Supervised Machine Learning for Training a Neural Network as 5:2 Compressor

Lavanya Maddisetti, Ranjan K. Senapati, Ravindra JVR

Abstract: ***Machine Learning has achieved substantial development in numerous applications like image processing, pattern recognition, approximate computing etc. This paper interlinks supervised machine learning algorithm and VLSI architectures to train a neural network as exact and approximate 5:2 compressors. Probabilistic pruning type of approximation technique has been employed on the exact 5:2 compressor. This approximation technique on compressors reduces the power consumption with variation in the outputs without affecting the error limit. The simulation of 5:2 compressors and training of neural network using machine learning algorithm has been done using Spectre simulator of Cadence Design Systems at 45nm CMOS technology node and Keras library with TensorFlow background respectively.***

Index Terms: Accuracy, Classification, Machine Learning, Neural Network.

I. INTRODUCTION

Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) have been the exciting areas of interest from past few years which can be applied to various fields. These three are the subsets of one another i.e. ML is subset of AI and DL is subset of ML. Thus, machine learning algorithms and techniques can be employed in DL applications and vice versa is not applicable. The algorithms utilized in ML applications trains the neural network (NN) using huge amount of data to perform particular task. The data fed to the NN is used to make decisions while passing through the three layers namely input layer, hidden layer and output layer. Different applications of the above three are categorized as depicted in Fig. 1.

High speed arithmetic circuits are the essential components to perform basic computations like addition and multiplication where these are the repeated operations that are carried out in Very Large Scale Integration (VLSI) systems and digital signal processors. Hence, compressors evolved into a significant element in the architectures of almost all the multipliers. As the need for low power and higher performance are in demand in the portable devices, mobile gadgets and so on, approximate computing has been into rapid advancement in many applications like digital image processing, multimedia, machine learning, deep learning, etc., where small loss in accuracy is acceptable.

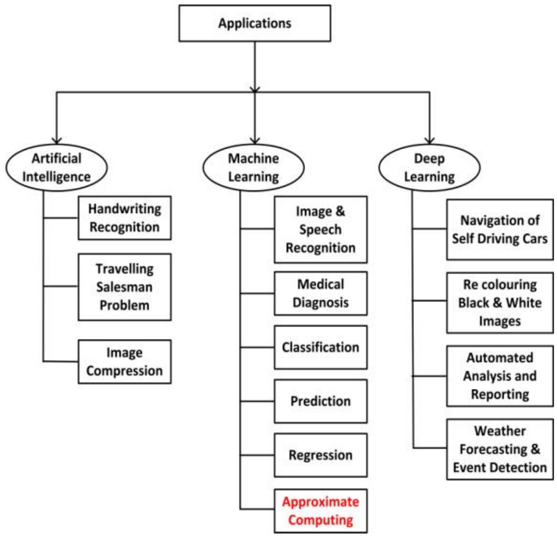

Fig. 1: Applications of Neural Networks

This research work deals with the ML application that examines the proposed approximate 5:2 compressor circuit. One of the classification algorithms of ML called supervised learning is employed to apply it to approximate 5:2 compressors. The subsequent sections of this paper are organized as follows. Section II discloses the work that was done previously in terms of different compressors, approximation techniques employed, error analysis, challenges in approximations and machine learning. The intention of the approximations on compressors and applying machine learning algorithm to these compressors is given in section III. Framework of training a neural network that can be applied to any application in shown in section IV. The analysis of the proposed work and simulations are demonstrated in sections V and VI respectively with conclusions in section VII.

II. RELATED WORK

The authors of [1] proposed 4:2 and 5:2 compressors using XOR-XNOR and 2x1 multiplexer circuits. In [2], Chang et.al presents XOR-XNOR circuits with different number of transistors and they have been employed in 4:2 and 5:2 compressors to operate them at ultra low supply voltages. Paper [3] discloses decimal compressors to handle decimal multipliers. The principal objective of the arrangement is to concentrate on compressors which are one of the fundamental elements of multiplier circuits that are being extensively used in high speed systems. The authors

Revised Manuscript Received on August 1, 2019.

Lavanya Maddisetti, research scholar in KL Deemed to be University, Vijayawada,AndhraPradesh,India e

Lavanya Maddisetti, Assistant Professor in Department of ECE, Vardhaman College of Engineering.

