IOSR Journal of Engineering (IOSRJEN)
ISSN: 2250-3021 Volume 2, Issue 6 (June 2012), PP 1352-1356
www.iosrjen.org

Efficient FPGA Implementation of Sigmoid and Bipolar Sigmoid Activation Functions for Multilayer Perceptrons

Manish Panicker[1], C.Babu[2]
(Department of Electronics and Communication, Amrita School of Engineering, Bangalore)

ABSTRACT : - *Artificial Neural Networks base their processing capabilities in a parallel architecture. This makes them extremely useful in pattern recognition, system identification and control problems. Multilayer Perceptron is an artificial neural network with one or more hidden layers. The Activation function determines the performance of a Multilayer Perceptron. In Multi Layer Perceptron, the most commonly used activation functions are sigmoid and bipolar sigmoid activation functions. In this paper we present a FPGA based digital hardware implementation of Sigmoid and Bipolar Sigmoid Activation function. The digital hardware was designed for 32 bit fixed point arithmetic and was modeled using Verilog HDL. The synthesis tool used was Xilinx 10.1 ISE and the design was implemented in Spartan 3 FPGA.*

Keywords *- FPGA, PLAN approximation, Multi Layer Perceptron, Neuron, Sigmoid Activation.*

I. INTRODUCTION

The human brain is probably the most complex and intelligent system in the world. It consists of the basic structural constituents known as neurons. Each neuron has a set of simple operations but in a network they exhibit complex global behavior. Artificial neural network (ANN) is an engineering approach to imitate the brain's activities. An ANN is configured through a learning process for a specific application, such as pattern recognition or data classification, and as in all biological systems, this learning process will require the adjustment of the synaptic connections between the neurons [1]. Multilayer Perceptron is an artificial neural network having one or more hidden layers.

Multilayer Perceptrons use sigmoid and bipolar sigmoid function as their activation function. Sigmoid and Bipolar sigmoid activation functions are normally used for gradient descent type algorithms [1].

Log sigmoid function $f(x) = 1 / (1+e^{-\zeta * x})$
Bipolar sigmoid function $f(x) = (1- e^{-\zeta * x}) / (1+e^{-\zeta * x})$

where parameter ζ determines slope in transition region.

The reason for the popularity of these activation functions is because they are easily differentiable, which is an important aspect in the back propagation type training algorithms [2]. Special attention must be paid to an efficient approximation of the sigmoid function in implementing FPGA-based reprogrammable hardware-based Artificial Neural Networks.

Hardware implementation of ANN to utilize the parallelism can follow analog, digital or mixed signal design technique. Matured and flexible digital design in Very large Scale Integration (VLSI) can be implemented on Application Specific Integrated Circuits (ASICs) or Field-Programmable Gate Arrays (FPGAs). ASIC design has some drawbacks like the ability to run only specific algorithm and limitations on the size of a network. Hence FPGA offers a suitable alternative that has flexibility with an appreciable performance. It maintains high processing density, which is needed to utilize the parallel computation in an ANN. Every digital module is concurrently instantiated on the FPGA and hence, operates in parallel. Thus the speed of the network is independent of the complexity [3].

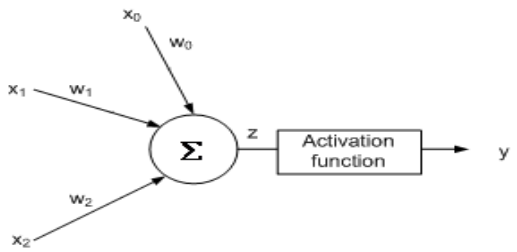

Fig 1 A simple Artificial Neuron Model

Fig 1 represents a basic simple artificial neuron model. The inputs to the neuron are x0, x1, x2 and the w0, w1, w2 are the corresponding weight values. The weight values are multiplied with their corresponding inputs and summed together. This is the input to the activation function block. The paper presents the FPGA implementation of Digital hardware for the efficient implementation of Sigmoid and Bipolar Sigmoid activation functions. The hardware for implementing the function is modelled using Verilog HDL. Piecewise linear approximation is used to approximate sigmoid activation function and second order approximation is used to approximate bipolar sigmoid activation function [3].

II. DATA REPRESENTATION

The hardware is designed for 32 bit fixed point arithmetic. It can therefore be used for more complex FPGA based neural networks. The fixed point arithmetic representation consisted of 1 sign bit, 15 integer bits and 16 fraction bits as shown in Fig 2. Using 32 bit fixed point arithmetic provides accuracy similar to its resource hungry floating point counterpart [4]. In both the hardware input and output is of 32bit fixed point.

