

Power Efficient Approximate Booth Multiplier

Suganthi Venkatachalam
Department of Electrical and
Computer Engineering,
University of Saskatchewan,
Canada.
suganthi.venkat@usask.ca

Hyuk Jae Lee
Department of Electrical and
Computer Engineering,
Seoul National University,
Korea.
hjlee@capp.snu.ac.kr

Seok-Bum Ko
Department of Electrical and
Computer Engineering,
University of Saskatchewan,
Canada.
seokbum.ko@usask.ca

Abstract—Power consumption is an important constraint in multimedia and deep learning applications. Approximate computing offers efficient approach to reduce power consumption. In this paper, novel approximation is proposed for radix-4 booth multiplication. Approximation is introduced in partial product generation and partial product accumulation circuits. Radix-4 partial product generation and accumulation approximation is proposed which remarkably enhances the performance. The proposed approximate booth multiplier achieves 41% area reduction and 49% power reduction compared to an exact booth multiplier. Also, it has better area, power and error metrics compared to existing works on approximate multipliers. The proposed multiplier is evaluated with an image processing application- in Discrete Cosine Transform (DCT) encoding part of JPEG compression and found to perform almost similar to exact multiplication unit.

Keywords—approximate computing, booth multiplication, error analysis, low power.

I. INTRODUCTION

Due to the demand for high computational performance in signal processing and machine learning applications, approximate computing [1] is a potential solution considered in recent times. The idea of approximate computing is to replace exact computational units with approximate ones.

Approximate adders and multipliers are discussed widely in literature. Approximate full adder also known as Lower part OR Adder (LOA) is discussed in [2]. Approximation in carry select adder is proposed in [3]. A segmentation based error tolerant adder is analyzed in [4]. Multipliers are an integral part and more resources consuming operation in applications such as digital signal processing. Extensive research is performed in approximation of multipliers [5-7]. Instead of exact 4-2 compressors, approximate 4-2 compressors are used for dadda multiplier structure in multipliers in [5]. The effects of eliminating one or more rows in the partial product accumulation is extensively analyzed in [6], where unsigned *AND* array multipliers and booth multipliers are approximated. Approximation driven by bit significance in [7] is performed using clustering of two or more rows in the partial product array.

Moreover, approximation of multipliers using truncation to produce fixed width multipliers and using Voltage Over Scaling (VOS) [8] is proposed in the literature. In [8], timing

induced approximation by over scaling the voltage is discussed. Fixed width multipliers produce n most significant bits output for $n \times n$ inputs. Truncation and rounding are performed to produce fixed word size output introducing quantization error. Various techniques are applied to reduce the quantization error after truncation in fixed point multipliers and can be found in literatures [9-11]. In [9], a simple truncation and rounding technique for fixed point multiplier is introduced. In [10], a self-compensation fixed width multiplier with a Fast Fourier Transform application is discussed. A probabilistic estimation bias derived from probabilistic analysis of partial product array to reduce the truncation error is introduced in [11].

Booth multipliers reduce number of partial products almost by half and are widely used. Although truncation of fixed point booth multipliers has received its due attention, very few works have focused on approximation of booth multipliers [6, 12]. In [12], approximation is introduced in generation of radix-8 partial products of booth multipliers. In case of radix-4 booth multiplier approximation, the approximate technique used in [6] relies on approximation by eliminating the generation of few partial product rows and thereby reducing the accumulation hardware.

In our paper, a novel approximation technique is introduced in the generation of radix-4 recoded partial products. After the generation of partial products and corresponding correction terms, *OR* gates are used for efficient approximation of the generated exact and approximate recoded partial products. Although the use of *OR* gates is a popular method of approximation [2, 7], in this work, varying inputs *OR* gates are used in well-fitting positions to get better performance in terms of design and accuracy.

