Reimagining Cryogenic Computing with Superconducting and Topological Devices

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Shamiul Alam
August 2025

Copyright © 2025 by Shamiul Alam
All rights reserved.

ACKNOWLEDGEMENTS

I would like to extend my sincerest gratitude to my Ph.D. advisor, Prof. Ahmedullah Aziz, whose exceptional guidance and unwavering support have been instrumental throughout the journey of completing this dissertation. Prof. Aziz's expertise, encouragement, and insightful feedback have profoundly shaped my research endeavors and academic growth throughout my graduate studies. I am deeply grateful for the invaluable mentorship provided by him, which has not only enriched my scholarly pursuits but also fostered my personal and professional development. I consider myself extremely fortunate to have had Prof. Aziz as my Ph.D. advisor.

I am also indebted to the members of my doctoral dissertation committee, Prof. Benjamin Blalock, Prof. Garrett Rose, and Prof. George Siopsis. Their valuable contributions, constructive criticism, and scholarly insights have significantly enhanced the quality of this dissertation.

Furthermore, I extend my heartfelt appreciation to the esteemed scholars and experts in the field who generously shared their knowledge, expertise, and valuable perspectives, including Prof. Vijaykrishnan Narayanan (Pennsylvania State University), Prof. Sumeet Kumar Gupta (Purdue University), Prof. Nathaniel Cady (SUNY Polytechnic Institute), Prof. Kai Ni (University of Notre Dame), Prof. Nikhil Shukla (University of Virginia), Prof. Akhilesh Jaiswal (University of Wisconsin Madison), Prof. Sumitha George (North Dakota State University), Dr. Md Shafayat Hossain (Princeton University), Dr. Adam McCaughan (National Institute of Standards and Technology), and Dr. Kazi Asifuzzaman (Oak Ridge National Laboratory).

I would also like to thank my colleagues at the Nanoelectronic Devices and Integrated Circuits (NorDIC) Lab, University of Tennessee Knoxville, for their relentless support and encouragement.

I wish to acknowledge the financial support received from the Air Force Research Lab, the Department of Energy, the National Science Foundation, and The University of Tennessee-Oak Ridge Innovation Institute, which has been instrumental in facilitating this research.

Last but certainly not least, I express my profound gratitude to my beloved family, whose unwavering love, encouragement, and sacrifices have been my constant source of strength and motivation. I extend special thanks to my parents, Mr. Abdus Samad and Mrs. Laizu Begum, for their unwavering support in every stage of my life. I am deeply grateful for their selfless love and lifelong sacrifices. I would also like to thank my loving wife, Mrs. Zeba Faria Reshme, whose unwavering support and sacrifices have been indispensable to my academic and personal success. Throughout the journey of pursuing my Ph.D., she has been a pillar of strength, offering encouragement and understanding in times of difficulty and uncertainty. Her unwavering support has gone beyond mere words, as she has actively shared the burden of the challenges and obstacles we encountered along the way. Whether it was late-night study sessions, moments of self-doubt, or the stresses of balancing work and life, she stood by my side. She has been my partner on this journey, sharing both the triumphs and the challenges. I am also grateful to my elder brother, Mr. Atikur Rahman, and my sister, Mrs. Tanjina Ferdous, for their unconditional love and support.

To my friends and colleagues, thank you for your encouragement, camaraderie, and unwavering belief in my abilities.

This dissertation is a testament to the collective efforts and support of all those who have played a part, whether big or small, in this academic endeavor. Your contributions have been invaluable, and I am truly grateful for your role in shaping this dissertation and enriching my academic journey.

ABSTRACT

Cryogenic computing - operating at or below 4 K - has emerged as a promising solution to address the scalability issues of quantum computers, which currently rely on room temperature controller and memory units. In addition, cryogenic computing offers unique advantages for high-performance computing and deep space exploration. Various technologies, including non-superconducting, superconducting, and hybrid, have been explored. However, each of these technologies has significant limitations. Non-superconducting devices suffer from low speed and high power consumption, while superconducting systems based on Josephson junctions (JJs) face severe scalability issues due to flux trapping and inductive coupling. Hybrid systems, meanwhile, lack efficient interface circuitry to integrate superconducting and non-superconducting technologies.

This dissertation utilizes two unique technologies: superconducting devices (ferroelectric superconducting quantum interference devices (FE-SQUIDs) and heater cryotrons) and topological devices exploiting the quantum anomalous Hall effect, to develop a suitable cryogenic computing platform. First, we demonstrate a non-volatile FE-SQUID-based memory with voltage-controlled operation, high scalability, and separate read-write paths. Using the same device, we develop a voltage-controlled superconducting logic family that addresses the limited fanout and cascadability issues of existing current-controlled logic circuits. We further present a cryogenic in-memory computing architecture based on FE-SQUIDs to reduce data movement and cooling costs associated with *von Neumann* systems. Additionally, we design a ternary content addressable memory (TCAM) using FE-SQUIDs which supports exact search without any peripheral circuitry, as well as nearest-match operations through Hamming distance calculation.

We also explore the reconfigurable potential of heater cryotron devices. A single cryotron circuit is shown to execute multiple 1-, 2-, and 3-input logic operations without any structural changes, enabling logic camouflaging for enhanced hardware security.

Finally, this dissertation is the first to utilize topological devices in circuit and system-level applications. Leveraging the quantum anomalous Hall effect, we develop a scalable, non-volatile, and ultra-low-power memory system. This topological memory is also

capable of performing both basic and complex in-memory computing tasks, further advancing the development of a practical and efficient cryogenic computing platform.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1
1.1 Cryogenic Computing 1
1.2 Overview of Existing Technologies 5
1.2.1 Non-Superconducting Technologies 5
1.2.2 Superconducting Technologies 6
1.2.3 Hybrid Technologies 7
1.3 Contributions of this Dissertation 8
1.4 Organization of this Dissertation 10
CHAPTER 2: CRYOGENIC MEMORY ARRAY BASED ON FERROELECTRIC SUPERCONDUCTING QUANTUM INTERFERENCE DEVICE 11
2.1 Introduction 11
2.2 Ferroelectric SQUID 12
2.3 Heater Cryotron-based Selector 14
2.4 Ferroelectric SQUID-based Memory Array 15
2.5 Modeling and Simulation Framework 17
2.6 Simulated Memory Operations 18
2.7 Comparative Analysis 19
2.8 Summary 19
CHAPTER 3: VOLTAGE-CONTROLLED LOGIC FAMILY WITH FERROELECTRIC SUPERCONDUCTING QUANTUM INTERFERENCE DEVICE 21
3.1 Introduction 21
3.2 Design Methodology of Logic Circuits 22
3.3 Working Principle and Functional Verification 23
3.4 Summary 27
CHAPTER 4: IN-MEMORY MATRIX-VECTOR MULTIPLICATION USING FERROELECTRIC SUPERCONDUCTING QUANTUM INTERFERENCE DEVICE 28
4.1 Introduction 28
4.2 In-Memory MVM With FE-SQUID 29
4.3 Accelerating Classification Tasks With FE-SQUID-based MVM 31

4.4 Comparative Analysis 35
4.5 Summary 35
CHAPTER 5: TERNARY CONTENT ADDRESSABLE MEMORY USING FERROELECTRIC SUPERCONDUCTING QUANTUM INTERFERENCE DEVICE 37
5.1 Introduction 37
5.2 Design Principle of FE-SQUID-based TCAM 40
5.3 Working Principle of FE-SQUID-based TCAM 42
5.4 System-level Benchmarking 45
5.5 Summary 48
CHAPTER 6: RECONFIGURABLE LOGIC CIRCUIT USING SUPERCONDUCTING HEATER CRYOTRON 49
6.1 Introduction 49
6.2 Device Fabrication and Characterization 50
6.3 Design Methodology of Reconfigurable Logic 53
6.4 Working Principle and Functional Verification 54
6.5 Summary 58
CHAPTER 7: NON-VOLATILE MEMORY ARRAY BASED ON QUANTUM ANOMALOUS HALL EFFECT 61
7.1 Introduction 61
7.2 Quantum Anomalous Hall Effect (QAHE) 61
7.3 QAHE-based Memory 63
7.4 Design of Cross-point Memory Array 65
7.5 Modeling and Simulation Framework 68
7.6 Write and Read Operations 69
7.7 Summary 71
CHAPTER 8: QUANTUM ANOMALOUS HALL EFFECT-BASED IN-MEMORY BOOLEAN LOGIC OPERATIONS 73
8.1 Introduction 73
8.2 QAHE-based Logic-in-memory System 76
8.3 Functional Verification 78

8.4 Variation Analysis 80
8.5 Summary 81
CHAPTER 9: IN-MEMORY BIT-SERIAL ADDITION USING QUANTUM ANOMALOUS HALL EFFECT-BASED MAJORITY LOGIC 83
9.1 Introduction 83
9.2 QAHE-based In-Memory Majority 85
9.3 QAHE-based In-memory Adder 86
9.4 Comparative Analysis 90
9.5 Summary 91
CHAPTER 10: CONCLUSION AND FUTURE WORKS 92
10.1 Summary 92
10.2 Future Works 93
10.2.1 hTron-based Current-Splitter Circuitry 93
10.2.2 Superconducting Neuromorphic Computing with FE-SQUID and hTron 94
10.2.3 Probabilistic Computing with hTron 95
LIST OF REFERENCES 96
VITA 122

LIST OF TABLES

Table 2.1. Comparison with major cryogenic memory technologies 20
Table 3.1. Design methodology of FE-SQUID-based logic circuits 23
Table 4.1. Comparison with existing approaches 36
Table 7.1. Comparison with the state-of-the-art cryogenic memories..... 72
Table 9.1. Implementation of 3-input and 5-input in-memory majority logic..... 86

LIST OF FIGURES

Figure 1.1. (a) Existing architecture of current quantum computers where control processor and memory are placed at room temperature. The use of (b) 77 K and (c) 4 K control processor and memory blocks to facilitate the scaling of quantum computers to thousands of qubits to achieve the dream capability. 2

Figure 1.2. (a) Exponential growth of energy demand in high-performance computing every year. (b) Comparison of different technologies in terms of achieving the DOE's exa-scale goal. (c) Comparison of superconducting and CMOS technologies in terms of power consumption and switching delay. 4

Figure 1.3. Overview of existing cryogenic computing technologies. 6

Figure 1.4. (a) Illustration of a hybrid memory system, where high scalability of non-supercomputing technology is combined with high speed and low power of superconducting technology. (b) Example of a hybrid memory system for cryogenic applications proposed in Ref. 121. Here, 3-transistor DRAM is used as the memory element and JJ-based circuits are used as the access scheme. 8

Figure 2.1. (a) Device structure and circuit symbol of FE-SQUID. (b) Polarization-voltage characteristics for the PZT ferroelectric along with the validation of the developed compact model with the experimental results. (c) Current-voltage characteristics of FE-SQUID. Two polarization states of the ferroelectric leads to two levels of IC. (d) Effects of ferroelectric polarization on different device characteristics of FE-SQUID. 13

Figure 2.2. (a)-(b) Gate current -driven superconducting → resistive switching and (c) I-V characteristics of a heater cryotron. 15

Figure 2.3. (a) Schematic of FE-SQUID-based memory array. (b) Biasing scheme to access the (2, 2) cell in the array for write and read operations. 16

Figure 2.4. Modeling approach of FE-SQUID. 18

Figure 2.5. Temporal dynamics of the ferroelectric voltage of accessed (ACC), half-accessed (HA), and unaccessed (UA) cells during (a) write '0' and (b) write '1' operations. Switching of PFE and IC of the cells during (c) write'0' and (d)

write‘1’ operations. Time dynamics of (e) RBL and RWL biasing, and (f) SQUID voltage of the accessed cell. 19
Figure 3.1. (a) Schematic of a Copy gate using a FeSQUID. The output of this configuration is not sufficiently large to drive the input of the next gate and hence, this configuration is not cascadable. (b) Modified version of the Copy gate of (a) where a hTron is used to convert the output voltage to the same level of the input so that two or more logic gates can be cascaded. 24
Figure 3.2. (a) Schematic of FeSQUID-based Copy gate. Working principle of the Copy gate for (b) logic ‘0’ and (c) logic ‘1’ at the input. (d) Schematic of FeSQUID-based NOT gate. Working principle of the NOT gate for (e) logic ‘0’ and (f) logic ‘1’ at the input. Green and red colors correspond to the superconducting and resistive states, respectively. (g) Simulated results for the 1-input Copy and NOT gates. 24
Figure 3.3. Working principle of FeSQUID-based AND gate for inputs (a) ‘00’, (b) ‘01’/‘10’, and (c) ‘11’. Operation of FeSQUID-based OR gate for (d) ‘00’, (e) ‘01’/‘10’, and (f) ‘11’ at the input terminals. (g) Schematic of XOR gate built cascading FeSQUID-based NOT, AND, and OR gates. (h) Simulated results for the 2-input AND, OR, and XOR gates. 26
Figure 4.1. (a) FE-SQUID memory array with modified peripheral circuitry to perform MVM. Schematics of (b) sense amplifier and (c) voltage comparator. (d, e) Two examples of MVM operation 31
Figure 4.2. (a) Mechanism of training weights and (b) storing them as memory states in the array. (c) FE-SQUID-based ML model to perform handwritten digit recognition. (d) Mechanism of encoding input as the read current levels. (e) Confusion matrix for the predicted digits. 32
Figure 4.3. (a) Effects of number of rows and columns on the accuracy of handwritten digit recognition task. (b) Comparison of accuracy for the three classification tasks tested using our framework. 34
Figure 5.1. Introduction to content addressable memory. Illustrations of the search mechanism with a (a) binary and (b) ternary CAM. (c) Block diagram showing

the steps how a CAM can be used for AI applications such as classification and pattern recognition tasks. 38
Figure 5.2. The proposed cryogenic TCAM based on FE-SQUID. (a) Circuit schematic of the proposed TCAM cell. (b) Illustration of the array-level organization of the proposed TCAM cells, where data will be stored in a row. (c) Values of RWL currents for two modes of the proposed TCAM. Definitions of the representation of (d) input search data in terms of RBL currents and (e) stored data in terms of FE-SQUID's device states. 41
Figure 5.3. Working Principle of the proposed TCAM in the exact search mode. Illustration of the working principle of the TCAM when (a) $Data$ = 0 and $Search$ = 0, (b) $Data$ = 0 and $Search$ = 1, (c) $Data$ = 0 and $Search$ = d, (d) $Data$ = 1 and $Search$ = 0, (e) $Data$ = 1 and $Search$ = 1, and (f) $Data$ = 1 and $Search$ = d. (g) ML voltage, power, and energy consumption for 1-bit exact search with the proposed TCAM. 43
Figure 5.4. ML voltage levels obtained for different number of data lengths and different amounts of mismatch between the input search and stored data. 45
Figure 5.5. (a) Illustration of the HDC algorithm for language recognition. The encoded search vector of the sample text that needs to be recognized is compared to all pre-trained class vectors. The closest vector is then inferred as the result, and the respective class label is used to answer the search. (b) The inference accuracy of language recognition over the vector size. (c) Energy consumption of a single vector comparison with 10,000 bits using FE-SQUID-based and SRAM-based TCAM cells. Using the discussed energy estimation scheme, the FE-SQUID energy is constant with respect to block size. 46
Figure 6.1. (a) False-colored scanning electron micrograph (SEM) of a fabricated hTron device. The fabricated device has a heater gate with an area of 200×200 nm^2 and a channel with a width of 4 μm. (b), (c) Illustration of the switching mechanism of a hTron device. (d) Measured distributions of the channel critical current (I_{Ch}^{C}) for different values of I_G. (e) Extracted median values of I_{Ch}^{C}

distributions for each I_G. (f),(g) Measured time dynamics of the switching of a hTron device with a load resistance (R_L) of 1 $k\Omega$. 51
Figure 6.2. (a) Schematic of the hTron-based reconfigurable logic circuit that can perform 1-input copy and NOT, 2-input AND and OR, and 3-input majority operations with only a suitable bias current (I_{B1}). We use 1 $k\Omega$ as the load resistor (R). (b) Definition of logic states. (c) Biasing conditions to perform different logic operations with the same circuit. 54
Figure 6.3. Copy operation with the proposed circuit when the input is (a) logic '0' and (b) logic '1'. (c)-(e) Simulated time dynamics of the copy operation..... 55
Figure 6.4. NOT operation with the proposed circuit when the input is (a) logic '0' and (b) logic '1'. (c)-(e) Simulated time dynamics of the copy operation..... 56
Figure 6.5. AND operation with the proposed reconfigurable circuit for (a) '00', '01', and '10' and (b) '11' input combinations. (c)-(e) Simulated time dynamics for the AND operation. 57
Figure 6.6. Illustration of the OR operation for (a) '00' and (b) '01', '10' and '11' input combinations, respectively. (c)-(e) Time dynamics of the OR operation where the channel switches for the '11' combination due to the choice of I_{B1}..... 58
Figure 6.7. Illustration of majority logic when there are (a) greater and (b) smaller number of logic '1' inputs. (c)-(e) Simulated time dynamics of the Majority operation. 59
Figure 7.1. (a) Schematic of a QAH insulator, where the red circles represent the electron and the red arrows show the spin of the electrons. (b) Illustration of the Hall resistance *Rxy* vs. the bias current, I_{Bias} at zero external magnetic field. I_{C-} and I_{C+} are two critical values of I_{Bias} which determine the hysteretic switching of R_{xy} between $-h/e^2$ and h/e^2. (c) Schematic of a tBLG moiré heterostructure where tBLG is encapsulated between hBN and a flake of few-layer graphite is used as gate. (d) R_{xy} data, plotted as a function of I_{Bias} at T = 4 K and B = 0 T, exhibit current-driven hysteretic switching. 62
Figure 7.2. (a) R_{xy} *vs.* I_{Bias} at zero external magnetic field. (b) Hall voltage, V_{xy} plotted as a function of I_{Bias}. (c), (d), (e), & (f) The procedures of write '0', write '1', read

‘0’, and read ‘1’ operations, respectively, along with the state of electron spins (red arrows). (g) Summary of the key idea for the QAHE-based memory. 64
Figure 7.3. (a) Device structure and (b) characteristics of an MIEC material-based selector. (c) Illustration of the overall memory array structure, custom-designed for the QAHE-based memory devices. 66
Figure 7.4. (a) Behavioral representation of the modeling approach for the QAHE observed in the tBLG moiré heterostructure. (b) Temperature dependence of R_{xy} with an experimental matching. 68
Figure 7.5. (a) Arrangement and types of memory cells in a conventional cross-point memory array under the *V/2* biasing scheme. I_{Bias} through different memory cells generated by applying appropriate V_{ACC} during (b) write ‘0’ and (c) write ‘1’ operations. Time dynamics of R_{xy} illustrating state transition only in the acessed cell during (d) write ‘0’ and (e) write ‘1’ operations. (f) I_{Bias} through different cells during read operation. Time dynamics of R_{xy} and V_{xy} during (g) read ‘0’ and (i) read ‘1’ operations. Time dynamics of the amplifier and comparator outputs during (h) read ‘0’ and (j) read ‘1’ operations. (k) Summary of I_{Bias} through different cells during write and read operations. 70
Figure 8.1. (a) Significant improvement in computing speed and energy requirement with the lowering of operating temperature. Computing speed and energy requirement have been normalized by the value at room temperature (300 K). Illustration of (b) traditional *von Neumann* architecture and (c) compute-in-memory (CiM) architectures. 74
Figure 8.2. (a) QAHE-based cryogenic CiM structure. (Wires of different colors are not connected to each other unless shown using black dots at the intersections between two wires.) (b) Schematics of the reference generator and the sense amplifier. (c) Truth table for the 2-input logic operations and the corresponding voltage levels for different logic combinations. (d) Conditions for choosing the reference voltages for read and in-memory logic operations such as NAND, NOR, and XOR. 77

Figure 8.3. Biasing scheme for the WLs and BLs during (a) memory read operation and (b) logic operations. Voltage applied to BL1 and the voltage generated at the output of the amplifiers during (c) read '0' and (d) read '1' operations. Time dynamics of the sense enable signal (SEN) and output of the sense amplifier during (e) read '0' and (f) read '1' operations. (g) Voltage at the outputs of the amplifiers for different logic combinations ('00', '01', '10', and '11') in the rows. Time dynamics of the sense enable signal (SEN) and output of the sense amplifier during (h) 2-input NAND, (i) 2-input NOR, and (j) 2-input XOR operations. 79
Figure 8.4. (a) Input distribution of the cell read current with a 10% variation of the nominal value (−2.02 nA) to run a 10000-point Monte-Carlo simulation. Distributions of (b) the cell Hall voltage and (c) the amplified row voltages (sum of active cell voltages) resulting from the variation considered in the read current shown in (a)..... 81
Figure 9.1. (a) Truth tables for 3-input and 5-input Majority logic. Implementation of (b) AND and OR gates, and (c) full adder using Majority logic. (d) Logic level and (e) in-memory implementation of full adder using NAND, NOR, and Majority. The use of Majority logic reduces the logical depth and number of cycles. 84
Figure 9.2. Calculation of (a) C_{out} and $\overline{C_{out}}$ and (b) Sum using 3-input and 5-input Majority logic, respectively..... 87
Figure 9.3. Simulation results for Majority-based in-memory 1-bit full adder. 87
Figure 9.4. Steps of in-memory bit-serial addition in a QAHE-based cryogenic memory system. 89
Figure 9.5. Comparison of our proposed QAHE-based in-memory majority logic with other approaches, highlighting the advantage of using QAHE. 91

CHAPTER 1
INTRODUCTION

1.1 Cryogenic Computing

Silicon-based complementary metal-oxide-semiconductor (CMOS) technology now allows us to accommodate more than tens of billion transistors on a single chip (for example, Cerebras Wafer Scale Engine 2 with 2.6 Trillion transistors [1]). However, this aggressive scaling comes with a price- an exorbitant power dissipation reaching the physical limit. For example, in 2018, about 205 TWh of electricity was consumed by the US data centers [2]. To circumvent this issue, numerous 'beyond-CMOS' technologies, such as superconducting single flux quantum (SFQ) electronics, are being explored. SFQ circuits and systems based on superconductive devices [e.g., Josephson junctions (JJ) and superconducting quantum interference device (SQUID)] are generally faster and more energy-efficient than the CMOS counterparts [3], thanks to the dissipation-less current flow in superconductors. Ultrafast and energy-efficient superconducting SFQ circuits, coupled with lossless and low dispersion interconnects, have significantly improved performance in different applications such as digital radio frequency receivers [4], [5], high-end computing [6], and so on. However, the large-scale development of superconducting circuits and systems needs a suitable cryogenic computing platform that can preferably operate at or below 4 K.

Going beyond the conventional computing paradigm, the importance of a suitable cryogenic computing platform is felt in quantum computing as well. Quantum computers promise to solve commercial and scientifically important problems that the current classical computers can hardly handle in a realistic timeframe. For instance, accurate simulation of large molecules for the development of new drugs and eco-friendly manufacturing processes now requires unrealistic time due to the exponential scaling of the complexity in classical computers. However, these simulations can be performed with a quantum computer in polynomial time [7]. Moreover, quantum computers can significantly accelerate number theory, algebraic, and optimization problems [8].

A typical quantum computer has three major components: quantum substrate (qubits), a control processor, and a memory block [9]. **Fig. 1.1**(a) illustrates the organization of a

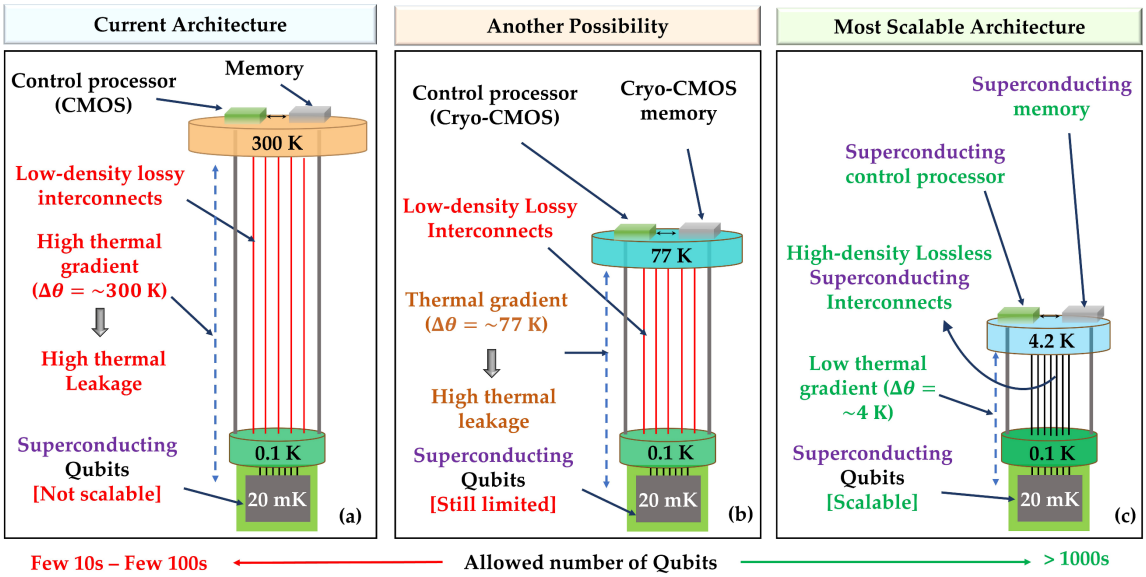


Figure 1.1. (a) Existing architecture of current quantum computers where control processor and memory are placed at room temperature. The use of **(b)** 77 K and **(c)** 4 K control processor and memory blocks to facilitate the scaling of quantum computers to thousands of qubits to achieve the dream capability.

scalable quantum computer proposed in Ref. [10], [11]. In quantum computing, qubit is the basic unit of quantum information realized with a two-state device. To protect the quantum states of the noise-sensitive qubits, they are placed at a few milli-Kelvin temperature[12]–[17]. In the current lab size quantum computers, control processor and memory blocks are placed at room temperature. Long control cables are required to connect the qubits with the control processor and the memory. Although this architecture works fine for a small number of qubits, it cannot be scaled to a few hundred qubits because it would require a large number of wires to establish a connection between the room-temperature components and the qubits at cryogenic temperature, particularly because each qubit might need an individual control [9]. For example, a recent experiment reported the requirement of 200 wideband coaxial cables, 45 microwave circulators, and a rack of electronic circuits to control only 53 qubits. Moreover, the wires and interconnects connecting the room temperature blocks with the qubits would have a large temperature gradient (around 300K) which generates thermal leakage to the noise-sensitive qubits.

For a large-scale quantum computer with thousands of qubits, the resulting effect of this thermal leakage can be destructive to the qubit states. Note, the quantum computer needs to be scaled up to thousands of qubits to utilize its full potential. For example, to run Shor's algorithm for prime factorization[8], which is a textbook quantum algorithm, on a 1024-bit number in a quantum computer, two quantum registers (one with 2048 qubits and another with 1024 qubits) are required. To circumvent this scaling issue, the control processor and the memory blocks should be placed at a temperature very close to that of the qubits [18]. There are two possible options to place the control processor and the memory blocks- 77 K (**Fig. 1.1**(b)) and 4 K (**Fig. 1.1**(c)). Placing these blocks at 4 K will have lower thermal leakage and require a smaller number of wires and interconnects, and hence, better scalability compared to the 77 K option. Placing them at 4K will also allow the use of dissipation-less superconducting interconnects. Although there is a small thermal leakage, placing them at 77 K will need $\simeq$ 200 times lower cooling cost than the 4 K option [13].

Cryogenic computing is also perfectly suited for high-performance computing (HPC) [19]. It has been known for many years that the energy demand of data centers, and supercomputing facilities around the world will reach an alarming level by the year 2030 [20]. As shown in **Fig. 1.2**(a), the energy demand is forecast to grow exponentially every year. On top of it, the energy demand for AI systems is increasing exponentially. A recent report suggested that AI could soon need as much electricity as an entire country [21]. These alarming statistics make it more compelling to look for drastically different platforms to design the computing fabric for HPC systems. Moreover, as shown in **Fig. 1.2**(b) (adapted from [22]), no CMOS-based technologies are capable of simultaneously meeting the power budget and performance target of the Department of Energy's exascale goal. However, superconducting technology has the potential to achieve this goal even with lower power requirements thanks to the extreme energy efficiency of superconducting devices. As shown in **Fig. 1.2**(c), the superconducting devices outperform CMOS technology in terms of power consumption and switching delay by a significant margin [23], [24]. Besides, the use of superconducting interconnects solves the issues of resistive dissipation, residual heating, and voltage degradation. Considering

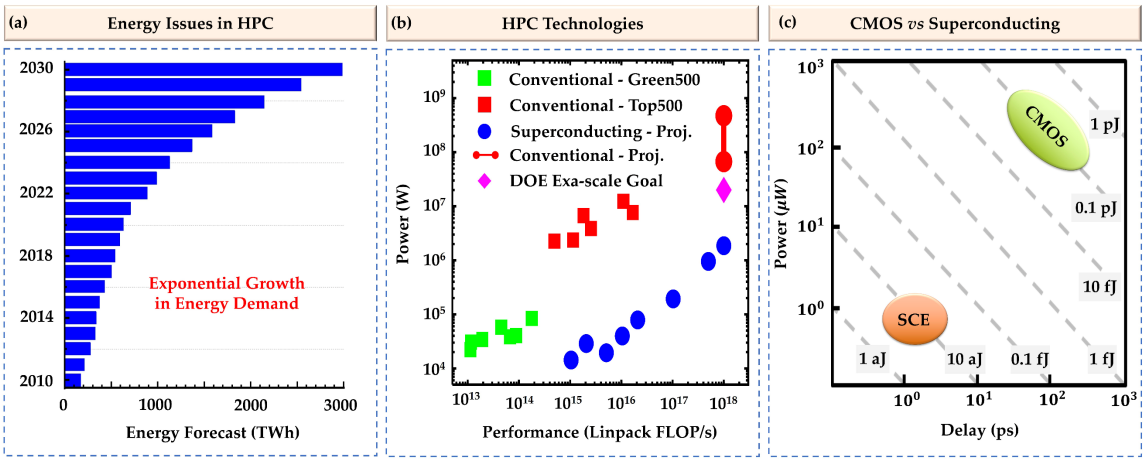


Figure 1.2. (a) Exponential growth of energy demand in high-performance computing every year. **(b)** Comparison of different technologies in terms of achieving the DOE's exa-scale goal. **(c)** Comparison of superconducting and CMOS technologies in terms of power consumption and switching delay.

these, the overall superconducting HPC system will exhibit substantial energy savings even after considering the cooling cost [25].

Finally, cryogenic computing is a natural fit for designing space electronics. The recent boom in the space industry pioneered by private space exploration companies, notably SpaceX, has opened endless opportunities for future space applications, which inevitably require more and more computing power. It is even possible to place an HPC server in space [26]. JJ and SQUID-based cryo-equipment have been used in several spacecrafts [22], [27]. The utilization of cryo-devices has enabled these spacecrafts to work with the electromagnetic radiation emitted by celestial objects over a wavelength range that is difficult to work with from the ground. Superconducting electronics, integrated into digital receivers featuring high-speed analog-to-digital converters, digital signal processing units, and ultra-low power readout electronics, further enhance their utility in cryogenic sensors and systems like transition edge sensors and microcalorimeters.

1.2 Overview of Existing Technologies

To develop a suitable cryogenic computing platform, various technologies like non-superconducting, superconducting, and hybrid technologies have been explored. **Fig. 1.3** presents a summary of the existing cryogenic technologies.

1.2.1 Non-Superconducting Technologies

Room temperature non-superconducting technologies like CMOS transistors, memristors, ferroelectrics, and spintronic devices, have been characterized at cryogenic temperatures to examine their capability as a cryogenic computing platform. The use of non-superconducting technologies for cryogenic applications provides high technological maturity and high scalability. Moreover, it has been reported that the cryogenic environment improves the performance of CMOS devices and circuits. The speed improved by 40% to 50% and the power dissipation reduced by 30% (depending on the circuit operation) for digital circuits at 4 K compared to the room temperature [28]. Cryogenic temperatures also reduce the leakage current substantially and improve carrier mobility [29]. More details on the cryogenic performance of non-superconducting technologies can be found in [12], [13], [23], [30], [31].

Despite notable improvement in the performance of non-superconducting technologies in cryogenic environments, they suffer from low-speed and high-power consumption issues. For cryogenic temperatures, there is a tight power budget. For example, the upper power limit for 0.1 K and 10 K temperatures are 10 mW and 100 mW, respectively. Therefore, high-power consumption of non-superconducting technologies becomes a significant challenge and limits the scalability within the power limit. As a result, the use of non-superconducting technologies at cryogenic temperature in different applications might become the eventual bottleneck, limiting the scalability. Additionally, for targeted applications like quantum computing, high-performance computing, and space electronics, the low speed of non-superconducting technologies is a major concern. For example, in quantum computing, the control processor and memory need to be fast enough to satisfy the time constraints dictated by the short qubit coherence time.

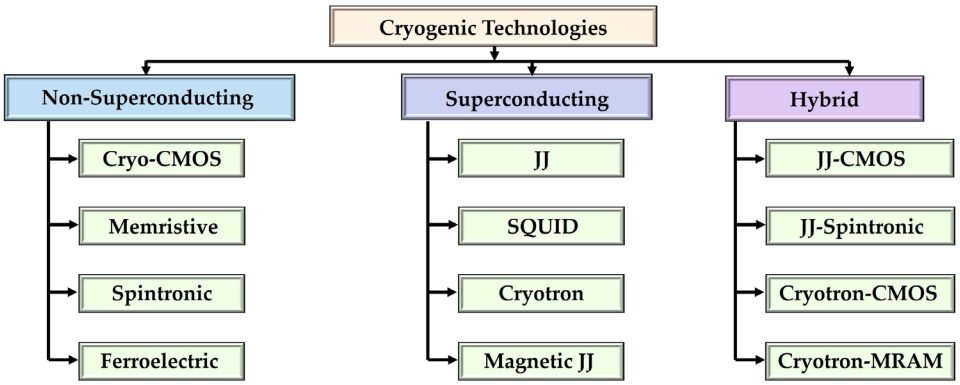


Figure 1.3. Overview of existing cryogenic computing technologies.

1.2.2 Superconducting Technologies

Unlike the non-superconducting technologies, superconducting devices and circuits can operate at ultra-low temperature (4 K) with high speed (hundreds of gigahertz) and low switching energy (0.1-1 atto-Joule) [9] (**Fig. 1.2**(c)). Therefore, superconducting technology is an obvious choice for cryogenic applications and various technologies like JJ, SQUID, magnetic JJ, cryotron, superconducting memristors, and superconducting transistors have been extensively explored. The superconducting devices and circuits are compatible in terms of speed and power requirements of the targeted cryogenic applications. However, they suffer from several pressing challenges. Trapping of magnetic flux is one of the major issues for the superconducting circuits [32]–[34]. Typically, the trapping of magnetic flux occurs during the cooling process when the metal films transform into a superconducting state. The trapped fluxes affect the superconducting circuits in two ways- (i) can either penetrate JJs or (ii) can couple with magneto-sensitive gates [32]. The first step to prevent flux trapping is to shield the earth's magnetic field which cannot solve the problem completely. Therefore, it is necessary to keep some space on the superconducting chips so that frozen vortices remain far from the magneto-sensitive portions of the chips [33]. Now, most of the JJ and SQUID-based superconducting circuits need inductive coupling which are magneto-sensitive. Therefore, this flux trapping imposes a limitation on the integration density of JJ and SQUID-based circuits. Another major issue arises from the requirement of current

bias for the superconducting devices which limits the fanout capability of superconducting logic circuits to one. To achieve a higher number of fanouts, additional circuitry (current splitter) is needed which negatively impacts the area efficiency and energy efficiency of the system.

1.2.3 Hybrid Technologies

To combine the advantages and solve the challenges of non-superconducting and superconducting technologies, hybrid systems have been proposed. The hybrid design, which utilizes the best features of each technology, is a promising approach to develop a suitable cryogenic computing system [35]. The hybrid method has mostly been utilized for developing cryogenic memory systems. In this approach, highly scalable non-superconducting memories are used as the storage element, and JJ-based superconducting devices are used to access them. **Fig. 1.4**(a) illustrates the concept of implementing a hybrid memory system that combines high scalability and maturity of non-superconducting technologies with high-speed and low-power operation of superconducting devices. **Fig. 1.4**(b) shows the high-level block diagram of a hybrid memory system where a dynamic random-access memory (DRAM) was used as the storage element [35], [36]. Similar memories have also been proposed with static random-access memories (SRAMs) [37] and magnetic random-access memories (MRAMs) [38], [39].

However, finding a suitable electrical interface between the non-superconducting and superconducting devices is a major concern for hybrid systems. The interface circuits are mainly cryogenic amplifier circuits that amplify millivolt signals of Josephson circuits to volt-level signals required for CMOS circuits. Superconducting amplifiers [40], semiconducting amplifiers [41], and hybrid superconductor-semiconductor amplifiers [35] are being explored to find suitable Josephson-CMOS interface circuits. Designing these interfaces is challenging and requires future research. Their speed and energy demand must not impose any additional bottlenecks.

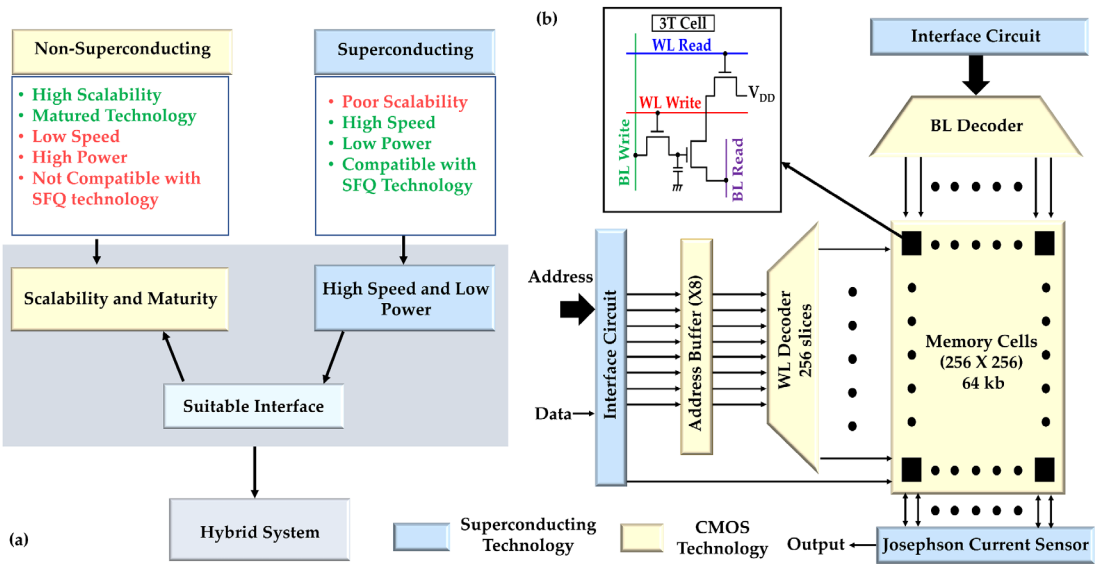


Figure 1.4. (a) Illustration of a hybrid memory system, where high scalability of non-supercomputing technology is combined with high speed and low power of superconducting technology. **(b)** Example of a hybrid memory system for cryogenic applications proposed in [35]. Here, 3-transistor DRAM is used as the memory element and JJ-based circuits are used as the access scheme.

