CMOS VLSI Hyperbolic Tangent Function & its Derivative Circuits for Neuron Implementation

Hussein CHIBLE, Ahmad Ghandour
Phd School of Sciences and Technology
Lebanese University - EDST
Beirut, Lebanon

Abstract- The hyperbolic tangent function and its derivative are key essential element in analog signal processing and especially in analog VLSI implementation of neuron of artificial neural networks. The main conditions of these types of circuits are the small silicon area, and the low power consumption. The objective of this paper is to study and design CMOS VLSI hyperbolic tangent function and its derivative circuit for neural network implementation. A circuit is designed and the results are presented.

Keywords – Non Linear function, derivative, Analog signal processing, neurons, CMOS VLSI implementation

I. INTRODUCTION

Neural Networks (NN) are particularly attractive for VLSI implementations as each parallel element (neuron or synapse) is relatively simple, allowing the complete integration of large networks on a single chip. Moreover, as noted by several authors, Neural Networks are most efficiently implemented by asynchronous analog circuits [1-5].

Analog implementations are generally faster (due to the asynchronous operation) and require less hardware (lower transistor count) than digital VLSI implementations. Analog VLSI Neural Networks are heavy parallel analog systems, which used and demonstrated in solving a wide range of real world problems [6].

[7] Presents a number of different implementations for the first derivative of the sigmoid function. The implementation of the sigmoid function employs a powers-of-two piecewise linear approximation. The best implementation scheme for the derivative is suggested based on overall speed performance (circuit speed and training time) and hardware requirements.

The CMOS circuit implementation of the feed forward neural primitives of a generic Multi Layer Perceptron network is presented in [8]. Basically the approach is based on current mode computation and is aimed at a low power/low voltage circuit implementation; moreover, it is easily scalable to implement network of any size. Experimental results are reported.

A new CMOS VLSI implementation of an asymmetric programmable sigmoid neural activation function, as well as of its derivative, is presented in [9]. It consists of two coupled PMOS and NMOS differential pairs with different programmable bias currents that set the upper and lower limits of the sigmoid. The circuit works in the weak inversion region, for low power consumption and exponential envelope, or in strong inversion to achieve higher speeds. The results obtained from the theoretical transfer function, and from the simulations of the circuit implemented in AMI’s 0.35mm technology, show a very good match.

[10] Presents a piecewise linear recursive approximation scheme is applied to the computation of the sigmoid function and its derivative in artificial neurons with learning capability. The scheme provides high approximation accuracy with very low memory requirements. The recursive nature of this method allows for the control of the rate accuracy/computation-delay just by modifying one parameter with no impact on the occupied area. The error analysis shows an accuracy comparable to or better than other reported piecewise linear approximation schemes. No multiplier is needed for a digital implementation of the sigmoid generator and only one memory word is required to store the parameter that optimizes the approximation.

In this paper, the CMOS VLSI implementation of the sigmoid function or the hyperbolic tangent function and the equivalent derivative are proposed and presented. Especially, this circuit can be used for the neuron module of the
multi layer perceptron implementation. In the next section the circuit proposed is presented and then results are presented. The characteristics of the circuit proposed in the small silicon area and the power consumptions.

II. THE PROPOSED HYPERBOLIC TANGENT FUNCTION AND ITS DERIVATIVE CIRCUITS

The hyperbolic tangent function and its derivative circuit are composed of three sub-circuits: 1) subcircuit1 based on one differential pair (See Figure 1); 2) subcircuit2 based on two differential pairs (See Figure 2); 3) subcircuit3 based on current mode multiplier (See Figure 3). The whole circuit is shown in Figure 4.