In [13], approximate adders are evaluated and a metric called normalized mean error distance (NMED) is proposed. The performance of approximate multipliers in terms of accuracy is discussed in this paper using NMED and mean relative error distance (MRED). MRED is the mean of relative error distances of all possible input combinations. Area, power and delay performance of the proposed multiplier is analyzed and compared with existing approximation works. The approximate multiplier is tested in the Discrete Cosine Transform (DCT) of JPEG compression of a standard image, and found that it achieves similar PSNR values as an exact multiplier.

The rest of the paper is organized as follows. Section II discusses radix 4 booth multiplier and approximation methods applied to the booth multiplier. Section III covers results and discussion. In section III, the proposed multiplier is compared with exact multiplier and existing approximate multipliers in terms of area, power, NMED and MRED. In section IV, approximated booth multipliers are applied in DCT application of image multiplication and tested. Conclusion and future works are discussed in section V.

II. APPROXIMATION IN RADIX-4 BOOTH MULTIPLIER

A. Radix-4 Booth encoding and partial product generation

Consider two n -bits signed inputs A and B , which produce signed $2n$ bits output P_{out} . The 2's complement A , B and P_{out} can be given as

$$\begin{aligned} A &= -a_{N-1}2^{N-1} + \sum_{n=0}^{N-2}\{a_n\}2^n \\ B &= -b_{N-1}2^{N-1} + \sum_{n=0}^{N-2}\{b_n\}2^n \\ P_{out} &= -P_{2n-1}2^{2n-1} + \sum_{n=0}^{2n-2}\{P_n\}2^n \end{aligned} \quad (1)$$

The input B is grouped into bits $\{b_{2i+1}, b_{2i}, b_{2i-1}\}$ which would belong to one of the encoded values of $\{-2, -1, 0, 1, 2\}$. The encoded values are recoded to signals and based on the signals, partial products are generated. The three signal recoding scheme proposed in [13] is used in this paper. Table I shows three recoded signals neg_i , two_i and $zero_i$ corresponding to $\{b_{2i+1}, b_{2i}, b_{2i-1}\}$ and the partial products P_{ij} generated for all possible values of $\{neg_i, two_i, zero_i\}$ and inputs $a_j a_{j-1}$. The corresponding partial product generation circuit and related K-Map is shown in Figure 1. The generated partial product matrix is shown in Figure 2. In figure 2, the partial products generated after partial product encoding are accumulated with its corresponding correction terms and sign values.

B. Approximation in Booth multipliers

In this paper, an 8- bit approximate booth multiplier is designed and analyzed. The proposed approximation technique can also be applied to large multipliers. The lower part of the partial product matrix is further divided into lower part major (LP_{major}) and lower part minor (LP_{minor}), and varying approximations are applied as shown in Figure 3. The approximation techniques applied to LP_{major} and LP_{minor} are explained in detail in the following sections.

C. LP_{minor} approximation

Approximation is applied at both partial product generation stage and partial product accumulation stage. Novel method to approximate partial product generation is introduced in this paper. The K-map in Figure 1(a) is modified as shown in Figure 4(a). As can be seen, 4 out of 32 cases are altered and highlighted, which greatly reduces the amount of logic circuits. The probability of error is 1/8. By intentionally introducing few errors in partial product generation, the logic complexity is reduced from equation (2) to equation (3). Exact partial product generation logic equation can be given as

$$\begin{aligned} P_{ij} &= \sim zero_i (neg_i (\sim a_j \cdot \sim two_i + \sim a_{j-1} \cdot two_i) \\ &\quad + \sim neg_i (a_j \cdot \sim two_i + a_{j-1} \cdot two_i)) \end{aligned} \quad (2)$$

After approximation, the logic equation is reduced as in (3).

