International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || Volume 5 Issue 5 *|| May 2016 || PP.1-7*

New binary memristor crossbar architecture based neural networks for speech recognition

Van-Tien Nguyen[1], Minh-Huan Vo[2]

[1,2]*Department of Electrical Electronic Engineering, HCMC University of Technology and Education, Viet Nam.*

ABSTRACT : *In this paper, we propose a new binary memristor crossbar architecture based neural networks for speech recognition. The circuit can recognize five vowels. The proposed crossbar is tested by 1,000 speech samples and recognized 94% of the tested samples. We use Monte Carlo simulation to estimate recognitition rate. The percentage variation in memristance is increased from 0% to 15%, the recognition rate is degraded from 94% to 82%.*

KEYWORDS – *Memristors, Crossbar, Speech recognition, Binary memristors.*

I. INTRODUCTION

According to Moore's Law, transistor count per chip will double within two years. The IC technology development in recent years has shown the validity of Moore's law. However, according to estimates in the near future, the technology will reach the limit of Moore's Law, which means that the chip size will reach critical values to ensure accuracy and stability. So, there are many new methods being studied to replace Moore's law. And memristor of Leon. Chua was mentioned in 1971 [1], which was completed by Stanley- Williams in 2008 [2]. The future of technology has opened up new development. This technology is even better than CMOS technology has been thriving.

Anyone with a basic knowledge in electrical engineering knows that there are four fundamental circuit variables: Current i, Voltage v, Charge q, and Flux f. Then it is clear that with these four parameters, there can be six possible combinations for relating them to each other. So far we have complete understanding and control over five of these combinations in which three of them are passive twoterminal fundamental circuit elements, namely the resistor R, the capacitor C and the inductor L. Unlike the active components which can generate energy, these three components are passive elements which are only capable of storing or dissipating energy but not generating it. The relationship between 'voltage and current', 'voltage and charge', and 'current and flux' are defined by a resistor, capacitor and an inductor, respectively. No device was there to relate the charge and the magnetic flux until Leon Chua introduced his new circuit element called "memristor". In 2008, a research group at HP Labs lead by Stanley Williams succeeded to fabricate the device in nanometer scale. Since then, the research being conducted on memristors gained momentum and the number of publications have boosted quite rapidly. Memristor have two types, are analog memristors and binary memristors. The analog memristor can change the value memristance depend on voltage or electric current applied to it. However, installing memristance value is difficult, not exactly. On the contrary, memristance of binary memristor is easy to install, and more exact. Binary memristors have two state either a high resistance state (HRS) or a low resistance state (LRS), so they can be stored only'1' or '0' in binary memristors.

Recent research focuses on using crossbar architecture to simulate synaptic systems. Thus, an application uses memristor for speech recognition [3]. Our research focuses on recognizing five vowels: 'a', 'e', 'i', 'o' and 'u' from the human voice. To do this, First, a voice signal will be extracted features by MFCC [4]. There are 48 feature values. Then, they are trained by neutral network to recognize 5 vowels. After training, weightings are quantized in 4 bits, their values were stored in binary memristor crossbar circuit. The memristor can achieve either a high-resistance state (HRS) or a low-resistance state (LRS). It means that memristor can store '1' and '0' with two states. This memristor plays a role as a 2-terminal switch to change the resistance between high resistance state (HRS, logic "0") and low resistance state (LRS, logic "1"). By doing so, we can recognize each vowel by multiplication of input signal and weight stored in binary memrisor. The summation of the multiplication results decides the biggest output among 5 outputs that will represent input signal. We suggest a new binary memristor crossbar circuit based neural network model for recognizing five vowels. In addition, statistical simulations are performed, and the simulation results are discussed and finally summarized in this paper.