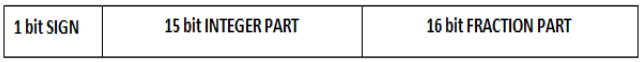

Fig 2 Fixed point data Representation of Input and Output

III. PLAN APPROXIMATION OF SIGMOID FUNCTION

A LUT can be used to implement the sigmoid activation function by means of discrete value. But it consumes large area and time, which may affect the speed of computation. On-chip realization of log-sigmoid function increases the size of hardware considerably. To optimize the area to some extent, the inbuilt RAM available in FPGAs can be used to realize LUT based activation function. It reduces area and improves speed [5]. But if the precision is to be improved, then hardware friendly PWL and other approximation approaches of activation function is to be utilized.

Piecewise linear (PWL) approximation is a method to obtain low values for both maximum and average error with low computational complexity [5].

Table 1 PLAN approximation of Sigmoid activation function

OPERATION	CONDITION
Y=1	**$\|X\| \geq 5$**
$Y = 0.03125 \cdot \|X\| + 0.84375$	**$2.375 \leq \|X\| < 5$**
$Y = 0.125 \cdot \|X\| + 0.625$	**$1 \leq \|X\| < 2.375$**
$Y = 0.25 \cdot \|X\| + 0.5$	**$0 \leq \|X\| < 1$**

The PLAN approximation (piecewise linear approximation of a nonlinear function) was proposed by Amin, Curtis and Hayes–Gill [5]. The PLAN approximation uses digital gates to directly transform from x to y. The approximation of the sigmoid function is presented in TABLE 1. The calculations need only be performed on the absolute value of the input x. After simplifying the shift and addition operations implicit in TABLE 1, the bit-level logic equations become effective to implement.

IV. SECOND ORDER APPROXIMATION OF BIPOLAR SIGMOID FUNCTION

A simple second-order piecewise nonlinear function which has a hyperbolic tan like transition between the lower and upper saturation regions is given by

$$H(z) = \begin{cases} z(\beta - \theta z) & for \;\; 0 \leq z \leq L \\ z(\beta + \theta z) & for \;\; -L \leq z \leq 0 \end{cases} \quad ; \; (1)$$

where β and θ determine the slope and gain of the nonlinear function between $-L \leq z \leq L$. L is the saturation point of the function [6].

Hence a sigmoid like Bipolar function can be realized by

$$G(z) = \begin{cases} 1 & for\ L \leq z \\ H(z) & for\ -L < z < L \\ -1 & for\ z \geq L \end{cases} \quad ; (2)$$

and $\beta = 2 / L$, $\theta = 1 / L^2$; (3)

The function can be implemented using a binary shifter, fixed point adder and multilplier as the inputs are 32 bit fixed point as shown in Fig 3. Bipolar Sigmoid function provides better convergence speed when compared to sigmoid activation function [6]. The input and output of the digital hardware is represented in 32bit fixed point arithmetic.

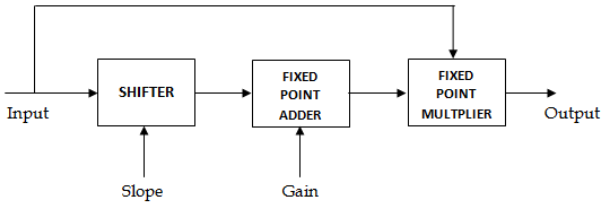

Fig 3 Digital Hardware Implementation of Approximated Bipolar Sigmoid function

V. RESULTS

The digital hardwares were modeled using Verilog HDL and simulated using MODEL SIM 6.5 Xilinx 10.1 ISE was used as synthesis tool for implementing the designs in Spartan 3 FPGA.

5.1 PLAN approximated Sigmoid Activation Function

The Digital hardware for the realization of PLAN approximated Sigmoid Activation function was designed. Fig 4 shows the Modelsim simulation results and TABLE 2 consists of the device utilization summary of the design.

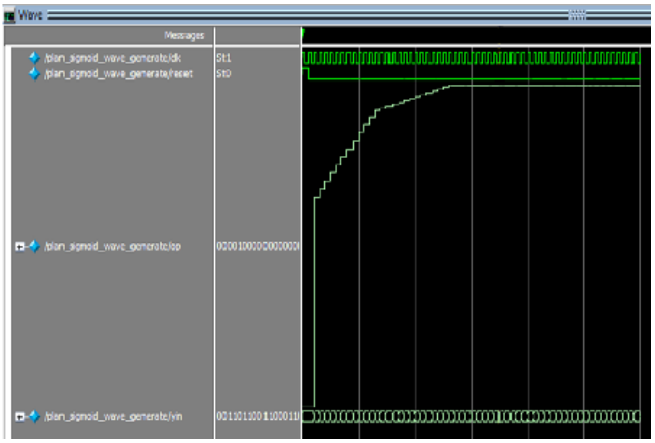

Fig 4 Modelsim Simulation of PLAN approximated Sigmoid Activation Function with ζ=1

Table 2 Device Utilization Summary of PLAN approximated Sigmoid activation function

	Used	Available	Utilization
Number of slices	142	4656	3%
Number of 4 input LUT	227	9312	2%
Number of IOBS	66	232	28%

5.2 Approximated Bipolar Sigmoid Activation function

The Second order Approximated Bipolar Sigmoid function was implemented. The obtained values were compared with the actual function values and plotted. Fig 5 shows the comparison between the actual bipolar sigmoid values and the values obtained from the designed hardware.

