A Neural Network trained to perform as an Imprecise 4:2 Compressor using Machine Learning Algorithm

Lavanya Maddisetti1,2,*, Ranjan K. Senapati1, JVR Ravindra2

1Department of ECE, KL Deemed to be University, Vijayawada, India
2Center for Advanced Computing Laboratory, Vardhaman College of Engineering, Telangana, India

(*Corresponding author: lavanya.maddisetty@gmail.com)

Abstract: In many fields, such as speech recognition, image processing, Internet of Things (IoT) systems, arithmetic circuits, etc., artificial neural networks have become a popular solution for a wide range of problems. In this research work, a 4:2 compressor-based arithmetic circuit is considered and an approximation technique called probabilistic pruning is applied to it. This proposed compressor is compared with various existing approximate 4:2 compressors. Training of a neural network as proposed inexact 4:2 compressor is performed using a supervised machine learning algorithm. The accuracy values of the proposed compressor are acceptable at architectural and neural network levels. The absolute difference between the train set and the predicted test set accuracy values obtained is very less (1.73%) when compared with the other (inexact 4:2 compressors) trained deep neural network models.

1. Introduction

Low power, high speed with smaller silicon area is the demanding parameters from decades to design arithmetic circuits like 4:2 compressor [1]. The inexact compressor is the one that consumes less power and is accurate than the other existing inexact 4:2 compressors.

Two analyses have been shown in this work; the first is the proposed architecture level approach where simulations of compressors are done to find the area, power, and energy consumption. The second approach is the training of Neural Network (NN) as the proposed inexact 4:2 compressor. The truth table of the proposed 4:2 compressor is given as the dataset to train the NN and it is split into train and test sets to predict the output of the compressor circuit.

Our contributions:

- The inexact 4:2 compressor is proposed with the least number of transistors and observed the power consumption, energy, and area of it by comparing it with the existing inexact compressors.
- Error analysis is performed for the existing and proposed inexact 4:2 compressors.
- Training of NN as an inexact 4:2 compressor is done using supervised Machine Learning (ML) algorithm by giving the truth table of the proposed compressor as the dataset.

2. Related work

The authors of [1] proposed a high speed, low power compressors of different orders which operate at ultra-low voltages with multiplexers in the proposed architecture. Two CMOS+ gate-based circuits were proposed in [2] to reduce power, area, and delay of 4:2 compressor. In [3], the authors used a compressor structure in a multiplier design. Also, a sparse adder was proposed to use it in the final addition of a multiplier. M. Ben has given a general introduction to approximate computing in [4] and proposed a taxonomy to show the merits of different approximation techniques. The author discussed the opportunities and challenges of non-deterministic approximate computing. The authors of [5] have
presented a study on combining three approximation techniques on more than one layer of system stack to get more benefits and said that benefits can be applied to different domains.

The authors of [6] have presented the way they have developed the spanning circuits, architecture, and software by applying techniques of approximate computing. In [7-8], 4:2 compressors were proposed and they were placed in a Dadda multiplier to get low power multipliers. Finally, these multipliers were used in image processing applications.

Due to approximations on digital circuits, the difference in the outputs of exact and inexact circuits can be calculated from the equations proposed in [9]. The work of [10] also resembles the same as [7-8] but the authors have used different operand lengths of approximate multipliers and image filtering application was shown in their research work. The ML techniques applied to CAD algorithms were given in [11] to make automation in the operation in the tools. In [12], the authors have proposed an approximate compressor using inexact logic minimization and embedded it in the multiplier for power efficiency and finally, an algorithm was given to simulate them in Google Colaboratory.

3. Proposed method

This section presents a proposed inexact 4:2 compressor employing a probabilistic pruning type of approximation. A feed-forward neural network is trained to behave as the proposed 4:2 compressor.

