Artificial neuron operations and spike-timing-dependent plasticity (STDP) using memristive devices for brain-inspired computing

Takao Marukame, Reika Ichihara, Masafumi Mori, Yoshifumi Nishi, Shin-ichi Yasuda, Tetsufumi Tanamoto and Yuichiro Mitani

Corporate R&D Center, Toshiba Corporation
1 Komukai-Toshiba-cho Saiwai-ku, Kawasaki 212-8582, Japan
Phone: +81-44-549-2316  E-mail: takao.marukame@toshiba.co.jp

Abstract

The use of memristive devices for artificial neurons is promising from the viewpoints of computation architecture and learning protocol. We propose an energy-efficient multiplier accumulators based on memristive array architecture. Spike-timing-dependent plasticity (STDP) is then demonstrated in metal-oxide memristive devices. We established STDP protocols that can control analog resistive changes periodically with long range stabilities.

1. Introduction

Neuromorphic devices have attracted much attention for artificial intelligence applications based on brain-inspired deep neural networks (DNN) [1, 2]. The growing needs of artificial intelligence applications based on brain-inspired computing will require efficient hardware systems for updating the connection weights of network, which is the most time- and energy-consuming part of the DNN algorithm. Recently, binary connection has been applied to enable efficient DNN computing through time-domain analog-digital mixed-signal processing utilizing memristive cells in small circuit areas with high energy efficiency [2]. However, continuous weights are still necessary for learning process. In this work, we propose two approaches at the circuit and device levels to exploit an efficient memristive system for a continuously weighted synapse for artificial neurons.

2. Artificial neurons with sequential synapse operation

To realize compact hardware with real-value multiplier accumulators (MACs), we propose sequential synapse operations for neural networks on an arrayed structure (Fig. 1).

The sequential bit calculation method has been proposed and demonstrated in a classical neuromorphic chip to save energy [3]. The calculation is represented by the equations (1) and (2), where \( w[k] \) is a k-th bit of fixed-point weight \( w \), \( x \) is input, \( b \) is bias, \( f \) is a non-linear function with saturation regions (e.g., sigmoid) and \( w[m] \) represents the MSB of \( w \).

\[
\begin{align*}
y_i &= f \left( \sum_{k=0}^{\infty} (w[k] \cdot x_k) - b \right) \\
u &= w_{ij} \cdot x_i = -2^m w_j [m] + \sum_{k=1}^{m} 2^{m-k} w_{ij}[m-k] \cdot x_i
\end{align*}
\]

The calculation of \( w_{ij} \) and \( x_i \) is carried out in divided multiplications of one bit of \( w \) and all bits of \( x \) followed by the summation. The initial summation \(-2^m w_{ij}[m] \cdot x_i \) can be negative or zero. In the following steps for \( w[m-1:0] \) in horizontal direction shown in Fig. 1(b), the calculated summation values monotonically increase. As Fig. 2(a) illustrates, if the temporal summation reach a threshold to judge that \( f \) reaches its saturation, the sequence can end the calculation and thus save the energy cost. Figure 2(b) shows that the average number of computation steps is reduced to half of the default value in a restricted Boltzmann machine (RBM) learning with MNIST dataset [4].

We validated the proposed calculations by implementing the sequence in FPGA, and we found that the 8 MACs with 256 inputs would consume all resources of the 530k logic elements. In contrast, our array-based architecture would require less than 1/100 of the equivalent circuit area resource. In Fig. 3, steps in MAC operations by the bit-sequential algorithm are shown as an example. While sequential architecture with parallelized accumulators takes over a thousand of steps, proposed memristive array with parallel calculation in vertical direction (see Fig. 1(b)) requires only 8 steps in this example and \( m-1 \) steps at most.

