

Implementation of a programmable neuron in CNTFET technology for low-power neural networks

Seyed Moosa Seyed Aalinejad

Department of Electronics, Urmia University of Technology, Urmia, West Azerbaijan, Iran

Received 27 November 2019;

revised 06 March 2020;

accepted 06 March 2020;

available online 19 March 2020

Abstract

Circuit-level implementation of a novel neuron has been discussed in this article. A low-power Activation Function (AF) circuit is introduced in this paper, which is then combined with a highly linear synapse circuit to form the neuron architecture. Designed in Carbon Nanotube Field-Effect Transistor (CNTFET) technology, the proposed structure consumes low power, which makes it suitable for the implementation of high-throughput Neural Networks (NNs). The main advantage of the proposed AF circuit is its higher accuracy for the generation of hyperbolic tangent function compared to the previously reported works. Moreover, the programmability feature for the slope and the position shifting enhances the adaptability of the designed neuron for different types of neural systems, especially Multi-Layer Perceptrons (MLPs). There is also excellent compatibility between the synapse and activation circuits, which illustrates another notable privilege of the proposed neuron. Simulations using HSPICE for CNTFET 32 nm standard process have been carried out for the designed scheme to indicate the correct operation. Based on the results, all of the claimed advantages can be proved clearly while the power dissipation is $6.11\mu\text{W}$ from the 0.9V power supply. Also, an accuracy of 98% has been achieved for the AF circuit.

Keywords: Activation Function; Artificial Neural Networks; CNTFET; Logistic Function; Neuron; Synapse.

How to cite this article

Seyed Aalinejad SM. Implementation of a programmable neuron in CNTFET technology for low-power neural networks. *Int. J. Nano Dimens.*, 2020; 11 (2): 120-129.

INTRODUCTION

As the definition suggests, the set of procedures that are utilized together for the modeling of the human brain are the basis of the systems which have found their reputation as Neural Networks (NNs). Such systems are extensively used to predict human behavior and the dynamics of natural events around us. The classification of NNs that deals explicitly with the animal brain is known as Artificial Neural Networks (ANNs). In such systems, the main objective is the approximation of the complicated functions with multiple inputs [1]. The general architecture of an ANN in its purest form has been shown in Fig. 1(a). By considering such a model, every circle will represent a neural node, while the arrows are used for the interconnection of the different neurons.

Moreover, one of the famous implementations of such a scheme known as Perceptron can be realized using the architecture of Fig. 1(b). Four

essential parts of a Perceptron NN are as follows:

- 1- The input values that are taken from nature.
- 2- The weights and bias that emphasize the importance of each input value.
- 3- The Sum node that accumulates the weighted input values.
- 4- The Activation Function (AF) that performs the decision task.

The inputs constitute a vector that will be injected into the network. The second stage is the modeling of synaptic operations, while the summations are performed in a node. Finally, the accumulated data are evaluated by AF to classify the input vectors [2].

The first event which put the ANNs in the center of attention was the introduction of the primary computing neuron in 1943 [3]. McCulloch and Pitts set the start point by their research. After that, a variety of models were developed by scientists to describe the mathematical relations for the biological NNs [2]. However, the next step,

* Corresponding Author Email: seyed_moosa@uut.ac.ir

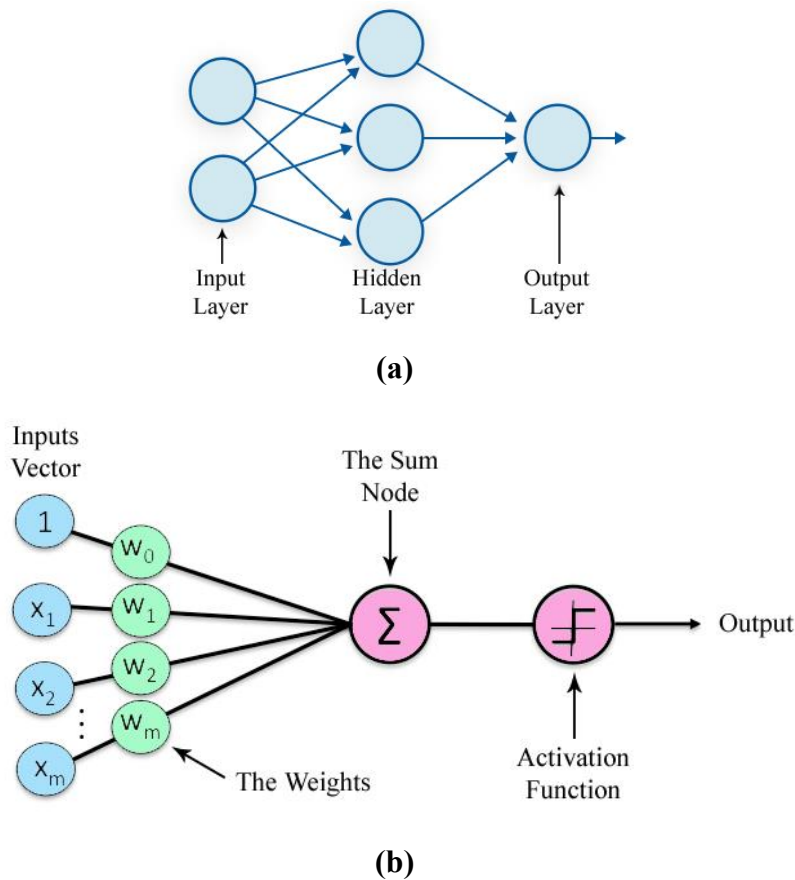


Fig. 1. General Model of (a) multilayer ANN (b) Perceptron network.

which was the hardware realization of ANNs, was boosted by the rapid growth of silicon technologies at the beginning of the 1990s [4, 5]. In recent decay, more researches have been reported in literature where the emphasis was put on power reduction of ANNs [6-11]. The main differences between those works include the type of AF (step or sigmoid function), the accuracy, and the power dissipation of AF. On the other hand, the common point between most of the previous works is the operational principle, which is based on voltage-mode.

