

A Multi-Frequency Performance Evaluation of an 8-Bit Fixed-Bias Truncated Approximate Multiplier Design on FPGA

¹Kuldeep Singh, ²Vinod Kumar, ³Ajay Kumar

^{1,2,3}Department of Electrical and Electronics Engineering, Guru Jambheshwar University of Science and Technology, Hisar 125001, India

E-mail: ¹kuldeepsingh@gjust.org, ²vinodspec@yahoo.co.in, ³ajaypunia506@gmail.com

Abstract - An RTL-to-FPGA design and implementation analysis for an 8-bit fixed-bias truncated approximate multiplier has been evaluated on the XC7Z020-CLG484-1 device fabricated at 28 nm. The main central approximate core eliminates partial-product columns 0–5 (CUT=6) and inserts the corrected column $C_{FB}=2^6=64$ at the first reserved column. To observe the different architectural sound effects, both the exact and fixed-bias cores are evaluated with isolation at same input/core/output register boundaries, FPGA target for synthesis and implementation settings, clock constraints, and fabric-only DSP policy. A fully comprehensive evaluation of all 65,536 unsigned 8-bit test cases have been run for simulation and verification. A vast range of targets from 50 to 200 MHz are analyzed for observing post-route timing. The obtained results at 200 MHz reveal that, in case of the fixed-bias model structure reported +0.012 ns WNS while in the exact reference was -0.132 ns, LUT utilization was 68 versus 84 LUTs respectively, and consumes 0.021 versus 0.029 W dynamic power in fixed bias versus exact reference respectively. The study state that the PDP values decreases from 0.1488 to 0.1047 nJ and LUT-based ADP from 431.09 to 339.18 LUT•ns respectively. During the evaluation across the full sweep, the average changes reported in the fixed bias and exact reference models are 27.4% in dynamic power, 21.1% in LUT count, 28.0% in PDP, and 21.7% in ADP.

Keywords: Approximate computing, fixed-bias multiplier, truncated multiplier, RTL-to-FPGA, ZedBoard, Zynq-7000, timing closure, WNS, power-delay product, area-delay product.

I. INTRODUCTION

It has been acknowledge that in to implement the digital signal processing, image processing, neural-network inference, communication systems, and embedded edge hardware the multiplication is a main arithmetic primitive operation. In FPGA hardware implementation for accelerators,

the use of multiplication operator is a best fit to achieve arithmetic latency, LUT utilization, routing complexity, switching activity, and the ability of the implemented design to close timing at a required clock period within the boundary values. Multiplication cost is reflected in considerable change in these parameters. In the approximate arithmetic operations the tolerance in errors of many signal- and data-processing workloads can be obtained by replacing the exact arithmetic operation with an operation with controlled numerical error [1-2]. The Partial-product truncation method is a prime approximation process as bit region discarded by it is clearly distinguishable in the multiplier matrix. All the discarded terms are considered nonnegative for all unsigned operands; therefore the uncompensated truncation presents a negative bias. Fixed-bias compensation method has a proved low-complexity rectifying technique in which a constant is added close to the truncation boundary. Many of the compensation method are operand-dependent, this rectifying solution does not need a variable estimator and may optimize the hardware reduction involved in the partial-product rejection [3-4].

The requirements of RTL simulation or timing synthesis is a reliable comparison process for hardware design. The FPGA implementation changes in logic packing, placement and routing, and total LUT count with variable clock constraints. The work reported in this paper therefore is completed by combining the arithmetic characterization of RTL-to-FPGA front-end and back-end process and a very wide fourteen-point frequency sweep for implementation. The main contribution in the study is a controlled FPGA implementation framework in which the analysis for accuracy, timing, power, and resource trade-offs can be repeatedly achieved under same pre-defined boundaries and implementation environment [5-6].

The literature study stats that truncated and fixed-width multipliers have applications for energy- and area-efficient DSP hardware realization. Petra *et al.* demonstrated a variable-correction truncated multipliers structure and it was optimized for specially mean-square error [7]. Garofalo *et al.*

has characterized model for a maximum error tolerance under the categories of truncated multipliers [8]. An approximate multiplication technique through a partial-product process was demonstrated and explained that selective discarded low-significance terms may overcome the limitation for high cost implementation maintaining numerical accuracy [9]. The multiplier with scalable truncation- and rounding-based approximation, has been explored for structured design and energy-accuracy achievement points [10]. The multiplier architectural complexity in an approximate compressor-based multiplication process can be defined and stated for optimization of hardware [11].

The exhaustive literature survey stated that the arithmetic accuracy of an operation cannot be defined in a single metric hardware definition strongly depend on its modeling style and implementation conditions [12]. When implementing a multiplier mapped to DSP blocks resources and another multiplier mapped to LUT and/or carry fabric primitive components may totally disturb the comparison behavior. In the reported study, it is therefore utilized a common primitive components LUT based implementation mechanism, practices same timing boundaries, and reports post-routed results rather than synthesis-only metrics for authenticity [13-14].