II. METHODS

The recognition model consists of two main processes: weight installation and recognition process. First, in weight installation, voice input is processed and trained in neural network model. These weights will be quantized in 4 bits. The obtained bits will be stored into the binary memristor. Second, in recognition process, voice input will be processed, and then applied to weighted memristor circuit to determine output. In the first process, the voice signal is extracted features by MFCC method, including preprocess, framing, windowing, DFT, mel frequency log. After that, they are trained in neural network by Matlab. Neural network model has one neural in hidden layer, transfer function is logsig [5]. The training process have 5 times for five vowels. Output is '1' for vowel which trained, else output is '0'. The process of recognition will be performed in each sound. In the recognition process, the input will be quantized in 4 bits with 16 levels from 0 to 15. The input before training was normalized to training input value in the range -15 to 15. After the training process for each vowel, we will have 48 weights correspondingly. We have $[x]_a$ is 'a' voice input, $[w]_a$ is 'a' weight. $[w]_a$ will be quantized to 4 bits. However, the $[w]_a$ contain both positive values and negative values $[w]_a = [w]_a^+ + [w]_a^-$. Therefore, before the quantization, we should process negative values $[w]_a = [-1\ 1]_{w_a} * [w]_a^+$. An array $[-1\ 1]_{w_a}$ was created to process negative values. Thus the value output depends on $O_a = [x]_a * [w]_a$ Or $O_a = [x]_a * [-1\ 1]_{w_a} * [w]_a^+ = [x]'_a * [w]_a^+. Here, [x]'_a$ is the new input values after multiplying with array $[-1\ 1]_{w_a}$. The training process will achieve the significant positive values of $[w]_a^+$, $[w]_e^+$, $[w]_i^+$, $[w]_o^+$ and $[w]_u^+$. The weights will be adjusted proportionally to the corresponding coefficient so that its value in the range (0; 15), then quantized to 4 bits. The bit value "1", "0" will be stored in two memristor with memristance values of R_{on} and R_{off}.

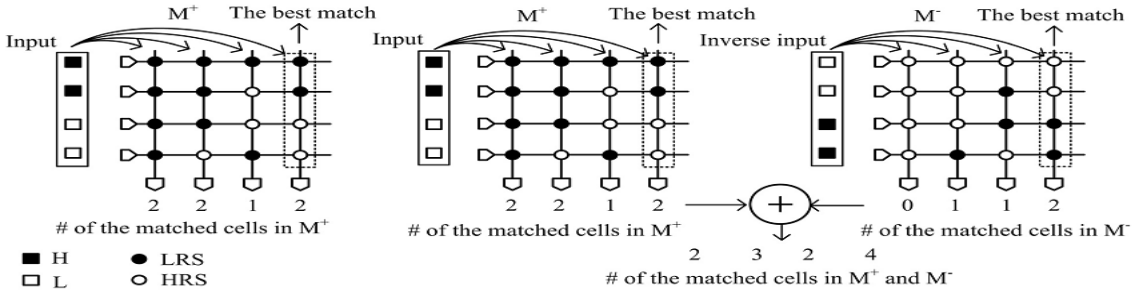

Fig 1: Proposed crossbar architecture for recognition based the comparison signal.

Previous researches proposed crossbar architecture for speed recognition based the comparison signal [6]. This is not true if the input signal is unstable or voice samples come from various people [6]. If test sample is best matched with trained samples, output is the biggest. Figure 1a shows that data input is '1100', we have 4 columns, that are the weights, first column is '1111', second column is '0111', third column is '1001' and fourth is '1100'. In Figure 1b, the best matched column with the input vector of '1100' is the fourth column. The number of matched cells is as large as 4 for the fourth column. By adding cells in M^+ and M^- arrays, we can find the best matched cells. Hence, we determine the best matched column with the input vector among four columns.

