

Optimizing the Distribution of Memristance Values of Memristive Synapses for Reducing Power Consumption in Analog Memristor Crossbar-Based Neural Networks



Son Ngoc Truong

Abstract: Memristor circuits have become one of the potential hardware-based platforms for implementing artificial neural networks due to a lot of advantageous features. In this paper, we compare the power consumption between an analog memristor crossbar-based a binary memristor crossbar-based neural network for realizing a two-layer neural network and propose an efficient method for reducing the power consumption of the analog memristor crossbar-based neural network. A two-layer neural network is implemented using the memristor crossbar arrays, which can be used with analog synapse or binary synapse. For recognizing the test samples of MNIST dataset, the binary memristor crossbar-based neural work consumes higher power by 19% than the analog memristor-based neural network. The power consumption of the analog memristor crossbar-based neural network strongly depends on the distribution of memristance values and it can be reduced by optimizing the distribution of the memristance values. To improve the power efficiency, the bias resistance must be selected close to high resistance state. The power consumption of the analog memristor-based neural network is reduced by 86% when increasing the bias resistance from 20KΩ to 160KΩ. For the bias resistance of 160KΩ, analog memristor crossbar-based neural network consumes less power by 89% than the binary memristor crossbar-based neural network.

Keywords: memristor, memristor crossbar, memristive synapse, handwritten digit recognition.

I. INTRODUCTION

VLSI (Very-large-scale Integration) implementation of artificial neural networks has been extensively studied for last decades [1]-[4]. The VLSI designs are predominantly based on the CMOS technology, however, CMOS technology seem to be saturated because of CMOS scaling limits [5],[6]. Therefore, beyond-CMOS devices have been invented for implementing artificial neural networks. In 1971, professor Leon O. Chua found the fourth circuit element which was named “Memristor” [7]. Memristors have attracted wildly researchers since they were experimentally demonstrated in TiO₂ film in 2008 [8]. The flexibly

modifiable conductance of memristor is potential for mimicking the functionalities of biological neurons in human’s brain [9]. Memristor’s conductance, which is so-called “memristance” can be modified according to the current flowing through the device. According to such characteristic, memristor has potentially used in realizing the artificial neural networks. Memristors can be used as analog devices or binary devices [10]. Analog memristors are mainly based on the interface-switching behavior, in which the memristance value can be changed gradually and precisely controlled by the applied voltage or current [10]. Under this characteristic, the analog data can be stored in memristors with high accuracy [10]. Fig 1(a) shows a cross-sectional view of an interface-switching memristor based on LaAlO₃ film [11]. A measurement of the voltage and current relationship of the memristor device in Fig. 1(a) is shown in Fig. 1(b). Here, we sweep the memristor’s voltage between -3V and +3V. The memristance is gradually changed when the positive voltage is applied to the device, as shown in Figure 1(b). Another behavior of the memristor device is known as filamentary-switching mechanism [10]. Memristors have only two states; high resistance state (HRS) and low resistance state (LRS) without intermediate values.

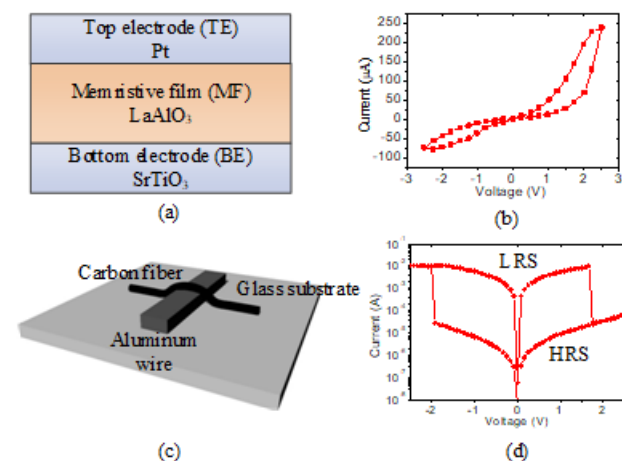


Fig. 1.(a) An interface-switching memristor fabricated on LaAlO₃ film, (b) the nonlinear voltage and current relationship of the memristor device shown in Fig. 1(a). (c) A filamentary-switching memristor and (d) the switching behavior of the memristor presented in Fig. 1(c).

