Rectifying Resistive Memory Devices as Dynamic Complementary Artificial Synapses

Dan Berco*

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore

Brain inspired computing is a pioneering computational method gaining momentum in recent years. Within this scheme, artificial neural networks are implemented using two main approaches: software algorithms and designated hardware architectures. However, while software implementations show remarkable results (at high-energy costs), hardware based ones, specifically resistive random access memory (RRAM) arrays that consume little power and hold a potential for enormous densities, are somewhat lagging. One of the reasons may be related to the limited excitatory operation mode of RRAMs in these arrays as adjustable passive elements. An interesting type of RRAM was demonstrated recently for having alternating (dynamic switching) current rectification properties that may be used for complementary operation much like CMOS transistors. Such artificial synaptic devices may be switched dynamically between excitatory and inhibitory modes to allow doubling of the array density and significantly reducing the peripheral circuit complexity.

Keywords: artificial neural networks, brain inspired computing, dynamic artificial synapses, memristors, rectifying synapses

INTRODUCTION

Ever since the scientific community's revival of interest in memristors (Chua, 1971; Chua and Kang, 1976) was triggered by publications in the last decade such as Strukov et al. (2008), these devices have been extensively used for the implementation of artificial neural networks (ANN) in brain-inspired computational platforms. Within this domain, crossbar array architectures are promising candidates for achieving high-densities (~10^{15} bits/cm^{2}) similar to the human cerebral cortex (~10^{14} synapses), when configured in 3D stacking (Küglera et al., 2009), due to the nanoscale device dimensions (Aratani et al., 2007; Jo et al., 2010). In addition, crossbar arrays are very efficient, in terms of calculation time and energy expenditure, when performing matrix-vector dot product operations, that form the basis for machine learning algorithms (Hu et al., 2012). Several demonstrations of resistive random access memory (RRAM) array implementations proved to be very successful in tasks such as image classification (Prezioso et al., 2015). The sizes of these networks range from small scales of only few neurons (Kim et al., 2012; Prezioso et al., 2015), medium levels (Park et al., 2015; Hu et al., 2016) and up to larger scale that incorporate hundreds of neurons (Yao et al., 2017) and even up to 10^{5} synaptic connections (Burr et al., 2015).

Resistive random access memory arrays are designed to imitate the functionality of biologic synaptic networks. A chemical synapse is a gapped connection between two neurons through which communication takes place (Pereda and Faber, 1996; Kandel et al., 2000;
Nicholls et al., 2001). A typical neuron can have several thousands of synapses which mostly connect axons in a presynaptic neuron to dendrites in postsynaptic neurons. Inter-neural signaling occurs by the release of neurotransmitters from the presynaptic neuron into the gap (i.e., synaptic cleft) that in turn is collected by receptors in the postsynaptic neuron. The molecular neurotransmitters are kept in sacs called synaptic vesicles. During signaling, these vesicles are released into the synaptic cleft and bind to receptors on the postsynaptic neuron. Once the signal is delivered, the transmitters are evacuated from the receptors through potential mechanisms such as enzymatic degradation, or absorbed back into the presynaptic neuron by specific transporters. The postsynaptic potential response is classified as being either excitatory or inhibitory and determined by the type of neurotransmitter (Glutamate or γ-aminobutyric acid) (Nakanishi, 1992). Two key characteristics resulting from this behavior are the so-called long-term potentiation (LTP) and long-term depression (LTD) and the synaptic weight (connection strength) is modulated by this neural activity. A recent study showed that actually both types of neurotransmitter could be released simultaneously during synaptic activity (Root et al., 2014). Moreover, neurotransmitters have been shown to be able to actually exchange roles during early stages of brain development (Ben-Ari, 2002).

Biologic neural networks have evolved over hundreds of millions of years to easily and efficiently perform tasks that state of the art computers find difficult. The operation of a man-made ANN should thus be true to the source as much as possible with respect to a building block artificial device. In order to achieve this target, artificial synaptic devices (ASDs) should imitate as much as possible the traits that biological synapses have. Although current understanding of neural networks and synapses is nowhere near complete, one may assume that such ASDs would be the best option for future ANN implementations. Some of these basic features include having a large dynamic range and multilevel operation to match the analog nature of the biological synaptic weight changes. These traits translate to higher accuracies, more degrees of freedom for weight adjustment and robustness during network training. A positive correlation exists between device dimensions and the number of states it can support (e.g., multilevel resistance) (Kuzum et al., 2013). However, sizing up the device will increases the overall current and power consumption as well as reduce the potential density.

