IMPACT: International Journal of Research in Engineering and Technology
ISSN (P): 2347–4599; ISSN (E): 2321–8843
Vol. 10, Issue 6, Jun 2022, 1–8

HIGH PERFORMANCE MAC BASED ON MULTI-LEVEL APPROXIMATE COMPRESSORS WITH BALANCED ERROR ACCUMULATION

Aneesath. K P[1], Manivannan. R[2], Arun Prasath. N[3] & G. Ranjithkumar[4]

[1]PG Student, EASA College of Engineering and Technology, Coimbatore, Tamilnadu, India

[2,4]Assistant Professor, Department of EEE, EASA College of Engineering and Technology, Coimbatore, Tamilnadu, India

[3]Senior Assistant Professor, Department of ECE, EASA College of Engineering and Technology, Coimbatore, Tamilnadu, India

Received: 07 Jun 2022 **Accepted:** 07 Jun 2022 **Published:** 07 Jun 2022

ABSTRACT

It is presented a unique approximate computing approach for implementing energy-efficient multiply-accumulate (MAC) processing. We use different approximate multipliers in an interleaved manner to mitigate mistakes in the opposite direction during accumulate operations, in contrast to previous efforts that suffer from error accumulation restricting the approximation range. We first build approximation 4-2 compressors that generate error in the opposite direction while minimizing computing costs for balanced error accumulation. Positive and negative multipliers are then carefully built based on the probabilistic analysis to produce a similar error distance. Further, additional to 4-2 compressors 5-2, 6-3 and 7-3 approximate compressors and counters are also used to optimize area and power. Simulation findings on a variety of real-world applications show that the proposed MAC processing extends the range of approximate components, resulting in a more energy-efficient computing situation. Even when compared to state-of-the-art alternatives, the suggested interleaving scheme reduces the latest CNN accelerator's core-level energy consumption by more than 35% without compromising recognition accuracy.

KEYWORDS: *Approximate Computing, Convolutional Neural Network, Image Processing, Low-Power Circuit Design, Multiplier*

INTRODUCTION

Approximate computing has been highlighted as a viable technique to reduce the energy consumption of certain signal processing algorithms, such as machine learning and multimedia digital signal processing, that are known to have error-tolerable properties [1]–[6]. These multimedia-related algorithms mostly involve heavy matrix multiplications, necessitating the development of a cost-effective approximation multiply-accumulate (MAC) operator that consumes the least amount of energy for a given amount of computational mistakes [7].Studies adders have also been proposed [8], however due to the difficulty of implementing each arithmetic unit, previous research has focused on easing multiplier costs, which dissipate far more energy than the addition operation [9]–[16]. Internal compressors, which are unavoidable for constructing a fast multiplier by minimising the number of partial products (PPs) [17], account for a substantial share of multiplier expenditures.

One of the key design criteria in practically any electronic device, especially portable ones like smart phones, tablets, and other gadgets, is energy minimization. It is extremely desirable to achieve this minimization with the least amount of performance (speed) loss possible. These portable devices' digital signal processing (DSP) blocks are essential for achieving a variety of multimedia applications. The arithmetic logic unit is the computational heart of these blocks, with multiplications accounting for the majority of arithmetic operations in these DSP systems. As a result, increasing the speed and power/energy efficiency of multipliers is critical to increasing the efficiency of processors.

PROPOSED APPROXIMATE MULTIPLIERS WITH MODIFIED COMPRESSORS

AND gates are used to create the PPs in their most basic form. The PPs will then be separated into two rows using a two-step process. PP reduction uses 4:2, 5:2, 6:3 and 7:3 compressors for minimizing the logical complexity. Approximation is done in the LSB part to minimize area and power with tolerable error. Fig. 1 illustrates the proposed block diagram.

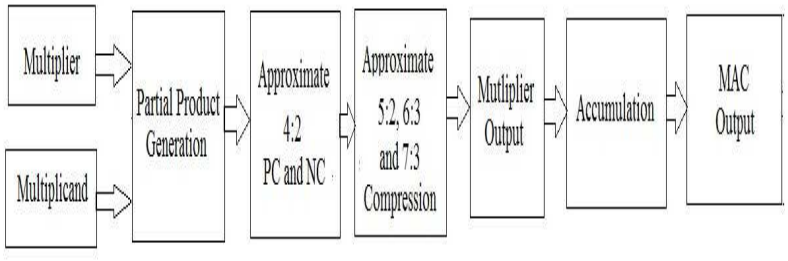

Figure 1: Block Diagram of Proposed Multiplier.