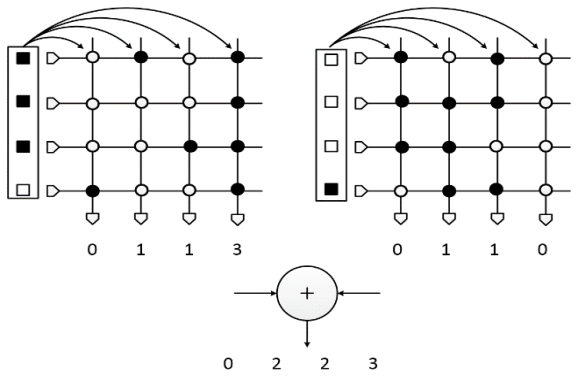

Fig 2: Previous crossbar architecture causes error.

However, if the test voice is not completely matched with trained samples then it is not matched with trained samples, certantly. In figure 2, data input is '0111' as 7, having 4 columns, first is '1000' as 8, second '0001' as 1, third is '0100' as 4 and last column is '1111' as 15. Like this, data input is 7 nearly same with first column, is 8. But we apply the memristor crossbar architecture, results is 0 as a bad result. The output of first column is 0, it is the smallest among 4 columns. So, this architecture is reason that causes low recognition rate. In addtion, we cann't recognize a lot of samples of many human with this architecture because each person has private speech. Therefore, to raise recognition rate with various human speech, we propose a new memristor crossbar architecture, it is based neural network model. The output of neural network is caculated $net = \sum_{i=0}^{48} x_i * w_i$. Each input is multiplied by each weight, then sum of multiplication results is output. Figure 3 shows proposed memristor crossbar architecture.

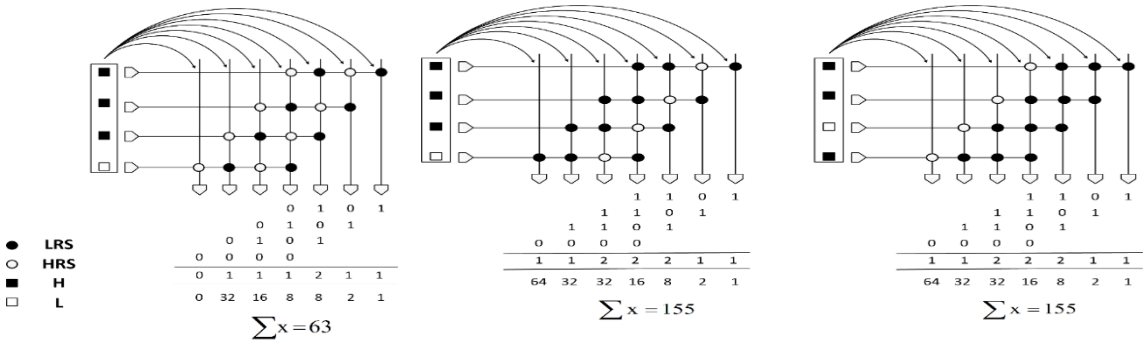

Fig 3: New crossbar architecture based neural network.

To multiply input and the weight, memristors are arranged as figure 3. In figure 3a and 3b, data input is '0111'. The weights of rows are '0101' as 5 and '1101' as 11. This works similarly to multiplication of two 4-bit numbers. From that, we have 7 column and 7 factors of 1, 2, 4, 8, 16, 32 and 64. The result in figure 3a is 7∗5 = 63 and in figure 3b is 7∗11=155. The results show that if b < c then a∗b < a∗c as multiplying two integer numbers. In figure 3c, data input is '1101' as 11 and the weight is '0111' as 7. The result is 11∗7 =155 like as the result of figure 3b. The results show that a∗b = b∗a. This is interchange between two integer numbers. The results show that the new memristor crossbar architecture simulates neural network model very accuracy. So, for recognizing five vowels, we make 4-bit 48 channel inputs corresponding to 48 features of voice input. Each channel is included 4-bit binary values. In Figure 4, $V_{a,1}$ is the voltage of the 'x1' column for recognizing'a'. $V_{a,2}$ is the voltage of the 'x2' column for recognizing'a'. Similarly, $V_{a,4}$, $V_{a,8}$, $V_{a,16}$, $V_{a,32}$ and $V_{a,64}$ are the voltages of the 'x4' , 'x8', 'x16', 'x32' and 'x64' columns in the 'a' crossbar array. Here, 'x1' is the weight of this column and voltage of this colume is as much as 1.

