REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
APRIL 2013	CONFERENCE PAPER (POST PRINT)	SEP 2011 – SEP 2012

4. TITLE AND SUBTITLE

THE CIRCUIT REALIZATION OF A NEUROMORPHIC COMPUTING SYSTEM WITH MEMRISTOR-BASED SYNAPSE DESIGN

5a. CONTRACT NUMBER
FA8750-11-1-0271

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62788F

6. AUTHOR(S)

Beiye Liu (UPitt), Yiran Chen (UPitt), Tingwen Huang (UPitt), and Bryant Wysocki (AFRL)

5d. PROJECT NUMBER
T2NC

5e. TASK NUMBER
PI

5f. WORK UNIT NUMBER
TT

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Electrical and Computer Engineering Department
University of Pittsburgh
4200 Fifth Ave
Pittsburgh, PA 15213

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/Information Directorate
Rome Research Site/RITB
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TP-2013-012

12. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA Case Number: 88ABW-2012-3651
DATE CLEARED: 27 JUNE 2012

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Conventional CMOS technology is slowly approaching its physical limitations and researchers are increasingly utilizing nanotechnology to both extend CMOS capabilities and to explore potential replacements. Novel memristive systems continue to attract growing attention since their reported physical realization by HP in 2008. Unique characteristics like non-volatility, re-configurability, and analog storage properties make memristors a very promising candidate for the realization of artificial neural systems. In this work, we propose a memristor-based design of bidirectional transmission excitation/inhibition synapses and implement a neuromorphic computing system based on our proposed synapse designs. The robustness of our system is also evaluated by considering the actual manufacturing variability with emphasis on process variation.

15. SUBJECT TERMS
Bidirectional synapse, memristor, Hopfield network, pattern recognition

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			NATHAN MCDONALD
U	U	U	UU	10	19b. TELEPHONE NUMBER (*Include area code*) N/A

The Circuit Realization of a Neuromorphic Computing System with Memristor-based Synapse Design

Beiye Liu[1], Yiran Chen[1], Bryant Wysocki[2], Tingwen Huang[3]

[1]ECE Dept., University of Pittsburgh, Pittsburgh, PA, USA
bel34@pitt.edu, yic52@pitt.edu
[2]U.S. Air Force Research Lab, Rome, NY, USA
Bryant.Wysocki@rl.af.mil, XXX
[3]Texas A&M University at Qatar, Doha, Qatar
tingwen.huang@qatar.tamu.edu

Abstract. Conventional CMOS technology is slowly approaching its physical limitations and researchers are increasingly utilizing nanotechnology to both extend CMOS capabilities and to explore potential replacements. Novel memristive systems continue to attract growing attention since their reported physical realization by HP in 2008. Unique characteristics like non-volatility, reconfigurability, and analog storage properties make memristors a very promising candidate for the realization of artificial neural systems. In this work, we propose a memristor-based design of bidirectional transmission excitation/inhibition synapses and implement a neuromorphic computing system based on our proposed synapse designs. The robustness of our system is also evaluated by considering the actual manufacturing variability with emphasis on process variation.

Keywords: Bidirectional synapse, memristor, Hopfield network, pattern recognition

1 Introduction

Although the existence of the memristor was predicted in 1971 by Professor Chua [1], it was not until 2008 that the first realization of a physical memristive system was reported by HP Labs [2]. Often regarded as the fourth fundamental circuit element, memristors are characterized by their pinched hysteresis loop such that all two terminal, non-volatile, resistive switching memories can be classified as memristors [3]. Memristors theoretically build upon the relationship between the magnetic flux(φ) and the electric charge(q) through the device as [1]:

$$d\varphi = M \cdot dq \tag{1}$$

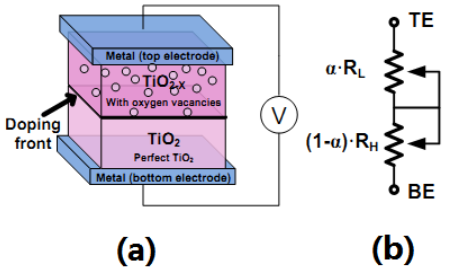


Fig. 1. (a) Structure of TiO_2 memristor (b) Conductivity equivalent circuit [4].