Because of their simple and regular topologies, compressor designs have been primarily used to realize practical multipliers [17]–[21]. The precise p-q compressor decreases the number of PPs from p to q with the assisted carry-in/out values after creating the initial PPs with bit-wise AND operations between the multiplicand A and multiplier X, as stated in [17]. Four PPs along the same I -th bit position designated as ai, bi, ci, and di yield two PPs over two columns (yi and yi+1) for the following stage in the case of the 4-2 compressor, which is extensively employed in real multi-stage multipliers.

$$\begin{aligned} y_i &= (a_i \oplus b_i) \oplus (c_i \oplus d_i) \oplus z_i, \\ y_{i+1} &= (a_i \oplus b_i \oplus c_i \oplus d_i) z_i + \overline{(a_i \oplus b_i \oplus c_i \oplus d_i)} d_i, \\ z_{i+1} &= (a_i \oplus b_i) c_i + \overline{(a_i \oplus b_i)} a_i, \end{aligned} \quad (1)$$

Where zi and zi+1 are the exact 4-2 compressor's carry-in and carry-out signals at the i-th bit position, respectively [20]. It's worth noting that achieving an identical 4-2 compressor essentially demands two full adders, which dominates the multiplier's total complexity [21]. As a result, modern approximation multipliers often simplify compressor designs for pre-defined approximate portions, allowing for imprecise operations.

$$\begin{aligned} \tilde{y}_i &= \overline{a_i \oplus b_i} + \overline{c_i \oplus d_i}, \\ \tilde{y}_{i+1} &= \overline{(a_i + b_i)} + \overline{(c_i + d_i)}, \end{aligned} \quad (2)$$

Where ˜yi and ˜yi+1 are the approximate results of the simplified compressor.

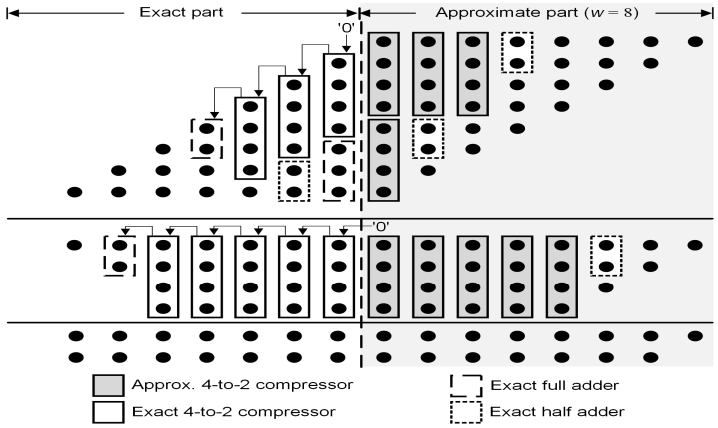

Figure 2: The Basic Design of 8 × 8 Approximate Multiplication.

The 4-2 compressor-based 8-bit approximate multiplier is theoretically illustrated in Fig. 2, where w denotes the number of approximate bits. While increasing w resulted in more energy-efficient multiplier designs, the number of faults also grows significantly. We build two types of approximate multipliers based on the error direction using the suggested approximation compressors: the positive multiplier (PM) and the negative multiplier (NM) (NM). The suggested implementation of 8 × 8 positive multiplier, is shown in Fig. 3(a). On the other hand, in the suggested 16 × 16 PM design, we combine PC and NC to reduce peak errors while still providing a biased error distribution in the positive direction. The error value of EMUL = AM(A, X)AX is then generated by the simplified multiplication procedure, where AM(A, X) reflects the results of the approximate multiplication of the input multiplicand A and multiplier X. It is also possible to perform a probabilistic analysis of EMUL based on the estimated error of each approximation compressor.

