



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 5, May 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Optimized Multiplier-Less FFT Structures for High Performance Machine Learning Hardware

Modalavalasa Hari Krishna¹, Makkena Madhavi Latha²

Ph.D Scholar, Department of ECE, JNTUH-CEH, Jawaharlal Nehru Technological University, Hyderabad, India¹

Professor, Department of ECE, JNTUH-CEH, Jawaharlal Nehru Technological University, Hyderabad, India²

ABSTRACT: Machine learning (ML) provides the intelligent systems with the ability to make autonomous decisions. Many advanced Machine Learning algorithms are developed for complex real-time applications. Complex Machine Learning algorithms require huge computational and storage hardware resources. Convolution is the one of the most used operation in machine learning algorithms, particularly in deep learning 90% computations of the algorithm are convolutions. FFT is the one of the best approaches to reduce the computational complexity of the convolution. Among the different FFT architectures Pipelined single-path delay feedback (SDF) FFT has lower memory and power requirements. In Radix-n SDF FFT structures complex twiddle-multiplications are the most time and power consuming operations. In this paper novel multiplier-less structures are proposed for delay, area and power efficient FFT operations. In this approach complex multipliers are replaced with add-shift modules and dedicated signal routing between real and imaginary components. Three novel VLSI structures are developed for multiplication with the twiddle factors $(-i, 0.707-0.707i$ and $0.707+0.707i)$ with delay, area and power optimizations. All the three proposed modules are developed in Verilog and compared against existing multiplier-based twiddle factor multiplications in terms of power, area and delay. Proposed optimized modules can reduce the major constraints of machine learning hardware (delay and memory) and reduce the training period by improving the operating frequency of ML Hardware.

KEYWORDS: FFT, Machine Learning Hardware, Multiplier-less FFT, Single-path Delay Feedback FFT.

I. INTRODUCTION

Artificial Intelligence (AI) is one of the major research areas in present technological world. Machine Learning (ML) algorithms act as brain of AI systems to make intelligent decisions without the need of traditional programming or manual support [1]. ML algorithms explore the data to understand the hidden structures and extract the hidden relationships to create ML models. These ML models can understand the future inputs and predict the corresponding future output using the previously extracted relationships or knowledge. ML algorithms are classified into three broader classes such as Supervised Machine Learning (SML), Unsupervised Machine Learning (UML) and Reinforcement Machine Learning (RML). SML algorithms are more powerful and used for both regression and classification but they required labelled data. If data is present but labels are not available then UML algorithms can be applied to classify data by understanding the internal relationships among the data samples. Semi-Supervised algorithms are there which are combination of UML with small portion of SML algorithms to improve the accuracy in absence of proper labelled datasets. RML algorithms are only possible option in absence of any dataset prior to deployment of ML models into real application environment. These RML models are very useful to learn from unknown environments which don't have any statistical nature [1][2]. Deep Learning (DL) is the subset of ML which has neural networks structures in its algorithms. As the name suggests DL algorithms dig deep into data without need of any prior feature extraction. The accuracy of the ML or DL models depends the algorithms and data. Mainly complex algorithms with huge datasets exhibit best performance. But the training period of complex algorithms with huge datasets is very high and requires specialized hardware. Existing hardware resources such as Central Processing Units (CPUs) and Graphical Processing units (GPUs) are not enough to handle complex deep learning algorithms. Most of the existing developers using clusters of GPUs or cloud computing for training of their models. Latest mobile platform processors come with a combination of CPUs and GPUs which can handle the inference complexity of tiny ML models. A lot of research is going on development of algorithms with controlled-complexity and high-performance hardware with high computational and memory resources. Most of the ML and DL algorithms requires huge amount of

convolution operations in their structures. In Convolutional Neural Networks (CNN), 90% of total computations are convolution operations. This shows the computational complexity of DL algorithms and number of computations (importantly multiplications) required [3][4].

II. OBJECTIVES AND GOALS

The accuracy of ML or DL models depends on training procedure and data. Precise and detailed training requires huge training time up to few years. The possible two ways under research to reduce this training time are controlling of complexity of algorithms and improving the capabilities of the hardware. Apart from these two another new approach is optimization of implementation of algorithm without changing its methodology by basic components in its structure [5][6]. As most of computations in ML and DL algorithms are convolutions, we focused on convolution operation. Fast Fourier Transform (FFT) can be used to reduce the intense computations complexity of the convolution. Single-path Delay Feedback (SDF) FFT are more optimized than memory based FFT structures in terms of area and power [7][8].

Radix-n SDF FFT can be implemented without any extra multipliers except for twiddle multiplication modules

