuous sensing and feedback scheme to stop RRAM programming when the required conductance is achieved.

The work proposed in this thesis demonstrates that RRAM can be efficiently employed to implement energy-efficient integrate and fire neurons. We further design an RRAM-based reprogrammable synapse. The precise RRAM programming mechanism shows that the Cycle-to-Cycle and Device-to-Device variations that are pertinent to RRAM devices can be resolved effectively using circuit-level techniques. Finally, we implement a spiking neural network to evaluate the performance of the proposed integrate and fire neuron and programmable synapse.