MINISTRY OF EDUCATION AND TRAINING
HCMC UNIVERSITY OF TECHNOLOGY AND ENGINEERING

LE MINH

RESEARCH ON OPTIMIZING AND IMPLEMENTING NEUROMORPHIC COMPUTING SYSTEM APPLYING MEMRISTORS

Major: Electronic Engineering
Major code: 9520203

SUMMARY OF PH.D. THESIS

HO CHI MINH CITY – 2025

This thesis is completed at HCMC University of Technology and Engineering

Supervisor 1: Assoc. Prof. Truong Ngoc Son

Reviewer 1:
Reviewer 2:
Reviewer 3:

The dissertation will be defended before the university-level Dissertation Evaluation Committee at Ho Chi Minh City University of Technology and Engineering on the date of / / .

LIST OF PUBLICATIONS

A. International Journal:

(1) **M. Le,** T. K. H. Pham, and S. N. Truong, “Noise and Memristance Variation Tolerance of Single Crossbar Architectures for Neuromorphic Image Recognition,” *Micromachines,* vol. 12, p. 690, June 2021 (SCIE Q2, ISSN: 2072-666X)

(2) **M. Le** and S. N. Truong, “Research on the Impact of Data Density on Memristor Crossbar Architectures in Neuromorphic Pattern Recognition,” *Micromachines,* vol. 14, p. 1990, 27 October 2023 (SCIE Q2, ISSN: 2072-666X)

(3) **M. Le** and S. N. Truong, “A Memristor Crossbar-Based XNOR-Net Architecture for Image Recognition”, *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25127–25132, August 2025. (Scopus Q2, ISSN: 2241-4487)

(4) **M. Le** and S. N. Truong, "Optimizing Memristor Crossbar for Neuromorphic Image Recognition by Introducing a Column-Wise Constant Term Circuit", *Journal of Semiconductor Technology and Science*, pp. 598-609, October 2025. (SCI Q3, ISSN: 1598-1657)

B. International Conference:

(5) **M. Le** and S. N. Truong, “Neuromorphic Character Recognition using The Single Memristor Crossbar Array,” in *2021 International Conference on System Science and Engineering (ICSSE)*, Ho Chi Minh City, Vietnam, 2021, pp. 433-436. (eISBN: 978-1-6654-4848-2)

(6) **M. Le** and S. N. Truong, “Memristor Crossbar Circuits for Neuromorphic pattern Recognition,” in *2021 18th International SoC Design Conference (ISOCC)*, Jeju Island, Korea, 2021, pp. 221-222. (eISBN: 978-1-6654-0174-6)

(7) **M. Le** and S. N. Truong, “Statistical Study on Data Dependency of Memristor Crossbar Architectures for Neuromorphic Computing,” in *2023 8th International Scientific Conference on Applying New Technology in Green Buildings (ATiGB)*, Danang, Vietnam, 2023, pp. 57-61. (eISBN: 979-8-3503-4397-7)

CHAPTER 1. INTRODUCTION

1.1 Overview

The Neuromorphic Computing System, which was proposed in 1990 by Professor Carver A. Mead [1], aims to develop nanoscale circuits capable of emulating the information processing mechanisms of the human brain, specifically by implementing artificial neural networks directly on hardware rather than through software based on CPU or GPU. Initially, neuromorphic computing systems were designed and implemented on very-large-scale integration (VLSI) systems and achieved remarkable results [2-5]. However, the fabrication technologies for current VLSI systems are based on CMOS technology, which is approaching limits in scaling and technical development [6-8]. Therefore, an alternative circuit element is required to overcome.

Since being proposed in 1971 by Professor Leon Chua [9] and successfully demonstrated in 2008 by Professor R. S. Williams and his colleagues [10], memristor arrays have emerged as a promising solution for implementing neuromorphic computing systems, owing to their ability to memory and modify values similar to synaptic characteristics [11–13], their capability to be interconnected in 2D arrays [14–21], and their potential for 3D architectures to realize the high connectivity of biological neural systems [22–26]. The approach of using analog memristor arrays to implement artificial neural networks has achieved certain results [27–33], but still faces many difficulties and challenges related to memristor nonlinear characteristics [31–33, 36–45]. In contrast, binary memristor arrays with two resistance states offer higher efficiency, particularly in binary neural networks [24, 44, 46–48], are less affected by memristance variations [44] due to high HRS/LRS ratio [25, 49, 50], and can perform logic operations based on Kirchhoff's law and Ohm's law [13, 51]. Therefore, binary memristor arrays are more promising and feasible for designing neuromorphic computing systems, particularly in speech recognition [44, 52] and image recognition [47, 51, 53-55].

Nationally, research on memristor-based applications has not been strongly developed. The studies on character recognition [56] or neural network implementation [57] were designed using two memristor arrays, which are not optimized in terms of array size and power consumption in comparison with architectures employing a single memristor array. In recent years, significant progress has been achieved in the research and fabrication of memristive devices and arrays [58, 59].

To implement pattern recognition applying binary memristor arrays, a number of binary memristor array architectures have been proposed, including the complementary memristor array architecture [44], the twin memristor array architecture [47], and the single memristor array architecture [48]. However, the impacts of factors including memristance variation, noise, and data density on the operation of these architectures, have not yet been analyzed and evaluated. Therefore, analyzing and evaluating these influencing factors, as well as research on a novelty optimal structure of binary memristor array that minimizes the number of memristors while ensuring the array's operational efficiency for implementing neuromorphic computing systems, is essential and promising.

1.2 Research Necessity and Significance

Research on optimizing binary memristor array architectures for neuromorphic computing systems implementing image recognition is essential and promising in applying emerging technologies to replace traditional CMOS technology, which is approaching its limitations, in designing neuromorphic computing circuits. This dissertation proposes an optimized design for the binary memristor array architecture applied in neuromorphic computing systems for image recognition tasks. In particular, a single binary memristor array is employed and optimized to fully implement the equation of the XNOR array and stabilizes the recognition performance.

In terms of scientific significance, this study investigates and proposes an optimized design of the binary memristor array architecture for image recognition in neuromorphic computing systems, overcoming the limitations of the previous study by addressing the impact of data density on the recognition performance. The optimized memristor array design is further applied to the development of a multilayer XNOR network for handwritten character recognition. From a practical perspective, this research provides a foundation for more in-depth studies on designing neuromorphic computing systems using memristor arrays as an alternative to CMOS circuits. It also serves as a basis for developing hardware to implement artificial intelligence applications and contributes to efforts to model human brain functions at the circuit level.

