Asian Journal of Applied Science and Technology (AJAST)
(Open Access Quarterly International Journal) Volume 2, Issue 2, Pages 345-350, 2018

Low Power Gate Level Pruning for Activation Functions

M.R.Archana[1] and Mr.S.Vellaisamy[2]

[1]*PG Scholar, Department of ECE, SCAD College of Engineering and Technology, Cheramahadevi, Tamilnadu, India.*
[2]*Assistant Professor, Department of ECE, SCAD College of Engineering and Technology, Cheramahadevi, Tamilnadu, India.*

Article Received: 27 January 2018 *Article Accepted: 23 February 2018* *Article Published: 15 April 2018*

ABSTRACT

Inexact and approximate circuit design is a promising approach to improve performance and energy efficiency in technology-scaled and low-power digital systems. Such strategy is suitable for error-tolerant applications involving perceptive or statistical outputs. Probabilistic pruning technique has been proposed for an efficient approximate tangent function. The approximation is based on a mathematical analysis considering the maximum allowable significance-activity product (SAP) and unessential nodes. Implementation of the proposed approximation schemes presented, which shows that the proposed structure compares favorably with previous architectures in terms of area and delay.

1. INTRODUCTION

Neural networks are complex non-linear models, built from components that individually behave similarly to a regression model. They can be visualized as graph and some sub-graphs may exist with behavior similar to that of logic gates. Although the structure of a neural network is explicitly designed beforehand, the processing that the network does in order to produce a hypothesis (and therefore, the various logic gates and other processing structures within the network) evolves during the learning process. This allows a neural network to be used as a solver that "programs itself", in contrast to typical algorithms that must be designed and coded explicitly. Evaluating the hypothesis defined by a neural network may be achieved via feed-forward, which amounts to setting the input nodes, then propagating the values through the connections in the network until all output nodes have been calculated completely. The learning can be accomplished by using gradient descent, where the error in the output nodes is pushed back through the network via back-propagation, in order to estimate the error in the hidden nodes, which allows calculation of the gradient of the cost-function.

2. LITERATURE SURVEY

Suresh Cheemalavagu, Pinar Korkmaz, Krishna V. Palem, Bilge E. S. Akgul and Lakshmi Chakrapani present a Probabilistic CMOS Switch and its Realization by Exploiting Noise. By viewing noise as a resource rather than as an impediment, we demonstrate an entirely novel approach to ultra low-energy computing. The subject of this study is the probabilistic inverter, ubiquitous to the design of digital systems, whose behavior is rendered probabilistic by noise.

Krishna V. Palem presents the mathematical technique of randomization yielding probabilistic algorithms is shown, for the first time, through a physical interpretation based on statistical thermodynamics, to be a basis for energy savings in computing. Concretely, at the fundamental limit, it is shown that the energy needed to compute a single probabilistic bit or PBIT is proportional to the probability of computing a PBIT accurately. This result is established through the introduction of an idealized switch, for computing a PBIT, using which a network of switches can be constructed.

Interesting examples of such networks including AND, OR and NOT gates (or as functions, Boolean conjunction, disjunction and negation respectively), are constructed and the potential for energy savings through randomization is established. To quantify these savings, novel measures of "technology independent" energy complexity are introduced—these parallel conventional machine-independent measures of computational complexity such as the algorithm's running time. Networks of switches can be shown to be equivalent to Turing machines and to Boolean circuits, both of which are widely-known and well-understood models of computation. These savings are realized using a novel way of representing a PBIT in the physical domain through a group of classical microstates. A measurement and thus detection of a microstate yields the value of the PBIT. While the eventual goal of this work is to lead to the physical realization of these theoretical constructs through the innovation of randomized (CMOS based) devices, the current goal is to rigorously establish the potential for energy savings through probabilistic computing at a fundamental physical level, based on the canonical thermodynamic models of idealized mono atomic gases developed by Boltzmann, Gibbs and Planck.