The potential ability to implement symmetric weight changes may play a role in the simplification of an ANN peripheral control system. Biologic neural networks are very adaptive and can easily compensate for asymmetric weight changes especially when hundreds of neurons are involved in determining the weight of a synaptic junction. State machine based control systems on the other hand, are best suited to operate with well known and predictable parameters. In order to deal with asymmetric or random parametric distributions, an elaborate feedback system must be implemented and incorporated into the controller. In this sense, an ASD having a potential for incrementally small changes in both the up and down directions may simplify the controller design. A desirable corresponding weight parameter (e.g., conductance in RRAM) should thus be both symmetric with regards to the direction of change (increase or decrease) and differentially linear in magnitude of step change. Nonetheless, RRAMs (being a promising candidate in terms of low-power and high-density) (Wong et al., 2012) and other types of memristors show non-linear conductance as well as asymmetric conductivity changes (Lee et al., 2010) in response to successive set and reset pulses (Alibart et al., 2012; Chen et al., 2015; Wang et al., 2016) that complicate the task of designing a control state machine and sensing circuitry. This in turn affects both the potentially achievable network accuracy and overall performance.

An even more critical attribute expected from an ideal ASD would be the ability to reconfigure dynamically during real time operation between the excitatory and inhibitory response modes in a similar manner to a biologic synapse. Current RRAM devices are not able to reproduce this feature since they are passive devices by nature (once formed) and intentionally operated to comply with the linearity requirements as much as possible. It is virtually impossible for an RRAM to display this dynamic attribute without adding an additional control terminal to modulate the material properties by the field effect. In other words, it should have a bipolar conductance that is distinct characteristic of active devices. Adding a control terminal would damage the linearity and seriously downgrade the high-integration capabilities.

A common architectural solution is used to work around the dynamic reconfiguration issue by employing a differential approach. Instead of representing a synaptic gap by a single RRAM, two devices are used in a differential manner (Bichler et al., 2012; Prezioso et al., 2015). In this way, the synaptic weight-current is evaluated through a differential amplifier to determine whether the synaptic gap represents an excitatory or an inhibitory state. Needless to say, the prospective array density is reduced to half in addition to the added complexity required from the control and sensory circuitry. Recent publications have demonstrated ASDs with dynamic capabilities (mimicking either LTP or LTD behaviors as a function of a modulation bias) through the use of FET structures (Kim et al., 2013; Tian et al., 2015, 2016, 2017; Yang et al., 2015). As mentioned previously, these devices rely on at least one additional modulation terminal and an associated bias voltage to control the conductance polarity through the field effect. However, they are operated using very large biases (tens of volts) that seriously compromise their integration possibility with modern CMOS architectures. Moreover, the physical dimensions of these devices are very large (tens of micrometers), in addition to having a lateral structure (as opposed to vertical stacking) that is not optimal for high-density 3D integration. State of the art ASDs are thus still far from being able to truly reproduce the behavior of biologic synapses.

Another type of biologic synapses is based on fast conductive links between neurons capable of transmitting and receiving electrical signals (Pereda and Faber, 1996; Kandel et al., 2000). These links contain numerous ion channels (i.e., connexons) scattered along plasma membranes to form the connection between the pre- and post-synaptic neurons (Connors and Long, 2004; Michael et al., 2004). Some of these junctions demonstrate a rectifying behavior while being stimulated by electric pulses.
resulting in a preferred direction for ion flow (Hormuzdi et al., 2004; Landisman and Connors, 2005; Haas et al., 2011). In this light, another subclass of micrometer sized RRAMs was shown to have a unidirectional current rectification property along with non-volatile, multilevel resistive states (Yoon et al., 2015; Kim et al., 2016). In addition, Kim et al. (2018) demonstrated a rectifying micrometer ASD with transient (volatile) current-voltage dependence. Unfortunately, all these devices fall short in terms of both size and high-density integrability that play a key role in the implementation of ideal ANN.