1.3 Research Objectives

The objectives of the dissertation are as follow:

- To analyze and evaluate the effects of factors including noise in data and memristance variation on the performance of various binary memristor array architectures in image recognition, thereby providing a basis for developing and proposing an optimized memristor array architecture for neuromorphic computing systems in image recognition tasks.

- To analyze and evaluate the impact of data density on the performance of binary memristor array architectures in implementing XNOR arrays for image recognition, and to identify the causes of this impact as well as the aspects of the binary memristor arrays that require optimization for image recognition.

- To optimize and propose a circuit design for the optimized binary memristor array architecture applied to neuromorphic computing systems for image recognition that fully implements the XNOR array equation with a single binary memristor array, mitigates the impact of data density, and stabilizes the recognition performance. Based on this design, the study further applies the optimized binary memristor to an XNOR neural network for handwritten character recognition using the MNIST dataset.

1.4 Research Object

The research objects encompass neuromorphic computing systems, memristors, memristor arrays, integrated circuits utilizing memristor arrays to implement neuromorphic computing systems, various binary memristor array architectures, factors influencing the array's performance, and optimal design solutions for binary memristor array architectures in image recognition applications.

1.5 Research Mothodology

The research methods include: comprehensive theoretical study; analysis and synthesis of fundamental theories; computational and design methods; simulation methods; and analysis and synthesis of results.

CHAPTER 2. THEORETICAL BASIS

This chapter introduces the theoretical foundations relevant to the dissertation's research problem, covering the structure and modeling of memristors, their classification, neuromorphic computing systems, and architectures of binary memristor arrays.

CHAPTER 3. ANALYSIS OF FACTORS INFLUENCING THE PERFORMANCE OF BINARY MEMRISTOR ARRAY ARCHITECTURES

3.1 Introduction

In this chapter, the dissertation analyzes the challenges of using memristor arrays with multi-level programmed memristors (analog), two-level memristors (binary or digital), as well as the factors affecting the performance of binary memristor array architectures. This analysis serves as the basis for selecting an optimal memristor array architecture for neuromorphic computing system designs in next studies.

3.2 Challenges of Implementing Neuromorphic Computing Systems with Analog Memristor Arrays

The nonlinear conductance characteristics of memristors make it difficult to program analog memristor values to precisely achieve the desired levels. he challenges in designing learning circuits and corresponding algorithms that allow on-hardware training also constitute a significant obstacle. Variations in memristance also have a significant impact on recognition. Therefore, there is an urgent need to develop a new approach that utilizes binary memristors for designing neuromorphic computing systems.

3.3 Impact of Noise and Memristance Variation on the Performance of Binary Memristor Array Architectures in Neuromorphic Systems

3.3.1 Analysis of the Mathematical Models of Binary Memristor Array Architectures

The complementary memristor array architecture employs two complementary arrays $M+$ and $M-$ to recognize an input vector A by implementing the XNOR function, as described in Equation (2.6) [44].

The output Y is calculated by following equation:

$$Y = A \cdot (M+) + A' \cdot (M-)$$

$$= \begin{bmatrix} a_0 & a_1 & \cdots & a_{(n-1)} \end{bmatrix} \cdot \begin{bmatrix} g_{0,0} & g_{0,1} & \cdots & g_{0,(m-1)} \\ g_{1,0} & g_{1,1} & \cdots & g_{1,(m-1)} \\ \vdots & \vdots & \vdots & \vdots \\ g_{(n-1),0} & g_{(n-1),1} & \cdots & g_{(n-1),(m-1)} \end{bmatrix}$$

$$+ \begin{bmatrix} a'_0 & a'_1 & \cdots & a'_{(n-1)} \end{bmatrix} \cdot \begin{bmatrix} g'_{0,0} & g'_{0,1} & \cdots & g'_{0,(m-1)} \\ g'_{1,0} & g'_{1,1} & \cdots & g'_{1,(m-1)} \\ \vdots & \vdots & \vdots & \vdots \\ g'_{(n-1),0} & g'_{(n-1),1} & \cdots & g'_{(n-1),(m-1)} \end{bmatrix} \quad (3.2)$$

$$= \begin{bmatrix} y_0 & y_1 & \cdots & y_{m-1} \end{bmatrix}$$

$$trong \text{ đó } y_j = y_j^+ + y_j^- = \sum_{i=0}^{n-1} (a_i g_{i,j} + a'_i g'_{i,j})$$

The vector Y contains the output column currents that represent the similarity score between the input A and the columns of memristor arrays.

In the twin crossbar array architecture, the XNOR operation in Equation (2.6) is re-expressed as Equation (2.7)[47].

In Equation (2.7), A' is a constant, which has no interaction with the array M, and is omitted when performing the XNOR function [47]. As a result, Equation (2.7) is simplified as:

$$Y = \overline{A \oplus M} = A \cdot (M +) - A' \cdot (M +)$$
$$= [y_0 \quad y_1 \quad \cdots \quad y_{m-1}]$$
$$where: y_j = y_j^+ - y_j^- = \sum_{i=0}^{n-1} (a_i g_{i,j} - a'_i g_{i,j}) \tag{3.3}$$

Therefore, the twin crossbar architecture employs 2 identical memristor arrays, $M +$ and $+$, to store patterns and implement the XNOR operation with the input A and obtain the output vector Y containing output column currents.

In the single crossbar architecture, the XNOR function is re-espressed as described in Equation (2.9) [48]. In Equation (2.9), A' is a constant and can be obmitted. Therefore, to implement the XNOR function, Equation (2.9) is simplified to one AND operation as described in Equation (2.10) [48]. In Equation (2.10), $I = [i_0 \quad i_1 \quad \cdots \quad i_{(n-1)}]$ is the input vector of size 1×n containing bipolar inputs. Equation (2.10) of the single memristor crossbar architecture can be represented as the matrix as:

$$Y = \overline{A \oplus M} = (A - A') \cdot M = I \cdot M$$
$$= [i_0 \quad i_1 \quad \cdots \quad i_{(n-1)}] \cdot \begin{bmatrix} g_{0,0} & g_{0,1} & \cdots & g_{0,(m-1)} \\ g_{1,0} & g_{1,1} & \cdots & g_{1,(m-1)} \\ \vdots & \vdots & \vdots & \vdots \\ g_{(n-1),0} & g_{(n-1),1} & \cdots & g_{(n-1),(m-1)} \end{bmatrix} \tag{3.4}$$
$$= [y_0 \quad y_1 \quad \cdots \quad y_{m-1}]$$
$$where \ y_j = \sum_{k=0}^{n-1} i_k g_{k,j}$$