1.3 Contributions of this Dissertation

The primary goal of this dissertation is to utilize the unique properties of topological and superconducting devices to develop a suitable cryogenic computing platform. We utilize the unique quantum anomalous Hall effect (QAHE) observed in topological devices, ferroelectric-controlled SQUID device (FE-SQUID), and heater cryotron to develop non-volatile memory, voltage-controlled Boolean logic family, reconfigurable logic circuit, and in-memory computing systems for cryogenic applications. We adopt experimentally calibrated simulation approach to verify the designed circuits and systems using the above-mentioned unique devices. We develop compact models for devices and simulation frameworks to perform circuit and system-level simulations. A brief overview of the major contributions of this dissertation is as follows-

(a) We utilize the unique characteristics of FE-SQUID to design a non-volatile memory array for cryogenic applications, which provides high scalability and the advantages

of separate read-write paths [42].

(b) We use this same device to design the very first voltage-controlled superconducting Boolean logic family which solves the single fanout and limited cascadability issues of the existing current-controlled designs [43].

(c) We leverage the highly energy-efficient read operation and separate read-write path capability of FE-SQUID-based memory system to develop a cryogenic in-memory matrix-vector multiplication system, which can be helpful for the hardware implementation of machine learning algorithms and can reduce the cooling cost with higher energy efficiency [44].

(d) We develop cryogenic ternary content addressable memory using FE-SQUID technology which allows exact search operation without needing additional complex peripheral circuitry [45].

(e) We design a reconfigurable logic circuit using only one heater cryotron device which allows us to perform various 1-, 2-, and 3-input logic operations without needing any modifications in the layout of the circuit. This provides higher area efficiency and better security against reverse engineering attacks [46].

(f) We utilize topological devices for exploring their circuit and system-level applications for the very first time. We use the QAHE phenomenon observed in topological devices to develop a non-volatile and ultra-low-power cryogenic memory system [47].

(g) We design a novel peripheral circuitry for the QAHE-based memory system to empower it for performing the basic Boolean logic operations where we demonstrate the unique benefits of QAHE and QAHE-based memory array for performing in-memory computing [48].

(h) We utilize the same memory array to perform a complex in-memory computing task, which is bit-serial addition. Here, we first design a QAHE-based in-memory majority function where the use of QAHE phenomenon allows us to use a simple voltage comparator circuit as the peripheral circuitry [49].

1.4 Organization of this Dissertation

The rest of the dissertation is organized as follows. We first discuss the use of a superconducting device (FE-SQUID) for designing various computing components for cryogenic environment. The use of FE-SQUID to design a non-volatile memory array, a voltage-controlled Boolean logic family, in-memory matrix-vector multiplication system, and ternary content addressable memory are presented in Chapter 2, Chapter 3, Chapter 4, and Chapter 5, respectively. Then, we discuss how we utilize the superconducting heater cryotron device to design a cryogenic reconfigurable logic circuit in Chapter 6. Next, we focus on the utilization of topological devices to design various circuits and systems. In Chapter 7, we present how we use the unique QAHE phenomenon observed in topological devices to design a non-volatile memory system. In Chapter 8 and Chapter 9, we discuss how we utilize the QAHE-based memory system to perform various in-memory Boolean logic operations and bit-serial addition with majority logic, respectively. Chapter 10 concludes the dissertation and provides an overview of future works.

CHAPTER 2
CRYOGENIC MEMORY ARRAY BASED ON FERROELECTRIC SUPERCONDUCTING QUANTUM INTERFERENCE DEVICE

2.1 Introduction

As highlighted in Chapter 1, a suitable cryogenic memory, capable of operating at/below 4 K, is a prime requirement for developing a large-scale quantum computing system with thousands of qubits, an energy-efficient high-performance computing system, a compatible aerospace electronic system. In addition, a key requirement for the storage device of a quantum computer is its high storage capacity. The state-of-the-art quantum algorithms require a large number of arbitrary rotations which eventually necessitate a large program memory [9]. Furthermore, to preserve the data integrity of the qubit states, the qubits undergo continuous error correction schemes [9], [12]; it requires extensive memory and bandwidth. In this quest, decades of research efforts have given rise to three major categories (and several sub-variants) of cryogenic memories based on non-superconducting, superconducting, and hybrid technologies [9], [13], [23], [35], [50]–[52]. However, each of these technologies suffers from its own unique challenges. Non-superconducting memories provide high scalability and high technological maturity but suffer from low-speed and high-power consumption. On the other hand, superconducting memories are extremely energy efficient but poorly scalable. Till date, the largest JJ-based superconducting memory that has been experimentally demonstrated has a storage capacity of only 16 KB. To combine the advantages of these two technologies, hybrid memories were proposed, but they need a suitable interface circuit to connect the non-superconducting circuits with the superconducting circuits, which remains a challenge.

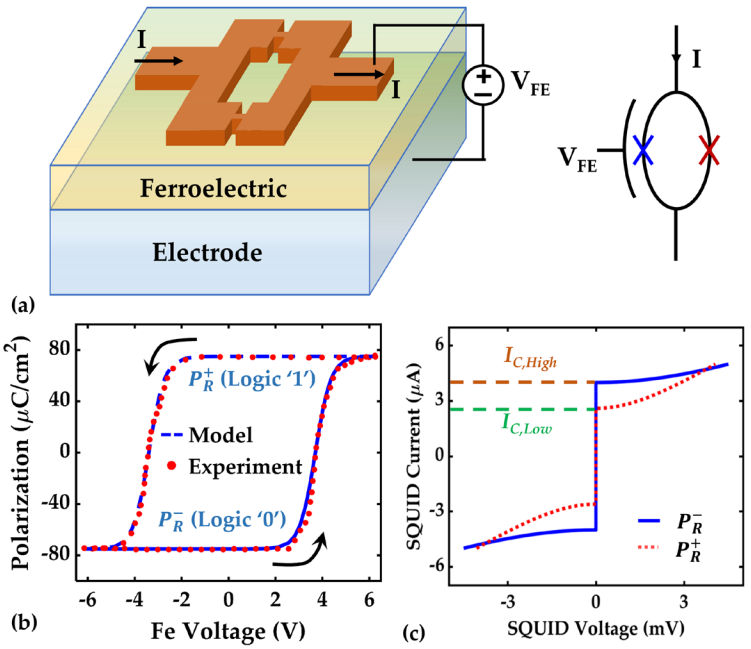
To address the issues of existing cryogenic memory technologies, we utilize a unique device, ferroelectric SQUID (FE-SQUID), to develop a suitable cryogenic memory system, which we discuss in this chapter. This device allows us to control the superconducting to resistive switching of SQUID with the ferroelectric polarization which can be controlled by suitable voltage biases. FE-SQUID-based cryogenic memory provides several advantages, including non-volatility and voltage-controlled write

operation due to the use of ferroelectric material, separate read-write path capability, high scalability, and high energy efficiency due to the use of SQUID.

2.2 Ferroelectric SQUID

Superconducting devices such as JJ and SQUID are typically fast (hundreds of giga-hertz operating frequency) and highly energy-efficient (atto-joule switching energy) and hence, have been utilized in different applications [53]. Researchers have tried continuously to introduce gate-tunability to these superconducting devices. The efforts include utilizing ionic liquid [54], [55], integrating ferromagnetic components [56]–[58], and so on. Although the integration of ferromagnetic components allows non-volatile tunability, magnetic tunability limits the scalability of the circuits and systems. Therefore, it has remained an unsolved challenge to introduce a convenient gating mechanism and voltage control in these devices. Recently, a new technique has been demonstrated that uses a ferroelectric material to tune the superconductivity of SQUID [59]. Ferroelectric materials show voltage-controlled non-volatile switching of the polarization. Therefore, the incorporation of the ferroelectric material introduces voltage-controlled non-volatile switching of the critical current of the SQUID which opens a lot of possibilities for utilizing this device at the circuit/system level. This device also offers better scalability compared to the existing superconducting devices like JJ, SQUID, magnetic JJ, and so on.

In FE-SQUID, a planar SQUID built with two weak links in parallel is fabricated on top of a ferroelectric material (schematic and symbol shown in **Fig. 2.1**(a)). The choice of superconducting and ferroelectric materials depends on lattice matching between them [60], [61]. Ref. [59] demonstrated a FeSQUID device by fabricating a 15 nm thick $\alpha Mo_{80}Si_{20}$ planar SQUID and 70 nm thick $PbZr_{0.2}Ti_{0.8}O_3$ (PZT) interface on a 15 nm thick $SrRuO_3$ bottom electrode. Ferroelectrics are dielectric materials that can retain polarization switchable by external voltage/electric field. **Fig. 2.1**(b) shows the voltage-controlled non-volatile switching of the P_{FE} of a PZT ferroelectric. In a ferroelectric material, internal P_{FE} translates into a surface charge that induces an electric field and injects direct charge [62]. Now, when a SQUID is fabricated on top of a ferroelectric, the superconducting layer screens the charge bound in the interface. This bound charge in the



(d) Device Characteristics of FE-SQUID					
Logic State	P_{FE}	SC Energy Gap (Δ)	Critical Temperature (T_C)	Critical Current (I_C)	Normal Resistance (R_N)
'0'	P_R^-	1.05 meV	6.8 K	4 μA	0.95 $k\Omega$
'1'	P_R^+	1.35 meV	5.8 K	2. 6 μA	1.75 $k\Omega$

Figure 2.1. (a) Device structure and circuit symbol of FE-SQUID. **(b)** Polarization-voltage characteristics for the PZT ferroelectric along with the validation of the developed compact model with the experimental results. **(c)** Current-voltage characteristics of FE-SQUID. Two polarization states of the ferroelectric leads to two levels of I_C. **(d)** Effects of ferroelectric polarization on different device characteristics of FE-SQUID.

surface of the interface ($\oiint P_R . dA$, where P_R is the remnant P_{FE} and A is the surface area) directly depends on the remnant P_{FE} – negative P_R (P_R^-) increases the bound charge in the surface of the interface whereas positive P_R (P_R^+) results in an opposite effect [59], [60]. The resulting change in the carrier density impacts the critical temperature (T_C) and hence, the superconducting energy gap (Δ). The dependence of Δ on T_C is given by Bardeen–Cooper–Schrieffer (BCS) theory [63], [64]-

$$\Delta\,(T) = 1.763 k_B T_C \; tanh\left(2.2\sqrt{\frac{T_C}{T} - 1}\right), \tag{1}$$

where T is the temperature and k_B is the Boltzmann constant. According to the Ambegaokar–Baratoff (AB) theory [65], $\Delta(T)$ determines the critical current (I_C) as given by-

$$I_C = \frac{\pi\Delta}{2q_e R_N}\tanh\left(\frac{\Delta}{2k_B T}\right), \tag{2}$$

where q_e is the charge of an electron and R_N is the normal state resistance of the SQUID. Therefore, based on the two non-volatile polarization states of the ferroelectric (controlled by voltage), two levels of I_C are observed in the *I-V* characteristics of the SQUID- P_R^- and P_R^+ states of the ferroelectric lead to high ($I_{C,high}$) and low ($I_{C,low}$) I_C for the SQUID, respectively (**Fig. 2.1**(c)). The effects of P_{FE} states on different characteristics (such as T_C, $\Delta(T)$, I_C, and R_N) are shown in **Fig. 2.1**(d).

2.3 Heater Cryotron-based Selector

To develop a working FE-SQUID-based memory array, *hTron* [66], [67] is used as the access device which allows us to individually access any specific cell in the array. *hTron* is a four-terminal device, where the gate and the channel are superconducting nanowires (**Fig. 2.2**(a)). The channel and gate are normally superconducting and can be made resistive by applying a sufficient current through the gate. The gate here acts as the heater for the channel. When the gate current (I_G) exceeds its switching current (I_{SW}), it starts to generate thermal phonons which are carried to the channel to create a hotspot. This effect suppresses the superconductivity of the channel and reduces its critical current. Now, with the increase of the gate current, when the channel critical current becomes lower than the channel current, the channel switches to the resistive state. **Fig. 2.2**(c) shows the

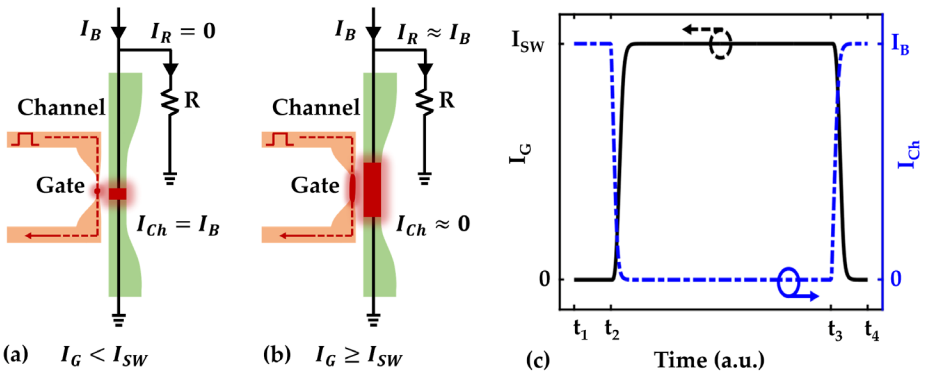


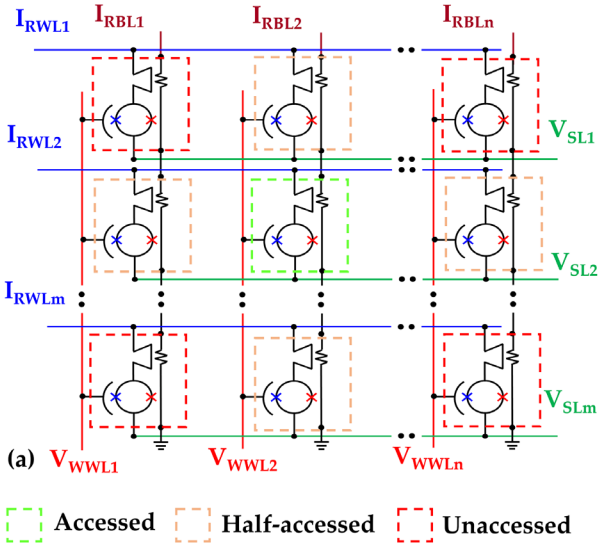
Figure 2.2. (a)-(b) Gate current -driven superconducting → resistive switching and **(c)** *I-V* characteristics of a heater cryotron.

gate current-controlled switching of the channel. *hTron* is utilized to control the current flow through any specific path in the FE-SQUID-based memory array.

2.4 Ferroelectric SQUID-based Memory Array

Fig. 2.3(a) shows the schematic of FE-SQUID-based memory array, where each cell is formed by connecting an *hTron* in series with FE-SQUID. Here, we connect all the ferroelectric gate terminals of FE-SQUIDs in the same column with a write word line (WWL). Similarly, all the gate terminals of *hTrons* in the same column are connected with a read bit line (RBL). The channels of *hTrons* are connected row-wise with a read word line (RWL). Finally, the other terminals of FE-SQUIDs (that are not connected with *hTron*) are connected row-wise with a source line (SL).

Fig. 2.3(b) shows the biasing scheme to write and read a specific cell, (2, 2) marked by a green rectangle, in an $m \times n$ FE-SQUID-based memory array. As shown in **Fig. 2.1** (b), we can apply suitable voltage biases (positive and negative) across the ferroelectric material to perform the write operations (write '1' and '0', respectively) in an FE-SQUID. As a result, to write into a specific memory cell that is manipulating the polarization state of the ferroelectric material of that FE-SQUID cell, we utilize $\frac{V}{2}$ biasing scheme [47]. We apply suitable biases ($\pm V_{WRITE}$ or $\pm \frac{V_{WRITE}}{2}$) to the write word lines (WWLs) and source lines (SLs) in a way so that only the accessed cell ((2, 2) cell in this cell) gets $\pm V_{WRITE}$ across its ferroelectric. For this biasing, half-accessed cells, that are in



(b)

Biasing to Access (2, 2) Cell	
Write Operation	
WWL2	V_{WRITE}
Other WWLs	$V_{WRITE}/2$
SL2	0 V
Other SLs	$V_{WRITE}/2$
Read Operation	
RWL2	I_{READ}
Other RWLs	0 A
RBL2	0 A
Other RBLs	I_{SW}

Figure 2.3. (a) Schematic of FE-SQUID-based memory array. **(b)** Biasing scheme to access the (2, 2) cell in the array for write and read operations.

the same row or column with the accessed cell (marked by orange rectangles), get $\pm$ ½V_{WRITE}, and unaccessed cells, that are not either in the same row or column with the accessed cell (marked by red rectangles), get 0 V across them. To ensure that only the accessed cell undergoes the intended write operation, we need to carefully choose the write voltage following this condition - (½$|V_{WRITE}| <$ |coercive voltage, $V_C| <$ $|V_{WRITE}|$).

As shown in **Fig.2.1**(c), if a suitable current ($I_{C,Low} < I < I_{C,High}$) is applied to the SQUID, we will get superconducting (0 V) and resistive (nonzero voltage) state for P_R^- and P_R^+, respectively. For read operations, we utilize this characteristic and apply suitable bias current through the SQUID of the accessed cell. First, all the WWLs and SLs are kept grounded. Then, RBLs are first biased (with current) in such a way that all the *hTron*s in the same column with the accessed cell remain superconducting, but the other *hTron*s turn resistive. Next, when the read current (I_{READ}) is applied to the corresponding RWL, the *hTron* of the accessed cell channels it through that cell only and the SQUID generates either $V_{SQUID} = 0$ (for P^-) or $V_{SQUID} = I_{READ}R_N$ (for P^+). Here, I_{READ} needs to be selected carefully. The condition to select I_{READ} is- $I_{C,Low} < I_{READ} < I_{C,High}$. These

voltage levels can be sensed using cryogenic sense amplifiers (SA) to identify the stored binary data [51].

2.5 Modeling and Simulation Framework

We develop a circuit compatible compact model for FE-SQUID in Verilog-A and calibrate that with experimental data. **Fig. 2.4** shows the modeling approach for FE-SQUID. We first model the non-volatile switching of ferroelectric polarization using the following Preisach model equation-

$$P_{FE} = P_S \tanh\left[\frac{1}{2E_C}\ln\left(\frac{P_S+P_R}{P_S-P_R}\right)\ (E_{FE} \mp E_C) + P_{off}\right], \quad (3)$$

where P_S, P_R, and P_{off} are the saturation, remnant, and offset polarizations, respectively. E_C and E_{FE} represent the coercive and applied electric fields across the ferroelectric, respectively. Here, we extract the values of different parameters used in equation (3) from the experimental result (reported in Ref. [59]) for a Led Zirconate Titanate (PZT) ferroelectric. **Fig. 2.1**(b) shows the validation of the P_{FE}-V characteristics of PZT obtained from the experiment reported in [59]. Next, we phenomenologically (look-up-table-based approach) implement the dependence of T_C on P_{FE} using the information mentioned in **Fig. 2.1**(d) obtained from [59]. Then we use equation (1) to calculate Δ(T) and subsequently equation (2) to calculate I_C of the SQUID. Finally, we model the I-V characteristics of the SQUID by utilizing the resistively and capacitively shunted junction (RCSJ) model, as given by the following equation-

$$V = \begin{cases} 0, & I < I_C \\ I \times R_N, & I \geq I_C \end{cases} \quad (4)$$

Next, we develop a LUT-based phenomenological compact model for the *hTron* device. In each time step, the model takes applied gate and channel currents as input and based on the value of the gate current, reads the corresponding channel critical current from the LUT. Finally, the model compares the applied channel current with the channel critical current obtained from the LUT and determines whether the channel will switch or not. Using these two compact models of FE-SQUID and *hTron* devices, we develop a simulation framework in HSPICE, which we utilize to demonstrate the write and read operations in the proposed memory array.

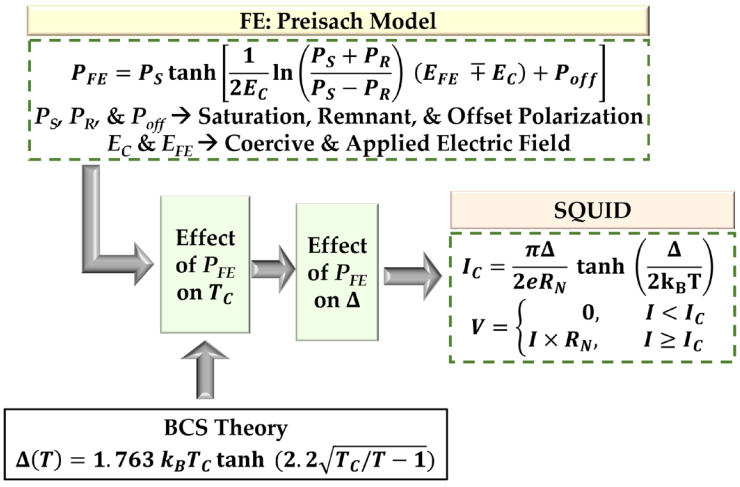


Figure 2.4. Modeling approach of FE-SQUID.

2.6 Simulated Memory Operations

Fig. 2.5 shows the simulated write and read operations of (2, 2) cell in the FE-SQUID memory array. **Figs. 2.5**(a) and (b) show the ferroelectric voltage of the accessed, half-accessed, and unaccessed cells during write '0' and '1' operations, respectively, resulting from the biasing scheme mentioned in **Fig. 2.3**(b). During write '0' ('1') operation, we initialize all the cells in the array in the logic '1' ('0') state. As seen in **Figs. 2.5**(d) and (e), P_{FE} state (and hence, I_C) of only the accessed cell is switched where other cells remain in the previous state.

And, to read from any memory cell, we utilize the *hTron*s connected in series with the FE-SQUIDs to flow the read current only through the accessed cell. We first apply I_{SW} $(100\ \mu A)$ to RBLs in a way so that all the *hTron*s except the one connected with the accessed cell become resistive. Therefore, the applied current to the RWL flows through the accessed cell only. **Fig. 2.5**(c) shows the RBL and RWL biasing, and **Fig. 2.5**(f) shows the corresponding SQUID voltage of the accessed cell during read operation.

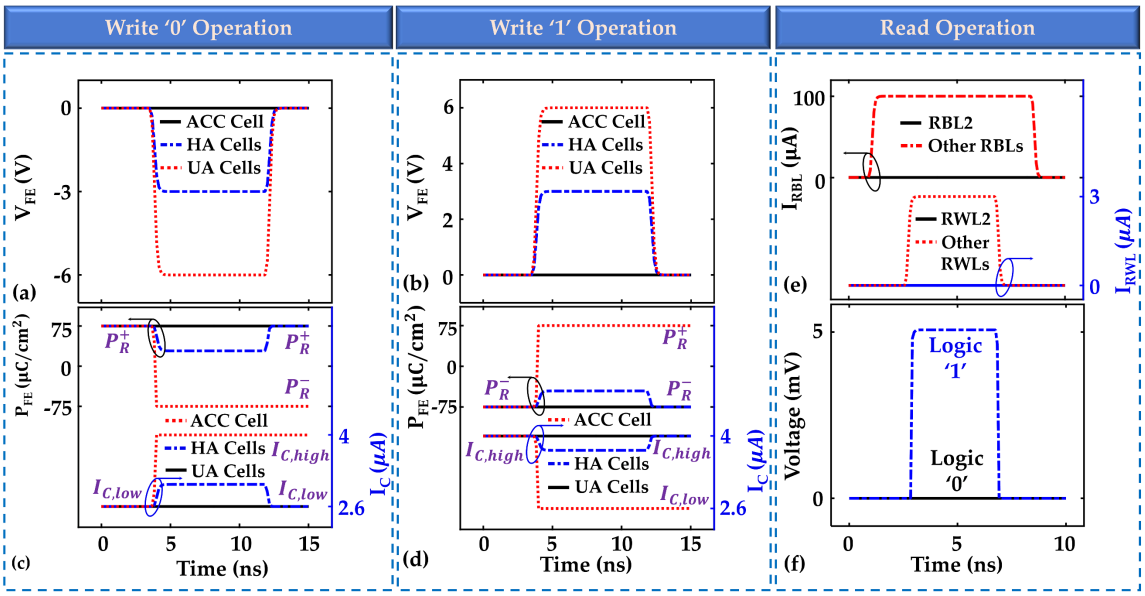


Figure 2.5. Temporal dynamics of the ferroelectric voltage of accessed (ACC), half-accessed (HA), and unaccessed (UA) cells during **(a)** write '0' and **(b)** write '1' operations. Switching of P_{FE} and I_C of the cells during **(c)** write'0' and **(d)** write'1' operations. Time dynamics of **(e)** RBL and RWL biasing, and **(f)** SQUID voltage of the accessed cell.

2.7 Comparative Analysis

Our proposed cryogenic memory array based on FE-SQUID combines the benefits of non-superconducting ferroelectric and superconducting SQUID technologies and hence, provides several advantages, including non-volatility, high scalability, voltage-controlled operation, and separate read and write path capability. To highlight the advantages of the FE-SQUID-based memory system, in **Table 2.1**, we present a qualitative comparison with the major cryogenic memory technologies reviewed in [23].

2.8 Summary

We proposed a scalable cryogenic memory array, combining the unique characteristics of an FE-SQUID, and an *hTron*. FE-SQUID-based memory array offers non-volatility, scalability, operation controlled by voltage, high energy efficiency, and the capability of separate read and write paths by combining the benefits of non-superconducting ferroelectric and superconducting SQUID technologies. Moreover, contrary to the conventional SC memory topologies, our design averts using inductive coupling and

Table 2.1: Comparison with Major Cryogenic Memory Technologies

Technology / Criteria	Cryogenic CMOS	JJ	Magnetic JJ	Hybrid	FE-SQUID (This Work)
Speed	Low	High	High	Medium	High
Power	High	Low	Low	Medium	Low
Non-volatility	Yes/No	No	Yes	Yes/No	Yes
Inductive Coupling	No	Yes	Yes	No	No
Magnetic Bias	No	No	Yes	No	No
Separate Read-Write Paths	Yes/No	No	No	No	Yes
Voltage Control	Yes	No	No	Yes/No	Yes
Scalability	High	Poor	Unknown	Medium	High

bulky peripheral circuits. The programming voltage can be reduced by adopting a different ferroelectric material (*e.g.*, HZO), or by using a thinner ferroelectric layer.

CHAPTER 3
VOLTAGE-CONTROLLED LOGIC FAMILY WITH FERROELECTRIC SUPERCONDUCTING QUANTUM INTERFERENCE DEVICE

3.1 Introduction

The idea of superconducting electronics was first put forward in the 1950s with the effort of building magnetic field-modulated superconducting wires [53], [68]. The requirement of extremely low temperature was one of the major roadblocks for the superconducting devices to be used in practical applications which now becomes an advantage, thanks to the recent interests in superconducting qubit-based quantum computers and space electronics [9], [17], [23], [69]. Superconducting logic circuits can significantly improve the performance of the control processor in quantum computers and aircrafts for space exploration [9]. Even in classical computing, superconducting processor shows immense potential.

Over the last few decades, several superconducting logic devices [46], [68], [70]–[78] have been introduced. All of these except JJ have been limited within the basic characterization. JJ has been successfully utilized in a wide variety of applications due to its high speed (hundreds of gigahertz) and low power consumption (sub-aJ/bit). However, JJ-based circuits suffer from challenges: (i) the manipulation of single flux quanta causes issues while driving large impedances or fanning out digital signals, (ii) difficulty in the fabrication of uniform JJs over a wafer and the effects of magnetic field limit the distance between two JJs and hence, the integration density [79]. Another disadvantage (particularly to logic circuits) of two-terminal JJ and SQUID is that the absence of resistance in superconductors hinders the voltage biasing and hence, superconducting devices cannot offer convenient gating mechanisms like the semiconducting transistors. Therefore, there has been a significant amount of effort to implement a gate-tunable superconducting device [68], [72]–[74], [77].

In this chapter, we present a voltage-controllable cryogenic Boolean logic family based on FE-SQUID, which will allow the close integration of logic and memory blocks since both logic and memory can be designed using the same device. Therefore, there will be no mismatch between the logic and memory blocks in terms of speed, power

consumption, and fabrication process which will also reduce the 'memory wall' bottleneck [44], [48], [49], [80], [81]. The major contributions of this work are outlined below-

- Designing the first voltage-controlled superconducting logic circuits.
- Solving the issues of single fanout and the need for extra current splitter circuits in current-controlled superconducting logic designs.
- Utilizing the interaction between FE-SQUIDs and *hTron*s to ensure that both input and output values match, allowing one logic gate's output to drive the next input.

3.2 Design Methodology of Logic Circuits

We use the voltage-controllability of SQUID's superconductivity (shown in **Figs. 2.1**(a)-(c)) to design the basic Boolean logic gates for cryogenic applications. **Table 3.1** illustrates how we utilize FE-SQUID to design logic gates. Here, we use the polarization states (P_R^- and P_R^+) of the ferroelectric to define the logic states (logic '0' and '1', respectively) which can be switched by applying suitable voltage input. Although the coercive voltage of PZT is $\sim$3 V (**Fig. 2.1**(b)), in our design, we use much larger voltage as the input (-6 V or $+6$ V for logic '0' and '1', respectively) to ensure that the ferroelectric remains in either P_R^- or P_R^+. These voltage levels can be reduced using different ferroelectric materials with scaled dimensions. In our design, logic '0' and '1' correspond to $I_{C,high}$ and $I_{C,low}$, respectively (**Fig. 2.1**(c)). Now, if we bias the FeSQUID with a bias current ($I_{C,low} < I < I_{C,high}$), FeSQUID with logic '0' ('1') input corresponds to superconducting (resistive) state of the SQUID. Based on the state of the SQUID, the bias current will flow either through the SQUID or through the external circuitry. We harness this phenomenon to design the FeSQUID-based logic gates.

For any logic gate, a crucial requirement is the capability of handling fan-out. For a cascadable logic system, the outputs of the logic gates should be able to drive the inputs of the next stage. Now, for FE-SQUID, to drive one gate by the output of another gate, the output voltage needs to be sufficiently large to set the intended polarization in the ferroelectric of the FE-SQUID in the next stage. As seen in **Fig. 2.1**(d), the I_C values of

Table 3.1: Design Methodology of FE-SQUID-based Logic Circuits

Biasing Condition to SQUID (Fixed): $I_{C,low} < I < I_{C,high}$			
Logic State	**V_{FE}**	**P_{FE}**	**Device State**
'0'	-6 V	Negative (P_R^-)	Superconducting
'1'	+6 V	Positive (P_R^+)	Resistive

FESQUID are 2.6 μA and 4 μA, and the values of R_N are 0.95 $k\Omega$ and 1.75 $k\Omega$ for P_R^- and P_R^+, respectively. For proper operation of the copy gate shown in **Fig. 3.1**(a), the value of the external resistance (R) needs to be chosen in a way so that when the FE-SQUID becomes resistive, almost all the bias currents flow through R. Also, we have to choose a bias current so that it satisfies the condition- $I_{C,low} < I < I_{C,high}$. Here, we choose 3.2 μA and 10 Ω for I and R, respectively. For these chosen values, we get $\sim 31\ \mu V$ (0 V) for logic '1' ('0') at the output of the copy gate shown in **Fig. 3.1**(a) which are not sufficient to drive the FE-SQUID of the next stage. To circumvent this issue, we utilize *hTron* [51], [66] at the output of each logic gate to develop a cascadable logic system based on FE-SQUIDs. **Fig. 2.2** shows the schematic and the I-V characteristics of *hTron*. After using *hTron* (the working principle of the designed logic gates is discussed later), the output voltage of the copy gate becomes $-6\ V$ and $+6\ V$ for logic '0' and '1', respectively as required to drive the FeSQUID of the next stage (**Fig. 3.1**(b)).

3.3 Working Principle and Functional Verification

In this section, we discuss the working principle and functional verification of the FE-SQUID and *hTron*-based Boolean logic circuits. For the verification, we utilize the same simulation framework discussed in Section 2.5 of Chapter 2, which was developed with the compact models of FE-SQUID and *hTron*.

First, we developed 1-input Copy gate where the input is applied as the voltage across the ferroelectric. **Fig. 3.2**(a) shows the schematic of the Copy gate. In case of logic '0' (-6 V) applied to the input and suitable bias current (I) is applied, the FE-SQUID shows $I_{C,high}$ (>I) and becomes superconducting (**Fig. 3.2**(b)). Therefore, almost all the bias current flows through FE-SQUID and the *hTron* gate does not get enough current to switch the channel to its resistive state (**Fig. 3.2**(b)). Due to the proper biasing of *hTron*,

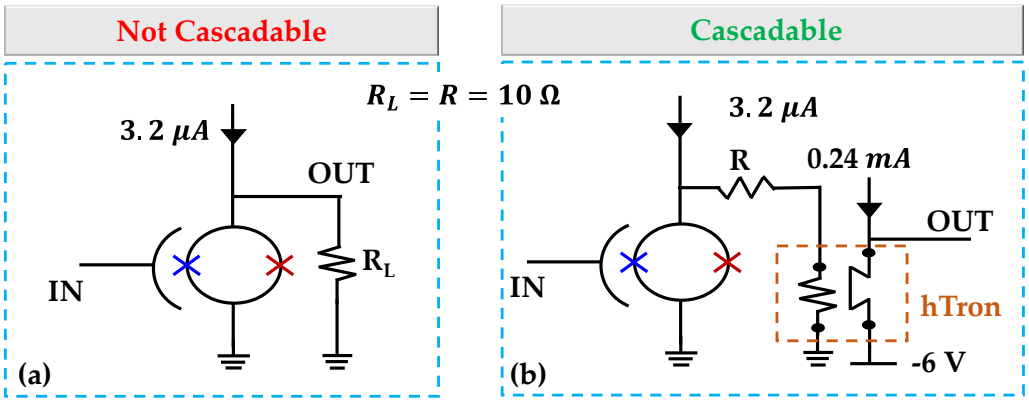


Figure 3.1. (a) Schematic of a Copy gate using a FeSQUID. The output of this configuration is not sufficiently large to drive the input of the next gate and hence, this configuration is not cascadable. **(b)** Modified version of the Copy gate of (a) where a *hTron* is used to convert the output voltage to the same level of the input so that two or more logic gates can be cascaded.

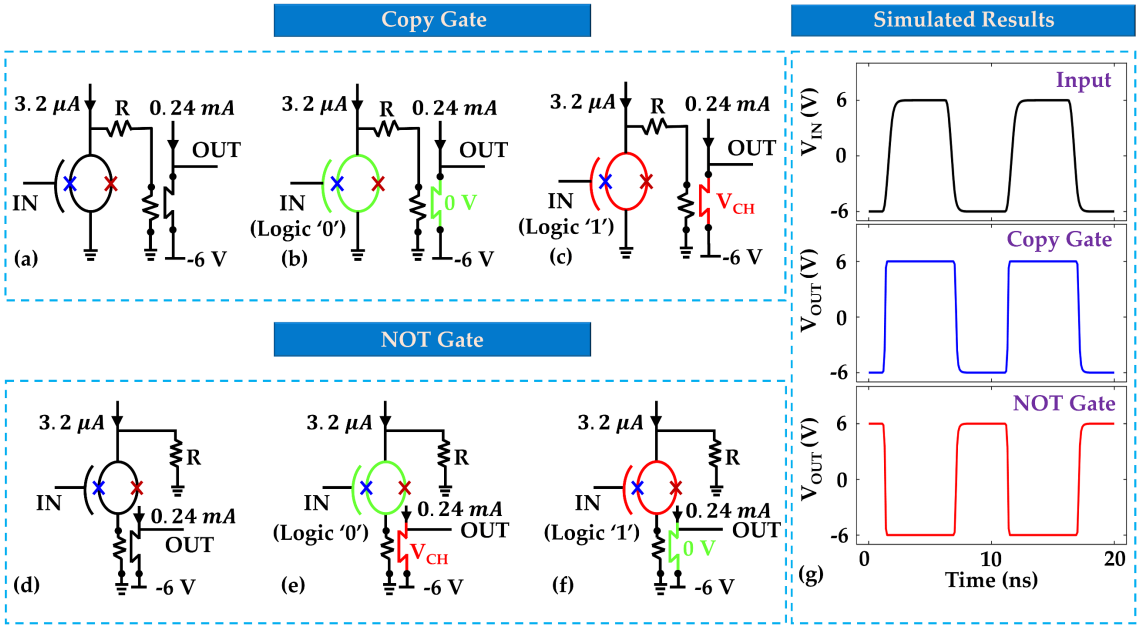


Figure 3.2. (a) Schematic of FeSQUID-based Copy gate. Working principle of the Copy gate for **(b)** logic '0' and **(c)** logic '1' at the input. **(d)** Schematic of FeSQUID-based NOT gate. Working principle of the NOT gate for **(e)** logic '0' and **(f)** logic '1' at the input. Green and red colors correspond to the superconducting and resistive states, respectively. **(g)** Simulated results for the 1-input Copy and NOT gates.

we get logic '0' (−6 V) at the output. Now, for logic '1' (+6 V) at the input, FE-SQUID becomes resistive, driving most of the bias current to the gate of *hTron* which switches the channel to its resistive state (**Fig. 3.2**(c)), and hence, we obtain logic '1' (+6 V) at the output.

Next, we design the NOT gate whose schematic is presented in **Fig. 3.2**(d). In the NOT gate, the gate of the *hTron* is connected in series with the FE-SQUID. Therefore, for logic '0' (−6 V), FE-SQUID becomes superconducting and passes enough current to the gate of *hTron* which consequently switches the channel to its resistive state and we get logic '1' (+6 V) at the output (**Fig. 3.2**(e)). For logic '1' (+6 V) at the input, the exact opposite scenario occurs for the FE-SQUID and *hTron*, and we get logic '0' (−6 V) at the output (**Fig. 3.2**(f)). **Fig. 3.2**(g) presents the simulation results for the 1-input Copy and NOT gates, showing the successful operations.