$$P_{ij} = \sim a_j \cdot neg_i + a_j \cdot \sim neg_i \cdot zero_i \quad (3)$$

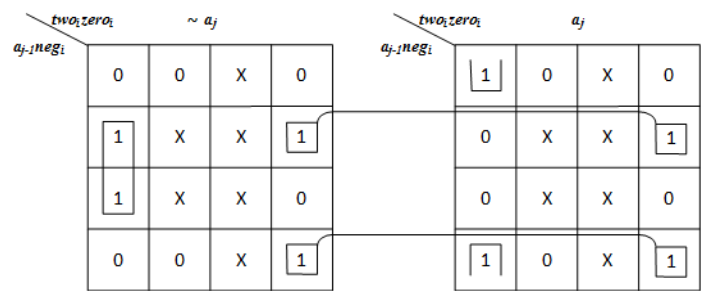
After novel approximation in generation, the generated values are accumulated using OR gates as in Figure 3. Row-wise clustering using OR gates is discussed in [7]. Similar technique is applied to accumulate the partial products. After approximated partial product generation in LP_{minor} , the approximate partial products are accumulated column-wise using varying inputs OR gate, based on the number of elements in each column.

D. LP_{major} approximation

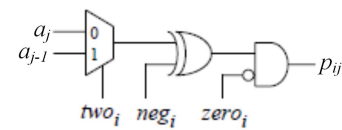
An approximation is proposed, where the correction term is ORed with least partial product of its associated row as in Figure 3. The main purpose of this approximation is to maintain the alignment which considerably enhances the delay performance of the LP_{major} . Exact half-adder, full-adder and 4-2 compressors are used to accumulate and produce sum and carry rows. Sum and carry in LP_{major} are added using LOA of [2].

TABLE I. BOOTH RECODING AND PARTIAL PRODUCT GENERATION

b_{2i+1}	b_{2i}	b_{2i-1}	neg_i	two_i	$zero_i$	Partial product P_{ij} for $a_j a_{j-1}$			
						00	01	10	11
0	0	0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	0	1	1
0	1	0	0	0	0	0	0	1	1
0	1	1	0	1	0	0	1	0	1
1	0	0	1	1	0	1	0	1	0
1	0	1	1	0	0	1	1	0	0
1	1	0	1	0	0	1	1	0	0
1	1	1	0	0	1	0	0	0	0



(a)



(b)

Fig. 1. Radix-4 Booth Partial product generation – (a) K-map and (b) corresponding logic circuit.

approximate circuits to perform such operations would drastically reduce resources consumption, while taking advantage of the tolerance to imprecise data.

To analyze the impact the approximation in a real life application, matrix multiplication blocks using approximate multipliers are designed and used in forward DCT transform of the JPEG encoding application. Standard Lena image is taken as a benchmark. Peak Signal to Noise ratio (PSNR) is used to evaluate the quality of the images. PSNR is more consistent with human visual system and based on mean square error (MSE) of the original image and compressed image. PSNR can be given as

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE} \right) \quad (4)$$

where 255 is the maximum value, a pixel can hold in the grayscale image.

Figure 5 shows the original image and images after JPEG compression using exact and approximate multipliers. The PSNR values after using exact and approximate multipliers-the proposed, ACBM and PPBM in JPEG application are 31.81, 31.77, 31.69 and 18.51 dB respectively. It is found that the proposed multiplier and ACBM shows PSNR almost similar to the exact multiplier. It can be suggested that approximate multiplier can be used in applications such as image processing to significantly reduce the power with minimal loss in quality.



Fig. 5. JPEG Image compression. (a) Original input image. JPEG images using (b) exact multiplier with a PSNR of 31.81 dB (c) proposed multiplier with a PSNR of 31.77 (d) ACBM with a PSNR of 31.69 dB (e) PPBM with a PSNR of 18.51.

V. CONCLUSIONS AND FUTURE WORK

With significant area and power improvement compared to equivalent exact multiplier, a minimal error approximate booth multiplier was investigated. A novel radix-4 approximation is introduced. Partial product generation and accumulation circuits are approximated. Partial product generation circuit in booth multiplier is an important and power consuming process. The exact partial product generation circuit is modified by introducing an error probability of 0.125 and reducing the logic complexity, resulting in high performance circuit. In partial products accumulation, OR gates are efficiently used to reduce area and power while also maintaining reasonable accuracy.