However, considering the fact that each neuron is at last composed of one synapse circuit along with the AF block, there will be several choices for implementation of each section [5]. For instance, in [6, 8, 11], the AF circuit was designed in transresistance mode while in [9]; voltage-mode is generally selected for the realization of both synapse and AF. Besides the differences, there must always be excellent compatibility between

synapse and AF for the precise operation of the neuron and, thereafter, the whole network [6, 8].

Following the submicron technology advances, especially for Carbon Nanotubes (CNTs) in which their diameter size does not exceed the scale of nanometer [12], more options have been provided for the researchers in size and power reduction of electronic systems. Among the different classes of CNTs, Single-Wall CNTs (SWCNTs) have demonstrated excellent properties making them the right choice for the fabrication of Integrated Circuits (ICs) [13]. However, a long road has been traversed so that CNT Field-Effect Transistor (CNTFET) based circuits came into realization [14].

Nowadays, circuit-level implementation of CNTFET based circuits has found more popularity, and there are several works reported in the literature to fulfill the promise [15, 16]. A similar story also goes on for the hardware realization of ANNs. Those works can be classified into two distinct groups. The first group has focused on

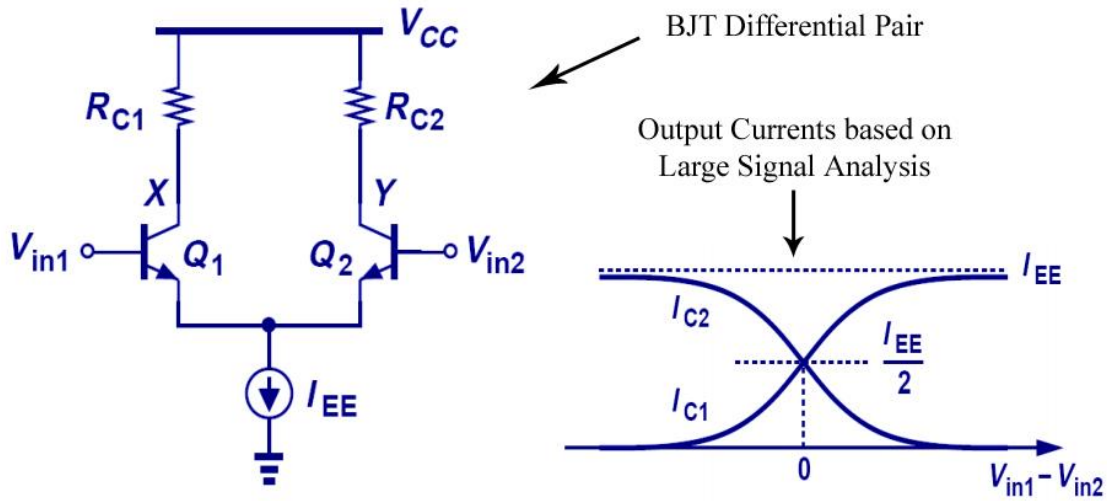


Fig. 2. BJT differential pair along with the output waveforms obtained from the large-signal analysis.

CNTFET modeling with the help of ANNs [17-19], while the researchers of the second group emphasized the design and fabrication of CNTFET based NNs [20-22]. However, the thorough investigation of those works depicts that there are few works that have considered the development of a new architecture for a neuron in CNTFET technology.

In this work, a novel circuit for the AF has been proposed to generate the hyperbolic tangent function. Meanwhile, the synapse circuit reported in [8, 11]; which is an analog multiplier, has been redesigned and adapted with the proposed AF circuit so that a complete architecture of a neuron can be obtained. The designed structures have been simulated using HSPICE for CNTFET 32nm standard process and 0.9V power supply. Low power consumption and high accuracy are the advantages of the proposed neuron. Section 2 pertains to the implementation of the AF and its adaptability with the redesigned synapse circuit. The simulation results, along with comparative analysis, are discussed in section 3. Finally, the conclusions are summarized in section 4.

EXPERIMENTAL

Activation Function Circuit

The most popular class of functions, which is used for the implementation of AF circuits are sigmoid functions [11]. Logistic and hyperbolic tangent functions constitute two common waveforms of this category, and depending on the application, the proper function will be employed.