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Son Ngoc Truong, HCMC University of Technology and Education, Ho Chi Minh City, Vietnam.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Optimizing the Distribution of Memristance Values of Memristive Synapses for Reducing Power Consumption in Analog Memristor Crossbar-Based Neural Networks

Filamentary-switching memristors are also called binary memristors which can store '1' or '0' represented by LRS or HRS [10]. Fig. 1(c) shows a schematic of a binary memristor which is made of carbon fiber [12]. Fig. 1(d) shows a switching characteristic of the binary memristor shown in Fig. 1(c). In Fig. 1(d), the memristance switches between HRS and LRS when a negative or positive voltage is applied to the device. Memristors can be fabricated between two layers of metal to form crossbar arrays, in which any two crossing wires are connected by a memristor. Memristor crossbar arrays have been demonstrated effective in realizing synapse arrays of artificial neural networks. To realize an artificial neural network, we can consider the memristor crossbar architecture proposed in the previous work. It utilizes only one memristor crossbar array for implementing the signed synaptic weight array [13]. The proposed memristor crossbar architecture can be used to realize the neural network in which memristors are used as analog or binary devices [14], [15]. In the previous works, the recognition rate and the variation tolerance of the analog memristor crossbar-based neural network and the binary memristor crossbar-based neural network are compared [10], [14]. The binary memristor crossbar-based neural network exhibits a better recognition rate and variation tolerance than the analog memristor crossbar-based neural network [10], [14]. In this work, we compare the power consumption between the analog memristor-based neural network and the binary memristor-based neural network and analyze the power consumption of the analog memristor-based neural network with varying the distribution of memristance values of memristor synapses. From the analysis results, we proposed an optimal distribution of memristance values for reducing the power consumption of analog memristor-based neural network.

MEMRISTOR CROSSBAR CIRCUITS FOR REALIZING NEURAL NETWORKS

A conceptual diagram of a two-layer neural network for an application of handwritten digit recognition is shown in Fig 2(a). The input layer has 784 neurons for 784 pixels of handwritten digits obtained from MNIST dataset. The hidden layer has 1024 neurons. The 10 neurons of the output layer for recognizing the 10 handwritten digits. Fig. 2(b) shows a schematic of the crossbar circuit for realizing the two-layer neural network conceptually shown in Figure 2(a). The crossbar has 784 rows for 784 input neurons and 1024 columns that implement the 1024 neurons in the hidden layer. Activation function circuits used in this neural network are the sigmoid functions realized by Op-amp circuits [14]. The outputs of the first layer are propagated to the second crossbar array as shown in Fig. 2(b). In Fig. 2(b), the output of the k th neuron in the first layer is calculated as follows [13]

$$V_{O,k} = \sum_{j=1}^m w_{j,k} V_{IN,j} \quad (1)$$

$$\text{where } w_{j,k} = R_0 \left(\frac{1}{R_B} - \frac{1}{M_{j,k}} \right)$$

Here $M_{j,k}$ is the conductance of the crosspoint located at the j th row and the k th column in crossbar array [13]. $w_{j,k}$ is a synaptic weight between the j th input and the k th neuron. The

synaptic weight $w_{j,k}$ is obtained by adjusting the memristance value $M_{j,k}$, m is the number of the inputs.

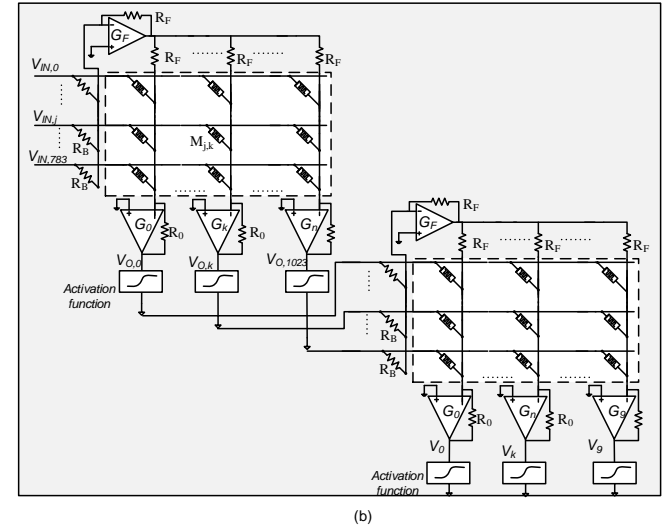
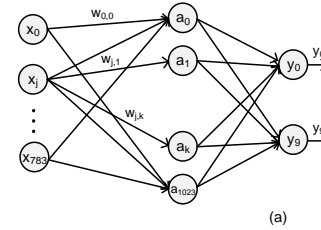


