A Memristor-Crossbar/CMOS Integrated Network for Pattern Classification and Recognition

L. Zhang

Institute of Microelectronics
Tsinghua University
Beijing 100084, China

Z.J. Chang

Medical School
Tsinghua University
Beijing 100084, China

Abstract-**A novel circuit model based on a trainable memristor-crossbar network integrated with a CMOS circuit for pattern classification and recognition is proposed and analyzed in this paper. The configurable memristors along each column wires of the crossbar are trained by a standard pattern input from the row wires of the crossbar to represent the pattern. After the training, the crossbar network can classify unknown patterns input from the row wires, and the output current from each column wire will be normalized by the CMOS circuits to denote the probability to classify the unknown patterns with respect to the standard pattern associated with the column wire. The probabilities can be further processed by a winner-take-all competition circuit for decision making. The circuit simulation results demonstrate that the proposed circuit based on our experimentally demonstrated memristor devices can classify patterns by calculating the probabilities and recognize patterns with distortions. Moreover, the circuit delay for classifying a pattern remains below 1 μs even when the pattern scales up to large dimensions. The large-scale parallel signal processing by the memristor-crossbar/CMOS circuit enables it to classify and recognize patterns with high dimensionality and complexity at a much faster speed than the software-based computers.**

Keywords-pattern classification; recognition; memristor; crossbar; CMOS analog integrated circuits; probability

I. INTRODUCTION

It is challenging for machines to recognize large dimensional complex patterns in speech, handwritten characters, medical images, bioinformatics, and stocks, etc. Although human brains may classify and recognize those patterns in a seemingly effortless fashion, it is immensely difficult for a computer to achieve the same tasks. The software-based computing time for the pattern classification increases exponentially with the pattern dimensions (so-called the curse of dimensionality), which makes the classification of the large dimensional pattern extremely time-consuming and even prohibitive for a modern computer and hardware [1]. The complementary metal-oxide-semiconductor (CMOS) based analog integrated circuits have also been designed and fabricated for pattern classification purpose [2-4], but the circuits needed to recognize large dimensional patterns are complex, expensive, and energy-consuming.

In this paper, we propose an electronic circuit for large dimensional patterns classification and recognition based on the experimentally demonstrated memristors [5], configurable resistors [6], and crossbar networks [7]. The conductance of the memristors at the cross point of the crossbar circuit can be modified to an arbitrary analog value proportional to the amplitude of the input signal from the pattern during the training process, which is similar to the memory function of a synapse in a neural network. The crossbar network of the memristor is a simple network architecture that can be easily expanded to large-scale to process the signals from a pattern with large dimensional inputs in parallel efficiently, and the device density can potentially reach $> 10^{11}/\text{in}^2$ in the crossbar circuits fabricated by nanoscale lithography [7]. The memristor-crossbar array is integrated with a Si-based CMOS circuit, and the output of the circuit gives the probability distribution to classify the input pattern with respect to the memorized patterns. The electronic pattern classification is a high-speed parallel process and can recognize patterns efficiently with high dimensionality and complexity. The low-cost fabrication process for the high-density nanoscale circuits and modern CMOS technology also make the proposed circuit model practically viable and cost-efficient.

II. CIRCUIT ARCHITECTURE

The pattern recognition circuit aims to classify patterns based on the knowledge of the patterns established statistically from supervised or unsupervised training processes, and the circuit classifies the pattern based on the probability belonging to each class. An arbitrary pattern can be abstracted as a vector $\vec{y} = (y_1, y_2, \cdots, y_N)$, where y_i represents statistically independent variables of a pattern at ith dimension with $i = 1, 2, \cdots, N$. The pattern can also be represented by a normalized probability $p_i = y_i / \sum_{i=1}^{N} y_i$ at the ith dimension. Assuming that the patterns can be classified into M different classes denoted by j with $j = 1, 2, \cdots, M$, based on the law of total probability we can have,

$$p_i \sum_j p_i^j p_j \tag{1}$$

where p_i^j is the probability for the pattern with the probability function p_i to be classified to class j, sometimes referred as the likelihood function. Eqn. (1) can be modified to an alternative format by multiplying p_i^k on both sides and take the summation with respect to i,

$$\sum_i p_i p_i^k = \sum_i \sum_j p_i^j p^j p_i^k = \sum_j \sum_i p_i^j p_i^k p^j \tag{2}$$

The goal of the circuit is to classify an input pattern by obtaining p^j based on eqn. (2). The latency for solving the equation increases exponentially with the dimensions of the patterns in the software-based methodologies, therefore it is necessary to design an electronic circuit that can solve the problem efficiently.