3. Spike-timing-dependent plasticity (STDP)

Memristive devices applied in an analog manner are promising for use as a synapse to control weight values while needing only small area [5, 6]. This is in contrast with digital manner which requires many bit cells as explained in Sec. 2. In metal-oxide devices such as CMOS process compatible TiO\(_2\) and AlO\(_x\), high density MAC operations have been demonstrated [6] which follow the simple principle shown in eq. (3) based on the conductance \( G \) of cells and applied voltage \( V \) through the cell:

\[
l_i = \sum_{j=1}^{n} g_{ij} \cdot v_i
\]

Experimental STDP was observed in memristive devices by Jo et al. [7]. Subsequently, STDP learning algorithms and circuits have been explored for application in robust spiking neural networks and unsupervised learning [5]. However, the application of STDP to MAC operations such as those described above has not yet been clearly investigated or demonstrated. We thus have evaluated the switching dynamics of TiO\(_2\)/AlO\(_x\) memristive devices to demonstrate STDP with the experimental set-ups shown in Fig. 4 [5, 7]. Intermediate states were observed for both conductance decrease (depression) and increase (potentiation) as shown in Fig. 5, exhibiting STDP characteristics. By using different STDP protocols with varying pulse widths, voltages, etc., we have successfully demonstrated the depression and potentiation cycles in the metal-oxide device (Fig. 6), indicating that STDP can hold the long range periodic stabilities of the memristive devices and therefore can be used as a learning method. Moreover, as shown in Fig. 7, the current flow as a neuron output corresponds to
\((1/R_1 + 1/R_2 + 1/R_3)V = (G_1 + G_2 + G_3)V\), which is consistent with the calculation principle of eq. (3). As a step towards achieving practical brain-inspired computing, scalability to a large number of neurons/synapses and robustness to noise and variability should be investigated. To take full advantage of neuromorphic approaches having tolerance against variability, we need to develop hardware modeling that will enable simulation which takes into account of the variability of cells [1].

4. Conclusions
We have proposed and analyzed artificial neurons using memristive devices with our energy-efficient MAC architecture. Using the STDP sequence, analog resistances with reduced variability can be well controlled depending on pulse protocols, which will give us promising ways to achieve high density and non-volatile computing.

Acknowledgements
We would like to thank Drs J. Deguchi, D. Miyashita and S. Kawanaka of Toshiba Memory Corporation and Prof T. Asai of Hokkaido Univ. and Dr A. Schmid of EPFL for their kind support and discussions.

References

Fig. 1 Neural networks: (a) two-layer model (b) proposed architecture using memristive devices to exploit the sequential synapse MAC algorithm.

Fig. 2 Proposed MAC operations: (a) illustration of the proposed sequence, (b) steps in MAC operations up to steps (black circle) by reaching the threshold, which were taken from 256 neurons in RBM learning. The green dashed line shows the numbers of steps in default calculations. The blue solid line shows the average of the black data.

Fig. 3 An example of bit-sequential MAC operations: (a) the fully-sequential calculation (black triangle) for one neuron extracted from data of Fig. 2(b), (b) a magnification of (a), showing the case of the proposed memristive array (red circle) with parallel calculations in vertical direction.

Fig. 4 Definition of pulse timing applied to TE and BE as post and pre pulses, respectively: (a) potentiation configuration (b) depression configuration. The pulse width \(\tau\) is tested from 100 \(\mu\)s to 100 ns.

Fig. 5 Demonstration of STDP in a memristive synapse: (a) the direct plots from measurements (b) a well-known STDP visualization taken from (a). Spike timing \(\Delta t\) is defined with \(\Delta t = (\tau_{\text{post}} - \Delta T)\) for depression and \(\Delta t = (\Delta T - \tau_{\text{post}})\) for potentiation. The lines serve as a guide for the eyes.

Fig. 6 Experimental depression and potentiation cycles in a TE/TiO\(_x\)/AlO\(_x\)/BE stimulated by two different STDP protocols: (a) potentiation is linear, depression is not (b) potentiation and depression are balanced. (a-1) and (b-1) are magnifications of (a-2) and (b-2), respectively.

Fig. 7 Synapse and neuron characteristics of 3 cells in a crossbar line driven by STDP: (a) resistance changes of synapses by STDP pulses (b) current flows from the three synapses corresponding to \(I = (G_1 + G_2 + G_3)V\).