Fig. 2. (a) A two-layer neural network for handwritten digit recognition, (b) the schematic of a memristor crossbar-based two-layer neural network.

Equation 1 is for calculating the output of the k th neuron. The synaptic weights are determined by the value of R_0 , R_B , and the memristance values, $M_{j,k}$. The R_0 is for amplifying the synaptic weight. R_B is a bias resistance value. The sign and amplitude of synaptic weights can be achieved by adjusting the values of memristances while keeping R_B constant. R_B is assigned between the LRS and HRS. The memristor crossbar circuit in Figure 2(b) can be used as an analog memristor crossbar-based neural network or a binary memristor-based neural network [13], [14]. For analog synapses, synaptic weights obtained from training process are converted to memristance values using the equation 1, where memristance values are in range of from low resistance state to high resistance state. It is interesting that the architecture in Fig. 2 can be used to implement the binary neural network, in which the synaptic weights are either -1 or $+1$ [14]. The synaptic weight of -1 and $+1$ can be represented by the memristor with memristance value is calculated by Equation 1 and Equation 2 as follows

$$w_{j,k} = \begin{cases} +1, & \text{if } M_{j,k} = \text{HRS} \\ -1, & \text{if } M_{j,k} = \text{LRS} \end{cases} \quad (2)$$

$$\text{where } R_0 = 20K\Omega, R_B = \frac{2}{\left(\frac{1}{\text{HRS}} + \frac{1}{\text{LRS}} \right)}$$

The value of R_0 and R_B are fixed to guarantee that the synaptic weight is -1 or $+1$ for the memristance value is LRS or HRS respectively.

II. PROPOSED METHOD FOR REDUCING THE POWER CONSUMPTION IN ANALOG MEMRISTOR CROSSBAR-BASED NEURAL NETWORKS

For realizing the binary neural network, the memristance values are calculated by using equation 2, where the value of R_B is fixed to guarantee that the HRS and LRS represent the synaptic weight of +1 and -1 respectively. By doing this the power consumption of binary memristor crossbar-based neural network is independent of the bias resistance, R_B . For analog synapses, memristance values obtained from Equation 1 are distributed around the value of R_B . The value of R_B is selected as long as the minimum and the maximum of synaptic weights can be represented by the memristance values which are in range of the LRS and HRS. Selecting the value of R_B strongly affects the distribution of memristance values.

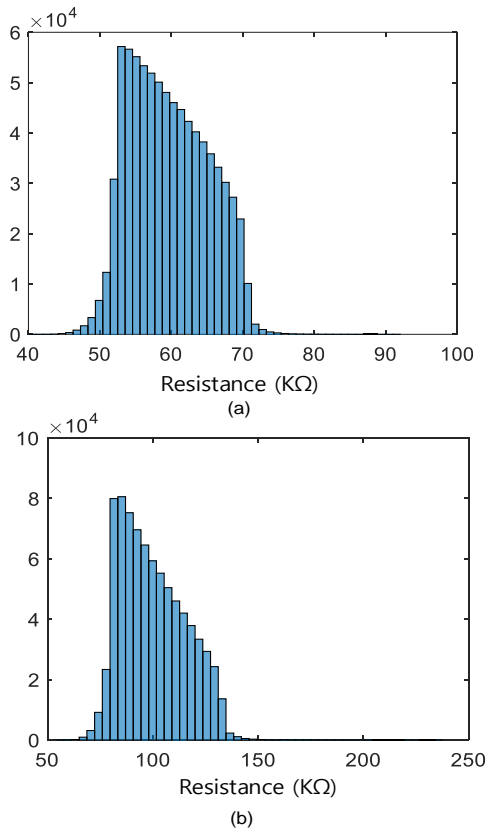


Fig. 3. The distribution of memristance value with (a) $R_B = 60K\Omega$ and (b) $R_B = 100K\Omega$