Here, the hyperbolic AF is selected as it can easily be generated by means of differential amplifiers. Such a function is described by the following mathematical expression:

$$f(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}} = \frac{e^{2t} - 1}{e^{2t} + 1} \tag{1}$$

Now if the Bipolar Junction Transistor (BJT) based differential pair of Fig. 2 is considered, the large signal analysis demonstrates that [23]:

$$I_{C2} - I_{C1} = I_{EE} \tanh\left(\frac{V_{in1} - V_{in2}}{2V_T}\right) \tag{2}$$

where I_{C1} and I_{C2} represent the collector currents for each transistor, while I_{EE} defines the bias current. Also, V_{in1} and V_{in2} denote the input voltages, and V_T illustrates the thermal voltage of the BJT. As (2) demonstrates, a hyperbolic expression exists for each differential pair structure. The same hypothesis with more complicated relations can also be proved for Metal–Oxide–Semiconductor (MOS) transistors.

The normalized value of the hyperbolic tangent function is bounded in the interval of -1 and 1. Therefore, the scheme of Fig. 2 is used as the basis for the implementation of the AF circuit. The proposed scheme has been shown in Fig. 3, which is composed of 10 transistors. The core of the structure is formed by $M_1, M_2, M_3, M_4,$ and

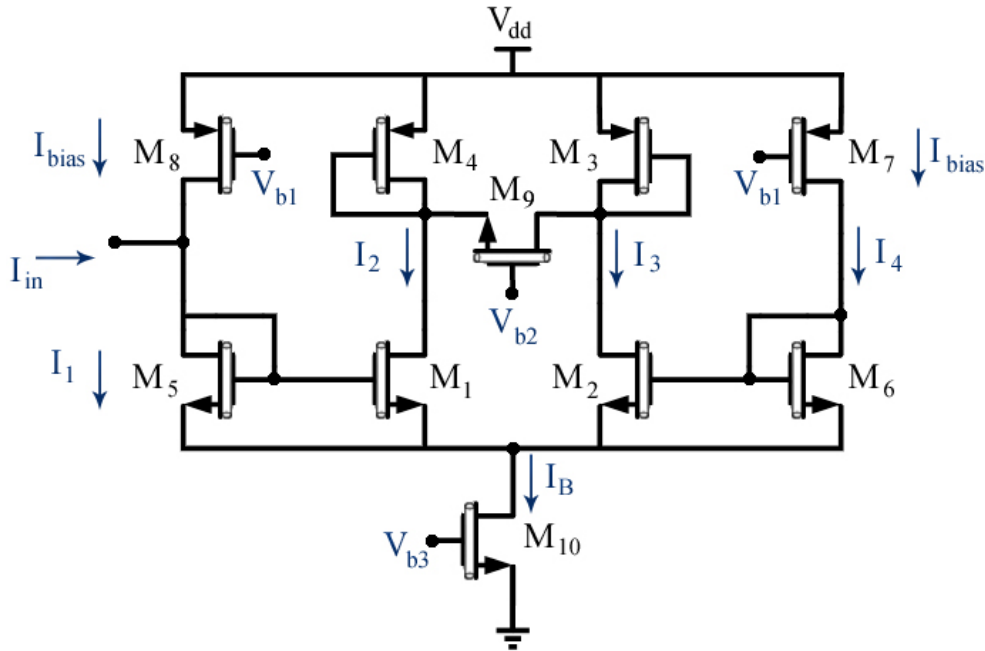


Fig. 3. The proposed Activation Function circuit.

M_{10} , which constitute the differential amplifier. On the other hand, M_5 , M_6 , M_7 , and M_8 play the role of the bias circuit for the architecture. In a balanced condition where, $I_{in} = 0$, the same bias current will flow through the differential pair by considering the fact that the size of counterpart transistors ($M_5 - M_6$ and $M_7 - M_8$) are identical. It must be mentioned that M_9 acts as a controller to adjust the accuracy of the output waveform.

The input current (I_{in}) comes from the synapse circuit. As a consequence, the current of the diode-connected transistor M_5 , which is denoted by I_1 in Fig. 3, will be:

$$I_1 = I_{in} + I_{bias} \quad (3)$$

This current will be copied to M_1 via the current source. On the other side, the current passing through M_6 (I_4) will be equal to I_{bias} . Hence, by means of the right-side current mirror, I_{bias} will be copied to M_2 . The output waveform will be obtained through the drain-source voltage of M_9 (V_{ds9}). On the other hand, it is obvious that:

$$V_{out} = V_{ds9} = V_{d3} - V_{d4} \quad (4)$$

The diode-connected transistors M_5 and M_4 will demonstrate resistances equal to $\frac{1}{g_{m3}}$ and $\frac{1}{g_{m4}}$,

respectively. Because the differential pair is assumed to be balanced on both sides, then the corresponding sizes of M_3 and M_4 will be identical. Therefore, it can be written as:

$$\frac{1}{g_{m3}} = \frac{1}{g_{m4}} = R \quad (5)$$

The whole structure is fed via M_{10} , which provides the bias current. According to Fig. 3, it is obvious that:

$$I_B = I_1 + I_2 + I_3 + I_4 \quad (6)$$

As discussed in previous paragraphs:

$$\begin{cases} I_2 = I_{in} + I_{bias} \\ I_3 = I_{bias} \\ I_4 = I_{bias} \end{cases} \quad (7)$$

Substitution of (7) and (3) in (6), results in:

$$I_B = 2I_{in} + 4I_{bias} \quad (8)$$

Therefore, I_{in} will be equal to:

