Variability Analysis of Memristor-based Sigmoid Function

Nursultan Kaiyrbekov, Olga Krestinskaya and Alex Pappachen James
Electrical and Computer Engineering Department
Nazarbayev University, Astana, Kazakhstan

Abstract—Activation functions are widely used in neural networks to decide the activation value of the neural unit based on linear combinations of the weighted inputs. The effective implementation of activation function is highly important to enhance the performance of a neural network. One of the most widely used activation functions is sigmoid. Therefore, there is a growing interest to enhance the performance of sigmoid circuits. In this paper, the main objective is to modify existing current mirror based sigmoid model by replacing CMOS transistors with memristive devices. We present the performance, variation of transistor sizes and temperature. The area, power and noise in the modified CMOS-memristive sigmoid circuit are shown. The application of memristors in the sigmoid circuit ensures the reduction of on-chip area, and power dissipation by 7%. The proposed sigmoid circuit was simulated in SPICE using TSMC 180nm CMOS design process.

Index Terms—Sigmoid, CMOS, Memristor, Artificial Neural Network

I. INTRODUCTION

One of important modules of artificial neural network is the activation function, which is normally represented by sigmoid function [1], [2], [3], [4], [5], [6]. A sigmoid function is applied to the output resulted in linear combination of input signals of neural network. Sigmoid functions are popular in neural networks mainly because of their derivative properties: they are computationally easy to perform. Sigmoid functions are widely used in learning algorithms for clustering [7], pattern recognition [8], function approximation [9], etc. Therefore, the improvement of a sigmoid activation function circuit is a rising need in designing new models in IC industry to meet specifications of minimized area, low power consumption, temperature independence and resistance to noise [10], [11], [12], [13].

There were several designs proposed for this purpose [14], [15], [5], [6]. However, large number of transistors, area and power consumption are the main drawbacks of listed circuits. This paper presents a modification of current mirror based sigmoid presented in [16]. The main advantage of this model is in ability of programming limits of sigmoid by adjusting two bias currents [16]. As the memristor is a promising solution used in various architectures [17], [18], [19], [20], [21], [22], introducing memristors further improves the circuit is terms of on-chip area and power dissipation.

This paper is structured as following. Section II goes through the background information about the existing circuits. In Section III, the methodology of the proposed design is discussed. Section IV illustrates simulation results and discussions. In Section V, the conclusion is made and several ideas for future works will be provided.

II. BACKGROUND

Generally, sigmoid function is one of nonlinear approaches for the activation function. It is described be Eq. 1.

\[
S(x) = \frac{1}{1 + e^{-x}}
\]  

(1)

The graphic representation of ideal sigmoid activation function is shown in Fig. 1.

![Sigmoid Function](image)

Fig. 1. Sigmoid Function

Activation functions are known to be used for neural networks applications [14]. Artificial neuron can be described by the following relationship:

\[
Y = \sum (\text{weight} \times \text{input}) + \text{bias}
\]  

(2)

where, neural networks can learn their biases and weights using backpropagation algorithm for cost functions [15], [5]. For particular weighted inputs, activation functions decide whether certain neuron should be fired (activated) or not. One of the obvious solution is to use step function. The output is high for values greater than threshold. However, it is binary, which is not applicable for multiple classes. The solution must be analog, therefore the linear function is considered, so that values in between can be taken into account; and its gradient is constant. This means that the error correction made by backpropagation is also constant and does not depend on change in input [15]. Therefore, sigmoid function has been proposed as a solution: it is real, continuous, non-linear and it can be differentiated. Despite the fact that it has the same
vanishing gradient problem as hyperbolic tangent function, it is still used widely in neural network implementations [23], [24], [5].

In paper [16], basic circuit of the sigmoid function is built as in Fig.2. Sigmoid function can be implemented using only 6 pairs of CMOS transistors. The circuit is regulated by input DC voltage. This circuit is simulated for 0.18μm technology, i.e. Length of transistors is set to 0.18μm, and Width is changed accordingly to satisfy specifications. The table providing width and length values is shown in Table I.

Here, the extra circuit has been added to avoid offset in the output. This is done to achieve zero-centered, symmetric sigmoid [16]. Also, biasing Voltage source has been connected to the gates of transistors M1 and M9 to turn them on.

| MOSFET  | TYPE | W/L (μm/μ) |
|---------|------|------------|
| M1      | PMOS | 1.5/0.18   |
| M2      | PMOS | 1/0.18     |
| M3, M5, M11 | PMOS | 4/0.18 |
| M4      | PMOS | 3/0.18     |
| M6      | NMOS | 2/0.18     |
| M9, M10, M12 | NMOS | 1/0.18 |
| M7, M8  | NMOS | 4/0.18     |