Ranjan K. Senapati, with KL Deemed to be University. He is a professor in the Department of Electronics and Communications Engineering

Ravindra JVR, with the Center for Advanced Computing Research Laboratory (C-ACRL), Department of Electronics and Communications Engineering, Vardhaman College of Engineering, Shamshabad, Telangana

of [4] proposed a new 5:2 compressor which gives glitchless output with increase in the speed of operation. A new 5:2 compressor with 58 transistors is discussed in [5]. In this paper, new design approaches have been investigated for low power 5-2 compressor circuits that acquire adequate drivability at ultra low voltages based on the progressive CMOS process technology. A 5:2 compressor with a 4:2 and 3:2 compressors was proposed in [6] and it is placed in a multiplier. In [7], a 5:2 compressor with probabilistic pruning type of approximation was proposed which has been used in this paper. Different challenges and opportunities in using approximate computing was given in [8]. Approximations have been implemented in [9-10] on existing 4:2 compressor to reduce the power consumption and the approximated compressor is placed in Dadda multiplier. Finally, this is used in image processing applications. In [11], the author proposed a compressor and an algorithm to explain the usage of proposed compressor in a multiplier which has been applied to image filtering and adaptive least mean squares filtering. Approximations can be applied to any digital circuit in various ways like probabilistic pruning and inexact logic minimization, where the authors of [12] used the latter technique in data path elements. When approximations are implemented in any architecture, the difference in output bits of exact and approximate circuits can be calculated using the equations given in [13-14]. Peter.A et.al of [15] has given approaches and challenges of ML in design automation i.e ML applications to VLSI designs. The authors of [16] have given a methodology for the design of hardware using approximations. Automation in approximate circuits using gate level pruning is shown in [17]. In [18], the author explained that the chip design process is common for all the applications. Hence, if this data process is made common then chip designing can be made automated. The implementation of basic component of deep neural network was performed using logarithmic number systems in [19].

III. PROBLEM STATEMENT AND SOLUTION

Low Power and low energy consumption of the VLSI architecture is the primary concern from decades with a little compromise in the accuracy. Hence approximations are to be implemented on 5:2 compressors with the intention to reduce the circuitry of partial product reduction stage of a multiplier. These approximations were performed in the literature on various architectures with impreciseness in the circuits output.

At present, the main intention of this research work is to lower the error rate and improve the accuracy of the approximate 5:2 compressor. For this, approximations in the form of pruning of XNOR2 terminal has been applied on the exact 5:2 compressor which is based on XOR-XNOR and 2x1 multiplexer circuits.

The transistor count and in turn the size of the circuitry is decreasing in 5:2 compressor with the removal of XNOR2 terminal. If there are different approximate 5:2 compressors with different number of errors then machine learning algorithm should choose the approximate compressor which has less error rate such that selected compressor can be placed directly in a multiplier to reduce the power consumption of it. This approach of approximating 5:2 compressors is applied to machine learning applications to train to select an ideal approximate compressor which has less error rate.

IV. FRAME WORK

Fig. 2. displays the flow to train a neural network which can be applied to any application. To find the accuracies of approximate 5:2 compressors, Supervised Learning (SL) technique is considered. This classification algorithm consists of two parts namely data pre-processing and building Artificial Neural Networks (ANN).

