International Journal of Applied Engineering Research ISSN 0973-4562 Volume 15, Number 11 (2020) pp. 1075-1080
 https://dx.doi.org/10.37622/IJAER/15.11.2020.1075-1080

Inexact Floating Point Adders Analysis

S.Aruna Mastani[1], Riyaz Ahamed Shaik[2]

Assistant Professor, Department of Electronics and Communication Engineering, Jawahrlal Nehru Technological University, Anantapur, India.

PG Scholar, Department of Electronics and Communication Engineering, Jawahrlal Nehru Technological University Anantapur, India.

Abstract

With increased complexity in arithmetic circuits, due to the increasing demands for inexact processing and enhanced integration capacity, power has become a key imperative in nano-scale integrated circuit structure. An inaccurate circuit provides a promising way to deal with a complete reduction in both dynamic and static power dissipation for error-tolerant applications as a growing computational worldview. In exploring the various approximate adders, Hardware Optimized and Error Reduced Approximate Adder(HOERAA) is superior to the many of the adders till the state. The basic idea is to use this HOERAA adder instead to LOA adder in the Inexact Floating Point 32bit adder proposed by Weiqiang Liu et al.[6] and made the detailed comparison of various parameters like area, power Maximum operating frequency. The total implementation is done on (xc7a100tcsg324-3) FPGA device with Xilinx vivado 2019.1 tool of Artix7 family of FGPA. Where we noticed that the area consumption is 9.12 percent less when compared to the exact floating-point adder and 3.48 percent less compared to the existing floating-point adder[6]. The above-mentioned constraint of area are addressed by the strategies of implementation (default) and synthesis (default)

Keywords: Inexact Circuits, Floating-Point (FP) adders, Low Area, mantissa adder

I. INTRODUCTION

As the development and advancement of innovative integrated computerized circuits. Computational capability has increased significantly. Power has become a key limitation in this scenario. The arithmetic unit is the core of a processor and the entire processor's output is determined by its capacity. Subsequent inquiry into inaccurate fixed-point adder led to the conclusion that approximate hardware [1],[2] handling with a general error can be increasingly successful in speed, field, and energy approximately multiple times than a precise chip. Other approximate adders [3],[4],[5],[6] are some of the available approximate adders which have an certain error range and performance difference in terms of area, power and frequency, where [7] bring forward an another approximate adder which is better in terms of area and error range. Common structures (exact adders) apply completely precise results to a wide range of computations with incurred overhead of area. In applications where errors can be tolerated, it is conceivable to perform the calculation with inexact circuits. In these cases, inexact processing [10] is an appealing way to deal with constraints of low power and area.

This paper is an analysis of the research by the authors[10] on inexact Floating-Point adder which includes an inexact adder of which it requires inexact modules; it is analysed as follows: (1) The design of exact and inexact floating-point adders are implemented with the algorithm which is obtainable at reference[10] and the architecture present in[9], (2) The inexact exponent subtractor and inexact mantissa adder part is implemented with inexact adders which are listed in[7] are considered while implementing the inexact Floating-Point adder, (3) constraints like area, power, maximum operating frequency, are analysed and came for a conclusion to recommend the usage of HOERAA inexact adder for implementing the inexact Floating-Point adder.

This paper is coordinated into various sections. section 2 gives a background of Floating-point format, Section 3 describes about the implemented work, section 4 explains about results and followed by a conclusion in Section 5.

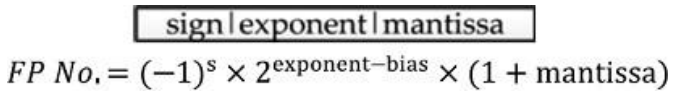
II. BACKGROUND

A. Floating-Point Format

The IEEE 754-2008 is the most widely used standard for the Floating Point community. The basic and extended forms that

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 15, Number 11 (2020) pp. 1075-1080
 https://dx.doi.org/10.37622/IJAER/15.11.2020.1075-1080

are retained by this standard are: half-precision (16 bits), single-precision (32 bits), double-precision (64 bits), extended-precision (80 bits) and quad-precision (128 bits). As below, Fig 1 a general IEEE Floating Point format, where the exponent part has an bias of 2^{E-1}-1, where E represents the number of exponent bits. The single-precision and double-precision formats are most utilized in the present computing.



$$FP\ No. = (-1)^{s} \times 2^{\text{exponent}-\text{bias}} \times (1 + \text{mantissa})$$

Fig. 1. General IEEE 754 FP format.