Having discussed the 1-input gates, next we discuss the 2-input AND and OR gates where we use two FE-SQUIDs and apply the inputs as voltages across the ferroelectric materials of the FE-SQUIDs. **Fig. 3.3**(a) captures the schematic of the AND gate where we connect two FE-SQUIDs in parallel and hence, if any of the FE-SQUIDs become superconducting ('00', '01', and '10' cases), most of the bias current flows through that and deprives the *hTron* gate to get enough current to switch the channel to its resistive state (**Figs. 3.3**(a) and (b)). Consequently, we get logic '0' (−6 V) at the output. However, for the '11' case at the input, both the FE-SQUIDs switch to the resistive state driving the bias current through the *hTron* gate which makes its channel resistive and creates logic '1' (+6 V) at the output (**Fig. 3.3**(c)). On the contrary, in the OR gate, two FE-SQUIDs are connected in series (**Fig. 3.3**(d)). Therefore, only when both the FE-SQUIDs remain superconducting ('00' case), the *hTron* gate does not experience enough current to switch the superconducting channel and creates an output of logic '0' (−6 V)

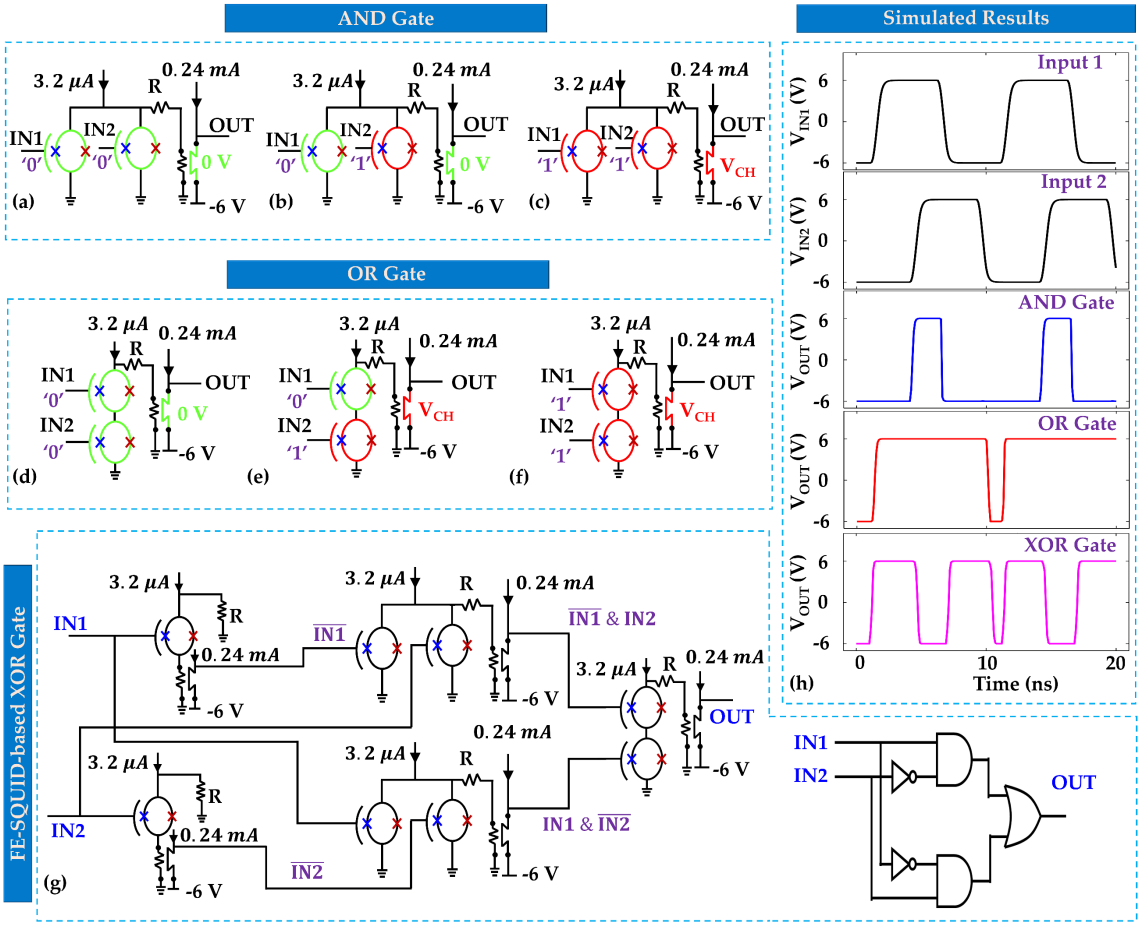


Figure 3.3. Working principle of FeSQUID-based AND gate for **(a)** '00', **(b)** '01'/'10', and **(c)** '11' inputs. Operation of FeSQUID-based OR gate for **(d)** '00', **(e)** '01'/'10', and **(f)** '11' at the input terminals. **(g)** Schematic of XOR gate built cascading FeSQUID-based NOT, AND, and OR gates. **(h)** Simulated results for the 2-input AND, OR, and XOR gates.

(**Fig. 3.3**(d)). For other cases (‘01’, ‘10’, and ‘11’), at least one FE-SQUID becomes resistive and hence, the *hTron* channel switches to the resistive state and we get logic ‘1’ (+6 V) at the output (**Figs. 3.3**(e) and (f)).

Lastly, to demonstrate the capability of the FE-SQUID-based logic family to handle higher fan-out and cascadability, we simulate a 2-input XOR gate using the FE-SQUID-based NOT, AND, and OR gates. The schematic of the XOR gate is shown in **Fig. 3.3**(g). **Fig. 3.3**(h) shows the simulation results for the 2-input AND, OR, and XOR gates.

3.4 Summary

In summary, using voltage-controlled superconductivity in FE-SQUIDs, we developed a CMOS-like voltage-controlled superconducting logic family (1-input Copy and NOT gates, and 2-input AND and OR gates), which can potentially revolutionize the implementation of (i) classical control architecture of quantum computers, (ii) exascale high-performance computing systems, and (iii) space electronics. Moreover, our designed logic circuits are transferrable to FE-SQUIDs with different ferroelectric and superconducting materials as well as other voltage-controlled superconducting devices such as JJ field effect transistor [82], Dayem bridge transistors [77], [83], etc.

CHAPTER 4
IN-MEMORY MATRIX-VECTOR MULTIPLICATION USING FERROELECTRIC SUPERCONDUCTING QUANTUM INTERFERENCE DEVICE

4.1 Introduction

Although silicon (Si)-based classical computing has enjoyed immense progress in last few decades, it cannot solve a number of commercially and scientifically important problems (such as accurate simulation of large molecules, integer factorization, optimization, etc.) in realistic timeframe [9], [23]. However, a quantum computer (QC) has the ability to solve these problems in polynomial time since it leverages quantum properties for storing and processing data [7]. Qubit is the fundamental block of a QC whose states are manipulated to solve any problem. Qubits are extremely sensitive to noise and hence, to protect the states, they are placed at extremely low temperature (tens of milli-Kelvin) [9]. Control processor and memory are the two other essential components of a QC. Currently, the control processor and memory are kept at room temperature which requires a large number of long cables to connect the qubits with the control processor and memory. Now, to realize a QC with the full capacity, it must be scaled to thousands of qubits. To realize such a large-scale QC, the control processor and memory must be placed close to the qubits [84]. Therefore, cryogenic electronics and memory are crucial requirement for a QC to achieve the dream capability [9], [47], [67].

While the superconducting cryogenic electronics provides high energy-efficiency and fast speed in conventional computing tasks, this still suffers from the *von Neumann* and *memory wall* bottlenecks [48], [85]. Due to the use of physically separate memory and processing units, large energy and time overheads are required for the back-and-forth data movement. Therefore, the existing cryogenic electronics will not perform well for implementing computationally intensive machine learning (ML) algorithms which entails unique benefits for the error correction [86] and control [87] of the qubit states . To circumvent this, compute-in-memory (CiM) architectures have been proposed which achieves computation capability inside the memory array with a slight modification in the peripheral circuitry [48], [88]. Another advantage of CiM is that most of the memories

leverage Ohm's law (multiply) and Kirchoff's law (accumulate) for read operation and hence, makes the implementation of matrix-vector multiplication (MVM) straightforward since MVM is necessarily a multiply and accumulate operation. Note, MVM is the most frequently performed operation in ML algorithms. Therefore, the implementation of in-memory MVM operation with a suitable and scalable cryogenic memory is of paramount importance for realizing a large-scale practical QC.

Although almost all the memory platforms have been utilized to implement in-memory MVM at room temperature, only quantum anomalous Hall effect (observed in topological insulators)-based memories have been utilized for in-memory logic [48] and MVM operations at cryogenic environment [88]. In this chapter, we discuss how we implement in-memory MVM operation with ferroelectric superconducting quantum interference device (FE-SQUID)-based cryogenic memory [42]. In FE-SQUID, the ferroelectric material stores the memory states which are voltage-controllable, and the SQUID is used to read the memory states which is current-controllable. Therefore, FE-SQUID features several unique advantages such as voltage-controlled robust non-volatile memory states, independent read-write paths, and high scalability. Moreover, since two completely separate paths are used for write and read operations, FE-SQUID-based memory does not impose any constraint on the value of the read bias. It allows us to use input levels with large separation to perform MVM operation that provides better distinguishability in decision-making and hence, improves the accuracy in different classification tasks. We evaluate our FE-SQUID-based MVM to perform three classification tasks using MNIST handwritten digit [89], fashion [90], and emotion [91] datasets and we achieve an average accuracy of 93.83%, 80.49%, and 92.5%, respectively. Finally, we show that there is a tradeoff between accuracy and number of rows and columns of the array used while performing the classification tasks.

4.2 In-Memory MVM With FE-SQUID

As mentioned earlier, cryogenic implementation of ML algorithms is useful for QC and space exploration. For any ML algorithm, the most frequent and computationally intensive operation is MVM. In this section, we demonstrate how we can perform MVM operation in an FE-SQUID memory array with the help of a modified peripheral circuitry.

Fig. 4.1 (a) shows the schematic of the FE-SQUID-based in-memory MVM system. The write and read operation in this memory system are performed in the same mechanism discussed in Chapter 2.

Here, we map the input matrix ($[I_1\ I_2\ .\ .\ I_n]$) as the read current applied to the RWLs. For logic '0' and '1' inputs, we use $10\ \mu A$ and $25\mu A$ read currents, respectively. Thanks to the use of FE-SQUID, we have the flexibility to use current levels with large distinguishability. We only need to make sure that the chosen current levels are larger than $I_{C,High}$. This is a major advantage over other memory arrays thanks to the separate read-write path capability. And the weight matrix ($[W_1\ W_2\ .\ .\ W_n]$) is mapped as the memory state of the FE-SQUID cells. We use one column of the array to map one element of the weight matrix. To demonstrate the proof of concept, we consider that both the input and weight matrices consist of three elements, and we use three bits for each element of the weight matrix which require a 3×3 array. MSB and LSB of an element of the weight matrix are stored in the first and last rows of each column, respectively. **Fig. 4.1**(a) shows the mapping into the memory array. Now, based on the current and the memory states of the FE-SQUID cells, different levels of voltages will develop across each row which is fed to sense amplifier. For the 3×3 memory array, sense amplifiers in each row generate outputs with two bits. **Fig. 4.1**(b) shows the schematic of sense amplifier which uses two voltage comparators with suitable reference voltages and one 2:1 multiplexer to select the reference voltage for the second voltage comparator. **Fig. 4.1**(c) shows the schematic of Josephson junction FET-based voltage comparator [75] that we use in this work. Finally, we use two left shifters for the first and second rows and add the outputs of all the rows to determine the MVM output. Note, if we use larger array, for n number of rows, we will need $n-1$ number of left shifters. **Figs. 4.1**(d) and (e) show two examples of MVM operation performed with FE-SQUID-based memory.

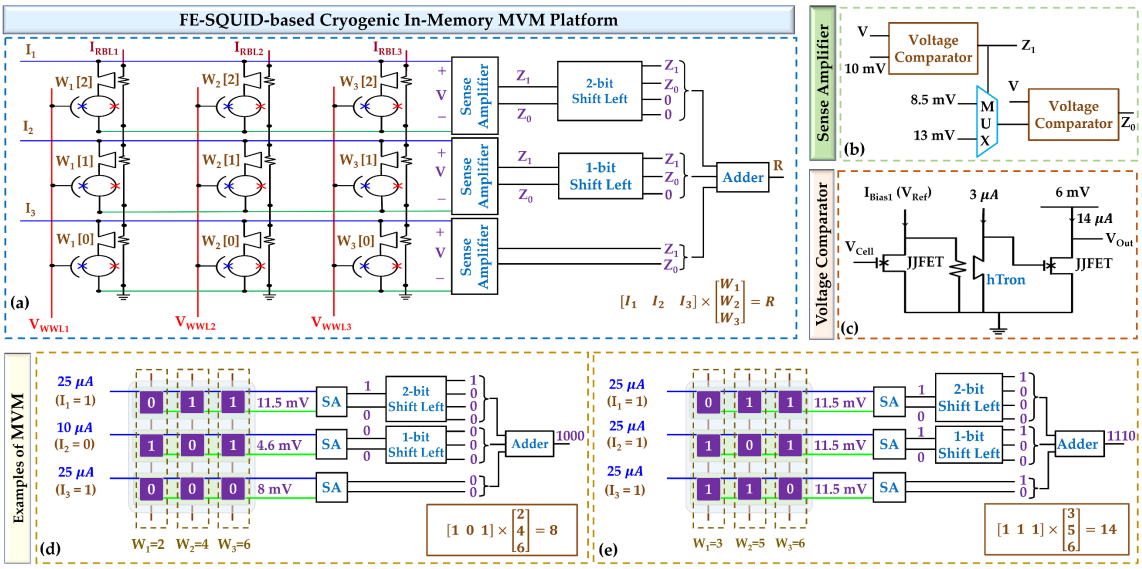


Figure 4.1. (a) FE-SQUID memory array with modified peripheral circuitry to perform MVM. Schematics of **(b)** sense amplifier and **(c)** voltage comparator. **(d, e)** Two examples of MVM operation.

4.3 Accelerating Classification Tasks With FE-SQUID-based MVM

To demonstrate the efficacy of FE-SQUID-based in-memory MVM, we develop a ML algorithm to utilize its advantages. Our model is a modified version of a multiclass perceptron [92] where the input and weights are discrete. To test this model, we use the handwritten digits dataset from the UCI Machine Learning Repository [93]. **Fig. 4.2**(c) illustrates how we utilize our FE-SQUID-based in-memory MVM for the handwritten digit recognition. This dataset contains 8×8 greyscale images of handwritten digits. For our model, the image is first flattened and treated as a vector of 64 values. To prepare the dataset for training, we encode the input in a way so that it can be used in the memory. We use read current ranging between 13 μA and 43 μA to map each element of the input matrix into 16 different levels, as shown in **Fig. 4.2**(d).

After preparing the input, we determine the possible weights that we can use in the model. Since our input has 64 values (mapped into 64 rows), we use 64 columns of the FE-SQUID-based memory. Each cell can show one of two resistances (0.95 $k\Omega$ and 1.75 $k\Omega$) based on the P_{FE} state. Next, we determine all the possible unique combinations

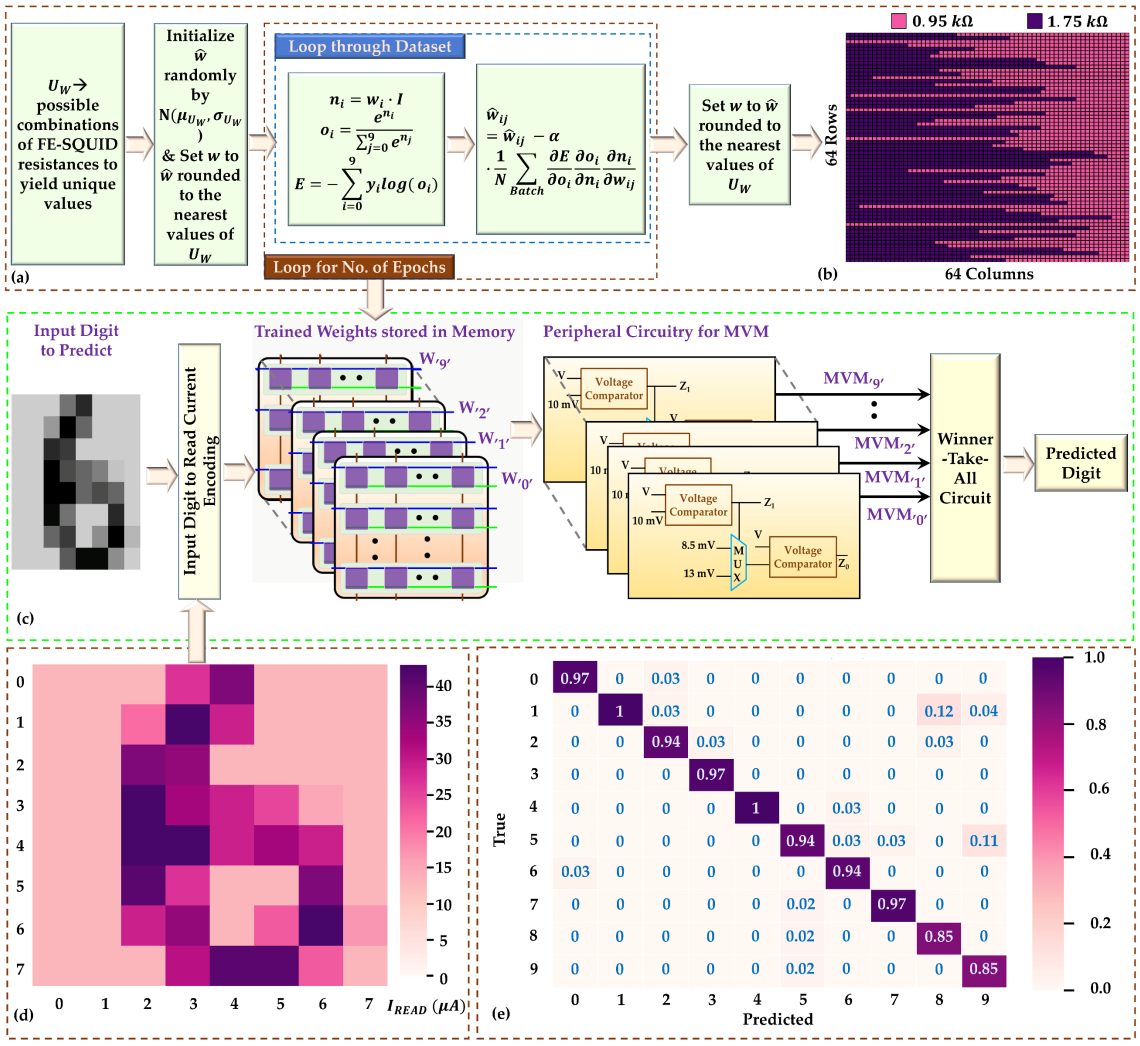


Figure 4.2. (a) Mechanism of training weights and **(b)** storing them as memory states in the array. **(c)** FE-SQUID-based ML model to perform handwritten digit recognition. **(d)** Mechanism of encoding input as the read current levels. **(e)** Confusion matrix for the predicted digits.

of FE-SQUID resistances and store them as the set U_W. For our approach, we define two sets of weights- (i) w for the weights to be stored in the memory array and (ii) $\widehat{w}$ for the weights used during training. $\widehat{w}$ is initialized to a normal distribution with the same mean and standard deviation as U_W. w is set equal to $\widehat{w}$ where each value is rounded to the nearest values in U_W. To train the model, we use stochastic gradient descent [94] modified for our memory array, as shown in **Fig. 4.2**(a). Essentially, this process iteratively optimizes the weights with respect to the error in the model which is performed by randomly selecting a batch of size N from the training data and feeding them through the model. Once this is done, we calculate the average gradient of the error with respect to the weights for the batch. Finally, we update the weights by subtracting the gradient from them multiplying with a user-defined learning rate, α. While calculating the gradients, we use the weights that can be mapped in the memory array (w) and apply the gradients to our training weights ($\widehat{w}$). Every time $\widehat{w}$ is updated, we update w as well. We continue to update the weights until we have gone through the entire training set, and we repeat this entire process for a defined number of epochs. Once training is complete, we convert the weights back to their corresponding combinations of FE-SQUID resistances such that they can be stored in the memory array, as shown in **Fig. 4.2**(a). Finally, the weights can be stored in the memory cells to perform inference with our model. **Fig. 4.2**(b) shows the weights for digit '6' recognition as mapped into the FE-SQUID array.

As previously discussed, we used 8×8 handwritten digit for the initial test of our model. After training, we achieve 93.89% accuracy on testing digits that were not used for training. **Fig. 4.2**(e) shows the confusion matrix of the recognition results, with only '8' and '9' being misclassified somewhat regularly, with 85% accuracy for each.

Next, we test how the matrix size of the input digits affect the recognition accuracy. For this, we use different matrix sizes (16×16 and 28×28) for input digits. The 28×28 dataset is the original MNIST dataset [89], while the 16×16 images are from the same dataset but with antialiasing [95]. We also test how lower number of weight columns affects the performance. This would decrease the possible levels of weights but with the benefit of less power consumption and delay. **Fig. 4.3**(a) shows the effects of input size

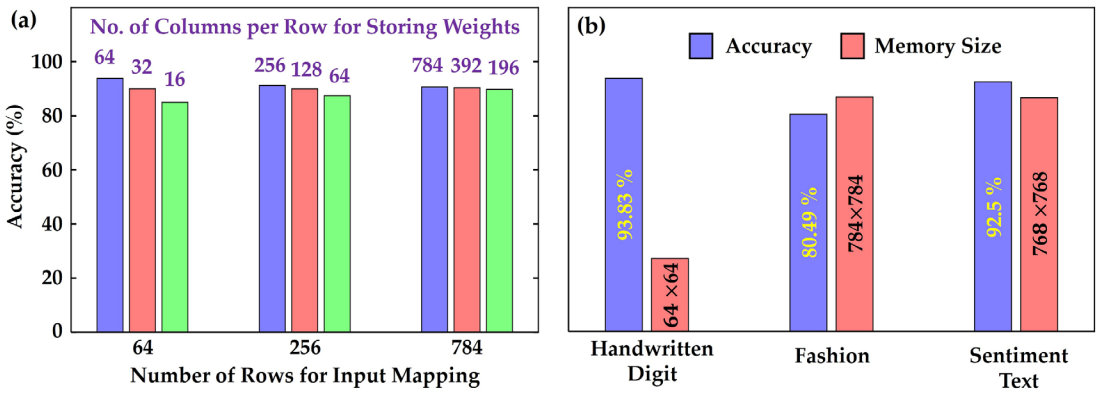


Figure 4.3. (a) Effects of number of rows and columns on the accuracy of handwritten digit recognition task. **(b)** Comparison of accuracy for the three classification tasks tested using our framework.

and the number of columns per row on the accuracy. As shown in **Fig. 4.3**(a), we achieve an average of 91.19% and 90.6% testing accuracy for 16×16 and 28×28 datasets, respectively. From **Fig. 4.3**(a), we also see that decreasing the number of columns has a much larger negative effect on accuracy for inputs with a smaller matrix size.

Also, we test our model on two more classification tasks. The first one is Fashion-MNIST dataset [90], which has 28×28 pixel images of ten classes of fashion items. This dataset is much more challenging than digit recognition while having the same input size. While using a 784×784 memory array, we achieve an accuracy of 80.49% on the testing set. Finally, we test our algorithm in another novel application- sentiment analysis. We use the emotion dataset [91] for training and testing. This dataset contains short text passages classified as one of six emotions- sadness, anger, fear, joy, love, or surprise. This problem is significantly more complex than the previous applications, so we choose to take a different approach to this problem. Instead of feeding the raw data into the model, we first use a feature extractor to obtain a consistently sized input for the memory array. For feature extraction, we use DistilBERT [96], and to obtain the features, we use the output of the last hidden layer from a pretrained DistilBERT model via HuggingFace [97]. Once the text is processed in this way, we get a vector of length 768 that represents the features of the text that we can map into the memory array as the row current. This

approach gives us 92.5% accuracy on the testing dataset. Note, for both these tasks, the approach shown in **Fig. 4.2**(a) is used for training. **Fig. 4.3**(b) summarizes the results obtained for the performed tasks.

4.4 Comparative Analysis

Finally, we compare our FE-SQUID-based approach with existing room-temperature memristor-based approaches. **Table 4.1** shows the comparison. For handwritten digit and fashion MNIST datasets, we compare our work with existing memristor-based neuromorphic approaches (handwritten digit- [98]–[100] and fashion- [100], [101]). For the sentiment analysis, we use two datasets (IMDB dataset [102] and emotion dataset [91]) and compare with CMOS-based [103] and memristive [102] approaches for IMDB dataset and with software-based deep learning approach [91] for emotion dataset. For all the classification tasks, our FE-SQUID-based approach outperforms the existing approaches in terms of accuracy.

4.5 Summary

This chapter discusses a technique to perform cryogenic in-memory MVM operation utilizing an FE-SQUID-based memory array. To test this novel approach, we perform three classification tasks (handwritten digits, fashion, and emotion datasets) using an experimentally calibrated simulation framework. Our FE-SQUID-based classification system achieves up to 17% better accuracy than existing room temperature memristor and CMOS-based approaches. Considering the typical operating speed of a SQUID (>100 GHz [53]), our FE-SQUID-based memory cells will consume less than 5 atto-Joules of energy (on average) for 1-bit MVM operation. The cooling cost is not a concern for our target application (QC systems), as the qubits are already cooled to a cryogenic temperature. FE-SQUID is fully compatible with the existing commercial process technology used for fabricating single-flux quantum (SFQ) circuits [104]. Besides the QC systems, our proposed in-memory computing technique can also be directly adopted for space electronics, and cryogenic high-performance computing systems [53].

Table 4.1: Comparison with Existing Approaches

Datasets	Existing Works			This Work
	Approach	**Accuracy (%)**	**Ref.**	**Accuracy (%)**
Handwritten digit [12]	Memristive	82	[27]	93.4
		89.4	[28]	
		85	[29]	
Fashion [13]	Memristive	69	[29]	80.5
		75.2	[30]	
IMDB Sentiment [31]	Memristive	84.3	[31]	86.8
	CMOS	81	[32]	
Emotion [14]	Software	81	[14]	92.5

CHAPTER 5
TERNARY CONTENT ADDRESSABLE MEMORY USING FERROELECTRIC SUPERCONDUCTING QUANTUM INTERFERENCE DEVICE

5.1 Introduction

A suitable cryogenic controller and memory system can facilitate the scaling of quantum computing systems up to thousands of qubits, by solving several existing challenges, including (i) the requirement of a large number of wires and interconnects to connect the qubits with currently used room temperature controller and memory, (ii) the possibility of a large amount of thermal noise propagation from room temperature to noise sensitive qubits, and (iii) the heat and noise generation by the lossy wires and interconnects [10], [23], [84]. Moreover, cryogenic systems based on superconducting devices, such as JJs and SQUIDs, provide unparalleled speed (hundreds of gigahertz) and energy efficiency (sub-atto-joule switching energy) [22]. This can be extremely useful for developing energy-efficient high-performance computing systems and space electronics.

One of the critical challenges in cryogenic systems is the need for high-speed, low-power memory solutions that can complement the computing infrastructure [23]. In this quest, superconducting, non-superconducting, and hybrid technologies have been explored. However, all of these technologies have unique challenges, and therefore, developing a suitable and scalable memory system for cryogenic applications has remained an ongoing pursuit. Content addressable memory (CAM) and ternary content addressable memory (TCAM) are specialized storage systems that can accelerate certain computing tasks by introducing the capability of comparing input data against stored data [105], [106]. CAM performs searches based on binary data ('0' and '1') returning exact match/mismatch output, where TCAM allows an additional state (don't care, 'd') for the input data (**Figs. 5.1**(a) and (b)). The ternary nature of TCAMs allow for more flexible searches, which is useful in several applications including pattern recognition, network routing, and database management. Additionally, CAM and TCAM can be empowered to calculate the Hamming distance (HD) between the input search and stored data. This capability allows us to identify nearby matches (the lowest HD implies the closest match) along with the exact matching (HD = 0). CAMs and TCAMs with HD calculation

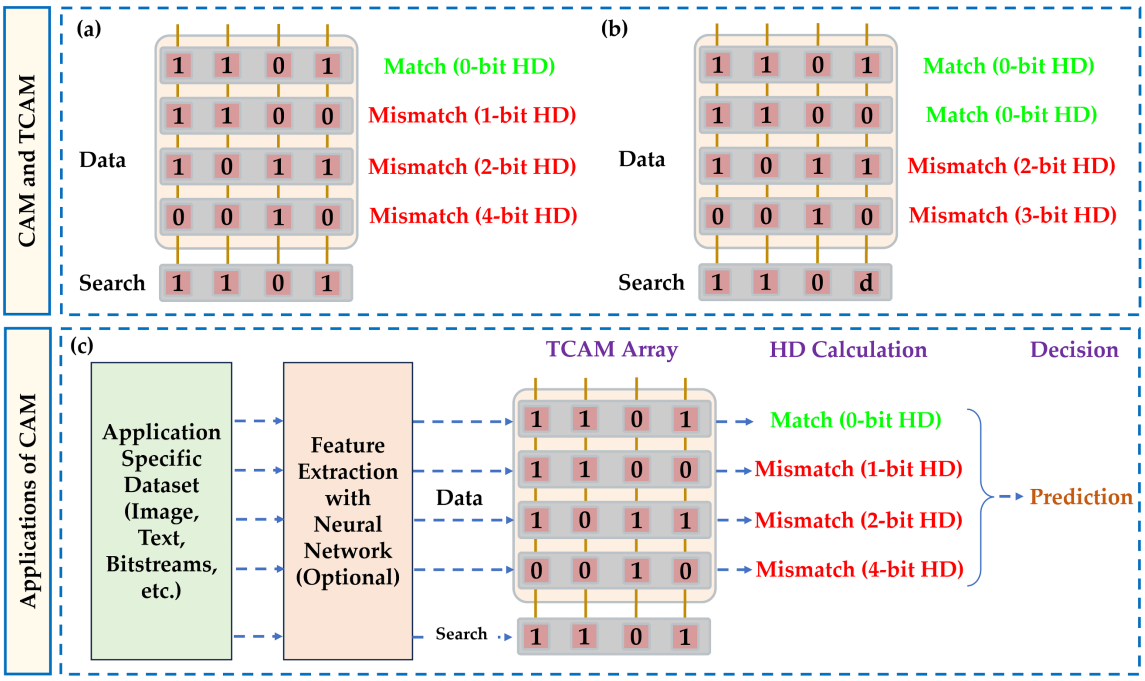


Figure 5.1. Introduction to content addressable memory. Illustrations of the search mechanism with a **(a)** binary and **(b)** ternary CAM. **(c)** Block diagram showing the steps how a CAM can be used for AI applications such as classification and pattern recognition.

capability can be powerful tools in error detection and correction for both classical and quantum domains, approximate searching, and other data-intensive applications. Additionally, TCAMs have recently been used to accelerate classification tasks, where the TCAM stores the features extracted from the dataset (images, texts, etc.) as the stored data and can provide the search result whenever the features for any search condition are applied as the input data (**Fig. 5.1**(c)) [106]. Based on the search results, TCAM can predict the classification result. Here, the parallel search capability of TCAMs becomes extremely useful for improving the speed and efficiency of the classification tasks.

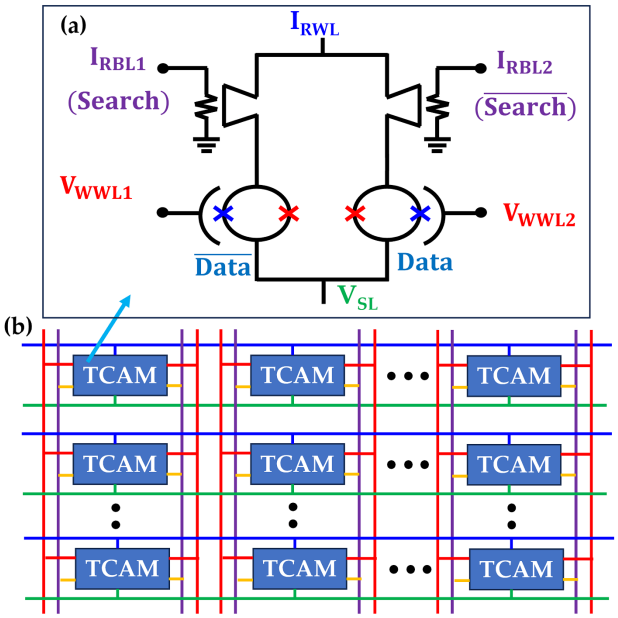
TCAMs can also be useful for cryogenic applications like quantum computing by efficiently handling quantum error correction decoding, high-performance computing by improving latency and energy efficiency, artificial intelligence (AI) by accelerating pattern recognition and classification, space exploration by assisting in data storage and retrieval, and decision-making algorithms, and data centers by improving large-scale search operations like database queries and search engines. This chapter presents the design and implementation of a cryogenic TCAM with exact search and HD calculation capabilities. Here, we utilize the FE-SQUID-based cryogenic memory system that combines the ultra-high speed and energy efficiency of SQUIDs with the voltage-controlled non-volatility of ferroelectric materials [42]. In the proposed TCAM, there are two modes- one for exact searching and another for HD calculation. Due to the use of FE-SQUID, we get several advantages from the proposed TCAM when compared with the existing designs (both room temperature and cryogenic), including (i) the exact searching mode provides binary decisions for matching and mismatching without needing any peripheral circuitry, thanks to the superconducting behavior of SQUIDs, and (ii) extremely energy-efficient TCAM operation (1.36 aJ and 26.5 aJ average energy consumption for 1-bit CAM and TCAM search. respectively). Finally, we demonstrate the efficiency and advantages of our proposed TCAM by performing hyperdimensional computing (HDC) algorithm-based language recognition task. By implementing the associative memory that supports a highly parallelized search operation for the inference step with FE-SQUID-based TCAM arrays, a single 10,000-bit vector comparison consumes just 89.4 fJ energy. We also demonstrate that the FE-SQUID-based TCAM

consumes over one order of magnitude less energy compared to the cryogenic 5 nm FinFET SRAM-based TCAM to perform the same operation.

5.2 Design Principle of FE-SQUID-based TCAM

Fig. 5.2(a) shows the schematic of the proposed cryogenic TCAM. The proposed TCAM is a modified version of FeSQUID-based memory cell, demonstrated in [42]. In the FE-SQUID-based memory cell, to store one bit of data, one FE-SQUID and one hTron are connected in series and multiple cells are connected in parallel along each row of the memory array. However, here, we use two memory cells connected in parallel to each other to design a 1-bit TCAM. The operation of this TCAM cell is controlled by one read word line (RWL), two read bit lines (RBLs), two write word lines (WWLs), and one sense line (SL). Now, to build a large-scale TCAM, these TCAM cells need to be connected in parallel along the row where the shared RWL in each row will act as the match line (ML) to exhibit the matching/mismatching result (**Fig. 5.2**(b)).

The proposed TCAM has two modes- one for exact search (binary result) and another for HD calculation (analog result). The mode of the TCAM can be selected by choosing a suitable bias current for the RWL of each TCAM. For exact search mode, the value of RWL current (I_{RWL}) needs to be chosen in a way so that we get superconducting and resistive states from SQUIDS for '0' and '1' data, respectively. According to **Fig. 5.2**(c), the range for this I_{RWL} is $I_{C,low} < I_{RWL} < I_{C,high}$. Due to the use of the superconducting state of SQUID and the parallel connection of all the TCAMs in a row (as shown in **Fig. 5.2**(b)), if there is any SQUID with a superconducting state, the ML of that row will have zero voltage. This provides us with the binary decision from the ML- either zero or some non-zero voltage for any amount of mismatch or complete matching, respectively. Now, for HD calculation mode, we do not want binary decisions for matching and mismatching, rather we want analog decisions with different levels of ML voltage for different levels of matching between the input and stored data. Therefore, we need to avoid the superconducting state of SQUIDs and that's where the use of FE-SQUIDs becomes extremely useful because FE-SQUIDs can show either superconducting/resistive states or only resistive state with two different resistance values for two states. In the HD calculation mode, we choose the value of I_{RWL} in a way so that



(c) Modes	I_{RWL}
Exact Search	3.2 μA
HD Calculation	5 μA

(d) Search Bit Definition		
Search	I_{RBL1}	I_{RBL2}
0	0 μA	10 μA
1	10 μA	0 μA
d	10 μA	10 μA

(e) Data Bit Definition		
Data	Device State	
	Exact Search	HD Calculation
0	SC (0)	Non-SC (0.9 $k\Omega$)
1	Non-SC (1.9 $k\Omega$)	Non-SC (1.9 $k\Omega$)

Figure 5.2. The proposed cryogenic TCAM based on FE-SQUID. **(a)** Circuit schematic of the proposed TCAM cell. **(b)** Illustration of the array-level organization of the proposed TCAM cells, where data will be stored in a row. **(c)** Values of RWL currents for two modes of the proposed TCAM. Definitions of the representation of **(d)** input search data in terms of RBL currents and **(e)** stored data in terms of FE-SQUID's device states.

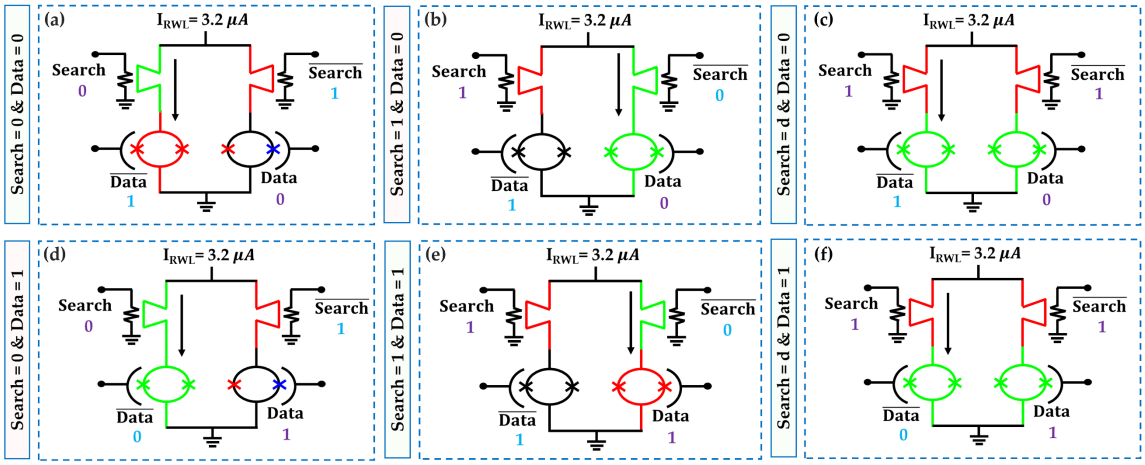
FE-SQUIDs show two different resistance values for '0' and '1' states. According to **Fig. 5.2**(c), the required range of I_{RWL} for this mode is - $I_{RWL} > I_{C,high}$. **Fig. 5.2**(c) shows the values of I_{RWL} used for two TCAM modes in this work.

The FE-SQUIDs are utilized to store the data as the ferroelectric polarization (similar to the memory operation) for the TCAM operations. One FE-SQUID stores the data (*Data*) while the other stores the inverted version of data ($\overline{Data}$). To store *Data* and $\overline{Data}$, we need to apply suitable voltages across the ferroelectric material of FE-SQUIDs through the WWLs (V_{WWL1} and V_{WWL2}). **Fig. 5.2**(e) shows the definition of *Data* and the corresponding device states. The input search data is applied through a combination of two gate currents (RBL currents, I_{RBL1} and I_{RBL2}) of the two hTrons, which determine the switching of the hTrons. **Fig. 5.2**(d) shows the values of RBL currents for different input search data (*Search* and $\overline{Search}$).

5.3 Working Principle of FE-SQUID-based TCAM

Any TCAM search can be divided into two major operations- (i) storing the data inside the memory through memory write operation and (ii) searching any input data against the stored data through memory read operation. To write data into a FE-SQUID-based memory cell, we follow the same mechanism explained in Chapter 2. As illustrated in **Fig. 2.1**(c), applying a current within the range $I_{C,low} < I < I_{C,high}$ to the SQUID will result in either a superconducting (0 V) state for P_R^- or a resistive (nonzero voltage) state for P_R^+, depending on the polarization state of the ferroelectric. This property is exploited for the exact search mode by applying a suitable current bias through the SQUID. Moreover, as seen in **Fig. 2.1**(c), if a current larger than $I_{C,high}$ is applied, the SQUID will show two different resistance values (and hence, two voltages) depending on the polarization states. This characteristic of SQUID is used for the HD calculation mode. RBL currents are responsible for determining the switching of hTrons connected in serial with each FeSQUID. The search data are represented with the RBL currents.

