Design of an Energy-Efficient Approximate Compressor for Error-Resilient Multiplications

Xilin Yi, Haoran Pei, Ziji Zhang, Hang Zhou, and Yajuan He
School of Electronic Science and Engineering
University of Electronic Science and Technology of China (UESTC)
Chengdu 610054, China

***Abstract*— Digital Multiplier is a fundamental component in many digital signal processing (DSP) systems, which takes up the most part of the computational resources. As many DSP applications have an inherent tolerance for inexact computations, approximate multiplication is considered as an appropriate substitution to obtain energy-performance-accuracy tradeoffs, especially in those applications that require high energy-efficiency in computing. Meanwhile, reducing the supply voltage is proved to be an efficient way to further lower the total energy consumption. In this paper, a novel approximate 4-2 compressor and its circuit implementation is proposed for error-resilient multiplication with a low supply voltage. Simulation results indicate that the approximate multiplier with our proposed approximate 4-2 compressor consumes the least energy per operation with the same computational accuracy when compared with other multipliers for the operand length of 8 bits. It achieves 26.7% reduction on energy-delay product (EDP) when compared with the exact multiplication.**

Keywords—approximate computing; 4-2 compressor; low power; energy-efficient; digital multiplier.

I. INTRODUCTION

With the booming development of Internet of Things (IoT) and portable devices, a great deal of challenge poses on the design of VLSI circuits to obtain the maximized energy efficiency. Fortunately, many applications do not require a full precision computation, such as digital signal processing (DSP), multimedia, fuzzy logic, and neural networks [1]. Therefore, approximate arithmetic circuits design is a promising approach to reduce circuit delay and power consumption with an acceptable loss of accuracy.

Digital multipliers are always fundamental arithmetic units with important influence on performance of the entire system. A fast multiplier is typically composed of three parts: partial products generation, partial products compression, and a fast carry propagation adder (CPA) for the result. Among these three parts, the compression tree occupies the largest proportion in terms of area, delay and power consumption. Thus, energy-efficient compressors are required in the multiplier design [2].

This work was supported by the National Natural Science Foundation of China under Grant 61874023, 61604036 and 61534002, and the Fundamental Research Funds for the Central Universities under Grant ZYGX2018J030.

Intuitively, the partial products in a compression tree are accumulated with two operand adders, which use full adders, sometimes known as 3-2 compressors. Compared with full adders, 4-2 compressors have a higher compression efficiency so that are widely used in a partial products reduction tree. In approximate multiplier design, many approximate 4-2 compressors have been proposed to achieve high energy efficiency [3]-[6]. An approximation method is proposed in [3], which cuts the carry chains between two 4-2 compressors of the same stage. It reduces the hardware complexity greatly but incurs a high error rate, and thus, a poor computational accuracy. An approximate Wallace-Booth multiplier is proposed in [4] based on the approximate 4-2 compressors shown in [3] to reduce the number of partial products as well as the probability of the error occurred. To further improve precision, a logical approximation approach using Karnaugh map to simplify computational logic is proposed in [5], and it achieves a lower error rate with three types of approximate compressors. However, the energy consumption is significantly increased. To obtain a better trade-off between circuit performance and accuracy, Suganthi *et. al.* alters the partial products to introduce terms with different probabilities using a series of novel approximate half-adders, full-adders and 4-2 compressors [6].

In this paper, an approximate 4-2 compressor is proposed with a complementary error strategy in order to obtain a reduced logic complexity with acceptable computation accuracy. In addition, the circuits are designed at a low supply voltage to achieve even higher energy efficiency.

The rest of paper is organized as follows. Section II introduces the conventional exact 4-2 compressors. The design and analysis of the proposed approximate compressors are shown in Section III. Simulation results for the approximate multipliers with approximate compressors are provided in Section IV. Section V gives the conclusion.

II. CONVENTIONAL EXACT 4-2 COMPRESSOR

A 4-2 compressor is used to count the number of '1' in inputs. The function of a 4-2 compressor is given as follows:

978-1-7281-0397-6/19/$31.00 ©2019 IEEE

$$X1+X2+X3+X4+C_{in}=2(Carry+C_{out})+Sum \quad (1)$$

where *X1*, *X2*, *X3*, *X4* and C_{in} are inputs, *Sum*, *Carry* and C_{out} are three outputs.

As indicated in (1), the output *Sum* has the same weight as input bits, while *Carry* and C_{out} are weighted one binary bit higher. Fig. 1 illustrates the implementation of conventional 4-2 compressor, which is composed of two full-adders. Thus, the expression of output *Sum*, *Carry*, and C_{out} are obtained as:

$$\begin{aligned} Sum &= X1\oplus X2\oplus X3\oplus X4\oplus C_{in.} \\ Carry &= (X1\oplus X2\oplus X3\oplus X4)\cdot C_{in} \\ &\quad +\overline{(X1\oplus X2\oplus X3\oplus X4)}\cdot X4. \\ C_{out} &= (X1\oplus X2)\cdot X3+\overline{(X1\oplus X2)}\cdot X1. \end{aligned} \quad (2)$$

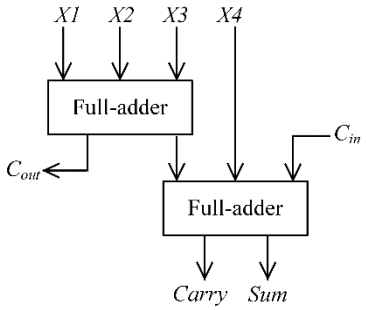

Fig. 1. Conventional exact 4-2 compressor based on full-adders.