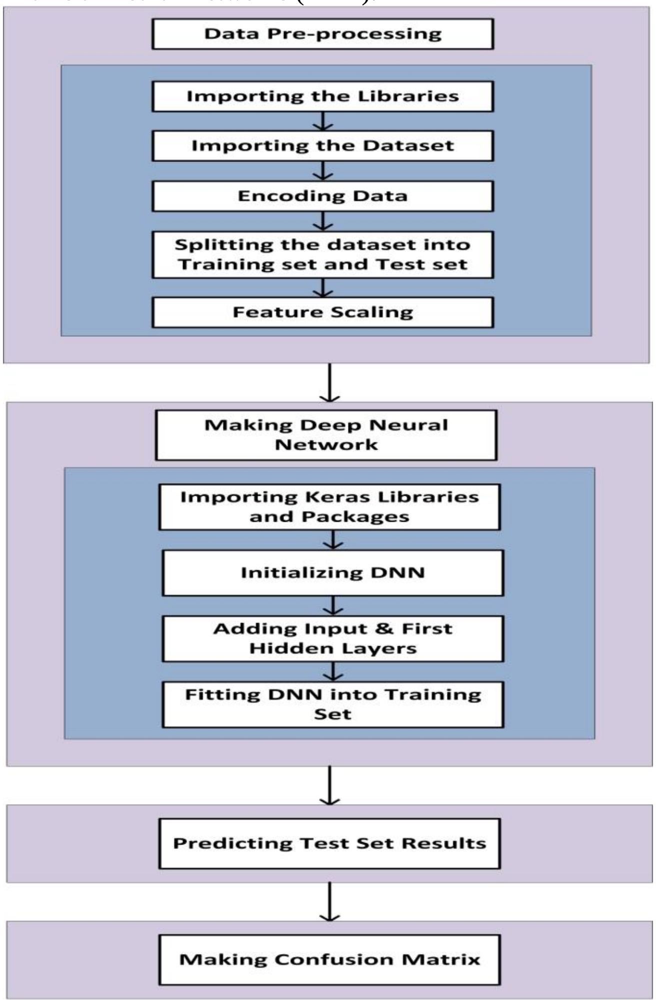

Fig. 2. Framework to train a Neural Network

In supervised learning, independent variables (inputs) are used to predict the dependent (output) variables. To predict the output variables, the critical inputs that are required for prediction are to be identified called encoding the data. Here the input variables are C_{in1}, C_{in2}, X_1, X_2, X_3, X_4, X_5 and the output is 'Sum'. Only Sum output is considered since it is generated only if all the input variables are utilized in the compressor operation i.e. to predict the Sum output all seven inputs are required. In this ML application, to split the dataset into training set and test set, encoding of data is not required since all the inputs are critical to obtain the 'Sum' output. The final step in part one is feature scaling which has to be done to perform parallel computations such that one independent variable does not dominate the other.

The second part of the classification algorithm is creating ANN for which the Keras library and its packages are to be imported; this builds ANN based on TensorFlow. The sequential and dense modules are required for initializing and building the ANN respectively. Initializing ANN is defining it as a sequence of layers and building ANN is adding input and the hidden layers which are done in five steps using stochastic gradient descent.

a) Initialize the weights randomly closer to smaller numbers.
b) The first observation row goes into neural network and each feature goes into input node. The number of nodes present is nothing but number of independent variables i.e. 7.
c) Forward Propagation: From left to right the neurons are activated by the activation function in such a way that the higher the values of the activation function is, the more it will pass from the nodes on the left to the nodes on the right. Rectifier (RELU) and sigmoid functions are chosen as activation functions for hidden and output layers respectively.
d) Backward Propagation: The algorithm compares the actual result with the predictive result and the error is generated which is then back propagated from output side to input side and all weights are updated by learning rate parameter according to the error that is generated.
e) Steps 'a' to 'd' are repeated either after each observation or after batch of observations. When the whole training set pass through ANN that makes an epoch.

Now fitting ANN into training set has to be done since only ANN has been built without making the connections to the training set. The last execution to be done is predicting the test set by making the confusion matrix. If the accuracies obtained from trained and test sets are equal then that completes validation of a model.

V. PROPOSED WORK

This research work interrelates VLSI architectures particularly compressors and machine learning. The 5:2 compressor proposed in [7] has been considered for training the neural network. Training of neural network (NN) as proposed approximate 5:2 compressor has been performed by taking the truth table of the approximate 5:2 compressor as dataset. The implementation of training part of the NN has been made using supervised machine learning algorithm. The dataset of approximate 5:2 compressor proposed in [7] has been considered as input to train the NN. Fig. 3 shows the neural network as proposed 5:2 compressor with the input as truth table. C_{in1}, C_{in2}, X_1, X_2, X_3, X_4, X5 are the inputs to the NN from the truth table and these inputs passes through the two hidden layers to get the final output.