In the designed model, a precise CUT=6 in case of fixed-bias 8×8 multiplier model with a +64 correction, a full RTL-to-FPGA methodology flow starting with specification, RTL verification, synthesis, XDC constraints, placement, clock routing, routing, Static Timing Analysis, power analysis, bit stream generation for programming FPGA hardware, and ending on-board verification; and fully recommended arithmetic evaluation with all possible 2256 input pairs has

been investigated exhaustively; A constrained over a span of 50–200 MHz post-route characterization of fixed biased vs an exact reference has been successfully completed. The reported results for PDP and ADP calculations demonstrate the practical hardware-efficiency benefit of fixed-bias compensation.

II. FIXED-BIAS TRUNCATED MULTIPLIER ARCHITECTURE

A functional block diagrams architecture of Exact Reference and fixed-bias truncated multiplier is designed using Fig. 1. The product term in the exact multiplier for two unsigned 8-bit operands A and B, may be expressed as the weighted sum of 64 elementary partial products:

$$P_{EX} = \sum_{i=0}^{7} \sum_{j=0}^{7} A_i B_j 2^{(i+j)} \quad (1)$$

For the final result; the exact multiplier model retains completely all partial-product columns from weight 2^0 through 2^{16} and used for a zero-error baseline. In second case; with a CUT=6, the terms true for $i+j < 6$ are discarded. Then the ordinary truncated product and the discarded regions are expressed using the following equations:

$$P_{OT} = \sum_{i+j \geq 6} A_i B_j 2^{(i+j)}, D_{CUT} = \sum_{i+j < 6} A_i B_j 2^{(i+j)} \quad (2)$$

The omitted region compensation is made by the constant correction partially without recalling the removed partial products or defining estimation logic by operand dependent. The simulation, synthesis and implementation results obtained through the same wrapper block with input/output registers used for exact and fixed bias multipliers. As the core block is kept pure combinational, hence the timing after post route measured through same register-to-register edges.

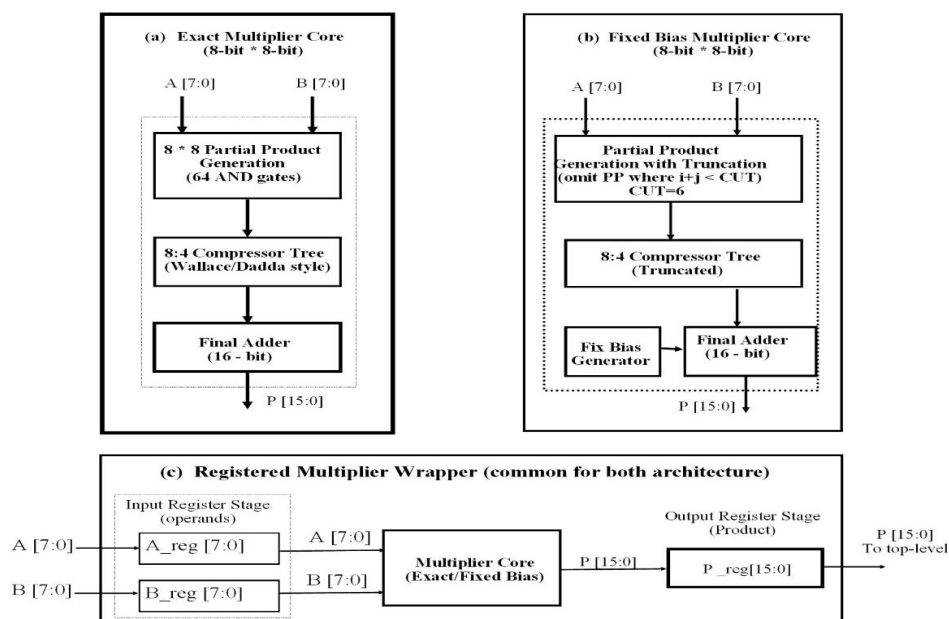


Figure 1: Functional block diagrams of Exact Reference and fixed-bias truncated multiplier

III. RTL-TO-FPGA IMPLEMENTATION

The Figure 2 describes (i) concept of RTL-2-FPGA flow with front-end/back-end partition, (ii) the detailed Zed Board sequence of processing steps, and (iii) the practically utilized RTL-to-FPGA overall workflow using Zed Board dedicated to this study. The front end FPGA design for digital systems starts from the functional specification, next followed by design entry of HDL-VHDL/Verilog description, testbench-based behavioral simulation and verification, RTL elaboration, synthesis, and technology mapping. Synthesis process the behavioral description of the design under test into a technology-specific gate level netlist that contains LUTs, flip-flops, carry-chain elements, and related programmable-logic resources and others. The back end part reads the synthesized netlist along with XDC file where timing and I/O constraints defined before optimization for placement, clock routing, programmable I/Os routing, static timing analysis, power estimation, DRC, and bitstream generation.