B. Floating-Point Adder Architecture

A floating point adder architecture contains a number of hardware blocks comprises with exponents alignment of mantissa addition, exponent subtraction, normalization and rounding as shown in Fig 2. The first two operands are unpacked to the floating point adder format and hidden bits are added to each mantissa. The exponents are evaluated for finding the larger number and accordingly swapped, and hidden bit 1 is added to mantissa. After this exponent difference is calculated. With obtained exponent difference the mantissa bits are right shifted. After the mantissa addition shifting is done by the normalizes to restore the results in an standard IEEE format. Detection of leading zero is a key step for a normalization processes, in which left shift is done with number of leading zero count, after that rounding is done and results are stored. If the special cases are detected such as over flow, under flow and not a number these are represented as flags. For further implementation steps refer the algorithm which is provided in Fig 3.

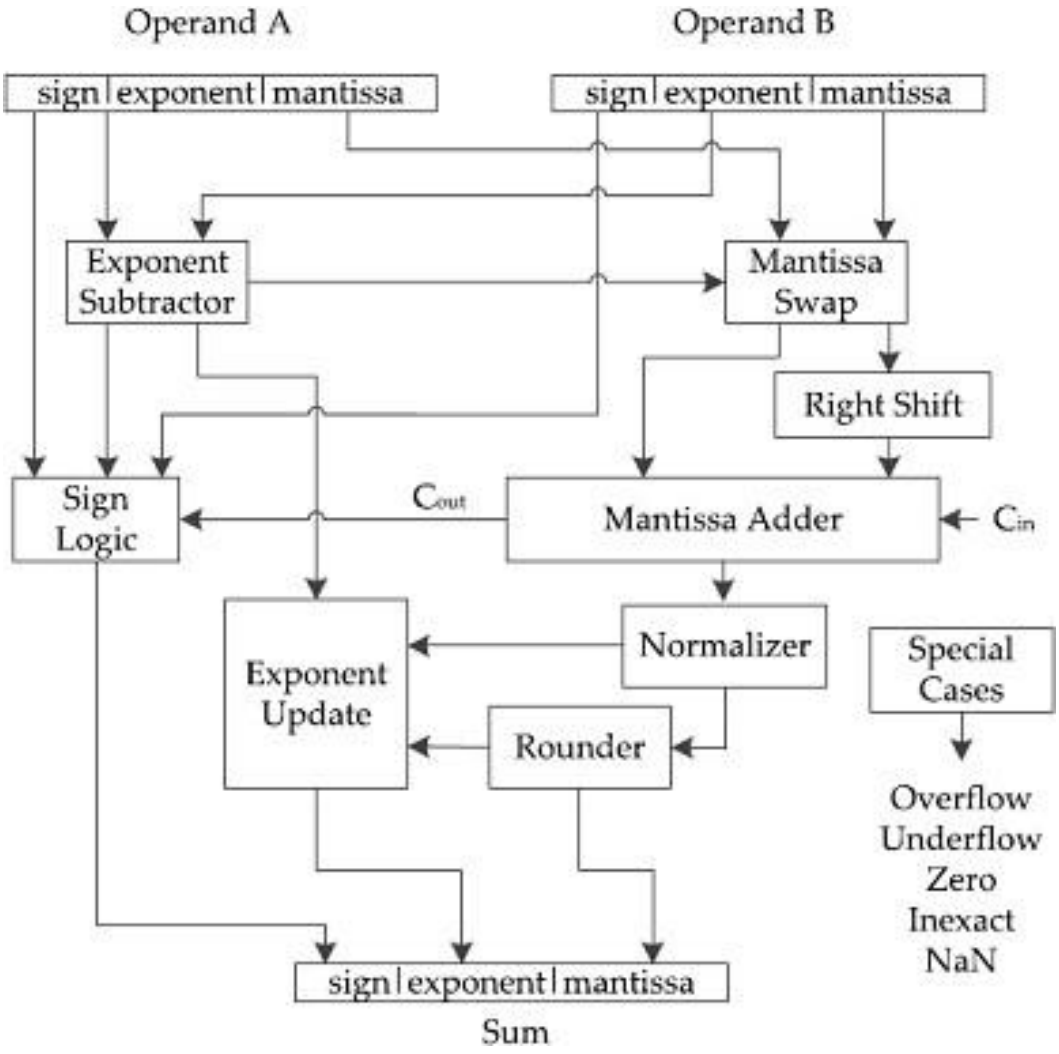


Fig. 2. Accurate FP Architecture.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 15, Number 11 (2020) pp. 1075-1080
 https://dx.doi.org/10.37622/IJAER/15.11.2020.1075-1080

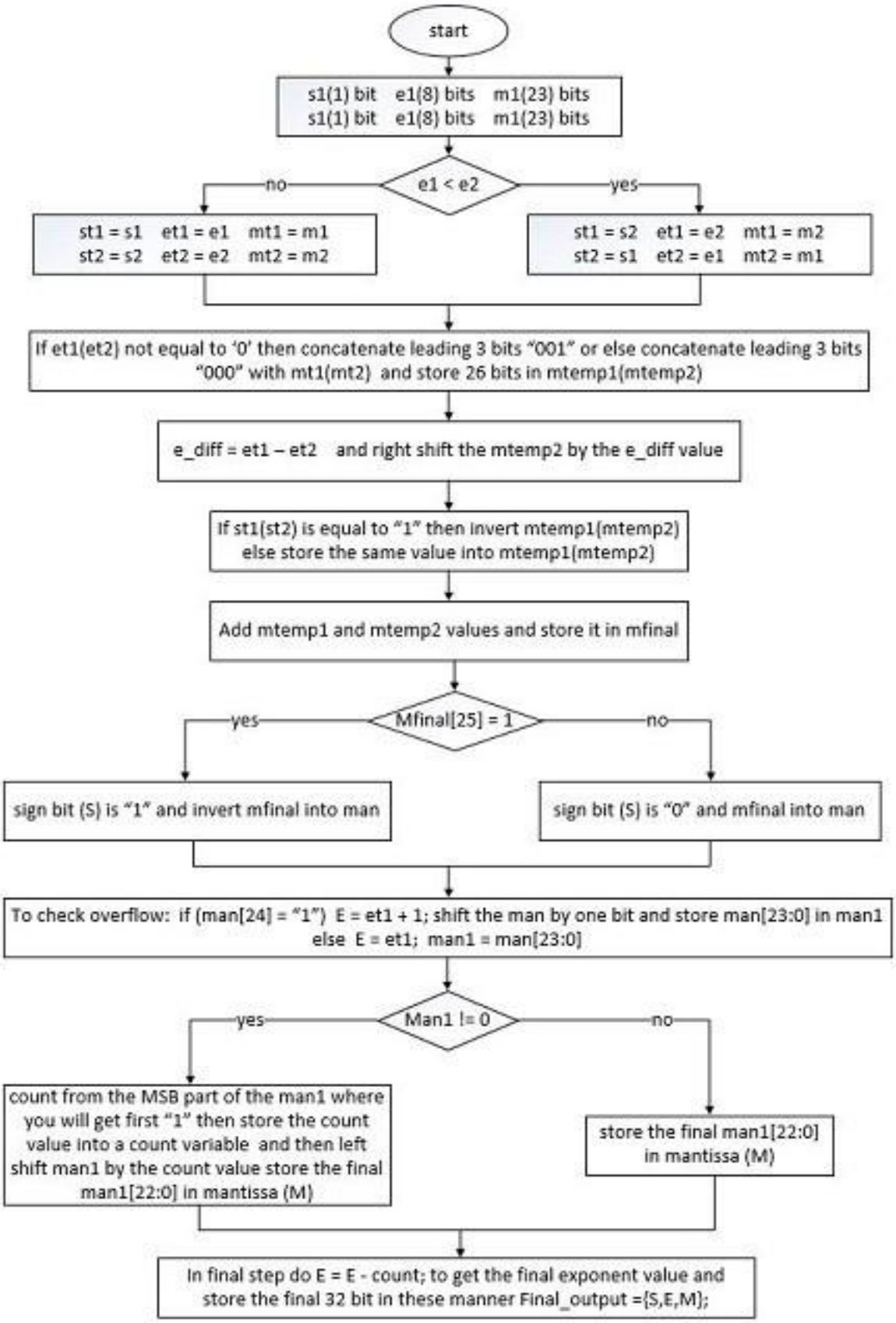


Fig. 3. Flowchart of single precision FP Adder.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 15, Number 11 (2020) pp. 1075-1080
© Research India Publications. https://dx.doi.org/10.37622/IJAER/15.11.2020.1075-1080