3. PROPOSED SYSTEM

In this project, an efficient approximation scheme for hyperbolic tangent function is proposed. The approximation is based on a mathematical analysis considering the maximum allowable error as design parameter. Hardware implementation of the proposed approximation schemes presented, which shows that the proposed structure compares favorably with previous architectures in terms of area and delay. The proposed structure requires less output bits for the same maximum allowable error when compared to the state-of-the-art. The number of output bits of the activation function determines the bit width of multipliers and adders in the network. Therefore, the proposed activation function results in reduction in area, delay, and power in VLSI implementation of artificial neural networks with hyperbolic tangent activation function. The above block is composed of three main blocks to approximate the hyperbolic tangent function in all three regions, including saturation, processing, and pass region. General arithmetic operations in each region can be described as follows.

1. Pass Region: In this region, fractional part of input is passed to the output. Based, a shift to left by N out−N f bits before passing the input to output is required.
2. Processing Region: For inputs in the processing region, a bit-level input mapping is required. The number of bit-level mapping blocks required is equal to the number of input ranges in this region. For each input range in the processing region, log N2 bits after none bit of input should be mapped to output bits using the bit-level mapping. Using logN2 bits after none bit covers all sub-ranges. The number of sub-ranges (N) is calculated while the output of each sub range is found. The bit-level mapping can be implemented using a combinational circuit.
3. Saturation Region Approximation: In this region, hyperbolic tangent function is approximated by the maximum value re presentable by output bits, and can be realized by setting all output bits to one.

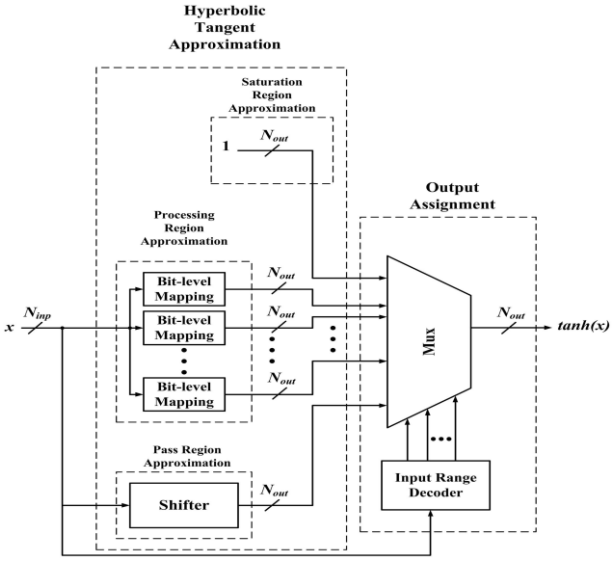

Fig.3.1.Block diagram of proposed structure

For a specific maximum allowable error, the proposed structure requires less number of output bits compared to the previously developed architectures. Therefore, bit width of multipliers and adders in the hidden layers of the network using proposed structure as its activation function is lower. Multipliers and adders with lower bit width have lower area, delay, and power consumption. To evaluate the efficiency of proposed structure, it is used to implement a 4-3-2 network for an optical template matching block diagram is shown in Fig. It is capable of recognizing six different input patterns and classifying them as four different classes. The general neural network Based on the proposed approximation scheme, hybrid architecture for hardware implementation of hyperbolic tangent activation function was presented. The synthesis results showed that the proposed structure compares favorably to the previously developed architectures in terms of area, delay, and area× delay.

4. RESULTS AND DISCUSSION

Xilinx Co-Founders, Ross Freeman and Bernard Vonder schmitt, invented the first commercially viable field programmable gate array in 1985 – the XC2064. The XC2064 boasted a mere 64 configurable logic blocks (CLBs), with two 3-input lookup tables (LUTs). Xilinx continued unchallenged and quickly growing from 1985 to the mid-1990s, when competitors sprouted up, eroding significant market-share. By 1993, Actel was serving about 18 percent of the market.

