

Low Power Adder based ANN

S N Prasad

Research Scholar, Jain University,
Associate Professor, Dept. of ECE,
Reva Institute of Technology & Management,
Bangalore

S.Y. Kulkarni

Principal
M.S. Ramaiah Institute of Technology,
Bangalore

ABSTRACT

This paper presents an overview of datapath realizations of the Hardware neural network models which perform massive parallel operations for best results and real time applications. Digital implemented neural models processing element – adder with low power consumption is proposed for real-time multimedia applications. Proposed adder is illustrated in the 2-3-1 tree layer artificial neural network (ANN). Designs were modeled with Verilog HDL and implemented in FPGA domain by targeting the Virtex 7 device.

Keywords

ANN, Low Power Adder, Datapath, Verilog HDL, VLSI

1. INTRODUCTION

One of the important creations of this modern technology is computational intelligence with sound biological understanding to model uncertainty, imprecision, evolutionary behavior and complex models for multimedia processing applications. Computational intelligence with the advantage of neural networks enables complementary features to develop synergistic systems which cover techniques for object-based recognition & coding, pattern detection & recognition, conversion & synchronization, content based indexing and etc [1].

ANN being one of the modern technologies enhances the computational abilities of real-time processing systems, due to its structure with large number of connections. It also provides greater ability to deliver simple and powerful solutions in the areas that challenged the conventional computing approaches like pattern recognition, image processing and speech synthesis. Similar to high performance computation on large information processing systems, the demand for power is also increasing. Remarkable advancement in the VLSI technology facilitates large amount of functional features on electronic devices [2].

The biological neural network provides strong performance in tasks such as vision, speech, image and etc than the digital computer. It is observed that the visual system of a human does more image processing than the world's supply of super-computers [3]. Unlike the conventional processors, ANN contains large collection of processing elements (PEs), which in turn have adders, multipliers and memory cache to process and carry the information from one element to another. Hardware implementation of ANNs facilitates high functional capabilities to meet the desired constraints of real-world problems. Advancement in VLSI technology enables enough flexibility in achieving the desired goals for a given application [2].

Several approaches like variable supply voltage, clock gating and some algorithmic optimizations were made in the past to mitigate the large power consumption. But due to scaling of technology, approaches have been saturated and there is a need to develop the architectures specific to application by adopting the optimizations/constraint aware datapath architectures at the lowest hierarchical level in the design cycle. Several architectures and designs of 4-2, 5-3 low power compressors

capable of performing at low voltages were presented in [4]. Importance of optimized Low power Adder at the systolic array based digital filter level for QRS detector was illustrated in [5]. In [6], author has demonstrated the value of datapath optimizations by exploiting the benefits of pre-computed summations in optimized LUTs using efficient compressors to improve the performance of addition. Low power digital filter using low power adders and multipliers was implemented in [7]. In this brief, we have attempted such an approach to reduce the power consumption of the design. Here low power adder is proposed for the 3 layered neural network models. This illustration reviews the importance of datapath architectural optimizations made at the lowest hierarchical level of the design. Such an approach can be applied at any levels of design abstraction - provided the architectures specific to the constraint. The other sections of the paper are arranged as per description. Section 2 briefs the mathematical modeling of ANN. Layered neural network architecture its datapath is analyzed in section in 3. Section 4 evaluates and analyzes the results of the 3-layered ANN implementation. Paper is concluded in section 5 and references are provided in last section.

2. ANN

ANN concept is emerged from human brain principles and made adapted to digital computers. First formal model of the neurons were proposed by McCulloch and Pits using the mathematical rule [8]. Proposed neural model allows '0' and '1' states & assumes that they are operating synchronously under discrete time mode. Weights and thresholds of the neurons are fixed and a generic model of such neuron is shown in Figure 1.

Neuron model shown in Figure 1 consists of a processing element with several synaptic connections associated with certain weight. As shown in Figure 1, the signal flow inputs and outputs are unidirectional. The output of the neuron model is decided the expression (1).