III. IMPLEMENTED WORK

The Inexact floating point adder is implemented at a behavioural level containing the various implementation of inexact and exact Exponent subtractor, sign logic, mantissa swap, inexact mantissa adder(IMA). Inexact Mantissa adder was implemented using inexact adders[3][7], the circuit-level inexact designs are explained in the following sections.

A. Exponent subtraction

Exponent subtraction will read two exponent bits and calculate the difference between those two exponent values, these difference is used for shifting the bits of mantissa values, it is implemented as exact Exponent subtractor for our analysis. where as in the exact exponent subtraction the error at the final output is minimalist, if we use inexact exponent subtraction then the error will be high, because the exponent subtraction is a major contributor among the entire floating point addition, so consideration of exact exponent subtractor will define the final output.

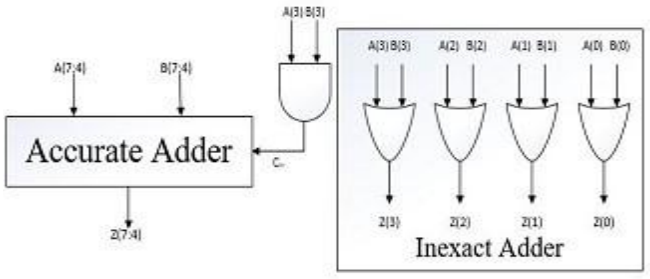


Fig. 4. LOA Adder.

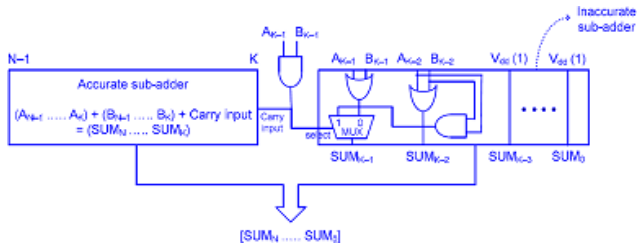


Fig. 5. HOERAA Adder.

B. Inexact Mantissa Addition

The Inexact Mantissa Adder is implemented by various adders for analysis and are explained here. where the first inexact adder which was implemented by the author[3] was LOA adder, for calculating the inexact mantissa value and the figure of LOA adder is shown in fig 4,and the second inexact adder is HOERAA inexact adder which was implemented by the author[6],the HOERAA adder was shown in Fig 5, As mantissa error is considerable then the inaccurate addition can be done at the LSB part which will not make a much difference in the result. Therefore approach to inexact design can be considerable. To analyse the design functionality in terms of inexact mantissa adder, implementation of these inexact mantissa adder is implemented by[3]and[7] results are explained in the below section

C. Normalization

The total or difference of adders of exponents and mantissa may be small, where the data may require change. Therefore, we do normalization to ensure that the addition of the data falls into the correct range. It also involves a reduction of the exponent. Normalization involves the calculation of the number of left changes, which is accompanied by a leading zero counter. Because the leading zero count logic is used to evaluate the change of the mantissa bits where the mantissa bits MSB part is processed by accurate adders. As the output from the mantissa adder is inexact we can implement an inexact normalizer to calculate the leading zero count.

D. Rounding

Rounding retains the three extra bits (i.e., guard bit, round bit and sticky bit) in the case of exact adders of the bits obtained from the previous normalization stage, which will yield exact n bits. However the implementation of this technique in inexact adders is not mandatory. Logical OR operations that include an error in the LSB location are applied to the least important bits in this approach. Rounding can also be overlooked when the inaccurate architecture in the Floating Point Adder is enforced. however there are inexact result in the n least significant bits then rounding can be ignored.

E. Architectures Implemented

In this Paper for analysing different architectures like exact and inexact floating point adders are implemented, for implementation of this architecture in Fig 6 inexact mantissa adder, inexact normalizer and exact exponent subtractor is considered.