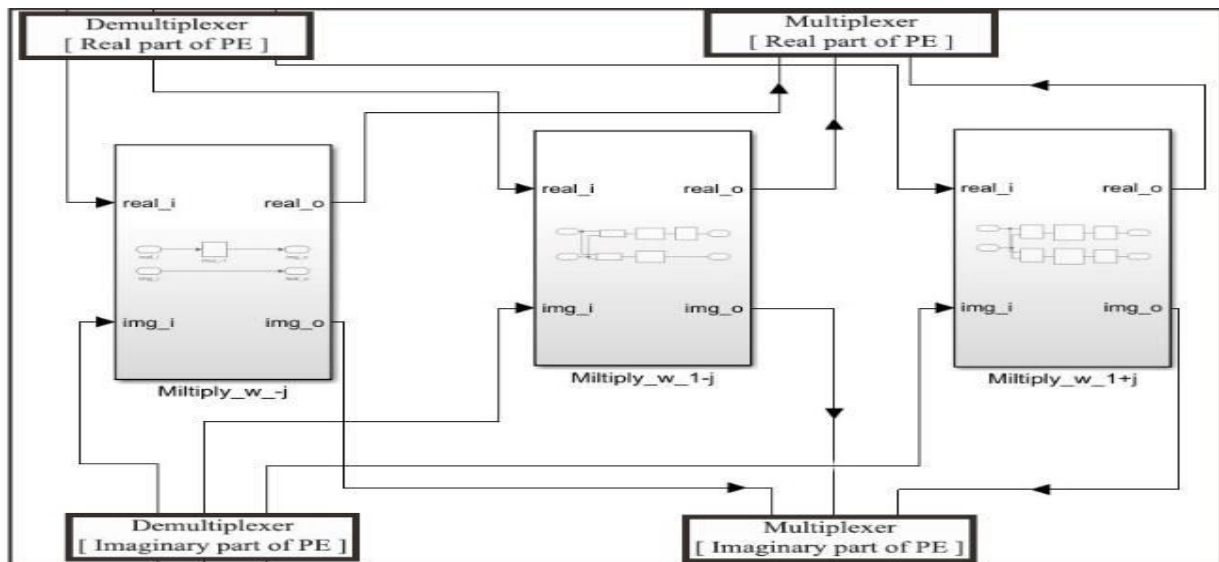


Fig.1. Real and imaginary signal routing between real and imaginary computations (demultiplexer to multiplexer) of each PE in Pipelined single-path delay feedback FFT

between demultiplexer and multiplexer inside each Processing Unit (PE) as shown in fig.1. This twiddle factor multiplication requires complex multipliers which consumes huge silicon area and more computational cycles. This computational overload from Twiddle multiplication is very high as the butterfly structures are basic blocks of FFT [9][10]. In this paper novel VLSI blocks are designed to perform twiddle multiplication without using any complex multipliers.

III. METHODOLOGY

The complex multiplier modules in twiddle multiplication of SDF FFT are replaced with dedicated VLSI modules for each twiddle factor. In proposed modules complex multipliers are replaced with combination of customized bit-shifters, adders and novel signal routing between real and image components of complex values. In this paper optimized VLSI modules are designed for multiplication with major twiddle factors. First proposed module is to multiply complex signal with i (or $-i$), second module is to multiply with $0.707*(i-1)$ and third module is to multiply with $0.707*(i+1)$. The fig.2 shows the implementation of module-1 (multiply with $-i$) in both existing (with multiplier) and proposed (without multiplier) methodologies. Fig.3 and fig.4 show the block diagram representation of module-2 (multiply with $0.707*(i-1)$) and module-3 (multiply with $0.707*(i+1)$) in both with and without multiplier designs where $x0.707$ block in proposed designs shows the actual multiplier logic for multiplication with 0.707 and shown in fig.5. The novel data routing between real and imaginary components of the complex signal is clearly represented using the arrows in the block diagrams of the proposed multiplier-less designs. All the routing logics are derived from the basic multiplication concepts of Complex values and multiplication concepts are designed with fundamental rules of Boolean algebra using shifters and adder networks.

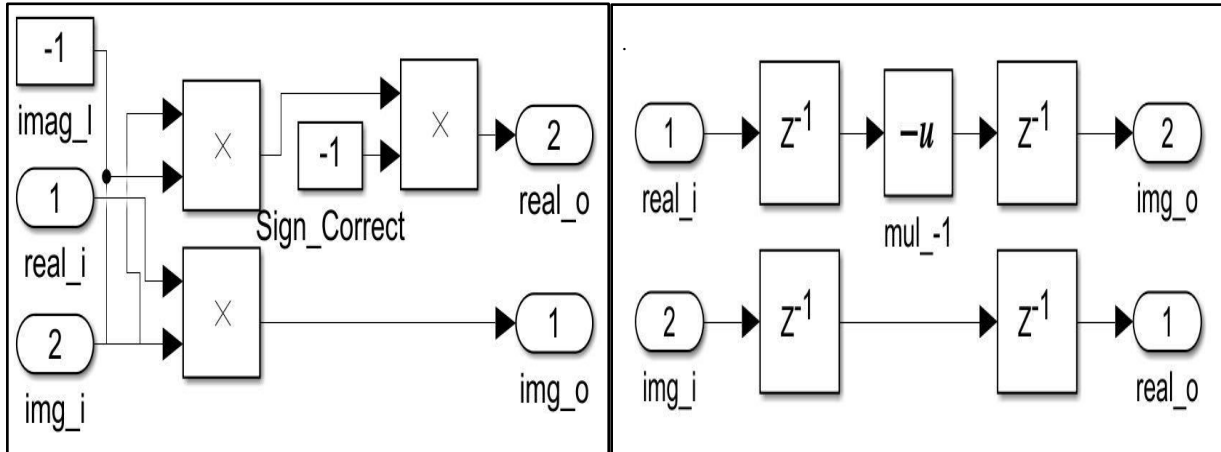


Fig.2. Block diagrams of module-1 (Multiplication with $-i$) with multiplier (Left side) and without multiplier (Right side)