Fig. 3 shows the distribution of memristance values representing synaptic weights of an analog memristor-based neural network. In Fig. 3(a), the value of R_B is $60K\Omega$, memristance values are distributed around the value of R_B . When R_B is $100K\Omega$, the distribution of memristance values goes forward to high resistance state, as shown in Fig. 3(b). It emerges that the distribution of memristance values depends on the value of R_B . It is due to the fact that memristance values are calculated by equation 1 in which R_B can be considered as a bias for generating negative of positive synaptic weights. Because the power consumption of the crossbar circuit is low when memristance values are high, in this paper, we propose a method for reducing the power consumption of crossbar circuit by selecting the optimal distribution of memristance value. It can be done by selecting the value of R_B close the high resistance state. And, the value of R_B is selected as long as the maximum value of

memristance converted from the maximum of synaptic weight must be in range of low resistance state and high resistance state.

III. RESULT AND DISCUSSION

The memristor crossbar circuit shown in Figure 2 (b) is used to implement an analog neural network and a binary neural network for the application of handwritten digit recognition. For the analog neural network, the crossbar bar circuit is trained using well-known back-propagation algorithm by Matlab. The obtained synaptic weights are converted to the values of memristance by using Equation 1. Memristor crossbar circuit is then simulated by Cadence Spectre circuit simulation [16]. For the binary neural network, the crossbar circuit is trained using by the modified version of back-propagation algorithm [17]. The obtained synaptic weight are converted to the memristance values using Equation 2. The MNIST dataset is used in the simulation to estimate the power consumption of the crossbar circuit when it implements the analog synapse array and the binary synapse array. For recognition of 10,000 test samples, the binary memristor-based neural network consumes 6.63W. The power consumption of the analog memristor-based neural network depends on the value of bias memristance, R_B since the memristance values are distributed around R_B , as aforementioned.

Table- I: The power consumption of the analog and binary memristor crossbar-based neural networks

Bias resistance, R_B (KΩ)	Power consumption (W)	
	Binary memristor-based NN	Analog memristor-based NN
20	6.63	5.383
40		2.723
60		1.836
80		1.392
100		1.126
120		0.949
140		0.822
160		0.727

Table I shows the power consumption of the binary memristor crossbar-based neural network and the analog memristor crossbar-based

neural network for recognizing 10,000 test samples, where the bias resistance of R_B varies from $20K\Omega$ to $160K\Omega$. The power consumption of binary memristor crossbar-based is independent to the value of bias resistance, as aforementioned. By contrast, the power consumption of analog memristor crossbar-based decreases as the value of R_B increases, as shown in Table I. When R_B is $20K\Omega$, the analog memristor-based neural network consumes 5.28W. The analog memristor crossbar-based neural network consumes less power by 19% than the binary memristor crossbar-based neural network for recognizing 10,000 test samples. When R_B is $160K\Omega$, the analog memristor-based neural network consumes 0.727W, 89% lower than the binary memristor crossbar-based neural network. It can be inferred that the power consumption of the analog memristor crossbar-based neural network is reduced significantly when increasing the value of bias resistance, R_B . In particular, when the bias resistance increases from $20K\Omega$ to $160K\Omega$, the power consumption can be saved by 86%.

Optimizing the Distribution of Memristance Values of Memristive Synapses for Reducing Power Consumption in Analog Memristor Crossbar-Based Neural Networks

IV. CONCLUSION

In this work, we compared the power consumption of the analog and binary memristor crossbar circuits for realizing the two-layer neural network and improve the power-efficiency of analog memristor crossbar-based neural network by optimizing the memristance distribution. For the analog memristor crossbar-based neural network, the memristance values are distributed around the bias resistance value and in range of from LRS to HRS. Therefore, controlling the value of bias resistance results in the change of memristance distribution. When the bias resistance R_B is set to be $20K\Omega$, the analog memristor crossbar-based neural network consumes less power by 19% than the binary memristor crossbar-based neural network. When R_B increases, the power consumption of the analog memristor-based neural network decreases significantly. For the bias resistance of $160K\Omega$, analog memristor crossbar-based neural network consumes less power by 89% than the binary memristor-crossbar neural network. The power consumption of memristor-based neural network is reduced by 86% when increasing the bias resistance from $20K\Omega$ to $160K\Omega$.