After implementing all these architectures we have analysed the output metrics like area power maximum operating frequency, detailed explanation was given in results section

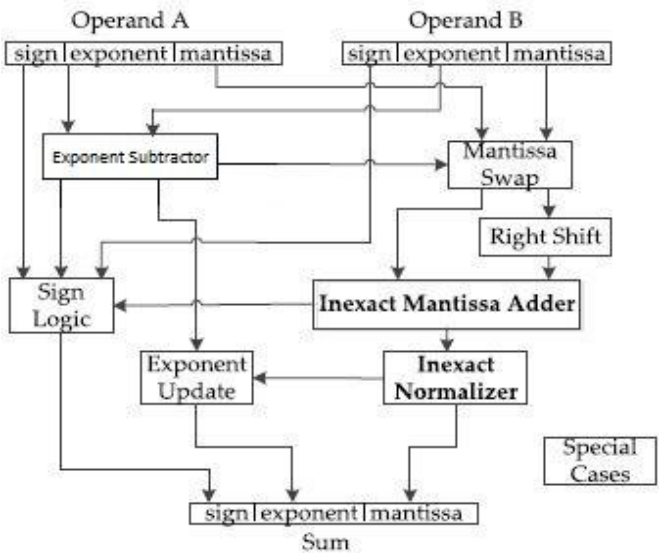


Fig. 6. Inexact Mantissa Floating Point Architecture.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 15, Number 11 (2020) pp. 1075-1080
© Research India Publications. https://dx.doi.org/10.37622/IJAER/15.11.2020.1075-1080

IV. RESULTS

An inexact design can be efficient only when the generated outputs from the inexact module are less error pron, in the analysis observations are done on the errors coming from the inexact floating point adders, and it will depend on the number of inexact bits that are considered in the exponent. These errors cannot be neglected as they may cause significant impact on the results when the number of inexact bits will keep increasing.

In this paper inexact architectures are implemented, and compared. In the analysis of Inexact Floating Point architectures we have analysed a scenario of only mantissa adder as inexact of 4bits.

Where we noticed that, for only mantissa adder is inexact adder (HOERAA), considering for implementation of inexact Floating Point Unit, the area consumption is 9.12% less when compared to the exact floating point adder and 3.48% less compared to the existing Inexact floating point adder[8]

In this paper for implementation xilinx vivado tool is used with the falmily Artix 7 (xc7a100tcsg324-3) to analyse the results and their synthesis reports are tabulated below. The architectures are coded in Verilog HDL.

V. CONCLUSION

In this paper Inexact Floating Point Adder results has been analysed with the Exact Floating point adder results. In the Inexact floating point adder LOA and HOERAA adders are implemented in inexact mantissa Adder modules. so we can conclude by saying that implementation of Inexact Floating Point Adder, using Inexact HOERAA adder is suggested, because Area consumption and the error form HOERAA adder is less when compared to LOA Adder which yields to the better results.

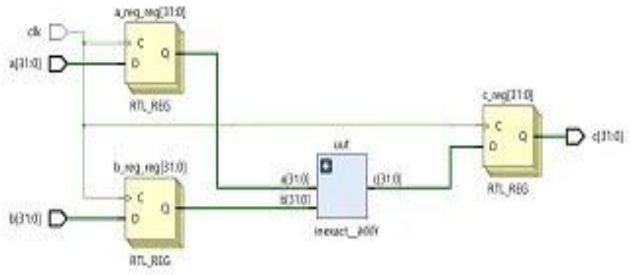


Fig. 7. Inexact Adder Top Schematic.

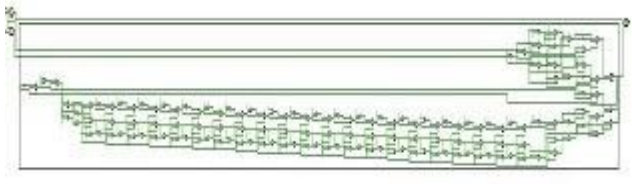


Fig. 8. Inexact Adder Schematic.