The proposed model has significant area and power improvement compared with an exact multiplier. The design requires small area and power and has better error metrics compared to the equivalent state of the art works. A real life application is chosen to discover the effect of approximation. In future, the approximate multipliers are to be tested with more real-life application scenarios and error compensation block is to be designed to further improve the accuracy of the proposed multiplier.

ACKNOWLEDGMENT

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Department of Electrical Engineering, University of Saskatchewan.

This work was also supported by the Korean Federation of Science and Technology Societies (KOFST) grant funded by the Korean government (MSIP: Ministry of Science, ICT and Future Planning)

REFERENCES

- [1] J. Han and M. Orshansky, "Approximate computing: An emerging paradigm for energy-efficient design," *2013 18th IEEE European Test Symposium (ETS)*, Avignon, 2013, pp. 1-6.
- [2] H. R. Mahdiani, A. Ahmadi, S. M. Fakhraie and C. Lucas, "Bio-Inspired Imprecise Computational Blocks for Efficient VLSI Implementation of Soft-Computing Applications," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 4, pp. 850-862, April 2010.
- [3] K. Du, P. Varman and K. Mohanram, "High performance reliable variable latency carry select addition," *2012 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Dresden, 2012, pp. 1257-1262.
- [4] Ning Zhu, W. L. Goh and K. S. Yeo, "An enhanced low-power high-speed Adder For Error-Tolerant application," *Proceedings of the 2009 12th International Symposium on IC*, Singapore, 2009, pp. 69-72.
- [5] A. Momeni, J. Han, P. Montuschi and F. Lombardi, "Design and Analysis of Approximate Compressors for Multiplication," in *IEEE Transactions on Computers*, vol. 64, no. 4, pp. 984-994, April 2015.
- [6] G. Zervakis, K. Tsoumanis, S. Xydis, D. Soudris and K. Pekmetzi, "Design-Efficient Approximate Multiplication Circuits Through Partial Product Perforation," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 10, pp. 3105-3117, Oct. 2016.
- [7] I. Qiqieh, R. Shafik, G. Tarawneh, D. Sokolov and A. Yakovlev, "Energy-efficient approximate multiplier design using bit significance-driven logic compression," *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2017, Lausanne, 2017, pp. 7-12.
- [8] R. Venkatesan, A. Agarwal, K. Roy and A. Raghunathan, "MACACO: Modeling and analysis of circuits for approximate computing," *2011 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, San Jose, CA, 2011, pp. 667-673.
- [9] M. J. Schulte and E. E. Swartzlander, "Truncated multiplication with correction constant [for DSP]," *Proceedings of IEEE Workshop on VLSI Signal Processing*, Veldhoven, 1993, pp. 388-396.
- [10] Hong-An Huang, Yen-Chin Liao and Hsie-Chia Chang, "A self-compensation fixed-width booth multiplier and its 128-point FFT applications," *2006 IEEE International Symposium on Circuits and Systems*, Island of Kos, 2006, pp. 4 pp.-3541.
- [11] C. Y. Li, Y. H. Chen, T. Y. Chang and J. N. Chen, "A Probabilistic Estimation Bias Circuit for Fixed-Width Booth Multiplier and Its DCT Applications," in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 58, no. 4, pp. 215-219, April 2011.
- [12] H. Jiang, J. Han, F. Qiao and F. Lombardi, "Approximate Radix-8 Booth Multipliers for Low-Power and High-Performance Operation," in *IEEE Transactions on Computers*, vol. 65, no. 8, pp. 2638-2644, Aug. 1 2016.
- [13] J. Liang, J. Han and F. Lombardi, "New Metrics for the Reliability of Approximate and Probabilistic Adders," in *IEEE Transactions on Computers*, vol. 62, no. 9, pp. 1760-1771, Sept. 2013.
- [14] E. de Angel and E. E. Swartzlander, "Low power parallel multipliers," *VLSI Signal Processing, IX*, San Francisco, CA, 1996, pp. 199-208.
- [15] N. Ahmed, T. Natarajan and K. R. Rao, "Discrete Cosine Transform," in *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90-93, Jan. 1974.