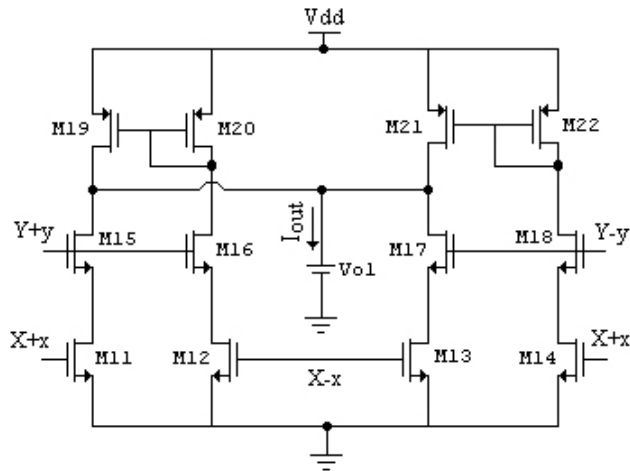


Fig. 4. The employed synaptic circuit of [8].

$$I_{in} = \frac{I_B}{2} - 2I_{bias} \quad (9)$$

By means of the obtained value for I_{in} , the output voltage of the AF circuit concerning the changes of input current from the synapse can easily be obtained. To do this, (4) can be rewritten as:

$$V_{out} = V_{d3} - V_{d4} = R(I_3 - I_2) \quad (10)$$

and with the help of (7), (10) will be simplified as:

$$V_{out} = -RI_{in} = R\left(2I_{bias} - \frac{I_B}{2}\right) \quad (11)$$

As (11) suggests, along with the increment of input current from the synapse, the output voltage will also rise to high-level value. Also, based on (2), the waveform will have a hyperbolic shape ranging from a fixed negative value to a certain positive constant quantity.

The synapse and Compatibility Issue

For the correct operation of the AF, a suitable synapse circuit has to be employed to fill out the architecture of the neuron. It will also help to investigate the precision of AF structure. The simplest structure for implementation of the synapse is the analog multiplier. In this article, the synaptic scheme reported in [8] has been employed due to its higher operating range compared with the other designs. The scheme

has been shown in Fig. 4, which is redesigned in CNTFET technology and is adjusted to operate with the power supply of 0.9V.

However, as discussed in the previous section, the adaptability between synapse and AF circuit is an important issue which needs to be analyzed carefully. When a synapse and an AF are connected to each other, there must be no loading effect from the AF circuit on the synapse structure. Because the proposed AF scheme operates in transresistance mode, the output current from synapse will act as the input signal for AF. As a consequence, it is mandatory for the AF to have an input resistance much smaller than the output resistance of the synapse. Fig. 5 illustrates the small-signal equivalent circuit for the proposed AF structure.

According to Fig. 5, the input resistance for the proposed AF circuit will be equal to:

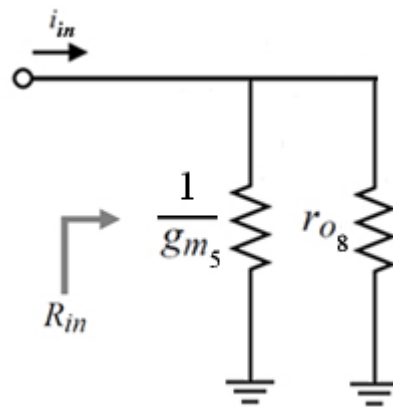


Fig. 5. The small-signal equivalent circuit for the proposed AF scheme.

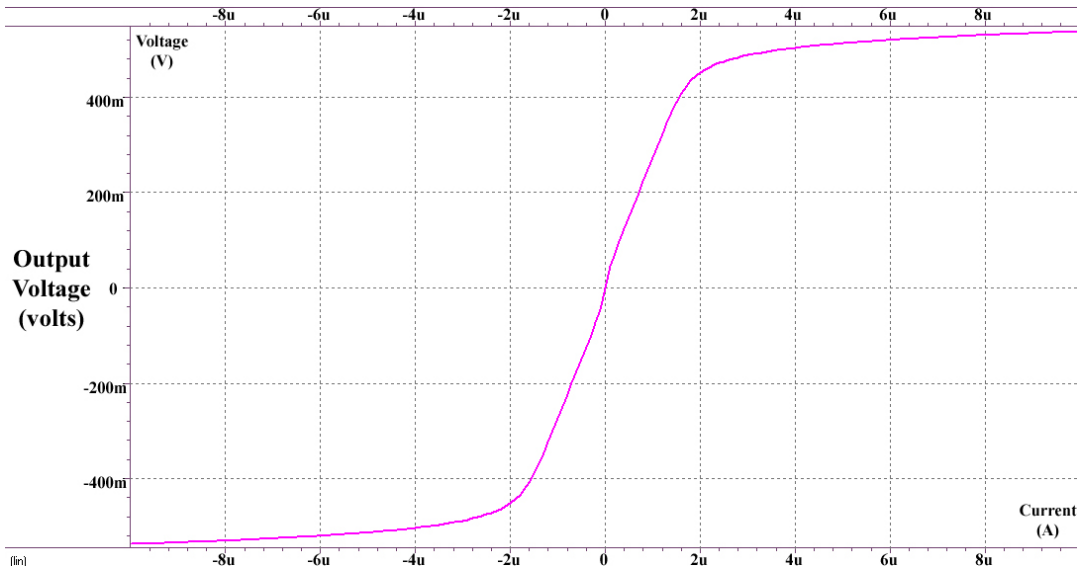


Fig. 6. Simulation results for the correct operation of the proposed AF circuit.