Recently, Berco et al. (2018) presented a proof of concept for a nanoscale current rectifying dynamic RRAM. This ASD, of merely a few nanometers in size, dissipates only several picowatts of power during operation while having the ability to dynamically flip its current rectification direction thus effectively implementing both excitatory and inhibitory synaptic functionalities without the need for a modulation terminal. This operation mode allows for doubling the array size when compared to the common differential approach discussed previously (Bichler et al., 2012; Prezioso et al., 2015). Conductance-weight changes may be implemented by using a digital methodology (as opposed to the ubiquitous analog model). In this manner, a group of nanoscale ASDs are grouped together to represent a single artificial synapse where each member of the group plays the role of a rectifying connexon at the expense of layout resources. Setting the number of current-rectifying devices to a specific direction, being either positive rectification (PR) or negative rectification (NR), in relation to the others (being in the opposite direction), effectively determines the total conductance for each rectification direction. These artificial rectifying connexons (ARCs) may be individually toggled and the overall synaptic weight digitally manipulated in a similar manner to the LTP and LTD in biologic synapses.

**ARTIFICIAL SYNAPTIC DEVICE IMPLEMENTATION**

Memristor arrays for hardware implementations of ANN are extremely efficient in performing matrix vector dot products as weighted-sum operations. A widely used RRAM-based differential array architecture is depicted in Figure 1A (Bichler et al., 2012; Prezioso et al., 2015). In this approach, two passive devices are used to determine a single synaptic weight in a differential manner thus allowing for both positive and negative parametric values. Programming the analog conductance levels may be done either during network training or on the fly to emulate the LTP and LTD synaptic behaviors. The sensing circuitry is based on a differential amplifier driving an activation function module (marked as $f$). However, analog RRAM operation (conductivity adjustment) (Lee et al., 2010; Alibart et al., 2012; Chen et al., 2015; Wang et al., 2016) usually requires complex pulsing schemes to account for its non-linear nature which complicates the design of a digital controller. An ARC-based implementation is given in Figure 1B (Berco et al., 2018). The operation principal allows for dynamically switching of

![Figure 1](https://www.frontiersin.org)
Berco RRAM – Dynamic Complementary Artificial Synapses

FIGURE 2 | (A) Schematic diagram of OV-O distribution in the RSL of an ARC after forming where the gap in the CF (indicated by a blue arrow) yields a rectifying behavior. (B) A digital approach for synaptic weight adjustment demonstrated by Spice simulation (using a generic behavioral model for a group of 10 ARCs). The LTP conductance (black curve) depends on the ratio of PR-ARCs to NR-ARCs with a bias of 0.3 V and the LTD conductance (blue curve) depends on the ratio of NR-ARCs to PR-ARCs with a bias of –1.5 V while the inset shows the circuit under simulation where individual ARC are consecutively flipped from PR to NR and vice versa. (C) An abstract model for synaptic operation as proposed by Berco et al. (2018).

the rectification direction resulting in either an excitatory or inhibitory weight parameter at each junction. In this manner, a single active device can be set to either push or pull current (much like the CMOS couple shown in the inset) effectively doubling the prospective density and simplifying the peripheral circuitry considerably.

A sample response of an ARC-based ANN to a generic input (1,1,0) is demonstrated in Figure 1C. In this example, some ARCs are configured as a PR-ARC (current direction marked by a red arrow) while the others are configured as an NR-ARC (current direction marked by a blue arrow). A high-voltage input “1” will result in current flowing into the output neuron through the PR-ARC. In the same manner, a low-voltage input “0” will result in current flowing out of the output neuron through the NR-ARC. ARCs that are configured in an opposite direction (in relation to the input value) will produce a zero current response. Figure 1D depicts the response of the exact same network configuration to a different input vector (1,0,0). In this manner, both network training and learning may be implemented.

Figure 2A depicts a schematic diagram of a lateral structured RRAM device configured as an ARC for illustrative purposes. The device is comprised of a metal-oxide-based resistive switching layer (RSL) placed between two conductive electrodes. The forming of a conductive filament (CF, an aligned path of current conducting defects) is done using a specific current compliance limit that yields an uneven distribution of oxygen vacancies (OV) and oxygen species (O) (Berco and Tseng, 2016; Berco et al., 2018). The figure depicts positively charged OV being pushed away from the anode and accumulated near the cathode while the negatively charged O ions are drawn to the anode. The forming process is arrested by the current compliance setting before a continuous CF from anode to cathode is able to form. The resulting gap (indicated by a blue arrow) yields a current rectifying behavior (Berco et al., 2018).