The proposed pattern classification circuit is shown in fig. 1. The circuit can be divided into two parts: (1) signal inference circuit composed of a crossbar memristor network and CMOS operational amplifiers (opamp), and (2) a winner-take-all (WTA) decision-making competition circuit.

In the inference stage, an input pattern can be represented by a vector $\overrightarrow{V^I} = \left(V_1^I, V_2^I, \cdots, V_N^I\right)$, where V_i^I denotes the voltage input to the ith row wire in the memristor-crossbar circuit (shown in fig. 1) with $i = 1, 2, \cdots, N$. V_i^I is applied on a memristor with a conductance of g^I on each row wire. A memristor with a configurable conductance g_{ij} at each cross point of the crossbar connects the ith row wire and the jth column wire. The voltage on the jth column wires is denoted by V_j^P with $j = 1, 2, \cdots, M$. V_j^P is applied on a resistor with a conductance of g_j^O on each column wire and also to the negative input terminal of the opamp A_{1j}. The structure of opamp is illustrated in fig. 2(a). It uses the complementary input differential pairs composed of M_0-M_4 and M_7-M_{11} to provide the rail-to-rail input dynamic range, and employs the Class-AB output stage realized by M_5, M_6, M_{12}, and M_{13} to achieve the maximized output current swing. M_6 and M_{13} also provide a duplicated output to the WTA competition circuit, and transistors M_{14}-M_{16} are serving for the biasing purposes. V_j^O denotes the output voltage from the opamp A_{1j} connected to the jth column wire of the crossbar with $j = 1, 2, \cdots, M$. V_j^O and its duplication are output from the "o" and "oc" terminals of opamp A_{1j}, respectively.
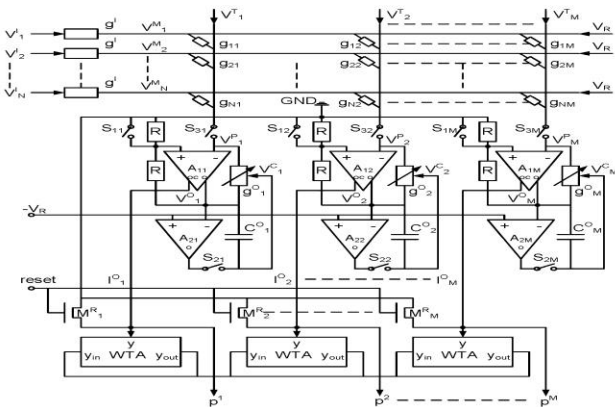

FIGURE I. STRUCTURE OF THE PROPOSED PATTERN CLASSIFICATION CIRCUIT.

Following the Kirchhoff's Current Law at both the row and column wires of the crossbar circuit shown in fig. 1,

$$\left(V_i^I - V_i^M\right)g^I - \sum_j \left(V_i^M - V_i^P\right)g_{ij} = 0 \quad (3)$$

$$\sum_i \left(V_i^M - V_k^P\right)g_{ik} - \left(V_k^P - V_k^O\right)g_k^O = 0 \quad (4)$$

where V_i^M is the voltage on the ith row wire. Assuming the condition $g_i \gg \sum_j g_{ij}$ is held in the circuit design, it can be obtained from eqn. (3) that $V_i^M \approx V_i^I + \sum_j V_j^P g_{ij} / g^I$, which can substitute V_i^M in eqn. (4),

