5:3 compressor based neural network with LM Algorithm in Multiplier as Application

Lanka Sai Charan*, M Vaishnavi Reddy, N Pooja Reddy and J V R Ravindra
C-ACRL, Department of Electronics and Communications Engineering, Vardhaman College of Engineering, Telangana, India

*lankasaicharan@ieee.org

Abstract. In conventional approach for VLSI, implementing large numbers of operations in parallel is possible with NN. The fundamental task of a neural network hardware is not dependent on the implementation technology and are quite constructive in simulating the digital circuits. These NN representations can be incorporated in various applications where in the behaviour of these circuits are essential to get a solution for discrete problems. Therefore, supervised learning algorithms are used to train the computer software. One such algorithm is the L-M method which is one of the powerful and widely used approaches. This paper accounts to the designing and training of a neural network model as computing technique in 5:3 compressor with Lavenberg-Marquardt Algorithm. The performance parameters are significantly remarkable for real time implementation, are presented in the paper.

1. Introduction

Advancements in the neural network-based applications always need efficient hardware in terms of power, delay, and throughput. Thus, employing such hardware always challenges VLSI research. Introduction of on-chip computation, less data movement - based architectures, developing efficient MAC (Multiply and Accumulate) units and in-memory/near memory concepts [1] are employed in neural network accelerators [2] to perform such powerful computation of neural networks [3]. Each neuron is a combination is multiplying of weights and inputs, and the result is passed to the activation function. Thus, using various algorithms [4], these neurons developed network are trained and utilized to develop various advance technologies like artificial intelligence [5]. Back propagation technique is used in hardware circuits [6] as a learning method for neural network.

A neural network is designed to build a training model that comprises a synthesized model that is competent enough of comprehending the outputs at high accuracy, precision which are extensive to machine learning and artificial intelligence applications that have implemented networks as transitional blocks of functionality. The efficiency of neural networks is depicted by regression, gradient, testing error parameters specifying as the ratio of the number of error outputs by the number of exact outputs such as MSE, SSE MAE, etc. The following paper expilicates about the training of a 5:3 compressor-based neural network, along with testing the exemplar, elucidating error parameters, accuracy, and few additional terms such as gradient, adjustment factor (μ) that are measuring factors of a training model.
The rest of the paper is followed by literature survey in section 2, which continuous with proposed work in section 3. Section 4 is inclusive of simulation results, the application of 5:3 compressor in digital circuits in section 5, concluding the paper in section 6.

2. Literature Survey

Research on neural networks and its applications in VLSI domain has commenced 3-4 decades back. The hardware realization of neural networks and its applications were majorly concerned with a decrease in energy, memory, and delay factors. Some of the survey points are as follows-

In [7], the performance evaluation of the on-chip interconnects using the DIL model is prepared using neural network, where performance of neural networks is evaluated at various on-chip interconnect lengths. Conventional methods consume more time in simulating repeatedly at different values of lengths. To reduce this time, neural networks were used where the performance of the system is determined using a regression algorithm. The main purpose of this paper is to provide an automatic performance evaluation method by avoiding traditional time-consuming methods.

Similarly, a high energy efficient MAC-based mixed-signal neuron architecture is presented in [8] by using a small-signal voltage mode multiplication. This design attained high energy efficiency without effecting accuracy parameters in applications of image/speech recognition. This architecture achieved energy efficiency higher than 20 times than that of prior specified neuron designs and around 3 times in case of memristor-based designs.

In [9], neural networks are used to evaluate system parameters, biasing field, and related parameters of a nano device for enhancing better performance devices with low operating power. Parallelism is applied for fast learning thus making an optimized design much faster.

3. 5:3 Compressor Neural Network design

The conception of neural networks is explored by the way biological neurons in human brain function, the way it processes data, links various scenarios. A Neural Network is precisely an interconnected artificial neuron energized by activation functions, weights, training algorithms supporting the network to generate desired outputs. The classification of neural networks relies on considerations of layers, data flow, neurons exercised and their densities, depth activation filters etc [10]. This paper accounts for a training model of a 5:3 compressor that operates with the Levenberg-Marquardt algorithm for high speed and low memory.

A neural network design comprises of input, hidden, and output layers. The peripheral layers are always constant, the hidden layer count keeps altering with the requirement of application. The design evaluates the function with proper selection of weights, bias, training algorithm, and activation function.