In Figure 4, 'x2', 'x4', 'x8', 'x16', 'x32' and 'x64' mean that the weight factors are 2, 4, 8, 16, 32 and 64 respectively, for the corresponding to columns in the 'a' crossbar array. Here, VC_a can be calculated with the weighted summation of $64V_{a,64} + 32V_{a,32} + 16V_{a,16} + 8V_{a,8} + 4V_{a,4} + 2V_{a,2} + V_{a,1}$. Similarly, VC_e is the weighted summation for recognizing 'e'. The value of V_e is caculated by the weighted summation of $64V_{e,64} + 32V_{e,32} + 16V_{e,16} + 8V_{e,8} + 4V_{e,4} + 2V_{e,2} + V_{e,1}$. The voltages of V_a, V_e V_i, V_o and V_u are inputs in the winner-take-all circuit. They are compared each other to determine which vowel is the biggest in vowels, that is the voice input. Figure 4 shows that: $Output_a$, $Output_e$, $Output_i$, $Output_o$, and $Output_u$ are the outputs of the winner-take-all circuit. We can measure the voltage level to recognize the voice.

Figure 5a shows the schematic of the binary memristor crossbar circuit. The circuit has 48 input channels, 4-bit binary values in each channel and each 4-bit binary weight is stored into each row. The 4-bit weight is set into 4 memristors. Each row has 4 memristors, another rows are shifted left to create 7 columns. In testing process, there are 48 input channels after extracting the MFCC correspond 48 features to voice. These input channels have value in the range (-15, 15). Because the input voltage value has both negative and positive