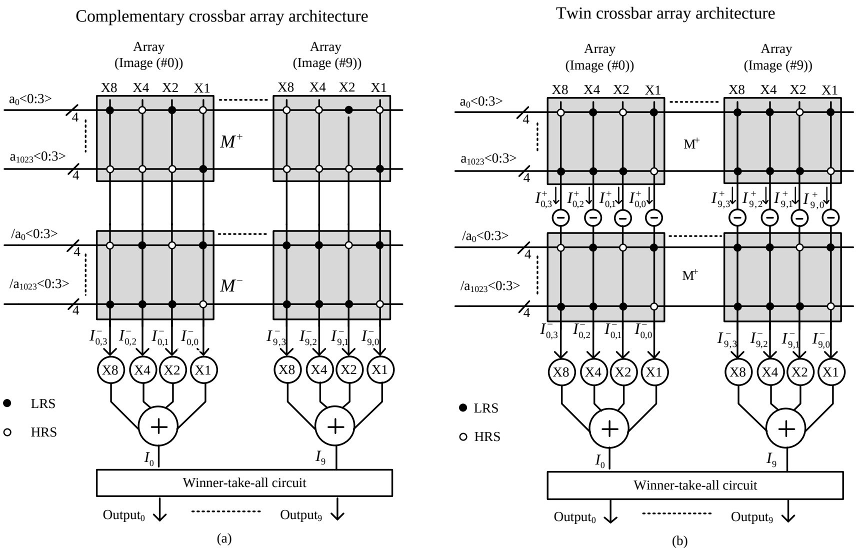
Thereby, the mathematical model of the single memristor crossbar array uses only one memristor array to perform the XNOR operation, producing the output vector Y containing the column currents.

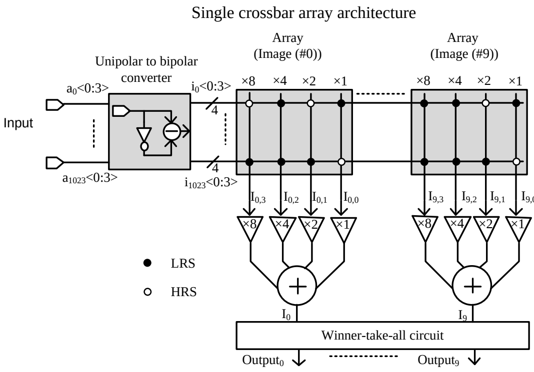
3.3.2 Implementation of Image Recognition Using Binary Memristor Array Architectures

The dissertation implements an image recognition application for 10 images with the complementary memristor array architecture, the twin memristor array architecture, and the single memristor array architecture [81]. The ten grayscale images of size 32×32 is illustrated in Figure 3.4.

Figure 3.1: Ten graycale images for recognizing with binary memristor array architectures.

Each grayscale image is converted from 32×32 format to 1×1024 vector. Each pixel is the digitalized by 4 bits [44, 47]. The 4 bits of the pixel number i, $\mathrm{a_i} < 0\!:\!3 >$, is stored to 4 memristors at the intersections of row number i with 4 columns. These columns have multiplication weights of 8, 4, 2, and 1 for calculating the output column currents. The ten images are stored to memristor arrays ($M+$ and $M-$; $M+$ and $M+$; M) in ten groups where each group has 4 columns.





(c)

Figure 3.2: The block diagrams of (a) the complementary crossbar architecture, (b) the twin crossbar architecture, and (c) the single crossbar architecture recognizing 10 grayscale images of size 32 ×32.

The block diagrams of three memristor crossbar architectures recognizing ten images are shown in Figure 3.5. To be recognized, the input image is converted into vector A of size 1×1024 pixels. Each pixel is subsequently digitalized by 4 bits to implement the XNOR function with memristor arrays pre-storing dataset, as expressed in equations (3.2), (3.3), (3.4). The output column currents are next applied to Winner-take-all circuit to identify the maximum current I_k.

3.3.3 Analysis of the Impact of Noise and Memristance Variation on the Performance of Binary Memristor Array Architectures

To assess the effect of noise on the recognition performance of the three memristor array architectures, Gaussian noise is added to the input images prior to recognition. Figure 3.6 illustrates the input images affected by Gaussian noise with a signal-to-noise ratio (SNR) of -10 dB. The signal-to-noise ratio (SNR) of the Gaussian noise added to the images is varied from -10 dB to 4 dB. Figure 3.7(a) compares the recognition rates of the three architectures on input images with added Gaussian noise. The results shown in Figure 3.7(a) indicate that the single array and the twin array architectures maintain relatively high recognition rates, whereas the recognition rate of the complementary array architecture drops significantly at the SNR of -10 dB.

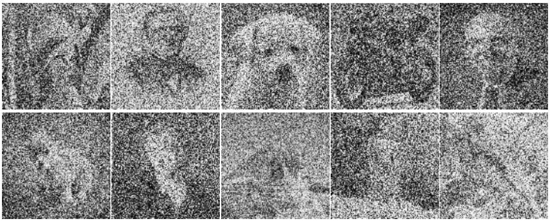


Figure 3.3: Ten grayscale images with Gaussian noise of SNR of -10 dB.

Specifically, at an SNR of -10 dB, the recognition rates of the single memristor array, the twin memristor array, and the complementary memristor array architectures are 91%, 89%, and 4%, respectively. These advantages stem from the fact that the twin memristor array architecture employs the subtraction operation when executing the XNOR function, as shown in Equation (3.3), which helps to reduce the variations in column output currents caused by noise in the input images. The single memristor array architecture is derived from the twin array architecture, as described in Equation (3.4), allowing the noise signal to be partially suppressed through the unipolar-to-bipolar converter. Therefore, the single memristor array architecture demonstrates the highest noise immunity and maintains a superior recognition rate compared to the other two architectures.

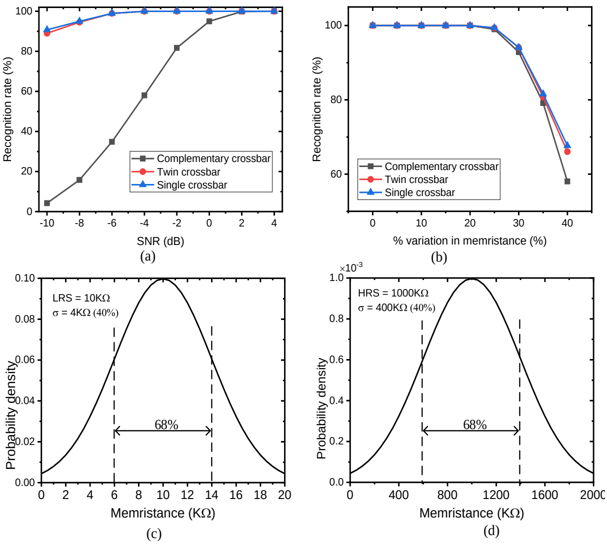


Figure 3.4: (a) The recognition rates of three architectures under Gaussian noise in input images, (b) the recognition rates of three architectures under memristance variation, (c) statical distribution of LRS, (d) statical distribution of HRS.