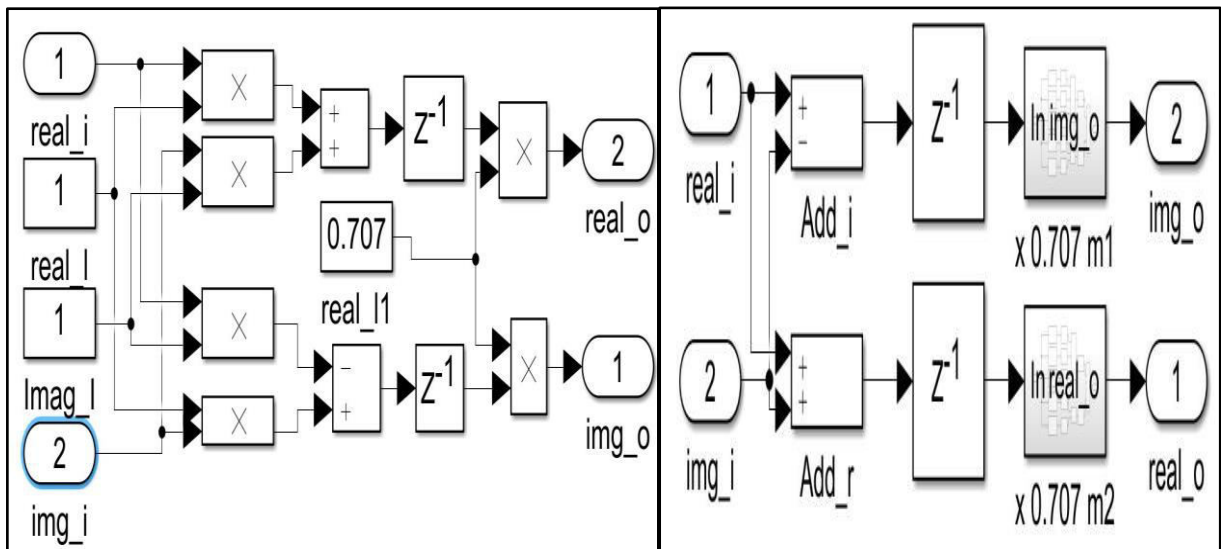


Fig.3. Block diagrams of module-2 (Multiplication with $0.707-i*0.707$) with multiplier (Left side) and without multiplier (Right side)

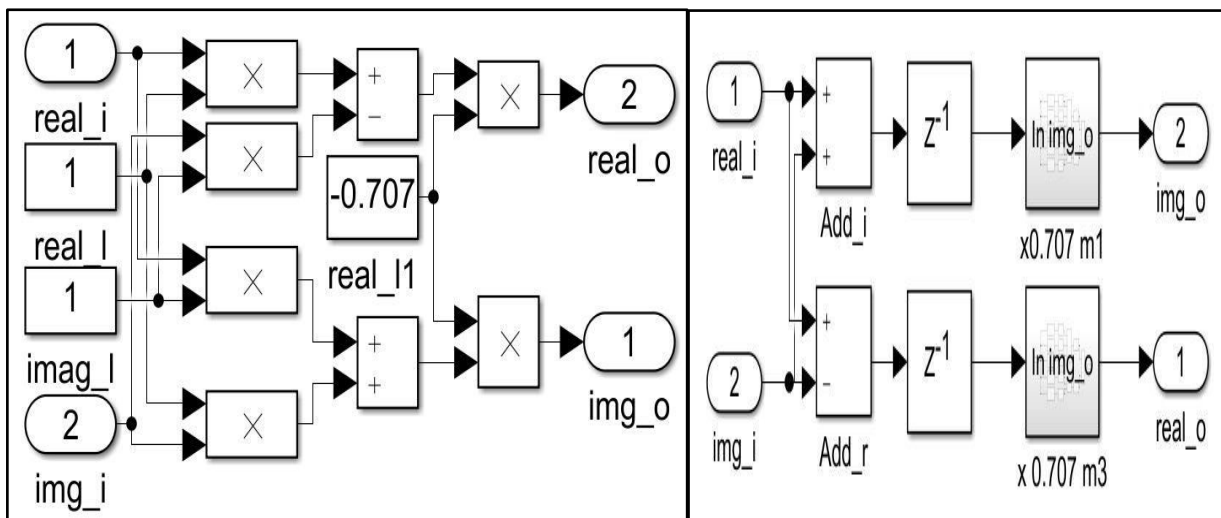


Fig.4. Block diagrams of module-3 (Multiplication with $0.707+i*0.707$) with multiplier (Left side) and without multiplier (Right side)

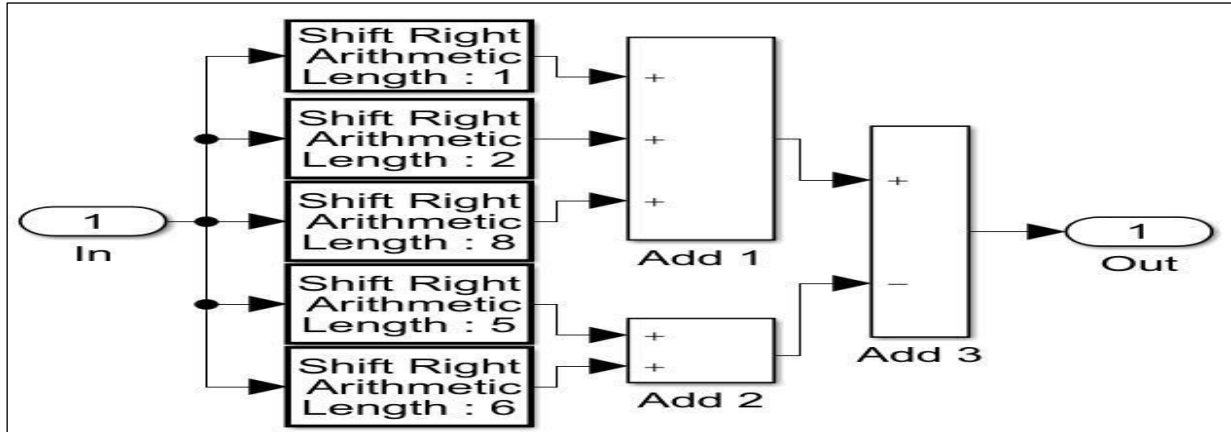


Fig.5. Block diagram of proposed multiplication module with gain factor 0.707 using shifters and adders.