![Figure-1: 5:3 Compressor design](image-url)
The input for 5:3 compressor design as shown in Figure 1, of 5bit size is given as input set with 64 samples consisting of twice 32 input combinations of 5 elements as input layer to Neural network, and the output layer with 64 samples consisting of 3 elements the model is trained thoroughly, rigorously by the Levenberg-Marquardt training (TRAINLM) algorithm. TRAINLM marks its significance by producing a high-speed network that is consuming less time. In order to be an ideal representative of a network in the digital counterpart, each neuron, along with its count is taken into consideration by the algorithm. The network is chosen with a single hidden layer with 18 neurons in it. Thus, a feed-forward back propagation neural network is engaged for the digital circuit. To finalize the training model the parameter considerations are as follows - to provide the network for large training space with 85% of inputs as the training set, along with the maximum limit of epoch value of 1000 for better generalization that terminates the training when the improvement by the mean squared error (MSE) of the validation samples. The training set circuit is shown in Figure 2, the inputs are propagated to the neural network for training, testing, and validation operations. The performance of a network is decided by the error parameters along with neuron count per layer, number of iterations taken by the network out of the epochs provided, the least the value, the efficient is the design.

![Neural network Model](image)

Figure 2: Neural network Model

Compared with the real-valued signals, binary signals have higher possibility rates of performance especially in the case of spiking neurons [11]. The decision making in the selection of neuron count depends on the functional techniques such as underfitting and overfitting. If the test set performance is better than the training set performance, the neuron count must be decreased gradually to equalize the performance rates apart from this observation a typical intuition by many is increasing the accuracy by increasing the neurons.

4. Simulation Results

A neural network is alleged to be completely valid when the set accuracies of train and test are approximately equal. The novel feed-forward back-propagation network for 5:3 compressor is designed in the MATLAB Neural Fitting tool R2017a. The competence of the neural network is illustrated by the parameters in the Table.
While implementing a design it is important to consider the attributes that make the circuit to be impeccable, one such aspect is to find performance of a function in terms of error parameters such as MSE or SSE. The training performance rate is obtained as 6.1425E-23 at epoch 9 as shown in Figure (3) indicating negligible value. Second, the adjustment factor $\mu$ that gets formulated at each iteration with the computation process of the Jacobian Matrix (J) for every iteration within the epoch range. Initially, the network gets started with small ($\mu$) value and gets regulated with each iteration ($\mu$) dec or ($\mu$) inc. The optimization factor that directs the steepest ascent of the loss function is known as Gradient as shown in Figure (4) (5). The overall observation of the network outcomes results in elevated performance of a 5:3 compressor neural network with TRAINLM algorithm. Compressor used as transitional circuitry in electronics proves to be part of deep learning that enhances the circuit functionality.

Figure 3: Performance of Neural Network

Figure 4: Training State of Neural Network
| S. No | Parameter | Value       |
|-------|-----------|-------------|
| 1     | Epoch     | 9           |
| 2     | MSE       | 6.142E-23   |
| 3     | SSE       | 1.85E-09    |
| 4     | (μ)       | 1.00E-12    |
| 5     | Gradient  | 5.26E-12    |

Table-1: Parameter values for 18 neurons based neural network

Figure 5: Error Histogram of Neural Network

| S. No | Neurons | Training MSE | Testing MSE |
|-------|---------|--------------|-------------|
| 1     | 8       | 2.87E-10     | 3.87E-1     |
| 2     | 10      | 2.02E-10     | 3E-10       |
| 3     | 15      | 1.10E-16     | 5.79E-17    |
| 4     | 18      | 6.142E-23    | 1.112E-22   |
| 5     | 20      | 2.01E-1      | 3.10E-1     |
| 6     | 22      | 2.036E-25    | 2.328E-17   |
| 7     | 25      | 2.39E-25     | 3.086E-25   |
| 8     | 30      | 6.60E-15     | 8.46E-15    |

Table-2: Neurons v/s MSE (Mean Squared Error)
Neural network is built by considering various neuron counts as described in Table 2. From the table, it is found that 18 neurons gave a better MSE values for both training and testing cases. Training MSE is valued at 6.142E-23 and testing MSE is valued at 1.112E-22. With least count of neurons and better performance, the neural network with 18 neurons in the hidden layer is considered to be a better choice from Table 2.

5. Application: 8 × 8 Multiplier

Machine learning (ML) is employed in many applications where large data needs to be compiled and efficient results must be extracted from it. In the field of VLSI, ML is used in finding more accurate output generating circuits. Various ML algorithms are used in finding efficient circuits by considering parameters like error distance, methods of groupings of 1’s and 0’s in K-maps [12] etc. Designing an efficient 8 × 8 multiplier by considering above mentioned parameters, is the main goal of this paper. Hence by maintaining the accuracy, an algorithm is proposed in building an 8x8 multiplier with the novel 5:3 compressor resulting in least error outputs.