$$R_{in} = \frac{1}{g_{m5}} \parallel r_{o8} \quad (12)$$

where $\frac{1}{g_{m5}}$ is the equivalent resistance of the diode-connected transistor M_5 . On the other hand, r_{o8} represents the drain-source resistance of M_8 . For the typical bias currents, $r_{o8} \gg \frac{1}{g_{m5}}$ which simplifies (12) to:

$$R_{in} \cong \frac{1}{g_{m5}} \quad (13)$$

On the other side, the output resistance for the synapse circuit will be obtained by the following expression:

$$R_{out} \cong r_{o19} \parallel r_{o21} \quad (14)$$

Comparison of (13) and (14) demonstrates that $R_{in} \ll R_{out}$. Therefore, it can be stated that there is moderate compatibility between the synapse and AF circuit.

RESULTS AND DISCUSSIONS

Simulations using HSPICE for CNTFET 32nm standard process and 0.9V power supply have been carried out to investigate the correct operation of the designed AF scheme. The result,

which is shown in Fig. 6, depicts that the proposed circuit successfully generates hyperbolic tangent waveform on the output nodes.

In order to check the position shifting feature, the bias voltages for M_7 and M_8 (V_{b1}) were swept, and the simulations have been performed for different values of V_{b1} . The result, which is demonstrated in Fig. 7, illustrates the programmability feature of the proposed AF circuit. Moreover, the corresponding sizes of M_1 and M_2 were changed (by varying the ratio of $\frac{W}{L}$) to pursue the capability of the designed scheme for the slope changing of hyperbolic function. To do this, parallel transistors have been utilized to increase the proportional ratio. Fig. 8 shows the result, which justifies the theoretic assumptions.

To calculate the accuracy of the proposed AF circuit, the simulation results obtained from HSPICE were compared with the hyperbolic tangent function. The error function is defined by subtracting the simulation results from the ideal function. The result is shown in Fig. 9, where the error function fluctuates from -9mV to 11mV value. As the value of simulated hyperbolic function ranges from -500mV to 500mV, the following expression will calculate the accuracy of the designed AF circuit:

$$Accuracy \cong \frac{[0.5 - (-0.5)] - [0.011 - (-0.009)]}{0.5 - (-0.5)} \times 100\% = 98\% \quad (15)$$

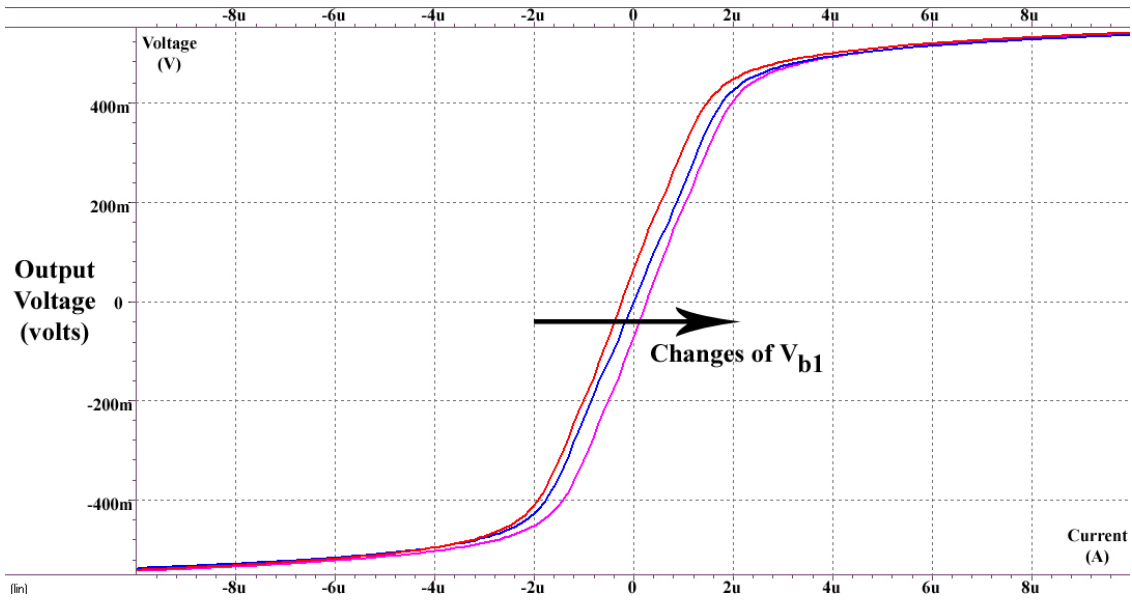


Fig. 7. Simulation results for the position shifting feature of the proposed AF circuit.