IV. RESULTS AND DISCUSSIONS

The three multiplier modules are designed in both existing multiplier-based and proposed multiplier-less methodologies using the XILINX VIVADO 2019.2 tools with Verilog programming. Functionality of the designed modules are verified in Vivado simulator using complex signals. Block diagrams of Model-1 shown in fig-2 are designed using Verilog and elaborated RTL schematic designs are shown in fig.6 (Multiplier-based design) and fig.7 (Multiplier-less design) where all the inputs are taken as 16-bit fixed digit floating values. Fig.8 and fig.9 represents the Vivado elaborated RTL schematic designs of multiplier-based and multiplier-less models of module-2 shown in fig.3. The elaborated RTL schematic diagrams of the module-3, using the multiplier-based methodology is shown in fig.10 and using multiplier-less methodology is shown in fig.11.

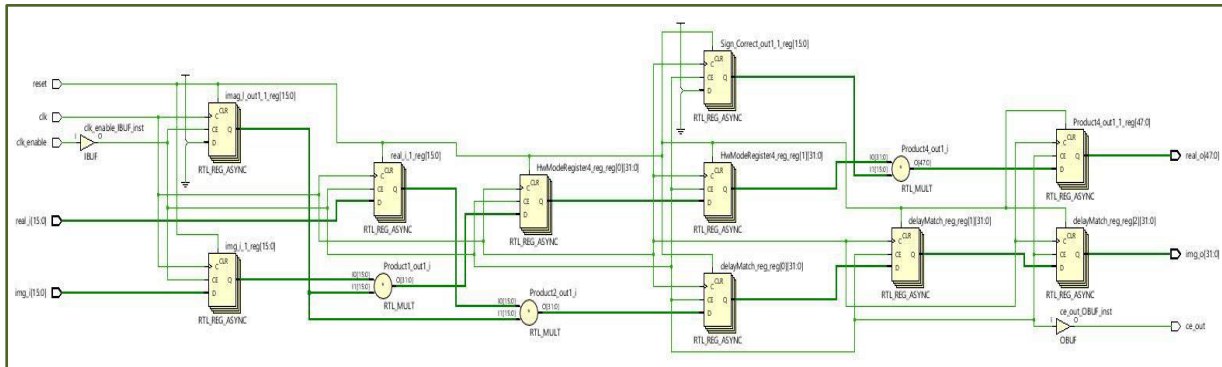


Fig.6. Elaborated RTL schematic diagram of module-1 (Multiplication using -i) using existing multiplier-based design.

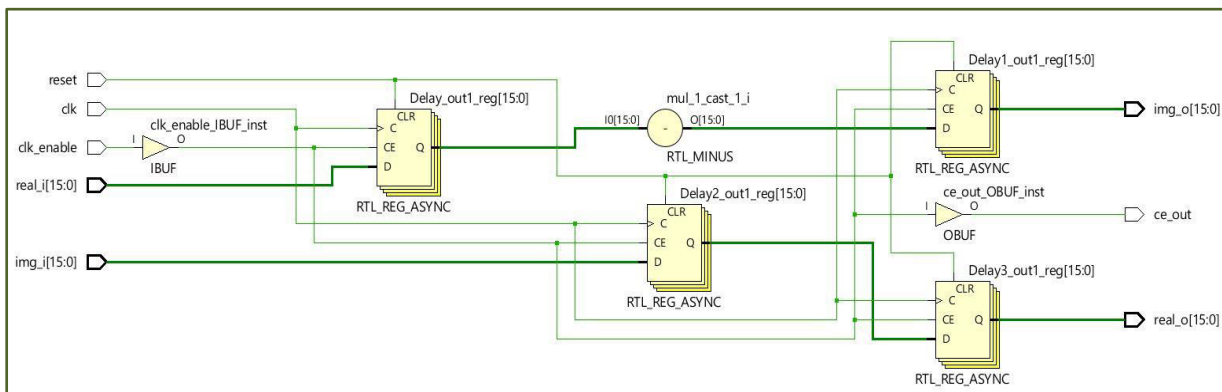


Fig.7. Elaborated RTL schematic diagram of module-1 (Multiplication using -i) using proposed multiplier-less design

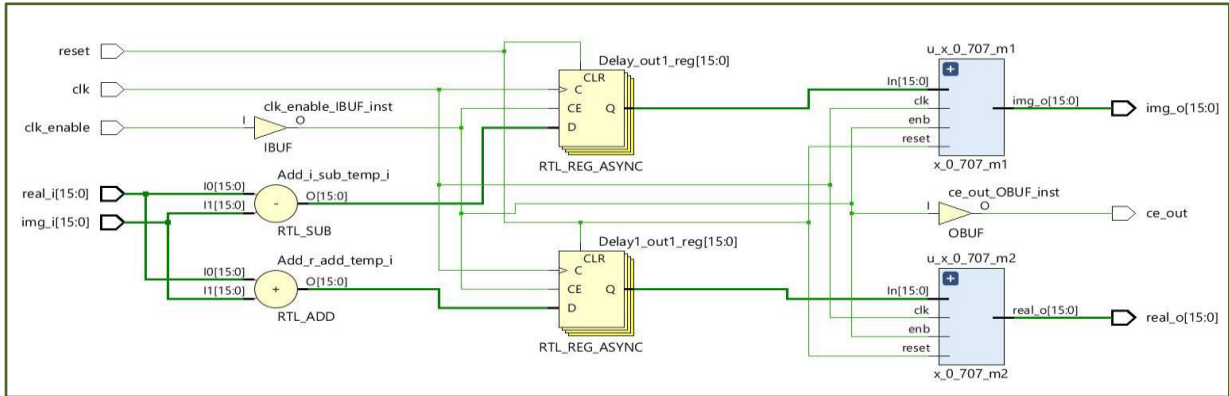


Fig.8. Elaborated RTL schematic diagram of module-2 (Multiplication using $0.707*(i-1)$) using existing multiplier-based design