Next, the memristance variation is adjusted from 0% to 40% according to a Gaussian distribution, as illustrated in Figures 3.7(c) and 3.7(d). The LRS and the HRS are assumed to be 10 KΩ and 1 MΩ. As shown in Figure 3.7(b), when the memristance variation reaches 40%, the single memristor array architecture attains a recognition accuracy of 67.8%, which is the highest compared to 58% and 66% for the complementary and the twin memristor array architectures, respectively. This is because the twin array architecture consists of two identical memristor arrays and employs the subtraction operation during the XNOR function, therefore can partially compensate for variations in column output currents caused by the memristance variation. By using only one memristor array to execute the XNOR function, the single memristor array architecture exhibits the best tolerance to the memristance variation.

3.4 Chapter Conclusion

The dissertation demonstrates that the single memristor array architecture exhibits superior tolerance to input noise and memristance variation, along with an advantage in array size, compared to the complementary and the twin memristor array architectures. Therefore, conducting further in-depth research and optimization of binary memristor arrays, based on the single memristor array architecture, represents an important and feasible approach for designing neuromorphic computing systems.

CHAPTER 4. ANALYSIS OF DATA DENSITY DEPENDENCE FOR OPTIMIZING MEMRISTOR ARRAY ARCHITECTURE IN NEUROMORPHIC COMPUTING SYSTEMS

4.1 Introduction

This chapter further analyzes and evaluates the impact of data density on the performance of binary memristor array architectures, identifies its effects and causes, and provides a foundation for optimizing binary memristor array architecture.

4.2 The single memristor crossbar array architecture in application of binary image recognition

The single array architecture executes the XNOR operation between the input vector A and a single memristor array M according to the equation (4.1) [48]:

$$Y = \overline{A \oplus M} = AM + A'M' = AM + A'(1-M) = (A-A')M = IM \quad (4.1)$$

Here, $I = (A - A')$ is the bipolar input vector, A' the omitted constant. Figure 4.1 shows the block diagram and schematic of the application for recognizing 10 binary images of size 32×32 with the single memristor array architecture.

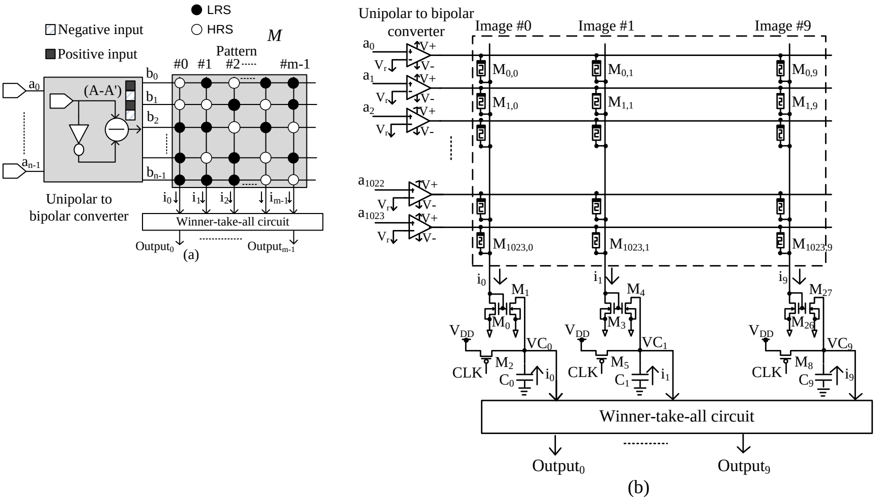


Figure 4.1: (a) The block diagram and (b) the schematic of the single memristor crossbar array in recognition of 10 binary images.

4.3 Data Density of Binary Images

The data density of a binary image is defined as the ratio of the number of 1 bits (white pixels) to the total number of bits in the image. To create a set of binary sample data, the dissertation uses 10 grayscale images of size 32×32, which are converted into black-and-white images with three different data density ratios, as illustrated in Figure 4.5 [83].

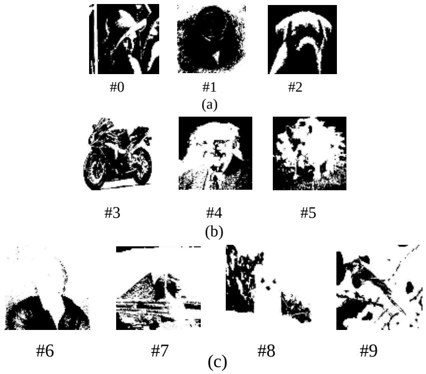


Figure 4.2: Ten binary images with varying data densities: (a) low density 0.25, (b) medium density 0.5, and (c) high density 0.75.

The data densities are 0.25, 0.5, and 0.75, corresponding to 25%, 50%, and 75% of the 1-bits in the images, respectively.

4.4 Analysis of Data Density Effects on the Performance of Memristor Array Architectures in Binary Image Recognition

4.4.1 Impact on Output Column Currents

The binary images shown in Figure 4.5 are sequentially recognized using both the single memristor array and the complementary memristor array architectures. The LRS (bit 1) and HRS (bit 0) are assumed to be 100 kΩ and 10 MΩ, respectively; bipolar voltages are set to 1 V and -1 V, while unipolar voltages are set to 1 V and 0 V. Figure 4.6 shows the output column currents of the two architectures.

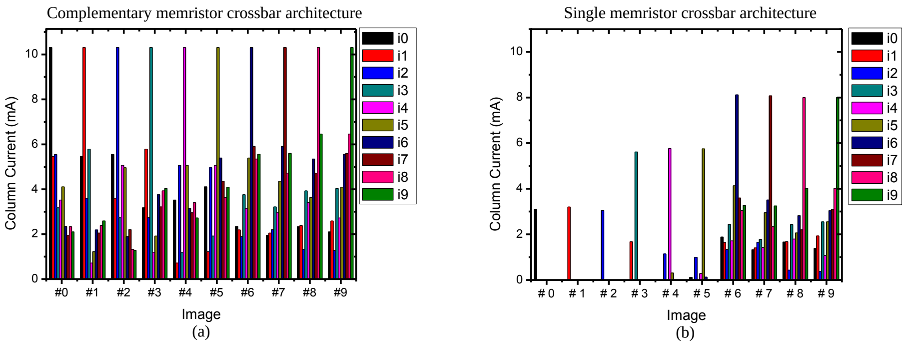


Figure 4.3: Output column currents for binary images #0 to #9 recognizing with the complementary memristor array architecture (a) and the single memristor array architecture (b).