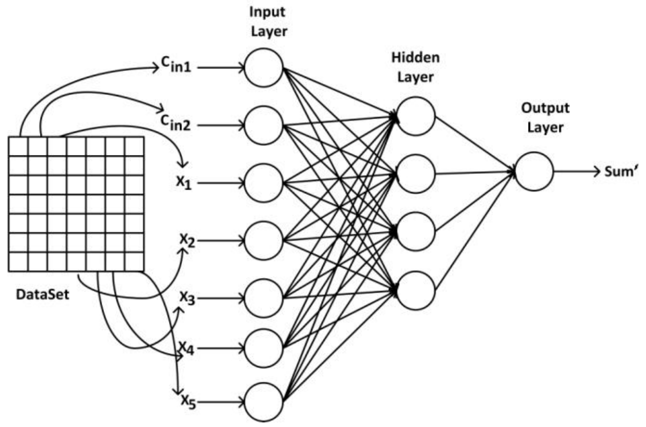

Fig. 3: Neural Network as 5:2 compressor

Since classification algorithm is not suitable for multiple binary outputs, only Sum output has been considered for execution.

VI. SIMULATIONS AND DISCUSSIONS

The execution of ML application to approximate compressors has been done using Anaconda Python with Jupyter Notebook as Integrated Development Environment (IDE). Data pre-processing, the first part of ML application (classification problem), has been performed importing numpy and pandas libraries and, ANN has been built (second part of ML application) importing Keras library with TensorFlow backend.

A. Building Artificial Neural Network

After data pre-processing and building ANN, training has been accomplished in Keras library for 1000 epochs (is a round when the whole training set passed through ANN) which gives accuracy value for each of the approximate 5:2 compressor. Lastly, predictions on the test sets has been performed using predict classifier of Keras library which conveys the probabilities of the incorrect output bits when compared with the exact ones. Importing the Confusion Matrix (CM) from Scikit-learn library (a machine learning library) gives the number of correct and incorrect predictions from which accuracy of the test set can be calculated. The accuracy evaluation has been performed on the exact and proposed 5:2 compressors by providing seven inputs (C_{in1}, C_{in2}, X_1, X_2, X_3, X_4, X_5) and Sum output.

B. Training of Exact 5:2 Compressor as Neural Network

Once the ANN has been fit in to the training set, the accuracy of the train set for an exact 5:2 compressor with Sum output is obtained as 55.88% for 1000 epochs and by executing Scikit library, CM observations are obtained as

$$\mathrm{CM} = \begin{pmatrix} 0 & 14 \\ 1 & 11 \end{pmatrix}$$

The accuracy of the test set is calculated with eq. (1),

$$\text{Accuracy} = \frac{\text{No. of correct predictions}}{\text{Total no. of predictions}} \tag{1}$$

$$\text{No. of correct predictions} = 0 + 11 = 11$$
$$\text{No. of incorrect predictions} = 1 + 14 = 15$$

$$\text{Total no. of predictions} = 11 + 15 = 26$$

Thus, accuracy is obtained as 42.307% from eq. (1). For the complete validation of a model, train and test set accuracies should be equal but here the difference in accuracies between them is 13.57%. The accuracy and loss graphs for the exact 5:2 compressor with sum output in the dataset are shown in Fig. 4. and Fig. 5 respectively.