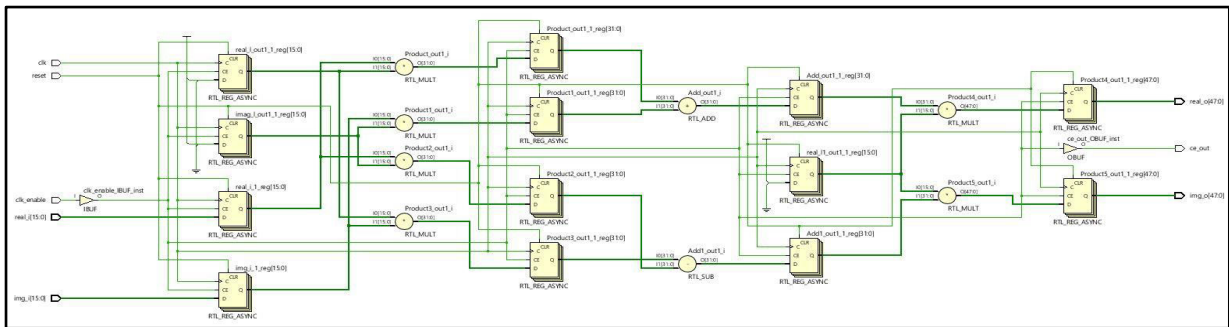


Fig.9. Elaborated RTL schematic diagram of module-1 (Multiplication using $0.707*(i-1)$) using proposed multiplier-less design

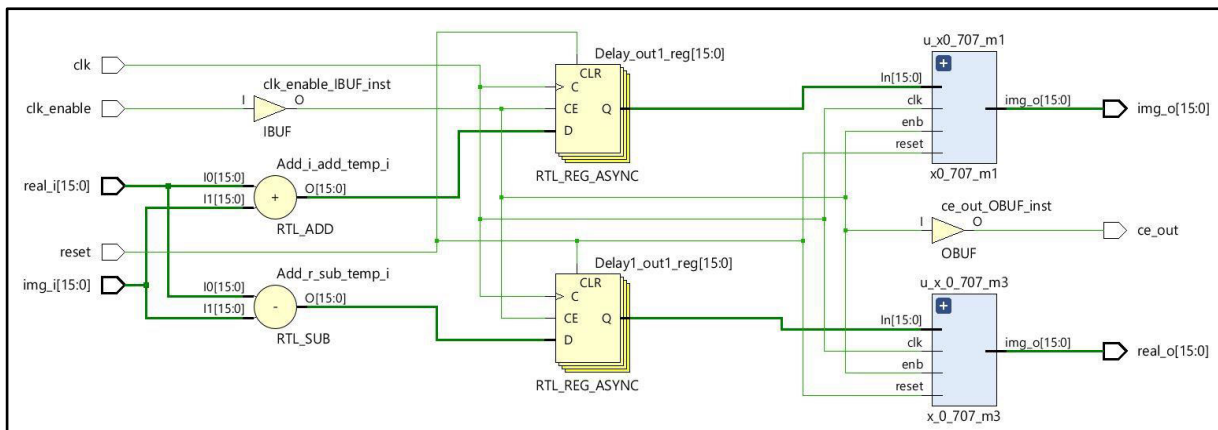


Fig.10. Elaborated RTL schematic diagram of module-3 (Multiplication using $0.707*(i+1)$) using existing multiplier-based design

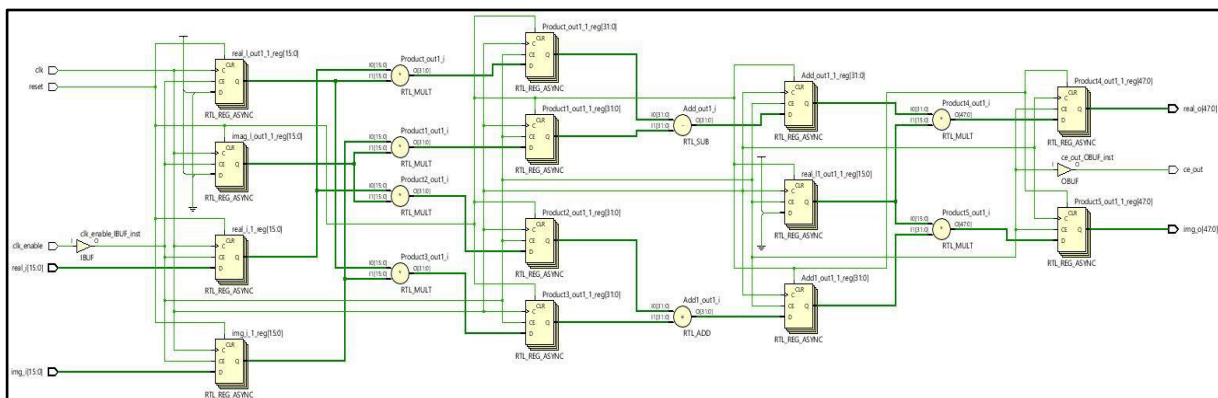


Fig.11. Elaborated RTL schematic diagram of module-3 (Multiplication using $0.707*(i+1)$) using proposed multiplier-less design