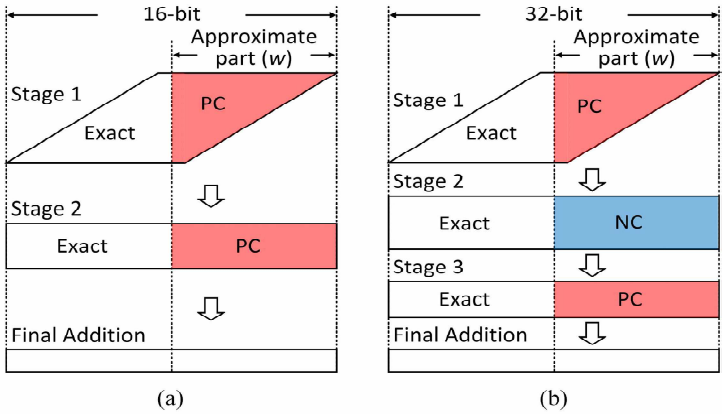

Figure 3: The Conceptual Architecture of (a) The Proposed 8 × 8 Positive Multiplier, and (b) The Proposed 16 × 16 Positive Multiplier.

To be more exact, we calculate the multiplier's estimated column-level error ei by adding all of the approximation compressors' predicted errors in the I -th bit position. The approximate multiplier will then generate the mistakes by taking into account the weight of each bit location.

5-2 Compressor

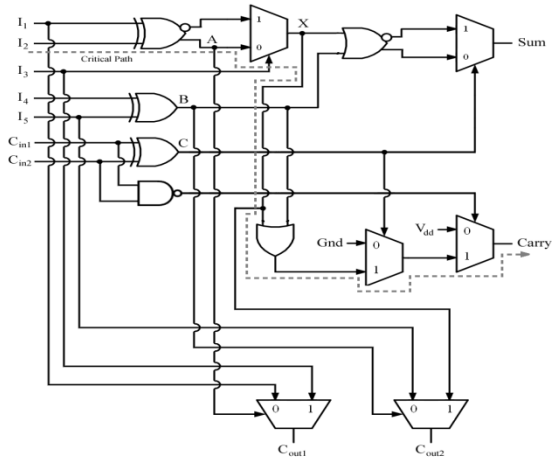

Figure 4: 5-2 Compressor.

The diagram of a 5:2 compressor is shown in Figure 4. Both outputs are generated concurrently by the two output XOR–XNOR gates, which are developed and thoroughly detailed. Their varied outputs have little effect on the delay and glitch effect of TG waveforms due to their comparable pathways. The architectures of NAND and NOR gates have also been designed using static CMOS. The AND/OR gates will be obtained if the outputs of these gates are inverted, and because these gates are not positioned in the critical route, they will not effect the overall system latency.

6:3 and 7:3 Counter

Figures 5 and 6 demonstrate how to build the 6:3 and 7:3 counter circuits, respectively. The critical route latency is lowered to seven basic gates by using bigger CMOS gates. This 6:3 counter beats conventional designs because there are no XOR gates on the critical route. The increased wiring complexity is one disadvantage of this approach. The suggested 6:3 counter based on bit stacking operates roughly 30% quicker than any existing counter designs since it has no XOR gates on its critical route. As a result of this new approach of counting using bit stacking, a counter can be built with a significant performance boost while consuming less power.

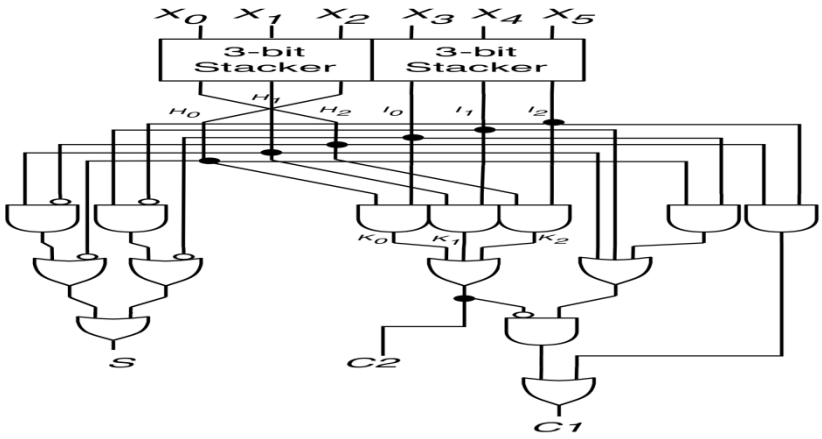

Figure 5: 6:3 Counter.