In Fig.4, \( g_m \) stands for transconductance, \( v_{gs} \) is gate to source voltage of the transistor and \( r_o \) is the output resistance. This model is used to get an expression for output current:

\[
I_d = K_n/2 *(V_{GS} - V_{TN})^2 (1 + \lambda V_{DS}) \
\]

Transconductance value is determined as

\[
g_m = \frac{2I_d}{V_{GS} - V_{TN}} = \sqrt{2K_nI_d} \
\]

and finally, output resistance is:

\[
r_o = \frac{1}{g_o} = \frac{1 + \lambda V_{DS}}{\lambda I_D} \
\]

III. METHODOLOGY

A. Modified Circuit Design of Sigmoid Activation Function

In the proposed design, transistors acting as resistors (in saturation mode) are replaced with memristors. For this purpose, memristor model created in shown in [25] is used. Fig.3 illustrates the proposed design. Two memristors are added to the circuit instead of transistors M2 and M10. The rest of transistors have the same width and length values as in original circuit.
IV. SIMULATION RESULTS AND DISCUSSION

A. Generalized Sigmoid function

First, the generalized sigmoid function was constructed. The standard formula for a generalized sigmoid is defined as:

\[ Y(t) = A + \frac{(K - A)}{(C + Qe^{-Bt})^{1/v}} \]  

(7)

By analyzing the output from Fig.3, the equation referred to that function could be derived:

\[ Y(x) = -510 + 490 \times 2 \times 10e^{-1} 
\]

The plot of this equation was constructed using MATLAB, and two signals are compared, which is illustrated in Fig. 6.

Therefore, mathematical model of an output from Fig.3 was derived.

B. Temperature analysis

The temperature analysis has been done for both original and memristor-based circuit. The temperature is varied between -50°C and 50°C with increment of 10°C.

The proposed and original circuits have the same reaction to the change in temperature, as it can be seen from Fig.7 and Fig.8. The values are in close range to ±500μA.

C. Area and Power calculations

1) Area: Area of transistors in original circuit is calculated as follows:

\[ A_{\text{transistor}} = 3WL \]

Therefore, summing up area of all 12 transistors in Table I will result in

\[ A_{\text{original}} = 14.85\mu m^2 \]
TABLE II
SIMULATION RESULTS AND COMPARISON OF CONVENTIONAL CMOS SIGMOID AND PROPOSED MEMRISTOR-BASED SIGMOID CIRCUIT.

|                  | Area (μm²) | Power (mW) | THD(%) |
|------------------|------------|------------|--------|
| Sigmoid using CMOS | 14.85      | 3.2        | 6.08   |
| Sigmoid using memristors | 13.775     | 3.0        | 5.67   |

Now, for the proposed circuit, assuming memristors to have dimensions 50nm x 100nm, the overall area is going to be:

\[ A_{\text{proposed}} = 13.775 \mu m^2 \]

The area has been reduced by 7.2%.

2) Power: Power is calculated element-wise: average power for every element is summed up to get overall circuit power. This is done in SPICE. For original circuit the average power between -3V and 3V is observed to be 5.2mW. For the proposed circuit, due to 2 memristors replacing transistors, the average power is significantly reduced to 3mW.

D. THD analysis

Total harmonic distortion analysis has been done for both original and proposed sigmoid. A 5V sine AC signal is applied as input with amplitude 1V and frequency 1kHz. The total harmonic distortion is 5.67% for proposed circuit compared to 6.08% of original circuit. Reduction of Total Harmonic Distortion in 0.41% is achieved.

The table showing different parameters comparing original and proposed circuits is presented in Table II.

V. CONCLUSION

This paper presented modified sigmoid circuit with two transistors acting in saturation mode being replaced by memristor models. Current mirror based sigmoid was recreated, and compared with proposed model. The slight improvement of efficiency in terms of area and noise has been achieved. Moreover, significant power reduction was possible by introducing memristors. The future works will include improvement of total harmonic distortion, as 3% might be too large for some application requiring minimum noise. The one way of doing it is to vary W/L ratio. Also, the possibility of adding more memristors should be investigated, with keeping in mind that the ability of programming sigmoid limits has to be kept.