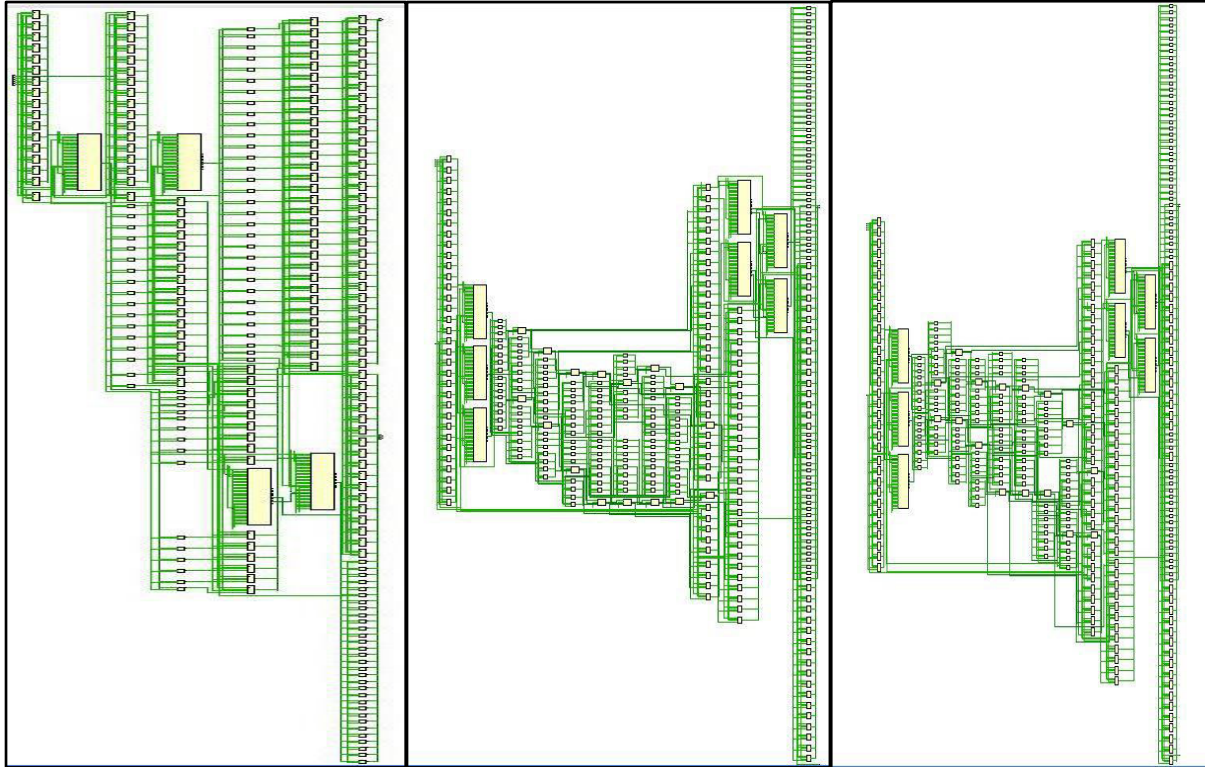


Fig.12. Synthesized technology schematic diagrams of module-1 (left), module-2 (middle) and module-3 (right) in existing multiplier-based design.

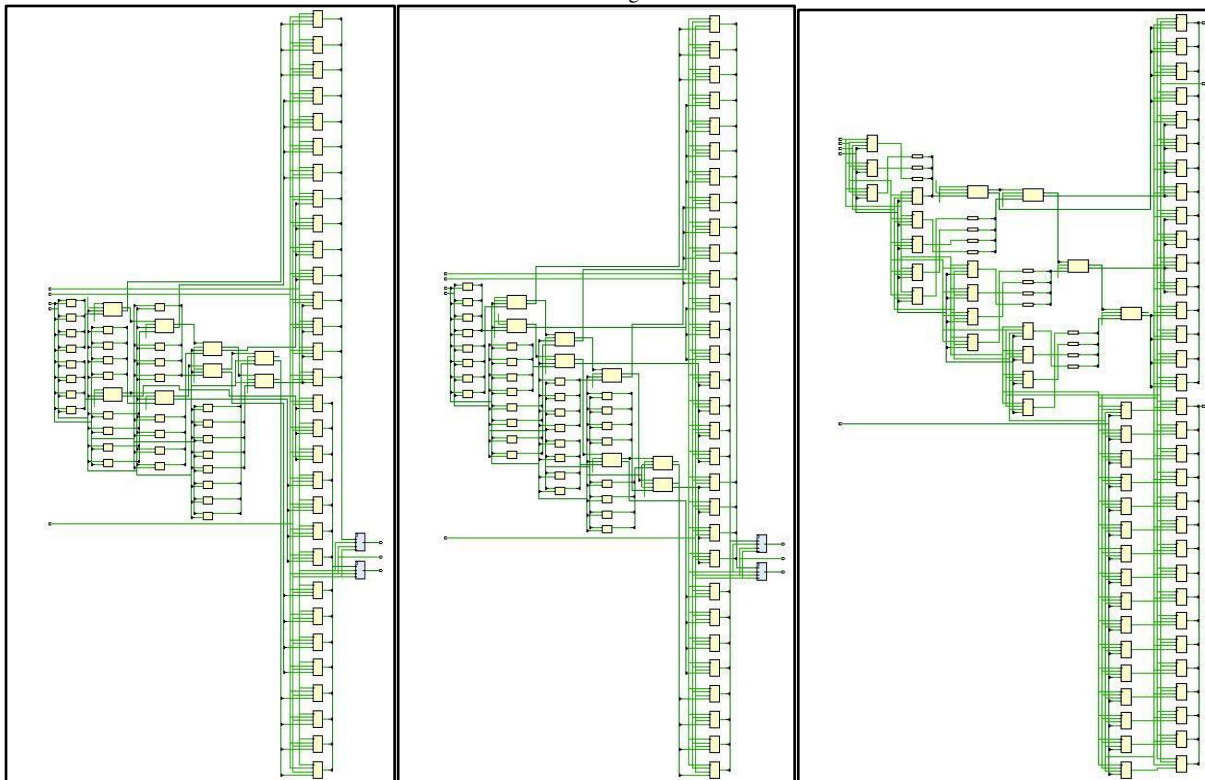


Fig.13. Synthesized technology schematic diagrams of module-1 (left), module-2 (middle) and module-3 (right) in proposed multiplier-less design.