Based on the application of RWL and RBL currents, the ML voltage will depend on whether the input search matches or mismatches the stored data. **Fig. 5.3** shows the working principle of the exact search mode of the proposed TCAM. Here, first, *Data* and $\overline{Data}$ are stored in the two FE-SQUIDs with the help of WWLs. Then, when *Search* and $\overline{Search}$ are applied through the RBL currents (chosen according to **Fig. 5.2**(d)), and the suitable current is applied to the RWL (chosen according to **Fig. 5.2**(c)), one of the hTrons become superconducting while the other remains in a highly resistive state ($50\ k\Omega$). For matching conditions ($Data = 0$ and $Search = 0$ (**Fig. 5.3**(a)), $Data = 1$ and $Search = 1$ (**Fig. 5.3**(e)), the FE-SQUID connected in series with the superconducting hTron switches to its resistive state. As a result, the RWL current flows through this branch and we get a non-zero voltage drop in the ML (shown in **Fig. 5.3**(g)). On the other hand, for the mismatching conditions ($Data = 0$ and $Search = 1$(**Fig. 5.3**(b)), $Data = 1$ and $Search = 0$ (**Fig. 5.3**(d))), the FE-SQUID connected in series with the superconducting hTron remains in its superconducting state. As a result, we get zero voltage drop in the ML (shown in **Fig. 5.3**(g)).



(g) Search	Data	ML Voltage	Decision	Power	Energy
0	0	2.83 mV	Match	9.05 nW	2.71 aJ
0	1	0 V	Mismatch	0 W	0
1	0	0 V	Mismatch	0 W	0
1	1	2.83 mV	Match	9.05 nW	2.71 aJ
d	0	80 mV	Match	256 nW	76.8 aJ
d	1	80 mV	Match	256 nW	76.8 aJ

Figure 5.3. Working Principle of the proposed TCAM in the exact search mode. Illustration of the working principle of the TCAM when **(a)** $Data = 0$ and $Search = 0$, **(b)** $Data = 0$ and $Search = 1$, **(c)** $Data = 0$ and $Search = d$, **(d)** $Data = 1$ and $Search = 0$, **(e)** $Data = 1$ and $Search = 1$, and **(f)** $Data = 1$ and $Search = d$. **(g)** ML voltage, power, and energy consumption for 1-bit exact search with the proposed TCAM.

Now, for the 'd' case in the input search, we apply non-zero currents to both RBLs. As a result, in these cases shown in **Fig. 5.3**(c) and (f), both the hTrons switch to their resistive states which leads to the division of the applied RWL current between two branches. The current that flows through each of the branches ($I_{RWL}/2$) is not sufficient to switch the state of the FE-SQUID to its resistive state. However, since both the hTrons become resistive, we get a non-zero voltage drop in the ML(shown in **Fig. 5.3**(g)), irrespective of the stored data in the FeSQUIDs. **Fig. 5.3**(g) shows value of ML voltages, power consumption, and energy consumption for all possible combinations of a 1-bit exact search operation in the proposed TCAM.

Now, we discuss another capability of our proposed TCAM, which is the HD calculation. **Fig. 5.4**(a) shows a 4-bit TCAM where five different combinations of 4-bit data are stored in the TCAM array, and one specific input search data is applied. As explained earlier in this section and **Fig. 5.3**, in the exact search mode, due to the choice of RWL current, we will get zero and non-zero ML voltages for any mismatch and complete match, respectively. However, in the HD calculation mode, we select the RWL current in a way so that the SQUID always gets a current that is larger than its both critical currents ($I_{C,low}$ and $I_{C,high}$). Therefore, the SQUID never remains in the superconducting state. For the two states, we get two different values of resistance, as shown in **Fig. 2.1**(c). We design and choose the operating modes of the proposed TCAM in a way so that we get the highest ML voltage for the complete matching condition and the voltage drops with the increase in the mismatch. **Fig. 5.4** (a) shows the ML voltages for this mode for different amounts of mismatch in the 4-bit TCAM. We also extend our simulation to larger TCAM arrays and the values of ML voltage for different amounts of mismatch in different sizes of TCAM array in **Fig. 5.4** (b). The ML voltage for n-bit TCAM can be calculated using the following equation:

$$V_{ML} = (n_{bits} \times I_{RWL}) \times \frac{1}{n_{bits} \times \frac{1}{50\ k\Omega} + n_{match} \times \frac{1}{1.8\ k\Omega} + (n_{bits} - n_{match}) \times \frac{1}{0.9\ k\Omega}}$$

(a)

TCAM Modes→				Exact Search		HD Calculation	
TCAM Array (Data Bits)				ML Voltage	Decision	ML Voltage	Decision (HD)
1	1	0	0	1.52 mV	Match	9.15 mV	0
1	0	0	0	0 mV	Mismatch	7.22 mV	1
1	0	1	0	0 mV	Mismatch	5.96 mV	2
0	0	1	0	0 mV	Mismatch	5.08 mV	3
0	0	1	1	0 mV	Mismatch	4.42 mV	4
1	1	0	0				
Search Bits							

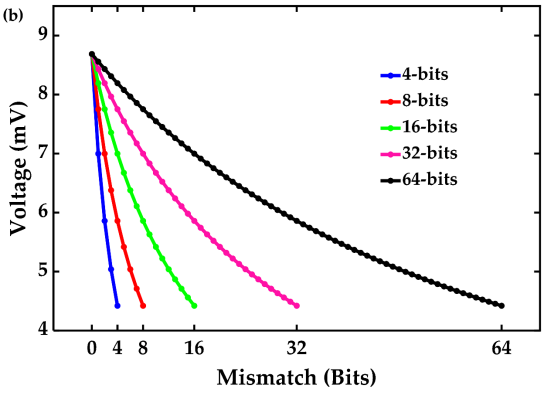


Figure 5.4. ML voltage levels obtained for different number of data lengths and different amounts of mismatch between the input search and stored data.

where, n_{bits} and n_{match} represent the size of TCAM cells in a row and the number of matching bits in that row, respectively. To decode the HD from the ML voltage levels, any cryogenic voltage-based comparator and sense amplifier can be utilized [75].

5.4 System-level Benchmarking

With the ever-growing interest in machine learning, the size and data requirements of today's models are enormous. However, many application scenarios have strict constraints for the system, e.g., in cryogenic environments, embedded systems, or IoT devices. Here, efficient and computationally lightweight algorithms are key. For these reasons, researchers have explored techniques to simplify models by quantizing conventional machine-learning approaches. While reducing the model size, the fundamental requirements of large datasets and large training efforts are unchanged. For these reasons, an emerging brain-inspired machine-learning concept called HDC has been proposed and gained rapid interest. In HDC, information is stored as patterns in large vectors with thousands of elements. By utilizing randomness and similarity, HDC is inherently resilient to noisy data and errors in computing operations. This enables the utilization of emerging hardware, which often may suffer from imprecise operation, e.g., due to variability. HDC has been showcased in various classical machine-learning tasks such as language recognition [107] (illustrated in **Fig. 5.5**(a)), image classification [108], or wafer map defect pattern classification [109].

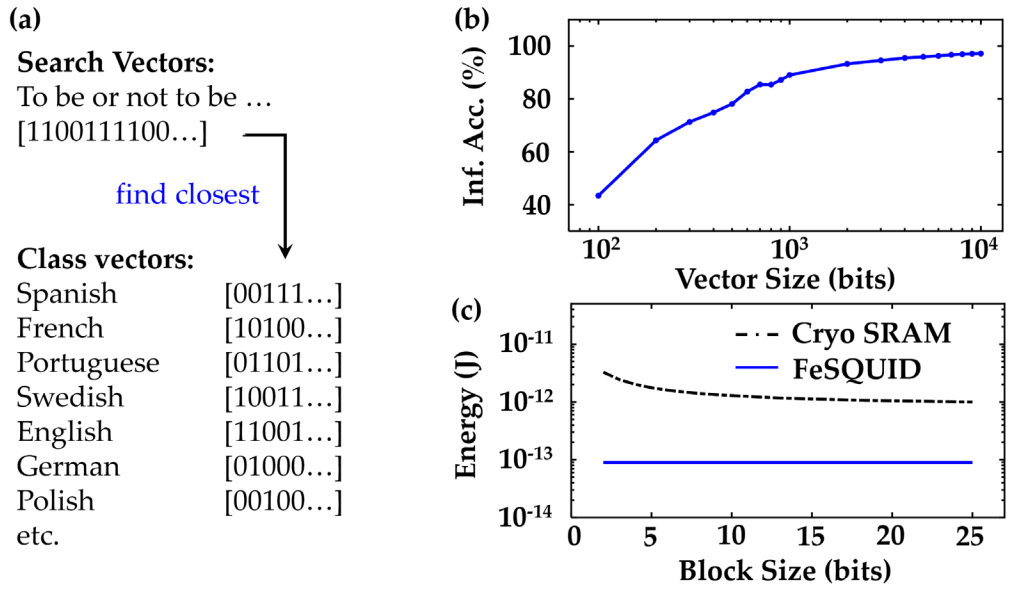


Figure 5.5. (a) Illustration of the HDC algorithm for language recognition. The encoded search vector of the sample text that needs to be recognized is compared to all pre-trained class vectors. The closest vector is then inferred as the result, and the respective class label is used to answer the search. **(b)** The inference accuracy of language recognition over the vector size. **(c)** Energy consumption of a single vector comparison with 10,000 bit using FE-SQUID-based and SRAM-based TCAM cells. Using the discussed energy estimation scheme, the FE-SQUID energy is constant with respect to block size.

The data type of vector elements can be selected, and real numbers, integer numbers, bipolar values, or binary values can be used. Using binary makes HDC very efficient as the operations can be implemented using basic logic gates. For the similarity metric used during inference, the Hamming distance needs to be calculated, which is a combination of bitwise XOR popcount. This can be implemented using TCAM arrays that store the class vector bits and are searched with the bits of the query vector [110]. The used vector size affects the precision of the final model, which is shown in **Fig. 5.5**(b). For instance, training the language recognition model with 10,000 bits yields an inference accuracy of 97.1% while using half the size yields 95.9%. Estimating the used energy consumption of a single vector comparison can be done using the following equation:

$$E = n_{bits} \times 5\mu A \times V_{ML} \times 0.3ns$$

where n_{bits} is the size of vectors and V_{ML} is the match line voltage depending on the amount of match. Here, we consider the switching time of hTron as the only delay during the search operation in the TCAM cell. We use *0.3 ns* for the superconducting to resistive switching of hTron devices, as reported in [66].

As most of the vectors in HDC are orthogonal (i.e., a similarity of around 50%), we assume the number of matching bits to be half of the total number of bits. This gives a energy consumption of 89.4 fJ per vector comparison with 10,000 bits, and 44.7 fJ with 5,000 bits. To put the numbers in perspective, we have implemented a block of TCAM cells using conventional SRAM technology with 5 nm FinFETs. These SRAMs are also operated at 4K temperature, like FE-SQUID-based TCAMs. By cutting the large vectors into smaller sections, blocks are formed that calculate the HD in the analog domain. Hence, the results in **Fig. 5.5**(c) show a decreasing energy consumption for SRAM-based cells. As fewer blocks are needed when increasing the block size, the comparison of an entire vector becomes more efficient. In our energy estimation scheme for FE-SQUID-based TCAM, the energy depends on V_{ML}, which is only subject to the ratio between block size to the number of mismatches, which is fixed to 50% here. Still, the energy consumption of FE-SQUID-based TCAMs is over one order of magnitude smaller than cryo SRAM-based ones. For instance, using vectors with 10,000 bits and a block size of

10 bits, SRAM-based TCAM consumes 1.29 pJ compared to the 89.4 fJ consumed by FE-SQUID-based design.

5.5 Summary

In this Chapter, we discuss a cryogenic TCAM architecture leveraging the unique properties of FE-SQUID technology. The proposed TCAM addresses key challenges faced by current cryogenic computing systems, such as the need for ultra-low power consumption, high-speed operation, and scalability. By combining the non-volatility of ferroelectric materials with the superconducting efficiency of SQUIDs, our design achieves substantial energy savings, with an average energy consumption of 1.36 aJ and 26.5 aJ for 1-bit binary and ternary searches, respectively. In addition to supporting exact match search operations, our TCAM enables Hamming distance (HD) calculation, which opens up new possibilities in error detection and correction, quantum error correction decoding, approximate search algorithms, and AI-based classification tasks. We demonstrate the advantages of FE-SQUID-based TCAM array by benchmarking against TCAM with 5 nm FinFET-based cryogenic SRAM array for HDC-based language recognition task. The proposed FE-SQUID-based TCAM needs over one order of magnitude lower energy. The parallel search capability and efficient data processing of this energy-efficient TCAM make it highly applicable to a range of cryogenic applications, including large-scale quantum computing, high-performance computing, space exploration, and data-intensive tasks in data centers.

CHAPTER 6
RECONFIGURABLE LOGIC CIRCUIT USING SUPERCONDUCTING HEATER CRYOTRON

6.1 Introduction

JJ-based logic families are the primary building blocks of superconducting Boolean logic. Thanks to the ultra-fast (>100 GHz) and ultra-low power (sub-aJ/bit switching energy) operation of JJ-based circuits [64], [111], a superconducting processor has the potential to solve many existing issues of its CMOS counterpart. However, JJ-based circuits suffer from a number of challenges including difficulty in cascading, fabrication challenges, limited scalability due to flux trapping and sensitivity to magnetic field, and so on [79], [112]. To solve some of these challenges, three-terminal cryotron devices were developed which show input gate current-driven switching of the channel between its superconducting and resistive states [79]. *hTron* [66] is a member of this cryotron family. These devices can drive large impedances ($> 100\ k\Omega$) and support a large number of fanouts due to their transition to a highly resistive state. Moreover, these devices do not require superconducting loops like JJs and hence, are free from flux trapping and resulting scalability issues [79]. *hTron* devices have already been used as an access device in cryogenic memories [42], [51], [113], as an interface between superconductors and semiconductors [66], to design logic circuits [75], [78], [79], [112], in cryogenic neuromorphic systems [114]–[117], and so on.

Despite being used in several sensitive and critical applications, the hardware security techniques for superconducting processors are yet to be developed. Therefore, the superconducting processors are alarmingly vulnerable to different adversarial attacks including integrated circuit (IC) counterfeiting, IC masking, IC overproduction, intellectual property (IP) piracy, reverse engineering, etc. To prevent these attacks, different hardware security measures including IC camouflaging, logic locking, and/or adding a watermark, among others are used. Among these, logic locking [118] and IC camouflaging [119] have recently been utilized for developing secured superconducting hardware. In the logic locking technique, additional gates (AND, OR, XOR, etc.), key inputs, and additional on-chip memory are introduced into the original design to prevent

external attacks. However, as reported in [118], the logic locking technique requires an additional 20% area overhead for only one OR gate. On the other hand, in IC camouflaging technique in [119], dummy JJs are introduced to make the layouts of all the gates look identical to prevent reverse engineering and other attacks. Both of these techniques have a significant negative impact on the area, delay, and power consumption of the circuit.

In this chapter, we present a superconducting reconfigurable logic circuit with a single *hTron* device that can perform the basic Boolean logic operations, including 1–input copy and NOT, 2–input AND and OR, and 3–input majority. Typically, the number of devices required to perform any logic operation increases with the number of inputs [79], [112]. However, in our proposed circuit, we perform all the above-mentioned single- and multi-input logic operations with only one *hTron* device. We apply different biasing conditions in the same circuit to perform all these logic operations. Thanks to the reconfigurability of the proposed circuit, it requires significantly less area compared to the existing superconducting logic circuits based on JJs, SQUIDs, and cryotrons. Moreover, as we are using the same circuit with the same layout to perform different tasks, we can build a camouflaged system (without compromising on any performance metric) where all the logic gates will look identical which can improve the hardware security of the system against reverse engineering attacks.

6.2 Device Fabrication and Characterization

hTron is a four-terminal current-driven superconducting device where two of the terminals form the gate and the other two form the superconducting channel (**Figs. 6.1** (a)–(c)). The devices described here were fabricated using a combination of electron beam lithography and optical lithography. First, a layer of superconducting tungsten silicide (WSi) with a thickness of 4 nm was deposited via magnetron sputtering on a silicon wafer. Contact pads of 5 nm Ti followed by 30 nm Au were deposited using a liftoff process and optical lithography. The WSi was then patterned using electron beam lithography and etched in an RIE with an SF_6 chemistry to form the channel of the device. For the dielectric spacer layer that separates the heater from the channel, 25 nm of SiO_2 was deposited via sputtering over the entire wafer. After this, another WSi layer

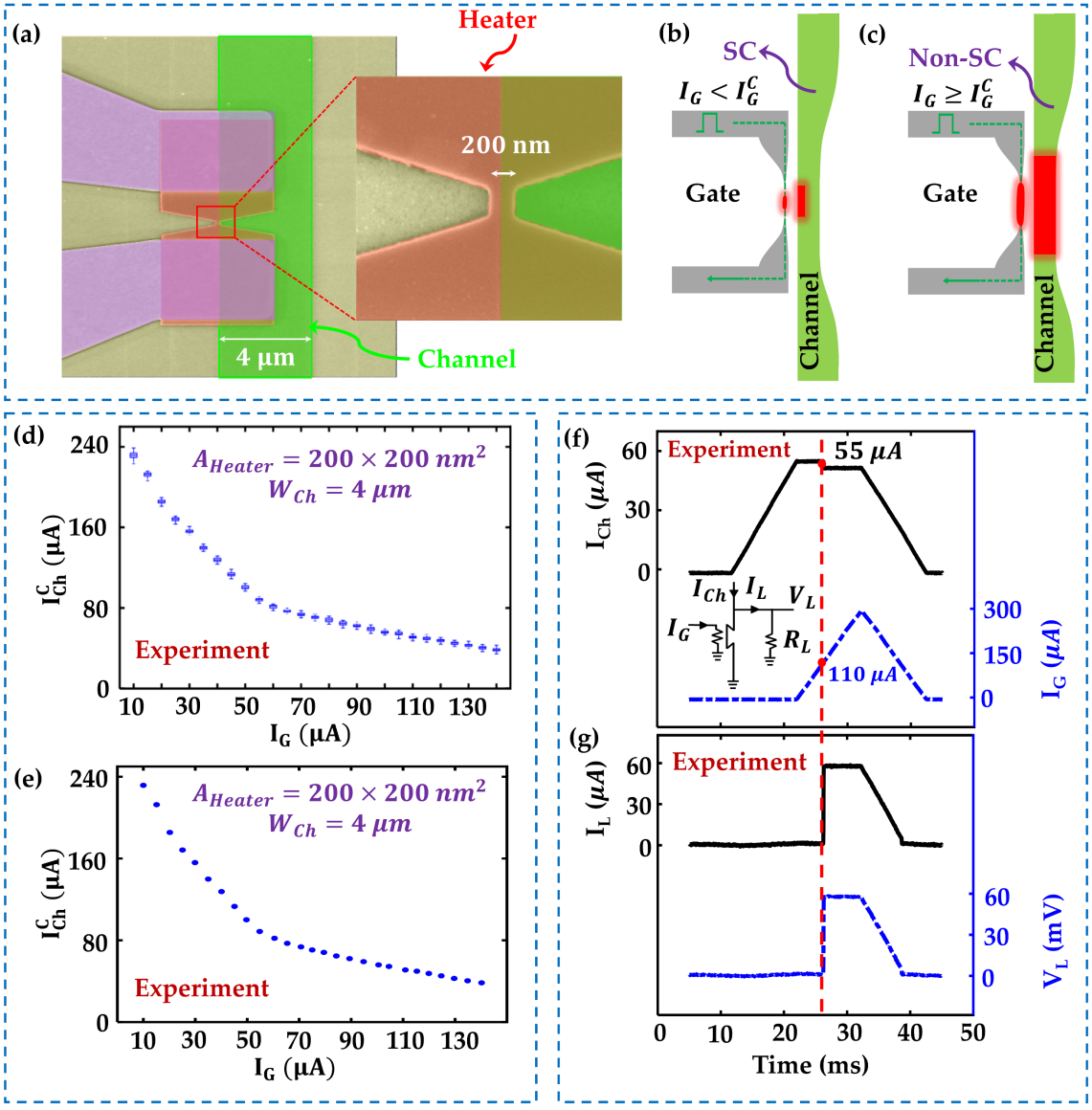


Figure 6.1: (a) False-colored scanning electron micrograph (SEM) of a fabricated hTron device. The fabricated device has a heater gate with an area of $200 \times 200\ nm^2$ and a channel with a width of 4 μm. **(b), (c)** Illustration of the switching mechanism of a hTron device. **(d)** Measured distributions of the channel critical current (I_{Ch}^C) for different values of I_G. **(e)** Extracted median values of I_{Ch}^C distributions for each I_G. **(f),(g)** Measured time dynamics of the switching of a hTron device with a load resistance (R_L) of 1 $k\Omega$.

with the same thickness of 4 nm was deposited via sputtering and once again patterned and etched to form the heater input gate for the device. Contact pads were added to this upper layer with added with the same liftoff process as the first layer. Lastly, openings were etched through the SiO_2 dielectric using an RIE with a CHF_3:O_2 chemistry so that electrical contact could be made with the bottom (channel) layer.

Initially, the channel remains superconducting for a given channel bias current (I_{Ch}) and no gate current (I_G). However, when I_G is applied, the gate becomes resistive and generates thermal phonons which are carried to the channel by the dielectric spacer. The channel remains superconducting until I_G exceeds a specific threshold and becomes resistive (**Fig. 6.1**(b)). However, when the gate becomes resistive, the presence of these thermal phonons suppresses superconductivity in the channel, and the more gate current is added, the greater the suppression of I_{Ch}^C is observed (as shown in **Fig. 6.1**(d)). The reason of suppressing the superconductivity is that the thermally-generated phonons are capable of breaking Cooper pairs in the superconducting channel (**Fig. 6.1**(c)). When I_G exceeds a specific threshold (I_G^C) such that the channel critical current (I_{Ch}^C) (a function of I_G, shown in **Fig. 6.1**(d)) becomes smaller than the applied I_{Ch}, enough phonons with sufficient energy ($> 2\Delta$, where Δ is the superconducting energy gap [63]) are generated which can break Cooper pairs in the channel and cause the entire channel to switch to the high impedance resistive state, driving I_{Ch} to the external circuitry (**Fig. 6.1**(c)).

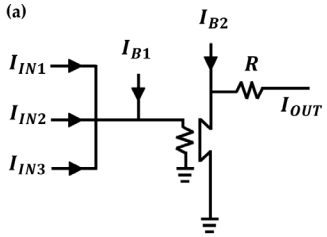
While characterizing the device, we first apply a fixed current to the gate (I_G). Then we ramp up the channel bias current (I_{Ch}) from 0 A to above I_{Ch}^C so that the channel switches. For each I_G, we perform the ramping of I_{Ch} 50 times and record the I_{Ch}^C for each measurement. **Fig. 6.1**(d) shows the measured distribution of I_{Ch}^C for I_G ranging from $10\ \mu A$ to $135\ \mu A$. We also extract the median of I_{Ch}^C for each I_G (**Fig. 6.1**(e)) and use that as the nominal I_{Ch}^C in our circuit design. Next, we perform a transient measurement to show the time dynamics of the channel switching from its superconducting to resistive state. The inset of **Fig. 6.1**(f) shows the transient measurement setup with a load resistance of $1\ k\Omega$. Here, we first ramp I_{Ch} up to $55\ \mu A$ and then start ramping up I_G to capture the gate-driven switching (**Fig. 6.1**(f)). As shown in **Fig. 6.1**(g), initially the

channel is in its superconducting state and hence, no current flows through the load resistor ($I_L = V_L \approx 0$). However, when I_G exceeds I_G^C ($110\ \mu A$ for this case), the channel switches to its resistive state and drives I_{Ch} to the load resistor which creates a nonzero voltage across it. The turn-on and turn-off times for the *hTron* device considered in this work are 300 ps and 15 ns, respectively. However, the turn-on and turn-off time can be reduced by using higher input energy and different superconducting materials, respectively [66]. The WSi film used as the superconducting material had a critical temperature (T_C) of 3.5 K and all the measurements were performed at 0.9 K. To obtain a higher operating temperature, a different material such as NbN or NbTiN can be used which show a T_C of 9-12 K depending on the stoichiometry.

6.3 Design Methodology of Reconfigurable Logic

To design a reconfigurable logic circuit for achieving secured superconducting hardware, we utilize the gate current-controlled superconducting to resistive transition of *hTron*. We use the same measurement setup (shown in the inset of **Fig. 6.1**(f)) and measured device characteristics (**Fig. 6.1**(e)) to design the logic circuit. The only difference is that instead of one input current to the gate, we use three input currents to perform up to three input logic operations. **Fig. 6.2**(a) shows the schematic of the designed reconfigurable logic circuit where I_{IN1}, I_{IN2}, and I_{IN3} are the input currents corresponding to the number of inputs of the logic operations, I_{B1} is a common gate bias used for all the operations, and I_{B2} is the channel bias. In our design, we define the two logic states ('0' and '1') with two current levels (0 μA and $I_G^C/2$, respectively) (**Fig. 6.2**(b)). **Fig. 6.2**(c) shows the biasing scheme that we use to perform different logic functions with the same circuit shown in **Fig. 6.2**(a). For all the logic operations, I_{B2} is kept fixed at 55 μA (chosen based on the measured device data shown in **Figs. 6.1**(d)–(g)). Now, to perform different logic functions such as 1–input copy and NOT, 2–input AND and OR, and 3–input majority, we only need to apply different bias currents to I_{B1} (range is shown in **Fig. 6.2**(c)). The combination of the input currents and I_{B1} will determine the switching of the channel under the fixed I_{B2} bias and accordingly, we will get zero or nonzero current at the output.

(a)



(b) Logic Definition	
Logic State	I_{IN} and I_{OUT}
'0'	0 ($0\ \mu A$)
'1'	$I_G^C/2$ ($55\ \mu A$)

(c) Reconfigurable Boolean Logic Functions			
I_{B1}	I_{B2}	Logic Function	
$I_G^C/2 < I_{B1} < I_G^C$	$55\ \mu A$	Copy	1-Input
$I_{B1} < -I_G^C$	$55\ \mu A$	NOT	
$0 < I_{B1} < I_G^C/2$	$55\ \mu A$	AND	2-Input
$I_G^C/2 < I_{B1} < I_G^C$	$55\ \mu A$	OR	
$0 < I_{B1} < I_G^C/2$	$55\ \mu A$	Majority	3-input

Figure 6.2: (a) Schematic of the hTron-based reconfigurable logic circuit that can perform 1-input copy and NOT, 2-input AND and OR, and 3-input majority operations with only a suitable bias current (I_{B1}). We use $1\ k\Omega$ as the load resistor (R). **(b)** Definition of logic states. **(c)** Biasing conditions to perform different logic operations with the same circuit.

6.4 Working Principle and Functional Verification

To demonstrate the functionality of different logic functions with our proposed circuit, we adopt a simulation–based approach. We first develop a LUT-based compact model for *hTron* in Verilog–A and calibrate the model with the measured data shown in **Figs. 6.1**(d)–(g). In each time step, the model takes applied gate and channel currents as input and based on the value of the gate current, reads the corresponding I_{Ch}^C from the LUT. Finally, the model compares the applied channel current with the I_{Ch}^C obtained from the LUT and determines whether the channel will switch or not. As already mentioned, our proposed circuit is similar to the measurement setup. The only difference is the use of multiple input currents applied to the gate. Note, all the input currents and I_{B1} are added together and seen as one bias current by the gate. Therefore, utilizing the experimentally calibrated device model, our simulation replicates the behavior that we would observe in the experiment.

First, we discuss the 1–input copy gate. Here, we use the $IN1$ terminal to apply the input current and we apply $65\ \mu A$ as I_{B1} and $55\ \mu A$ as I_{B2}. As shown in **Fig. 6.3**(a), for a logic '0' at the input ($I_{IN1} = 0$), the total applied current to the gate is I_{B1} ($65\ \mu A$) which is less than I_G^C ($110\ \mu A$) and hence, the channel remains superconducting. Therefore, I_{B2} flows through the channel and we get logic '0' ($I_{OUT} = 0$) at the output. On the other hand, as shown in **Fig. 6.3**(b), when a logic '1' is applied ($I_{IN1} = I_G^C/2 = 55\ \mu A$), the

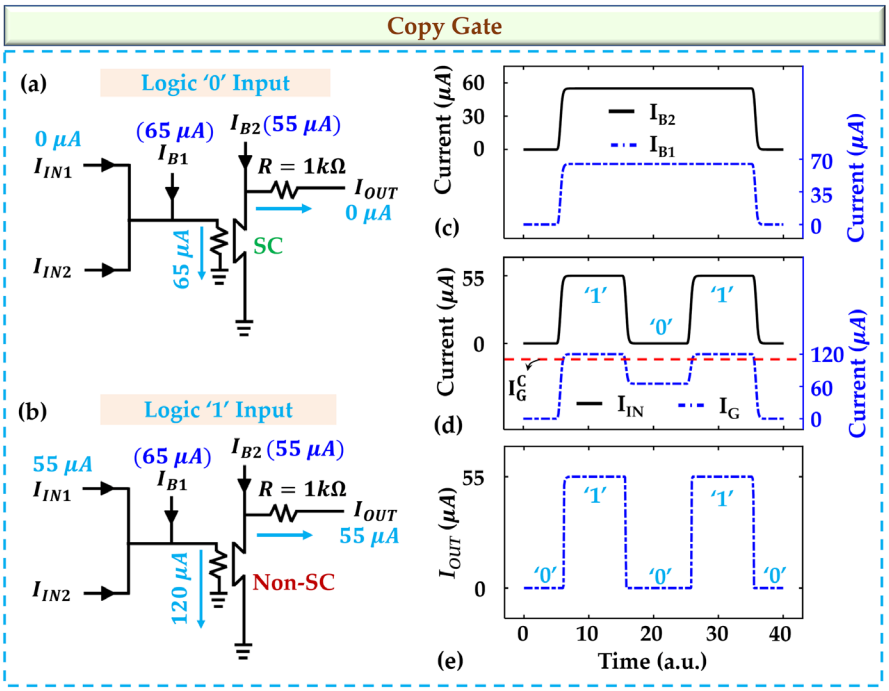


Figure 6.3: Copy operation with the proposed circuit when the input is **(a)** logic '0' and **(b)** logic '1'. **(c)-(e)** Simulated time dynamics of the copy operation.

total gate current ($I_G = 120\ \mu A$) exceeds I_G^C and switches the channel to its non-superconducting state which is highly resistive. Therefore, I_{B2} flows through the external resistor ($R = 1\ k\Omega$) and we get logic '1' ($I_{OUT} = I_G^C/2 = 55\mu A$) at the output. **Figs. 6.3**(c)-(e) show the simulated time dynamics of the copy gate which verifies the functionality of the logic function.

Next, we discuss the 1–input NOT gate where we use the same input terminal and same I_{B2}. The only difference is that we apply $-120\ \mu A$ as I_{B2}. Now, when a logic '0' ($I_{IN1} = 0$) is applied, only I_{B1} flows through the gate which itself is greater than I_G^C (**Fig. 6.4**(a)). As a result, the channel of the device switches to the non-superconducting state and drives I_{B2} to the external resistor. Therefore, we get logic '1' ($I_{OUT} = I_G^C/2 = 55\mu A$) at the output. However, when the input is logic '1' ($I_{IN1} = 55\ \mu A$), the device gets a total gate current of $-65\ \mu A$ which cannot switch the channel (**Fig. 6.4**(b)). Hence, the

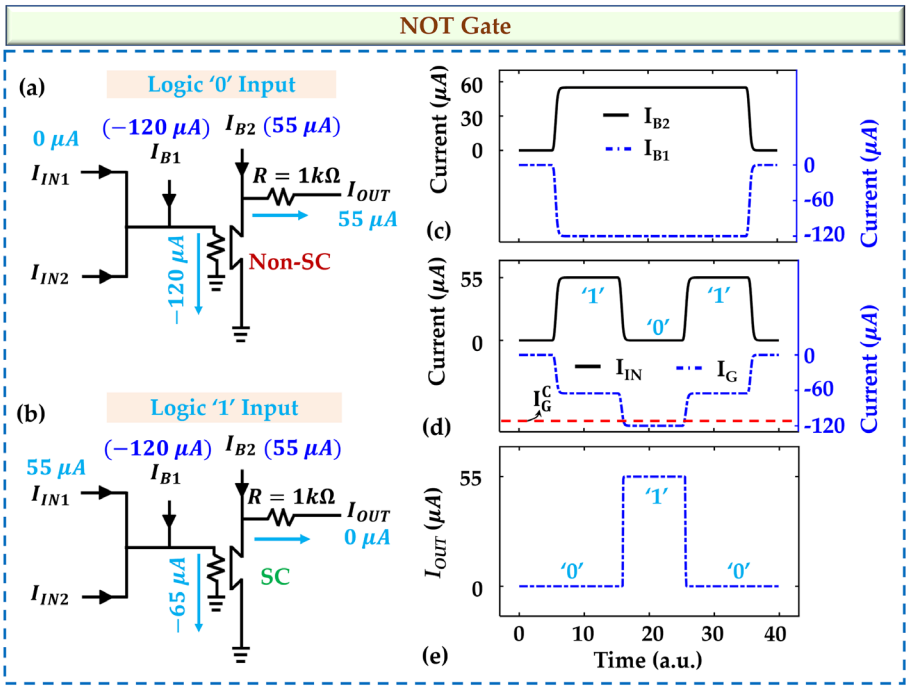


Figure 6.4: NOT operation with the proposed circuit when the input is **(a)** logic '0' and **(b)** logic '1'. **(c)-(e)** Simulated time dynamics of the copy operation.

channel remains superconducting and we get 0 A at the output. **Figs. 6.4**(c)-(e) verify the NOT gate functionality.

With the same circuit, we can also perform 2–input Boolean logic functions such as AND and OR logic where we use two input terminals to apply the input currents (I_{IN1} and I_{IN2}). To perform AND operation, we apply $I_{B1} = 10\ \mu A$. As a result, for input combinations of '00' and '01' (or '10'), the gate gets $10\ \mu A$ and $65\ \mu A$, respectively which are not sufficient enough to switch the state of the hTron channel (**Fig. 6.5**(a)). Therefore, the channel remains in its superconducting state and drives $0\ \mu A$ (logic '0') to the output. On the other hand, for the '11' input combination, a total current of $120\ \mu A$ flows through the gate and switches the channel to the non-superconducting state (**Fig. 6.5**(b)). Therefore, I_{B2} flows through the external resistor and we get logic '1' ($55\ \mu A$) at the output. The time dynamics of AND functionality are shown in **Figs. 6.5**(c)–(e).

Next, to perform the OR operation, we apply $I_{B1} = 65\ \mu A$. As a result, whenever one of the two inputs is logic '1', the gate terminal gets sufficient current to switch the

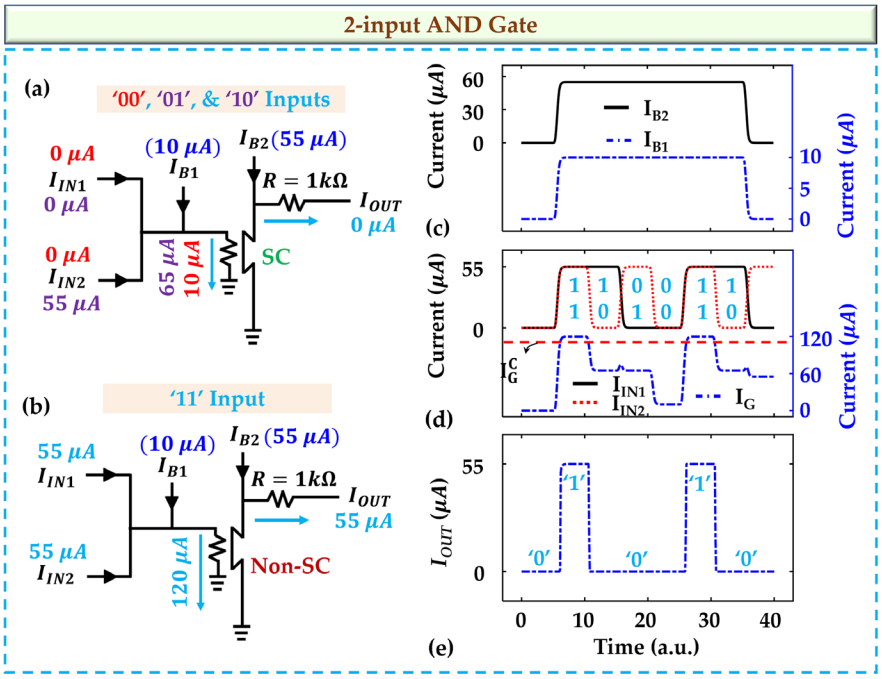


Figure 6.5: AND operation with the reconfigurable circuit for **(a)** '00', '01', and '10' and **(b)** '11' input combinations. **(c)-(e)** Simulated time dynamics for the AND operation.

channel of the device. Therefore, only for the input combination of '00', the channel remains superconducting and for '01', '10', and '11' combinations, the channel switches to its non-superconducting state (**Figs. 6.6**(a) and (b)). As seen in **Figs. 6.6**(c)-(e), we get the successful OR operation.

Majority is a Boolean logic operation that generates an output of logic '1' if there is a greater number of logic '1' among an odd number of inputs. The majority logic can be utilized to perform basic logic operations such as AND, OR, XOR, etc., and requires a lower number of gates to develop arithmetic-intensive systems [120]. Compared to the AND implementation, majority reduces the logical depth of the arithmetic-intensive circuits by up to 33% [49]. For example, while building a 1–bit full adder, the use of majority can reduce the logic level by 50% and 57.14% compared to the NAND and NOR implementations, respectively, and can reduce the number of required cycles by 60% [49].

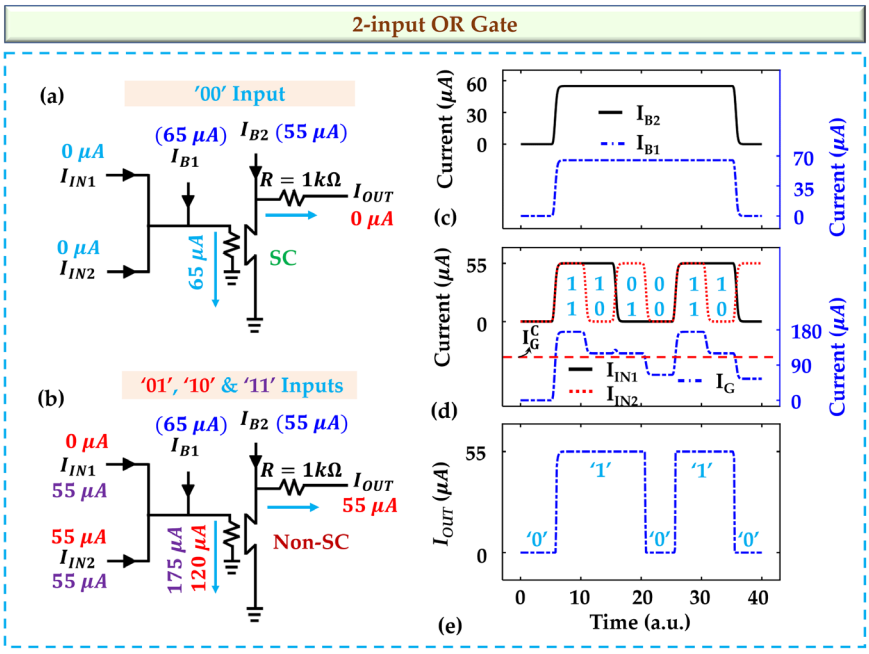


Figure 6.6: Illustration of the OR operation for **(a)** '00' and **(b)** '01', '10' and '11' input combinations, respectively. **(c)-(e)** Time dynamics of the OR operation where the channel switches for the '11' combination due to the choice of I_{B1}.

