A Low-Power and High Accuracy Approximate Multiplier with Reconfigurable Truncation

[1]N. Sree Bhavani, [2]N. Sreeja, [3]N. Ravi Kiran, [4]D. Naresh

[1]Student, [2]Student, [3]Student, [4]Asst Professor Department of Electronics & Communication Engineering
Guru Nanak Institutions Technical Campus (Autonomous) Ibrahimpatnam, Hyderabad, Telangana, 501506.

ABSTRACT: Multipliers are among the most critical arithmetic functional units in many applications, and those applications commonly require many multiplications which result in significant power consumption. For applications that have error tolerance, employing an approximate multiplier is an emerging method to reduce critical path delay and power consumption. An approximate multiplier can trade off accuracy for lower energy and higher performance. In this paper, we not only propose an approximate 4-2 compressor with high accuracy, but also an adjustable approximate multiplier that can dynamically truncate partial products to achieve variable accuracy requirements. In addition, we also propose a multiplier and accumulation (MAC)unit. The proposed MAC using approximate multiplier can adjust the accuracy and power required for multiplications at run-time based on the users' requirement.

Keywords: Approximate, Lower energy, Higher performance, Multiplier and Accumulator (MAC) unit, 4-2 Compressor

I. Introduction:

Multipliers are among the most critical arithmetic functional units in many applications, such as digital signal processing (DSP), computer vision, multimedia processing, image recognition, and artificial intelligence. Those applications commonly need numerous multiplications that result in huge power consumption. The high-power consumption is a challenge for implementing those applications, especially on mobile devices. Therefore, many studies have proposed techniques to reduce the power consumption of multiplier circuits. One solution to reduce the power consumption of a multiplier is to approximate multiplication if the targeted applications allow error tolerance, or in other words, if they are related to human senses. Due to the human sensory limitations, such as limited viewing spectrum and hearing range, the accurate computing results are not necessary.

II. Literature Survey

A.G. M. Strollo *et al.* proposed approximate 4–2 compressors for low-power multipliers, highlighting Recursive Multipliers (RMs) as energy-efficient designs with configurable power–accuracy trade-offs. This work explores approximate recursive multipliers using 2×2 building blocks, addressing an open research problem. Two novel 2×2 approximate multipliers are introduced, achieving 52% area reduction and 25% delay improvement with bounded error. Using these, three 8×8 multipliers are designed with varying accuracy levels. The proposed designs show significant improvements in MRED and power-delay product compared to existing methods. When applied to convolutional neural networks, the designs demonstrate up to 32.64% power savings with improved classification accuracy.

III. Proposed System

Even though the input width of the multiplier is designed to be 8-bit, the proposed technique can still be extended to larger multipliers. The proposed approximate multiplier contains three stages. In the first stage, each partial product is generated by two 2-input AND gates as shown previously in FIGURE 5 with the gate sharing technique applied to further reduce hardware costs. Depending on the requirements, the accuracy of the generated partial product can be determined

based on the Trunc signal. In our proposed approximate multiplier, to make the control more efficient as well as to reduce the hardware costs, we design a 4-bit Trunc signal with each bit to control more than one partial product column, which we call ‘‘3-4-4-4 partition’’, specifically, each bit from MSB to LSB to control column 14th∼12th, 11th ∼8th, 7th∼4th and 3rd ∼0th respectively, corresponding to the color of khaki, sky blue, green and black in Stage 2 in FIGURE 7. For example, if the Trunc (3−0) is 01012, column 14th
∼12th and 7th ∼4th are accurate, and column 11th ∼8th and 3rd ∼0th are truncated.

IV. Design Methodology

The Multiply–Accumulate (MAC) unit is a fundamental component in many Digital Signal Processing (DSP) applications, enabling efficient execution of operations such as convolution, filtering, and inner products. Advanced DSP algorithms, including Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT), rely heavily on repeated multiplication and addition operations; hence, overall system performance is largely determined by the speed of these arithmetic computations.