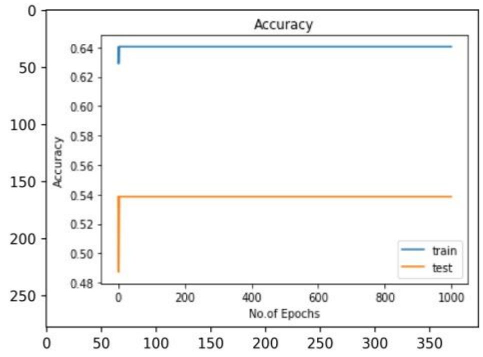

Fig. 4 : Train and Test set Accuracy representation of an exact 5:2 compressor

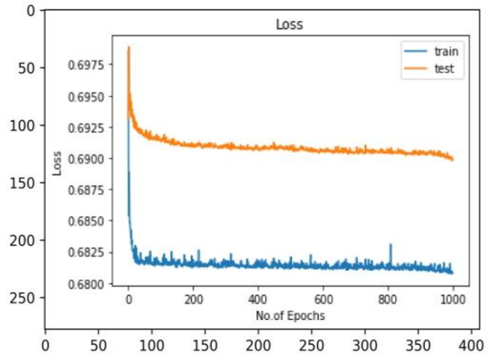

Fig. 5 : Train and Test set Loss representation of an exact 5:2 compressor

Evaluation of ANN employing K4 cross validation has been performed on the same dataset to make the difference between train and test sets much lesser, and obtained the trained accuracy as 48.91% after evaluation. Now the difference between the train (after evaluation) and test sets can be found as 6.603% which is less when compared with the previous difference. The accuracy and loss graphs for the exact 5:2 compressor with sum output in the dataset after evaluation are shown in Fig. 6. and Fig. 7 respectively.

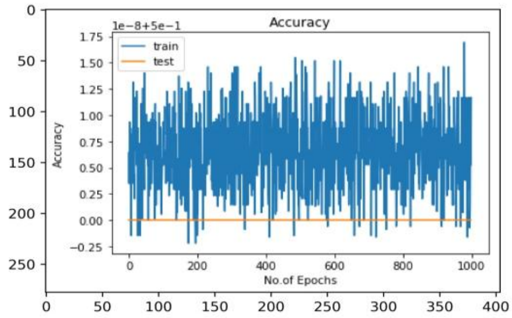

Fig. 6 : Train and Test set Accuracy representation of an exact 5:2 compressor after evaluation

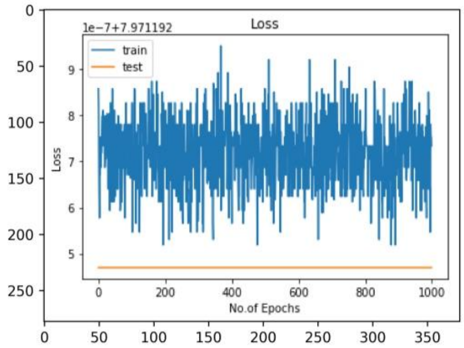

Fig. 7 : Train and Test set Loss representation of an exact 5:2 compressor after evaluation

C. Training of proposed 5:2 Compressor as neural network

The accuracy of the train set for the proposed 5:2 compressor with Sum output is obtained as 52.94% for 1000 epochs confusion matrix observations are obtained as

$$CM = \begin{pmatrix} 10 & 0 \\ 16 & 0 \end{pmatrix}$$

Hence accuracy of the test set is found to be 38.46%. The accuracy and loss graphs for the proposed 5:2 compressor with Sum output in the dataset are shown in Fig.8. and Fig. 9 respectively

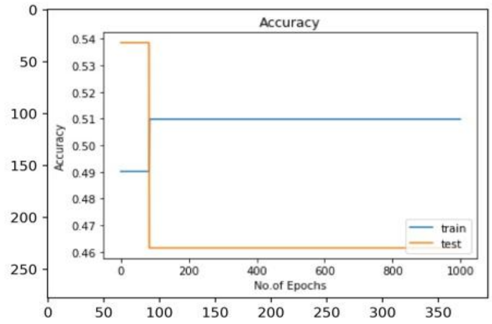

Fig. 8 : Train and Test set Accuracy representation of the proposed 5:2 compressor

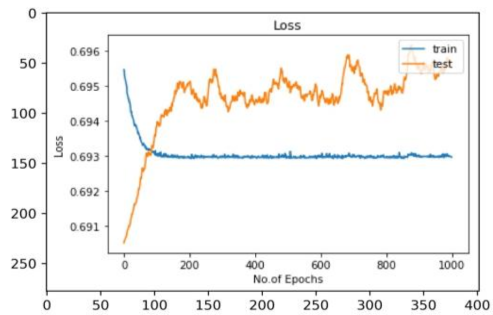

Fig. 9 : Train and Test set Loss representation of the proposed 5:2 compressor

Retrieval Number J93330881019/2019©BEIESP
DOI: 10.35940/ijitee.J9333.0881019

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The difference between the train and test set accuracies is high which is14.48%. Thus, evaluation has been performed using “K4 Cross Validation” from which the trained accuracy is obtained as 46.74% and the difference between evaluated train accuracy and test set accuracy is 8.28%. The accuracy and loss graphs for the proposed 5:2 compressor with Sum output in the dataset after evaluation are shown in Fig. 10 and Fig. 11 respectively.