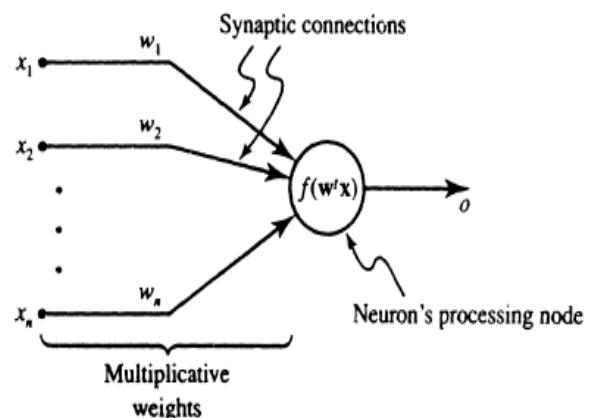


Figure 1: Generic artificial neuron [2]

$$o = f(w^t x) \text{ or } o = f(\sum_{i=1}^n w_i x_i) \quad (1)$$

Where 'w' is the weight factor and is defined as in expression (2) and 'x' is the input vector given as in expression (3).

$$w \triangleq [w_1 \ w_2 \ \dots \dots \ w_n]^t \quad (2)$$

$$x \triangleq [x_1 \ x_2 \ \dots \dots \ x_n]^t \quad (3)$$

'f' in expression (1) is the activation function or the PE function and the threshold value is explicitly mentioned but can be assumed that neuron has (n-1) connections coming from the actual inputs 'x1, x2, ..., xn'. Inspired by the natural neural network principles, many artificial neural networks can be constructed using mathematically modelled artificial neuron of Figure 1. To improve the performance of ANN's several connection methods were explored, which are shown in Figure 2. The most basic model is the single layer feed-forward network and its cascaded connection is called as multi-layer feed-forward network where the output of a layer is the input of following layer. In feed-forward networks the outputs are compared with desired output values and an error signal is adapted for network's weights. Third type of connection is the recurrent network where the neurons outputs are connected back to their inputs [2]. In this brief, a multi-layer feed forward network is illustrated.

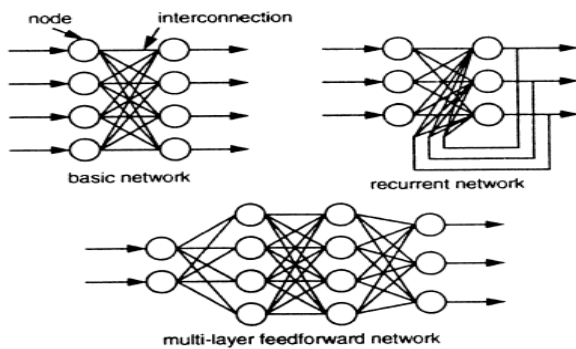


Figure 2: ANN connection models

3. MULTI-LAYER ANN ARCHITECTUE

This section represents the ANN's architecture, working and its datapath. Fully parallel feed-forward ANN architecture is illustrated to review the importance of datapath architectures. Figure 3 shows the 3 layered ANN where number of multipliers per neuron will be equal to number of connections to this neuron and the number of adders will be equal to number of connections to previous layer minus one [9].