3.1 Architectural approach of proposed 4:2 compressor

The probabilistic pruning type of approach is applied to the exact 4:2 compressor by removing a multiplexer present at the output of Sum [1]. Thus, the Sum output for the proposed 4:2 compressor can be taken from the multiplexer whose select line is connected to the output of the second XOR [1]. With this approximation, there is no difference in the outputs of Cout and Carry. The proposed approximate method gives the least number of errors with less transistor count. The Sum expression for the proposed 4:2 compressor is obtained as equation (1).

\[
\text{Sum} = (X_3 \oplus X_4)' (X_1 \oplus X_2) + (X_3 \oplus X_4) (X_1 \oplus X_2)'
\]  

(1)

The transistor-level circuit of the proposed 4:2 compressor is shown in Figure 1.

![Figure 1: Proposed Inexact 4:2 Compressor with 16 Transistors](image)

Minimum numbers of errors are present in the above circuit. i.e., Cout and Carry have `0' errors and the Sum output has `8' errors which give a 25% error rate considering all the three outputs. When the approximate design of [12] is examined, it also has a 25% error rate, but it is obtained considering only Sum output. The Cout and Carry outputs of [12] are giving 8 and 4 errors respectively when compared with the exact 4:2 compressor. Thus, the total errors of the proposed inexact circuit have `0,
3.2 Neural Network Model of an Inexact 4:2 Compressor

A supervised machine learning type of algorithm is used for training the neural network. The truth table of the proposed inexact compressor is given as dataset to the NN with $C_{in}$, $X_4$, $X_3$, $X_2$, $X_1$ inputs, and $C_{out}$, Carry, and Sum target labels. The architecture of the NN with inputs and outputs is shown in Figure 2.

![Figure 2: Neural Network Model as Inexact 4:2 Compressor with the activation functions, ReLU at input and Softmax at the output](image)

The truth table is divided into train and test sets to find the accuracy of the NN model. The accuracy value obtained from the train and test sets describe how best the NN model is performing as the proposed inexact compressor. The difference between the trained accuracy and the predicted test accuracy should be equal in ideal cases for the NN model to be complete. When these accuracy values are equal then the model is said to be completely validated.

4. Results and Discussion

The experimental work of 4:2 compressors was performed using the Spectre simulator of Cadence Virtuoso for schematic simulations. To find the area of these architectures Cadence Layout XL was used in 45nm CMOS technology node. The error analysis of compressors was performed using MATLAB. The training part of the NN was done using the Jupyter notebook with Pandas, Numpy, and Keras libraries.

4.1 Simulation Results of 4:2 compressors

In this subsection, the proposed inexact 4:2 compressor is compared with the exact and existing approximate 4:2 compressors. The comparison has been done in terms of transistor count, power, energy, and area for exact, existing inexact, and proposed inexact 4:2 compressors. The simulation results of the compressors are tabulated in Table 1.
Table 1: Simulation Parameters of 4:2 Compressors

| 4:2 Comp.  | Transistor Count | Power (nW) | Energy (pJ) | Area (nm²) |
|------------|------------------|------------|-------------|------------|
| Exact [1]  | 50               | 178.9      | 1861        | 29.893     |
| Inexact [8]| 40               | 157.4      | 1348        | 23.522     |
| Inexact [12]| 34               | 110        | 836.9       | 20.34      |
| Inexact [7]| 28               | 102.4      | 650.5       | 16.728     |
| Proposed   | 16               | 83.23      | 240.3       | 9.641      |

When the proposed compressor is compared with the design1 of [7], the improvement in power, energy, and the area is 18.72%, 63.05%, and 42.36% respectively. From Table 1, it is observed that the power, area, and energy parameters are varying according to the transistor count of each 4:2 compressor. The graph is shown in Figure 3 also explains that these parameters are varying for the transistor count.