$$\sum_i V_i^I g_{ik} + \frac{\sum_j V_j^P \left(\sum_i g_{ij} g_{ik}\right)}{g^I} - V_k^P \sum_i g_{ik} - \left(V_k^P - V_k^O\right)g_k^O = 0 \quad (5)$$
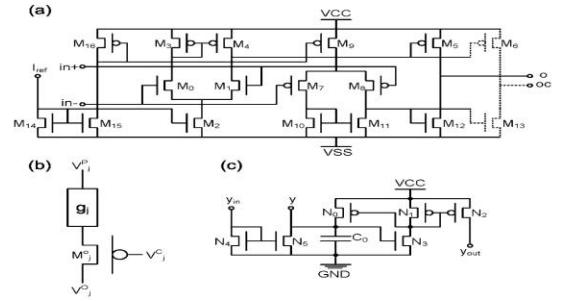

FIGURE II. (A) THE SCHEMATIC OF OPAMP A1J AND A2J SHOWN IN FIG. 1. THE ADDITIONAL DEVICES IN A1J ARE SHOWN IN DASHED LINES. (B) CIRCUIT REALIZATION OF THE CONFIGURABLE RESISTOR g_j^O. (C) WINNER-TAKE-ALL CIRCUIT UNIT.

When both $g_k^O = \sum_i g_{ik}$ and $V_k^O = 2V_k^P$ are enforced for all columns with $k = 1, 2, \cdots, M$ in the circuit, the underlined terms in eqn. (5) can be enforced to zero for all column wires, and eqn. (5) can be modified as

$$\sum_i V_i^I \frac{g_{ik}}{g_k^O} + \sum_j \left(\frac{g_j^O}{2g^I} V_j^O\right) \sum_i \frac{g_{ij} g_{ik}}{g_j^O g_k^O} = 0 \quad (6)$$

The conductance g_{ij} can be modified to be proportional to p_i^j in eqn. (2) by the training process as described in the following, then we have $p_i^j = g_{ij} / \sum_i g_{ij} = g_{ij} / g_j^O$. The input voltage V_i^I is also proportional to p_i in eqn. (2), comparing eqn. (2) and (6), p^j, the probability for an input pattern to be classified to class j, can be expressed as,

$$p^j = -\frac{V_j^O g_j^O}{2\sum_i V_i^y g^I} = \frac{I_j^O}{\sum_j I_j^O} \quad (7)$$

where I_j^O, the current on conductor g_j^O, determines the distribution of p^j.

Due to the virtual short property of opamp, the condition $V_k^O = 2V_k^P$ is ensured by doubling V_j^P by the resistor ladder

R at port "o" of A_{1j}. The condition $g_k^O = \sum_i g_{ik}$ is enforced by the transistors M_j^O, resistors g_j, capacitors C_j^O, as shown in fig. 1 and fig. 2(b), and the incorporated opamps A_{1j} and A_{2j} with their structures shown in fig. 2(a). When the circuit is operated in the normalization mode, all the switches S_{1j}, S_{2j}, and S_{3j} are closed, an all "1" input voltage vector $\overrightarrow{V^I} = (V_R, V_R, \cdots, V_R)$ is imposed on all the row wires with V_R representing the reference voltage, while the output voltage V_j^O is enforced to $-V_R$ on all the column wires by the feedback mechanism introduced by the second stage of opamp A_{2j}. At the equilibrium, g_j^O the series resistance of g_j and the transistors M_j^O, is self-adjusted to the value of $\sum_i g_{ij}$ in all column wires, and the gate-source voltage of M_j^O are memorized in capacitor C_j^O. When the circuit is operated in the recognition mode, the switches S_{1j} and S_{2j} are open, while S_{3j} still remains closed, the voltage memorized by the capacitor C_j^O biases M_j^O in the triode region, and the condition $g_j^O = \sum_i g_{ij}$ is thus sustained. Due to the current leakage through S_{2j}, the memorized voltages on C_j^O have to be refreshed periodically.