The target device, XC7Z020-CLG484-1, used as a core part inside the Avnet Zed Board. The RTL functional behavior correctness is tested verified in the process as pre-synthesis. XDC constraints file create clock with the defined parameters and board I/O configuration. Next in the implementation process the following steps performed physical optimization, placement, clock-resource assignment, routing, and post-route timing analysis. The WNS parameter is used for timing closure evaluation; greater than zero WNS is a satisfactory requirement for dedicated maximum frequency. To consider the effects of clocking and routing-dependent capacitance measured power must be taken after implementation. After that, the .bit generated file is configured and be downloaded on-board functional verification.

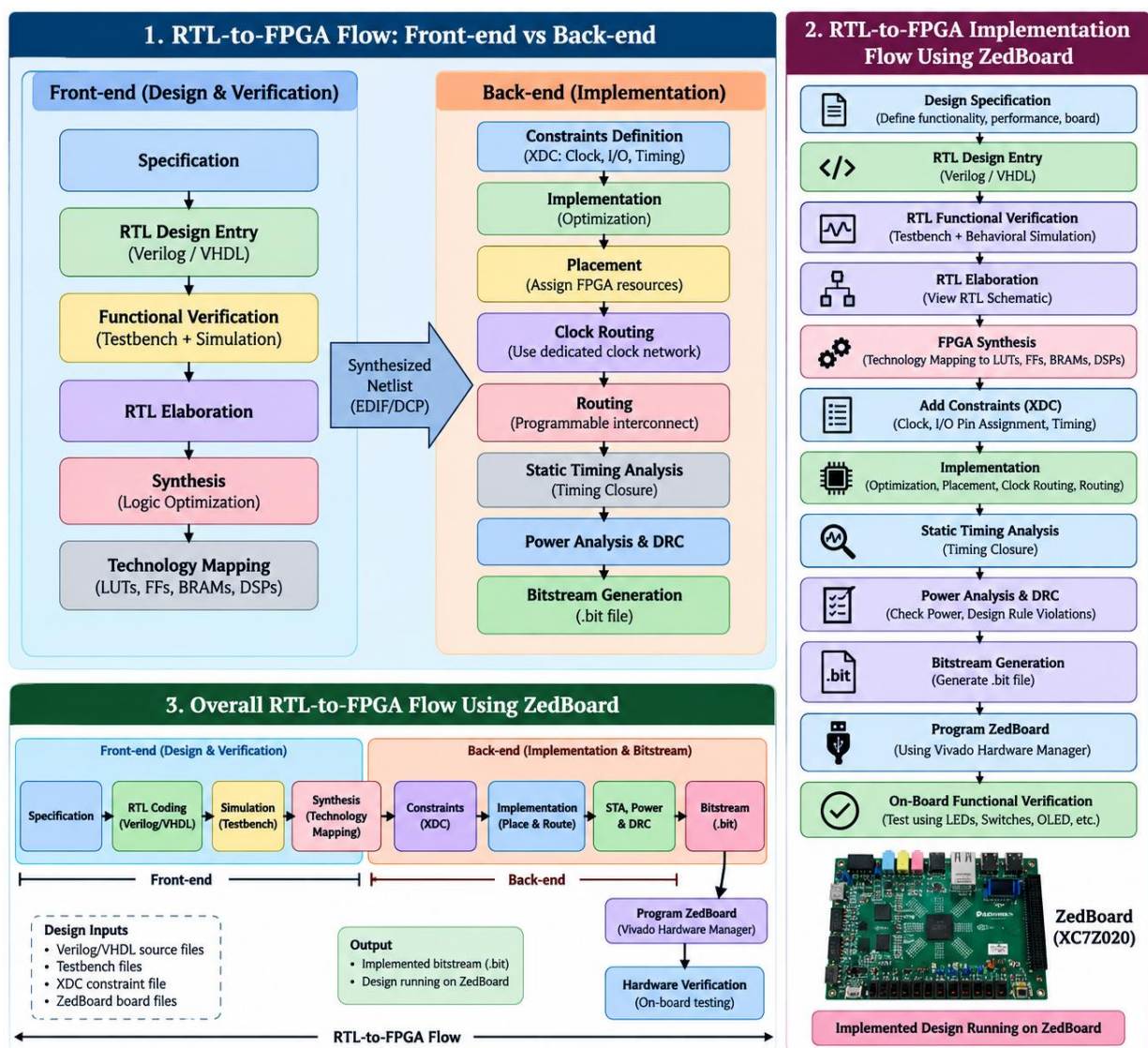


Figure 2: A detailed Conceptual and Practical RTL-to-FPGA implementation process flow using ZedBoard: 1) front-end/back-end partitioning, 2) Detailed implementation flow, and 3) Overall RTL-to-bitstream/hardware-validation sequence used in the design