REFERENCES

1. A. Ardakani et al, "VLSI Implementation of Deep Learning Neural Network Using Integral Stochastic Computing," IEEE Trans. Very Large Scale Integration System, vol. 25, iss. 10, pp. 2688-2699, Feb. 2017
2. M. Walker, P. Hasler, and L. Akers, "CMOS neural network for pattern association," IEEE Micro, vol. 9, no. 5, pp. 68-71, Oct. 1989
3. B. V. Benjamin et al, "Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations," Proc. IEEE, vol. 102, no.5, pp.-699-716, Apr. 2014
4. P. A. Merolla et al, "A million spiking- neuron integrated circuit with a scalable communication network and interface," Science, vol. 345, no. 6197, pp 668-673, Aug. 2014
5. P. M. Solomon, "Device Innovation and Material Challenges at the Limit of CMOS Technology," Annu. Rev. Mater. Sci., vol. 30, pp. 681-697, Aug. 2000
6. T. P. Brđanin and B. Dokić, "Strained Silicon Layer in CMOS Technology," Electronics, vol. 18, no. 2, pp. 63-69, Dec. 2014
7. L. O. Chua, "Memristor – the missing circuit element," IEEE Trans. Circuit Theory, vol. CT-18, no. 5, pp. 507-519, Sep. 1971.
8. D. B. Strukov, G. S. Sinder, D. R. Stewart, and R. S. Williams, "The missing memristor found," Nature, vol. 453, pp. 80-83, May 2008.
9. S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," Nano Letters, vol. 10, no. 4, pp. 1297-1301, Mar. 2010.
10. S. N. Truong, S. J. Ham, and K. S. Min "Neuromorphic crossbar circuit with nanoscale filamentary-switching binary memristors for speech recognition," Nanos. Res. Lett., vol. 9 no. 629, pp. 1-9, Nov. 2014.
11. S. N. Truong et al., "New pulse amplitude modulation for fine tuning of memristor synapses," Microelectronics Journal, vol. 55, pp. 162-168, 2016
12. S. N. Truong et al., "Experimental Demonstration of Sequence Recognition of Serial Memristors," Elec. Mat. Lett., vol. 13, no. 1, pp. 86-90, Jan. 2017
13. S. N. Truong and K. S. Min "New memristor-based crossbar array architecture with 50-% area reduction and 48-% power saving for matrix-vector multiplication of analog neuromorphic computing," J. Semi. Tech. Sci., vol. 14, no. 3, pp. 356-363, Jun. 2014.
14. K. V. Pham et al., "Memristor binarized neural networks," J. Semi. Tech. Sci., vol. 18 no. 5, pp. 568- 577, 10. 2018
15. Y. Zhang, X. Wang, E. G. Friedan, "Memristor-Based Circuit Design for Multilayer Neural Networks," IEEE Trans. Circuits and Systems-I, Regular Paper, vol. no. 2, pp. 677-686, Aug. 2017.
16. Virtuoso Spectre Circuit Simulator User Guide, Cadence, San Jose, CA, USA, 2004
17. T. Simons and D. J. Lee, "A Review of Binarized Neural Network," Electronics, vol. 8, no. 661, pp. 1-25.

AUTHORS PROFILE



Son Ngoc Truong received the B.S. and M.S. degrees in Electronic Engineering from HCMC University of Technology and Education, HCM City, Vietnam, in 2006 and 2011, respectively. He received the Ph.D degree in Electronic Engineering from Kookmin University, Seoul, Korea in 2017. He is currently lecturer at the Faculty of Electrical and Electronics Engineering, HCMC University of Technology and Education, HCM City, Vietnam. His research interest include memristor devices – modeling and application, circuit design, neuromorphic computing systems.