Figure 1(a) shows the conceptual structure of a TiO_2 memristor [2][4]. A perfect TiO_2 layer acts as an insulator or highly resistive conductor while the conductivity of the oxygen-deficient titanium dioxide (TiO_{2-x}) layer is much higher. The resistance of the entire memristive system can be controlled by translating the doping boundary between TiO_2 and TiO_{2-x}. As shown in Figure 1(b), the overall memristance can be described as:

$$M(\alpha) = \alpha \cdot R_l + (l - \alpha) \cdot R_h \tag{2}$$

Here, R_h and R_l represent the conductivities of TiO_2 and TiO_{2-x}, respectively. In general, memristors have the following unique properties that make them very promising devices for artificial neural system realization: First, memristance relies on the history of the total electric charge flowing through the device [1][5]. Second, memristors are non-volatile [6][7], which means that the memristance/resistance of the device is retained even after the system is powered off. No leakage or refresh power overheads are introduced into the memristor-based storage system. Lastly, memristors can be used as an analog device in which the resistance can be programmed continuously.

2 Principle of Memristor-based Synapse Design

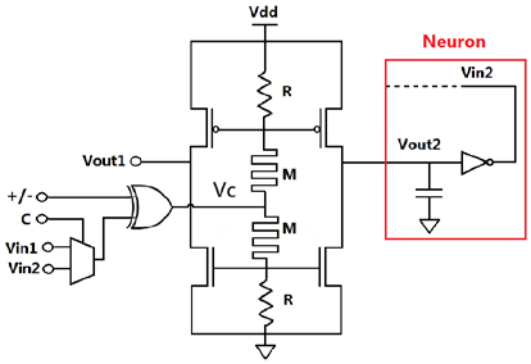
Previous circuit implementations of synapse designs mainly focus on basic functions such as multiplication, summation and comparison etc. [8]. In order to build more complex fully-connected networks (e.g., bidirectional associative memory) and to implement higher level applications such as pattern recognition, the synapse must be able to transmit both excitation and inhibition signals in two directions.

Figure 2(a) shows our proposed memristor-based bidirectional synapse designed to mimic the basic biological synapse structure. Besides the basic functions like multiplication and summation, our synapse design has the following new functions that could significantly simplify the corresponding neuron network designs:

Bidirectional transmission: in a real neural system, there exist two kinds of synapses (chemical synapse and electrical synapse) with quite different characteristics. Chemical synapse can only transmit signals in a single direction while an electrical synapse can transmit bidirectional signals. The bidirectional data transmission is es-

sential in some artificial neural networks where the neurons in the neuromorphic computing system need to communicate among each other. In our design, each synapse has two pairs of inputs/outputs (Vin1, Vin2/ Vout1, Vout2) that can connect two adjacent neurons. Such a structure allows the neuron to be either a message sender or a message receiver according to the control switch signal ‘c’.

Excitation/Inhibition: there are two ways neurons can communicate with each other; either through excitation or inhibition. In our design, a synapse can translate the absolute voltage amplitude of a neuron to either an excitation signal (pull-up current) or an inhibition signal (pull-down current). For example, in Figure 2(a), the input signal ‘Vc’ will turn on either a PMOS transistor or an NMOS transistor based on the input signals. When the PMOS transistor is turned on, the synapse output ‘Vout2’ will be connected to ‘Vdd’, denoting the excitation state. Alternatively, when the NMOS transistor is turned on, the ‘Vout2’ will be pulled down to ground, denoting the inhibition state. The amplitude of the output current is a function of the instantaneous resistance of the two memristors, as shown in Figure 1(b). A truth table depicting the operation of our synapse is shown in Table 1.