As shown in Figure 4.6(b), the single memristor array architecture exhibits a decrease in output column currents as the data density of the images decreases. In contrast, in the complementary memristor array architecture, the maximum output column currents remain approximately the same across all cases, as shown in Figure 4.6(a). Specifically, as the data density decreases from 0.75 to 0.25, the maximum output column current of the single memristor array architecture decreases from approximately 8.1 mA to about 3 mA, representing a reduction of roughly threefold. In contrast, in the complementary memristor array architecture, the maximum output column current remains approximately 10.3 mA and is unaffected by variations in data density.

The decrease in output column currents in the single memristor array architecture is due to the simplification of the XNOR operation equation (4.1), where the constant term A' was omitted. Specifically, when an input image vector A has low data density, A' exhibits high data density, resulting in a reduction of output column currents. Furthermore, the subtraction in equation (4.1) may produce negative values, which do not generate current at outputs.

4.4.2 Impact on Recognition Performance

Variations in column output currents cause the capacitors, initially charged from C_0 to C_9, to discharge at different rates, so that the voltages VC_0 to VC_9 decrease at varying speeds. These voltages are then input into Winner-take-all circuit to identify which voltage reaches the threshold value first. When the discharge voltage VC_k of the k capacitor falls below the 0.5 V threshold, the Winner-take-all circuit generates a locking pulse at $Output_k$ to conclude.

Figure 4.7 presents the analysis of capacitor discharge voltages when recognizing image #6, which has a high data density of 0.75, with the single memristor array architecture. The discharge voltage VC_6 exhibits the fastest decay, reaching the 0.5 V threshold within approximately 4 ns. Meanwhile, after 4 ns, the voltages of the other capacitors remain above 0.7 V, i.e., still higher than the 0.5 V threshold. Therefore, with a comparison threshold value of V_{REF} = 0.5 V, the Winner-take-all circuit can identify i_6 as the maximum output column current within approximately 7 ns. In conclusion, the recognition performance of the single memristor array architecture remains both accurate and stable when the input image exhibits a high data density.

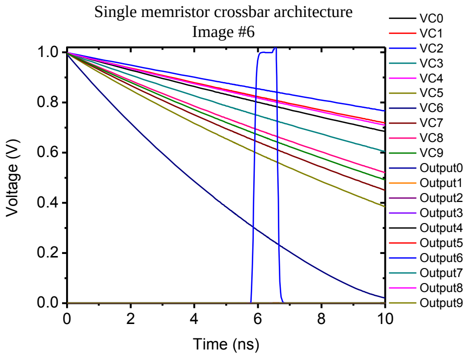


Figure 4.4: Discharge voltages of capacitors VC_0 to VC_9 when image #6 of high density 0.75 is recognized with the single memristor array architecture.

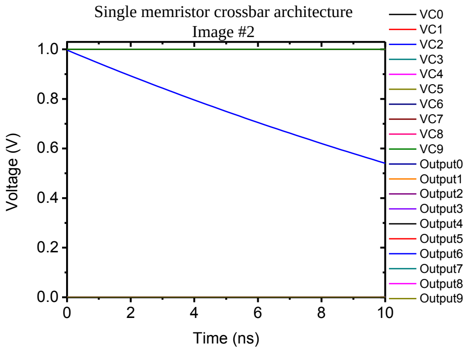


Figure 4.5: Discharge voltages of capacitors VC_0 to VC_9 when image #2 of low density 0.25 is recognized with the single memristor array architecture.

Figure 4.8 shows the discharge voltages of capacitors VC_0 to VC_9 when image #2, with a low data density of 0.25, is recognized with the single memristor array architecture. The discharge voltage VC_2 decreases more slowly compared to the high-density input image case, while the voltages of the other capacitors remain at their initially charged value of 1 V.

In particular, within the same recognition period of 7 ns, the discharge voltage VC_2 remains above 0.8 V, and none of the capacitor voltages have dropped below the 0.5 V threshold. Consequently, within the same 7 ns recognition period, when an image with a low data density of 0.25 is processed, the Winner-take-all circuit in the single memristor array architecture fails to function properly and cannot determine the column with the highest output current, leading to inaccurate recognition,

4.5 Chapter Conclusion

The dissertation demonstrates that a decrease in data density adversely impacts the recognition performance of the single memristor array architecture, leading to reduced output column currents and recognition accuracy in comparison with high data density input images within the same recognition period. This finding underscores the critical need to optimize the binary memristor array architecture in order to achieve stable and reliable recognition performance.

CHAPTER 5. DESIGN OF AN OPTIMIZED MEMRISTOR ARRAY ARCHITECTURE FOR NEUROMORPHIC COMPUTING SYSTEMS

5.1 Introduction

This chapter analyzes and proposes an optimal circuit design for the binary memristor array architecture based on the optimization of the mathematical equations, aiming to fully implement the XNOR operation, mitigate the impact of data density, and stabilize performance in image recognition. Furthermore, a multilayer XNOR neural network leveraging the optimized memristor array architecture is proposed for handwritten character recognition on the MNIST dataset.

5.2 Design of an Optimized Memristor Array Architecture for Neuromorphic Computing Systems

5.2.1 Design of a Column-wise Constant Term Circuit to Mitigate the Impact of Data Density

The XNOR operation in the single memristor array architecture is re-expressed as [48]:

$$\begin{aligned} Y = \overline{A \oplus M} &= AM + A'M' = AM + A'(1-M) \\ &= (A-A')M + A' = (A-A')M = IM \end{aligned} \quad (5.1)$$

Mathematically, A' is a constant and therefore has been omitted. However, at circuit-level perspective, the omission of the constant A' reduces the output current equally across all columns. Therefore, to resolve this issue, it is necessary to implement the full XNOR function as:

$$\begin{aligned} Y = \overline{A \oplus M} &= AM + A'M' = AM + A'(1-M) = (A-A')M + A' \\ &= (A-A') \cdot M + A' \cdot 1 \end{aligned} \quad (5.2)$$

The constant A′ is equivalently converted into the AND operation ($A' \cdot 1$) to retainThe logic value 1 represents a memristor in the low-resistance state (LRS) to generate current.