The dynamic nature of ARCs may be utilized for real time modulation of the ASD junction plasticity, using a plurality of devices, by changing the ratio of the number of PR to NR-ARCs. A digital approach (contrary to the common analog treatment)

![Figure 2A](image_url)

FIGURE 2 | Implementation of logic gates with ARCs. The directionality of the ARC determines either a push or pull functionality and as a result the logic output value. (A) OR gate. (B) AND gate.
Using this concept, a group of ARCs are treated as a single ASD (consuming more area). This implementation was verified with Spice simulations using a behavioral model based on the experimental data published by Berco et al. (2018). The circuit under simulation is composed of 10 ARCs connected in parallel to a single DC voltage source (Figure 2B inset) representing the input value (either “0” or “1”). The LTP simulation progresses by consecutive flipping an NR-ARC to a PR-ARC starting from \( n = 1 \) to \( n = 10 \) and calculating the overall conductance under positive bias of 0.3 V. The LTD simulation is done by consecutive flipping a PR-ARC to an NR-ARC in the same manner under a negative bias of \(-1.5\) V. The conductance results in Figure 2B show a good linear behavior when depicted as a function of the ratio of PR-ARCs (\( n_{PR} \)) to NR-ARCs (\( 10 - n \)) and vice versa. The positive slope for PR may thus be used to implement an excitatory synaptic weight change (increased current flow to the postsynaptic circuitry for a positive input vector) while the negative slope for NR may be used for an inhibitory synaptic weight (increased current flow from the postsynaptic circuitry for a negative input vector).

**REFERENCES**

Alibart, F., Gao, L., Hoskins, B. D., and Strukov, D. B. (2012). High precision tuning of state for memristive devices by adaptable variation-tolerant algorithm. *Nanotechnology* 23, 75201–75220. doi: 10.1088/0957-4484/23/7/075201

Aratani, K., Ohba, K., Mizuguchi, T., Yasuda, S., Shimoto, T., Tsushima, T., et al. (2007). A novel resistance memory with high scalability and nanosecond switching. *IEDM* 2, 783–786.

Ben-Ari, Y. (2002). Excitatory actions of gaba during development: the nature of the nurture. *Nat. Rev. Neurosci.* 3:728.
Berco, D., and Tseng, T. Y. (2016). A comprehensive study of bipolar operation in resistive switching memory devices. J. Comp. Elect. 15, 577–585. doi: 10.1007/s10825-015-0736-7

Berco, D., Zhou, Y., Golli, S. R., Kalaga, P. S., Kole, A., Hassan, M., et al. (2018). Nanoscale conductive filament with alternating rectification as an artificial synapse building block. ACS Nano 12, 5946–5955. doi: 10.1021/acsnano.8b02193

Richler, O., Suri, M., Querlioz, D., Vuillaume, D., DeSalvo, B., and Gamrat, C. (2012). Visual pattern extraction using energy efficient ‘2-PCM synapse’ neuromorphic architecture. IEEE Trans. Elec. Dev. 59, 2206–2214. doi: 10.1109/TED.2012.2197951

Burr, G. W., Shelby, R. M., Sidler, S., di Nolfo, C., Jang, J., Boybat, I., et al. (2015). Experimental demonstration and tolerancing of a large-scale neural network (165000 synapses) using phase-change memory as the synaptic weight element. IEEE Trans. Elect. Dev. 62, 3498–3507. doi: 10.1109/TED.2015.2439635

Chen, P.-Y., Lin, B., Wang, I.-T., Hou, T.-H., Ye, J., Vrudhula, S., et al. (2015). "Mitigating effects of non-ideal synaptic device characteristics for on-chip learning," in Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD), (Austin, TX: IEEE), 194–199.

Chua, L. O. (1971). Memristor missing circuit element. IEEE Trans. Circuit Theory 18, 507–519. doi: 10.1109/TCT.1971.1083337

Chua, L. O., and Kang, S. M. (1976). Memristive Devices and Systems. Proc. IEEE 64, 209–223. doi: 10.1109/PROC.1976.10092

Connors, B. W., and Long, M. A. (2004). Electrical synapses in the mammalian thalamus. Science 310, 1809–1813. doi: 10.1126/science.1144655

Lee, H. Y., Chen, Y. S., Chen, P. S., Gu, P. Y., Hsu, Y. Y., Wang, S. M., et al. (2010). "Evidence and solution of over-RESET problem for HfOx based resistive memory with sub-nS switching speed and high endurance," in Proceedings of the International Electron Devices Meeting, (San Francisco, CA: IEEE).