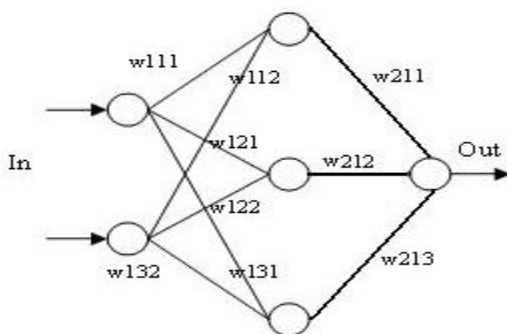


Figure 3: A 3-layer ANN

This structure consists of 3-layers, input, hidden and output layer. Each neuron consists of 'N' bit input data and 'N' bit input weight. The output of the neuron has two stages – multiplication and accumulation. Multiplication requires '2N' bits, but the input of the next stage neuron has 'N' bit input, therefore the multiplication has to be truncated to 'N' bit. Thus the transfer or activation function for the neuron is chosen to be the truncating/rounding unit. Such truncation introduces positive error between the range '0' to '2^{-N-1}'. However solutions are provided in [10] to minimize the error. Even though accuracy plays a critical role, several applications can tolerate inaccuracies to certain level so that their computational efforts are reduced. For example in image processing applications, if an processed image with less-accuracy provides an same information as that of the more-accurate design, then the additional complexity involved in accurate arithmetic design can be trade-off. Advantages of such truncation, reduces the complexity involved in the computations and its cumulative area, delay and power parameters.

As mentioned previously, the ANN architecture contains multipliers and adders. Here a typical Wallace tree based multiplier is utilized. And to implement the addition, a ripple carry addition (RCA) is incorporated. RCA contains cascade of full-adders. Typically or conventionally used Full adder architecture is shown in Figure 4.

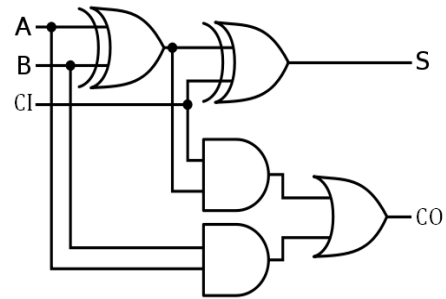


Figure 4: Conventional Full Adder architecture [11]

Conventional Full adder shown in Figure 4, consists of two XOR gates, two AND gates and one OR gate. As the number of cells are more in this architecture, interconnects between the gates will be more and results in the larger interconnect delays. In FPGA's, as the logics are mapped in-to Look-up-tables (LUTs), Multiplexers and carry path arithmetic's, more number of cells will lead to large number of LUTs and interconnects. Larger interconnect; large will be the delay and more LUTs (more area) leads to higher power consumption. To mitigate the effect of interconnect; the datapath architecture should be specific to LUT organization.

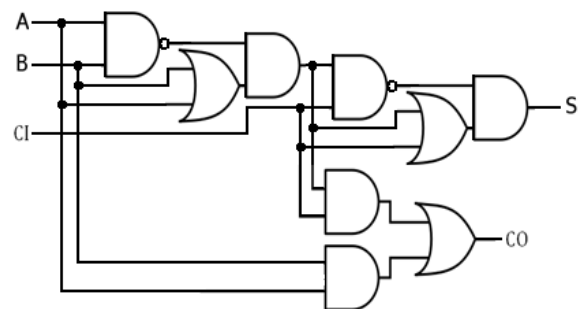


Figure 5: Proposed Full Adder architecture

Figure 5 show the proposed full-adder architecture. It consists of larger fan-in gates where the logics are mapped to minimum number of LUTs and requires lesser interconnects.

While implementing the design in FPGA domain, following optimizations and methodologies were applied to improve the results of the proposed design.

- Higher fan-in gates based architecture was developed
- Minimal number of gates are considered for the implementation
- Resource sharing
- Modeling the architecture in HDL as per the FPGA architectures

The proposed full adder architecture can also be utilized in ASIC domain, as the larger fan-in gates have higher transistor

stack which inturn have higher resistance and helps in reducing the leakage power.

4. RESULTS

The ANN designs were modeled using Verilog HDL and functionality was verified using the Mentor graphics model-sim simulator with the help of waveform editor. Figure 6 shows the verification results of the experiment. Designs were synthesized in Xilinx ISE tool by targeting the Virtex 7 (xc7v285t-3-ffg1157) FPGA device. Figure 7 shows RTL schematic view of the ANN block diagram and also the layout view of the logics mapped to FPGA resources. Power Analysis was done by considering all the standard input parameters using Xilinx Power Analyzer. Results of both the conventional and proposed ANNs were benchmarked as per the standard FPGA design methodology. The results of the conventional and proposed designs were tabulated in Table 1 and its chart view in Figure 8.