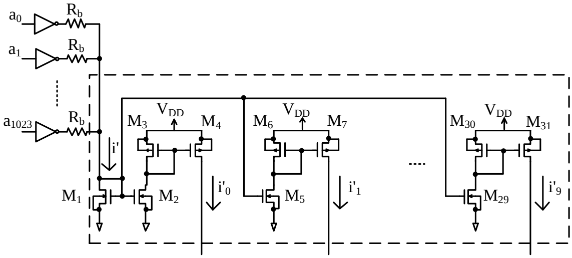


Figure 5.1: The proposed circuit implementing the AND $(A' \cdot 1)$ in Equation (5.2) to add the constant term to output column currents when recognizing 10 binary images.

The AND circuit $(A' \cdot 1)$is implemented separately as a peripheral circuit, using resistors instead of memristors to avoid altering the memristor array architecture, as proposed in Figure 5.1 [84].

The input images are 10 binary images of size 32×32. The input voltage vector $A = [a_0 \quad a_1 \quad \cdots \quad a_{1023}]$ is fed into inverters to obtain its complement vector $A' = [a'_0 \quad a'_1 \quad \ldots \quad a'_{1023}]$ and passed through resistors R_b to generate the additional current i' component for a single output column as:

$$i' = \sum_{k=0}^{1023} \frac{a'_k}{R_b} \qquad (5.3)$$

Here, resistor R_b = LRS. This additional current i' is subsequently duplicated using current mirrors to produce the corresponding additional currents for output columns i'_0 to i'_9 and distribute across the columns of the single memristor crossbar architecture M to stabilize the output column currents under varying data density.

5.2.2 Design of an Optimized Memristor Array Circuit for Image Recognition Applications

The optimized circuit design for the binary memristor array architecture for recognizing 10 binary images is illustrated in Figure 5.2 [84]. This optimized memristor array architecture fully implements the XNOR operation according to the following equation:

$$\begin{aligned} Y = \overline{A \oplus M} &= (A - A') \cdot M + A' \cdot 1 \\ &= [i_0 \quad i_1 \quad \cdots \quad i_{m-1}] \end{aligned} \qquad (5.4)$$

The memristor aray M is of size 1024×10 to pre-store 10 images of size 32×32 in 10 columns. The LRS and HRS are assumed to be 100 KΩ and 10 MΩ. The input vector A is applied into memristor array M to implement the first AND operation $(A - A') \cdot M$ and concurrently applied into the column-wise constant circuit to perform the second AND operation $(A' \cdot 1)$ in Equation (5.4) to obtain the output column currents $[i_0 \quad i_1 \quad \cdots \quad i_{m-1}]$.

Figure 5.4 illustrates the 10 output column currents from i_0 to i_9.

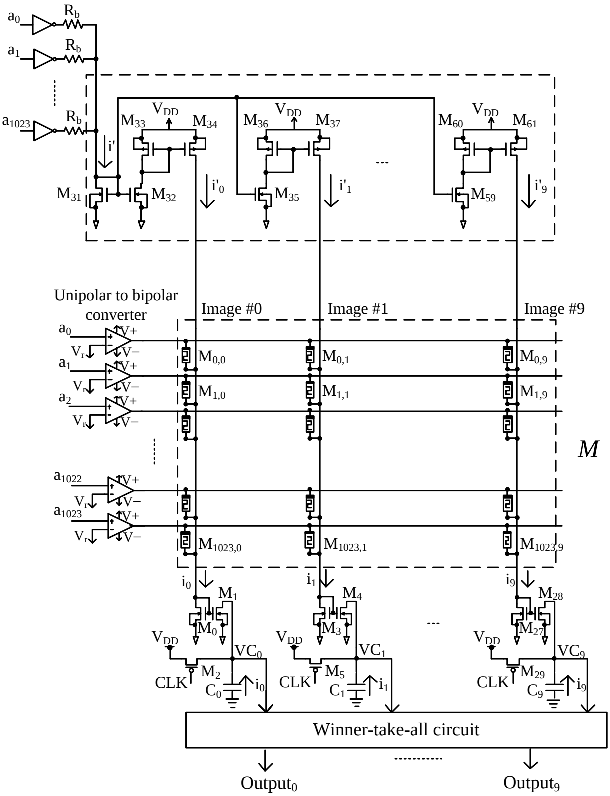


Figure 5.2: The proposed circuit of the optimized binary memristor array architecture implements the full XNOR array in Equation (5.3) to recognize 10 binary images.

The results in Figure 5.4 demonstrate that the output column currents corresponding to images with low data density of 0.25 (images #0, #1, #2) and medium data density of 0.5 (images #3, #4, #5) have been compensated and increased, reaching values comparable to those of high data density images of 0.75 (images #6, #7, #8, #9). Notably, the maximum column output currents are also uniformly stabilized, reaching approximately 10 mA across varying data densities.

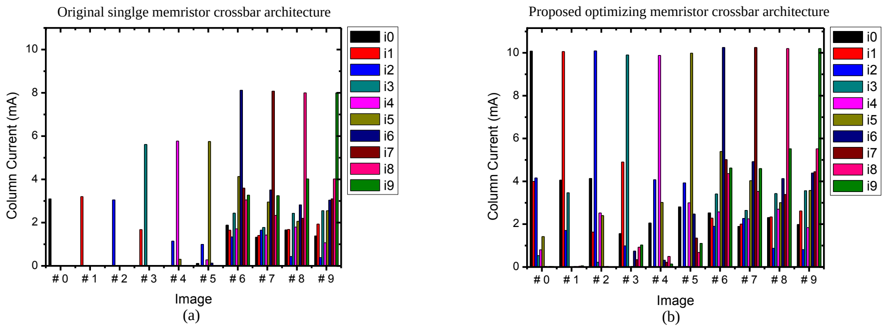


Figure 5.3: Output column currents when recognizing binary images from image #0 to image #9 across varying data densities with (a) the original single array architecture and (b) the proposed architecture.

The output column currents, i_0 to i_9, cause the precharged capacitors C_0 to C_9 in Figure 5.2 to discharge, leading to the different decreasing rates of voltages VC_0 to VC_9. In this study, the capacitors were 50 pF. Figure 5.6 shows the discharge voltages of the capacitors VC_0 to VC_9 along with the output pulses at $Output$ of the proposed architecture in Figure 5.2, corresponding to images #2 and #6. As illustrated in Figure 5.6(a), for #2 with a low data density of 0.25, the discharge voltages of the capacitors successfully reach the 0.5 V threshold, with VC_2 decreasing most rapidly to this threshold within approximately 4.2 ns. The WTA successfully generates the output pulse, the $Output2$, corresponding to the voltage VC_2 at 6.2 ns. For input image #6, which possesses a high data density of 0.75, Figure 5.6(b) illustrates that the discharge voltage VC_6 of capacitor C_6 reaches the 0.5 V threshold most rapidly, within approximately 4 ns. The WTA circuit successfully locks the output pulse, the $Output6$, corresponding to voltage VC_6 at 6.1 ns.