IV. RESULTS AND DISCUSSION

A. Multi-Frequency Post-Route Timing Results

Table I: Post-route exact-reference and fixed-bias results over fourteen timing constraints

MHz	WNS EX	WNS FB	Tcrit EX	Tcrit FB	Pdyn EX	Pdyn FB	LUT EX	LUT FB
50	12.584	12.855	7.416	7.145	0.007	0.005	71	56
62.5	8.725	8.640	7.275	7.360	0.009	0.006	71	56
75.09	6.035	6.029	7.295	7.301	0.011	0.008	71	56
87.5	4.407	4.543	7.023	6.887	0.012	0.009	71	56
100	3.501	3.355	6.499	6.645	0.014	0.010	71	56
112.5	2.030	2.613	6.860	6.277	0.016	0.012	71	56
125	1.466	1.274	6.534	6.726	0.018	0.013	71	56
133	0.800	0.910	6.719	6.609	0.019	0.014	71	56
150	0.560	0.268	6.110	6.402	0.022	0.016	71	56
162.6	0.192	0.304	5.958	5.846	0.023	0.017	80	60
175.131	0.304	0.123	5.406	5.587	0.025	0.018	81	58
180.995	0.233	0.194	5.292	5.331	0.026	0.019	75	64
190.5	-0.094	0.171	5.357	5.092	0.027	0.020	82	67
200	-0.132	0.012	5.132	4.988	0.029	0.021	84	68

In the Table I, the reported results states that WNS remains positive till the 180.995-MHz target, after that WNS=-0.094 ns at 190.5 MHz and -0.132 ns at 200 MHz in case of the Exact model architecture. In comparison with the fixed-bias core maintains positive sign and reported as +0.171 ns and +0.012 ns at both targets respectively. Thus, inside the steady sweep taken under consideration, fixed bias model is satisfying the timing-constraints at 200 MHz also but the exact model is not.

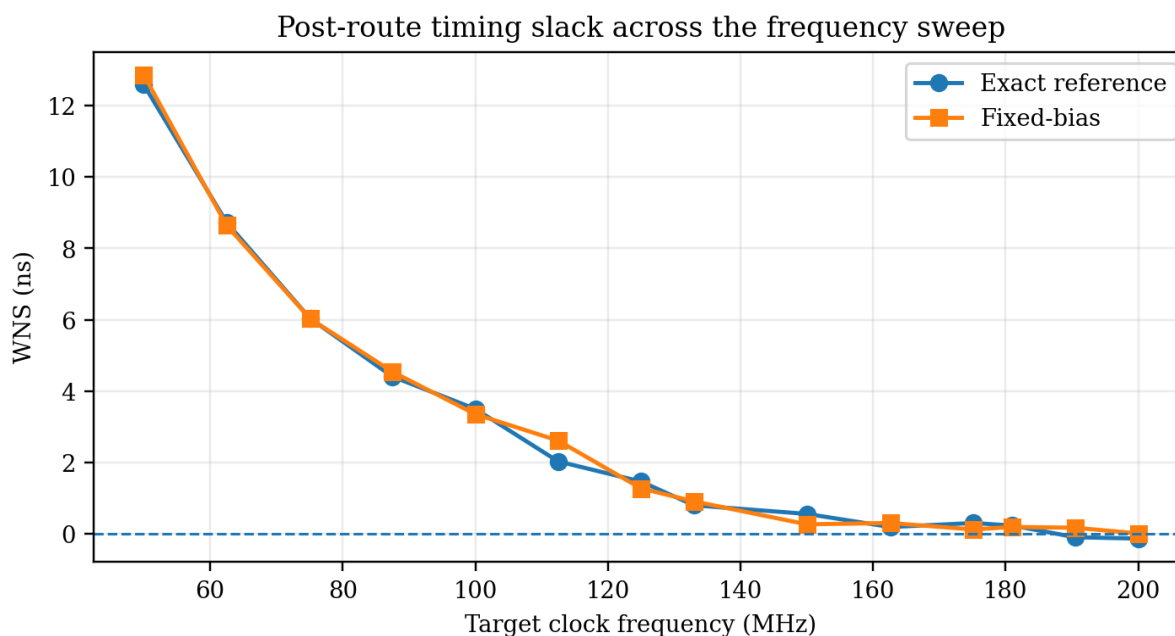


Figure 3: Post-route WNS versus target clock frequency. The zero-slack line marks the setup timing-closure boundary