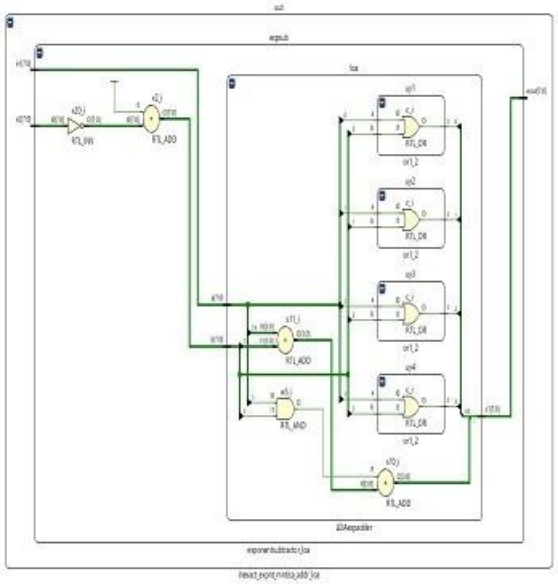


Fig. 9. Inexact Exponent LOA Schematic.

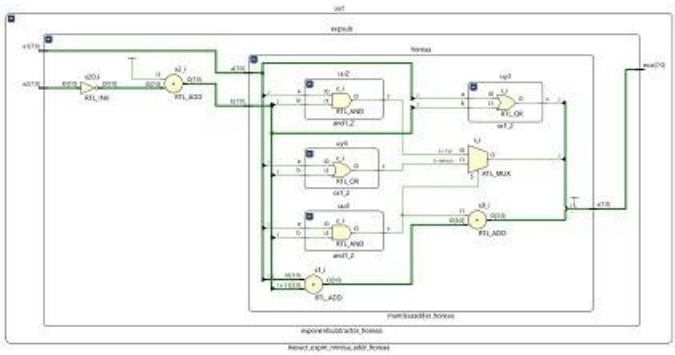


Fig. 10. Inexact Exponent HOERAA Schematic.

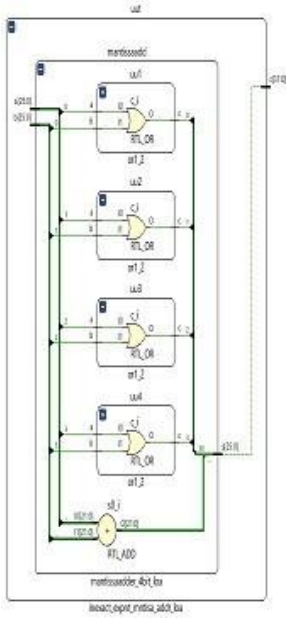


Fig. 11. Inexact Mantissa LOA Schematic.