(a)

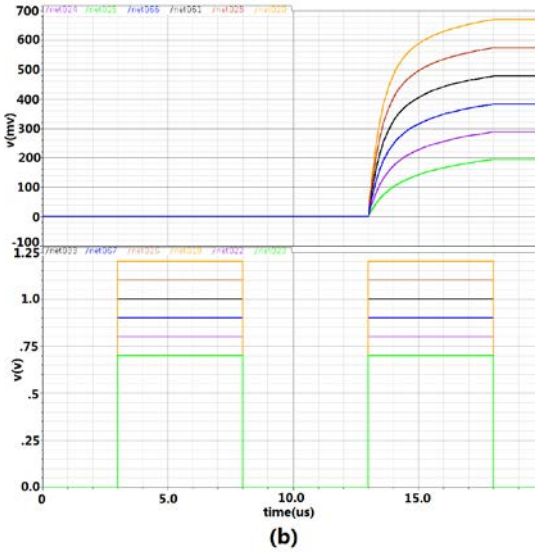


Fig. 2. (a) Bidirectional Synapse Circuit with One Neuron. (b) Weighted Output Current

Table 1. Excitation/Inhibition Synapse Truth Table

Vin1/Vin2	+/-	Vc	P-transistor	N-transistor	Vout1/Vout2
1	1	0	Pass	Cut off	Pull-up
1	0	1	Cut off	Pass	Pull-down
0	1	1	Cut off	Pass	Pull-down
0	0	0	Pass	Cut off	Pull-up

The neuron in our proposed design adopted a structure similar to some of our previous work [8], where a capacitance is serially connected with two inverters. The capacitance collects both the pull-up and pull-down signals generated from all the neurons which are connected to it through synapses, thus implementing the summation function. The neuron will change its state by comparing the inverter threshold to the voltage of the capacitance as:

$$\mathrm{N0} = \begin{cases} 1 & if\ \sum_{i=0}^{n} N_i \times W_i \geq threshold \\ 0 & otherwise \end{cases} \quad (3)$$

Eq. (3) signifies that the neuron (N0), collects the signals from all the other connected neurons (Ni) through the weighted connections (W_i). The neuron (N0) is excited (N0 = 1) if the voltage is higher than the threshold value and it is inhibited (N0= 0) if voltage is lower than the threshold value.

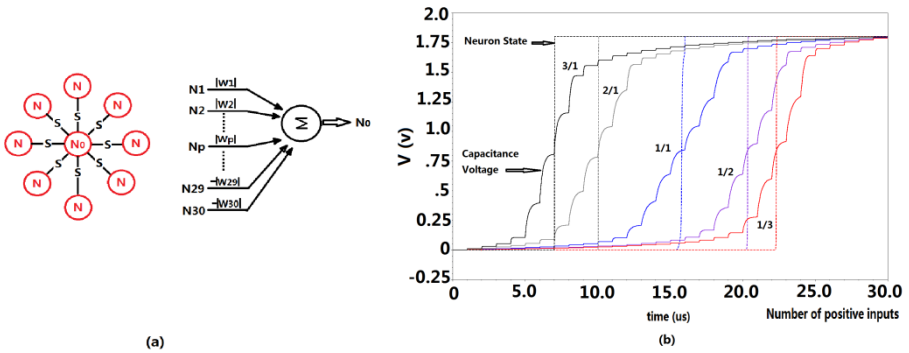


Fig. 3. (a) Signal Collection System. (b) Simulation Result.