REFERENCES

[1] Y. Zhang, X. Wang, and E. G. Friedman, “Memristor-based circuit design for multilayer neural networks,” IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 65, no. 2, pp. 677–686, Feb 2018.
[2] J. A. Starzyk and Basawaraj, “Memristor crossbar architecture for synchronous neural networks,” IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 61, no. 8, pp. 2390–2401, Aug 2014.
[3] M. Al-Nsour and H. S. Abdel-Aty-Zohdy, “Implementation of programmable digital sigmoid function circuit for neuro-computing,” in Circuits and Systems, 1998. Proceedings. 1998 Midwest Symposium on. IEEE, 1998, pp. 571–574.
[4] I. Yeo, S.-g. Gi, B.-g. Lee, and M. Chu, “Stochastic implementation of the activation function for artificial neural networks,” in Biomedical Circuits and Systems Conference (BioCAS), 2016 IEEE. IEEE, 2016, pp. 440–443.
[5] O. Krestinskaya, K. N. Salama, and A. P. James, “Analog backpropagation learning circuits for memristive crossbar neural networks,” in Circuits and Systems (ISCAS), 2018 IEEE International Symposium on. IEEE, 2018.
[6] K. Smagulova, O. Krestinskaya, and A. P. James, “A memristor-based long term memory circuit,” Analog Integrated Circuits and Signal Processing, pp. 1–6, 2018.
[7] L. Yen, E. Fouss, C. Dacaetecker, P. Francq, and M. Saerens, “Graph nodes clustering with the sigmoid commute-time kernel: A comparative study,” Data & Knowledge Engineering, vol. 68, no. 3, pp. 338–361, 2009.
[8] R. P. Lippmann, “Pattern classification using neural networks,” IEEE communications magazine, vol. 27, no. 11, pp. 47–50, 1989.
[9] Y. Ito, “Approximation of functions on a compact set by finite sums of a sigmoid function without scaling,” Neural Networks, vol. 4, no. 6, pp. 817–826, 1991.
[10] A. Jaya, G. Soemarno, and J. Puspita, “Classification of epileptiform waves based on frequency by using backpropagation neural network,” in Journal of Physics: Conference Series, vol. 1028, no. 1. IOP Publishing, 2018, p. 012048.
[11] B. S. Saljoughi and A. Hezarkhani, “A comparative analysis of artificial neural network (ann), wavelet neural network (wnn), and support vector machine (svm) data-driven models to mineral potential mapping for copper mineralizations in the shahr-e-babak region, kerman, iran,” Applied Geomatics, pp. 1–28, 2018.
[12] A. V. Mainkar et al., “Improved initialization for the multi layer perceptron,” Ph.D. dissertation, 2018.
[13] S. Harold and T. Yamakawa, “Isothermal brain predicting macculloch-pitts logic implicating walter freeman ionic diffusion and loth zadeh fuzzy logic that might be useful for coding autonomous vehicles to be coexisting with the pedestrians,” MOJ App Bio Biomech, vol. 2, no. 3, pp. 177–184, 2018.
[14] G. Bogason, “Generation of a neuron transfer function and its derivatives,” Electron Lett, pp. 1867–1869, 1993.
[15] R. W. David Rumelhart, Geoffrey Hinton, “Learning representations by back-propagation errors,” Nature, vol. 323, no. 9, 9 October, 1986.
[16] S. Tabarce, “Programmable analogue vlsi implementation for asymmetric sigmoid neural activation function and its derivative,” Electronic Letters, vol. 41, no. 15, 21st July, 2015.
[17] O. Krestinskaya and A. P. James, “Feature extraction without learning in an analog spatial pooler memristive-cmos circuit design of hierarchical temporal memory,” Analog Integrated Circuits and Signal Processing, pp. 1–9, 2018.
[18] O. Krestinskaya, T. Ibrayev, and A. P. James, “Hierarchical temporal memory features with memristor logic circuits for pattern recognition,” IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2017.
[19] A. James, T. Ibrayev, and O. Krestinskaya, “Design and implication of a rule based weight sparsity module in htm spatial pooler,” in Electronics , Circuits and Systems (ICECS), 2017 24th IEEE International. IEEE, 2017.
[20] N. Dastanova, S. Duisenbay, O. Krestinskaya, and A. P. James, “Bit-plane extracted moving-object detection using memristive crossbar-cam arrays for edge computing image devices,” IEEE Access, vol. 6, pp. 18954–18966, 2018.
[21] A. James, T. Ibrayev, O. Krestinskaya, and I. Dolzhikova, “Introduction to memristive htm circuits,” in Memristor and Memristive Neural Networks. InTech, 2018.
[22] A. Irmanova and A. P. James, “Neuron inspired data encoding memristive multi-level memory cell,” Analog Integrated Circuits and Signal Processing, pp. 1–6, 2018.
[23] R. Marugados, “Universal approximation of nonlinear system predictions in sigmoid activation functions using artificial neural networks,” 2014 IEEE International Conference, 18 December, 2014.
[24] A. Sharma, “Understanding activation functions in neural networks,” medium.com, Accessed on 10.02.2018, Available on https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-5491262884e0.html.
[25] Z. Biolek, D. Biolek, and V. Biolekova, “Spice model of memristor with nonlinear dopant drift,” Radioengineering, vol. 18, no. 2, 2009.
[26] S. Liu and L. W. Nagel, “Small-signal mosfet models for analog circuit design,” IEEE Journal of Solid-State Circuits, vol. 17, no. 6, pp. 983–998, 1982.