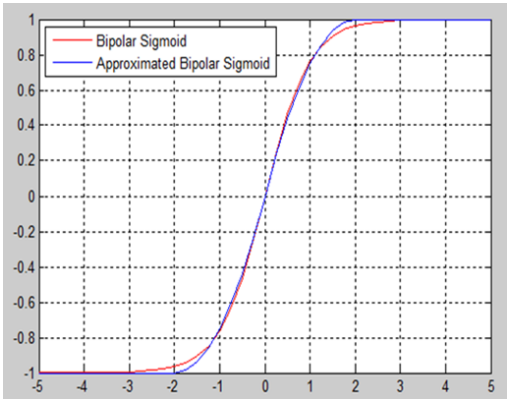

Fig 5 Comparison of actual and hardware computed value of Bipolar Sigmoid function with $\zeta = 2$

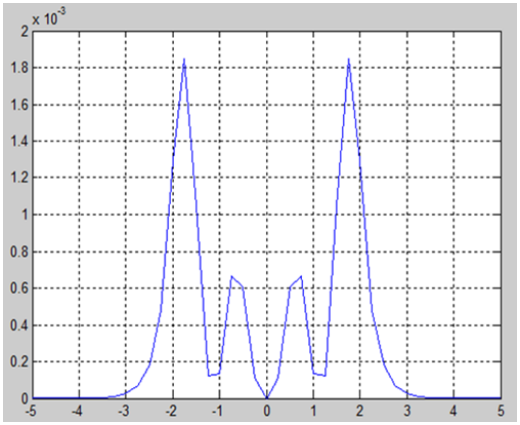

Fig 6 Mean Square Error Plot

The design was implemented in Spartan 3 FPGA. TABLE 3 shows the design utilization summary of the hardware for second order approximation of bipolar sigmoid activation function. It can be noticed from the error plots and the device utilization summary that the design is area efficient and it is also accurate. The maximum mean square error value was found to be 0.00187 as shown in the error plot in Fig 6.

Table 3 Device Utilization Summary of Second order Approximation of Bipolar Sigmoid activation function

	Used	Available	Utilization
Number of slices	112	4656	2%
Number of 4 input LUT	209	9312	2%
Number of IOBs	65	232	28%

CONCLUSION

The digital hardware for implementing sigmoid and bipolar sigmoid activation function was modeled using Verilog HDL. The design was synthesized using Xilinx 10.1 ISE tool and implemented in Spartan 3 FPGA. The Implemented hardware design was area efficient and was highly accurate. A comparison of

hardware obtained values and actual activation function values were done, the maximum mean square error value was found to be 0.00187 for bipolar sigmoid activation. The implemented design can be used in complex Multilayer perceptron networks that use fixed point arithmetic.

REFERENCES

Books

[1] R.Omondi, C. Rajapakse," *FPGA Implementation of Neural Networks", Springer* U.S., 2006, ISBN 10-0-387-28485-0

Journal Papers

[2] Hiroomi Hikawa , "A Digital Hardware Pulse-Model Neuron With Piecewise Linear Activation Function," *IEEE Trans.Neural Networks, vol.14,no.5*,pp.1028-1037,Sept.2003

[3] M. Banuelos Sauced, et. al.,"Implementation of neuron model using FPGAs". *Journal of Applied Research and Technology ISSN: 1665-6423, vol. 3 (10), 2003, pp. 248-255.*

[4] Antony W. Savich, Medhat Moussa, and Shawki Areibi, "The Impact of Arithmetic Representation on Implementing MLP-BP on FPGAs: A Study," *IEEE Trans. Neural Networks, vol.18, no.1*, Jan.2003.

[6] Kwan, H.K.: 'Simple sigmoid-like activation function suitable for digital hardware implementation', *Electronic. Letters., 1992*, 28, (15), pp. 1379–1380

Proceeding papers

[5] Amin, H., Curtis, K.M., and Hayes–Gill, B.R.: 'Piecewise linear approximation applied to nonlinear function of a neural network', *IEE Proc. Circuits, Devices Sys.*, 1997, 144, pp. 313–317

`