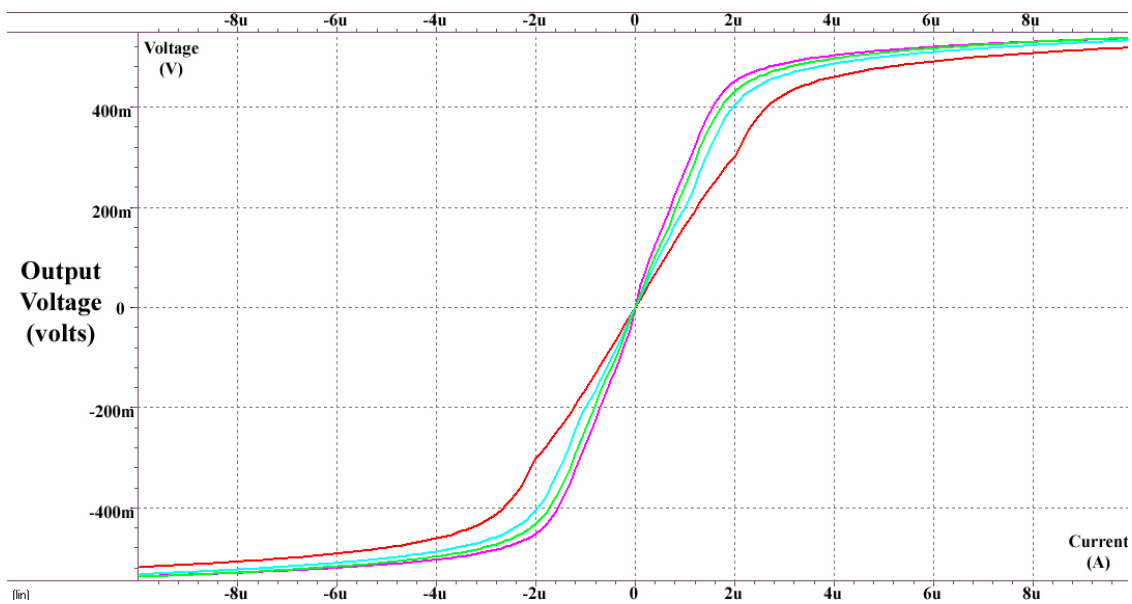


Fig. 8. Simulation results for the slope changing feature of the proposed AF circuit.

As (15) depicts, the accuracy equal to 98% has been achieved for the designed scheme, which illustrates the superiority of this work over previous structures.

In order to investigate the adaptability between AF function and synapse, the circuits of Fig. 3 and Fig. 4 were cascaded and simulated by varying the control parameters of the synapse circuit

(x and y). The results for the voltage changes at the output node of the synapse circuit are shown in Fig. 10. As the results demonstrate, the adaptability is at an excellent level, which proves the validity of provided mathematical expressions in the previous section.

Finally, the effect of temperature variations on the designed AF circuit has been analyzed.

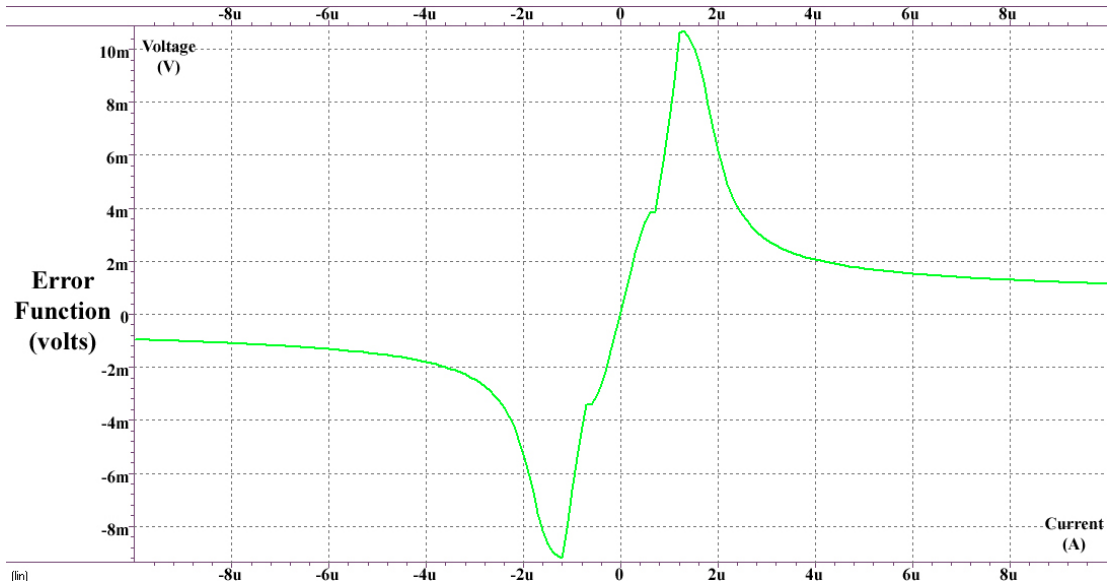


Fig. 9. Error function of the designed AF circuit.