The MAC unit enhances signal processing capabilities in microcontrollers for applications such as audio and servo control. It typically employs a 3-stage pipelined architecture optimized for high-speed multiplication. The unit supports signed and unsigned integer operations, as well as signed fixed-point fractional computations.

A typical MAC architecture consists of a multiplier, adder, and accumulator register. Input data is fetched from memory, processed through the multiplier, and accumulated via the adder, with results stored back in memory. For high-performance DSP systems, this entire operation is ideally completed within a single clock cycle.

In this work, an approximate multiplier is utilized within a 16-bit MAC architecture to improve computational efficiency. The 16-bit inputs are sourced from memory, multiplied, and the resulting product is accumulated to produce the final output.

The MAC unit function is represented by the following equation:

$$F = \sum P_i Q_i$$

V. Tools and Technologies Used

- Verilog HDL
- Xilinx Vivado
- FPGA-based simulation environment
- VLSI design methodology

Xilinx software is used by the VHDL/VERILOG designers for performing Synthesis operation. Any simulated code can be synthesized and configured on FPGA. Synthesis is the transformation of HDL code into gate level net list. It is an integral part of current design flows.

VI. Implementation

This work proposes a high-accuracy, low-power approximate 4–2 compressor with optimized carry and sum generation. Since carry errors contribute higher error distance, the carry is designed to be always accurate using shared intermediate signals. The sum is approximated using simplified logic, reducing area and power while maintaining bounded error. The design introduces minimal error (only when all inputs are 1) with an error distance of 1, making it negligible for practical applications.

An area-efficient implementation further reduces complexity by reusing intermediate signals and employing a MUX-based approach. Additionally, a dynamic input truncation technique is proposed to enable runtime control of accuracy and power by selectively truncating partial products, thereby lowering hardware activity and energy consumption.
The proposed design achieves significant improvements in area, delay, and power compared to exact compressors. Integrated into an 8×8 multiplier and pipelined MAC architecture, it delivers high throughput and efficiency, making it suitable for error-tolerant applications such as image processing and neural networks.

VII. Results

RTL Schematic

The HDL language is used to convert the description or summery of the architecture to the working summery by use of the coding language i.e Verilog, VHDL. The RTL schematic even specifies the internal connection blocks for better analyzing. The figure represented below shows the RTL schematic diagram of the designed architecture.

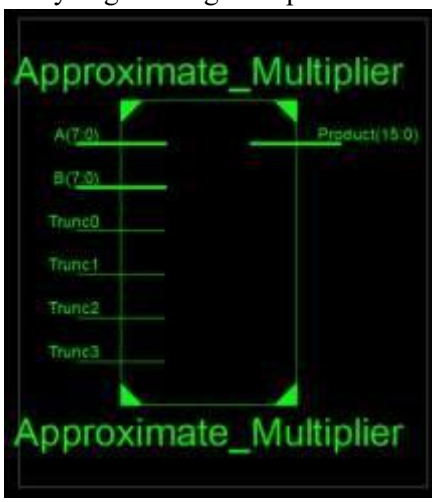

Fig. RTL Schematic of approximate multiplier

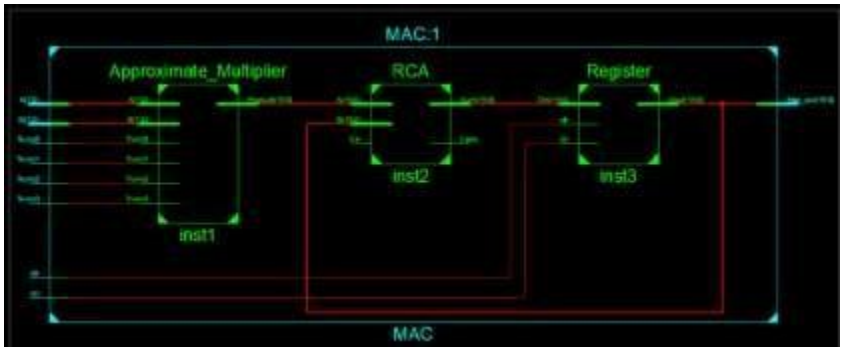

Fig. RTL Schematic of MAC using approximate multiplier

The technology schematic makes the representation of the architecture in the LUT format, where the LUT is consider as the parameter o area that is used in VLSI to estimate the architecture design. The LUT is consider as an square unit the memory allocation of the code is represented in there LUTs in FPGA.

