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 network 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.

1. 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).

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

II. 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 kth neuron in the first layer is calculated as follows [13]

\[ V_{O,k} = \sum_{j=1}^{m} w_{j,k} V_{I,j} \]

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 jth row and the kth column in crossbar array [13]. \( w_{j,k} \) is a synaptic weight between the jth input and the kth neuron. The synaptic weight \( w_{j,k} \) is obtained by adjusting the memristance value \( M_{j,k} \). m is the number of the inputs.

Equation 1 is for calculating the output of the kth 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. \( R_0 \) 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

\[ w_{j,k} = \begin{cases} 1, & \text{if } M_{j,k} = \text{HRS} \\ -1, & \text{if } M_{j,k} = \text{LRS} \end{cases} \]

where \( R_0 = 20k\Omega \), \( R_B = \frac{2}{\text{HRS} + \frac{1}{\text{LRS}}} \)

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.
III. 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.

![Graph](image)

**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 60KΩ, memristance values are distributed around the value of $R_b$. When $R_b$ is 100KΩ, 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.

IV. 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 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>
<thead>
<tr>
<th>Bias resistance, $R_b$ (KΩ)</th>
<th>Power consumption (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary memristor-based NN</td>
<td>Analog memristor-based NN</td>
</tr>
<tr>
<td>20</td>
<td>5.383</td>
</tr>
<tr>
<td>40</td>
<td>2.723</td>
</tr>
<tr>
<td>60</td>
<td>1.836</td>
</tr>
<tr>
<td>80</td>
<td>1.392</td>
</tr>
<tr>
<td>100</td>
<td>1.126</td>
</tr>
<tr>
<td>120</td>
<td>0.949</td>
</tr>
<tr>
<td>140</td>
<td>0.822</td>
</tr>
<tr>
<td>160</td>
<td>0.727</td>
</tr>
</tbody>
</table>

Table 1 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 20KΩ to 160KΩ. 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 20KΩ, 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 160KΩ, 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 20KΩ to 160KΩ, 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

V. 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 20KΩ, 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 160KΩ, 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 20KΩ to 160KΩ.

REFERENCES