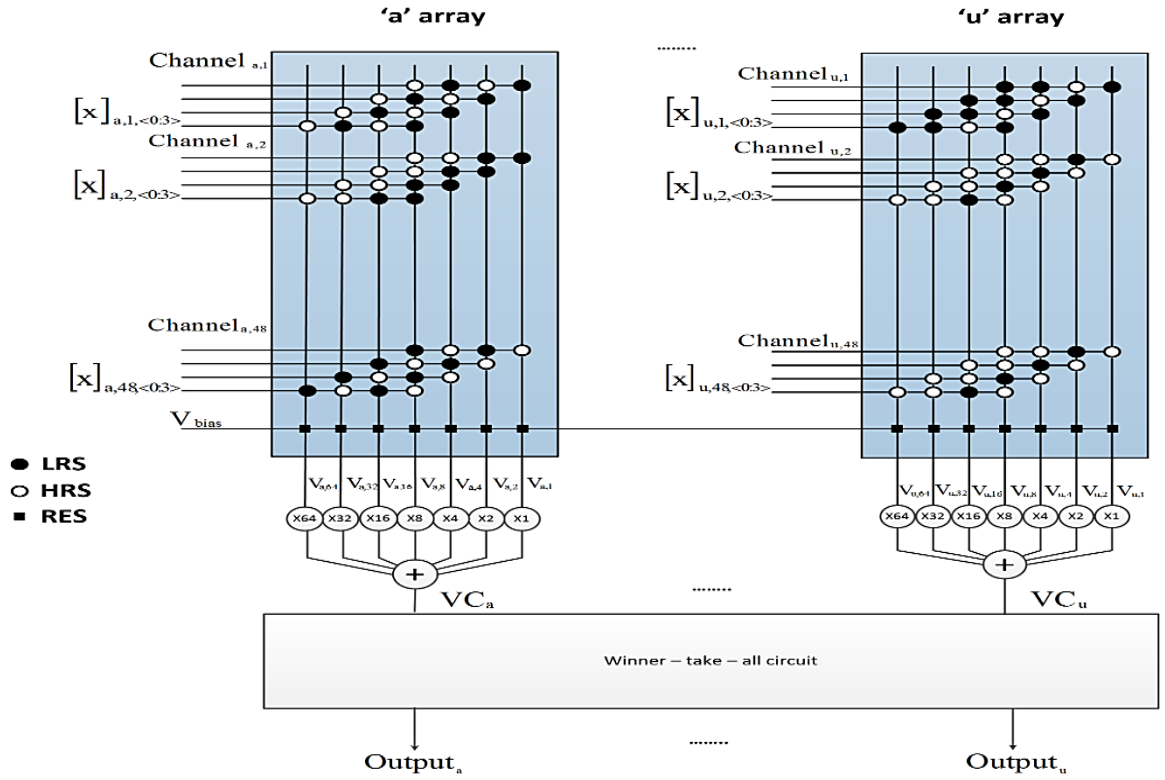

Fig 4: The block diagram of the proposed binary memristor crossbar circuit with 4-bit 48 input channels.

values, we add bias voltage to increase input voltage to positive voltage values. Figure 5b, this is multiplier circuit. The output is calculated by $S_a = \left(1 + \frac{R_2}{R_1}\right) V_a$ corresponding to the multipliers of 1, 2, 4, 8, 16, 32 and 64. Here, $R_2 = 0 * R_1, R_2 = 1 * R_1, R_2 = 3 * R_1, R_2 = 7 * R_1, R_2 = 15 * R_1,\ R_2 = 31 *\ R_1,\ R_2 = 63 * R_1$. Figure 5c is the adder circuit, $VC_a = S_{a,1} + S_{a,2} + S_{a,4} + S_{a,8} + S_{a,16} + S_{a,32} + S_{a,64}$. Here, the five capacitors of C_a, C_e, C_i, C_o, and C_u are represented to the five vowels 'a', 'e', 'i', 'o', and 'u'. We can determine that a certain vowel corresponding to the fastest-charged capacitor among the five capacitors of C_a, C_e, C_i, C_o and C_u is the biggest with the input of a human voice. The capacitor C_a can be charged by the weighted summation of VC_a, If the weighted summation of VC_a is large, C_a can be charged to VCC very fast. If the weighted summation of VC_a is small, it takes longer time to charge C_a to VCC. Then $VC_a, VC_e, VC_i, VC_o, VC_u$ are compared with a reference voltage V_{ref} through the comparison I_a, I_e, I_i, I_o and I_u. If VC_a is bigger than V_{ref} then D_a become high. VC_e, VC_i, VC_o and VC_u are smaller than V_{ref}, the ouputs of D_e, D_i, D_o and $D_u$become low. G_1, G_2, G_3 are the OR gates. A delay time τ between G_4 and G_5creates small CLK pulse. FF_a, FF_e, FF_i, FF_o and FF_u are flip flops with input D_a, D_e, D_i, D_o and D_u.The simulated waveforms of VC_a, VC_e, VC_i, VC_o and VC_u are shown in figure 6. Here, VC_a seems to be charged to VCC faster than the other capacitor nodes of VC_e, VC_i, VC_o and VC_u. So, the vowel 'a' is the best among the other vowels.

The timing diagram of important signals is shown in figure 6. When the CLK signal is high, all the capacitor nodes of VC_a, VC_e, VC_i, VC_o and $VC_u$are charged to VCC. At this time, VC_a, VC_e, VC_i, VC_o and $VC_u$are higher than V_{ref}.Thus, D_a, D_e, D_i, D_o and $D_u$can be high. If C_a is charged to VCC faster than C_e, C_i, C_o, C_u, VC_a gets the higher voltage level among VC_a, VC_e, VC_i, VC_o, and VC_u. If VC_a becomes higher than V_{ref}, D_a becomes high. So D_a can also be the fastest rising signal among $D_a, D_e, D_i, D_o, and\ D_u$. Since D_a generates the locking pulse that is the clock signal of D flip-flop circuits of FF1, FF2, FF3, FF4, and FF5, FF1 register leads to high output signal. So, we can determine which vowel is similar to the voice input. The signal of D_a will make $Output_a$ high and the other output signals become low, as shown in figure 6.