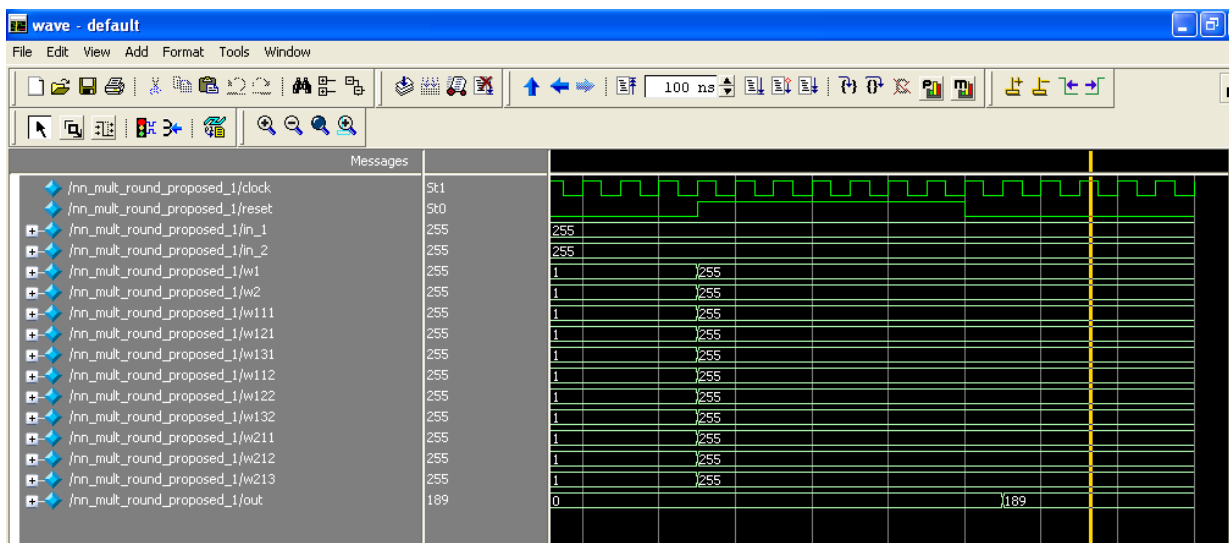


Figure 6: Simulation result of the ANN design

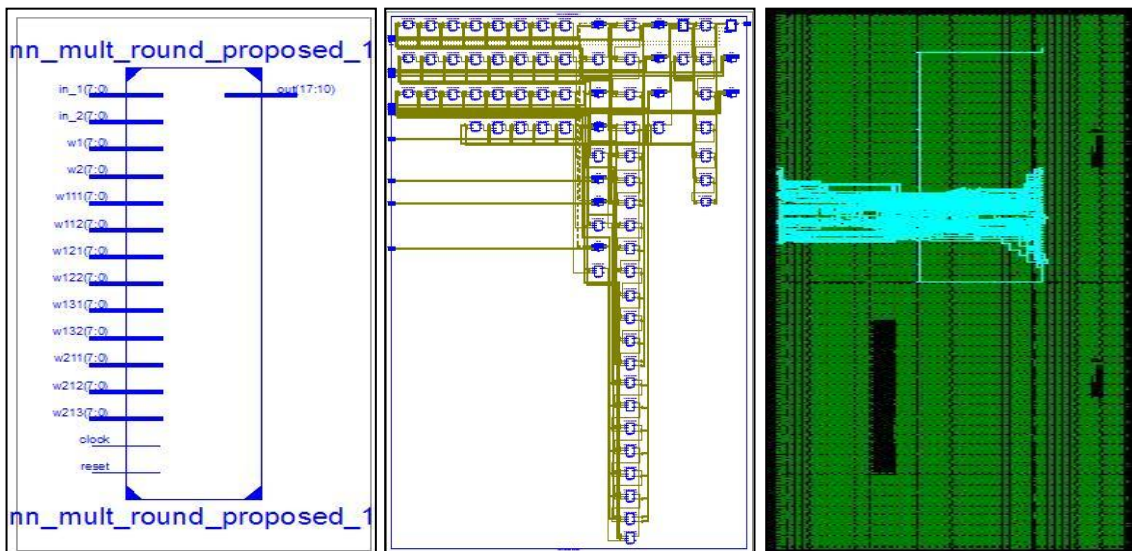


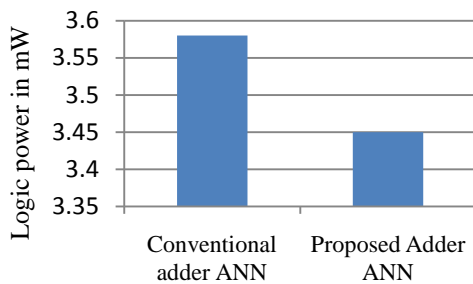
Figure 7: Schematic and FPGA Layout view of the ANN design

Table 1: Logic Power consumption Results of the conventional and proposed 3-layer ANN

PARAMETER	Conventional Adder ANN design	Proposed Adder ANN design	% change
Logic/DSP power (mW)	3.58	3.45	3.6

Note: mW: milli Watt

Table 1 gives the logic power consumption results of the existing and proposed 3-layer ANN architectures. Logic power result suggests that the proposed architecture has lesser power dissipation than the existing architecture. The use of higher fan-in gates has reduced the power consumption of the proposed design by 3.6% and this impact will be higher in intensive computation applications.



8: Logic power consumption

5. CONCLUSION

Low power adder based ANN was illustrated in this work. Here the importance of datapath architectural optimizations is presented by utilizing the proposed full adder architecture in the adder part of ANN. Proposed concept has resulted in 3.6% change in the logic power consumption at the ANN architecture level. Extending the concept to any bit-widths in the design will have similar improvements and due to datapath architectural optimizations; the proposed concepts impact will be higher for applications where intense computations are involved.

6. REFERENCES

[1] Hassanien, Aboul-Ella, et al. "Computational intelligence in multimedia processing: foundation and trends." *Computational Intelligence in Multimedia*

Processing: Recent Advances. Springer Berlin Heidelberg, 2008. 3-49.