The comparison for the probability p^j is realized by the WTA competition among the input currents I_j^O from different columns. In the WTA circuit shown in fig. 2(c), each unit associated to one of the column wires adopts the duplication of I_j^O from terminal "oc" of the opamp A_{1j} to the port "y" and charges the capacitor C_0. When the voltage exceeds the threshold of the transistor N_3, N_1 is turned on and feeds a current back to charge C_0 further and increases the output voltage at "y", meanwhile provides the leakage current at "yout" port to all the other units to decrease their competencies. The current mirror composed of N_4 and N_5 duplicates the leaking currents from all other units associated to other column wires input from "yin" to discharge C_0. As shown in fig. 1, all the WTA units are connected through the "yin" and "yout" ports to a global inhibitory node, and the transistors $M_1^R, M_2^R, \cdots, M_M^R$ in fig. 1 serve to initiate and reset the competition.

III. SIMULATION AND DISCUSSIONS

The proposed pattern recognition circuit has been designed and simulated using a 0.18-μm CMOS process. The conductance g^I has to be carefully selected to satisfy the condition $g^I \gg \sum_j g_{ij}$, but meanwhile confined by the maximal output current of the opamp A_{1j}. In this circuit, g^I is set to 50 kΩ.

During the training process, all the switches S_{1j}, S_{2j}, and S_{3j} are open, by applying the input voltage vector $\overrightarrow{V^I} = (V_1^I, V_2^I, \cdots, V_N^I)$ representing a standard pattern to the row wires, and setting the training voltage, V_k^T, on a corresponding column wire to a value exceeding the training threshold voltage, the configurable memristors at the cross points along the column can be trained to the conductance, g_{ij}, proportional to the component V_i^I of the input voltage vector $\overrightarrow{V^I}$. The voltages on the other column wires, $V_j^T (j \neq k)$, are set to a value below the training threshold voltage, therefore the configurable memristors at the cross points along the other column wires are not configured. More advanced feedback training mechanism can be included to tolerate and correct the defects in g_{ij} values to achieve the expected classification probability p^j.

The condition $V_k^O = 2V_k^P$ is satisfied by the resistor ladder with two 1 MΩ resistors in series. The opamp draws a biasing current of about 40 μA, and the aspect ratios of nMOS and pMOS in the opamps and WTA units are chosen as $W_n/L = 12/0.6$ and $W_p/L = 24/0.6$, respectively. Since the application of global inhibitory node, all transistors in the WTA units are designed with the standard sizes ($W_p = 24$ μm, $W_n = 12$ μm, and $L = 0.6$ μm), which in turn saves the physical areas. In order to remove the body effects and ensure the resistances of M_j^O being independent of their source voltages, V_{BS} of M_j^O is enforced to be zero by utilizing pMOS with shorted body and source terminals.

A. Pattern Classification

In the simulation, four characters, "A", "C", "L", and "U", as are shown in fig. 3(a), are represented by sixteen pixel patterns with their corresponding input voltage vectors $\overrightarrow{V_A^I} = (0,1,1,0,1,0,0,1,1,1,1,1,1,0,0,1)$, $\overrightarrow{V_C^I} = (1,1,1,1,1,0,0,0,1,0,0,0,1,1,1,1)$, $\overrightarrow{V_L^I} = (1,0,0,0,1,0,0,0,1,0,0,0,1,1,1,1)$, $\overrightarrow{V_U^I} = (1,0,0,1,1,0,0,1,1,0,0,1,1,1,1,1)$, where the black and white pixels in the patterns are denoted by "1" ($V_i^I = 50$ mV) and "0" ($V_i^I = 0$ V), respectively. Four columns in the crossbar are set to represent the characters, "A", "C", "L", and "U", respectively, and the conductance of the configurable memristors in the crossbar are set to 10^{-6} S or 0 S, corresponding to the 1 or 0 components of their corresponding vector V_i^I, respectively.