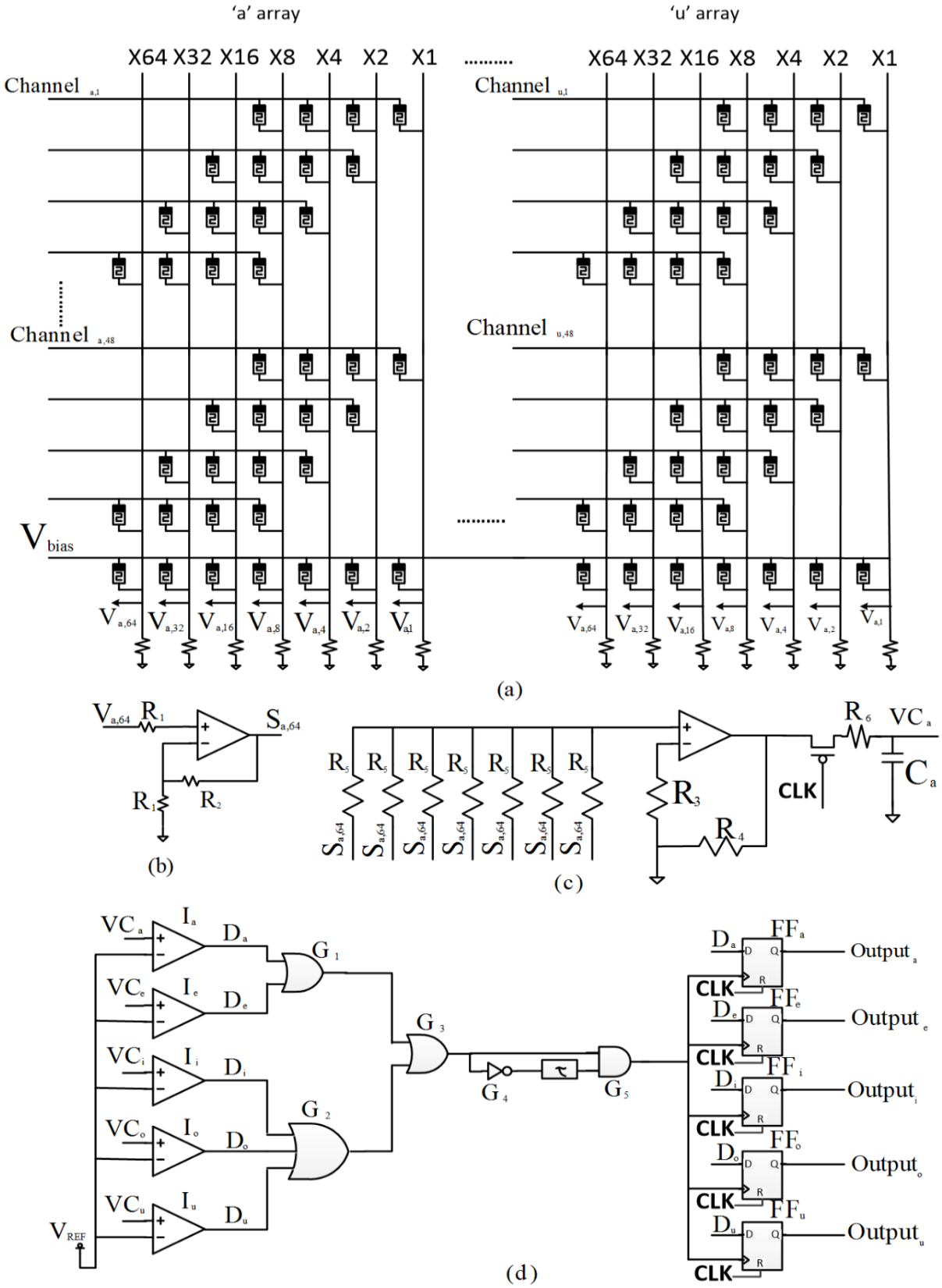

Fig 5: The schematics of the binary memristor crossbar circuit for speech recognition. (a) The schematic of the binary memristor crossbar circuit, (b) Voltage multiplier circuit, (c) Adder circuit, (d) The schematic of the winner-take-all circuit binary memristor circuit.