A. Subcircuit1 - Hyperbolic tangent Circuit

The hyperbolic tangent function is based on the famous differential pair and it is composed of 9 transistors $M_{n1}, M_{n2}, M_{n3}, M_{n4}, M_{n5}, M_{n6}, M_{n7}, M_{p}$, and $M_{n}$. The transistors $M_{n4}$ and $M_{n5}$ are biased in order to work in weak inversion, and then the output voltage is given by:

$$V_{no} = I_{1} \tanh \left( \frac{I_{syn} \times R_{MnMp}}{2nVt} \right)$$

Where, $I_{1}$ is based on $V_{b1}$ and $R_{MnMp}$ is a resistor based on the voltages $V_{n}$ and $V_{p}$.

Figure 1: subcircuit1-hyperbolic tangent function

B. Subcircuit2 - Two differential Pairs Circuit

This circuit is based on two differential pairs (see Figure2). It is used in order to create two outputs currents based on the output voltage of the subcircuit1. The circuit is composed of 12 transistors “$M_{1}, M_{2}, M_{3}, M_{4}, M_{5}, M_{6}, M_{7}, M_{8}, M_{10}, M_{11}, M_{12}$”. They must bias in weak inversion in order to get the following outputs:
Figure 2: subcircuit2 - Two differential pairs

\[ I_{o1} = I_1 \cdot \tanh \left( \frac{V_{no} - V_{ref1}}{2nVt} \right) \]

\[ I_{o2} = I_1 \cdot \tanh \left( \frac{V_{ref2} - V_{no}}{2nVt} \right) \]

If \( V_{no} - V_{ref1} \ll 2nVt \) \( \Rightarrow \) \( I_{o1} = a(V_{no} - V_{ref1}) \)

Where \( a = \frac{I}{2nVt} \)

The same for \( I_{o2} \) which is given by: \( I_{o2} = a(V_{ref2} - V_{no}) \)

C. Subcircuit3 – Current Mode Multiplier Circuit

The other part of the circuit is the four transistor multiplier. The current \( I_{o3} \) is given by [11]:

[Diagram of subcircuit3 - Current Mode Multiplier Circuit]
Figure 3: subcircui3-current mode multiplier.

\[ I_{\text{der}} = \frac{R_{Mn-Mp}}{2nVt} (I_{o1} \ast I_{o2}) \]

D. Merge Subcircuit2 & Subcircuit3 together

By substituting \( I_{o1} \) and \( I_{o2} \), \( I_{\text{der}} \) is given by:

\[ I_{\text{der}} = \frac{R_{Mn-Mp}}{2nVt} \left( a^2 (V_{no} - V_{ref}) (V_{ref2} - V_{no}) \right) \]

\[ \Rightarrow \]

\[ I_{\text{der}} = \frac{R_{Mn-Mp}}{2nVt} \left( \frac{I_1}{2nVt} \right)^2 (V_{no} - V_{ref}) (V_{ref2} - V_{no}) \]

\[ \Rightarrow \]

\[ I_{\text{der}} = \frac{R_{Mn-Mp}}{2nVt} \left( \frac{I_1}{2nVt} \right)^2 \left( (V_{no} - V_{ref}) - (V_{ref2} - V_{no}) \right) \]

Where \( V_{ref1} = V_{ref} + 0.1V \) & \( V_{ref2} = V_{ref} + 0.1V \), then the above current becomes:

\[ I_{\text{der}} = \frac{R_{Mn-Mp}}{2nVt} \left( \frac{I_1}{2nVt} \right)^2 \left( (V_{no} - V_{ref}) + 0.1V \right) \left( 0.1V - (V_{no} - V_{ref}) \right) \]

\[ \Rightarrow \]
\[ I_{der} = \frac{R_{Mn-Mp}}{2nVt} \left( \frac{I_1}{2nVt} \right)^2 \left( 0.1V + \left( \frac{V_{no} - V_{ref}}{0.1V} \right) \left( 0.1V - \left( \frac{V_{no} - V_{ref}}{0.1V} \right) \right) \right) \]

\[ \Rightarrow \quad I_{der} = \frac{R_{Mn-Mp} (0.1V \times I_1)^2}{(2nVt)^2} \left( 1 - \left( \frac{V_{no} - V_{ref}}{0.1V} \right)^2 \right) \]

**E. Merge Sub-circuits 1, 2, and 3 together**

Figure 4 shows the merging of Subcircuit1, Subcircuit2, and Subcircuit3 together, in order to make the circuit produce the hyperbolic tangent function and its derivative at the same time.