To demonstrate the function of the proposed synapse, we constructed a signal collection system comprised of 31 neurons and 30 synapses, as shown in Figure 3(a). Neuron (N0) collects the signals from its 30 neighboring neurons through 30 synaptic connections. In this case, the information (voltage) stored in the 30 neurons (N1-N30) is the same while the weights and excitation/inhibition properties of the 30 synapses (W1-W30) are different. In our experiments, we utilized different ratios of excitation weight vs. inhibition weight (Wex)/($Winh$) and tested for the ratios of how many excitation synapses vs. inhibition synapses ($Sexc$)/($Sinh$) were required to excite neuron N0. All the circuit simulations are conducted under the Cadence Virtuoso environment while configured for 90nm technology. The simulation's results are shown in Figure 3(b). It shows that more than 15 excitation inputs are needed to excite 'N0' when the excitation weight is the same as the inhibition weight. When the excitation weights decrease, the required number of excitation inputs increases accordingly. The changing of neurons state always happens at the threshold as:

$$\text{Wexc} \times Sexc = Winh \times Sinh. \tag{4}$$

As aforementioned, the state of a neuron changes when the capacitance voltage exceeds the threshold value.

3 Neuromorphic Computing System Implementation

Although many CMOS implementations of analog neural networks (ANN) have been proposed [9][10], they generally suffer from the following issues:

Digital/Analog conversion: In neural network designs, the weights of the connections between neurons are analog. There are generally two ways to build analog connections. In the first approach, the analog weight can be digitized and stored in memory. However, an increase of the neural network size requires the enhancement of the data precision, which may lead to the requirement for an unacceptably large memory capacity. In the second approach, the analog weight can be stored in an analog memory device, i.e., as the voltage level on a capacitance. Compared to the digital

solution, such an approach suffers from the reliability issues such as charge sharing-induced voltage level fluctuation or leakage induced voltage level degradation. Thus, memory refreshing schemes become essential, and compensate the fluctuating voltage level on the capacitance at certain required time intervals. Both read and write operations are performed during the refresh.

Power consumption: Although analog neural networks generally offer a more compact structure and small footprint than that of a digital network, the power consumption of the analog neural network is usually higher. The major aspects of component power consumption include the current through the memristor and any other analog devices, as well as analog peripheral circuitry such as sense amplifiers. These power consumption issues become more and more severe as the network increases in size and complexity.

Our memristor-based synapse design is constructed on analog computation and totally eliminates the conversion between the analog and digital values. Also, the non-volatility of the memristor effectively reduces the standby power required by the computing circuit by eliminating the need for a refresh cycle.

We note that the previous research on memristor-based neural networks mainly focused on system-level simulations using high level languages [11][12] or restricted the synapse design to that with single memristors. Our research, however, includes system level circuit evaluations besides on dual-memristor designs. In the next section we use the character recognition of printed text as a case study to demonstrate our design concept.

4 Circuit-level Evaluation

We implemented a Hopfield network [11][12] (see Figure 4) with our proposed memristor-based bidirectional synapses and applied it to a pattern recognition application. The network is designed to learn and recognize several standard text patterns on a 4x3 grid, i.e., three text letter images of ‘A’, ‘B’, ‘C’ as shown in Figure 5(a). During the learning procedure, we first translate all standard patterns to vectors, each composed of 12 elements ($X_1, X_2, X_3 \ldots, X_{12}$), where black or white pixels are represented by the neurons with values of ‘1’ or ‘0’. Then the weight matrix is trained with these vectors based on the Hebbian learning rule [13][14]:

$$W = X_1^T \cdot X_1 + X_2^T \cdot X_2 \cdots \cdots + X_m^T \cdot X_m - mI. \quad (5)$$

After that, the input patterns that initialize the network will finally converge to a local minimum, which corresponds to one of the standard patterns that are used in training the weight matrix. The entire network scheme is depicted in Figure 4.

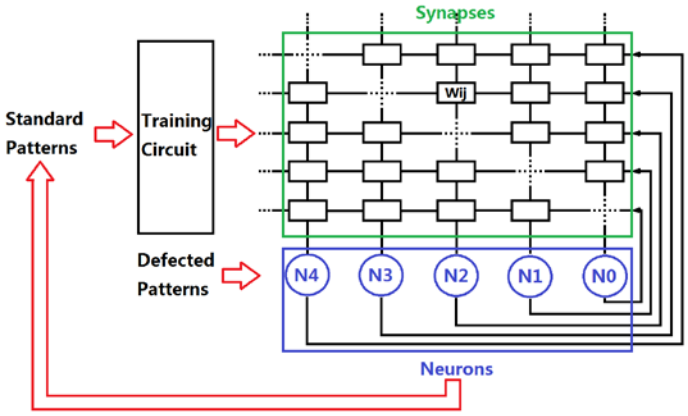


Fig. 4. Memristor-based Hopfield Network [9].