B. Critical Delay

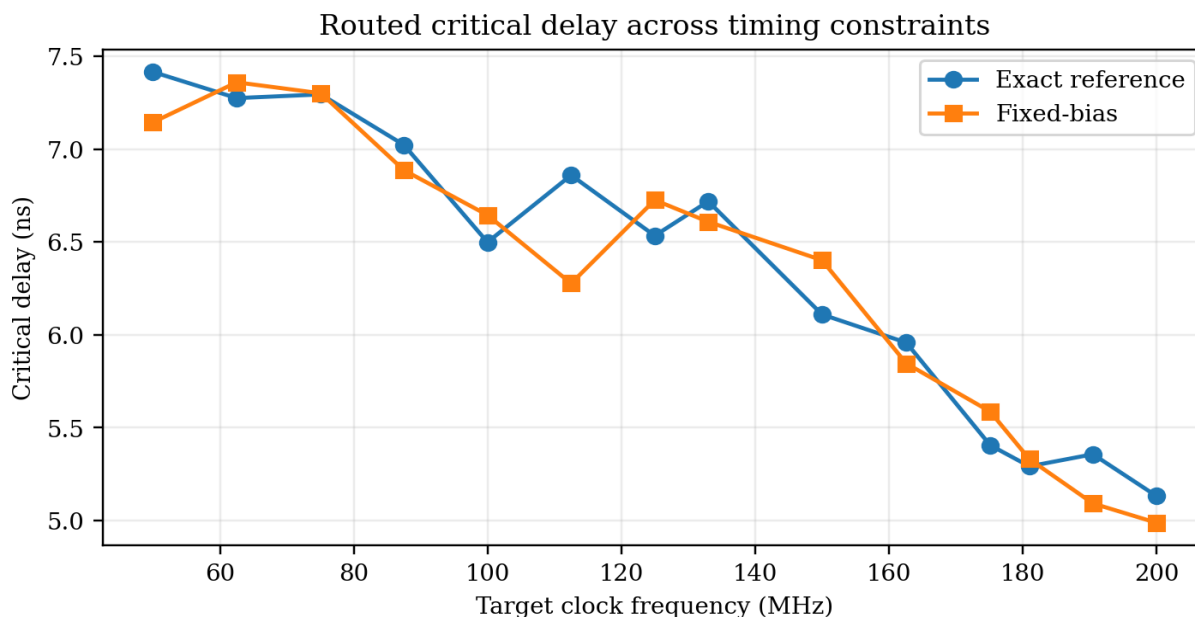


Figure 4: Routed critical delay for exact and fixed-bias implementations

In the Figure 4, at 200 MHz, the critical delay is 4.988 ns for fixed biased compared with 5.132 ns for the exact model, a 2.81% reduction in values can be seen. Further, at 190.5 MHz the reduction is 4.95%. But the at safe constraints, this pattern is not consistently approving the optimization; for example, at 100 MHz, the exact model gives 6.499 ns versus 6.645 for fixed bias. That why there is a need for a sweep rather considering single timing operating point.