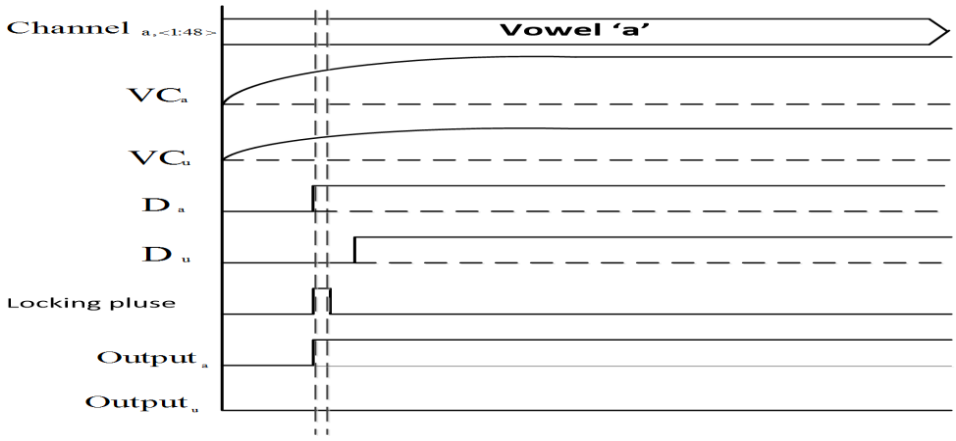

Fig 6: Voltage waveforms of the binary memristor crossbar and winner-take-all circuits.

III. RESULTS AND DISSCUSION

In this work, we used the MFCC method to process the voice signal, the MFCC includes preprocessing, framing, windown-ing, discrete fourier transforming, mel filter. Then, we received 48 MFCC values, that is feature of the voice signal and is data input for training process. We have 5 training times for 5 vowels. After the training process, the weights of each vowel will be quantized to 4-bit binary and stored into the memristor crossbar array by HRS or LRS. When test voice is quantized and converted to 4-bit voltage evels '-1', '0', '1'. In regconition circuit, voltage input is applied to binary memristors. Output voltage of winner take all circuit will decide the vowel which is tested.

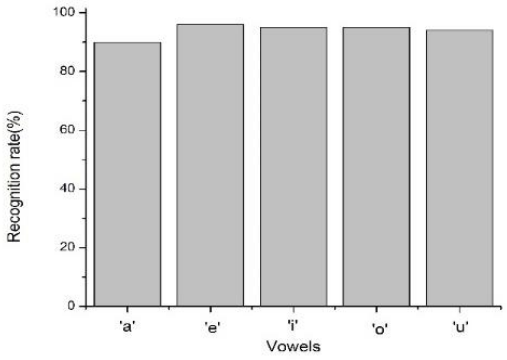

Fig7: The simulation results for the recognition rate of the proposed new binary memristor crossbar circuit.

The recognition rate shows in figure 7 with 1000 input voices for recognizing five different vowels. Each vowel is tested by 200 different voices. The average recognition rate of five different vowels is estimated to be around 94%. In the result, the recognition rate of 'e' is the highest at 96%. While recognition rate of vowel 'i', 'o' are 95%, 94%. The vowel 'u' and 'a' has the lowest recognition rate at 90%.