Here, we show that the same circuit with a single hTron device can be reconfigured to perform 3–input majority logic operation. Here, we use $I_{B1} = 10\ \mu A$. As a result, when there is a smaller number of logic '1' among the inputs, the gate terminal does not get sufficient current to switch the channel (**Fig. 6.7**(a)). On the other hand, when there is a greater number of logic '1' inputs, the total gate current exceeds I_G^C and the channel switches to its non-superconducting state (**Fig. 6.7**(b)). As seen in the transient results of **Figs. 6.7**(c)-(e), the same circuit can perform the 3–input majority logic function.

6.5 Summary

We demonstrate a single hTron-based reconfigurable logic circuit that can perform 1–input copy and NOT, 2–input AND and OR, and 3–input majority operations by only choosing suitable biasing conditions. Our proposed logic circuit benefits from useful features of the hTron including low turn-on-time, input-output isolation, low power operation, and so on [66]. Moreover, hTron devices do not need any superconducting loop like other superconducting devices such as JJs and SQUIDs. Therefore, our

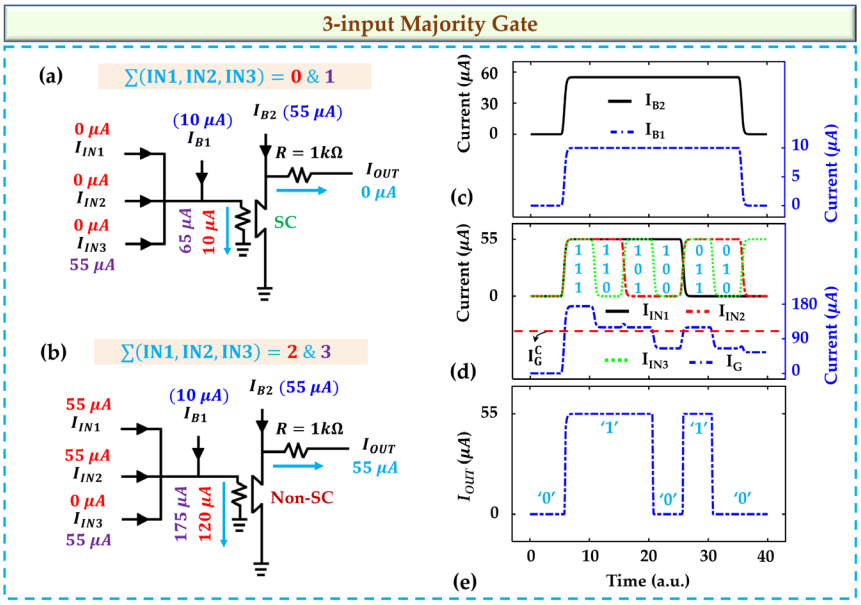


Figure 6.7: Illustration of majority logic when there are **(a)** greater and **(b)** smaller number of logic '1' inputs. **(c)-(e)** Simulated time dynamics of the Majority operation.

proposed logic circuit will not suffer from any flux trapping issue and will be able to solve the scalability issue of the JJ and SQUID-based circuits. Typically, the number of required devices in a logic gate increases with the number of inputs which is not the case for our design. We use the same circuit based on one hTron device to reconfigurably perform all these basic operations, enabling the implementation of a superconducting processing unit with a much smaller number of devices compared with the existing approaches. Moreover, since we are using the same circuit for different logic operations, we can develop a camouflaged processing unit where all the logic gates will have the same layout and will reduce the risk of reverse engineering.

Also, thanks to the high impedance ($\gg 1\ k\Omega$) non-superconducting state of hTron devices, the cascadability of the proposed logic circuit is not a concern like the other current-controlled superconducting logic gates. One hTron can drive other resistive gate terminals (typically has an impedance of $1\ k\Omega$) of the next stage logic circuits. Additionally, our proposed copy gate can be used to develop a current splitter circuit like the one proposed in [121] to enable a larger number of fanouts.

Finally, hTron devices suffer from reset time limitations due to the thermal recovery after each switching and so will be the case for our proposed circuit. To solve this issue, a different superconductor with a suitable thickness (for example, NbN with $\sim 1\, ns$ thermal reset time [122]) or the same superconductor with different thickness can be used to make the channel microwire. Furthermore, our proposed circuit is not limited to thermal hTron devices. Our proposal will be applicable to any device that shows gate-controlled switching between superconducting and non-superconducting states like JJ field effect transistors [74], [75], Dayem transistors [77], [83], ferroelectric SQUIDs [42], [112], and so on.

CHAPTER 7
NON-VOLATILE MEMORY ARRAY BASED ON QUANTUM ANOMALOUS HALL EFFECT

7.1 Introduction

Electronic bands of non-trivial topology give rise to quantum phenomena at the macroscopic scale. Under a strong magnetic field, two-dimensional electron systems exhibit such a quantum state- the quantum Hall effect, i.e. the quantization of Hall resistance. Back in 1988, Haldane [123] predicted that a quantum Hall state could arise even in the absence of an external magnetic field. This new state became known as the quantum anomalous Hall effect (QAHE). However, it took more than two decades for the first experimental realization of QAHE, when it was demonstrated in a magnetic topological insulator [Cr-doped $Bi(Sb)_2Te_3$] [124]–[131]. The experimental fingerprint of QAHE is straightforward: The Hall resistance is quantized to $h/\nu e^2$ (where ν, the so-called Chern number, is an integer that depends on the topological properties of the band structure) at zero magnetic field. Such simple and universal values stem from the current being carried by lossless edge channels which require neither high electron mobility [132], [133] nor external magnetic fields. Moreover, on a given edge, the current flows only in one direction. Thanks to such transport properties which are also immune to sample complexities [132], [133], QAHE can potentially be useful in spin-filtering[132], resistance metrology [134], and topological quantum computing [135]. However, how we can materialize these prospects and come up with a QAHE-based electronic device has remained elusive so far.

Here we bridge the gap between the QAHE physics and device architecture to build the framework of a scalable non-volatile memory. We utilize the quantization of Hall resistance in QAHE to design a memory cell and then construct a 3D cross-point memory array capable of efficient read and write operations.

7.2 Quantum Anomalous Hall Effect (QAHE)

We start by delineating the dynamics of a QAH insulator and how it can be used as a non-volatile memory cell. **Fig. 7.1**(a) illustrates the schematic of a QAH insulator with electric contacts to apply bias current and to measure both longitudinal and transverse

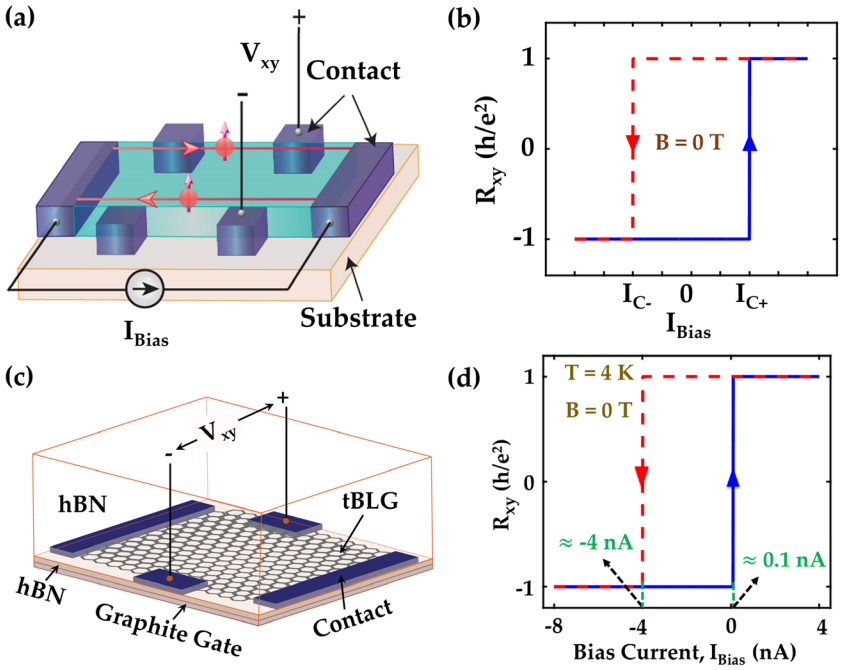


Figure 7.1: (a) Schematic of a QAH insulator, where the red circles represent the electron and the red arrows show the spin of the electrons. **(b)** Illustration of the Hall resistance R_{xy} vs. the bias current, I_{Bias} at zero external magnetic field. I_{C-} and I_{C+} are two critical values of I_{Bias} which determine the hysteretic switching of R_{xy} between $-h/e^2$ and h/e^2. **(c)** Schematic of a tBLG moiré heterostructure where tBLG is encapsulated between flakes of hBN and a flake of few-layer graphite is used as gate. **(d)** R_{xy} data, plotted as a function of I_{Bias} at T = 4 K and B = 0 T, exhibit current driven hysteretic switching.

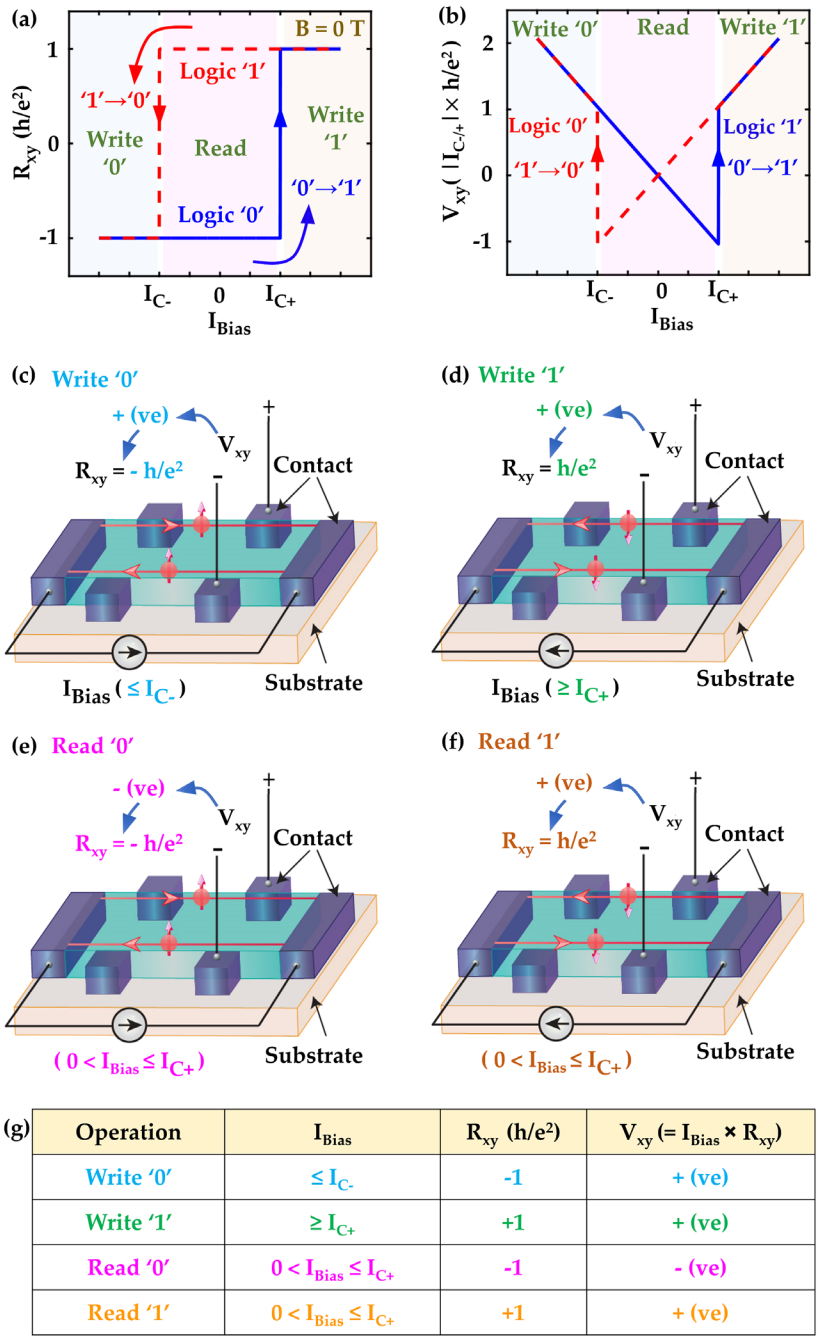
(Hall) voltages. The QAH states occur when an applied gate voltage tunes the bulk carrier density to close to zero. Then, the Hall resistance shows hysteretic switching with the change of magnetic field. Also, magnetic domains of the ferromagnetic materials (for example, tBLG) strongly interact with the applied electrical current, which provides electrical control over the polarization of the magnetic domains [136]. It has also been shown [137] that the switching of Hall resistance driven by DC bias current is similar to that driven by a magnetic field. We leverage this tunability in our design.

To design our memory cell, we use the intrinsic QAHE reported in twisted bilayer graphene (tBLG) on hexagonal boron nitride (hBN) moiré heterostructure, where the quantization of Hall resistance persists at temperatures as high as 6 K [137]. This is significantly higher compared to the magnetic topological materials, such as transition metal doped $(Bi, Sb)_2Te_3$ thin films [98]–[101] and $MnBi_2Te_4$ [142].With the ongoing thrust in material discoveries, room-temperature applications may become feasible. Very recent works on QAHE in $TbMn_6Sn_6$, $MnBi_2Te_4$ offer such prospects [143]–[145]. **Fig. 7.1**(c) schematically shows the tBLG moiré heterostructure where tBLG (with interlayer twisted angle of 1.1°) is encapsulated between flakes of hBN. **Fig. 7.1**(d) shows R_{xy} as a function of I_{Bias} for the tBLG heterostructure at zero external magnetic field and at T = 4 K. As seen from **Fig. 7.1**(d), we can deduce the values of I_{C-} and I_{C+} for the tBLG heterostructure (approximately -4 nA and 100 pA, respectively) which are crucial for the memory operation.

Broadly, our design can be implemented with any material platform harboring an electrically tunable QAHE. Note that the exact reason behind the electrical switching of QAHE in tBLG awaits a comprehensive understanding. Orientation of the graphene layers as well as the interaction between the electrical current and ferromagnetic domains in tBLG may play a role. It is also worthwhile to add that electrically tunable QAHE is not unique in tBLG; tri-layer graphene (twisted monolayer-bilayer graphene) also shows similar tunability [146], [147].

7.3 QAHE-based Memory

Fig. 7.2 captures the highlight of our idea to leverage the quantized Hall resistance of a QAH insulator as a non-volatile memory cell along with the write and read operations.



(g)

Operation	I_{Bias}	R_{xy} (h/e^2)	V_{xy} (= I_{Bias} × R_{xy})
Write '0'	$\leq I_{C-}$	-1	+ (ve)
Write '1'	$\geq I_{C+}$	+1	+ (ve)
Read '0'	$0 < I_{Bias} \leq I_{C+}$	-1	- (ve)
Read '1'	$0 < I_{Bias} \leq I_{C+}$	+1	+ (ve)

Figure 7.2: (a) R_{xy} vs. I_{Bias} at zero external magnetic field. **(b)** Hall voltage, V_{xy} plotted as a function of I_{Bias}. **(c)**, **(d)**, **(e)**, & **(f)** The procedures of write '0', write '1', read '0', and read '1' operations, respectively, along with the state of electron spins (red arrows). **(g)** Summary of the key idea for the QAHE based memory.

We define the quantized Hall resistances, $-h/e^2$ and $+h/e^2$, as logic '0' and logic '1' respectively. Based on this definition, I_{Bias} can be divided into three regions (**Fig. 7.2**(a)) – (i) $I_{Bias} \leq I_{C-}$: write '0' region, (ii) $I_{Bias} \geq I_{C+}$: write '1' region, and (iii) $I_{C-} < I_{Bias} < I_{C+}$: read region. **Fig. 7.2**(b) shows the bias current dependence of the Hall voltage, V_{xy} ($=I_{Bias} \times R_{xy}$). Importantly, as seen in **Fig. 7.2**(b), write operations of both logic '0' and '1' entail positive V_{xy} because I_{Bias} and R_{xy} are either both positive (when writing logic '1'; **Fig. 7.2** (c)), or both negative (when writing logic '0'; **Fig. 7.2** (d)). In sharp contrast, read operations of logic '0' and '1' manifest opposite sign of V_{xy}. For example, if we use a positive bias current ($0 < I_{Bias} < I_{C+}$), logic '1' and logic '0' correspond to positive and negative V_{xy}, respectively. The clear difference in the sign of the Hall voltage for logic '0' and '1' makes the sensing of the memory states simple and straightforward. **Fig. 7.2**(g) summarizes the key idea of a non-volatile memory utilizing the quantization of the Hall resistance in a QAH insulator, listing I_{Bias} for write and read operations, as well as the state of the Hall resistance and Hall voltage for the two memory states.

7.4 Design of Cross-point Memory Array

To evaluate the practicality of a memory device, it is crucial to consider an array level scenario. The unique properties of our QAHE based memory device necessitate a custom designed memory array. To ensure the highest storage density, we adopt the cross-point memory architecture [148] with slight modification in the inter-cell connection pattern. A major component of the cross-point array is a two-terminal selector device [149] that allows access to a specific cell for read/write operation and suppresses current flow through the other cells. The selector devices are connected in series with the memory elements in every cell. A multitude of selector devices exist [150]–[155] with diverse selectivity (ON/OFF ratio) and switching thresholds. For our QAHE-based memory, the Cu-containing mixed-ionic-electronic-conduction (MIEC) material will be a perfect selector, due to its high selectivity (~10^6), and ultra-low leakage (< 10 pA) [156]–[158]. **Figs. 7.3** (a) and (b) show the schematic and *I-V* characteristics of the MIEC material-based selector, respectively.

Fig. 7.3(c) shows the schematic view of our proposed array architecture. Every memory cell (QAHE device + selector) is sandwiched between orthogonally running

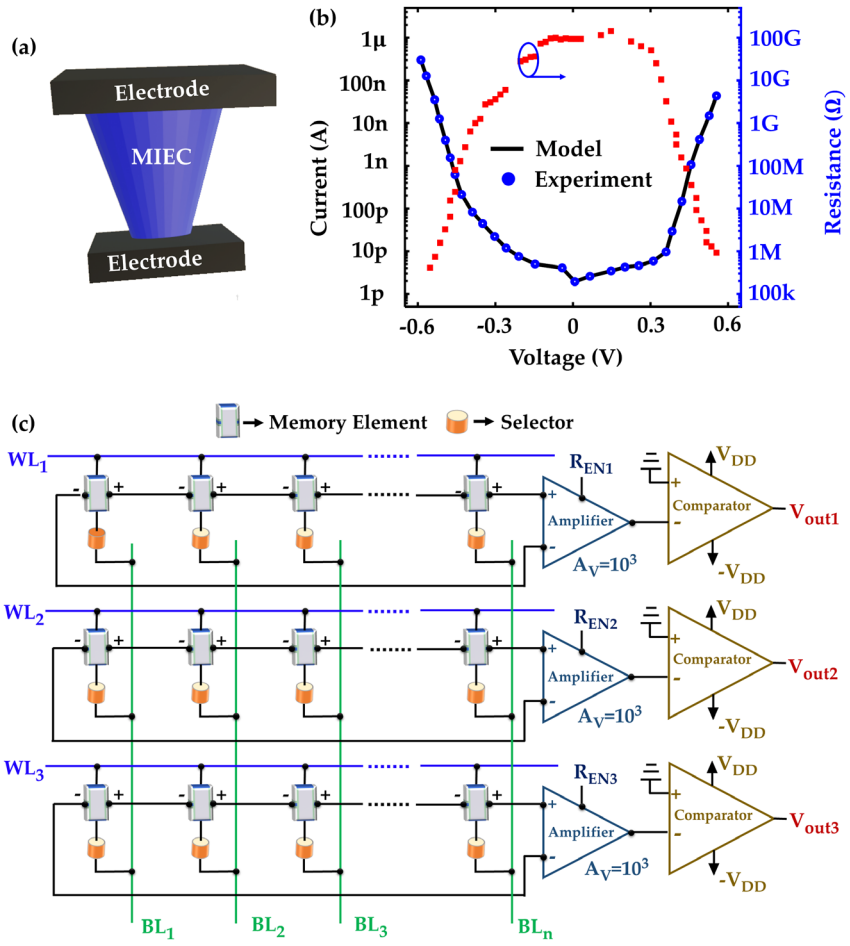


Figure 7.3: (a) Device structure and **(b)** characteristics of an MIEC material-based selector. **(c)** Illustration of the overall memory array structure, custom-designed for the QAHE based memory devices.

metal lines, namely word line (WL) and bit line (BL). The WLs and the BLs are shared along the rows and columns of the array, respectively. The difference between the DC voltages applied on the WL and BL of a particular a memory cell essentially dictates the effective voltage across the cell (V_{Cell}). Note, the substrate terminal of every cell is biased with a constant voltage to maintain an appropriate electron density (not illustrated in the figures for simplicity). Conveniently, we need only one control input, V_{Cell}, to read from or write into the QAHE devices. We utilize three different levels of V_{Cell} (with appropriate polarity) to write logic '1/0' and to read the stored data.

Our architecture entails accessing one cell per row at a time. The Hall-voltage terminals of the neighboring cells in the same row are electrically connected in series with each other. Importantly, the resultant voltage across the end terminals of the series connected cells in a row (V_{hr}) bears the signature of the stored data in the accessed memory cell, along with some residual contributions from the other cells in the same row. For example, consider that every row holds n cells, and the k^{th} cell of the i^{th} row is being accessed. The resultant terminal voltage for this row can be expressed as:

$$V_{hr-i} = \sum_{j=1}^{n} V_{xy-ij} - 2n \times V_{loss}$$

$$= I_{\text{Bias}-ik} \times R_{xy-ik} + \sum_{j=1,\ j\neq k}^{n} I_{\text{Bias}-ij} \times R_{xy-ij} - 2n \times V_{loss} \qquad (4)$$

Here, V_{loss} denotes the average loss component per contact, indicating the inherent non-ideality. The specialized biasing scheme of the memory array (discussed later) ensures that the losses and residual components are negligible compared with the Hall voltage of the accessed cell. As a result, the polarity of V_{hr} is dictated by the R_{xy} of the accessed cell. For the tBLG moiré heterostructure based QAHE device, we need to amplify the V_{hr} from a few tens of microvolts to tens of millivolts using cryogenic amplifiers. A suitable candidate for such an amplifier can be one of the cryogenic low-noise amplifiers reported in Refs. [159]–[161]. After the amplification, the millivolt level Hall voltage is used to determine the memory state of the accessed cell. Recall, during read operation, logic '0' and '1' memory states correspond to opposite polarity of Hall

voltage (**Fig. 7.2**(b)). Therefore, we feed the amplified V_{hr} to a cryogenic voltage comparator [162] to sense the memory state of the accessed cell.

7.5 Modeling and Simulation Framework

We develop a Verilog-A-based phenomenological model for the QAHE in tBLG moiré heterostructure to use in our analysis of the heterostructure as a memory cell. The model can also be calibrated for other QAH insulators using the values of bias currents (I_{C-} and I_{C+}) and the Chern number (ν). A behavioral representation of our model is shown in **Fig. 7.4**(a). In our phenomenological model, we also account for the temperature dependence of Hall resistance using the following equation [137]:

$$R_{xy} = \frac{h}{e^2} - R_1 e^{-\frac{\Delta}{T}}, \tag{5}$$

where, R_1 is a fitting constant, Δ (= 26 ± 4 K) is the energy required to create and separate an excitation of particle-antiparticle of the QAH state and T is the temperature. **Fig. 7.4**(b), which shows the temperature dependence of the Hall resistance obtained from our model, agrees reasonably with the measured values reported in [137].

We also develop a look-up-table-based phenomenological model (in Verilog-A) for the MIEC material-based selector and calibrate the model (shown in **Fig. 7.3**(b)) with the measured current-voltage (I-V) characteristics reported in [156]–[158].

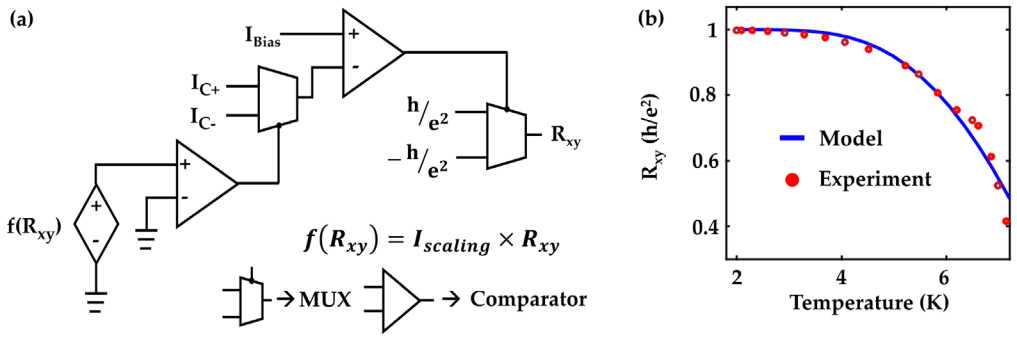
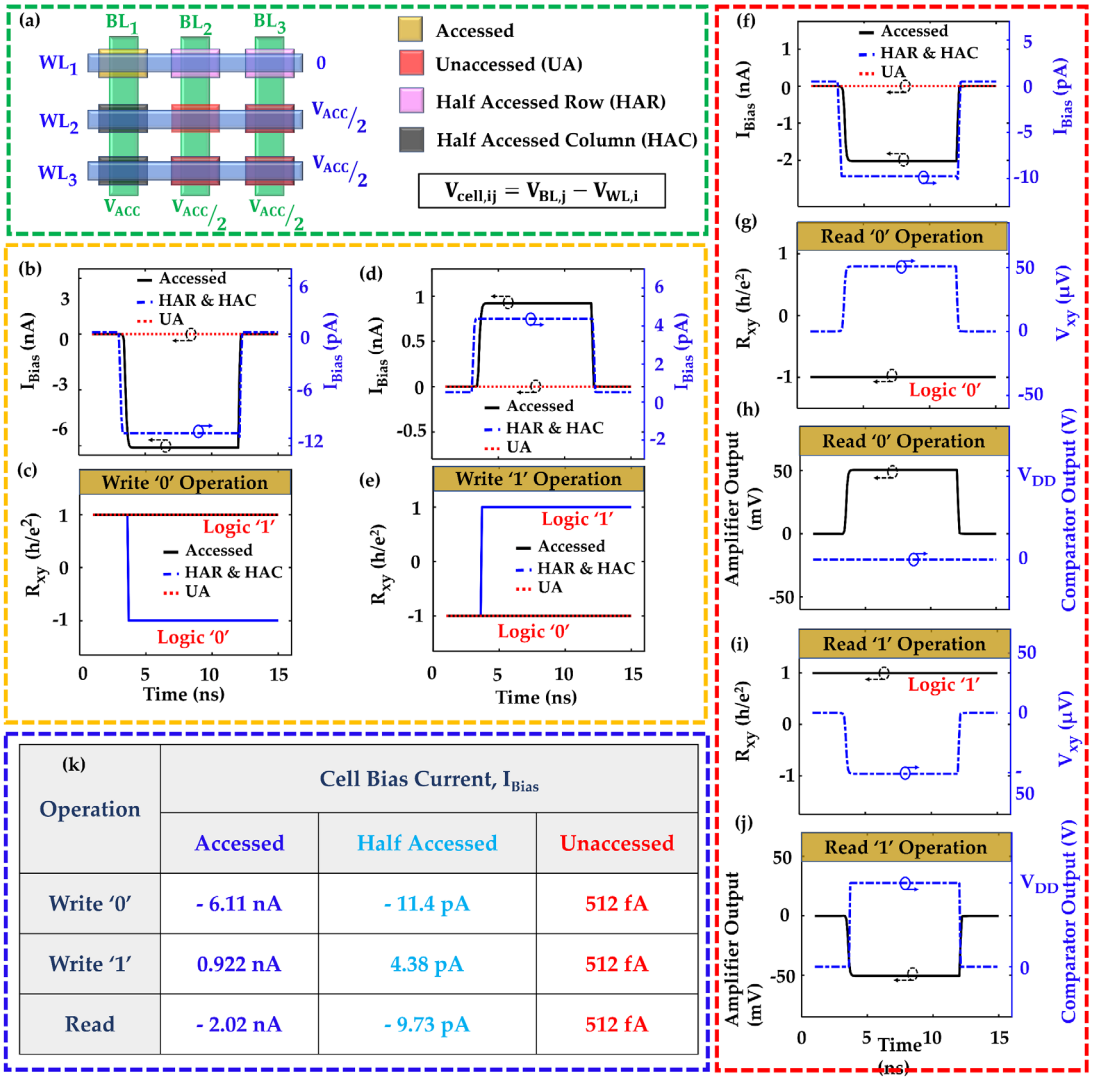


Figure 7.4: (a) Behavioral representation of the modeling approach for the QAHE observed in the tBLG moiré heterostructure. **(b)** Temperature dependence of R_{xy} with an experimental matching.

7.6 Write and Read Operations

In **Fig. 7.5**, we present the simulated memory operations (read/write) in our proposed QAHE based cross-point array. **Fig. 7.5**(a) illustrates four types of memory cells in a block of cross-point array. We utilize a standard biasing scheme for cross-pint arrays, commonly known as the *V/2* biasing [155]. Different levels of access voltage (V_{ACC}) are applied to read from or write into a specific cell. The row (column) that holds the accessed cell is called the half-accessed row (column), because the inactive cells in this row (column) receive half of the access voltage ($V_{ACC}/2$). All the other unaccessed cells ideally have zero voltage across them. Without any loss of generality, we assume M_{11} (**Fig. 7.3**(c)) to be the accessed cell and examine its read/write dynamics. **Fig. 7.5**(b) shows the cell current levels through the accessed, half-accessed, and unaccessed cells during 1→0 memory write operation. Only the accessed cell exhibits a transition ($+h/e^2$ → $-h/e^2$) in the Hall resistance (R_{xy}) (**Fig. 7.5**(c)), indicating a successful write operation in the accessed cell without disturbing the other cells. **Figs. 7.5**(d) and 7.5(e) show similar time dynamics for the 0→1 write operation. Note, the cell current level for the 0→1 operation is of the opposite polarity compared to the 1→0 operation (**Fig. 7.5**(d)).

Next, we examine the read dynamics (**Figs. 7.5**(f)-(j)) of the proposed memory cell. The current flow direction for the read operation is the same as the 1→0 write operation (**Fig. 7.5**(f)). However, the magnitude of the read current is about one-third of the critical current of transition (I_{C-}), which alleviates the possibility of accidental 1→0 data flip. In addition, the chosen polarity of the read current nullifies the possibility of accidental 0→1 transition. The read current through the QAHE device generates a Hall voltage (*Vxy*), whose polarity is dictated by the previously set Hall resistance (($+h/e^2$ or $-h/e^2$). In our design, positive/negative polarity of *Vxy* corresponds to logic 0/1, respectively (**Fig. 7.5**(g), (i)). It is worth noting, the read operation utilizes different pair of terminals than the write operation, allowing independent optimization of read/write peripheral circuits. Furthermore, the amplifier-comparator pairs are enabled only during the read operation to minimize peripheral power demand. The comparator in our design also provides digital outputs corresponding to the stored data in the accessed cell (**Figs. 7.5**(h) and (j)). The



(k)

Operation	Cell Bias Current, I_{Bias}		
	Accessed	Half Accessed	Unaccessed
Write '0'	- 6.11 nA	- 11.4 pA	512 fA
Write '1'	0.922 nA	4.38 pA	512 fA
Read	- 2.02 nA	- 9.73 pA	512 fA

Figure 7.5: (a) Arrangement and types of memory cells in a conventional cross-point memory array under the *V/2* biasing scheme. I_{Bias} through different memory cells generated by applying appropriate V_{ACC} during **(b)** write '0' and **(c)** write '1' operations. Time dynamics of R_{xy} illustrating state transition only in the accessed cell during **(d)** write '0' and **(e)** write '1' operations. **(f)** I_{Bias} through different cells during read operation. Time dynamics of R_{xy} and V_{xy} during **(g)** read '0' and **(i)** read '1' operations. Time dynamics of the amplifier and comparator outputs during **(h)** read '0' and **(j)** read '1' operations. **(k)** Summary of I_{Bias} through different cells during write and read operations.

array biasing scheme is designed to enforce high resistance mode of the selector in every memory cell but the accessed cell. This design technique ensures that the half-accessed and unaccessed cells encounter orders of magnitude less current compared with the accessed cell. The MIEC selector takes an overwhelming share of any voltage applied across the cell, averting the chances of a breakdown in the QAHE device. The bias currents that flow through the cells during different memory operations are shown in **Fig. 7.5**(k) which clearly demonstrate that the memory states of the half-accessed and unaccessed cells will not be disturbed during write or read operation in the accessed cell. Thus, our architecture lays out a device-to-array design pathway for the QAHE based unique memory devices.

7.7 Summary

Our cryogenic memory framework is transferable to the material structures that exhibit current-controlled switching of QAH states (*e.g.,* tri-layer graphene [146], [147]). Some of these material structures will necessitate an initialization step to first create a magnetic moment in the system. The additional analog components required for this purpose will be parts of the array peripherals. The underlying transition mechanisms of the QAHE devices may determine their endurance and reliable life span.

This work made the first mark to construct a working electronic device using topological properties of materials. Here we leverage QAHE, that does not require an external magnetic field, and design an elegant memory device that can be built with topological quantum materials. The proposed non-volatile memory architecture is a strong candidate for the cryogenic memory system, thanks to its ultra-low temperature compatibility. Note, the cryogenic memory block is a crucial component of quantum computing systems based on superconducting qubits [163]. Our proposed QAHE-based memory device offers a significant reduction in the cell area and 1000 times reduction in the cell read/write power compared with the state-of-the-art cryogenic memory devices [164]–[171] (see **Table 7.1** for detailed comparison). Our proposed memory device is a potential game-changer for scalable quantum computing systems [163] and space cryogenics [172].

Table 7.1: Comparison with the state-of-the-art cryogenic memories

Cryogenic Memories	**Cell Area**	**Power Consumption**
Superconducting Memory Chip [164]	$50 \times 46\ \mu m^2$	1.62 mW for 2Kb RAM [Approximated 0.79 µW per cell]
VT RAM [168]	$55 \times 55\ \mu m^2$	6.7 mW for 4 Kb RAM [Approximated 1.64 µW per cell]
SFQ CRAM [169]	$40 \times 45\ \mu m^2$	2.4 mW for 16 Kb RAM [Approximated 0.15 µW per cell]
High frequency RAM [173]	$55 \times 55\ \mu m^2$	Not Reported
Hybrid Josephson-CMOS Memory [165], [166]	$6.5 \times 7.5\ \mu m^2$	0.7 mW for reading and 1.4 mW for writing for 64 Kb RAM [Approximated 0.01 µW for reading and 0.02 µW for writing per cell]
Small Array of Josephson Junctions [170], [171]	$5 \times 5\ mm^2$ (Chip Size)	~ 5 nW (maximum) for reading and writing per cell (predicted)
QAHE-based Proposed Memory (This work)	$7 \times 8\ \mu m^2$	0.96 pW for writing and 0.1 pW for reading per cell (simulated)

CHAPTER 8
QUANTUM ANOMALOUS HALL EFFECT-BASED IN-MEMORY BOOLEAN LOGIC OPERATIONS

8.1 Introduction

Future scalability of CMOS technology will continue to become more challenging due to the exorbitant power dissipation resulting from the aggressive downsizing. This limitation is one of the major concerns in designing high-performance systems at the architecture level. Cryogenic electronics, where the processing and storage units are cooled to very low temperatures (4 K or below), is an extreme yet one of the most promising solutions to circumvent this high-power dissipation issue. Cryogenic operations lead to significant improvement in computing speed and energy efficiency compared to the room temperature operations (illustrated in **Fig. 8.1**(a)) [12], [13], [28].

Recent studies show that the cryogenic superconductive systems consume around 80 times less energy compared to the room-temperature 7 nm CMOS technology even including the cost required for helium cooling [25]. Nevertheless, the cooling cost remains a challenge for the cryogenic systems, calling for a low-power and low cooling-cost cryogenic system architecture [174]. For low-power operations, the compute-in-memory (CiM) architectures, where computation and memory operations are performed within a memory array, have been explored for cryogenic and room temperatures with both volatile (SRAMs [175]–[177] and DRAMs [176], [178], [179]) and non-volatile (resistive [180]–[182], spintronic [85], [183]–[185], ferroelectric [186]–[188], phase change [189], [190], and others [191]) memory platforms. CiM architectures avoid the power-consuming data movement between the processing and memory units and hence, can operate with a high energy efficiency [192].

Along with the excellent energy efficiency, CiM architectures can lead to throughput improvement. '*Memory wall*', the growing disparity of speed between the processing and memory units, is one of the major challenges for the *von-Neumann* architecture (shown in

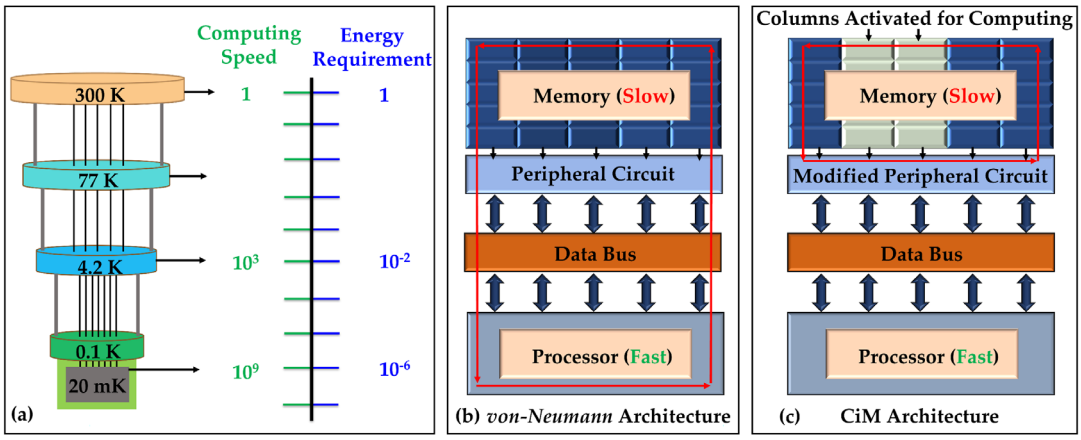


Figure 8.1: (a) Significant improvement in computing speed and energy requirement with the lowering of operating temperature. Computing speed and energy requirement have been normalized by the value at room temperature (300 K). Illustration of **(b)** traditional *von-Neumann* architecture and **(c)** compute-in-memory (CiM) architectures.