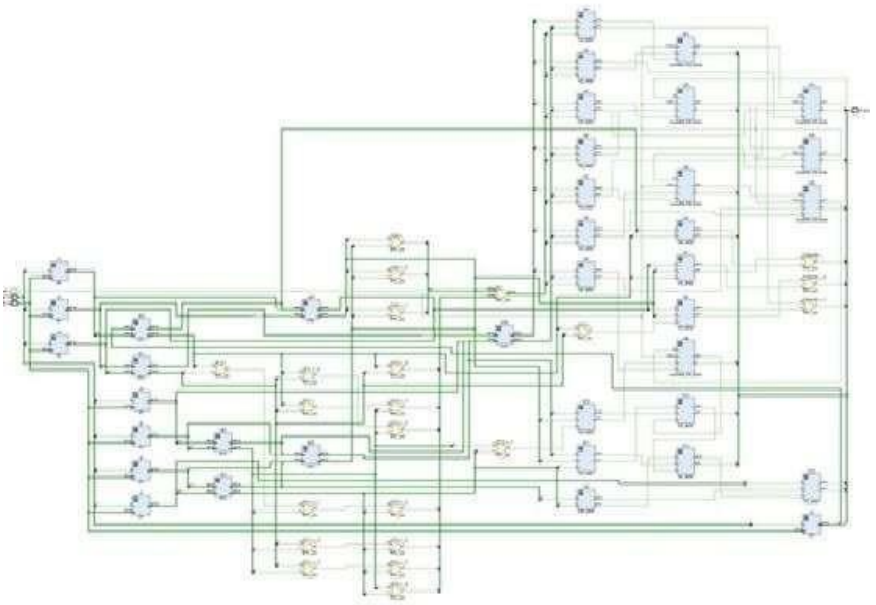

Fig. View Technology schematic of approximate multiplier

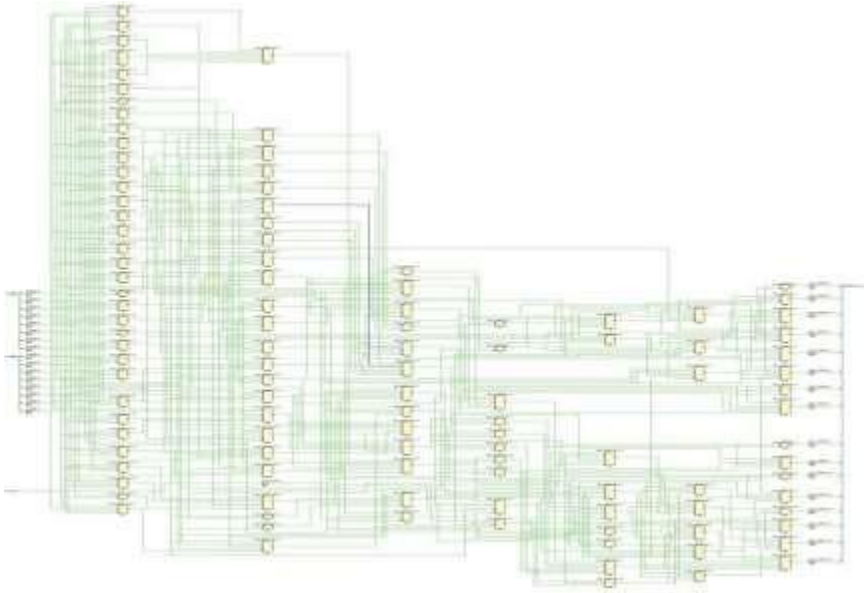

Fig. View Technology schematic of MAC using approximate multiplier

Simulation

The simulation is the process which is termed as the final verification in respect to its working where as the schematic is the verification of the connections and blocks. The simulation window is launched as shifting from implantation to the simulation on the home screen of the tool, and the simulation window confines the output in the form of the wave forms. Here it has the flexibility of providing the different radix number systems.