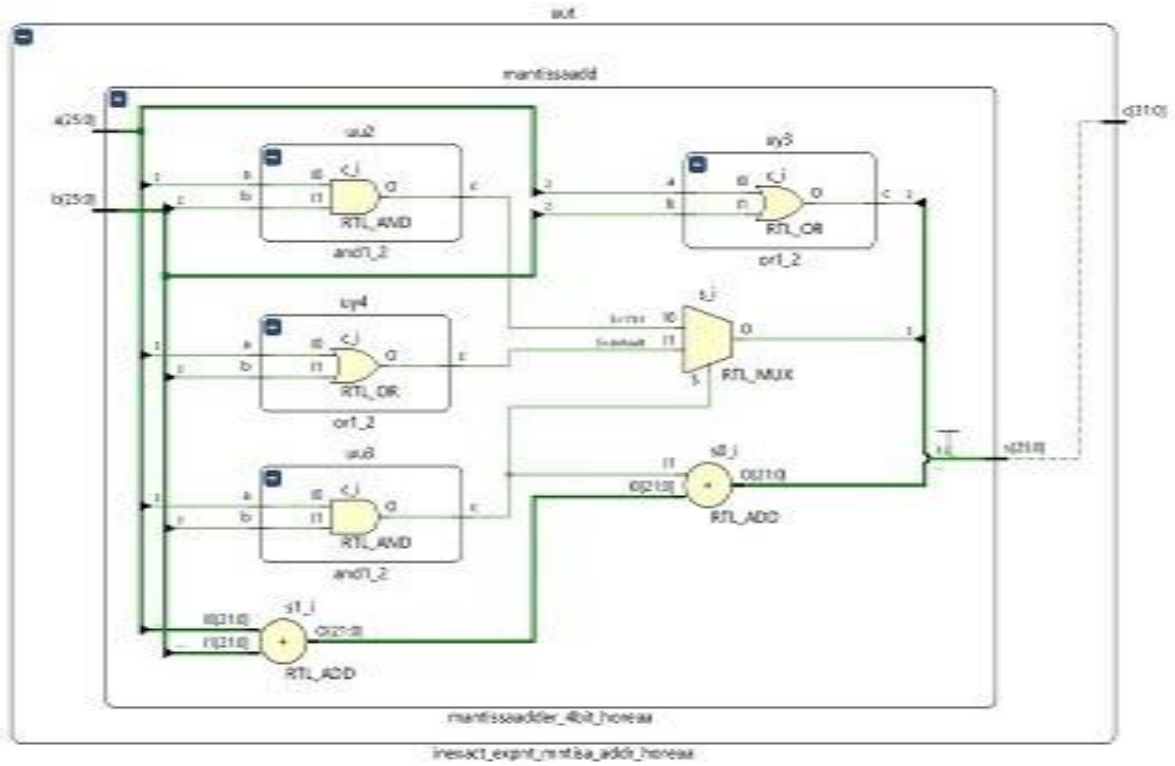


Fig. 12. Inexact Mantissa HOERAA Schematic.

Table I. Comparison of Exact, Inexact and Proposed FP Adders

FPU's	***Max Freq(MHz)***	***Power (W)***	***# LUTs***	***# FFs***
Exact FPU	90	0.126	274	96
Inexact FP Adder1	101	0.126	258	96
Inexact FP Adder2	93	0.126	249	92

REFERENCES

[1] K. Palem and A. Lingamneni, "Ten years of building broken chips: The physics and engineering of inexact computing," ACM Trans. Embedded Comput. Syst., vol. 12, no. 2, article 87, 2013.

[2] A.Lingamneni, K. Muntimadugu, C. Enz, R. Karp, K. Palem, and C. Piguet,"Algorithmic methodologies for ultra-efficient inexact archi- tectures for sustaining technology scaling," in Proc. ACM Int. Conf. Comput. Frontiers, 2012,

[3] H. Mahdiani, A. Ahmadi, S. Fakhraie, and C. Lucas, "Bio-inspired imprecise computational blocks for efficient VLSI implementation of softcomputing applications," IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 57, no. 4, pp. 850–862, Apr. 2010.

[4] Albicocco, P.; Cardarilli, G.C.; Nannarelli, A.; Petricca, M.; Re, M. Imprecise arithmetic for low power image processing. In Proceedings of the 46th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 4–7 November 2012.

[5] Gupta, V.; Mohapatra, D.; Raghunathan, A.; Roy, K. Low-power digital signal processing using approximate adders. IEEE Trans. CAD Integr. Circuits Syst. 2013, 32, 124–137.

[6] Balasubramanian, P.; Maskell, D. Hardware effcient approximate adder design. In Proceedings of the IEEE Region 10 Conference, Jeju, Korea, 28–31 October 2018.

[7] Padmanabhan Balasubramanian and Douglas L. Maskell, "Hardware Optimized and Error Reduced Approximate Adder," "www.mdpi.com journal electronics". Oct. 2019

[8] W. Liu, L, Chen, C. Wang, M. O'Neill, and F. Lombardi, "Inexact floating point adder for dynamic image processing," in Proc. 14th IEEE Conf. Nanotechnol., 2014, pp. 239–243.

[9] Weiqiang Liu, Senior Member, IEEE, Linbin Chen, Chenghua Wang, maire O. Neill, Senior Member, IEEE, and Fabrizio Lombardi, Fellow, IEEE "Design and Analysis of Inexact Floating-Point Adders," "IEEE TRANSACTIONS ON COMPUTERS, VOL. 65, NO. 1, JANUARY 2016,"

[10] Atul Rahman, Abdullah-Al-Kafi, Mr. Khalid, A.T.M. Saiful Islam, Mah- mudur Rahman, "Optimized Hardware Architecture for Implementing IEEE 754 Standard Double Precision Floating Point Adder/Subtractor" 2014 17th International Conference on Computer and Information Tech- nology (ICCIT)