Fig. 8.1(b)) in today's big data computing era. With the emergence of data-intensive applications (such as machine learning and artificial intelligence, etc.), the performance mostly depends on the memory units. Moreover, the data transfer between the memory and the processing units causes delay and requires more energy than the computation process itself, posing a bottleneck [193]. For example, *Google's* recent studies show that about 20-42% of the energy is required to drive the high-speed data bus connecting the processing and memory units [194], [195]. On the other hand, CiM architectures aim to significantly reduce the data transfer (**Fig. 8.1**(c)) and hence, emerge as an attractive solution to the '*memory wall'* bottleneck for data-intensive applications. Also, CiM architectures take advantage of the large internal memory bandwidth and hence solve the issue of limited throughput through massive parallelism.

Beyond traditional computing, quantum computers will also benefit from the CiM architectures. A quantum computing system also comprises several classical hardware components, including a control processor and a memory block [9], [163]. The state-of-the-art quantum computing systems necessitate that the control processor and the memory block operate at cryogenic temperature. The speed and energy demands of the existing cryogenic memory systems are not yet comparable to the cryogenic single flux

quantum (SFQ) control processors [9], [12], [51], [196]–[198]. Thus, an equivalent '*memory wall*' problem arises even in the quantum computing paradigm. Cryogenic CiM can help alleviate this disparity and reduce classically imposed overhead in quantum computing systems [9].

As mentioned earlier, almost all the volatile and non-volatile memory platforms have been explored for in-memory computing. Such CiM architectures, at room temp, have convincingly demonstrated significant mitigation of memory wall and show better performance with higher throughput and energy efficiency. At cryogenic temperatures, few works including volatile SRAM [177] and non-volatile STT-MRAM [85]-based CiMs have been characterized. It is now well established that the performances (latency and energy efficiency) of the CMOS devices (SRAMs and DRAMs) improve significantly at cryogenic temperatures [28], [113] which is supposed to reduce the '*memory wall'* bottleneck. However, the cryogenic temperatures are also supposed to increase the speed of the processing units by about 30-40% [12] which will increase the memory data access rate and worsen the '*memory wall'* issue. Compared to the volatile platforms, the non-volatile platforms show significantly higher energy efficiency at room temperatures [178], [199], which makes them very attractive for the cryogenic CiMs. However, in the only cryogenic characterization of non-volatile memory (STT-MRAM)-based CiM architecture [85], it has been reported that the cryogenic operations improve the latency but lower the energy efficiency compared to the room temperature operations. The characterization of FeFET at cryogenic temperatures also reported higher energy requirements compared to the room temperature since the program and erase voltages increase at cryogenic temperature [200].

While researchers, to date, have only characterized the room temperature memory-based CiM architectures at cryogenic temperatures, we present a CiM architecture based on a cryogenic device. Here, we utilize the QAHE observed in tBLG moiré heterostructures [137]. In Chapter 7, we have already discussed a cryogenic memory system harnessing the QAHE phenomenon, which provides ultra-low-power performance, high scalability, excellent distinguishability between two logic states, simple peripheral circuitry, and higher sense margin. In this work, we utilize the same

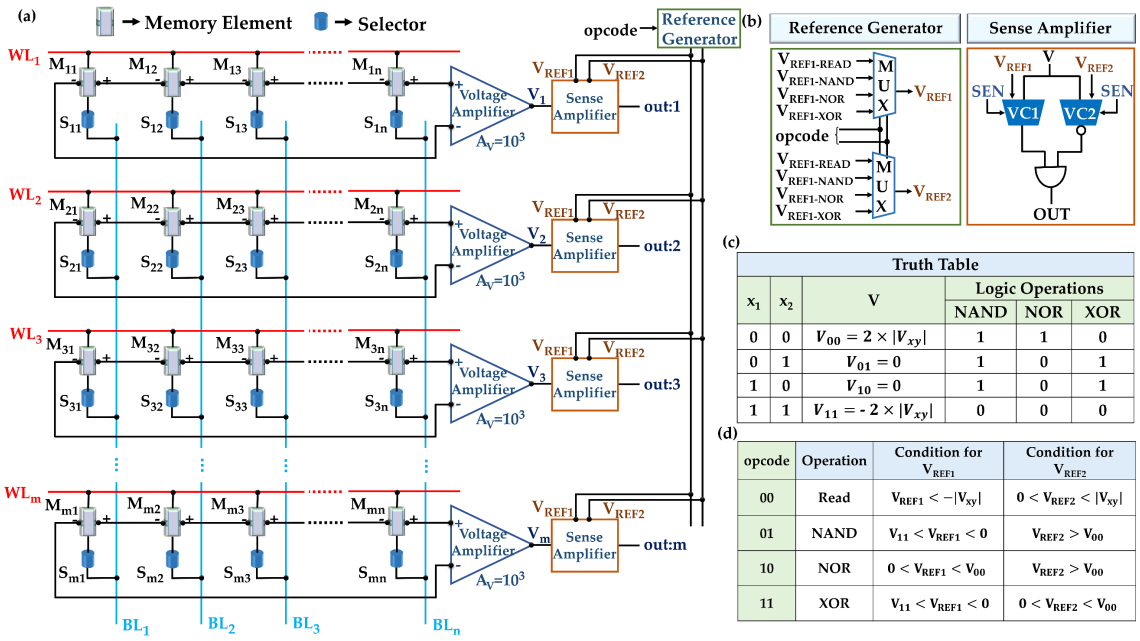
memory array with modified peripheral circuitry to demonstrate compute-in-memory. The key contributions made by this work are highlighted below-

- This is the first attempt to design a cryogenic CiM technology using a dissipation-less topological material system that we argue to be a natural fit for this application.
- In contrast to previous works that aimed to characterize conventional CiM technologies at cryogenic temperature [85], [177], we design this Cryo-CiM technique from scratch using physical phenomena native to cryogenic environment. Thus, we avert the limitations of the conventional technology platforms.
- Unprecedentedly, our architecture can compute all fundamental logic operations (NAND, NOR, XOR) within a single-cycle. We construct a peripheral assembly with dynamically customizable reference to facilitate on-chip sensing and interpretation of different Hall voltage levels which allows the design to achieve this feat.
- Our design continuously computes an analog sum of Hall voltages which dramatically improves performance and reduces power demand.

8.2 QAHE-based Logic-in-memory System

We utilize the crosspoint memory array (discussed in detail in Chapter 7) developed using the MIEC material-based selector and QAHE-based memory element to perform in-memory logic operations. Note, for the QAHE-based memory, to sense the memory state of a cell in the memory array, a simple voltage comparator, able to differentiate between positive and negative voltage, was sufficient. However, for the in-memory logic operations, a more complicated peripheral circuitry is required. **Fig. 8.2**(a) shows the complete crosspoint memory array with the modified peripheral circuitry for in-memory computing. In this work, we discuss the read and in-memory logic operations with this memory array. Please refer to Chapter 7 for a detailed discussion on the biasing schemes for accessing a specific cell in the array (write and read operations) and design of peripheral circuitry for the memory operations.

In this memory array, to activate a cell for read and logic operations, a suitable voltage is applied between the BL and word-line WL which creates the read current through the corresponding memory cell. As shown in **Fig. 7.2**(b), the read current creates two different polarities of V_{xy} for two memory states. We connect the Hall voltage terminals



Truth Table

x_1	x_2	V	Logic Operations		
			NAND	NOR	XOR
0	0	$V_{00} = 2 \times \lvert V_{xy} \rvert$	1	1	0
0	1	$V_{01} = 0$	1	0	1
1	0	$V_{10} = 0$	1	0	1
1	1	$V_{11} = -2 \times \lvert V_{xy} \rvert$	0	0	0

opcode	Operation	Condition for V_{REF1}	Condition for V_{REF2}
00	Read	$V_{REF1} < -\lvert V_{xy} \rvert$	$0 < V_{REF2} < \lvert V_{xy} \rvert$
01	NAND	$V_{11} < V_{REF1} < 0$	$V_{REF2} > V_{00}$
10	NOR	$0 < V_{REF1} < V_{00}$	$V_{REF2} > V_{00}$
11	XOR	$V_{11} < V_{REF1} < 0$	$0 < V_{REF2} < V_{00}$

Figure 8.2: (a) QAHE-based cryogenic CiM structure. (Wires of different colors are not connected to each other unless shown using black dots at the intersections between two wires.) **(b)** Schematics of the reference generator and the sense amplifier. **(c)** Truth table for the 2-input logic operations and the corresponding voltage levels for different logic combinations. **(d)** Conditions for choosing the reference voltages for read and in-memory logic operations such as NAND, NOR, and XOR.

of the memory cells in a row so that the input of the amplifiers of each row gets the sum of the Hall voltages of all the cells in that row (**Fig. 8.2**(a)). The sum of V_{xy} of the activated cells is first amplified from micro-volt level to milli-volt level using a cryogenic voltage amplifier [159]–[161]. Then this amplified voltage is processed in the sense amplifier which generates the final output for the read and logic operations. The sense amplifier, modified for the CiM purpose, consists of two simple voltage comparators (VC1 and VC2), one inverter, and one AND gate (shown in **Fig. 8.2**(b)). The two voltage comparators are used to compare the amplified voltage with two different reference voltages (V_{REF1} and V_{REF2}) set by the reference generator. In our simulation, we utilize a behavioral model for the voltage comparators which behave as (similar to all the conventional voltage comparators):

$$V_{out} = \begin{cases} 0, & V_{in} < V_{REF} \\ V_{DD}, & V_{in} \geq V_{REF} \end{cases} \quad (1)$$

However, these voltage comparators can be implemented using the Josephson junction-based voltage comparator [162] and voltage comparators with the CMOS devices characterized at cryogenic temperature [30], [176], [201]. As shown in **Fig. 8.2**(b), the reference generator consists of two 4-to-1 multiplexers (MUX) which generate the two reference voltages for the voltage comparators based on the logic operation to perform (represented by 'opcode'). The table in **Fig. 8.2**(d) captures the conditions for choosing the values for the reference voltages to perform different operations (such as memory read, NAND, NOR, and XOR). The noteworthy feature of this sense amplifier design is that it can be configured for different logic operations only by setting suitable reference voltages.

8.3 Functional Verification

To simulate the read and in-memory logic operations, we use the same simulation framework discussed in Section 7.5 of Chapter 7 and simulate a 4 × 4 memory array with the MIEC-based selector, QAHE-based memory element, and the modified peripheral circuitry. Using this framework, we first simulate the memory read operation to demonstrate that the peripheral circuitry, modified for the logic operations, can also be used for the memory read operation. **Fig. 8.3**(a) shows the biasing scheme to access the

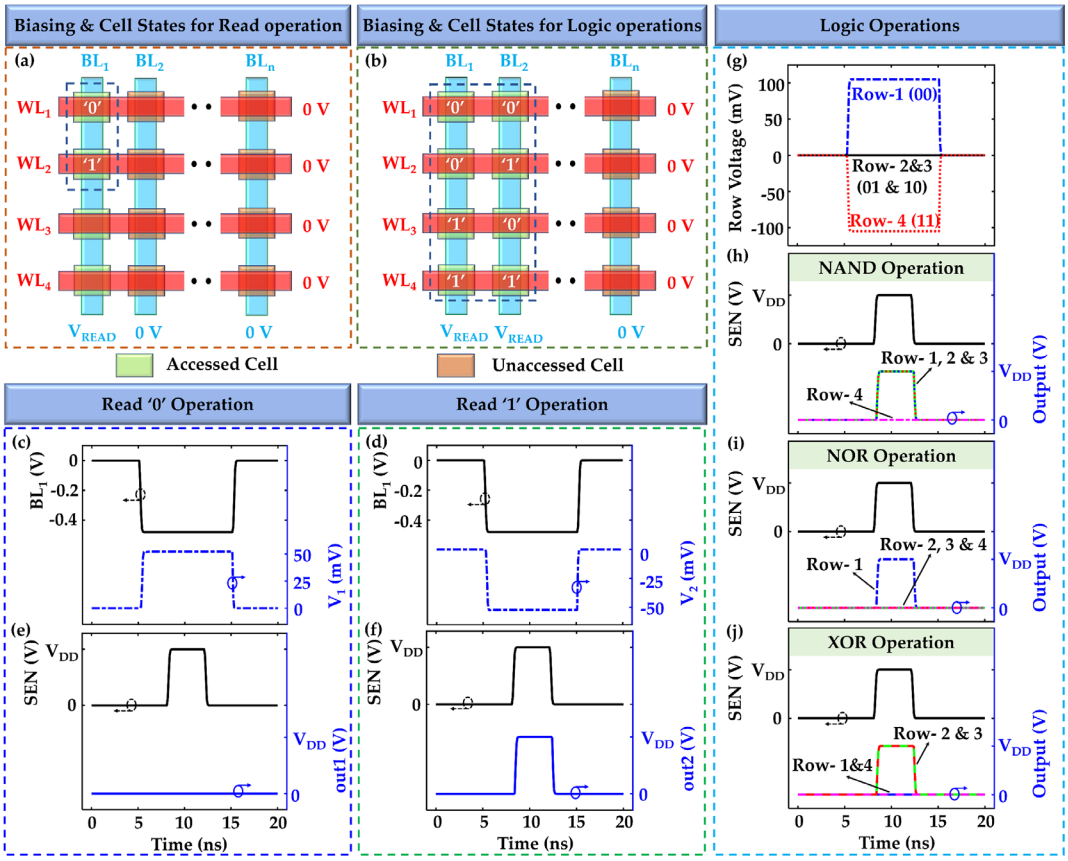


Figure 8.3: Biasing scheme for the WLs and BLs during **(a)** memory read operation and **(b)** logic operations. Voltage applied to BL_1 and the voltage generated at the output of the amplifiers during **(c)** read '0' and **(d)** read '1' operations. Time dynamics of the sense enable signal (SEN) and output of the sense amplifier during **(e)** read '0' and **(f)** read '1' operations. **(g)** Voltage at the outputs of the amplifiers for different logic combinations ('00', '01', '10', and '11') in the rows. Time dynamics of the sense enable signal (SEN) and output of the sense amplifier during **(h)** 2-input NAND, **(i)** 2-input NOR, and **(j)** 2-input XOR operations.

first column (keeping all the other columns unaccessed). To demonstrate the read operation for both logic states, we utilize the first two cells of the first column (logic '0' in M_{11} and logic '1' in M_{21}). **Figs. 8.3**(c) and (d) show the applied read voltage at BL_1 and the corresponding Hall voltage generated at the output of the amplifier during read '0' and read '1' operations, respectively. **Figs. 8.3**(e) and (f) show the proper read operation using the peripheral circuitry after activating the sense amplifier and choosing the suitable reference voltages for the read operation. We estimate an average of ~1.75 nW/cell programming power, and ~1 nW/cell read power. **Fig. 8.3**(b) shows the biasing scheme for the logic operations where we apply the read voltages to the first two columns and assume that the 1st, 2nd, 3rd, and 4th rows of the array store '00', '01', '10', and '11' logic conditions, respectively. **Fig. 8.3**(g) shows the voltage at the output of the amplifiers of the four rows where '00', '01'and '10', and '11' logic combinations entail $\sim$105 mV, 0 mV, and ~ -105 mV at the outputs of the amplifiers, respectively. The clear distinction in the voltage level corresponding to the logic combinations makes the logic operations and the design of peripheral circuitry simple. **Figs. 8.3**(h), (i), and (j) present the final outputs for all the logic combinations for NAND, NOR, and XOR operations, respectively. Irrespective of the type of logic operation, the average power remains ~2 nW per bit per operation (excluding peripheral power).

8.4 Variation Analysis

We also explore the effects of variation on the in-memory logic operations. In terms of variations, QAHE-based memories provide a clear advantage compared to other memories as shown in **Fig. 8.4**. In all the memories, there are two main sources of variation- (i) the input biases applied to write and read and (ii) the two logic states. Variation from these two sources mostly affects the separation between the memory states which leads to the requirement of complex sensing circuitry and read error in the extreme case. Furthermore, since CiM operations are essentially a modified version of memory read accesses, a reduced sensing margin due to variations inevitably affects the robustness of CiM operations. Advantageously, in QAHE-based memories, the two Hall resistance states are topologically protected and consequently remain constant at $\pm h/e^2$ as long as the temperature remains at 4 K/below [137]. Therefore, there is only one major

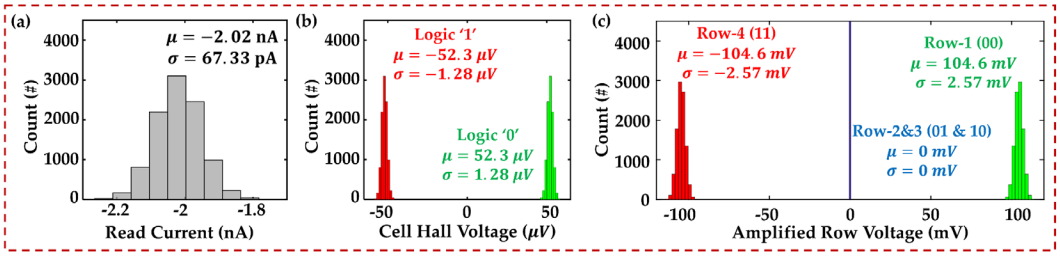


Figure 8.4: (a) Input distribution of the cell read current with a 10% variation of the nominal value (−2.02 nA) to run a 10000-point Monte-Carlo simulation. Distributions of **(b)** the cell Hall voltage and **(c)** the amplified row voltages (sum of active cell voltages) resulting from the variation considered in the read current shown in (a).

source of variation in the QAHE-based memories- the variation in the write or read bias. To demonstrate the effects of variation in our QAHE-based in-memory computing, we run a 10000-point Monte-Carlo simulation considering a 10% variation in the read current with its nominal value at -2.2 nA (**Fig. 8.4**(a)). The resulting distribution of the cell Hall voltage is shown in **Fig. 8.4**(b). Consequently, we can find the distributions for the amplified row voltages as depicted in **Fig. 8.4**(c) which clearly shows that there is no effect of read current variation on the 01/10 logic combinations because the resistance levels remain constant. On the other hand, for '00' and '11' logic combinations, we observe the effect of read current variation in the row voltages. Note that considering the variation, there is still a significant amount of separation between the voltage levels to set the reference voltages. Therefore, the outputs of the in-memory logic operations can be generated using a very simple sense amplifier with two reference voltages.

8.5 Summary

Cryogenic in-memory computing provides impressive energy efficiency that can solve the *memory wall* bottleneck for both conventional and quantum computing fields and the cooling issues of the cryogenic systems. Here, we have demonstrated a cryogenic CiM based on the QAHE observed in the tBLG moiré heterostructure. We have proposed modified peripheral circuitry with specific biasing schemes along with a reconfigurable sense amplifier for the memory read and in-memory logic operations such as NAND, NOR, and XOR. The sense amplifier can be configured (only suitable reference voltages

need to be chosen) to perform different in-memory operations in single-cycle. The QAHE-based CiM provides additional advantages such as- ultra-low-power requirement thanks to the nano-ampere level of switching current, better immunity to variations provided by the topologically protected Hall resistance states, and better scalability enabled by the small cell area and the simpler peripheral circuitry. Through this work, we commence an effort to embed CiM techniques in QAHE-based memory systems and possibly beyond.

CHAPTER 9
IN-MEMORY BIT-SERIAL ADDITION USING QUANTUM ANOMALOUS HALL EFFECT-BASED MAJORITY LOGIC

9.1 Introduction

As highlighted in Chapter 6, CiM architectures can solve both '*von Neumann*' and '*memory wall*' bottlenecks by processing data inside the memory array. However, unlike the CMOS implementation, CiM accommodates usually the homogeneity of logic gates (only NAND or only NOR) [120]. In this regard, in-memory Majority logic proves to be more efficient compared to other logic primitives since Majority along with NOT gate is considered to be functionally complete.

Majority logic is a Boolean logic primitive which generates an output of logic '1' when there is a greater number of logic '1' among the inputs. **Fig. 9.1**(a) shows the symbols and truth tables for 3-input and 5-input Majority logics. The logic expression of a 3-input Majority gate is-

$$Maj\ (a, b, c) = (a\&b)\ |\ (b\&c)\ |\ (c\&a) \quad (1)$$

where a, b, and c are the inputs. The idea of Majority logic was first put forward in 1960 [202] and several beyond CMOS devices such as spin waves [203], single electron tunneling [204], quantum dot cellular automata [205], nanomagnetic logic [206], superconducting Josephson junction [121], etc. have been used to implement Majority logic. The advantages of Majority logic are- (i) other Boolean logic gates such as AND and OR can be implemented using Majority (**Fig. 9.1**(b)), (ii) in some technologies (for example, quantum dot cellular automata), the implementation of Majority logic is simpler and more efficient compared with other logic primitives such as NAND, NOR, and XOR, and (iii) Majority logic enables the efficient implementation of arithmetic-intensive systems with lower number of gates [120], [203]–[205]. The utilization of Majority logic in arithmetic-intensive circuits reduces the logical depth up to 33% compared to the AND implementation [207].

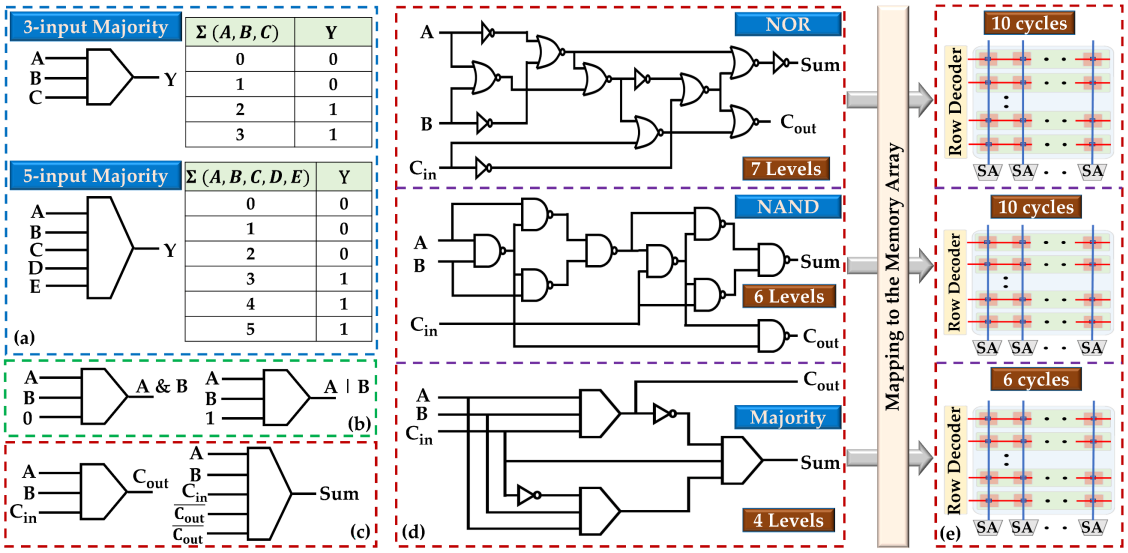


Figure 9.1: **(a)** Truth tables for 3-input and 5-input Majority logic. Implementation of **(b)** AND and OR gates, and **(c)** full adder using Majority logic. **(d)** Logic level and **(e)** in-memory implementation of full adder using NAND, NOR, and Majority. The use of Majority logic reduces the logical depth and number of cycles.

Usually in CiM architectures, one logic primitive (either NAND or NOR) is implemented and then, all the other logic functions are implemented based on that. Therefore, implementing an in-memory full adder requires the XOR gate to be implemented using either NAND or NOR which consequently increases the required logic levels. For example, the implementation of a XOR gate requires a sequence of four NAND operations. However, a 1-bit full adder can be implemented using one 3-input and one 5-input Majority gates (**Fig. 9.1**(c)) since the logic expression of a 1-bit full adder can be simplified to-

$$C_{out} = Maj\ (A, B, C_{in})$$

$$\text{and} \qquad Sum = Maj\ (A, B, C_{in}, \overline{C_{out}}, \overline{C_{out}}) \tag{2}$$

Fig. 9.1(d) shows the implementation of a 1-bit full adder using NAND, NOR, and Majority logic along with NOT gate which illustrates that the Majority gate-based implementation reduces the logic level by 50% (57.14%) compared to the NAND (NOR)-based implementations [120]. Now, when the full adders are mapped inside the memory system for in-memory implementation, the use of Majority logic reduces the

number of required cycles by 60% compared to the NAND and NOR-based implementations (**Fig. 9.1**(e)) [120].

In this chapter, we present a cryogenic in-memory Majority logic which is further used to design an in-memory full adder. Full adder is a critical component for the processing units in both room temperature and cryogenic electronics as it is required in the arithmetic logic unit, floating point unit, and address generation to access the memory unit [208]. Therefore, an efficient cryogenic full adder is required in the control processor of quantum computers, cryogenic electronics, and the processor used in the spacecrafts. We utilize the QAHE-based memory array (discussed in Chapter 7) to implement the in-memory Majority function and subsequently, a full adder. QAHE-based memory array shows two unique characteristics that benefits the implementation of in-memory Majority logic- (i) during read operation, we get positive (negative) Hall voltage (V_{xy}) across a cell for logic '0' ('1') state and (ii) V_{xy} of all the cells in a row are algebraically summed. These two features lead to positive (negative) voltage corresponding to the Majority '0' ('1') output irrespective of the number of inputs which dramatically simplifies the design of peripheral circuitry.

9.2 QAHE-based In-Memory Majority

We utilize the same QAHE-based crosspoint memory array of **Fig. 7.3** to implement in-memory Majority logic. To perform the Majority functionality, when odd number of cells are activated in a row, we get two distinguishable voltage levels (positive and negative) corresponding to the two logic outputs ('0' and '1', respectively). **Table 9.1** shows the truth table along with the corresponding voltage levels for 3-input and 5-input Majority. As evident from this table, the Majority function can be implemented in the QAHE-based memory using only a simple voltage comparator which is used for the memory read operation. This advantage is unique to the QAHE-based memory system and there are two major reasons- (i) V_{xy} with two opposite polarities for two memory states and (ii) adding the V_{xy} of all the cells in a row. Another major advantage of QAHE-based in-memory Majority is that we will get this same distinguishability (positive and

Table 9.1: Impmentation of 3-input and 5-input in-memory majority logic

	∑ Inputs	V_R (mV)	V_{out} (logic)	$V_{out} = \begin{cases} 0, & V_R > 0 \\ 1, & V_R < 0 \end{cases}$
3-input	0	150	0	
	1	50	0	
	2	-50	1	
	3	-150	1	
5-input	0	250	0	
	1	150	0	
	2	50	0	
	3	-50	1	
	4	-150	1	
	5	-250	1	

negative levels) in the voltage irrespective of the number of inputs (3, 5, 7, etc.). The cases of 3 and 5 inputs are shown in **Table 9.1**.

9.3 QAHE-based In-memory Adder

We implement the 1-bit full adder by implementing one 3-input and one 5-input Majority logic. **Fig. 9.2**(a) shows the array structure to calculate C_{out} where three QAHE cells in a row storing the three inputs (A, B, and C_{in}) are activated. In this step, we calculate and store both C_{out} and $\overline{C_{out}}$ since C_{out} will be used in the next bit addition as C_{in} and $\overline{C_{out}}$ is required to calculate *Sum* according to (2). Therefore, the sense amplifier should be able to calculate both out and $\overline{out}$ at the output. **Fig. 9.2**(b) shows the activation of five QAHE cells to calculate *Sum* where three cells store the three inputs and two cells store $\overline{C_{out}}$.

To verify the functionality of 1-bit full adder using the in-memory Majority logic with the QAHE-based memory system, we utilize the same simulation framework discussed in Section 7.5 of Chapter 7. The simulation results for all the possible input combinations are shown in **Fig. 9.3**. **Fig. 9.3**(a) shows the different levels of voltage at the output of the voltage amplifier after applying suitable I_{Bias} to the cells storing the inputs. Then, when we apply the enable signal and reference voltage to the sense amplifier, we get C_{out} and $\overline{C_{out}}$ at the output performing 3-input Majority (**Fig. 9.3**(b)). Then, we store C_{out} and

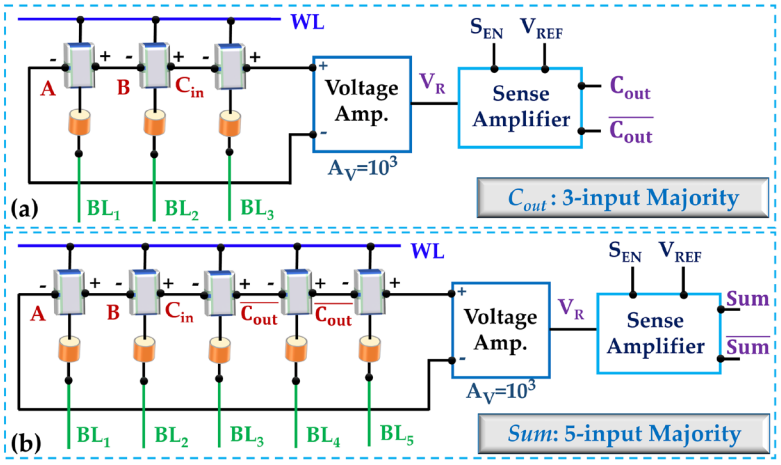


Figure 9.2: Calculation of **(a)** C_{out} and $\overline{C_{out}}$ and **(b)** Sum using 3-input and 5-input Majority logic, respectively.

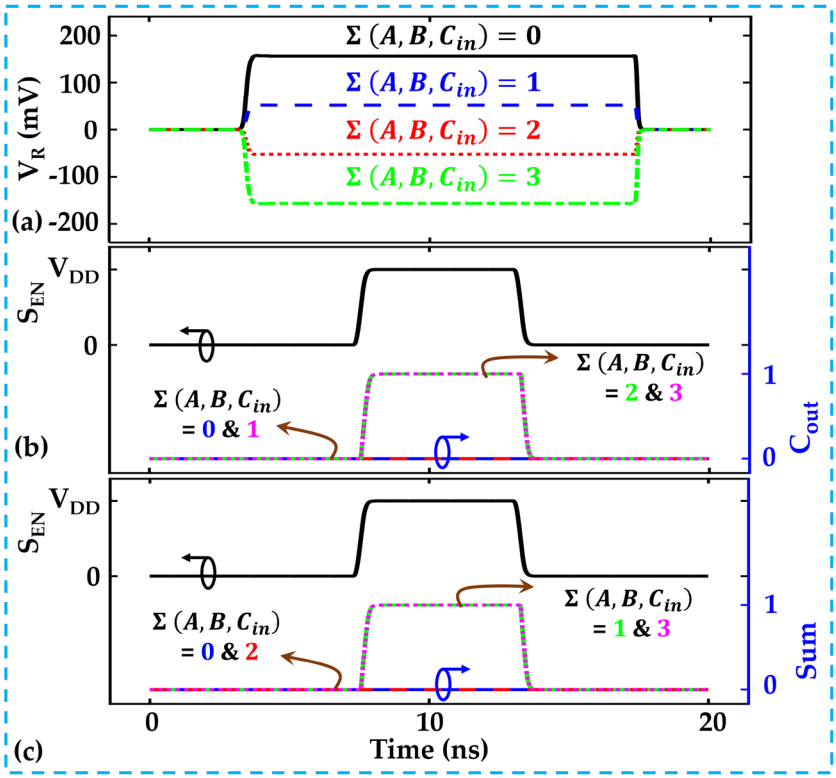


Figure 9.3: Simulation results for Majority-based in-memory 1-bit full adder.

$\overline{C_{out}}$, and in the next cycle, we activate five cells together to perform 5-input Majority logic to calculate *Sum* (**Fig. 9.3**(c)).

Finally, we show an example of performing bit-serial addition in QAHE-based memory array using the in-memory Majority logic (**Fig. 9.4**). We consider that each of the inputs (*A* and *B*) consists of eight bits and we perform the Majority functionality by activating different columns in the same row. Here, we divide the memory array into data columns to store the inputs and outputs, and compute columns to calculate C_{out} and Sum. We again divide the data columns to store the inputs (*A* in the first eight columns and *B* in the next eight columns) and the addition result (Sum in the next eight columns). We use the compute columns to store the copied versions of A and B, C_{in}, C_{out}, and $\overline{C_{out}}$. In the compute columns, we store both C_{out}, and $\overline{C_{out}}$ twice because we will need C_{out} for the next bit addition and two $\overline{C_{out}}$ to calculate Sum. In **Fig. 9.4**, we only show the addition of the first bits of A and B. **Fig. 9.4**(a) shows the initial state where the inputs (A, B, and C_{in}) are stored. In the next two cycles (**Figs. 9.4**(b) and (c), respectively), we copy $A[0]$ and $B[0]$ to the corresponding compute columns (labelled *A* and *B*, respectively). Afterwards, we calculate and store C_{out} and $\overline{C_{out}}$ by activating three compute columns storing $A[0]$, $B[0]$, and C_{in} (**Fig. 9.4**(d)). In the final step, we calculate Sum by activating five compute columns (**Fig. 9.4**(e)) and store the result in the designated location. Similarly, the addition of the next bit will be performed where C_{out} will be used as C_{in}. As shown in **Fig. 9.4**, the proposed adder adds an overhead of eight additional columns to perform the computation. Note, even if we consider larger input size (1024 or 2048 bits instead of 8) of the inputs for different applications, the number of additional columns required will still be eight due to the use of bit-serial addition mechanism [208]. Another major advantage of this proposed in-memory adder is that due to the use of QAHE-based memory array, no modification in the memory array and the sense amplifier is required whereas all the existing in-memory implementations of adder require some degree of modification in the peripheral circuitry. Moreover, in terms of variation tolerance, the use

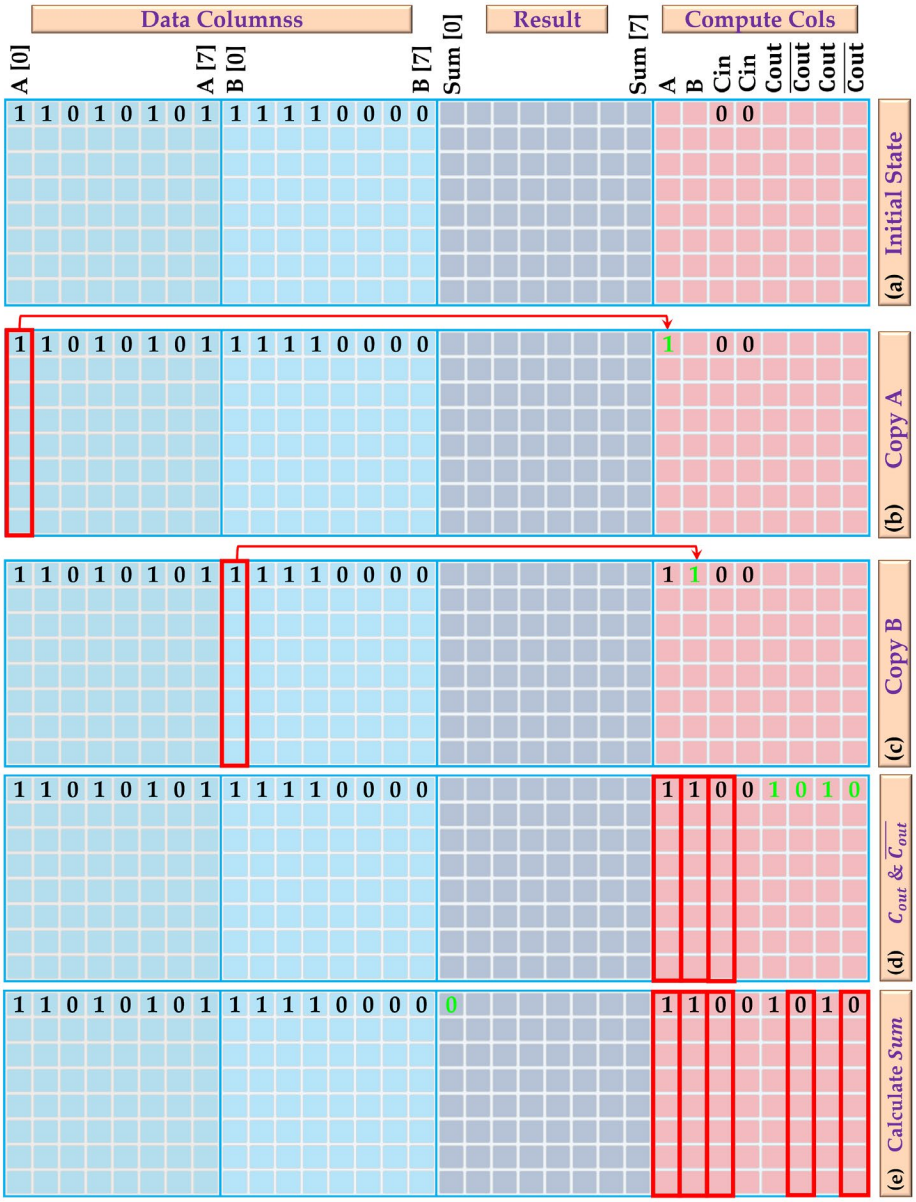


Figure 9.4: Steps of in-memory bit-serial addition in a QAHE-based memory system.

of QAHE-based memory array entails a major advantage because the two R_{xy} states are topologically protected and remain constant at $-h/e^2$ or $+h/e^2$ for any temperature below 4 K [48], [137].

9.4 Comparative Analysis

In this section, we compare our proposed QAHE-based in-memory full adder with NAND and NOR- based implementations. AdderNet [209], a deep neural network (DNN), shows that the computationally complex multiplication operations can be replaced with addition. AdderNet not only achieves same accuracy as the binary and convolutional neural networks but also improves latency because the latency of multiplication is higher than addition. Here, we show the superiority of our proposed QAHE-based in-memory addition hardware for implementing AdderNet. **Fig. 9.5**(a) shows the number of cycles required to implement different DNN models for different datasets with NAND-, NOR-, and the proposed QAHE-based in-memory adders. Here, VGG-Small, ResNet-20, and ResNet-32 models are used for CIFAR-10 and CIFAR-100 datasets, while ResNET-18 and ResNet-50 models are used for ImageNet datasets [209]. In all the cases, we show that the QAHE-based in-memory adder requires a significantly smaller number of cycles than NAND and NOR implementations.

Next, we show the advantage of using QAHE to implement the Majority logic, comparing with the memristor-based implementation. Memristors have already been utilized to develop non-volatile memory [210], CiM architecture [211], neuromorphic system [212], associative memory [213], [214], etc. at room temperature. However, we show that our proposed QAHE-based Majority logic can offer unique advantage compared to the memristor-based implementations. Here, we use the one-transistor one resistor (1T1R) memristive array framework from [210] to simulate the Majority logic. **Fig. 9.5**(b) shows that for 3-input Majority logic, the memristive memory provides a separation of around 68 μA (depends on the memristor and read voltage) for the sensing purpose. Note, all the conventional memory platforms (such as memristive, ferroelectric, spintronic, phase-change, etc.) will require differentiating between two current levels with same polarity (positive/negative) to detect the majority output. However, in our proposed QAHE-based majority logic, we get two voltage levels with opposite polarities

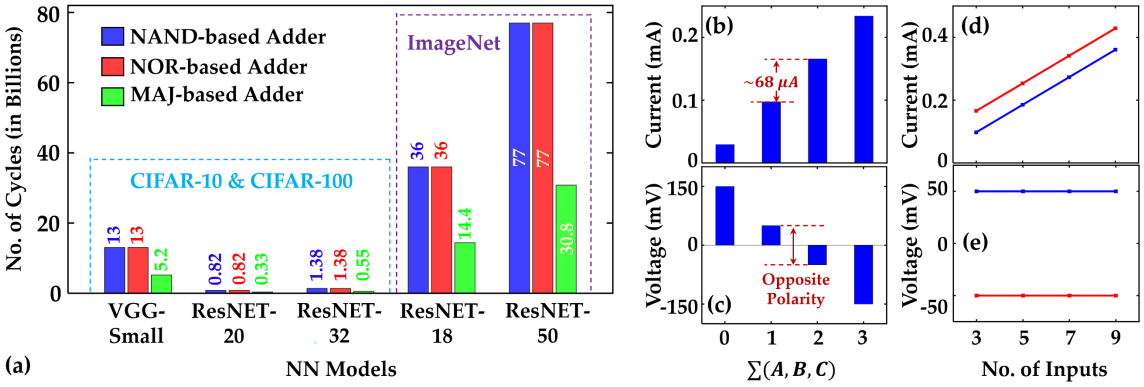


Figure 9.5: Comparison of our proposed QAHE-based in-memory majority logic with other approaches, highlighting the advantage of using QAHE.