Michael, V. L., Bennett, R., and Zukin, R. S. (2004). Electrical coupling and neuronal synchronization in the mammalian brain. Neuron 41, 495–511. doi: 10.1016/S0896-6273(04)00431-1

Nakashashi, S. (1992). Molecular diversity of glutamate receptors and implications for brain function. Science 258, 597–603. doi: 10.1126/science.1329206

Nicholls, J., Martin, A., Wallace, B., and Fuchs, P. (2001). From Neuron to Brain. Sunderland, MA: Sinauer Associates.

Park, S., Chu, M., Kim, J., Noh, J., Jeon, M., Hun Lee, B., et al. (2015). Electronic system with memristive synapses for pattern recognition. Nat. Sci. Rep. 5:10123. doi: 10.1038/srep10123

Perea, A. E., and Faber, D. S. (1996). Activity dependent short-term enhancement of intercellular coupling. J. Neurosci. 16, 983–992. doi: 10.1523/JNEUROSCI.16-03-00983.1996

Prezioso, M., Merrilkh-Bayat, F., Hoskins, B. D., Adam, G. C., Likharev, K. K., and Strukov, D. B. (2015). Training and operation of an integrated neuromorphic network based on metal-oxide memristors. Nature 521, 61–64. doi: 10.1038/nature14441

Root, D. H., Mejias-Aponte, C. A., Zhang, S. W., Hofman, A. F., Lupica, C. R., et al. (2014). Single rodent mesolimbic aconex release glutamate and GABA. Nat. Neurosci. 17, 1543–1551. doi: 10.1038/nn.3823

Strukov, D. B., Snyder, G. S., Stewart, D. R., and Williams, R. S. (2008). The missing memristor found. Nature 453, 80–83. doi: 10.1038/nature06932

Tian, H., Cao, X., Xie, Y., Yan, X., Kostecki, A., DiMarzio, D., et al. (2017). Emulating bilingual synaptic response using a junction-based artificial synaptic device. ACS Nano 11, 7156–7163. doi: 10.1021/acsnano.7b03033

Tian, H., Deng, B., Chin, M. L., Yan, X., Jiang, H., Han, S.-J., et al. (2016). Dynamically reconfigurable ambipolar black phosphorus memory device. ACS Nano 10, 10428–10435. doi: 10.1021/acsnano.6b06293

Tian, H., Mi, W., Wang, X.-F., Zhao, H., Xie, Q.-Y., Li, C., et al. (2015). Graphene dynamic synapse with modulatable plasticity. Nano Lett. 15, 8013–8019. doi: 10.1021/acs.nanolett.5b03283

Wang, I.-T., Chang, C.-C., Chiu, L.-W., Chou, T., and Hou, T.-H. (2016). 3D Ta/TaOx/TiO2/Ti synaptic array and linearity tuning of weight update for hardware neural network applications. Nanotechnology 27:365204. doi: 10.1088/0957-4484/27/36/365204

Wang, H.-S. P., Lee, H. Y., Yu, S., Chen, Y. S., Wu, Y., Chen, P. S., et al. (2012). Metal-oxide RRAM. Proc. IEEE 100, 1951–1970. doi: 10.1109/JPROC.2012.2190369

Yang, Y., Chen, B., and Lu, W. D. (2015). Memristive physically evolving networks enabling the emulation of heterosynaptic plasticity. Adv. Mater. 27, 7720–7727. doi: 10.1002/adma.201503202

Yao, P., Wu, H., Gao, B., Eryilmaz, S. B., Huang, X., Zhang, W., et al. (2017). Face classification using electronic synapses. Nat. Commun. 8:15199. doi: 10.1038/ncomms15199

Yoon, J.-H., Kim, K. M., Song, J. Y., Seok, J. Y., Yoon, K. J., Kwon, D. E., et al. (2015). Pt/TaOx/HfO2-x/Ti resistive switching memory competing with multilevel NAND flash. Adv. Mater. 27, 3811–3816. doi: 10.1002/adma.201501167

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Berco. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.