Figure 3: Transistor Count, Power, Energy and Area of Exact and Inexact 4:2 Compressor

Table 2: Error Metrics of Compressors

| 4:2 Comp.  | Errors | ER (%) | MED            | NMED            | RED            |
|------------|--------|--------|----------------|-----------------|----------------|
| Exact [1]  | 0      | 0      | 0              | 0               | 0              |
| Inexact [8]| 5      | 31.25  | 3.3432x10⁶     | 1.044x10⁵       | 0.0604         |
| Inexact [12]| 8     | 25     | 5.535x10⁷      | 1.729x10⁶       | 1              |
| Inexact [7]| 12     | 37.5   | 7.551x10⁷      | 2.359x10⁶       | 1.3641         |
| Proposed   | 8      | 25     | 1.203x10³      | 37.6025         | 2.1736x10⁻⁵    |

ER is Error Rate, MED is Mean Error Distance, NMED is Normalized mean error distance

The observations from Table 2 are

- The proposed inexact 4:2 compressor and [12] have the same error rate. This is because in [12] only sum is considered for error rate whereas in the proposed circuit the error rate is 25% considering all the outputs (0+0+8 for C₀, Carry, and Sum).
- Although the transistor count is very less, the design1 of [7] was giving a 37.5% error rate which is higher than the rest of the circuits.
- The design of [8] has only 4 inputs Cᵢᵣ, X₄, X₃, X₂, X₁ inputs, and two outputs (Carry and Sum). Hence error analysis has been performed for 16 inputs instead of 32 inputs. Therefore, the error rate is 31.25%.
4.2 Simulations of a deep neural network model

Training of NN has been done by splitting 66% of the dataset into a train set and 34% of the dataset into a test set. The parameters used for building the NN is given in Table 3.

Table 3: Neural Network Parameters used in Training

| NN Parameters                | Values                      |
|-----------------------------|-----------------------------|
| Test size                   | 34% of the dataset          |
| Classifier                  | Sequential                  |
| Kernel initializer          | Uniform                     |
| Input activation function    | ReLU                        |
| No. of Hidden Layers        | 2                           |
| Output activation function   | Softmax                     |
| Optimizer                   | adam                        |
| Loss function               | Categorical cross-entropy   |
| Batch Size                  | 10                          |
| No. of Epochs               | 1000                        |

For the proposed inexact compressor, the train set accuracy is obtained as 61.90%. To assess the performance of the trained model, the prediction of its accuracy has been done by the test set. After prediction, the confusion matrix is obtained as

\[
\begin{pmatrix}
5 & 0 & 0 \\
4 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\]

Test set accuracy = \( \frac{\text{No. of correct outputs}}{\text{Total outputs}} \)

\[
= \frac{5 + 1 + 1}{11} = \frac{7}{11} = 63.63\%
\]

The absolute difference between the accuracies of train and test sets is 1.73%. This difference suggests that the NN model is perfectly built as a proposed inexact 4:2 compressor. The accuracy and loss graphs for the number of epochs are shown in Figure 4.

![Figure 4: Accuracy and Loss graphs for the Train and Test sets of the proposed inexact NN model](image-url)
To compare train and test accuracies of the other approximate designs, NN models are built using datasets, and accuracies are obtained as shown in Table 4.

| DNN Model       | Train Set Accuracy (%) | Confusion Matrix | Test Set Accuracy (%) | Absolute Difference (%) |
|-----------------|------------------------|------------------|-----------------------|-------------------------|
| Exact [1]       | 85.71                  | \[1 \ 0 \\ 3 \ 2\] | 81.81                 | 3.9                     |
| Inexact [8]     | 60                     | \[1 \ 0 \\ 3 \ 2\] | 50                    | 10                      |
| Inexact [12]    | 57.14                  | \[2 \ 4 \ 0 \\ 0 \ 4 \ 0 \ 0 \ 1 \ 0\] | 54.54                 | 2.6                     |
| Inexact [7]     | 57.14                  | \[6 \ 0 \ 0 \\ 4 \ 0 \ 0 \ 1 \ 0 \ 0\] | 54.54                 | 2.6                     |
| Proposed        | 61.9                   | \[5 \ 0 \ 0 \\ 4 \ 1 \ 0 \ 0 \ 0 \ 1\] | 63.63                 | 1.73                    |

From Table 4, it is clear that the proposed inexact 4:2 compressor is modeled perfectly since the absolute difference between train and test set accuracies obtained are higher for other models when compared with the proposed one.