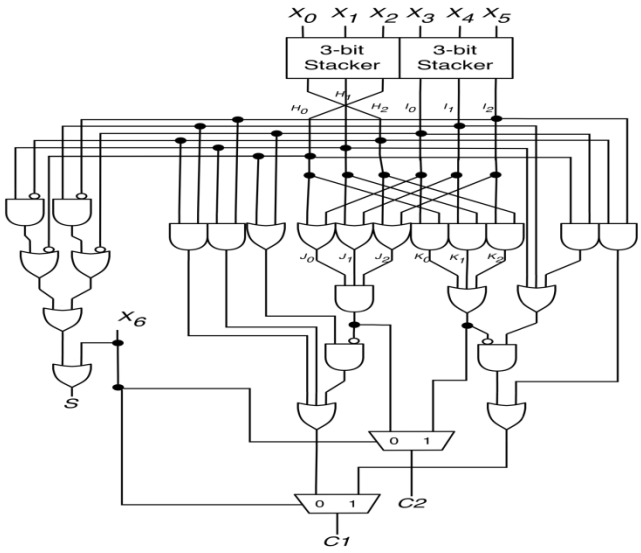

Figure 6: 7:3 Counter.

APPROXIMATE MAC PROCESSING

As the hardware complexity of multipliers, in general, dominates the overall computing expenses of today's parallel MAC operators, [20],. In this paper, we create a unique interleaving method of approximate multipliers with opposite error orientations to alleviate the biassed error accumulation generated by earlier approximate multipliers, resulting in a balanced error distribution during MAC operations. We first observe E(EPMUL) and E(ENMUL) for the provided approximate range w to estimate the blending ratio of two approximate multipliers, which is denoted as ρ = E(ENMUL)/E(EPMUL).

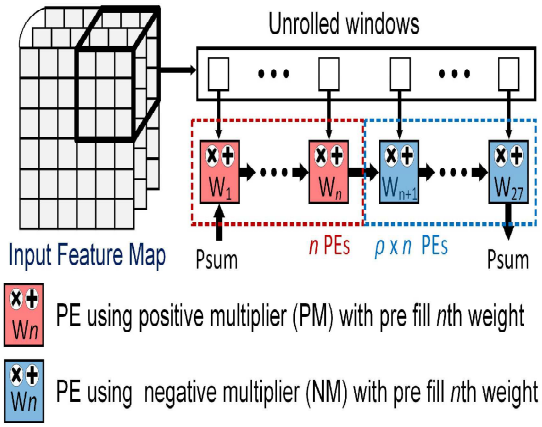

Figure 7: The Proposed MAC.

EXPERIMENTAL RESULTS

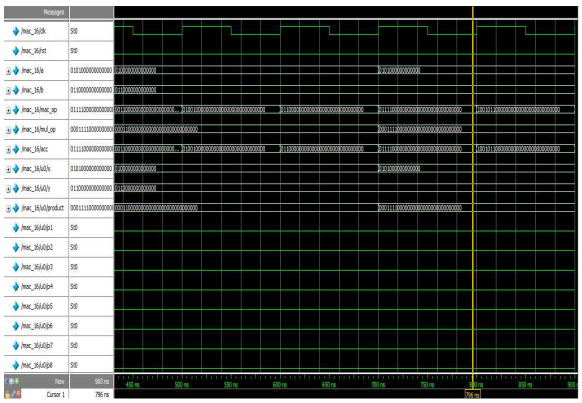

Figure 8: Simulation Result of Propose MAC.

Fig.8 shows the simulation result of proposed MAC unit. By utilizing the proposed approximate compressors, approximate multipliers is implemented depending on the error direction, denoted as the positive multiplier (PM) and the negative multiplier (NM). It also uses multi-level 52-63-73 compressors and counters for reducing the partial product to get optimized area and power.

Table 1: Comparison Results of MAC

Parameters	Area (Gate Count)	Power (mW)	Delay(ns)
Existing MAC	3415	162.19	6.237
Proposed MAC	836	136.42	5.479

Table 1 shows the various parameters like area, power and delay of MAC designed using proposed multi-level compressor based approximate multiplier. From different types of compressors and counters the proposed designed is optimized to get reduce area, delay and power values.

CONCLUSION

Novel approximate multipliers based on multi-level approximate compressors are proposed for use in MAC-oriented signal processing algorithms to reduce energy consumption. First, two compressor types are devised to reduce hardware costs while causing errors in different directions, and then the related approximation multipliers are extensively evaluated to forecast the quantity of errors in a probabilistic manner. Then 5:2, 6:3 and 7:3 approximate compressors are applied to existing design to reduce the complexity. In contrast to prior efforts that suffered from accumulated errors, the suggested interleaving method uses two simple multipliers based on the blending ratio to generate a narrow and balanced error distribution. The proposed method is better suitable for implementing approximate computing in various case studies, successfully saving processing energy with low performance loss, according to simulation results.