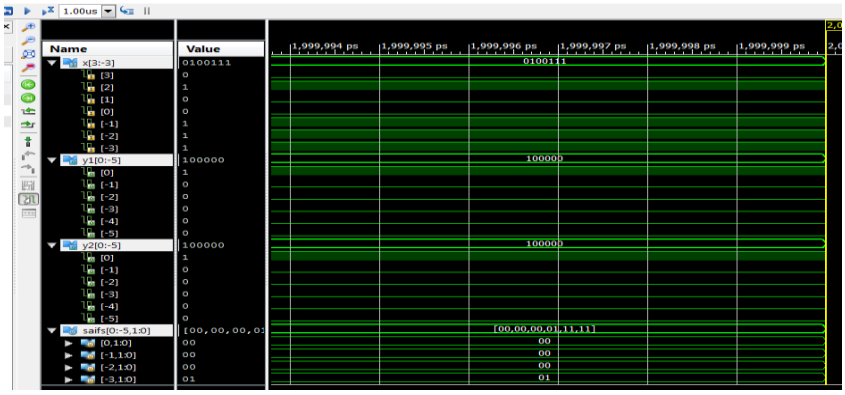

Fig.4.1. Activation function Calculation

RTL Schematic

After performing the synthesize process, the RTL schematic has been created automatically based on the functionality. The routing between the different cells can be viewed clearly by this schematic. This tutorial deals with the Xilinx ISE-synthesis tool. I have assumed that you know how to start a project in Xilinx ISE and run a simulation etc... Once you have completed simulation for your design successfully you want to test it in hardware. The first step to this is synthesizing your code. The synthesis of your VHDL code is done by XST (Xilinx Synthesis Technology) tool, which is included in Xilinx ISE software. XST creates Xilinx-specific net-list files called NGC files. Remember that NGC files are not always same, even for the same VHDL code. Depending upon the family and device you have chosen you will get different NGC files. Each NGC file has two parts: logical design data and constraints.

Let us learn how to synthesis a code and infer the reports generated by the software. Copy the 'counter' program from here. This is a 4 bit counter with reset input. Add this code to your Xilinx ISE project and click on "Synthesis-XST". This will start the Synthesis process. Once the synthesis is done without any errors, you can get a lot of details from the files generated by synthesis process.

5. CONCLUSION

A new approximation scheme for hyperbolic tangent was proposed in this project. The proposed approximation schemes based on a mathematical analysis considering maximum allowable error as a design parameter. Based on the proposed approximation scheme, hybrid architecture for hardware implementation of hyperbolic tangent activation function was presented. The synthesis results showed that the proposed structure compares favorably to the previously developed architectures in terms of area, delay, and area× delay. The proposed structure required less output bits for the same maximum allowable error compared to the previously developed architectures.

REFERENCES

[1] K. V. Palem, “Energy aware computing through probabilistic switching: A study of limits,” IEEE Trans. Comput., vol. 54, no. 9, pp. 1123–1137, Sep. 2005.

[2] Muthukumaran. N and Ravi. R, 'Hardware Implementation of Architecture Techniques for Fast Efficient loss less Image Compression System', Wireless Personal Communications, Volume. 90, No. 3, pp. 1291-1315, October 2016, SPRINGER.

[3] Muthukumaran. N and Ravi. R, 'The Performance Analysis of Fast Efficient Lossless Satellite Image Compression and Decompression for Wavelet Based Algorithm', Wireless Personal Communications, Volume. 81, No. 2, pp. 839-859, March 2015, SPRINGER.

[4] Muthukumaran. N and Ravi. R, 'VLSI Implementations of Compressive Image Acquisition using Block Based Compression Algorithm', The International Arab Journal of Information Technology, vol. 12, no. 4, pp. 333-339, July 2015.

[5] Muthukumaran. N and Ravi. R, 'Simulation Based VLSI Implementation of Fast Efficient Lossless Image Compression System using Simplified Adjusted Binary Code & Golumb Rice Code', World Academy of Science, Engineering and Technology, Volume. 8, No. 9, pp.1603-1606, 2014.

[6] Ruban Kingston. M,Muthukumaran. and N, Ravi. R, 'A Novel Scheme of CMOS VCO Design with reduce number of Transistors using 180nm CAD Tool', International Journal of Applied Engineering Research, Volume. 10, No. 14, pp. 11934-11938, 2015.

[7] Muthukumaran. N and Ravi. R, 'Design and analysis of VLSI based FELICS Algorithm for lossless Image Compression', International Journal of Advanced Research in Technology, Vol. 2, No. 3, pp. 115-119, March 2012.