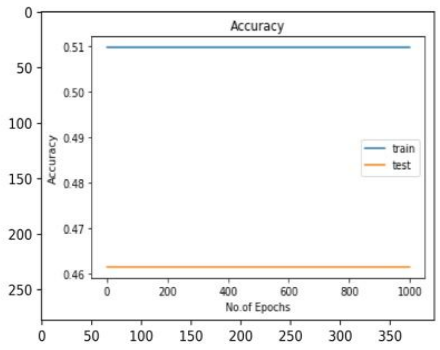

Fig. 10: Train and Test set Accuracy representation of the proposed 5:2 compressor after evaluation

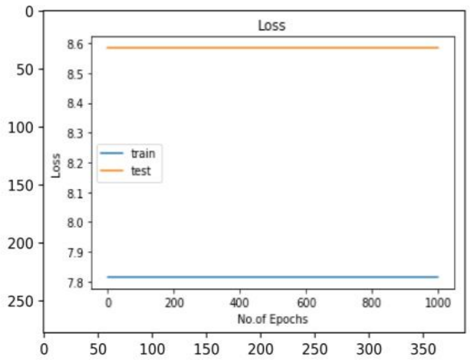

Fig. 11: Train and Test set Loss representation of the proposed 5:2 compressor after evaluation

The mean, variance, difference1 and difference2 of the exact and proposed 5:2 compressors are tabulated in Table I, where the difference1 indicates the difference between the train (before evaluation) and test set accuracies of the exact and proposed 5:2 compressors and difference2 indicates the difference between the train (after evaluating using K4 cross validation) and test set accuracies of the exact and proposed 5:2 compressors.

TABLE I: MEAN, VARIANCE AND DIFFERENCES OF EXACT AND PROPOSED 5:2 COMPRESSORS

Parameters	Exact	Proposed
Mean	0.51	0.47
Variance	0.154	0.135
Difference 1	13.57%	14.48%
Difference 2	6.60%	8.28%

Tuning of parameters has been done on exact and proposed 5:2 compressors to observe the best batch size, number of epochs and the optimizer. The same has been depicted in Table II.

TABLE II: BEST PARAMETERS OF EXACT AND PROPOSED 5:2 COMPRESSORS

Best Parameters	Exact	Proposed
Batch Size	32	25
Epochs	500	100
Optimizer	RMS prop	RMS prop

XIII. CONCLUSIONS

The principal object of the research work of this paper is to relate the approximate compressor circuits to machine learning applications. For the proposed approximate compressor, the accuracies of the train and test sets are approximately equal and, the error rate and transistor count was less which confirms that the connection between digital logic circuits and machine learning is valid. From this final outcome, the advantage that can be achieved is automation of approximate digital circuits. For automatic selection of the best approximate compressor, reinforcement learning algorithm can be used such that the improved approximate compressor circuit is placed in arithmetic circuits like multipliers where these approximate multipliers can be placed in multimedia and image processing applications.

ACKNOWLEDGEMENTS

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

REFERENCES

1. Veeramachaneni, et.al., “Novel architectures for high-speed and low-power 3-2, 4-2 and 5-2 compressors,” Proc. Int. Conf. on VLSI Design (VLSID), 2007, pp. 324-329.
2. Chang, et.al, “Ultra low voltage, low power CMOS 4-2 and 5-2 compressors for fast arithmetic circuits”, IEEE Trans. Circuits Syst. I, Fundam. Theory Appl., 2004, 51, (10), pp. 1985–1997.
3. Saha.P., Samantha, P., kumar, D., 2016, “ 4:2 and 5:2 Decimal Compressors,” proc. 7th International Conference on Intelligent Systems, Modelling and Simulation., Bangkok, Thailand, pp. 424-429.
4. Mehdi Ghasemzadeh, Sina Mahdavi, Abolfazl Zokaei, Khayrollah Hadidi, “A New Ultra High Speed 5-2 Compressor with a New Structure”, 23rd International Conference Mixed Design of Integrated Circuits and Systems, June 23-25, 2016, Poland.
5. D.Balobas and N.Konofaos, "Low- power high-performance CMOS 5-2 compressor with 58 transistors", Electronics Letters, Institution of Engineering and Technology, 2018, 54(5), pp. 278-280.
6. Lavanya. M, Ranjan K. Senapati, JVR Ravindra, "Low-Power Near-Explicit 5:2 Compressor for Superior Performance Multipliers", International Journal of Engineering Research and Technology, Vol(11), No.(4), 2018.
7. Lavanya.M, Ravindra JVR, “Performance metrics of Inexact Multipliers based on Approximate 5:2 Compressors”, 15th International SOC Design Conference (ISOCC), 2018, Daegu, South Korea, pp. 84-85.