Fig. Simulation wave forms of approximate multiplier

Fig. Simulation wave forms of MAC using approximate multiplier

VIII. Conclusion

This project presents MAC unit using approximate multiplier with novel approach of approximate 4:2 compressor architectures. In this paper, we not only propose an approximate 4-2 compressor with high accuracy, but also an adjustable approximate multiplier that can dynamically truncate partial products to achieve variable accuracy requirements. In addition, we also propose a multiplier and accumulation (MAC)unit. The proposed MAC using approximate multiplier can adjust the accuracy and power required for multiplications at run-time based on the users' requirement. Basically, it is quite challenging to design an approximate multiplier with absolute benefit, and the optimal answer is typically the one that best suits the target application.

IX. References

[1] A. Bosio, D. Ménard, and O. Sentieys, Eds. Approximate Computing Techniques: From Component-to Application-Level. Cham, Switzerland: Springer, 2022. [Online]. Available: https://link.springer.com/book/10.1007/978-3-030-94705-7

[2] A. G. M. Strollo, E. Napoli, D. De Caro, N. Petra, and G. D. Meo, "Comparison and extension of approximate 4-2 compressors for lowpower approximate multipliers," IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 67, no. 9, pp.

3021–3034, Sep. 2020.

[3] T. Kong and S. Li, "Design and analysis of approximate 4-2 compressors for high-accuracy multipliers," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 29, no. 10, pp. 1771–1781, Oct. 2021.

[4] 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.

[5] F. Sabetzadeh, M. H. Moaiyeri, and M. Ahmadinejad, "A majority-based imprecise multiplier for ultra- efficient approximate image multiplication," IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 66, no. 11, pp. 4200– 4208, Nov. 2019.

[6] H. Pei, X. Yi, H. Zhou, and Y. He, "Design of ultra-low power consumption approximate 4-2 compressors based on the compensation characteristic," IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 68, no. 1, pp. 461– 465, Jan. 2021.

[7] D. Esposito, A. G. M. Strollo, E. Napoli, D. de Caro, and N. Petra, "Approximate multipliers based on new approximate compressors," IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 65, no. 12, pp. 4169–4182, Dec. 2018.

[8] U. Anil Kumar, S. K. Chatterjee, and S. E. Ahmed, "Lowpower compressor-based approximate multipliers with error correcting module," IEEE Embdded Syst. Lett., vol. 14, no. 2, pp. 59–62, Jun. 2022.

[9] X. Yi, H. Pei, Z. Zhang, H. Zhou, and Y. He, "Design of an energyefficient approximate compressor for error-resilient multiplications," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), 2019, pp. 1–5.

[10] M. Ha and S. Lee, "Multipliers with approximate 4-2 compressors and error recovery modules," IEEE Embdded Syst. Lett., vol. 10, no. 1, pp. 6–9, Mar. 2018.

[11] M. Ahmadinejad, M. H. Moaiyeri, and F. Sabetzadeh, "Energy and area efficient imprecise compressors for approximate multiplication at nanoscale," AEU Int. J. Electron. Commun., vol. 110, Oct. 2019, Art. no. 152859

[12] D. Mohapatra, V. K. Chippa, A. Raghunathan, and K. Roy, ''Design of voltage-scalable meta-functions for approximate computing,'' in Proc. Design, Autom. Test Eur., Mar. 2011, pp. 1–6.