[8] Manoj Kumar. B and Muthukumaran. N, 'Design of Low power high Speed CASCADED Double Tail Comparator', International Journal of Advanced Research in Biology Engineering Science and Technology, Vol. 2, No. 4, pp.18-22, June 2016.

[9] N. Muthukumaran, 'Analyzing Throughput of MANET with Reduced Packet Loss', Wireless Personal Communications, Vol. 97, No. 1, pp. 565-578, November 2017, SPRINGER.

[10] P.Venkateswari, E.Jebitha Steffy, Dr. N. Muthukumaran, 'License Plate cognizance by Ocular Character Perception', International Research Journal of Engineering and Technology, Vol. 5, No. 2, pp. 536-542, February 2018.

[11] S. Cheemalavagu, P. Korkmaz, K. V. Palem, B. E. Akgul, and L. N. Chakrapani, "A probabilistic CMOS switch and its realization by exploiting noise," in Proc. IFIP Int. Conf. VLSI, Oct. 2005, pp. 535–541.

[12] P. Korkmaz, B. E. S. Akgul, K. V. Palem, and L. N. Chakrapani, "Advocating noise as an agent for ultra-low energy computing: Probabilistic complementary metal–oxide–semiconductor devices and their characteristics," Jpn. J. Appl. Phys., vol. 45, no. 4B, p. 3307, 2006.

[13] G. Karakonstantis and K. Roy, "Voltage over-scaling: A cross-layer design erspective for energy efficient systems," in Proc. 20th Eur. Conf. Circuit Theory Design (ECCTD), Aug. 2011, pp. 548–551.

[14] N. Muthukumaran, Mrs R.Sonya, Dr.Rajashekhara and Chitra V, 'Computation of Optimum ATC Using Generator Participation Factor in Deregulated System', International Journal of Advanced Research Trends in Engineering and Technology, Vol. 4, No. 1, pp. 8-11, January 2017.

[15] Keziah. J, Muthukumaran. N, 'Design of K Band Transmitting Antenna for Harbor Surveillance Radar Application', International Journal on Applications in Electrical and Electronics Engineering, Vol. 2, No. 5, pp. 16-20, May 2016.

[16] Akhil. M.S and Muthukumaran. N, 'Design of Optimizing Adders for Low Power Digital Signal Processing', International Journal of Engineering Research and Applications, Vol. 5, pp. 59-65, March 2014.

[17] Muthukumaran. N and Ravi. R, 'Quad Tree Decomposition based Analysis of Compressed Image Data Communication for Lossy and Lossless using WSN', World Academy of Science, Engineering and Technology, Volume. 8, No. 9, pp. 1543-1549, 2014.

[18] Marvin Mark. M and Muthukumaran. N, 'High Throughput in MANET using relay algorithm and rebroadcast probability', International Journal of Engineering Research and Applications, Vol. 5, pp. 66-71, March 2014.

[19] Marvin Mark. M and Muthukumaran. N, 'An Advanced Homogeneous Pattern for Mobile Ad Hoc Network', International Journal of Advanced Research in Computer Science and Management Studies, Vol. 2, Issue 4, pp. 6-8, April 2014.

[20] D. Ernst et al., "Razor: A low-power pipeline based on circuit level timing speculation," in Proc. 36th Annu. IEEE/ACM Int. Symp. Microarchitecture (MICRO), Dec. 2003, pp. 7–18.

[21] P. K. Krause and I. Polian, "Adaptive voltage over-scaling for resilient applications," in Proc. Design, Autom. Test Europe Conf. Exhibit. (DATE), Mar. 2011, pp. 1–6.

[22] S. Ghosh, S. Bhunia, and K. Roy, "CRISTA: A new paradigm for low power, variation-tolerant, and adaptive circuit synthesis using critical path isolation," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 26, no. 11, pp. 1947–1956, Nov. 2007.

[23] V. Gupta, D. Mohapatra, A. Raghunathan, and K. Roy, "Low-power digital signal processing using approximate adders," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 32, no. 1, pp. 124–137, Jan. 2013.