By substituting \( V_{no} \) in \( I_{der} \), we get:

\[ I_{der} = \frac{R_{Mn-Mp} (0.1V \times I_1)^2}{(2nVt)^2} \left( 1 - \left( \frac{I_{syn} \times R_{Mn-Mp}}{2nVt} - \frac{V_{ref}}{0.1V} \right)^2 \right) \]

---

**III. SIMULATION RESULTS**

The circuit discussed above have been designed and simulated by using WinSpice “Wspice3 Simulator for Windows” and by using the Parameters of the technology 0.35µm, which is used for analog implementation. The dimensions “width and the length” of the MOS transistors have been computed on the base of the technology parameters (e.g. the mobility of the electron, oxide capacitor, etc.). Also, the dimensions of the transistors are calculated and designed based on making the differential pairs work in weak inversion region.

**Sub-circuit 1 simulation**
Figure 5: $V_{on}$ versus $V_{in}(I_{syn})$ - Hyperbolic tangent function.

Sub-circuit 2 simulation

Figure 6 shows the DC transfer characteristics $I_{o1}$ versus $V_{in}[1V:2.3V]$ with steps 0.01 and $V_{b2}$ as a parameter varies in [0:3.3] with steps 0.3V, where $V_{ref} = V_{ref1}$.

Figure 6: DC transfer characteristics $I_{o1}$ versus $V_{in}$.

Figure 7 shows the DC transfer characteristics $I_{o2}$ versus $V_{in}[1V:2.3V]$ with steps 0.01 and $V_{b2}$ as a parameter varies in [0:3.3] with steps 0.3V, where $V_{ref} = V_{ref2}$.
The derivative will be the multiplication between the two currents in $I_{o1}$ (Figure 6) and $I_{o2}$ (Figure 7) as a function of $(V_{ref1} - V_{no})$ and $(V_{no} - V_{ref2})$.

**Sub-circuit 3 simulation**

Figure 8 shows the DC transfer characteristics between the output current $I_{out}$ and the input current $I_{o1}$ [-10u:10u] with step 0.1u and with $I_{o2}$ as a parameter varies in the range [-10u:10u] with step 0.5u.

Figure 9 shows the DC transfer characteristics between the output current $I_{out}$ and $I_{o2}$[-10u:10u] with step 0.1u and with $I_{o1}$ as a parameter varies in the range [-10u:10u] with step 0.5u.
Figure 9: DC transfer characteristics $I_{out}$ versus $I_{o1}$ [-10u:10u].

Figure 10 shows the DC transfer characteristics between the output current $I_{out}$ and $I_{o2}$[-1u:1u] with step 0.1u and with $I_{o3}$ as a parameter varies in the range [-10u:10u] with step 0.5u.

The whole proposed circuit simulation

The multiplication of Figure 6 & Figure 7 by using the current mode multiplier shown in Figure 3 is shown in Figure 11 (weak inversion), which shows clearly the derivative behavior of the hyperbolic tangent function.
IV. CONCLUSIONS

In this research, a circuit for hyperbolic tangent non linear function and its derivative has been presented. The circuit can be used especially in neural networks applications and also in other signal processing operations. The circuit can be divided into three sub-circuits. The low power consumption & the small silicon area are the main characteristics. The future work will be based on building a neural network based on the proposed circuits.

ACKNOWLEDGMENT

Thanks go to the Lebanese University “Funding Scientific Research Program” that funded this work and research.

REFERENCES