We evaluate the robustness and performance of our circuit by using the defected images in Figure 5(a) as input patterns. In the ideal case, after several iterations, the outputs of the neural network will converge to one of the standard patterns. Based on previous work on associative recall [15], the capacity limits of Hopfield network is ~15%, or about 2 patterns in a system of 12 neurons. But our initial simulation shows that our network works well in recognizing the defected patterns 'A', 'B' and 'C' (capacity of 25%), where 2 of the total 12 pixels are defective. After 2 iterations, the neuron states converged to a stable standard pattern that corresponds to the defected input pattern (shown in Figure 5(a)).

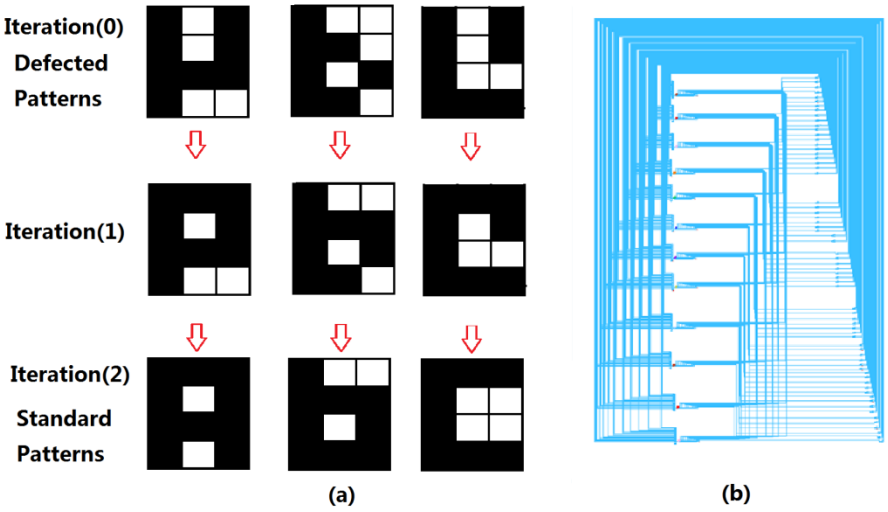


Fig. 5. (a) Defected/Standard Patterns. (b) Schematic of 12 Neurons Hopfield Network

We also conducted Monte-Carlo simulations to evaluate the impact of memristor process variations on the robustness of our networks. A large Hopfield network with 100 neurons was built to recognize larger sets of text patterns where the respective theoretical capacity is limited to about 15 patterns. Process variations were simulated by introducing Gaussian distribution noise to the memristance values of the memristive devices within the Matlab simulations. A system failure was defined as converging to a wrong standard pattern (ones that do not correspond to the input pattern), or by failing to converge to a stable point. The test results are shown in Figure 6. Here σ is the standard deviation of the memristance and P_f is the system failure rate.

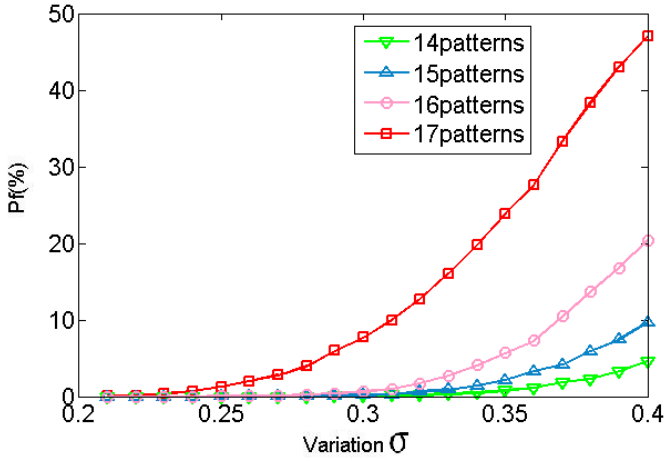


Fig. 6. Failure rate of memristor-based Hopfield network under different pattern numbers and process variation conditions.