REFERENCES

1. *P. N. Whatmough, S. K. Lee, H. Lee, S. Rama, D. Brooks, and G.-Y. Wei, "14.3 A 28 nm SoC with a 1.2 GHz 568 nJ/prediction sparse deep-neural-network engine with >0.1 timing error rate tolerance for IoT applications," in IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers, Feb. 2017, pp. 242–243.*
2. *V. Gupta, D. Mohapatra, A. Raghunathan, and K. Roy, "Low-power digital signal processing using approximate adders," IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 32, no. 1, pp. 124–137, Jan. 2013.*

3. *M. Kazhdan, "An approximate and efficient method for optimal rotation alignment of 3D models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 7, pp. 1221–1229, Jul. 2007.*
4. *M. J. Wainwright and M. I. Jordan, "Log-determinant relaxation for approximate inference in discrete Markov random fields," IEEE Trans. Signal Process., vol. 54, no. 6, pp. 2099–2109, Jun. 2006.*
5. *J. Jo, J. Kung, and Y. Lee, "Approximate LSTM computing for energy efficient speech recognition," Electronics, vol. 9, no. 12, p. 2004, Nov. 2020.*
6. *B. K. Mohanty, "Parallel VLSI architecture for approximate computation of discrete Hadamard transform," IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 12, pp. 4944–4952, Dec. 2020.*
7. *X. Si et al., "15.5 A 28 nm 64Kb 6T SRAM computing-in-memory macro with 8b MAC operation for AI edge chips," in IEEE Int. Solid- State Circuits Conf. (ISSCC) Dig. Tech. Papers, Feb. 2020, pp. 246–248.*
8. *H. Jiang, J. Han, and F. Lombardi, "A comparative review and evaluation of approximate adders," in Proc. 25th Ed. Great Lakes Symp. (VLSI), May 2015, pp. 343–348.*
9. *N.Arun Prasath, N.Kumaresan," A Novel Approach on Greedy Maximal Scheduling Algorithm on Embedded Networks," Networks and Complex Systems (IISTE), Vol.5,Issue 3,pp 15-20,March 2015*
10. *V. Leon, K. Asimakopoulos, S. Xydis, D. Soudris, and K. Pekmestzi, "Cooperative arithmetic-aware approximation techniques for energy efficient multipliers," in Proc. 56th Annu. Design Autom. Conf. (DAC), Jun. 2019, pp. 1–6.*
11. *V. Leon, G. Zervakis, S. Xydis, D. Soudris, and K. Pekmestzi, "Walking through the energy-error Pareto frontier of approximate multipliers," IEEE Micro, vol. 38, no. 4, pp. 40–49, Jul. 2018.*
12. *S. Narayanamoorthy, H. A. Moghaddam, Z. Liu, T. Park, and N. S. Kim, "Energy-efficient approximate multiplication for digital signal processing and classification applications," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 23, no. 6, pp. 1180–1184, Jun. 2015.*
13. *S. Hashemi, R. I. Bahar, and S. Reda, "DRUM: A dynamic range unbiased multiplier for approximate applications," in Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD), Nov. 2015, pp. 418–425.*
14. *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.*
15. *I. Qiqieh, R. Shafik, G. Tarawneh, D. Sokolov, and A. Yakovlev, "Energy-efficient approximate multiplier design using bit significance driven logic compression," in Proc. Design, Autom. Test Eur. Conf. Exhib. (DATE), Mar. 2017, pp. 7–12.*
16. *M. Wang et al., "An optimized compression strategy for compressor based approximate multiplier," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), Oct. 2020, pp. 1–5.*
17. *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.*

18. *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.*

19. *N. Arun Prasath, K. M. Valarmathy, B. Veerasamy, "A Novel Approach on Wireless DAQ System Using IOT," International Journal of Mechanical and Production Engineering and Research and Development (IJMPERD), Vol.10,Issue 3, pp.7371-7378, August 2020*

20. *Z. Yang, J. Han, and F. Lombardi, "Approximate compressors for error resilient multiplier design," in Proc. IEEE Int. Symp. Defect Fault Tolerance VLSI Nanotechnol. Syst. (DFTS), Oct. 2015, pp. 183–186.*

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