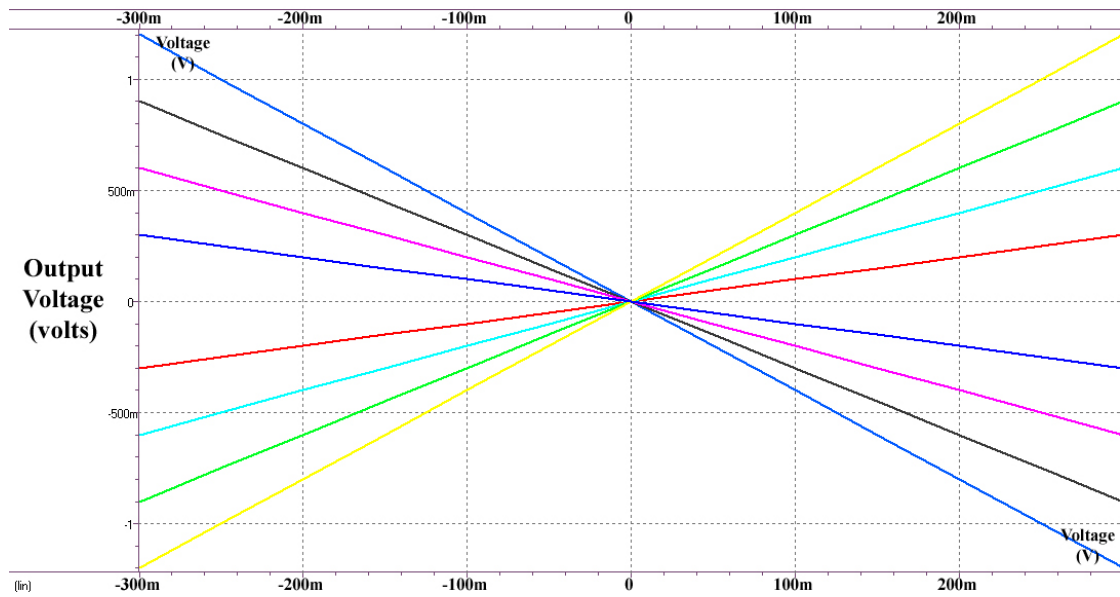


Fig. 10. Full adaptability between synapse and AF circuit.

To perform this, the temperature was swept from -20°C to 70°C in CNTFET technology, and the circuit was simulated using HSPICE. The result, which has been illustrated in Fig. 11, indicates the correct operation of the proposed scheme at a wide temperature range. Moreover, Table 1 summarizes the comparative analysis of this work along with previous designs. As the results indicate, this paper introduces the first

implemented neuron in CNTFET technology, while significant optimizations on power consumption and accuracy have been achieved here.

CONCLUSIONS

In this manuscript, a novel architecture for hardware realization of ANNs in CNTFET technology has been introduced. The proposed structure was composed of two building blocks where the AF

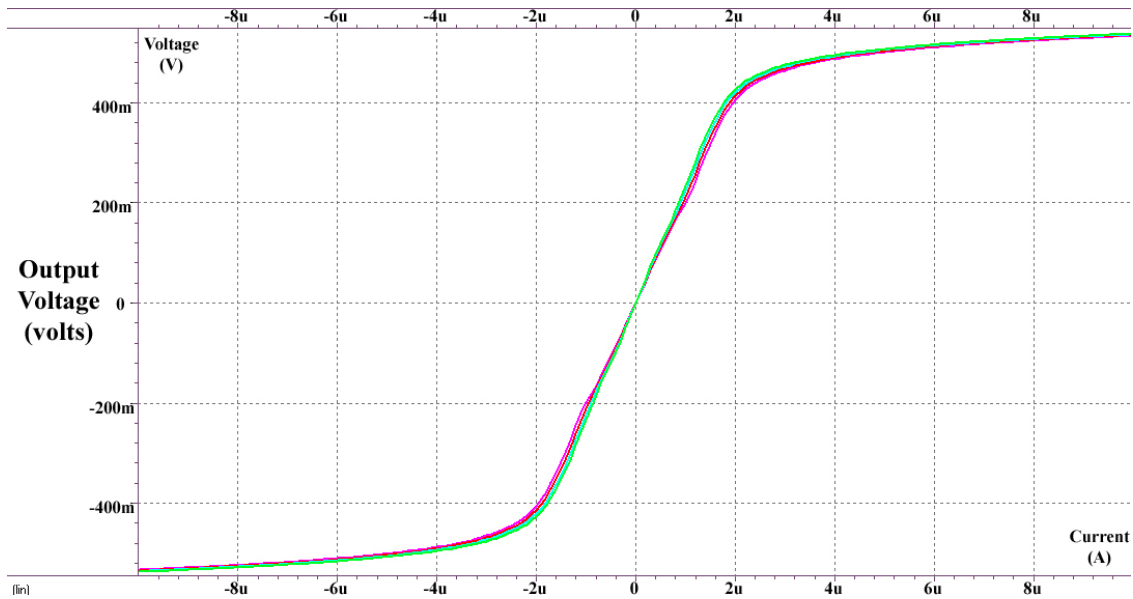


Fig. 11. The effect of temperature changes on the output of the AF circuit.

Table 1. Comparison of The proposed AF Circuit and Previous works.

	This Work	[6]	[7]	[8]	[9]	[10]	[11]
Process(nm)	32	350	180	350	90	180	180
Technology	CNTFET	CMOS	CMOS	CMOS	CMOS	CMOS	CMOS
Programmability	√	×	√	√	×	√	√
Accuracy	98%	NA	96%	NA	96%	93%	96%
Transistor Count	10	5	NA	10	6	3	5
Power(μW)	6.11	50	NA	99	473	62.5	72

circuit was newly designed, and the synapse was adapted from the previous works. High accuracy and low power consumption were the notable advantages of the proposed AF circuit. Moreover, there was excellent compatibility between synapse and AF, which qualifies this work to be widely employed in the hardware implementation of MLPs. Meanwhile, the programmability feature for the slope and the position shifting enhances the adaptability of the designed neuron for different types of neural systems.