8. Ankur.A, et.al, “Approximate Computing: Challenges And Opportunities,” IEEE International Conference on Rebooting Computing (ICRC), 2016, San Diego, USA.
9. Momeni.A, et.al, “Design and analysis of approximate compressors for multiplication,” IEEE Transactions on Comp., 64 (4), 2015, pp. 984–994.
10. Suganthi.V, et.al, “Design of Power and Area Efficient Approximate Multipliers,” IEEE Transactions on VLSI Systems, 2016, pp.1-5.
11. Darjn.E, et.al, “Approximate Multipliers Based on New Approximate Compressors,” IEEE Transactions on CAS-I: Regular Papers, PP (99), 2018, pp. 1-14.
12. Avinash.L, et.al, “Parsimonious Circuits for Error-Tolerant Applications through Probabilistic Logic Minimization”, International Workshop on PATMOS 2011, pp.204-213.
13. Liang.J, et.al, “New Metrics for the Reliability of Approximate and Probabilistic Adders,” IEEE Transactions on Computers, 63(9), 2013, pp. 1760-1771.
14. Zervakis.G, et.al, “Design-Efficient Approximate Multiplication Circuits through Partial Product Perforation,” IEEE Trans. on VLSI Systems, 24(10), 2016, pp. 3105-3117.
15. Peter. A, et.al, “Opportunities for Machine Learning in Electronic Design Automation," in ISCAS, 2018, Italy.
16. Ilaria.S, et.al, “Circuit carving: A methodology for the design of approximate hardware,” Design, Automation & Test in Europe Conference (DATE), 2018, Germany.
17. Schlachter.J et.al, “Design and Applications of Approximate Circuits by Gate-Level Pruning,” IEEE Transactions on VLSI Systems 25 (5), 2017, pp.1694-1702.
18. Sashi Obilishetty, “Digital Intelligence and Chip Design,” International Symposium on VLSI Design, Automation and Test (VLSI-DAT), 2018, Taiwan.
19. Kouretas, et.al, “Logarithmic number system for deep learning,” 7th International Conference on Modern Circuits and Systems Tech. (MOCAST), 2018, Greece

AUTHORS PROFILE

Lavanya Maddisetti is a research scholar pursuing Ph.D in KL Deemed to be Uiversity, Guntur, India. She has received B.Tech and M.Tech degrees from Jawaharlal Nehru Technological University, Hyderabad (JNTUH), India. She has three papers in IEEE Conferences and one Scopus Indexed Journal with one Indian patent. Her research interests include Low Power VLSI Circuits, Approximate Circuits, Machine Learning and Magnetic Tunnel Junction (MTJ) Logic circuits.

Ranjan K. Senapati is a professor in the department of ECE, KL Deemed to be University. He is a member of IEEE. He has published more than 50 papersin scopus and SCI indexing journals.

JVR Ravindra (S’01, M’08, SM’18) received the B.E. degree in Electronics and Communication Engineering from Osmania University, Hyderabad, India and Ph.D. degree from the Center for VLSI and Embedded Systems Technologies (CVEST), at International Institute of Information Technology (IIIT-H), Hyderabad, India in 2011. He joined the Faculty of Electronics and Communication Engineering, Vardhaman College of Engineering, Hyderabad, India where he is currently Professor and Dean (Acads.). He has authored more than 100 papers in refereed journals and conferences, holds three patents, graduated one Ph.D. student. His current research interests include Approximate Computing using Machine Learning, Modeling of Ultra Low power interconnects, Designing of High-Speed and Low power Arithmetic Circuits, Low Power DSP Architectures, Hardware Security, Wireless Sensor Networks.