To evaluate the circuits, the four different input voltage vectors V_A^I, V_C^I, V_L^I, and V_U^I were applied on the row wires of the crossbar. The output current I_j^O was simulated as the function of time. For example, when the vector V_U^I is input to

the circuit, the current I_j^O is shown as the function of time in fig. 3(b). After the initial increase for all I_j^O for a period of 567 ns, only I_U^O increases, while I_A^O, I_C^O, I_L^O, and decrease gradually. After initiating the WTA competition, within 160 ns the output probability p^j from the column with the largest I_j^O (I_U^O in this case) is set to 1, and the p^j from the rest of the columns are set to 0, the final output from the circuit can be expressed as a vector $\overrightarrow{p^*} = \left(\overrightarrow{p^A}, \overrightarrow{p^C}, \overrightarrow{p^L}, \overrightarrow{p^U} \right) = (0,0,0,1)$. When the different voltage vectors V_A^I, V_C^I, V_L^I, and V_U^I are input, the output probability vectors obtained from the circuit are $\overrightarrow{p^*} = (1,0,0,0)$, $\overrightarrow{p^*} = (0,1,0,0)$, $\overrightarrow{p^*} = (0,0,1,0)$, and $\overrightarrow{p^*} = (0,0,0,1)$, respectively, which demonstrates the primary classification function of the circuit.

B. Distorted Patterns

The proposed circuit is further assessed in terms of the responses to distorted patterns. Fig. 4(a) shows the distorted patterns when they transform from standard "A" to "C", and the hamming distance, which is defined as the count of the different components of the input voltage vectors of the distorted and original patterns, increases from 0 to 8 (the intrinsic hamming distance between standard "A" and "C" is 8.).

Fig. 4(b) characterizes the output of the circuit to the different input patterns transformed from "A" to "C" in terms of the output classification probability p^j (before WTA competition) as the function of the hamming distances. It can be seen that when the pattern transforms from "A" to "C", p^A, the probability to classify the pattern as "A", decreases from 1 to 0, meanwhile p^C, the probability to classify the pattern as "C", increases from 0 to 1. The patterns with approximately equal hamming distance from "A" and "C" may confuse the competition circuit and lead to erroneous decisions.
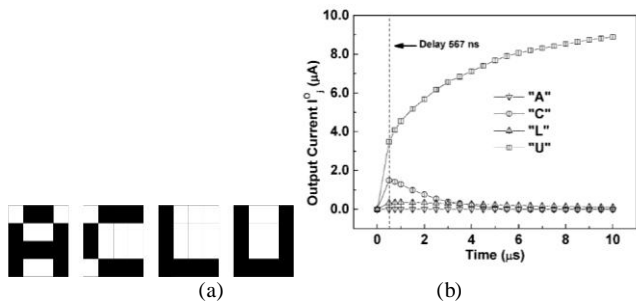
FIGURE III. (A) STANDARD PATTERNS "A", "C", "L", AND "U". (B) TRANSIENT RESPONSES OF THE OUTPUT CURRENT I_j^O ON THE INFERENCE STAGE WHEN THE STANDARD PATTERN "U" IS INPUT TO THE CIRCUIT.
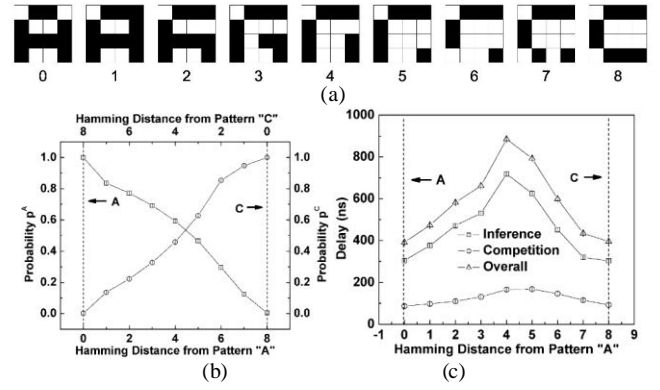
FIGURE IV. (A) INPUT PATTERN TRANSFORMS FROM "A" TO "C". THE DIGIT UNDERNEATH REPRESENTS THE HAMMING DISTANCE FROM THE STANDARD PATTERN "A". (B) CLASSIFICATION PROBABILITIES p^A AND p^C SHOWN AS THE FUNCTIONS OF THE HAMMING DISTANCE FROM THE STANDARD PATTERN "A". (C) CIRCUIT DELAY SHOWN AS THE FUNCTION OF THE HAMMING DISTANCE FROM THE STANDARD PATTERN "A".