Figure 6 shows that our design has a good immunity against process variations and performs well with a simulated P_f value close to zero even when the standard deviation reached values greater than 0.2 ($\sigma < 0.2$). The results confirm that increasing the number of text patterns quickly degraded the system's robustness showing much higher P_f values. When the number of patterns approaches the capacity limit, the slope of the failure rate rises quickly as the system robustness degrades. Increased process variations (σ) were also shown to degrade system robustness. However, in conventional CMOS circuit manufacturing, the parametric standard deviation is usually less than 10% [16].

5 Conclusion

In this paper, we proposed a novel memristor-based bidirectional synapse design that can transmit both excitation and inhibition signals. On top of that, we implemented a neural network circuit based on Hopfield networks and successfully demon-

strated character pattern recognition capabilities. Simulation results show that our design has very good immunity against process variations, offering reliable functionality under the normal variability range of the CMOS manufacturing process.

6 References

1. L. Chua, “Memristor-the missing circuit element,” IEEE Trans on Circuit Theory, vol. 18, pp. 507-519, 1971.
2. 2D. B. Strukov, et al., “The missing memristor found,” Nature, vol. 453, pp. 80-83, 2008.
3. L. Chua, “Resistance switching memories are memristors,” Applied Physics A: Materials Science & Processing, vol. 102, Issue 4, pp. 765-783, 2011.
4. H. Wang, H. Li, and R. Pino, “Memristor-based Synapse Design and Training Scheme for Neuromorphic Computing Architecture” International Symposium on Circuit and Systems, 2011.
5. X. Wang, et al., “Spintronic memristor through spin-torque-induced magnetization motion,” IEEE Electron Device Letters, vol. 30, pp. 294-297, 2009.
6. Y. Ho, G. Huang, and P. Li, “Nonvolatile memristor memory: device characteristics and design implications,” in Proceedings of the 2009 International Conference on Computer-Aided Design. ACM, 2009, pp. 485–490.
7. D. Niu, Y. Chen, C. Xu, and Y. Xie, “Impact of process variations on emerging memristor,” in Design Automation Conference (DAC), 2010, pp. 877–882.
8. G. Rose, R. Pino, Q. Wu, “A Low-Power Memristive Neuromorphic Circuit Utilizing a Global/Local Training Mechanism” in Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, 2011.
9. C. S, and H. Card, “Analog CMOS Deterministic Boltzmann Circuits” IEEE Journal of Solid-State Circuit, vol. 28, no. 8, 1993
10. D. Querlioz, O. Bichler, and C. Gamrat, “Simulation of a Memristor-Based Spiking Neural Network Immune to Device Variations” in Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, 2011
11. J. J.Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” in Proc. Nat. Acad. Sci., vol. 79, 1982, pp. 2554–2558.
12. S. Chartier, and R. Proulx, “NDRAM: Nonlinear Dynamic Recurrent Associative Memory for Learning Bipolar and Nonbipolar Correlated Patterns” in IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 16, NO. 6, NOVEMBER 2005
13. R. Rojas “Neural Networks” Springer-Verlag, Berlin, 1996
14. S. Haykin, Neural Networks: A Comprehensive Foundation. Englewood Cliffs, NJ: Prentice-Hall, 1999.
15. I. Kanter and H. Sompolinsky, “Associative recall of memory without errors,” Phy. Rev. A, vol. 35, pp. 380-392, 1987
16. R E. Pino, H. Li, Y. Chen, M. Hu, B. Liu. “Statistical Memristor Modeling and Case Study in Neuromorphic Computing” 49th Design Automation Conference (DAC), 2012: 585-590