(**Fig. 9.5**(c)). Moreover, for majority logic with higher inputs, it is seen that both the current levels (used to detect the majority output) increase (**Fig. 9.5**(d)) and require different reference levels to be set for the sensing operation. However, in the QAHE-based approach, we get voltage levels with opposite polarities and with same values (**Fig. 9.5**(e)) for any number of inputs. Therefore, the same sensing circuitry can be used to perform Majority logic with any number of inputs.

9.5 Summary

In-memory computing is envisioned to provide a unique advantage for cryogenic systems by improving the energy efficiency and consequently, reducing the cooling issue. Here, we proposed an in-memory adder based on Majority logic using the QAHE-based cryogenic memory. The QAHE-based memory allows us to implement in-memory Majority functionality without any modification in the peripheral circuitry. A simple voltage comparator can be used as the sense amplifier like the memory read operation. Finally, we utilize this in-memory Majority functionality to implement an in-memory adder which requires total $4n + 1$ cycles for inputs with n bits. This proposed QAHE-based in-memory adder can be a promising candidate to perform addition tasks in different cryogenic applications.

CHAPTER 10
CONCLUSION

10.1 Summary

This dissertation explored the utilization of two superconducting devices, namely FE-SQUID and *hTron*, and the unique QAHE phenomenon observed in topological materials for cryogenic applications. The scope of this dissertation spanned the applications of these promising superconducting and topological devices in designing cryogenic logic, memory, and in-memory computing systems. We utilized experimentally calibrated compact models to simulate and analyze the designed circuits and systems.

By combining the distinctive features of FE-SQUID and hTron, we developed a scalable cryogenic memory array that offers non-volatility, scalability, voltage-controlled operation, high energy-efficiency, and advantages of having separate read-write paths. This innovative memory design avoids the limitations of conventional superconducting memory topologies, such as inductive coupling and bulky peripheral circuits, thus presenting a promising solution for scalable cryogenic memory systems.

Furthermore, we introduced a voltage-controlled superconducting logic family based on FE-SQUIDs, which resolves issues such as single fanout and the need for additional current splitter circuits in existing current-controlled superconducting logic designs. Leveraging the interaction between FE-SQUIDs and hTrons, we achieved seamless cascadability and enhanced functionality in logic circuits.

We also developed a cryogenic in-memory computing system based on FE-SQUID technology which can perform matrix-vector multiplication with high energy efficiency. This can facilitate the energy efficient implementation of machine learning algorithms which are useful for control and error correction of qubits in quantum computing system. We performed different classification tasks with this system which provide better accuracy compared with the existing approaches.

Then, we modified the FE-SQUID-based memory system to design a unique cryogenic ternary content addressable memory which utilizes the unique characteristics of FE-SQUIDs. Firstly, the design TCAM provides binary decisions in the exact search mode for match and mismatch conditions, eliminating the need for additional complex

peripheral circuitry. Secondly, the TCAM call can also be programmed to calculate the Hamming distance to find out the nearest search.

We also demonstrated a single hTron-based reconfigurable logic circuit capable of executing various single- and multi–input logic operations, through suitable biasing conditions. Unlike conventional designs, our approach reduces the number of required devices in logic gates, leading to a more compact implementation of superconducting processing units. Additionally, by utilizing the same circuit for different logic operations, we introduce the possibility of developing a camouflaged processing unit with uniform layouts, mitigating the risk of reverse engineering. This innovation holds significant promise for advancing the efficiency and security of cryogenic computing systems.

Additionally, we leveraged the QAHE phenomenon in topological devices to design an ultra-low power, non-volatile memory system compatible with cryogenic temperatures. This QAHE-based memory device offers significant reductions in cell area and read/write power compared to conventional cryogenic memory devices, thus presenting a compelling solution for cryogenic memory systems.

Furthermore, we demonstrated the implementation of a cryogenic in-memory computing architecture based on QAHE, showcasing its ultra-low power requirements, immunity to variations, and scalability advantages. We also proposed an in-memory adder based on majority logic, utilizing QAHE-based cryogenic memory, which offers a promising solution for addition tasks in various cryogenic applications.

In summary, the findings presented in this dissertation represent significant advancements in cryogenic computing, offering innovative solutions to challenges faced by existing technologies. These contributions lay the groundwork for the development of scalable, energy-efficient, and high-performance cryogenic computing platforms, with potential applications in quantum computing, high-performance computing, and aerospace exploration.

10.2 Future Works

10.2.1 hTron-based Current-Splitter Circuitry

One of the major limitations of the existing superconducting logic circuits is that their operation depends on current bias. Therefore, these current-controlled logic circuits are

limited to single fanout because if one output current is connected to multiple inputs in the next stage, the output current will be divided and will hamper the functionality of the next stage logic gates. To achieve higher fanout, a current splitter circuit is used. The current splitter circuit takes a current as input and generates multiple current outputs with the same magnitude of the input current. This provides seamless fanout and cascadability for the current-controlled circuits.

The current splitter circuit is nothing but a combination of multiple buffer or copy gates. The number of buffers determines the number of fanout. Also, the generation of multiple current signals from one input with the same magnitude requires careful design of the buffer circuit. Our designed hTron-based reconfigurable logic circuit (discussed in Chapter 6) can be used to design a current splitter circuit. The high area efficiency of the designed reconfigurable circuit will improve the area efficiency of the current splitter circuitry, and in turn, will solve the scalability issues of the existing current-controlled superconducting logic circuits.

10.2.2 Superconducting Neuromorphic Computing with FE-SQUID and hTron

In the future, the use of superconducting devices such as FE-SQUIDs and hTrons represents a promising avenue for the construction of neuromorphic computing architectures. The extreme energy efficiency of superconducting devices and circuits will help us develop an efficient and scalable computing system like human brain. These devices offer a synergistic combination of non-volatility, ultra-low power dissipation, and high-speed operation – making them suitable for energy-efficient, cryogenic neuromorphic computing.

FE-SQUIDs are tunable synaptic units due to their hysteretic nature that is regulated by ferroelectric polarization states. Memory storage without the use of static power is aligned with the specifications for synaptic weight storage in spiking neural networks (SNNs). Furthermore, continuous tunability of the critical current using ferroelectric gating holds promises towards the use of analog or multi-level synapses that are critical to learning higher accuracy.

On the other hand, hTron—a current-controlled superconducting switch—can be used for neuron-like thresholding and firing dynamics. Its high-speed switching and single-

flux quantum (SFQ) pulse compatibility qualify it for realizing integrate-and-fire or leaky integrate-and-fire neuron models. When co-integrated with FE-SQUIDs, hTrons can support hybrid neural implementations where memory and weighting functions are done by FE-SQUIDs and spike generation and routing by hTrons.

10.2.3 Probabilistic Computing with hTron

Another promising direction is the application of hTron devices to probabilistic computing, a paradigm that is well adapted to tasks such as Bayesian inference, stochastic optimization, and machine learning. Due to the intrinsic thermal sensitivity and stochastic switching behavior of hTrons near their critical operating point, these devices can be engineered to function as hardware random number generators or probabilistic bit (p-bit) elements. By modulating the input current close to the switching threshold of the hTron, controlled randomness of its output can be triggered enabling natural integration of probabilistic logic gates and Boltzmann machines. Controlled randomness can be used to construct energy-efficient probabilistic circuits in the cryogenic regime with thermal noise and quantum fluctuations used as computational resources, rather than as a source of noise. Future research will include the creation of hTron-based probabilistic units, the determination of their tunability and reliability, and their integration into more complete stochastic computing architectures that are compatible with superconducting logic and memory technologies.

LIST OF REFERENCES

[1] H. Mujtaba, “Cerebras Wafer Scale Engine Is A Massive AI Chip Featuring 2.6 Trillion Transistors & Nearly 1 Million Cores.” [Online]. Available: https://wccftech.com/cerebras-unveils-7nm-wafe-scale-engine-2-largest-ai-chip-ever-built/. [Accessed: 08-Jul-2021].

[2] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, “Recalibrating global data center energy-use estimates,” *Science (80-.).*, vol. 367, no. 6481, pp. 984–986, 2020, doi: 10.1126/science.aba3758.

[3] K. K. Likharev and J. Lukens, “Dynamics of Josephson Junctions and Circuits,” *Phys. Today*, 1988, doi: 10.1063/1.2811641.

[4] O. A. Mukhanov, D. Kirichenko, I. V. Vernik, T. V. Filippov, A. Kirichenko, R. Webber, V. Dotsenko, A. Talalaevskii, J. C. Tang, A. Sahu, P. Shevchenko, R. Miller, S. B. Kaplan, S. Sarwana, and D. Gupta, “Superconductor digital-RF receiver systems,” *IEICE Trans. Electron.*, 2008, doi: 10.1093/ietele/e91-c.3.306.

[5] I. V. Vernik, D. E. Kirichenko, V. V. Dotsenko, R. Miller, R. J. Webber, P. Shevchenko, A. Talalaevskii, D. Gupta, and O. A. Mukhanov, “Cryocooled wideband digital channelizing radio-frequency receiver based on low-pass ADC,” in *Superconductor Science and Technology*, 2007, doi: 10.1088/0953-2048/20/11/S05.

[6] NSA, “Superconducting Technology Assessment,” 2005.

[7] M. B. Hastings, M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, “Improving quantum algorithms for quantum chemistry,” *Quantum Inf. Comput.*, 2014.

[8] P. W. Shor, “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer,” *SIAM J. Comput.*, 1997, doi: 10.1137/S0097539795293172.

[9] S. S. Tannu, D. M. Carmean, and M. K. Qureshi, “Cryogenic-DRAM based memory system for scalable quantum computers: A feasibility study,” in *ACM International Conference Proceeding Series*, 2017, doi: 10.1145/3132402.3132436.

[10] J. M. Hornibrook, J. I. Colless, I. D. C. Lamb, S. J. Pauka, H. Lu, A. C. Gossard, J.

D. Watson, G. C. Gardner, S. Fallahi, M. J. Manfra, and D. J. Reilly, “Cryogenic Control Architecture for Large-Scale Quantum Computing,” 2015, doi: 10.1103/PhysRevApplied.3.024010.

[11] S. S. Tannu, Z. A. Myers, P. J. Nair, D. M. Carmean, and M. K. Qureshi, “Taming the Instruction Bandwidth of Quantum Computers via Hardware-Managed Error Correction,” *Proc. 50th Annu. IEEE/ACM Int. Symp. Microarchitecture*, vol. 13, doi: 10.1145/3123939.

[12] B. Patra, R. M. Incandela, J. P. G. Van Dijk, H. A. R. Homulle, L. Song, M. Shahmohammadi, R. B. Staszewski, A. Vladimirescu, M. Babaie, F. Sebastiano, and E. Charbon, “Cryo-CMOS Circuits and Systems for Quantum Computing Applications,” *IEEE J. Solid-State Circuits*, 2018, doi: 10.1109/JSSC.2017.2737549.

[13] F. Ware, L. Gopalakrishnan, E. Linstadt, S. A. McKee, T. Vogelsang, K. L. Wright, C. Hampel, and G. Bronner, “Do superconducting processors really need cryogenic memories? The case for cold DRAM,” in *ACM International Conference Proceeding Series*, 2017, doi: 10.1145/3132402.3132424.

[14] M. Veldhorst, J. C. C. Hwang, C. H. Yang, A. W. Leenstra, B. De Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, “An addressable quantum dot qubit with fault-tolerant control-fidelity,” *Nat. Nanotechnol.*, 2014, doi: 10.1038/nnano.2014.216.

[15] M. Veldhorst, C. H. Yang, J. C. C. Hwang, W. Huang, J. P. Dehollain, J. T. Muhonen, S. Simmons, A. Laucht, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, “A two-qubit logic gate in silicon,” *Nature*, 2015, doi: 10.1038/nature15263.

[16] J. M. Chow, J. M. Gambetta, E. Magesan, D. W. Abraham, A. W. Cross, B. R. Johnson, N. A. Masluk, C. A. Ryan, J. A. Smolin, S. J. Srinivasan, and M. Steffen, “Implementing a strand of a scalable fault-tolerant quantum computing fabric,” *Nat. Commun.*, 2014, doi: 10.1038/ncomms5015.

[17] L. Dicarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf,

"Demonstration of two-qubit algorithms with a superconducting quantum processor," *Nature*, 2009, doi: 10.1038/nature08121.

[18] D. B. Tuckerman, M. C. Hamilton, D. J. Reilly, R. Bai, G. A. Hernandez, J. M. Hornibrook, J. A. Sellers, and C. D. Ellis, "Flexible superconducting Nb transmission lines on thin film polyimide for quantum computing applications," *Supercond. Sci. Technol.*, 2016, doi: 10.1088/0953-2048/29/8/084007.

[19] R. N. Das, V. Bolkhovsky, S. K. Tolpygo, P. Gouker, L. M. Johnson, E. A. Dauler, and M. A. Gouker, "Large Scale Cryogenic Integration Approach for Superconducting High-Performance Computing," *Proc. - Electron. Components Technol. Conf.*, pp. 675–683, Aug. 2017, doi: 10.1109/ECTC.2017.54.

[20] S. Long, Y. Li, J. Huang, Z. Li, and Y. Li, "A review of energy efficiency evaluation technologies in cloud data centers," *Energy Build.*, vol. 260, p. 111848, Apr. 2022, doi: 10.1016/J.ENBUILD.2022.111848.

[21] S. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin Uk, 2019.

[22] D. S. Holmes, A. L. Ripple, and M. A. Manheimer, "Energy-Efficient Superconducting Computing—Power Budgets and Requirements," *IEEE Trans. Appl. Supercond.*, vol. 23, no. 3, pp. 1701610–1701610, Feb. 2013, doi: 10.1109/tasc.2013.2244634.

[23] S. Alam, M. S. Hossain, S. R. Srinivasa, and A. Aziz, "Cryogenic memory technologies," *Nat. Electron. 2023 63*, vol. 6, no. 3, pp. 185–198, Mar. 2023, doi: 10.1038/s41928-023-00930-2.

[24] G. Krylov and E. G. Friedman, *Single Flux Quantum Integrated Circuit Design*. 2022.

[25] C. L. Ayala, T. Tanaka, R. Saito, M. Nozoe, N. Takeuchi, and N. Yoshikawa, "MANA: A Monolithic Adiabatic iNtegration Architecture Microprocessor Using 1.4-zJ/op Unshunted Superconductor Josephson Junction Devices," *IEEE J. Solid-State Circuits*, vol. 56, no. 4, pp. 1152–1165, Apr. 2021, doi: 10.1109/JSSC.2020.3041338.

[26] G. Leopold, "SpaceX Leverages HPC to Reach Orbit." [Online]. Available:

https://www.hpcwire.com/2015/01/02/spacex-leverages-hpc-reach-orbit/. [Accessed: 06-Apr-2024].

[27] J. Huang, R. Fu, X. Ye, and D. Fan, “A survey on superconducting computing technology: circuits, architectures and design tools,” *CCF Trans. High Perform. Comput.*, vol. 4, no. 1, pp. 1–22, Mar. 2022, doi: 10.1007/S42514-022-00089-W/METRICS.

[28] N. Yoshikawa, T. Tomida, M. Tokuda, Q. Liu, X. Meng, S. R. Whiteley, and T. Van Duzer, “Characterization of 4 K CMOS devices and circuits for hybrid Josephson-CMOS systems,” in *IEEE Transactions on Applied Superconductivity*, 2005, doi: 10.1109/TASC.2005.849786.

[29] W. H. Henkels, N. C. C. Lu, W. Hwang, T. V. Rajeevakumar, R. L. Franch, K. A. Jenkins, T. J. Bucelot, D. F. Heidel, and M. J. Immediato, “A low temperature 12 ns DRAM,” in *International Symposium on VLSI Technology, Systems and Applications*, pp. 32–35, doi: 10.1109/VTSA.1989.68576.

[30] E. Charbon, F. Sebastiano, A. Vladimirescu, H. Homulle, S. Visser, L. Song, and R. M. Incandela, “Cryo-CMOS for quantum computing,” *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 13.5.1-13.5.4, Jan. 2017, doi: 10.1109/IEDM.2016.7838410.

[31] P. R. Genssler, F. Klemme, S. S. Parihar, S. Brandhofer, G. Pahwa, I. Polian, Y. S. Chauhan, and H. Amrouch, “Cryogenic Embedded System to Support Quantum Computing: From 5-nm FinFET to Full Processor,” *IEEE Trans. Quantum Eng.*, vol. 4, 2023, doi: 10.1109/TQE.2023.3300833.

[32] Y. Polyakov, S. Narayana, and V. K. Semenov, “Flux trapping in superconducting circuits,” *IEEE Trans. Appl. Supercond.*, vol. 17, no. 2, pp. 520–525, Jun. 2007, doi: 10.1109/TASC.2007.898707.

[33] S. Narayana, Y. A. Polyakov, and V. K. Semenov, “Evaluation of flux trapping in superconducting circuits,” *IEEE Trans. Appl. Supercond.*, vol. 19, no. 3, pp. 640–643, Jun. 2009, doi: 10.1109/TASC.2009.2018248.

[34] K. Jackman and C. J. Fourie, “Flux Trapping Analysis in Superconducting Circuits,” *IEEE Trans. Appl. Supercond.*, vol. 27, no. 4, Jun. 2017, doi:

10.1109/TASC.2016.2642590.

[35] U. Ghoshal, H. Kroger, and T. Van Duzer, “Superconductor-Semiconductor Memories,” *IEEE Transactions on Applied Superconductivity*. 1993, doi: 10.1109/77.233542.

[36] M. Tanaka, M. Suzuki, G. Konno, Y. Ito, A. Fujimaki, and N. Yoshikawa, “Josephson-CMOS Hybrid Memory with Nanocryotrons,” *IEEE Trans. Appl. Supercond.*, vol. 27, no. 4, Jun. 2017, doi: 10.1109/TASC.2016.2646929.

[37] K. Kuwabara, H. Jin, Y. Yamanashi, and N. Yoshikawa, “Design and implementation of 64-kb CMOS static RAMs for Josephson-CMOS hybrid memories,” *IEEE Trans. Appl. Supercond.*, vol. 23, no. 3, 2013, doi: 10.1109/TASC.2012.2229331.

[38] J. B. Yau, Y. K. K. Fung, and G. W. Gibson, “Hybrid cryogenic memory cells for superconducting computing applications,” in *2017 IEEE International Conference on Rebooting Computing, ICRC 2017 - Proceedings*, 2017, doi: 10.1109/ICRC.2017.8123684.

[39] M.-H. Nguyen, G. J. Ribeill, M. V. Gustafsson, S. Shi, S. V. Aradhya, A. P. Wagner, L. M. Ranzani, L. Zhu, R. Baghdadi, B. Butters, E. Toomey, M. Colangelo, P. A. Truitt, A. Jafari-Salim, D. McAllister, D. Yohannes, S. R. Cheng, R. Lazarus, O. Mukhanov, K. K. Berggren, R. A. Buhrman, G. E. Rowlands, and T. A. Ohki, “Cryogenic Memory Architecture Integrating Spin Hall Effect based Magnetic Memory and Superconductive Cryotron Devices,” *Sci. Rep.*, 2020, doi: 10.1038/s41598-019-57137-9.

[40] H. Suzuki, A. Inoue, T. Imamura, and S. Hasuo, “Josephson driver to interface Josephson junctions to semiconductor transistors,” in *Technical Digest - International Electron Devices Meeting*, 1988, pp. 290–293, doi: 10.1109/iedm.1988.32814.

[41] U. Ghoshal, S. Kishore, A. Feldman, L. Huynh, and T. Van Duzer, “CMOS amplifier designs for Josephson-CMOS interface circuits,” *IEEE Trans. Appl. Supercond.*, vol. 5, no. 2, pp. 2640–2643, 1995.

[42] S. Alam, M. M. Islam, M. S. Hossain, K. Ni, V. Narayanan, and A. Aziz,

"Cryogenic Memory Array based on Ferroelectric SQUID and Heater Cryotron," *2022 Device Res. Conf.*, pp. 1–2, Jun. 2022, doi: 10.1109/DRC55272.2022.9855813.

[43] S. Alam, M. S. Hossain, K. Ni, V. Narayanan, and A. Aziz, "Voltage-controlled cryogenic Boolean logic gates based on ferroelectric SQUID and heater cryotron," *J. Appl. Phys.*, vol. 135, no. 1, p. 14903, Jan. 2024, doi: 10.1063/5.0172531/2932581.

[44] S. Alam, J. Hutchins, M. S. Hossain, K. Ni, V. Narayanan, and A. Aziz, "Cryogenic In-Memory Matrix-Vector Multiplication using Ferroelectric Superconducting Quantum Interference Device (FE-SQUID)," in *2023 60th ACM/IEEE Design Automation Conference (DAC)*, 2023, pp. 1–6, doi: 10.1109/DAC56929.2023.10247669.

[45] S. Alam, S. Thomann, S. S. Parihar, Y. S. Chauhan, K. Ni, H. Amrouch, and A. Aziz, "Energy-Efficient Cryogenic Ternary Content Addressable Memory using Ferroelectric SQUID," *ArXiV Prepr.*, Oct. 2024, doi: 10.48550/arXiv.2410.11091.

[46] S. Alam, D. S. Rampini, B. G. Oripov, A. N. McCaughan, and A. Aziz, "Cryogenic reconfigurable logic with superconducting heater cryotron: Enhancing area efficiency and enabling camouflaged processors," *Appl. Phys. Lett.*, vol. 123, no. 15, Oct. 2023, doi: 10.1063/5.0170187.

[47] S. Alam, M. S. Hossain, and A. Aziz, "A non-volatile cryogenic random-access memory based on the quantum anomalous Hall effect," *Sci. Rep.*, 2021, doi: 10.1038/s41598-021-87056-7.

[48] S. Alam, M. M. Islam, M. S. Hossain, A. Jaiswal, and A. Aziz, "CryoCiM: Cryogenic compute-in-memory based on the quantum anomalous Hall effect," *Appl. Phys. Lett.*, vol. 120, no. 14, p. 144102, Apr. 2022, doi: 10.1063/5.0092169.

[49] S. Alam, M. M. Islam, M. S. Hossain, A. Jaiswal, and A. Aziz, "Cryogenic In-Memory Bit-Serial Addition Using Quantum Anomalous Hall Effect-Based Majority Logic," *IEEE Access*, vol. 11, pp. 60717–60723, Jun. 2023, doi: 10.1109/ACCESS.2023.3285604.

[50] S. Tahara, I. Ishida, Y. Ajisawa, and Y. Wada, "Experimental vortex transitional

nondestructive read-out Josephson memory cell," *J. Appl. Phys.*, 1989, doi: 10.1063/1.343077.

[51] S. Alam, M. S. Hossain, and A. Aziz, "A cryogenic memory array based on superconducting memristors," *Appl. Phys. Lett.*, vol. 119, no. 8, p. 082602, Aug. 2021, doi: 10.1063/5.0060716.

[52] S. Alam, M. S. Hossain, and A. Aziz, "A non-volatile cryogenic random-access memory based on the quantum anomalous Hall effect," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, 2021, doi: 10.1038/s41598-021-87056-7.

[53] K. K. Likharev, "Superconductor digital electronics," *Phys. C Supercond. its Appl.*, vol. 482, pp. 6–18, Nov. 2012, doi: 10.1016/J.PHYSC.2012.05.016.

[54] D. Costanzo, H. Zhang, B. A. Reddy, H. Berger, and A. F. Morpurgo, "Tunnelling spectroscopy of gate-induced superconductivity in MoS2," *Nat. Nanotechnol. 2018 136*, vol. 13, no. 6, pp. 483–488, Apr. 2018, doi: 10.1038/s41565-018-0122-2.

[55] H. Zhang, C. Berthod, H. Berger, T. Giamarchi, and A. F. Morpurgo, "Band Filling and Cross Quantum Capacitance in Ion-Gated Semiconducting Transition Metal Dichalcogenide Monolayers," *Nano Lett.*, vol. 19, no. 12, pp. 8836–8845, Dec. 2019, doi: 10.1021/ACS.NANOLETT.9B03667.

[56] A. Chernyshov, M. Overby, X. Liu, J. K. Furdyna, Y. Lyanda-Geller, and L. P. Rokhinson, "Evidence for reversible control of magnetization in a ferromagnetic material by means of spin–orbit magnetic field," *Nat. Phys. 2009 59*, vol. 5, no. 9, pp. 656–659, Aug. 2009, doi: 10.1038/nphys1362.

[57] M. Liu, T. Nan, J. M. Hu, S. S. Zhao, Z. Zhou, C. Y. Wang, Z. De Jiang, W. Ren, Z. G. Ye, L. Q. Chen, and N. X. Sun, "Electrically controlled non-volatile switching of magnetism in multiferroic heterostructures via engineered ferroelastic domain states," *NPG Asia Mater. 2016 89*, vol. 8, no. 9, pp. e316–e316, Sep. 2016, doi: 10.1038/am.2016.139.

[58] H. M. Jafri, M. Sulaman, J. Wang, C. Yang, X. Shi, and H. Huang, "Phase-Field Simulation of Superconductor-Ferromagnet Bilayer-Based Cryogenic Strain Sensor," *J. Supercond. Nov. Magn.*, vol. 35, no. 2, pp. 409–414, Feb. 2022, doi:

10.1007/S10948-021-05959-Y/FIGURES/4.

[59] M. Suleiman, M. F. Sarott, M. Trassin, M. Badarne, and Y. Ivry, “Nonvolatile voltage-tunable ferroelectric-superconducting quantum interference memory devices,” *Appl. Phys. Lett.*, vol. 119, no. 11, p. 112601, Sep. 2021, doi: 10.1063/5.0061160.

[60] A. Crassous, R. Bernard, S. Fusil, K. Bouzehouane, D. Le Bourdais, S. Enouz-Vedrenne, J. Briatico, M. Bibes, A. Barthélémy, and J. E. Villegas, “Nanoscale electrostatic manipulation of magnetic flux quanta in ferroelectric/superconductor BiFeO 3/YBa 2Cu 3O 7-δ heterostructures,” *Phys. Rev. Lett.*, vol. 107, no. 24, p. 247002, Dec. 2011, doi: 10.1103/PHYSREVLETT.107.247002.

[61] A. Crassous, R. Bernard, S. Fusil, K. Bouzehouane, J. Briatico, M. Bibes, A. Barthélémy, and J. E. Villegas, “BiFeO3/YBa2Cu3O7−δ heterostructures for strong ferroelectric modulation of superconductivity,” *J. Appl. Phys.*, vol. 113, no. 2, p. 024910, Jan. 2013, doi: 10.1063/1.4774248.

[62] Q. Huang, Z. Chen, M. J. Cabral, F. Wang, S. Zhang, F. Li, Y. Li, S. P. Ringer, H. Luo, Y. W. Mai, and X. Liao, “Direct observation of nanoscale dynamics of ferroelectric degradation,” *Nat. Commun. 2021 121*, vol. 12, no. 1, pp. 1–7, Apr. 2021, doi: 10.1038/s41467-021-22355-1.

[63] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, “Theory of superconductivity,” *Phys. Rev.*, 1957, doi: 10.1103/PhysRev.108.1175.

[64] S. Alam, M. A. Jahangir, and A. Aziz, “A Compact Model for Superconductor-Insulator-Superconductor (SIS) Josephson Junctions,” *IEEE Electron Device Lett.*, vol. 41, no. 8, pp. 1249–1252, 2020, doi: 10.1109/LED.2020.3002448.

[65] V. Ambegaokar and A. Baratoff, “Tunneling between superconductors,” *Phys. Rev. Lett.*, 1963, doi: 10.1103/PhysRevLett.11.104.

[66] A. N. McCaughan, V. B. Verma, S. M. Buckley, J. P. Allmaras, A. G. Kozorezov, A. N. Tait, S. W. Nam, and J. M. Shainline, “A superconducting thermal switch with ultrahigh impedance for interfacing superconductors to semiconductors,” *Nat. Electron. 2019 210*, vol. 2, no. 10, pp. 451–456, Sep. 2019, doi: 10.1038/s41928-019-0300-8.

[67] S. Alam, M. S. Hossain, and A. Aziz, “A cryogenic memory array based on superconducting memristors,” *Appl. Phys. Lett.*, vol. 119, no. 8, p. 082602, Aug. 2021, doi: 10.1063/5.0060716.

[68] D. A. Buck, “The Cryotron—A Superconductive Computer Component,” *Proc. IRE*, vol. 44, no. 4, pp. 482–493, 1956, doi: 10.1109/JRPROC.1956.274927.

[69] M. F. Kessler, “The Infrared Space Observatory (ISO) mission,” *Adv. Sp. Res.*, 2002, doi: 10.1016/S0273-1177(02)00557-4.

[70] B. D. Josephson, “Possible new effects in superconductive tunnelling,” *Phys. Lett.*, 1962, doi: 10.1016/0031-9163(62)91369-0.

[71] J. Wang, J. Ma, H. Huang, J. Ma, H. M. Jafri, Y. Fan, H. Yang, Y. Wang, M. Chen, D. Liu, J. Zhang, Y. H. Lin, L. Q. Chen, D. Yi, and C. W. Nan, “Ferroelectric domain-wall logic units,” *Nat. Commun. 2022 131*, vol. 13, no. 1, pp. 1–8, Jun. 2022, doi: 10.1038/s41467-022-30983-4.

[72] S. B. Lee, G. D. Hutchinson, D. A. Williams, D. G. Hasko, and H. Ahmed, “Superconducting nanotransistor based digital logic gates,” *Nanotechnology*, vol. 14, no. 2, p. 188, Jan. 2003, doi: 10.1088/0957-4484/14/2/317.

[73] T. Akazaki, H. Takayanagi, J. Nitta, and T. Enoki, “A Josephson field effect transistor using an InAs-inserted-channel In0.52Al0.48As/In0.53Ga0.47As inverted modulation-doped structure,” *Appl. Phys. Lett.*, vol. 68, no. 3, p. 418, Aug. 1998, doi: 10.1063/1.116704.

[74] F. Wen, J. Shabani, and E. Tutuc, “Josephson Junction Field-Effect Transistors for Boolean Logic Cryogenic Applications,” *IEEE Trans. Electron Devices*, vol. 66, no. 12, pp. 5367–5374, Dec. 2019, doi: 10.1109/TED.2019.2951634.

[75] S. Alam, M. M. Islam, M. S. Hossain, and A. Aziz, “Superconducting Josephson Junction FET-based Cryogenic Voltage Sense Amplifier,” *2022 Device Res. Conf.*, pp. 1–2, Jun. 2022, doi: 10.1109/DRC55272.2022.9855654.

[76] R. C. Jaklevic, J. Lambe, A. H. Silver, and J. E. Mercereau, “Quantum interference effects in Josephson tunneling,” *Phys. Rev. Lett.*, 1964, doi: 10.1103/PhysRevLett.12.159.

[77] G. De Simoni, F. Paolucci, P. Solinas, E. Strambini, and F. Giazotto, “Metallic

supercurrent field-effect transistor," *Nat. Nanotechnol.*, vol. 13, pp. 802–805, 2018, doi: 10.1038/s41565-018-0190-3.

[78] S. Alam, A. N. McCaughan, and A. Aziz, "Reconfigurable Superconducting Logic Using Multi-Gate Switching of a Nano-Cryotron," *2023 Device Res. Conf.*, pp. 1–2, Jun. 2023, doi: 10.1109/DRC58590.2023.10186942.

[79] A. N. McCaughan and K. K. Berggren, "A superconducting-nanowire three-terminal electrothermal device," *Nano Lett.*, 2014, doi: 10.1021/nl502629x.

[80] A. Govindankutty, S. Alam, S. Das, N. Challapalle, A. Aziz, and S. George, "Ternary In-Memory Computing with Cryogenic Quantum Anomalous Hall Effect Memories," in *Proceedings of the Great Lakes Symposium on VLSI*, 2023, pp. 521–526, doi: 10.1145/3583781.3590236.

[81] A. Govindankutty, S. Alam, S. Das, A. Aziz, and S. George, "Cryogenic In-memory Binary Multiplier Using Quantum Anomalous Hall Effect Memories," *Proc. - Int. Symp. Qual. Electron. Des. ISQED*, vol. 2023-April, 2023, doi: 10.1109/ISQED57927.2023.10129345.

[82] W. Mayer, J. Yuan, K. S. Wickramasinghe, T. Nguyen, M. C. Dartiailh, and J. Shabani, "Superconducting proximity effect in epitaxial Al-InAs heterostructures," *Appl. Phys. Lett.*, vol. 114, no. 10, p. 103104, Mar. 2019, doi: 10.1063/1.5067363.

[83] M. Rocci, G. De Simoni, F. Giazotto, C. Puglia, D. D. Esposti, E. Strambini, V. Zannier, and L. Sorba, "Gate-controlled suspended titanium nanobridge supercurrent transistor," *ACS Nano*, vol. 14, no. 10, pp. 12621–12628, 2020, doi: 10.1021/acsnano.0c05355.

[84] K. Kang, D. Minn, S. Bae, J. Lee, S. Kang, M. Lee, H. J. Song, and J. Y. Sim, "A 40-nm Cryo-CMOS Quantum Controller IC for Superconducting Qubit," *IEEE J. Solid-State Circuits*, vol. 57, no. 11, pp. 3274–3287, Nov. 2022, doi: 10.1109/JSSC.2022.3198663.

[85] S. Resch, H. Cilasun, and U. R. Karpuzcu, "Cryogenic PIM: Challenges Opportunities," *IEEE Comput. Archit. Lett.*, vol. 20, no. 1, pp. 74–77, Jan. 2021, doi: 10.1109/LCA.2021.3077536.

[86] P. Wang, X. Peng, W. Chakraborty, A. I. Khan, S. Datta, and S. Yu, "Cryogenic

benchmarks of embedded memory technologies for recurrent neural network based quantum error correction," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 2020-December, pp. 38.5.1-38.5.4, Dec. 2020, doi: 10.1109/IEDM13553.2020.9371912.

[87] X. M. Zhang, Z. Wei, R. Asad, X. C. Yang, and X. Wang, "When does reinforcement learning stand out in quantum control? A comparative study on state preparation," *npj Quantum Inf. 2019 51*, vol. 5, no. 1, pp. 1–7, Oct. 2019, doi: 10.1038/s41534-019-0201-8.

[88] Y. Liu, A. Lee, K. Qian, P. Zhang, H. He, Z. Ren, S. K. Cheung, Y. Li, X. Zhang, Z. Ma, Z. Xiao, G. Yu, X. Wang, J. Liu, Z. Wang, K. L. Wang, and Q. Shao, "Cryogenic in-memory computing using tunable chiral edge states," *ArXiV Prepr.*, Sep. 2022, doi: 10.48550/arxiv.2209.09443.

[89] L. Deng, "The MNIST database of handwritten digit images for machine learning research," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, 2012, doi: 10.1109/MSP.2012.2211477.

[90] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms," *ArXiV Prepr.*, Aug. 2017, doi: 10.48550/arxiv.1708.07747.

[91] E. Saravia, H. C. Toby Liu, Y. H. Huang, J. Wu, and Y. S. Chen, "CARER: Contextualized Affect Representations for Emotion Recognition," *Proc. 2018 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2018*, pp. 3687–3697, 2018, doi: 10.18653/V1/D18-1404.

[92] S. I. Gallant, "Perceptron-Based Learning Algorithms," *IEEE Trans. Neural Networks*, vol. 1, no. 2, pp. 179–191, 1990, doi: 10.1109/72.80230.

[93] A. Asuncion and D. Newman, "UCI machine learning repository," *ArXiV Prepr.*, 2019, doi: 10.48550/arXiv.1910.03771.

[94] S. ichi Amari, "Backpropagation and stochastic gradient descent method," *Neurocomputing*, vol. 5, no. 4–5, pp. 185–196, Jun. 1993, doi: 10.1016/0925-2312(93)90006-O.

[95] M. A. Z. Dippé and E. H. Wold, "Antialiasing through stochastic sampling," vol.

19, no. 3, pp. 69–78, 1985, doi: 10.1145/325334.325182.

[96] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” *ArXiV Prepr.*, Oct. 2020, doi: 10.48550/arxiv.1910.01108.

[97] “HuggingFace’s Transformers: State-of-the-art Natural Language Processing,” *ArXiV Prepr.*, Oct. 2019, doi: 10.48550/arxiv.1910.03771.

[98] F. Liu, C. Liu, and F. Bi, “A Memristor based Unsupervised Neuromorphic System Towards Fast and Energy-Efficient GAN,” May 2018, doi: 10.48550/arxiv.1806.01775.

[99] M. Shahsavari, P. Falez, and P. Boulet, “Combining a volatile and nonvolatile memristor in artificial synapse to improve learning in Spiking Neural Networks,” *Proc. 2016 IEEE/ACM Int. Symp. Nanoscale Archit. NANOARCH 2016*, pp. 67–72, Sep. 2016, doi: 10.1145/2950067.2950090.

[100] Y. Wei, Q. Duan, R. Yuan, X. Yan, and Y. Yang, “Dropout neuronal unit with tunable probability based on NbOx stochastic memristor for efficient suppression of overfitting,” *Microelectron. Eng.*, vol. 259, p. 111778, Apr. 2022, doi: 10.1016/J.MEE.2022.111778.

[101] P. Zhou, J. K. Eshraghian, D.-U. Choi, and S.-M. Kang, “SPICEprop: Backpropagating Errors Through Memristive Spiking Neural Networks,” Mar. 2022, doi: 10.48550/arxiv.2203.01426.

[102] S. Wen, H. Wei, Y. Yang, Z. Guo, Z. Zeng, T. Huang, and Y. Chen, “Memristive LSTM Network for Sentiment Analysis,” *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 51, no. 3, pp. 1794–1804, Mar. 2021, doi: 10.1109/TSMC.2019.2906098.

[103] C. Chen, H. Ding, H. Peng, H. Zhu, Y. Wang, and C. J. R. Shi, “OCEAN: An On-Chip Incremental-Learning Enhanced Artificial Neural Network Processor with Multiple Gated-Recurrent-Unit Accelerators,” *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 8, no. 3, pp. 519–530, Sep. 2018, doi: 10.1109/JETCAS.2018.2852780.

[104] S. K. Tolpygo, V. Bolkhovsky, T. J. Weir, L. M. Johnson, M. A. Gouker, and W. D. Oliver, “Fabrication process and properties of fully-planarized deep-submicron Nb/Al-AlOx/Nb Josephson junctions for VLSI circuits,” *IEEE Trans. Appl.*

Supercond., vol. 25, no. 3, Jun. 2015, doi: 10.1109/TASC.2014.2374836.