Simulation results using HSPICE for CNTFET 32nm standard process and 0.9V power supply confirm the correct behavior of the designed neuron. According to the results, all of the claimed advantages were completely proved. Also, the power dissipation for the AF circuit was 6.11μW, while the accuracy of 98% has been achieved concerning the ideal hyperbolic tangent function.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

- [1] Hentrich M., (2015), Methodology and Coronary Artery Disease Cure, *SSRN*.
- [2] Zurada J., (1992), *Introduction to Artificial Neural Systems*, West Publishing Company, USA.
- [3] McCulloch W. S., Pitts W., (1943), A logical calculus of the ideas immanent in nervous activity. *The bull. Math. Biophys.* 5: 115–133.
- [4] Ota Y., Wilamowski B. M., (1994), VLSI Implementation of a Programmable Current-Mode Neural Network, *Proceedings: Intellig. Eng. Sys. Through Artific. Neural Networks, ANNIE'94 Conf., St. Louis, 4: 71-76*.
- [5] Cauwenberghs G., Bayoumi M., (1999), *Learning on Silicon, Adaptive VLSI Neural Systems*, Kluwer Academic Publishers.
- [6] Wojtyna R., Talaska T., (2006), Transresistance CMOS neuron for adaptive neural networks implemented in hardware, *Bull. Pol. Acad. Sci.* 54: 443-448.



- [7] Hosseinzadeh Namin A., Leboeuf K., Muscedere R., Wu H., Ahmadi M., (2009), Efficient hardware implementation of the hyperbolic tangent sigmoid function, *IEEE Int. Symp. Circ. Sys.* 2117-2120.
- [8] Azizian S., Fathi K., Mashoufi B., Derogarian F., (2011), Implementation of a programmable neuron in 0.35 μ m CMOS process for multi-layer ANN applications. *Eurocon-Int. Conf. Comp. Tool.* 1-4.
- [9] Khodabandehloo G., Mirhassani M., Ahmadi M., (2012), Analog implementation of a novel resistive-type sigmoidal neuron. *IEEE-Transact. Very Large Scale Integ. (VLSI) Systems* 20: 750-754.
- [10] Shamsi J., Amirsoleimani A., Mirzakuchaki S., Ahmadi A., Alirezaee Sh., Ahmadi M., (2015), Hyperbolic tangent passive resistive-type neuron. *2015 IEEE- Int. Symp. Circ. Sys. (ISCAS)*. 581-584.
- [11] Azizian S., Aziz Aghchegala V., Azizian S., Sefidgar Dilmaghani M., (2019), CMOS bulk-controlled fully programmable neuron for artificial neural networks. *IETE J. Res.* 65: 320-328.
- [12] Liu C., Cheng H., (2013), Carbon nanotubes: Controlled growth and application. *Mater. Today.* 16: 19-28.
- [13] Avouris P., Afzali A., Appenzeller J., Chen J., Freitag M., Klinker C., Lin Y.-M., Tsang J. C., (2004), Carbon nanotube electronics and optoelectronics. *Tech. Digest-IEEE Int. Elec. Dev. Meeting (IEDM)*. 525-529.
- [14] Franklin A. D., (2013), Electronics: The road to carbon nanotube transistors. *Nature.* 498: 443-444.
- [15] Rahnamaei A., Zare Fatin G., Eskandarian A., (2019), Design of a low power high speed 4-2 compressor using CNTFET 32 nm technology for parallel multipliers. *Int. J. Nano Dimens.* 10: 114-124.
- [16] Rahnamaei A., Zare Fatin G., Eskandarian A., (2019), High speed Radix-4 Booth scheme in CNTFET technology for high performance parallel multipliers. *Int. J. Nano Dimens.* 10: 281-290.
- [17] Yousefi R., Saghafi K., Moravvej-Farshi M. K., (2010), Application of neural space mapping for modeling ballistic carbon nanotube transistors. *Iranian J. Elec. Elec. Eng.* 6: 70-76.
- [18] Abdollahzadeh Badelbo R., Farokhi F., Kashaniniya A., (2013), Efficient parameters selection for CNTFET modelling using Artificial Neural networks. *Int. J. Smart Elec. Eng.* 2: 217-222.
- [19] Zhang J., Chang Sh., Wang H., He J., Huang O., (2014), Artificial Neural network based CNTFETs modeling. *Appl. Mech. Mater.* 667: 390-395.
- [20] Friesz A. K., Parker A. C., Zhou Ch., Philip Wong H.-S., Deng J., (2007), A Biomimetic carbon nanotube synapse circuit. *Biomed. Eng. Soc. (BMES) Annual Fall Meeting.*
- [21] Joshi J., Zhang J., Wang Ch., Hsu Ch.-Ch., Parker A. C., Zhou Ch., Ravishankar U., (2011), A biomimetic fabricated carbon nanotube synapse for prosthetic applications. *IEEE/NIH Life Sci. Sys. Appl. Workshop (LISSA).*
- [22] Najari M., El-Grouar T., Jelliti S., Mousa Hakami O., (2016), Simulation of a spiking neuron circuit using carbon nanotube transistors. *Proceed. Fifth Saudi Int. Meeting on Front. Phys. (SIMFP2016).*
- [23] Fathi A., Azizian S., Sharifan N., (2017), *Sensors and amplifiers: Sensor output signal amplification systems*, In Handbook of Research on Nanoelectronic Sensor Modeling and Applications, 423-504, IGI Global.