- [2] Fang, Xuefeng "Small area, low power, mixed-mode circuits for hybrid neural network applications" Diss. Ohio University, 1994.
- [3] C. A. Mead, Ismail M, "Analog VLSI Implementations of Neural Systems", Reading, MA: Addison-Wesley, 1989
- [4] Chang, Chip-Hong, Jiangmin Gu, and Mingyan Zhang. "Ultra low-voltage low-power CMOS 4-2 and 5-2 compressors for fast arithmetic circuits." *Circuits and Systems I: Regular Papers, IEEE Transactions on* 51.10 (2004): 1985-1997.
- [5] Murali, L., D. Chitra, and T. Manigandan. "Low Power Adder Based Digital Filter for QRS Detector." *The Scientific World Journal* 2014 (2014).
- [6] Sharifi, Fazel, et al. "A Flexible Design for Optimization of Hardware Architecture in Distributed Arithmetic based FIR Filters." *RadioElectronics & Informatics* 4 (2012) 25-30; arxiv: preprint arXiv:1403.4554.
- [7] Rashidi, B.; Pourormazd, M., "Design and implementation of low power digital FIR filter based on low power multipliers and adders on xilinx FPGA," *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, vol.2, no., pp.18,22, 8-10 April 2011.
- [8] W. S. McCulloch and W. H. Pitts, "A logic calculus of the ideas imminent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115-133,1943.
- [9] Sahin, Suhap, Yasar Becerikli, and Suleyman Yazici. "Neural network implementation in hardware using FPGAs." *Neural Information Processing*. Springer Berlin Heidelberg, 2006.
- [10] Dhafer r. Zaghar, "Reduction of the error in the hardware neural network", *Al-khwarizmi Engineering Journal*, Vol.3, No. 1 PP, 80-41 (2007)
- [11] Neil H Weste and David M Harris, "CMOS VLSI Design-A Circuits & System Perspective", *Pearson Education*, 2008.