C. Circuit Delay

High speed is one of the major advantages of the proposed pattern recognition circuits. The circuit delays, defined as the time for the circuit outputs being bifurcated toward different directions in the inference stage (as shown in fig. 3(b)), and the time for the outputs being settled to the stable values in the competition stage, are analyzed.

In the proposed circuit, the delay not only depends on parameters of the device elements, but also influenced by the input patterns, as shown in fig. 4(c). When pattern transforms from "A" to "C", both the delays on the inference and competition stages change as the function of the hamming distance of the patterns distorted from the standard patterns, and are maximized at the position with approximately equal hamming distance from "A" and "C". The delay on the inference stage takes 70%~80% of the overall delay.

D. Dimension Scaling

When the dimension of the input pattern, N, scales up, the equivalent circuit for characterizing the delay has been illustrated in fig. 5(a), where $\overline{g^I}$ is the average input conductance, $\overline{g}$ and $\overline{C}$ are the average conductance and capacitance at cross points of the crossbar, and R_L and C_L are the loading resistance and capacitance of opamp, respectively. The opamp can be modelled as a single pole system with the trans-conductance function of $g_m(s) = g_m^0 \omega_p / (s + \omega_p)$, where g_m^0 is the DC trans-conductance, and ω_p is the dominant pole of the opamp, and s is the Laplace operator. By applying the Kirchhoff's Law at the points A and B shown in fig. 6(a),

$$\frac{\left(\overline{g} + s\overline{C} \right) \overline{g^I} \left(V - v_n \right)}{\overline{g} \left(\overline{g} + s\overline{C} + \overline{g^I} \right)} = v_n - 2v_p \tag{8}$$

$$\frac{\left(v_p - v_n\right)g_m^0\omega_p}{s+\omega_p} + \frac{\left(\overline{g}+s\overline{C}\right)N\overline{g^I}\left(V-v_n\right)}{\overline{g}+s\overline{C}+\overline{g^I}} = v_p\left(2\left(\frac{1}{R_L}+sC_L\right)+\frac{1}{R}\right) \quad (9)$$

Based on our circuit design, the conditions $g_m^0\left(R_L \parallel R\right) \gg 1$, $\overline{g}R_L \sim 1$, and $N \gg 1$ are satisfied, we may have the simplified characterization function of

$$2N\left(\overline{g}+s\overline{C}\right)\left(s+\omega_p\right) - g_m^0\left(R_L \parallel R\right)\omega_p = 0 \;, \text{ and } s \text{ is solved as}$$

$$s = \sqrt{-\frac{\overline{g}\omega_p}{\overline{C}}+\left(\frac{\overline{g}}{2\overline{C}}+\frac{\omega_p}{2}-\frac{g_m^0\left(R_L \parallel R\right)\omega_p}{4N}\right)^2} - \frac{\overline{g}}{2\overline{C}}-\frac{\omega_p}{2}+\frac{g_m^0\left(R_L \parallel R\right)\omega_p}{4N} \quad (10)$$

$$\approx -\omega_p - \frac{g_m^0\left(R_L \parallel R\right)\overline{C}\omega_p^2}{4N\overline{g}}$$

The circuit 3-dB bandwidth approaches ω_p when the input vector dimension N scales up, therefore, the circuit delay initially increases with N and saturates at the intrinsic delay of the opamp. Fig. 5(b) shows the circuit speed characteristics based on the simulation on the inference stage when the input vector dimension scales up. It can be seen that the circuit delay drastically increases with N over the low dimensional range ($N < 8$), however, saturates approximately at 800 ns on the large dimensional range ($N > 16$). The results indicate that the proposed circuit is potentially capable to classify patterns with large dimensions at a much faster speed than the software-based computer.
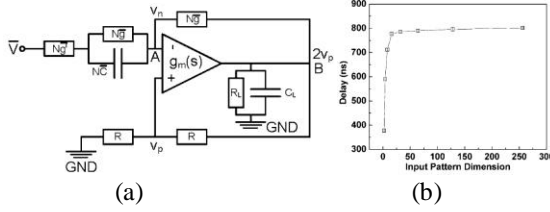

