1.1 Thesis Abstract

With the evolution of artificial intelligence (AI) and the internet of things (IoT), the semiconductor industry needs to fulfill the demands of high-speed, large data processing, higher integration density, and power efficiency. The fundamental requirement for efficient data processing is the availability of large and high-speed memory systems, such as complementary metal-oxide-semiconductor (CMOS) based static random-access memory (SRAM). Conventional CMOS technology, enhanced through device scaling, has improved parameters such as dynamic power dissipation, chip area, and delay. In modern high-performance computing, especially in AI applications that can involve moving up to terabytes of data per second, most power consumption arises from data transmission between memory and processing units. This is due to the separation between memory and processing units in traditional Von Neumann architectures. Hence, this high-speed data processing requirement cannot be met by traditional CMOS-based memory solutions and conventional computing architectures. These limitations of CMOS have prompted the exploration of emerging technologies for memory and information processing.

The emerging non-volatile memory (NVM) technologies have enabled us to look toward some of the memory-centric computing paradigms beyond the Von Neumann architecture. One of the new paradigms is neuromorphic computing to mitigate the bottlenecks of the conventional Von Neumann architectures. NVMs are well-suited for neuromorphic computing, a paradigm inspired by the human brain, which aims to mimic the functionality of the human brain by integrating memory and computation into a unified framework. Neuromorphic systems leverage non-volatile devices to replicate synaptic functionalities, enabling highly energy-efficient and parallel processing for AI. Furthermore, emerging NVM

technologies, which offer energy efficient solutions and data retention under zero power supply, have shown strong potential to replace CMOS-based memory technologies. Amongst different NVM solutions, magnetic tunnel junction (MTJ) and ferroelectric field-effect transistor (FeFET) based NVM are focused in this work.

This dissertation focuses on the design and optimization of novel devices to enhance the performance of spin-transfer torque magneto-resistive random-access memories (STT-MRAM) and FeFET-based NVM at the bit-cell level. Firstly, we have developed a novel silicon-on-insulator (SOI) based junctionless-accumulation-mode (JAM) ferroelectric FinFET along with a proposed process flow at a 3-nm node. This device can be utilized as a synaptic element in neuromorphic computing applications. Secondly, we have also developed a 1T-1R-based memory cell featuring distinct read and write paths that utilize a memristive variant of the ferroelectric field effect transistor (MFeFET) for data storage. This memory cell architecture offers a unique feature, *i.e.*, a non-destructive read, better scalability, enhanced energy and area efficiency, and compatibility with one transistor, one memristor memory (1T-1M) technologies like STT-MRAM and resistive random-access memory (RRAM). Furthermore, we developed a 3-D asymmetric FinFET device structure at a 3nm technology node using process emulation in TCAD. Subsequently, we utilized this asymmetric FinFET design to achieve performance improvements in the STT-MRAM cell. All simulations presented in this thesis were performed using Sentaurus TCAD simulation tool. The work in this dissertation is structured into five chapters, as outlined below:

Chapter 1, provides an overview of neuromorphic computing with a brief background of various non-volatile memory technologies. Additionally, this chapter reviews the literature on device modeling and logic circuits co-design of non-volatile memory cells, specifically the bit cell used in STT-MRAM and FeFET. We explore novel device architectures aimed at improving the performance of these memory cells. Furthermore, this chapter explores the fundamental physics, memory cell operation, and structural details of FeFET-based memory devices and MTJ. Finally, a brief summary is provided towards the formulation of the research problem statement for this thesis.

In ***Chapter 2,*** we present an SOI-based JAM ferroelectric FinFET designed to function as a synaptic weight device in neuromorphic computing architectures. The proposed device aims to enhance the non-volatile conductance range in the ON-state compared to existing junctionless and conventional ferroelectric FinFET devices. Our simulations demonstrate that the device offers linear conductance variation and symmetric switching characteristics, which are crucial for synaptic weight applications. Additionally, we have shown the experimental and process feasibility of this novel synaptic weight device through process emulation using Sentaurus TCAD. The proposed device offers zero process overhead and is fully compatible with the pristine FinFET process flow.

In ***Chapter 3,*** we introduce a 1T-1R memory cell that employs a memristive variant of the ferroelectric field effect transistor (MFeFET) for data storage. This 1T-1R memory architecture features distinct write and read paths, enabling non-destructive read operations and higher read currents at lower operating voltages. This novel memory architecture also offers lower read latency than STT-MRAM technologies.

After this chapter, in ***Chapter 4,*** we propose a novel asymmetric FinFET device in terms of drive-current from source to drain and drain to source. The proposed device is targeted towards increasing the efficiency of the STT-MRAM cell, which consists of an MTJ as the data storage element. The MTJ has an asymmetric bidirectional write current, which can be exploited to achieve power savings by using an asymmetric select/access device in the STT-MRAM bit-cell. Further, the simulation and process emulation carried out in this chapter show that the proposed device provides better asymmetry and easy compatibility with the existing FinFET fabrication flow, respectively. The proposed device provides an additional benefit of lower gate capacitance over existing symmetric and asymmetric FinFET device architectures. Low gate capacitance provided by the proposed asymmetric device highlights the other vital improvement, which could result in lower write-line capacitance in STT-MRAM bit-cells.

Chapter 5 concludes the thesis by highlighting the key contributions and summarizing the research results achieved. Furthermore, it outlines the potential directions for future work based on the findings presented in the thesis.

1.2 Introduction

Today's semiconductor industry demands faster and more extensive data processing capabilities, higher integration density, and improved power efficiency. The rapid growth of artificial intelligence (AI) and the internet of things (IoT) has intensified the need for high-performance and energy-efficient computing systems. However, traditional CMOS-based computing architectures are encountering numerous obstacles, including the imminent end of Moore's Law and performance limitations imposed by the Von Neumann bottleneck. As fabrication costs continue to rise and physical scaling reaches its limits, further miniaturization of devices becomes increasingly difficult. Additionally, the energy inefficiencies and latency caused by data transfer between memory and processors severely hinder performance improvements, even if scaling continues. These challenges have compelled the semiconductor industry to explore innovative solutions built on novel devices and emerging computing