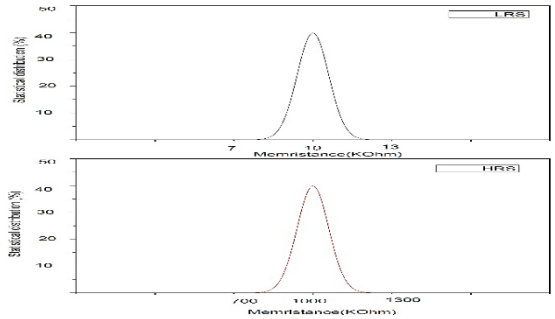

Fig 8: The statistical variation of memristance in HRS and LRS with the standard deviation (=σ) of 10%.

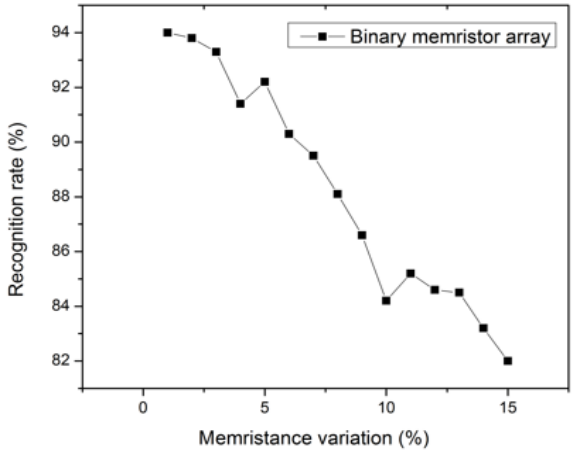

Fig 9: The recognition rate of the binary memristor crossbar with variation in memristance.

Figure 8 shows the statistical variation of memristance. The memristance of 1MΩ HRS and 10KΩ LRS have standard deviation (=σ) of 10%. The statistical variation was measured by Monte Carlo simulation that is performed by matlab software. Monte Carlo simulation estimates tolerant of recognized rate when memristance variation varies in range 0% to 15%. In figure 9, the recognition rate of the binary memristor crossbar is decreased very little only from 94% to 82% when the percentage statistical variation in memristance increase from 0% to 15%.

IV. CONCLUSION

In this paper, the new binary memristor crossbar based neural network model could recognize five vowels ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’ with 48 channels. Because each voice input has each feature so the proposed crossbar array can determine output signal by weights stored in memristance. We tested 1000 speech samples and verified to be able to recognize 94% of the total tested samples. It shows that using neural network apply in binary memristor crossbar gets better result than comparison among samples. In neural network, we only use a neural in hidden layer, so we have 5 times for training. In the further research, we will use more neural in hidden layers to raise recognition rate and focus on low power consumption, leakage current in binary memristor crossbar circuit.

REFERENCES

[1]. L. O. Chua, “Memristor – the missing circuit element,” *IEEE Trans. Circuit Theory, vol. CT-18, no. 5,* pp. 507-519, Sep. 1971.
[2]. D. B. Strukov, G. S. Snider, D. R.Stewart, and R. S. Williams, “The missing memristor found,” *Nature, vol. 453,* pp. 80-83, May 2008.
[3]. Son Ngoc Truong, Seok-Jin Ham and Kyeong-Sik Min, Neuromorphic crossbar circuit with nanoscale filamentary-switching binary memristors for speech recognition, *Nanoscale Research Letters 2014 9:629.*
[4]. Muda L, Begam M, Elamvazuthi I, Voice recognition algorithms using Mel frequency epstral coefficient (MFCC) and dynamic time warping (DTW) techniques, *J Comput 2010,2(3)*:138–143.
[5]. Raqibul Hasan and Tarek M. Taha, Enabling Back Propagation Training of Memristor Crossbar Neuromorphic Processors, *2014 International Joint Conference on Neural Networks (IJCNN)*.
[6]. Son Ngoc Truong, SangHak Shin, Sang-Don Byeon, JaeSang Song, Hyun-Sun Mo and Kyeong-Sik Min, Comparative Study on Statistical-Variation Tolerance Between Complementary Crossbar and Twin Crossbar of Binary Nanoscale Memristors for Pattern Recognition, *Nanoscale Research Letters* (2015*) 10:405.*