C. Dynamic Power and LUT Utilization

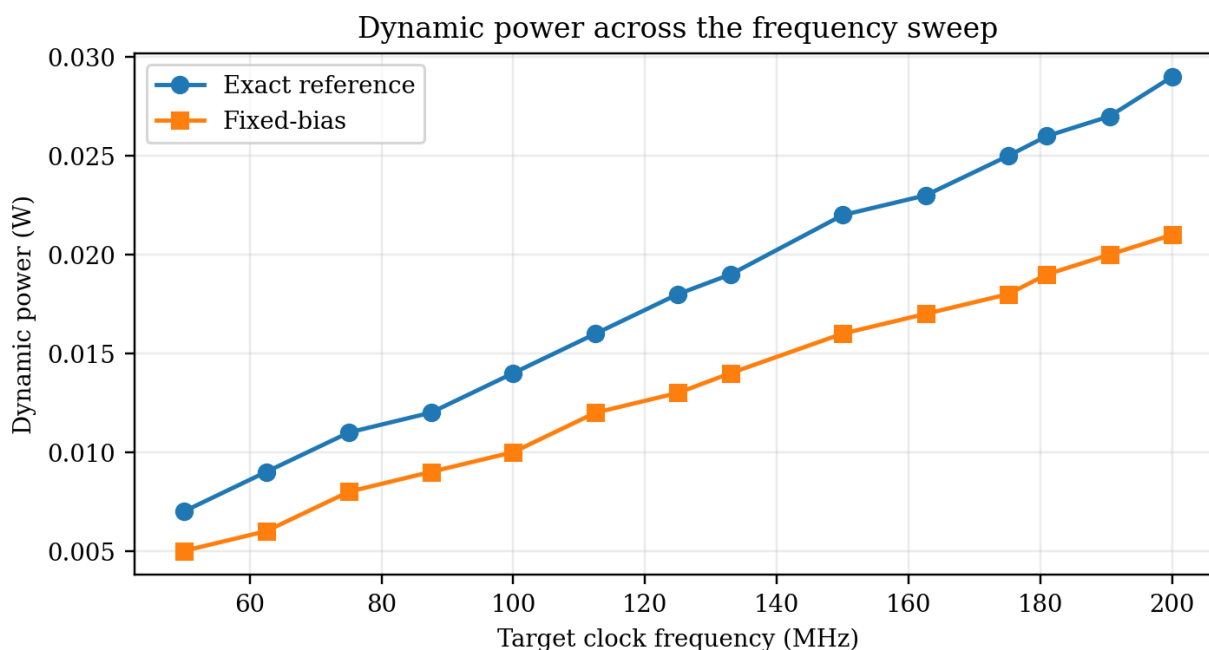


Figure 5: Dynamic power versus target clock frequency

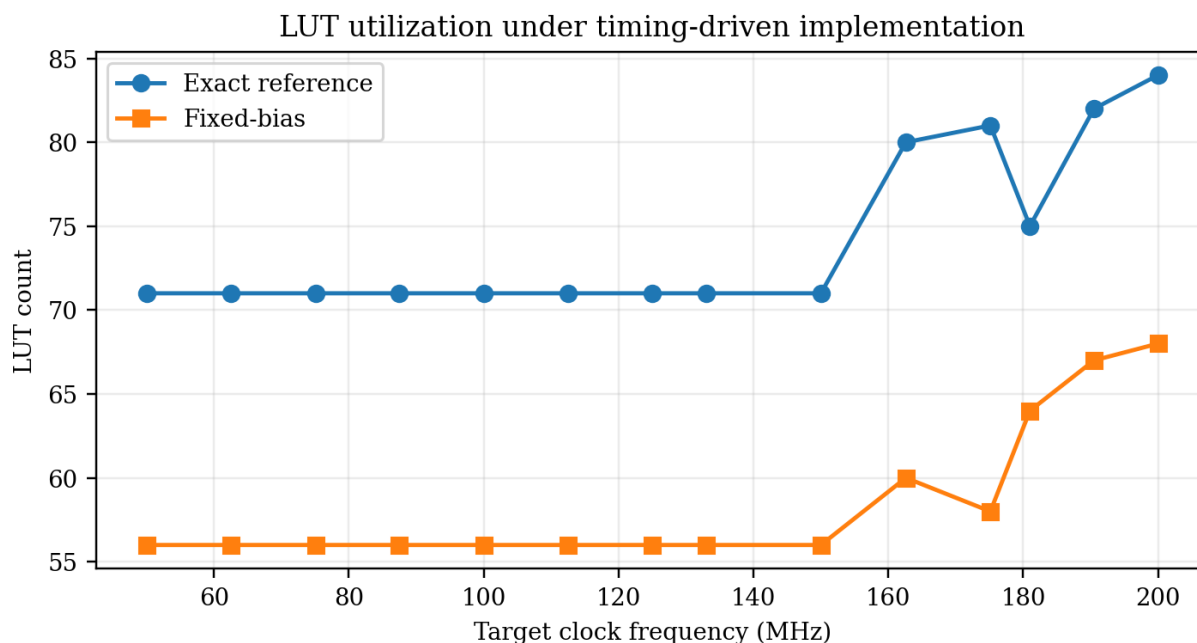


Figure 6: LUT utilization versus target clock frequency

Figure 5 shows the behavior of both the models for dynamic power against target clock frequency. The fixed biased model shows the best performance forever. The dynamic power reduced from 0.029 W to 0.021 W at 200 MHz, it is a 27.59% less. Figure 6, represents the LUT utilization, which is also consistently lower compared with exact reference, a total of 19.05% less at 200 MHz. over all, the average decrease in the dynamic power and LUT count are 27.4% and 21.1% respectively.