[105] M. M. Islam, J. Hutchins, S. Alam, M. S. Hossain, A. Jaiswal, and A. Aziz, "Quantum Anomalous Hall Effect-Based Variation Robust Binary Content Addressable Memory," *Midwest Symp. Circuits Syst.*, pp. 331–335, 2023, doi: 10.1109/MWSCAS57524.2023.10406068.

[106] K. Ni, X. Yin, A. F. Laguna, S. Joshi, S. Dünkel, M. Trentzsch, J. Müeller, S. Beyer, M. Niemier, X. S. Hu, and S. Datta, "Ferroelectric ternary content-addressable memory for one-shot learning," *Nat. Electron. 2019 211*, vol. 2, no. 11, pp. 521–529, Nov. 2019, doi: 10.1038/s41928-019-0321-3.

[107] A. Rahimi, P. Kanerva, and J. M. Rabaey, "A Robust and Energy-Efficient Classifier Using Brain-Inspired Hyperdimensional Computing," *Proc. Int. Symp. Low Power Electron. Des.*, pp. 64–69, Aug. 2016, doi: 10.1145/2934583.2934624.

[108] D. Kleyko, E. Osipov, A. Senior, A. I. Khan, and Y. A. Şekercioğlu, "Holographic Graph Neuron: A Bioinspired Architecture for Pattern Processing," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 6, pp. 1250–1252, Jun. 2017, doi: 10.1109/TNNLS.2016.2535338.

[109] P. R. Genssler and H. Amrouch, "Brain-Inspired Computing for Wafer Map Defect Pattern Classification," *Proc. - Int. Test Conf.*, pp. 123–132, 2021, doi: 10.1109/ITC50571.2021.00020.

[110] S. Thomann, P. R. Genssler, and H. Amrouch, "HW/SW Co-Design for Reliable TCAM- Based In-Memory Brain-Inspired Hyperdimensional Computing," *IEEE Trans. Comput.*, vol. 72, no. 8, pp. 2404–2417, Aug. 2023, doi: 10.1109/TC.2023.3248286.

[111] W. Chen, A. V. Rylyakov, V. Patel, J. E. Lukens, and K. K. Likharev, "Rapid single flux quantum t-flip flop operating up to 770 GHz," *IEEE Trans. Appl. Supercond.*, vol. 9, no. 2 PART 3, pp. 3212–3215, 1999, doi: 10.1109/77.783712.

[112] S. Alam, M. S. Hossain, K. Ni, V. Narayanan, and A. Aziz, "Voltage-controlled Cryogenic Boolean Logic Family Based on Ferroelectric SQUID," *ArXiV Prepr.*, Dec. 2022, doi: 10.48550/arxiv.2212.08202.

[113] G. H. Lee, D. Min, I. Byun, and J. Kim, "Cryogenic computer architecture

modeling with memory-side case studies," *Proc. - Int. Symp. Comput. Archit.*, pp. 774–787, Jun. 2019, doi: 10.1145/3307650.3322219.

[114] M. M. Islam, S. Alam, M. S. Hossain, and A. Aziz, "Dynamically Reconfigurable Cryogenic Spiking Neuron based on Superconducting Memristor," *Proc. IEEE Conf. Nanotechnol.*, vol. 2022-July, pp. 307–310, 2022, doi: 10.1109/NANO54668.2022.9928634.

[115] J. M. Shainline, S. M. Buckley, A. N. Mccaughan, J. T. Chiles, A. Jafari Salim, M. Castellanos-Beltran, C. A. Donnelly, M. L. Schneider, R. P. Mirin, and S. W. Nam, "Superconducting optoelectronic loop neurons," *J. Appl. Phys.*, vol. 126, no. 4, 2019, doi: 10.1063/1.5096403.

[116] M. M. Islam, S. Alam, M. S. Hossain, K. Roy, and A. Aziz, "A review of cryogenic neuromorphic hardware," *J. Appl. Phys.*, vol. 133, no. 7, p. 70701, Feb. 2023, doi: 10.1063/5.0133515/2876598.

[117] M. M. Islam, S. Alam, M. R. I. Udoy, M. S. Hossain, and A. Aziz, "A Cryogenic Artificial Synapse based on Superconducting Memristor," 2023, pp. 143–148, doi: 10.1145/3583781.3590203.

[118] T. Jabbari, G. Krylov, and E. G. Friedman, "Logic Locking in Single Flux Quantum Circuits," *IEEE Trans. Appl. Supercond.*, vol. 31, no. 5, Aug. 2021, doi: 10.1109/TASC.2021.3065301.

[119] H. Kumar, T. Jabbari, G. Krylov, K. Basu, E. G. Friedman, and R. Karri, "Toward Increasing the Difficulty of Reverse Engineering of RSFQ Circuits," *IEEE Trans. Appl. Supercond.*, vol. 30, no. 3, Apr. 2020, doi: 10.1109/TASC.2019.2901895.

[120] J. Reuben, "Rediscovering Majority Logic in the Post-CMOS Era: A Perspective from In-Memory Computing," *J. Low Power Electron. Appl. 2020, Vol. 10, Page 28*, vol. 10, no. 3, p. 28, Sep. 2020, doi: 10.3390/JLPEA10030028.

[121] N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Reversible logic gate using adiabatic superconducting devices," *Sci. Reports 2014 41*, vol. 4, no. 1, pp. 1–4, Sep. 2014, doi: 10.1038/srep06354.

[122] A. J. Kerman, J. K. W. Yang, R. J. Molnar, E. A. Dauler, and K. K. Berggren, "Electrothermal feedback in superconducting nanowire single-photon detectors,"

Phys. Rev. B - Condens. Matter Mater. Phys., vol. 79, no. 10, p. 100509, Mar. 2009, doi: 10.1103/PHYSREVB.79.100509.

[123] F. D. M. Haldane, "Model for a quantum hall effect without landau levels: Condensed-matter realization of the 'parity anomaly,'" *Phys. Rev. Lett.*, 1988, doi: 10.1103/PhysRevLett.61.2015.

[124] C.-Z. Chang, J. Zhang, X. Feng, J. Shen, Z. Zhang, M. Guo, K. Li, Y. Ou, P. Wei, L.-L. Wang, Z.-Q. Ji, Y. Feng, S. Ji, X. Chen, J. Jia, X. Dai, Z. Fang, S.-C. Zhang, K. He, Y. Wang, L. Lu, X.-C. Ma, and Q.-K. Xu "Experimental observation of the quantum anomalous Hall effect in a magnetic topological Insulator," *Science*, 2013, doi: 10.1126/science.1234414.

[125] J. G. Checkelsky, R. Yoshimi, A. Tsukazaki, K. S. Takahashi, Y. Kozuka, J. Falson, M. Kawasaki, and Y. Tokura, "Trajectory of the anomalous Hall effect towards the quantized state in a ferromagnetic topological insulator," *Nat. Phys.*, 2014, doi: 10.1038/nphys3053.

[126] X. Kou, S. T. Guo, Y. Fan, L. Pan, M. Lang, Y. Jiang, Q. Shao, T. Nie, K. Murata, J. Tang, Y. Wang, L. He, T. K. Lee, W. L. Lee, and K. L. Wang, "Scale-invariant quantum anomalous hall effect in magnetic topological insulators beyond the two-dimensional limit," *Phys. Rev. Lett.*, 2014, doi: 10.1103/PhysRevLett.113.137201.

[127] A. J. Bestwick, E. J. Fox, X. Kou, L. Pan, K. L. Wang, and D. Goldhaber-Gordon, "Precise quantization of the anomalous hall effect near zero magnetic field," *Phys. Rev. Lett.*, 2015, doi: 10.1103/PhysRevLett.114.187201.

[128] X. Kou, L. Pan, J. Wang, Y. Fan, E. S. Choi, W. L. Lee, T. Nie, K. Murata, Q. Shao, S. C. Zhang, and K. L. Wang, "Metal-to-insulator switching in quantum anomalous Hall states," *Nat. Commun.*, 2015, doi: 10.1038/ncomms9474.

[129] Y. Feng, X. Feng, Y. Ou, J. Wang, C. Liu, L. Zhang, D. Zhao, G. Jiang, S. C. Zhang, K. He, X. Ma, Q. K. Xue, and Y. Wang, "Observation of the Zero Hall Plateau in a Quantum Anomalous Hall Insulator," *Phys. Rev. Lett.*, 2015, doi: 10.1103/PhysRevLett.115.126801.

[130] C. Z. Chang, W. Zhao, D. Y. Kim, H. Zhang, B. A. Assaf, D. Heiman, S. C. Zhang, C. Liu, M. H. W. Chan, and J. S. Moodera, "High-precision realization of

robust quantum anomalous Hall state in a hard ferromagnetic topological insulator," *Nat. Mater.*, 2015, doi: 10.1038/nmat4204.

[131] A. Kandala, A. Richardella, S. Kempinger, C. X. Liu, and N. Samarth, "Giant anisotropic magnetoresistance in a quantum anomalous Hall insulator," *Nat. Commun.*, 2015, doi: 10.1038/ncomms8434.

[132] S. Oh, "The complete quantum hall trio," *Science*. 2013, doi: 10.1126/science.1237215.

[133] M. B. A. Jalil, S. G. Tan, and Z. B. Siu, "Quantum anomalous Hall effect in topological insulator memory," *J. Appl. Phys.*, 2015, doi: 10.1063/1.4916999.

[134] M. Götz, K. M. Fijalkowski, E. Pesel, M. Hartl, S. Schreyeck, M. Winnerlein, S. Grauer, H. Scherer, K. Brunner, C. Gould, F. J. Ahlers, and L. W. Molenkamp, "Precision measurement of the quantized anomalous Hall resistance at zero magnetic field," *Appl. Phys. Lett.*, 2018, doi: 10.1063/1.5009718.

[135] B. Lian, X. Q. Sun, A. Vaezi, X. L. Qib, and S. C. Zhang, "Topological quantum computation based on chiral Majorana fermions," *Proc. Natl. Acad. Sci. U. S. A.*, 2018, doi: 10.1073/pnas.1810003115.

[136] A. L. Sharpe, E. J. Fox, A. W. Barnard, J. Finney, K. Watanabe, T. Taniguchi, M. A. Kastner, and D. Goldhaber-Gordon, "Emergent ferromagnetism near three-quarters filling in twisted bilayer graphene," *Science (80-.).*, 2019, doi: 10.1126/science.aaw3780.

[137] M. Serlin, C. L. Tschirhart, H. Polshyn, Y. Zhang, J. Zhu, K. Watanabe, T. Taniguchi, L. Balents, and A. F. Young, "Intrinsic quantized anomalous Hall effect in a moiré heterostructure," *Science (80-.).*, vol. 367, no. 6480, pp. 900–903, Feb. 2020, doi: 10.1126/science.aay5533.

[138] E. O. Lachman, A. F. Young, A. Richardella, J. Cuppens, H. R. Naren, Y. Anahory, A. Y. Meltzer, A. Kandala, S. Kempinger, Y. Myasoedov, M. E. Huber, N. Samarth, and E. Zeldov, "Visualization of superparamagnetic dynamics in magnetic topological insulators," *Sci. Adv.*, 2015, doi: 10.1126/sciadv.1500740.

[139] I. Lee, C. K. Kim, J. Lee, S. J. L. Billinge, R. Zhong, J. A. Schneeloch, T. Liu, T. Valla, J. M. Tranquada, G. Gu, and J. C. S. Davis, "Imaging Dirac-mass disorder

from magnetic dopant atoms in the ferromagnetic topological insulator Crx(Bi0.1Sb0.9)2-xTe3," *Proc. Natl. Acad. Sci. U. S. A.*, 2015, doi: 10.1073/pnas.1424322112.

[140] W. Wang, F. Yang, C. Gao, J. Jia, G. D. Gu, and W. Wu, "Visualizing ferromagnetic domains in magnetic topological insulators," *APL Mater.*, 2015, doi: 10.1063/1.4921093.

[141] K. Yasuda, M. Mogi, R. Yoshimi, A. Tsukazaki, K. S. Takahashi, M. Kawasaki, F. Kagawa, and Y. Tokura, "Quantized chiral edge conduction on domain walls of a magnetic topological insulator," *Science (80-.).*, 2017, doi: 10.1126/science.aan5991.

[142] Y. Deng, Y. Yu, M. Z. Shi, Z. Guo, Z. Xu, J. Wang, X. H. Chen, and Y. Zhang, "Quantum anomalous Hall effect in intrinsic magnetic topological insulator MnBi2Te4," *Science (80-.).*, vol. 367, no. 6480, pp. 895–900, Feb. 2020, doi: 10.1126/science.aax8156.

[143] J.-X. Yin, W. Ma, T. A. Cochran, X. Xu, S. S. Zhang, H.-J. Tien, N. Shumiya, G. Cheng, K. Jiang, B. Lian, Z. Song, G. Chang, I. Belopolski, D. Multer, M. Litskevich, Z.-J. Cheng, X. P. Yang, B. Swidler, H. Zhou, H. Lin, T. Neupert, Z. Wang, N. Yao, T.-R. Chang, S. Jia, and M. Z. Hasan "Quantum-limit Chern topological magnetism in TbMn6Sn6," *Nature*, 2020, doi: 10.1038/s41586-020-2482-7.

[144] H. Deng, Z. Chen, A. Wolos, M. Konczykowski, K. Sobczak, J. Sitnicka, I. V. Fedorchenko, J. Borysiuk, T. Heider, L. Plucinski, K. Park, A. B. Georgescu, J. Cano, and L. Krusin-Elbaum, "Observation of high-temperature quantum anomalous Hall regime in intrinsic MnBi$_2$Te$_4$/Bi$_2$Te$_3$ superlattice," *Nat. Phys. 2020*, pp. 1–7, Jan. 2020, doi: 10.1038/s41567-020-0998-2.

[145] C. Liu, Y. Wang, H. Li, Y. Wu, Y. Li, J. Li, K. He, Y. Xu, J. Zhang, and Y. Wang, "Robust axion insulator and Chern insulator phases in a two-dimensional antiferromagnetic topological insulator," *Nat. Mater.*, 2020, doi: 10.1038/s41563-019-0573-3.

[146] H. Polshyn, J. Zhu, M. A. Kumar, Y. Zhang, F. Yang, C. L. Tschirhart, M. Serlin, K. Watanabe, T. Taniguchi, A. H. MacDonald, and A. F. Young, "Electrical switching of magnetic order in an orbital Chern insulator," *Nature*, vol. 588, no. 7836, pp. 66–70, Dec. 2020, doi: 10.1038/s41586-020-2963-8.

[147] S. Chen, M. He, Y. H. Zhang, V. Hsieh, Z. Fei, K. Watanabe, T. Taniguchi, D. H. Cobden, X. Xu, C. R. Dean, and M. Yankowitz, "Electrically tunable correlated and topological states in twisted monolayer–bilayer graphene," *Nat. Phys.*, pp. 1–7, Oct. 2020, doi: 10.1038/s41567-020-01062-6.

[148] R. Bez, "Innovative technologies for high density non-volatile semiconductor memories," in *Microelectronic Engineering*, 2005, doi: 10.1016/j.mee.2005.04.076.

[149] G. W. Burr, R. S. Shenoy, K. Virwani, P. Narayanan, A. Padilla, B. Kurdi, and H. Hwang, "Access devices for 3D crosspoint memory," *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.*, 2014, doi: 10.1116/1.4889999.

[150] A. Aziz, N. Jao, S. Datta, and S. K. Gupta, "Analysis of Functional Oxide based Selectors for Cross-Point Memories," *IEEE Trans. Circuits Syst. I Regul. Pap.*, 2016, doi: 10.1109/TCSI.2016.2620475.

[151] K. Virwani, G. W. Burr, R. S. Shenoy, C. T. Rettner, A. Padilla, T. Topuria, P. M. Rice, G. Ho, R. S. King, K. Nguyen, A. N. Bowers, M. Jurich, M. Brightsky, E. A. Joseph, A. J. Kellock, N. Arellano, B. N. Kurdi, and K. Gopalakrishnan, "Sub-30nm scaling and high-speed operation of fully-confined Access-Devices for 3D crosspoint memory based on mixed-ionic-electronic-conduction (MIEC) materials," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2012, doi: 10.1109/IEDM.2012.6478967.

[152] S. Hirose, A. Nakayama, H. Niimi, K. Kageyama, and H. Takagi, "Resistance switching and retention behaviors in polycrystalline La-doped SrTiO3 ceramics chip devices," *J. Appl. Phys.*, 2008, doi: 10.1063/1.2975316.

[153] L. Zhang, B. Govoreanu, A. Redolfi, D. Crotti, H. Hody, V. Paraschiv, S. Cosemans, C. Adelmann, T. Witters, S. Clima, Y. Y. Chen, P. Hendrickx, D. J.

Wouters, G. Groeseneken, and M. Jurczak, “High-drive current (>1MA/cm2) and highly nonlinear (>103) TiN/amorphous-Silicon/TiN scalable bidirectional selector with excellent reliability and its variability impact on the 1S1R array performance,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2015, vol. 2015-February, no. February, pp. 6.8.1-6.8.4, doi: 10.1109/IEDM.2014.7047000.

[154] W. Lee, J. Park, J. Shin, J. Woo, S. Kim, G. Choi, S. Jung, S. Park, D. Lee, E. Cha, H. D. Lee, S. G. Kim, S. Chung, and H. Hwang, “Varistor-type bidirectional switch (J MAX>10 7A/cm 2, selectivity~10 4) for 3D bipolar resistive memory arrays,” in *Digest of Technical Papers - Symposium on VLSI Technology*, 2012, pp. 37–38, doi: 10.1109/VLSIT.2012.6242449.

[155] A. Aziz, N. Shukla, S. Datta, and S. K. Gupta, “Implication of hysteretic selector device on the biasing scheme of a cross-point memory array,” in *International Conference on Simulation of Semiconductor Processes and Devices, SISPAD*, 2015, vol. 2015-October, pp. 425–428, doi: 10.1109/SISPAD.2015.7292351.

[156] K. Gopalakrishnan, R. S. Shenoy, C. T. Rettner, K. Virwani, D. S. Bethune, R. M. Shelby, G. W. Burr, A. Kellock, R. S. King, K. Nguyen, A. N. Bowers, M. Jurich, B. Jackson, A. M. Friz, T. Topuria, P. M. Rice, and B. N. Kurdi, “Highly scalable novel access device based on Mixed Ionic Electronic Conduction (MIEC) materials for high density phase change memory (PCM) arrays,” in *Digest of Technical Papers - Symposium on VLSI Technology*, 2010, doi: 10.1109/VLSIT.2010.5556229.

[157] R. S. Shenoy, K. Gopalakrishnan, B. L. Jackson, K. Virwani, G. W. Burr, C. T. Rettner, A. Padilla, D. S. Bethune, R. M. Shelby, A. J. Kellock, M. Breitwisch, E. A. Joseph, R. Dasaka, R. S. King, K. Nguyen, A. N. Bowers, M. Jurich, A. M. Friz, T. Topuria, P. M. Rice, and B. N. Kurdi, “Endurance and scaling trends of novel access-devices for multi-layer crosspoint-memory based on mixed-ionic-electronic-conduction (MIEC) materials,” in *Digest of Technical Papers - Symposium on VLSI Technology*, 2011.

[158] G. W. Burr, K. Virwani, R. S. Shenoy, A. Padilla, M. BrightSky, E. A. Joseph, M.

Lofaro, A. J. Kellock, R. S. King, K. Nguyen, A. N. Bowers, M. Jurich, C. T. Rettner, B. Jackson, D. S. Bethune, R. M. Shelby, T. Topuria, N. Arellano, P. M. Rice, B. N. Kurdi, and K. Gopalakrishnan, "Large-scale (512kbit) integration of multilayer-ready access-devices based on Mixed-Ionic-Electronic-Conduction (MIEC) at 100% yield," in *Digest of Technical Papers - Symposium on VLSI Technology*, 2012, doi: 10.1109/VLSIT.2012.6242451.

[159] S. Weinreb, J. C. Bardin, and H. Mani, "Design of cryogenic SiGe low-noise amplifiers," *IEEE Trans. Microw. Theory Tech.*, 2007, doi: 10.1109/TMTT.2007.907729.

[160] T. Arakawa, Y. Nishihara, M. Maeda, S. Norimoto, and K. Kobayashi, "Cryogenic amplifier for shot noise measurement at 20 mK," *Appl. Phys. Lett.*, 2013, doi: 10.1063/1.4826681.

[161] B. I. Ivanov, M. Trgala, M. Grajcar, E. Ilichev, and H. G. Meyer, "Cryogenic ultra-low-noise SiGe transistor amplifier," *Rev. Sci. Instrum.*, 2011, doi: 10.1063/1.3655448.

[162] R. F. Dziuba, B. F. Field, and T. F. Finnegan, "Cryogenic Voltage Comparator System for 2e/h Measurements," *IEEE Trans. Instrum. Meas.*, vol. 23, no. 4, pp. 264–267, 1974, doi: 10.1109/TIM.1974.4314288.

[163] S. K. Tolpygo, "Superconductor digital electronics: Scalability and energy efficiency issues," *Low Temperature Physics*. 2016, doi: 10.1063/1.4948618.

[164] P. F. Yuh, "A 2-kbit Superconducting Memory Chip," *IEEE Trans. Appl. Supercond.*, vol. 3, no. 2, pp. 3013–3021, 1993, doi: 10.1109/77.257228.

[165] Q. Liu, K. Fujiwara, X. Meng, S. R. Whiteley, T. Van Duzer, N. Yoshikawa, Y. Thakahashi, T. Hikida, and N. Kawai, "Latency and power measurements on a 64-kb hybrid Josephson-CMOS memory," in *IEEE Transactions on Applied Superconductivity*, 2007, doi: 10.1109/TASC.2007.898698.

[166] Y. J. Feng, X. Meng, S. R. Whiteley, T. Van Duzer, K. Fujiwara, H. Miyakawa, and N. Yoshikawa, "Josephson-CMOS hybrid memory with ultra-high-speed interface circuit," in *IEEE Transactions on Applied Superconductivity*, 2003, doi: 10.1109/TASC.2003.813902.

[167] S. Nagasawa, K. Hinode, T. Satoh, Y. Kitagawa, and M. Hidaka, “Design of all-dc-powered high-speed single flux quantum random access memory based on a pipeline structure for memory cell arrays,” *Supercond. Sci. Technol.*, 2006, doi: 10.1088/0953-2048/19/5/S34.

[168] S. Tahara, I. Ishida, S. Nagasawa, M. Hidaka, H. Tsuge, and Y. Wada, “4-Kbit Josephson Nondestructive ReadOut Ram Operated At 580 psec and 6.7 mW,” *IEEE Trans. Magn.*, 1991, doi: 10.1109/20.133751.

[169] A. Kirichenko, O. Mukhanov, A. F. Kirichenko, O. A. Mukhanov, and D. K. Brock, “A Single Flux Quantum Cryogenic Random Access Memory Rapid Single Flux Quantum Digital Electronics View project Inductance extraction and flux trapping analysis of superconducting circuits View project A Single Flux Quantum Cryogenic Random Access Memory,” in *Extended Abstracts of 7th International Superconducting Electronics Conference (ISEC’99)*, 1999, pp. 124–127.

[170] Y. Braiman, B. Neschke, N. Nair, N. Imam, and R. Glowinski, “Memory states in small arrays of Josephson junctions,” *Phys. Rev. E*, 2016, doi: 10.1103/PhysRevE.94.052223.

[171] N. Nair, A. Jafari-Salim, A. D’Addario, N. Imam, and Y. Braiman, “Experimental demonstration of a Josephson cryogenic memory cell based on coupled Josephson junction arrays,” *Supercond. Sci. Technol.*, vol. 32, no. 11, p. 115012, Oct. 2019, doi: 10.1088/1361-6668/AB416A.

[172] B. Collaudin and N. Rando, “Cryogenics in space: A review of the missions and of the technologies,” *Cryogenics (Guildf).*, 2000, doi: 10.1016/S0011-2275(01)00035-2.

[173] S. Nagasawa, H. Numata, Y. Hashimoto, and S. Tahara, “High-frequency clock operation of josephson 256-word x 16-bit rams,” *IEEE Trans. Appl. Supercond.*, 1999, doi: 10.1109/77.783834.

[174] I. Byun, D. Min, G. H. Lee, S. Na, and J. Kim, “CryoCore: A Fast and Dense Processor Architecture for Cryogenic Computing,” *Proc. - Int. Symp. Comput. Archit.*, vol. 2020-May, pp. 335–348, May 2020, doi: 10.1109/ISCA45697.2020.00037.

[175] Z. Jiang, S. Yin, J. S. Seo, and M. Seok, “C3SRAM: An In-Memory-Computing SRAM Macro Based on Robust Capacitive Coupling Computing Mechanism,” *IEEE J. Solid-State Circuits*, vol. 55, no. 7, pp. 1888–1897, Jul. 2020, doi: 10.1109/JSSC.2020.2992886.

[176] X. Zhang, V. Mohan, and A. Basu, “CRAM: Collocated SRAM and DRAM with In-Memory Computing-Based Denoising and Filling for Neuromorphic Vision Sensors in 65 nm CMOS,” *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 67, no. 5, pp. 816–820, May 2020, doi: 10.1109/TCSII.2020.2980125.

[177] P. Wang, X. Peng, W. Chakraborty, A. Khan, S. Datta, and S. Yu, “Cryogenic Performance for Compute-in-Memory Based Deep Neural Network Accelerator,” pp. 1–4, Apr. 2021, doi: 10.1109/ISCAS51556.2021.9401756.

[178] S. Li, C. Xu, Q. Zou, J. Zhao, Y. Lu, and Y. Xie, “Pinatubo: A processing-in-memory architecture for bulk bitwise operations in emerging non-volatile memories,” *Proc. - Des. Autom. Conf.*, vol. 05-09-June, Jun. 2016, doi: 10.1145/2897937.2898064.

[179] T. Yoo, H. Kim, Q. Chen, T. T. H. Kim, and B. Kim, “A Logic Compatible 4T Dual Embedded DRAM Array for In-Memory Computation of Deep Neural Networks,” *Proc. Int. Symp. Low Power Electron. Des.*, vol. 2019-July, Jul. 2019, doi: 10.1109/ISLPED.2019.8824826.

[180] NiLeibin, HuangHantao, LiuZichuan, J. V., and YuHao, “Distributed In-Memory Computing on Binary RRAM Crossbar,” *ACM J. Emerg. Technol. Comput. Syst.*, vol. 13, no. 3, Mar. 2017, doi: 10.1145/2996192.

[181] S. Yin, J. S. Seo, Y. Kim, X. Han, H. Barnaby, S. Yu, Y. Luo, W. He, X. Sun, and J. J. Kim, “Monolithically Integrated RRAM- And CMOS-Based In-Memory Computing Optimizations for Efficient Deep Learning,” *IEEE Micro*, vol. 39, no. 6, pp. 54–63, Nov. 2019, doi: 10.1109/MM.2019.2943047.

[182] M. K. Bashar, J. Vaidya, A. Mallick, R. S. S. Kanthi, S. Alam, N. Amin, C. Lee, F. Shi, A. Aziz, V. Narayanan, and N. Shukla, “An Oscillator-based MaxSAT solver,” Sep. 2021.

[183] D. Fan, S. Angizi, and Z. He, “In-Memory Computing with Spintronic Devices,”

Proc. IEEE Comput. Soc. Annu. Symp. VLSI, ISVLSI, vol. 2017-July, pp. 683–688, Jul. 2017, doi: 10.1109/ISVLSI.2017.116.

[184] W. Kang, H. Wang, Z. Wang, Y. Zhang, and W. Zhao, "In-Memory Processing Paradigm for Bitwise Logic Operations in STT-MRAM," *IEEE Trans. Magn.*, vol. 53, no. 11, Nov. 2017, doi: 10.1109/TMAG.2017.2703863.

[185] Z. He, S. Angizi, and D. Fan, "Exploring STT-MRAM based in-memory computing paradigm with application of image edge extraction," *Proc. - 35th IEEE Int. Conf. Comput. Des. ICCD 2017*, pp. 439–446, Nov. 2017, doi: 10.1109/ICCD.2017.78.

[186] D. Reis, M. Niemier, and X. Sharon Hu, "Computing in memory with FeFETs," *Proc. Int. Symp. Low Power Electron. Des.*, vol. 18, 2018, doi: 10.1145/3218603.

[187] K. Ni, B. Grisafe, W. Chakraborty, A. K. Saha, S. Dutta, M. Jerry, J. A. Smith, S. Gupta, and S. Datta, "In-Memory Computing Primitive for Sensor Data Fusion in 28 nm HKMG FeFET Technology," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 2018-Decem, pp. 16.1.1-16.1.4, Jan. 2019, doi: 10.1109/IEDM.2018.8614527.

[188] W. Qiao, Y. Zhao, J. Yang, C. Liu, P. Jiang, Q. Ding, T. Gong, Q. Luo, H. Lv, and M. Liu, "Non-volatile in Memory Dual-Row X(N)OR Operation with Write Back Circuit Based on 1T1C FeRAM," *2020 IEEE 15th Int. Conf. Solid-State Integr. Circuit Technol. ICSICT 2020 - Proc.*, Nov. 2020, doi: 10.1109/ICSICT49897.2020.9278353.

[189] P. Hosseini, A. Sebastian, N. Papandreou, C. D. Wright, and H. Bhaskaran, "Accumulation-based computing using phase-change memories with FET access devices," *IEEE Electron Device Lett.*, vol. 36, no. 9, pp. 975–977, Sep. 2015, doi: 10.1109/LED.2015.2457243.

[190] I. Chakraborty, G. Saha, and K. Roy, "Photonic In-Memory Computing Primitive for Spiking Neural Networks Using Phase-Change Materials," *Phys. Rev. Appl.*, vol. 11, no. 1, p. 014063, Jan. 2019, doi: 10.1103/PhysRevApplied.11.014063.

[191] L. Yin, R. Cheng, Y. Wen, C. Liu, and J. He, "Emerging 2D Memory Devices for In-Memory Computing," *Adv. Mater.*, vol. 33, no. 29, p. 2007081, Jul. 2021, doi:

10.1002/ADMA.202007081.

[192] O. Mutlu, “Intelligent Architectures for Intelligent Machines,” *2020 Int. Symp. VLSI Des. Autom. Test, VLSI-DAT 2020*, Aug. 2020, doi: 10.1109/VLSI-DAT49148.2020.9196490.

[193] V. Sze, Y. H. Chen, J. Einer, A. Suleiman, and Z. Zhang, “Hardware for machine learning: Challenges and opportunities,” *Proc. Cust. Integr. Circuits Conf.*, vol. 2017-April, Jul. 2017, doi: 10.1109/CICC.2017.7993626.

[194] S. Kanev, J. P. Darago, K. Hazelwood, P. Ranganathan, T. Moseley, G.-Y. Wei, and D. Brooks, “Profiling a warehouse-scale computer,” in *Proceedings of the 42nd Annual International Symposium on Computer Architecture*, 2015, pp. 158–169, doi: 10.1145/2749469.2750392.

[195] A. Boroumand, S. Ghose, Y. Kim, R. Ausavarungnirun, E. Shiu, R. Thakur, D. Kim, A. Kuusela, A. Knies, P. Ranganathan, and O. Mutlu, “Google Workloads for Consumer Devices: Mi-tigating Data Movement Bottlenecks,” *Proc. Twenty-Third Int. Conf. Archit. Support Program. Lang. Oper. Syst.*, vol. 18, 2018, doi: 10.1145/3173162.

[196] C. Ahn, S. Kim, T. Gokmen, O. Dial, M. Ritter, and H. S. P. Wong, “Temperature-dependent studies of the electrical properties and the conduction mechanism of HfOx-based RRAM,” in *Proceedings of Technical Program - 2014 International Symposium on VLSI Technology, Systems and Application, VLSI-TSA 2014*, 2014, doi: 10.1109/VLSI-TSA.2014.6839685.

[197] L. Lang, Y. Jiang, F. Lu, C. Wang, Y. Chen, A. D. Kent, and L. Ye, “A low temperature functioning CoFeB/MgO-based perpendicular magnetic tunnel junction for cryogenic nonvolatile random access memory,” *Appl. Phys. Lett.*, vol. 116, no. 2, p. 022409, Jan. 2020, doi: 10.1063/1.5129553.

[198] S. Alam, M. S. Hossain, S. R. Srinivasa, and A. Aziz, “Cryogenic Memory Technologies,” *ArXiV Prepr.*, Nov. 2021, doi: 10.48550/arXiv.2111.09436.

[199] S. Resch, S. K. Khatamifard, Z. I. Chowdhury, M. Zabihi, Z. Zhao, H. Cilasun, J. P. Wang, S. S. Sapatnekar, and U. R. Karpuzcu, “MOUSE: Inference in non-volatile memory for energy harvesting applications,” *Proc. Annu. Int. Symp.*

Microarchitecture, MICRO, vol. 2020-Octob, pp. 400–414, Oct. 2020, doi: 10.1109/MICRO50266.2020.00042.

[200] Z. Wang, H. Ying, W. Chern, S. Yu, M. Mourigal, J. D. Cressler, and A. I. Khan, "Cryogenic characterization of a ferroelectric field-effect-transistor," *Appl. Phys. Lett.*, 2020, doi: 10.1063/1.5129692.

[201] J. He, T. Teng Lu, G. Guo, J. Xu, and C. Luo, "A Low Power and High Speed CMOS Comparator at Croyogenic Temperature," *Proc. - 2020 2nd Int. Conf. Inf. Technol. Comput. Appl. ITCA 2020*, pp. 86–89, Dec. 2020, doi: 10.1109/ITCA52113.2020.00025.

[202] R. Lindaman, "A Theorem for Deriving Majority-Logic Networks within an Augmented Boolean Algebra," *IRE Trans. Electron. Comput.*, vol. EC-9, no. 3, pp. 338–342, 1960, doi: 10.1109/TEC.1960.5219856.

[203] F. Ciubotaru, G. Talmelli, T. Devolder, O. Zografos, M. Heyns, C. Adelmann, and I. P. Radu, "First experimental demonstration of a scalable linear majority gate based on spin waves," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 2018-December, pp. 36.1.1-36.1.4, Jan. 2019, doi: 10.1109/IEDM.2018.8614488.

[204] T. Oya, T. Asai, T. Fukui, and Y. Amemiya, "A Majority-Logic Nanodevice Using a Balanced Pair of Single-Electron Boxes," *J. Nanosci. Nanotechnol.*, vol. 2, no. 3–4, pp. 333–342, Jun. 2002, doi: 10.1166/JNN.2002.108.

[205] A. Imre, G. Csaba, L. Ji, A. Orlov, G. H. Bernstein, and W. Porod, "Majority logic gate for magnetic quantum-dot cellular automata," *Science (80-.).*, vol. 311, no. 5758, pp. 205–208, Jan. 2006, doi: 10.1126/science.1120506.

[206] S. Breitkreutz, J. Kiermaier, I. Eichwald, X. Ju, G. Csaba, D. Schmitt-Landsiedel, and M. Becherer, "Majority gate for nanomagnetic logic with perpendicular magnetic anisotropy," *IEEE Trans. Magn.*, vol. 48, no. 11, pp. 4336–4339, 2012, doi: 10.1109/TMAG.2012.2197184.

[207] L. Amaru, P. E. Gaillardon, and G. De Micheli, "Majority-Inverter Graph: A New Paradigm for Logic Optimization," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 35, no. 5, pp. 806–819, May 2016, doi: 10.1109/TCAD.2015.2488484.

[208] M. F. Ali, A. Jaiswal, and K. Roy, "In-Memory Low-Cost Bit-Serial Addition

Using Commodity DRAM Technology," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 67, no. 1, pp. 155–165, Jan. 2020, doi: 10.1109/TCSI.2019.2945617.

[209] H. Chen, Y. Wang, C. Xu, B. Shi, C. Xu, Q. Tian, and C. Xu, "AdderNet: Do we really need multiplications in deep learning?," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1465–1474, 2020, doi: 10.1109/CVPR42600.2020.00154.

[210] S. Alam, M. M. Islam, J. Hutchins, N. Cady, S. Gupta, G. Rose, and A. Aziz, "Threshold Switch Assisted Memristive Memory with Enhanced Read Distinguishability," in *IEEE 22nd International Conference on Nanotechnology (NANO)*, 2022, pp. 531–534, doi: 10.1109/NANO54668.2022.9928710.

[211] W. H. Chen, C. Dou, K. X. Li, W. Y. Lin, P. Y. Li, J. H. Huang, J. H. Wang, W. C. Wei, C. X. Xue, Y. C. Chiu, Y. C. King, C. J. Lin, R. S. Liu, C. C. Hsieh, K. T. Tang, J. J. Yang, M. S. Ho, and M. F. Chang, "CMOS-integrated memristive non-volatile computing-in-memory for AI edge processors," *Nat. Electron. 2019 29*, vol. 2, no. 9, pp. 420–428, Aug. 2019, doi: 10.1038/s41928-019-0288-0.

[212] G. S. Rose, M. S. A. Shawkat, A. Z. Foshie, J. J. Murray, and M. M. Adnan, "A system design perspective on neuromorphic computer processors," *Neuromorphic Comput. Eng.* , vol. 1, no. 2, p. 022001, Nov. 2021, doi: 10.1088/2634-4386/AC24F5.

[213] J. Sun, Y. Wang, P. Liu, S. Wen, and Y. Wang, "Memristor-Based Neural Network Circuit With Multimode Generalization and Differentiation on Pavlov Associative Memory," *IEEE Trans. Cybern.*, May 2022, doi: 10.1109/TCYB.2022.3200751.

[214] J. Sun, G. Han, Z. Zeng, and Y. Wang, "Memristor-Based Neural Network Circuit of Full-Function Pavlov Associative Memory with Time Delay and Variable Learning Rate," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 2935–2945, Jul. 2020, doi: 10.1109/TCYB.2019.2951520.

VITA

Shamiul Alam is a Ph.D. candidate in the Department of Electrical Engineering and Computer Science at the University of Tennessee, located in Knoxville, Tennessee, USA. He obtained his Master of Science degree in Electrical Engineering from the University of Tennessee Knoxville in 2023 and his Bachelor of Science degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology in 2017. Before his graduate studies, Shamiul worked as a Lecturer at Military Institute of Science and Technology, Dhaka, Bangladesh. His research interests primarily focus on device modeling and circuit design targeting room temperature and cryogenic computing. Shamiul has received numerous awards and accolades from the University of Tennessee for his research contributions, including the Tickle College of Engineering Graduate Research Award (2025), Gonzalez Family Award for Outstanding Graduate Research Assistant (2023), the Graduate Student Senate Award for Excellence in Graduate Research (2024), and the Volunteer of Distinction Award (2024). Additionally, his research has been recognized with the selection as 'Featured' and 'Editor's Pick' articles in Applied Physics Reviews and Applied Physics Letters, respectively. Another of his lead-authored article was selected in the 'Top 100 in Materials Science' and 'Top 100 in Physics' collections of Scientific Reports. Throughout his academic journey, Shamiul has been recognized for his academic excellence through multiple fellowships and awards such as the Tennessee's Top 100 Graduate Fellowship (2020-2024), Graduate Advancement Training and Education (GATE) Fellowship (2023-2024), Shipley-Swann Graduate Fellowship (2023-2024), and DAC Young Fellowship (2020).