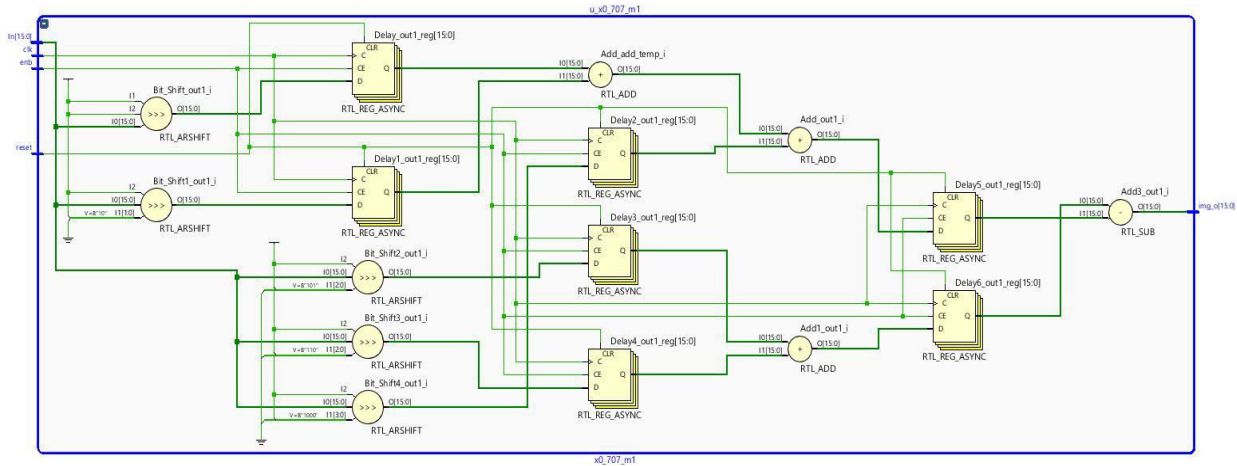


Fig.14. Elaborated RTL schematic diagram of inner multiplication module in proposed modules-2 and module-3 with gain factor 0.707

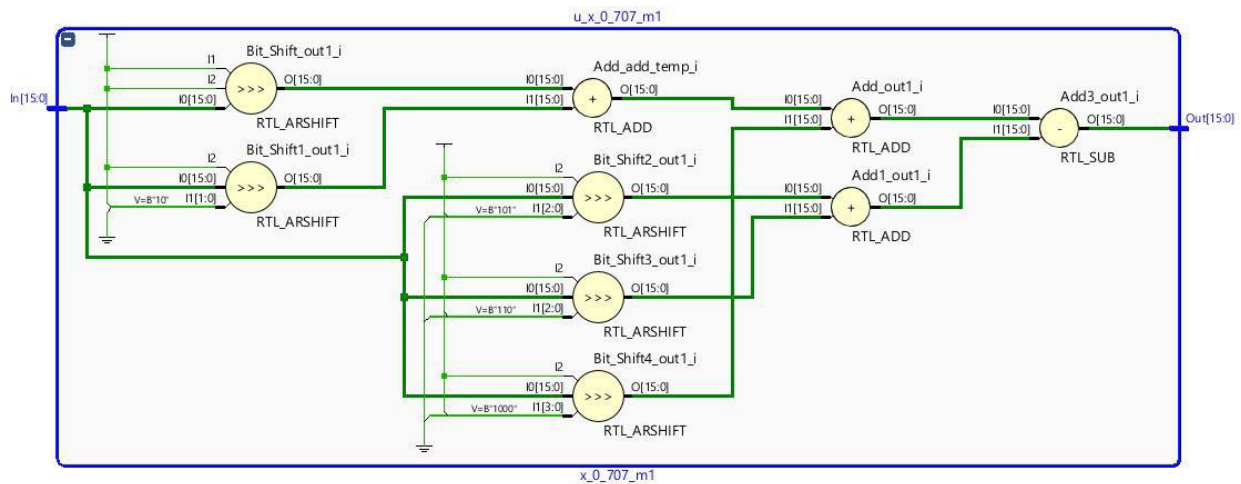


Fig.15. Synthesized technology schematic diagram of inner multiplication module in proposed modules-2 and module-3 with gain factor 0.707

The designed modules are synthesized using Vivado synthesizer to perform a comparative analysis between existing and proposed modules in terms of the power, area and delay. The technology schematic diagrams are mapped for Xilinx’s Zynq series Zed board. The technology schematics of all the 3 multiplier modules (Module-1, 2 & 3) in existing multiplier-based design are shown in fig.12. In the same technology mapping, the technology schematic diagrams of same multiplier modules (Module 1, 2 & 3) in proposed multiplier-less design are shown in fig.13. The reduce complexity in terms of the resource utilization or need of silicon area can be clearly observed from fig.12 & 13.