Next, the recognition rates under varying data densities are analyzed in Figure 5.8. The results demonstrate that the proposed optimized array architecture attains both a high and stable recognition rate across varying data densities. In contrast, the original single memristor array architecture experiences a marked degradation as the data density decreases. When the data density decreases to 0.4, the proposed optimized architecture effectively stabilizes and sustains a high recognition accuracy of 100%. In contrast, the original single memristor array architecture exhibits a substantial reduction of recognition rate to 20%.

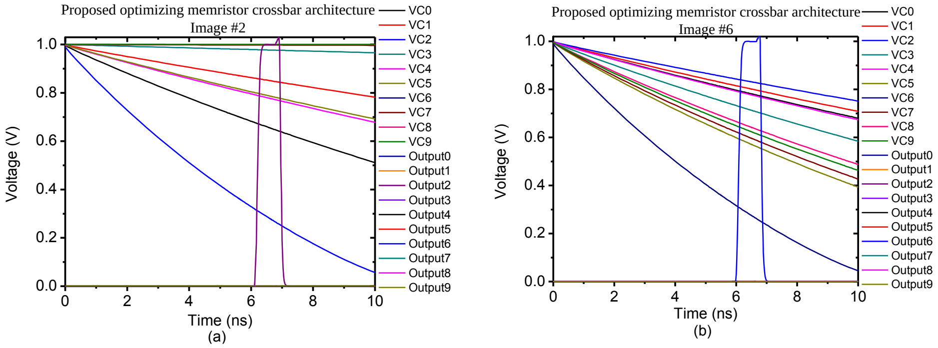


Figure 5.4: Output pulses along with capacitor discharge voltages of the proposed memristor array architecture with column-wise constant term circuit when recognizing: (a) image #2 with low data density (0.25), and (b) image #6 with high data density (0.75).

The proposed architecture exhibits an average power consumption of 95.2 mW, compared to 50.1 mW of the original single memristor array architecture.

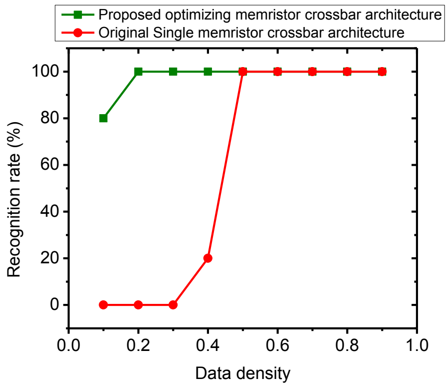


Figure 5.5: Comparison of recognition rates under varying data densities between the proposed optimizing architecture and the original single array architecture.

In conclusion, the proposed circuit design for the optimized binary memristor array architecture successfully eliminates the adverse impact of data density, thereby ensuring consistently high and stable performance in recognition tasks.

5.2.3 Comparison of the Proposed Optimized Binary Memristor Array with Other Architectures in Neuromorphic Image Recognition

The dissertation conducts a

comprehensive comparison of the proposed optimized architecture with the complementary, the twin, and the single memristor array architectures in recognizing 10 images of size 32×32. The results are summarized in Table 5.3.

Table 5.3: Comparison of binary memristor array architectures in recognizing 10 images of size 32x32.

Architectures	*Number of memristor /synapse*	*Recognition rate @ 40% memristance variation*	*Recognition rate @ data density of 0,4*
The complementary architecture [44], [54], [55]	2	50.7%	100%
The twin architecture [47]	2	58.8%	20%
The single architecture [48]	1	62.1%	20%
The proposed architecture	**1**	**62.1%**	**100%**

The results reveal that the proposed memristor array architecture is the robustest candidate among the architectures in terms of the number of memristors per synapse, tolerance of noise, and independence from the density of the data set.

5.2.4 Design of an XNOR Neural Network Applying the Proposed Optimizing Memristor Array Architecture

The proposed memristor array architecture is subsequently applied to design an XNOR neural network of size 784×500×500×10 for handwritten digit recognition in the MNIST dataset [85]. The schematic of the XNOR-Net is designed and presented in Figure 5.11. The trained weight matrices of the hidden layers W_h, W_l and the output layer W_O are mapped onto the memristor arrays M_h, M_l and M_O according to the proposed optimizing array architecture, the weight of +1 and -1 are coded to LRS and HRS, respectively.

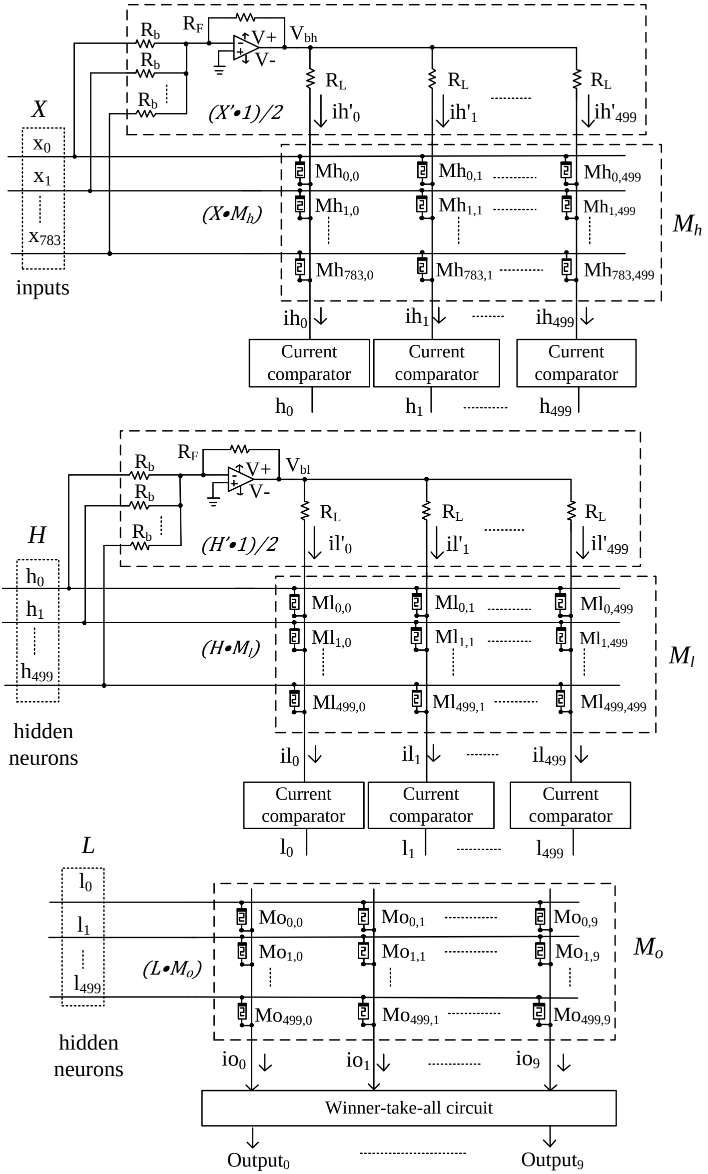


Figure 5.6: The schematic of the proposed XNOR neural network applying the optimized memristor array architecture for handwritten digit recognition on the MNIST dataset.