(a)　　　　　　　(b)

FIGURE V. (A) EQUIVALENT CIRCUIT FOR DELAY ANALYSIS. (B) CIRCUIT DELAY SHOWN AS THE FUNCTION OF THE DIMENSION OF INPUT PATTERNS. THE SQUARES ARE THE SIMULATED RESULTS ON INPUT DIMENSIONS OF 2, 4, 8, 16, 32, 64, 128, AND 256, RESPECTIVELY.

E. Energy Consumption

The energy consumption for the circuit is calculated, it takes ~1 pJ per dimension to recognize a pattern. The maximal power density at each cross point in the crossbar is ~1 μW/μm^2, which falls in the range of the power density in a standard Si CMOS circuit. The typical operation power for each opamp in the circuit is 40 μW. It is worthwhile to notice that the delay of the proposed pattern classification circuit is associated with the 3-dB bandwidth of the opamp. One can further miniaturize the power consumption by incorporating opamps with smaller bandwidth, however, at the cost of an increased delay and therefore lower speed. Careful trade-off has to be made between power and speed for specific applications.

IV. CONCLUSION

A novel circuit model for pattern classification has been proposed based on statistic theorem. The circuit architecture is composed of a memristor-crossbar/CMOS hybrid network, with each column of the crossbar representing a pattern class.

When a pattern is input from the row wires of the crossbar, the output current from each column is normalized by a CMOS analog circuit, and the probability to classify the input pattern to a class is denoted in terms of the normalized output current from each column. By comparing the output currents, the input pattern can be finally classified to a class with the largest probability in a winner-take-all competition circuit. The delay to classify a pattern is at the order of sub-μs and does not increase when the pattern dimension scales up, which endows the circuit with the capability to recognize patterns with large dimensions at a speed much faster than the software-based computers.

REFERENCES

[1] Sardar, S. & Babu, K.A., Hardware Implementation of Real-Time, High Performance, RCE-NN Based Face Recognition System. *Int. Conf. VLSI Design and Int. Conf. Embedded Systems*, pp.174-179, 2014.

[2] Payvand, M., Rofeh, J., Sodhi, A. & Theogarajan, L., A CMOS-memristive self-learning neural network for pattern classification applications. *IEEE/ACM Int. Symp. Nanoscale Architectures*, pp. 92-97, 2014.

[3] Yoo, J., Yan, L., El-Damak, D., Bin Altaf, M.A., Shoeb, A.H. & Chandrakasan, A.P., An 8-Channel Scalable EEG Acquisition SoC With Patient-Specific Seizure Classification and Recording Processor. *IEEE J. Solid-State Circuits*, 48(1), pp. 214-228, 2013.

[4] Goknar, I.C., Yildiz, M., Minaei, S. & Deniz, E., Neural CMOS-Integrated Circuit and Its Application to Data Classification. *IEEE Trans. Neural Networks Learning Systems*, 23(5), pp. 717-724, 2012.

[5] Strukov, D.B., Snider, G.S., Stewart, D.R. & Williams, R.S., The missing memristor found. *Nature*, 453(7191), pp. 80-83, 2008.

[6] Lai, Q.X., Zhu, Z.H., Chen, Y., Patil, S. & Wudl, F., Organic nonvolatile memory by dopant-configurable polymer. *Applied Physics Lett.*, 88(3), 2006.

[7] Jung G.Y., Ganapathiappan, S., Ohlberg, D.A.A., Olynick, D.L., Chen, Y., Tong, W.M. & Williams, R.S., Fabrication of a 34 x 34 crossbar structure at 50 nm half-pitch by UV-based nanoimprint lithography. *Nano Lett.*, 4(7), pp. 1225-1229, 2004.