D. Power–Delay and Area–Delay Products

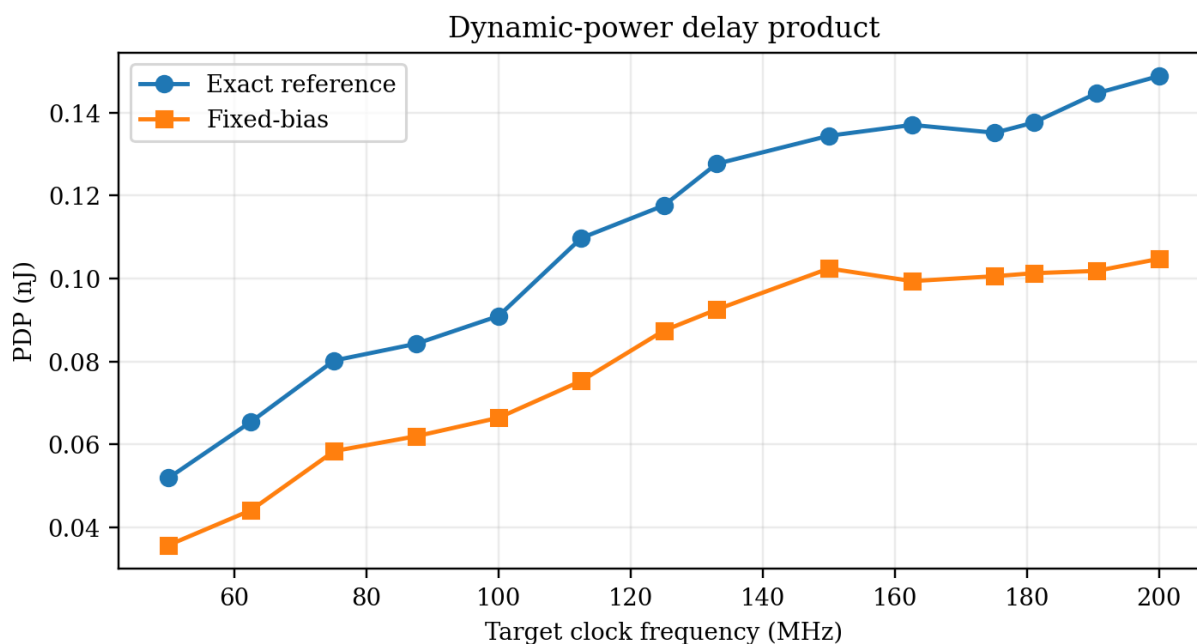


Figure 7: Dynamic-power delay product across the timing sweep

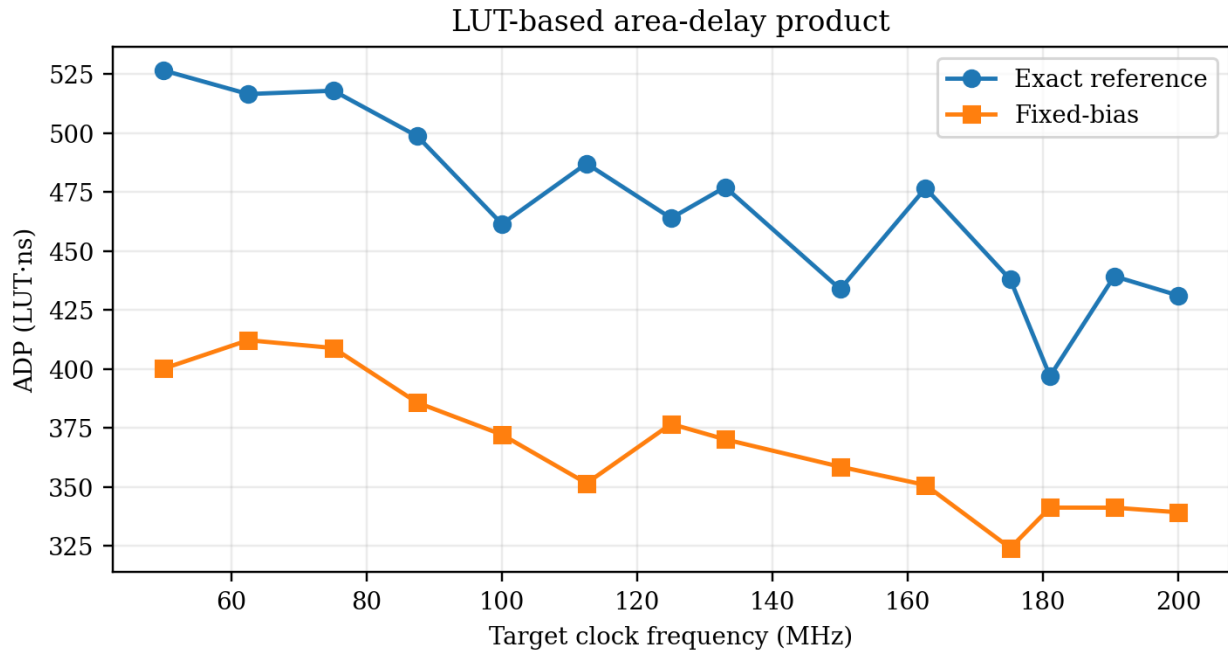


Figure 8: LUT-based area-delay product across the timing sweep

Table II: Representative composite-metric comparison

MHz	PDP EX	PDP FB	PDP ↓%	ADP EX	ADP FB	ADP ↓%
100	0.0910	0.0664	26.97	461.43	372.12	19.35
133	0.1277	0.0925	27.52	477.05	370.10	22.42
180.995	0.1376	0.1013	26.38	396.90	341.18	14.04
190.5	0.1446	0.1018	29.59	439.27	341.16	22.33
200	0.1488	0.1047	29.62	431.09	339.18	21.32

Now Figure 7, plots graphs for the PDP versus target clock frequency. It is reported that at 200 MHz, PDP declines from 0.1488 to 0.1047 nJ, that results to a 29.62% decrease, whereas Figure 8, depicting the ADP behavior with respect to target clock frequency and it decreases from 431.09 to 339.18 LUT·ns, which is a 21.32% decrease. These composite metrics also shown through the table II, states that the key advantage of fixed bias is a constantly less power–area achievements with highest timing-constraint margins.