Algorithm:
Step 1: Calculate all partial products and arrange them as shown.
Step 2: If any of the partial products is zero, consider remaining non-zero elements and proceed.
Case 1 – If there are 2 or 3 elements in the column including carry, perform normal addition.
Case 2 – If there are 4 elements in the column including carry, apply them to 4:3 compressor NN model.
Case 3 – If there are 5 elements in the column including carry, apply them to 5:3 compressor NN model. Check for inputs X3 X2 X1 at iterations (4,8,15,16,23,27) for output O1. If found, complement the output obtained from NN model. Check for inputs X4 X3 X2 X1 at iterations (7, 11) for output O2. If found, complement the output obtained from NN model.
Else, continue.
Case 4 – If there are more than 5 elements in the column including carry, apply above cases accordingly to reduce them. Repeat Step 2 until one attains a single row.

6. Conclusion

The concept of 5:3 compressor is implemented using neural networks in MATLAB with the Lavenberg-Marquardt algorithm. Various parameters like MSE, SSE, gradient is considered during analysis and had better results for 18 neuron count in the neural network with the training MSE valued at 6.142E-23 and the testing MSE valued at 1.112E-22. After simulation of a serial-in parallel-out circuit combined with the neural network, it was found that the NN model has worked accurately as that of the digital circuit implementation of 5:3 compressor thus becoming an efficient design of a 5:3 compressor.
References

[1] Junjie Mu and Bongjin Kim. A 65nm logic-compatible embedded and flash memory for in-memory computation of artificial neural networks. In 2020 IEEE International Symposium on Circuits and Systems (ISCAS), pages 1–4. IEEE, 2020.

[2] S Rasoul Faraji, Pierre Abillama, Gaurav Singh, and Kia Bazargan. Hbucnna: Hybrid binary-unary convolutional neural network accelerator. In 2020 IEEE International Symposium on Circuits and Systems (ISCAS), pages 1–5. IEEE, 2020.

[3] Eyal Kolman and Michael Margaliot. Knowledge extraction from neural networks using the all-permutations fuzzy rule base: the led display recognition problem. IEEE transactions on neural networks, 18(3):925–931, 2007.

[4] Laura Wang and Matt Luo. Machine learning applications and opportunities in ic design flow. In 2019 International Symposium on VLSI Design, Automation and Test (VLSI-DAT), pages 1–3. IEEE, 2019.

[5] A Dinu and PL Ogrutan. Opportunities of using artificial intelligence in hardware verification. In 2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME), pages 224–227. IEEE, 2019.

[6] Zheng Tang, OKIHIKO Ishizuka, and HIROKI Matsumoto. Backpropagation learning in analog t-model neural network hardware. In Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan), volume 1, pages 899–902. IEEE, 1993.

[7] Ajita Misra, Diksha Diksha, Yash Agrawal, and Vinay Palaparthy. Performance evaluation of on-chip interconnect system using prospective neural network design. In 2020 IEEE International Students’ Conference on Electrical, Electronics and Computer Science (SCEECs), pages 1–4. IEEE, 2020.

[8] Baibhab Chatterjee, Priyadarshini Panda, Shovan Maity, Ayan Biswas, Kaushik Roy, and Shreyas Sen. Exploiting inherent error resiliency of deep neural networks to achieve extreme energy efficiency through mixed-signal neurons. IEEE Transactions on Very Large-Scale Integration (VLSI) Systems, 27(6):1365–1377, 2019.

[9] Subir Kumar Sarkar, Ankush Ghosh, MA Gautham, A Sobha Rani, and Debasis Samanta. An efficient technique of integrating parallel neural networks for faster and power efficient nanodevices for ultradense vlsi cir8 cuits. In 2007 International Workshop on Physics of Semiconductor Devices, pages 232–235. IEEE, 2007.

[10] Ying Ma and Jose C Principe. A taxonomy for neural memory networks. IEEE transactions on neural networks and learning systems, 2019.

[11] Sungho Jo, Jijun Yin, and Zhi-Hong Mao. Random neural networks with state-dependent firing neurons. IEEE transactions on neural networks, 16(4):980–983, 2005.

[12] Lavanya Maddisetti and JVR Ravindra. Machine learning based power efficient approximate 4: 2 compressors for imprecise multipliers. In 2019 32nd International Conference on VLSI Design and 2019 18th Int Conf on Embedded Systems (VLSID), pages 221–226. IEEE, 2019.