Table-1
Table.1. Comparative analysis in terms of Area, power and delay between multiplier-based and multiplier-less designs

Module - Design \ parameter(Area, Power & Delay)	Slice LUTs (17600)	Slices (4400)	Resisters (35200)	DSPs (80)	Dynamic power (watt)	Static Power (watt)	Worst Negative Slack (ns)	Worst Hold Slack (ns)	Worst pulse width Slack (ns)	Data Path Delay (ns)
Module-1 with multiplier	48	30	147	4	0.002	0.09	12.525	0.136	9.5	5.991
Module-2 with multiplier	95	31	132	7	0.003	0.09	12.347	0.268	9.5	6.169
Module-3 with multiplier	96	37	133	7	0.003	0.09	12.499	0.372	9.5	6.017
Module-1 without multiplier	15	4	64	0	0.001	0.09	17.844	0.173	9.5	2.134
Module-2 without multiplier	112	35	116	0	0.002	0.09	15.865	0.173	9.5	3.853
Module-3 without multiplier	112	37	116	0	0.002	0.09	15.947	0.274	9.5	3.819

The inner bit-shifter and adder module for multiplication with 0.707 are synthesized separately and RTL schematic is shown in fig-14 and technology schematic for Zed board device is shown in fig.15. The detailed design analysis is performed after implementation of above modules for delay and power reports along with utilization reports and shown in table.1. From column-2 to column-4 of table-1, the resource utilization of proposed design is less than existing design which represents the savings in silicon area. Mainly the proposed designs not using any DSP units (Critical signal processing units in Zynq FPGAs) due to their multiplier less design and results in 0% utilization whereas multiplier-based designs consuming 8.8% total DSPs for single module and more than 23% of DSPs for one PE, which shows the critical resource savings from proposed designs. Along with the area, the proposed multiplier-less designs resulted in 50% savings in dynamic power with better signal routing and reducing the delay by 50% and improves the operating frequency of the entire design.

IV. CONCLUSION

High performance Radix-n single-path delay feedback FFT with lower memory requirement, is considered for high performance Machine Learning hardware. The major twiddle factor multiplication modules in Radix-n SDF FFT are designed in Verilog, synthesized using Vivado tools and implemented on Xilinx's Zynq series Zed board. The performance of the proposed multiplier-less design is compared and analysed against the multiplier-based design in terms of three major design constraints (area, power and delay). For all the 3 twiddle multipliers, the multiplier-based design consuming 23% of total DSP units where as proposed multiplier-less designs results in zero consumption of DSP units. The overall silicon area consumption also less than the existing design with greater savings in registers. Along with area, the proposed designs exhibiting 33% savings in dynamic power with dedicated signal routing between real and imaginary components. The delay is reduced by 50% using the multiplier-less design and results in the double operating frequency of the overall FFT. Using this improved operating frequency, the training speed of the entire machine learning or deep learning hardware will be improved and result in lower training periods.

REFERENCES

- [1] John D. Kelleher; Brendan Tierney, "MACHINE LEARNING 101, in Data Science", MITP, pp.97-150, 2018.
- [2] E. Alpaydin, "Introduction to Machine Learning", in MIT Press, Cambridge, Mass, USA, 2nd edition, 2010.
- [3] Vivienne Sze, Yu-Hsin Chen, Joel Emer, Amr Suleiman, Zhengdong Zhang, "Hardware for Machine Learning: Challenges and Opportunities", in doi: 5v52670.2161:viXra, Oct-2017.
- [4] M.Hari Krishna, M.Madhavi Latha, "VLSI Technology in Design of Future Hardware for Machine Learning: A Review", in SHODH SARITA: An International Bilingual Peer Reviewed Refereed Research Journal, Pp. 32-36, Vol. 4, Issue 29, ISSN – 2348-2397, January to March 2021.
- [5] D.PadmaShri, D.SubbaRao, G Vijay Goud, "VLSI Design of High Performance Complex Multiplier", in International Refereed Journal of Engineering and Science (IRJES), Volume 1, Issue 4, PP.68-75, e-ISSN: 2319-183X, p-ISSN: 2319-1821, 2015.
- [6] S. Lin et al., "FFT-based deep learning deployment in embedded systems," 2018 Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 1045-1050, doi: 10.23919/DATE.2018.8342166, 2018.
- [7] Y. S. Algnabi, F. A. Aldaamee, R. Teymourzadeh, M. Othman and M. S. Islam, "Novel architecture of pipeline Radix 22 SDF FFT Based on digit-slicing technique," 2012 10th IEEE International Conference on Semiconductor Electronics (ICSE), pp. 470-474, doi: 10.1109/SMElec.2012.6417188, 2012.
- [8] C. Ingemarsson, P. Källström, F. Qureshi and O. Gustafsson, "Efficient FPGA Mapping of Pipeline SDF FFT Cores," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 25, no. 9, pp. 2486-2497, Sept. 2017, doi: 10.1109/TVLSI.2017.2710479, 2017.
- [9] Phuong H. Lai, Manh Hoang, Viet Q. Tran, Tung V. Nguyen, Thien V. Truong, and Phong H. Nguyen, "Analysis and implementation of SDF Radix-2 FFT processor using VERILOG Hardware Description Language", in International Journal of Advanced Trends in Computer Science and Engineering, Volume-9, No.4, doi: 10.30534/ijatcse/2020/144942020, August-2020.
- [10] Pattan, Anwar Bhasha, and M. Madhavi Latha. "Low Power and High Performance Structures for Fast Fourier Transform Processor." International Journal of Computer Applications, Volume 132 – No.10, ISSN: 0975-8887, Dec-2015.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor

**Impact Factor:
7.512**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details