V. CONCLUSION

An Accomplish process for an approximate multiplier with 8-bit operands and CUT=6 for fixed-bias truncated mechanism has been presented and evaluated on the Zynq-7000 platform for complete RTL-to-FPGA. The study integrates one controlled experimental framework for both the design models to obtain the results. The investigated study concludes that the fixed bias is constantly reduces power and LUT metrics. At 200 MHz, 0.021 W dynamic power and 68 LUTs in case of fixed biased and compared with 0.029 W and 84 LUTs in case of the exact core realization, and having +0.012 ns WNS. In the reported results it can be seen clearly that the PDP and ADP values improve by 29.62% and 21.32%, respectively at 200 MHz.

REFERENCES

- [1] H. Jiang, F. J. H. Santiago, H. Mo, L. Liu, and J. Han, "Approximate Arithmetic Circuits: A Survey, Characterization, and Recent Applications," *Proc. IEEE*, vol. 108, no. 12, pp. 2108–2135, Dec. 2020, doi: 10.1109/JPROC.2020.3006451.
- [2] S. Hashemi, R. I. Bahar, and S. Reda, "DRUM: A Dynamic Range Unbiased Multiplier for Approximate Applications," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design (ICCAD)*, 2015, pp. 418–425, doi: 10.1109/ICCAD.2015.7372600.
- [3] P. Kulkarni, P. Gupta, and M. Ercegovic, "Trading Accuracy for Power With an Underdesigned Multiplier Architecture," in *Proc. 24th Int. Conf. VLSI Design*, 2011, pp. 346–351, doi: 10.1109/VLSID.2011.51.

- [4] R. Zendegani, M. Kamal, M. Bahadori, A. Afzali-Kusha, and M. Pedram, "RoBA Multiplier: A Rounding-Based Approximate Multiplier for High-Speed Yet Energy-Efficient Digital Signal Processing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 2, pp. 393–401, Feb. 2017, doi: 10.1109/TVLSI.2016.2587696.
- [5] S. Mittal, "A survey of techniques for approximate computing," *ACM Computing Surveys*, vol. 48, no. 4, 2016.
- [6] A.G.M. Strollo et al., "A modular Vedic multiplier architecture for model-based design and deployment on FPGA platforms," *Microprocessors and Microsystems*, vol. 76, 103106, 2020.
- [7] N. Petra, D. De Caro, V. Garofalo, E. Napoli, and A. G. M. Strollo, "Truncated Binary Multipliers With Variable Correction and Minimum Mean Square Error," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 6, pp. 1312–1325, Jun. 2010, doi: 10.1109/TCSI.2009.2033536.
- [8] V. Garofalo, N. Petra, and E. Napoli, "Analytical Calculation of the Maximum Error for a Family of Truncated Multipliers Providing Minimum Mean Square Error," *IEEE Trans. Comput.*, vol. 60, no. 9, pp. 1366–1371, Sep. 2011, doi: 10.1109/TC.2010.236.
- [9] G. Zervakis, K. Tsoumanis, S. Xydis, D. Soudris, and K. Pekmestzi, "Design-Efficient Approximate Multiplication Circuits Through Partial Product Perforation," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 10, pp. 3105–3117, Oct. 2016, doi: 10.1109/TVLSI.2016.2535398.
- [10] S. Vahdat, M. Kamal, A. Afzali-Kusha, and M. Pedram, "TOSAM: An Energy-Efficient Truncation-and Rounding-Based Scalable Approximate Multiplier," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 27, no. 5, pp. 1161–1173, May 2019, doi: 10.1109/TVLSI.2018.2890712.
- [11] A.Momeni, J. Han, P. Montuschi, and F. Lombardi, "Design and Analysis of Approximate Compressors for Multiplication," *IEEE Trans. Comput.*, vol. 64, no. 4, pp. 984–994, Apr. 2015, doi: 10.1109/TC.2014.2308214.
- [12] K. V. Gowreesrinivas, S. Srinivas, and P. Samundiswary, "FPGA implementation of a resource efficient Vedic multiplier using SPST adders," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10698–10702, 2023.
- [13] Xilinx, Vivado Design Suite User Guide: Design Analysis and Closure Techniques, UG906, documentation applicable to Vivado 2019.2.
- [14] Avnet, ZedBoard Hardware User Guide, Zynq-7000 All Programmable SoC development platform.

Citation of this Article:

Kuldeep Singh, Vinod Kumar, & Ajay Kumar. (2026). A Multi-Frequency Performance Evaluation of an 8-Bit Fixed-Bias Truncated Approximate Multiplier Design on FPGA. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(9), 92-99. Article DOI <https://doi.org/10.47001/IRJIET/2026.109010>