The neurons of the input layer X, in format of voltages $(V+, V-)$, is applied to the optimized memristor array M_h to execute the multiplication and addition between the input X and weight matrix W_h as in the equation (5.9):

$$I = \overline{X \oplus M_h} = X \cdot M_h + \frac{X' \cdot 1}{2} = [ih_0 \quad ih_1 \ldots ih_{k-1}] \tag{5.9}$$

The output column currents ih_j of the array M_h are obtained as:

$$ih_j = \sum_{i=0}^{m-1}\left(\frac{x_i}{Mh_{i,j}} + \frac{x'_i}{2LRS}\right) = \sum_{i=0}^{m-1} x_i \times \left(\frac{1}{Mh_{i,j}} - \frac{1}{2LRS}\right) \quad (5.10)$$

Each output column current ih_j of the hidden layer H is processed by a current comparator [83] implementing the binary activation function for the neuron h_j as follows:

$$h_j = f(ih_j) = \begin{cases} V+, & when\ ih_j \geq 0 \\ V-, & when\ ih_j < 0 \end{cases} \quad (5.11)$$

At the hidden layer L, the multiplications and additions between the input H and the weight matrix W_l are executed as processing at the hidden layer H.

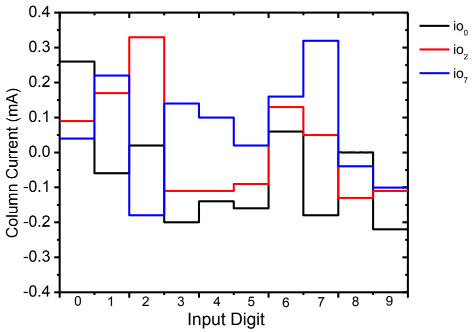


Figure 5.7: The output column currents io_0, io_2 and io_7 of the output layer O when digits 0 – 9 of the MNIST are sequentially applied into the input of the XNOR-Net.

At the output layer O, the neuron l_j of layer L, in voltages of $V+$ and $V-$, is applied into the memristor array M_O to perform the multiplications and additions between L and the output-layer weight matrix W_O in order to obtain the column currents of the output layer as Equation (5.20).

$$I_O = \overline{L \oplus M_O} = L \cdot M_O + \frac{L' \cdot 1}{2} = L \cdot M_O = [io_0\ \ io_1 \ldots io_9] \quad (5.20)$$

At the output layer, the memristor array M_O is designed without the column-wise constant term circuit to reduce the peripheral without affecting the output results of the XNOR neural network.The Winner-take-all circuit acts as the activation function for the output layer to determine the maximum output column current io_j, and generate the result pulse at the column j.

The proposed XNOR-Net demonstrates a recognition accuracy of 94% on the MNIST dataset. Figure 5.12 illustrates the output column currents io_0, io_2, and io_7 at the output layer O corresponding to the recognitions of digits 0-9.

5.3 Chapter Conclusion

In this chapter, the dissertation presents an optimization of the binary memristor array architecture, enabling the full implementation of the XNOR array using only a single memristor array. This optimization effectively mitigates the adverse effects of data density degradation on recognition performance, a challenge of the previous study. Furthermore, the dissertation proposes an XNOR neural network utilizing the optimized memristor array architecture for the handwritten character recognition with the MNIST dataset.

CHAPTER 6. CONCLUSIONS

6.1 Conclusions

The dissertation has accomplished the research objectives, as follows:

- First, the dissertation analyzed and assessed the effects of noise and memristance variation on the performance of binary memristor array architectures. In particular, the dissertation demonstrated that the single memristor array architecture exhibits superior tolerance of Gaussian noise and memristance variation, while benefiting from a reduced array size, compared to the other architectures. Therefore, further optimization of tne binary memristor array based on the single array architecture is promising and feasible.

- Second, the dissertation revealed that the single memristor array architecture is adversely affected by reductions in data density and identified the cause of the effect, in comparison with the complementary architecture. Specifically, a decrease in data density markedly reduces the column output currents of the single array architecture, leading to malfunction of the Winner-take-all circuit and consequently impairing the system’s performance. Additionally, the dissertation identified that the cause of this adverse effect is the omission of the constant term A' in the mathematical expression of the XNOR array.The dissertation provides a novel and significant finding, emphasizing the urgent need to optimize the binary memristor array architecture in order to mitigate the effects of data density, ensure stable recognition performance, and implement the full XNOR array with a single memristor array.

- Third, the dissertation optimized the binary memristor array architecture, addressing the effects of data density in the image recognition and ensuring stable recognition performance regardless of the dataset, while implementing the XNOR array with a single memristor array. The analysis results demonstrate that, even under a reduction in data density, the column output currents of the proposed memristor array architecture remain stable, the Winner-take-all circuit functions accurately and produces the output pulse within approximately 7 ns. Furthermore, the dissertation proposed and designed a multilayer XNOR neural network that utilizes the optimized binary memristor array architecture for handwritten character recognition using the MNIST dataset. At the output layer, the column-wise constant term circuit has been omitted without compromising the network's performance, thereby optimizing both size and power consumption. The proposed XNOR-Net demonstrates a recognition accuracy of 94% on the MNIST dataset.

6.2 Future work

The potential directions for further in-depth research and development include:

-First, continuing advanced research to address other challenges in the design of neuromorphic computing systems using memristor arrays, including the nonlinear characteristics of memristors, device variations, device faults, and leakage currents.

- Second, continuing research on applying the proposed optimized memristor array architecture to design binary convolutional neural networks.

- Third, completely designing peripheral circuits to enable the fabrication of a chip for neuromorphic computing systems in deep learning applications.