5. Conclusion

Our findings suggest that the proposed inexact method can be applied to any arithmetic combinational circuit in terms of the architectural and training of neural network approaches. The proposed inexact 4:2 compressor model has the least absolute difference between a train and test accuracies which proves that the NN model is validated completely. The power and area consumption of the proposed compressor is also less with the least error rate.

Acknowledgments

This research project was carried out at C-ACRL, Vardhaman College of Engineering. The authors would like to thank the management and faculty for their constant support throughout.

References

[1]. Sreehari Veeramachaneni, Kirthi M Krishna, Lingamneni Avinash, Sreekanth Reddy Puppala, M.B. Srinivas, (2007), Novel architectures for high-speed and low-power 3-2, 4-2 and 5-2 compressors, Proceedings of International Conference on VLSI Design (VLSID), pp. 324-329.

[2]. Raphael Dornelles, Guilherme Paim, Bianca Silveira, Mateus Fonseca, Eduardo Costa, Sergio Bampi, (2017), A Power-Efficient 4-2 Adder Compressor Topology, 15th IEEE (NEWCAS), Strasbourg, France, pp. 281-284.
[3]. Sreehari Veeramachaneni, M. B. Srinivas, 2013, Design of optimized arithmetic circuits for multiplier realization, IEEE Asia Pacific Conf. on Postgraduate Research in Microelectronics and Electronics (PrimeAsia), India.

[4]. M. Ammar Ben Khadra, 2017, An introduction to approximate computing. December, [Online]: https://arxiv.org/abs/1711.06115.

[5]. Anurag Agrawal, Jungwook Choi, Kailash Gopalakrishnan, Suyog Gupta, Ravi Nair, 2016, Approximate Computing: Challenges And Opportunities, IEEE International Conference on Rebooting Computing (ICRC), San Diego, USA.

[6]. S.Venkataramanis, S.T. Chakradhar, K. Roy, A. Raghunathan, 2015, Computing approximately, and efficiently, in Design, Automation & Test in Europe Conference & Exhibition, pp. 748-751.

[7]. Amir Momeni, Jie Han, Paolo Montuschi, Fabrizio Lombard, 2015, Design and analysis of approximate compressors for multiplication, IEEE Transactions Computers, 64(4), April, pp. 984-994.

[8]. Suganthi Venkatachalam, Seok-BumKo, 2017, Design of Power and Area Efficient Approximate Multipliers, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 25 (5), pp.1782-1786.

[9]. Jinghang Liang, Jie Han, and Fabrizio Lombardi, 2012, “New Metrics for the Reliability of Approximate and Probabilistic Adders, IEEE Transactions on Computers, 62(9), pp. 1760–1771

[10]. Darjn Esposito, Antonio Giuseppe Maria Strollo, Ettore Napoli, Davide De Caro, Nicola Petra, 2018, Approximate Multipliers Based on New Approximate Compressors, IEEE Transactions on Circuits and Systems I: Regular Papers, PP(99), pp. 1-14.

[11]. Peter A. Beerel, Massoud Pedram, 2018, Opportunities for Machine Learning in Electronic Design Automation," IEEE International Symposium on Circuits and Systems (ISCAS), Italy.

[12]. Lavanya Maddisetti, JVR Ravindra, 2019, Machine Learning based Power-Efficient Approximate 4:2 Compressors for Imprecise Multipliers, in 32nd International Conference on VLSI Design, New Delhi, India.