[13] K. Yin Kyaw, W. Ling Goh, and K. Seng Yeo, ''Low-power high-speed multiplier for error-tolerant application,'' in Proc. IEEE Int. Conf. Electron Devices Solid-State Circuits (EDSSC), Dec. 2010, pp. 1–4.

[14] M. de la Guia Solaz, W. Han, and R. Conway, ''A flexible low power DSP with a programmable truncated multiplier,'' IEEE Trans. Circuits Syst., vol. 59, no. 11, pp. 2555–2568, Nov. 2012.

[15] R. Zendegani, M. Kamal, M. Bahadori, A. Afzali-Kusha, and M. Pedram, ''RoBa multiplier: A rounding- based approximate multiplier for highspeed yet energy-efficient digital signal processing,'' IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 25, no. 2, pp. 393–401, Feb. 2017.

[16] C. S. Wallace, ''A suggestion for a fast multiplier,'' IEEE Trans. Electron. Comput., vol. EC-13, no. 1, pp. 14–17, Feb. 1964.

[17] A. Weinberger, ''4:2 carry-save adder module,'' IBM Tech. Discl. Bull., vol. 23, no. 8, pp. 3811–3814, 1981.

[18] 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.

[19] Z. Yang, J. Han, and F. Lombardi, ''Approximate compressors for errorresilient multiplier design,'' in Proc. IEEE Int. Symp. Defect Fault Tolerance VLSI Nanotechnol. Syst. (DFTS), Oct. 2015, pp. 183–186.

[20] C.-H. Lin and I.-C. Lin, ''High accuracy approximate multiplier with error correction,'' in Proc. IEEE 31st Int. Conf. Comput. Design (ICCD), Oct. 2013, pp. 33–38.

[21] P. J. Edavoor, S. Raveendran, and A. D. Rahulkar, ''Approximate multiplier design using novel dual-stage 4:2 compressors,'' IEEE Access, vol. 8, pp. 48337–48351, 2020.

[22] F. Sabetzadeh, M. H. Moaiyeri, and M. Ahmadinejad, ‘‘A majority-based imprecise multiplier for ultra-efficient approximate image multiplication,’’ IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 66, no. 11, pp. 4200– 4208, Nov. 2019.

[23] A. G. M. Strollo, E. Napoli, D. De Caro, N. Petra, and G. D. Meo, ‘‘Comparison and extension of approximate 4–2 compressors for lowpower approximate multipliers,’’ IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 67, no. 9, pp. 3021–3034, Sep. 2020.

[24] H. Xiao, H. Xu, X. Chen, Y. Wang, and Y. Han, ‘‘Fast and high-accuracy approximate MAC unit design for CNN computing,’’ IEEE Embedded Syst. Lett., early access, Dec. 21, 2021, doi: 10.1109/LES.2021.3137335.

[25] O. Akbari, M. Kamal, A. Afzali-Kusha, and M. Pedram, ‘‘Dual-quality 4:2 compressors for utilizing in dynamic accuracy configurable multipliers,’’ IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 25, no. 4, pp. 1352–1361, Apr. 2017.

[26] T. Yang, T. Ukezono, and T. Sato, ‘‘A low-power high-speed accuracycontrollable approximate multiplier design,’’ in Proc. 23rd Asia South Pacific Design Autom. Conf. (ASP-DAC), Jan. 2018, pp. 605–610.

[27] I. Hammad, L. Li, K. El-Sankary, and W. M. Snelgrove, ‘‘CNN inference using a preprocessing precision controller and approximate multipliers with various precisions,’’ IEEE Access, vol. 9, pp. 7220–7232, 2021.

[28] C. Guo, L. Zhang, X. Zhou, W. Qian, and C. Zhuo, ‘‘A reconfigurable approximate multiplier for quantized CNN applications,’’ in Proc. 25th Asia South Pacific Design Autom. Conf. (ASP-DAC), Jan. 2020, pp. 235–240.