III. Proposed Approximate 4-2 Compressor

A. Proposed 4-2 Compressor

As indicated in [3], the approach that omits C_{in} and C_{out} helps to further reduce logic complexity. As C_{in} of the 4-2 compressor with lowest weight is always '0', C_{out} is required only when all the four inputs are '1' with the occurrence probability of 1/16. In this case, if the exact outputs (C_{out}='1', *Carry*='1', *Sum*='0') is replaced by approximate output (C_{out}='0', *Carry*='1', *Sum*='1'), C_{out} must be '0', thus can be pruned. Since C_{in} of each 4-2 compressor is connected to the C_{out} of lower weighted compressor, it can be also ignored with acceptable loss of accuracy. In this case, (1) can be rewritten as:

$$X1+X2+X3+X4=2Carry+Sum. \quad (3)$$

As observed, the above equation can always be satisfied except the case that all the inputs (*X*1~*X*4) are '1', which happens only once out of 16 cases. Table I summarizes all the combinations of the 4 inputs (*X1*-*X4*), the calculated values (*Value*), carry bit (*Carry*), and sum bit (*Sum*). Thus, the corresponding logical expression can be derived from Table I as follows:

$$\begin{aligned} Sum &= X1\oplus X2\oplus X3\oplus X4+X1\cdot X2\cdot X3\cdot X4. \\ Carry &= X1\cdot X2+X3\cdot X4+(X2+X1)\cdot X4+(X1+X2)\cdot X3. \end{aligned} \quad (4)$$

Suppose all inputs are independent and uniformly distributed, thus, each bit of the multiplier and the multiplicand has a half chance of being either '1' or '0'. As each partial product bit is generated with an AND function between one multiplier bit and one multiplicand bit. Therefore, the probability of one partial product bit being '1' is a quarter. With this calculation, the probabilities that the corresponding input pattern occurs (*Prob.*) are listed in Table I as well. As noticed, the probability that all inputs are '1' is the least, which has the probability of 1/256 only.

In order to simplify the logic further based on (4), *Carry* is simplified as *Carry*' in TABLE I and its logical expression can be derived accordingly.

$$Carry'=X1\cdot X2+X3\cdot X4. \quad (5)$$

TABLE I. Truth Table of Approximate 4-2 Compressor

X1	*X2*	*X3*	*X4*	*Value*	*Carry*	*Sum*	*Prob.*	*Carry'*	*Sum'*	*Diff.*
0	0	0	0	0	0	0	81/256	0	0	0
0	0	0	1	1	0	1	27/256	0	1	0
0	0	1	0	1	0	1	27/256	0	1	0
0	0	1	1	2	1	0	9/256	1	0	0
0	1	0	0	1	0	1	27/256	0	1	0
0	1	0	1	2	1	0	9/256	0	1	-1
0	1	1	0	2	1	0	9/256	0	1	-1
0	1	1	1	3	1	1	3/256	1	1	0
1	0	0	0	1	0	1	27/256	0	1	0
1	0	0	1	2	1	0	9/256	0	1	-1
1	0	1	0	2	1	0	9/256	0	1	-1
1	0	1	1	3	1	1	3/256	1	1	0
1	1	0	0	2	1	0	9/256	1	0	0
1	1	0	1	3	1	1	3/256	1	1	0
1	1	1	0	3	1	1	3/256	1	1	0
1	1	1	1	4	1	1	1/256	1	1	-1

It should be noted that the approximate situation of *Carry* occurs at *Carry* and *Sum* are not '1' at the same time. Therefore, we can adjust the result of *Sum* as *Sum'* by changing a '0' to '1' to minimize the difference between the exact decimal value of the addition of the inputs and that produced by the approximate compressor (*Diff* in TABLE I). Thus, we get *Sum*' as shown in TABLE I. Its logical expression can be also derived from this truth table as follows:

$$Sum'=(X1\oplus X2)+(X3\oplus X4)+X1\cdot X2\cdot X3\cdot X4. \quad (6)$$

For the purpose of logic reduction with intermediate terms sharing, the logical expression of *Carry'* in (5) and *Sum*' in (6) can be further transformed into (7) with De Morgan's laws.

$$\begin{aligned} Carry' &= \overline{\overline{(X1\cdot X2)}\cdot\overline{(X3\cdot X4)}}. \\ Sum' &= \overline{(X1\odot X2)\cdot(X3\odot X4)\cdot\overline{(X1\cdot X2)}\cdot(X3\cdot X4)}. \end{aligned} \quad (7)$$

B. Implementation of the proposed 4-2 Compressor

This section deals with the specific implementation of the circuit, which has a more direct impact on the hardware performance. As we all know, supply voltage reduction helps to reduce the power consumption greatly, while deteriorates the circuit performance on the other hand. As indicated in [7-8], it is promising to achieve the highest energy efficiency in the near-threshold field. Therefore, to obtain a higher energy efficiency, design techniques targeted at low supply voltage are considered. A 0.18-μm standard CMOS technology is employed in this work. The supply voltage is targeted at 1.0V.

Fig. 2 illustrates a gate-level implementation of our proposed approximate 4-2 compressor according to (7).

Two fundamental circuits used in Fig. 2 are XNOR and NAND gates. Thus, we illustrate the design techniques on these two gates at a low supply voltage. A transistor-level circuit implementation is shown in Fig. 3. Because reverse short channel effect (RSCE) becomes relatively strong in the near-threshold region, we explore optimal transistor length to achieve minimal delay. As shown in Fig. 3 (a) and (b), the channel length of PMOS is adjusted from standard 180nm to 190nm. In addition, since the imbalance between the NMOS and PMOS strength will become more serious as the supply voltage decreases. We not only increase the channel width of PMOS, but also avoid PMOS connected in series such as NOR gates. Using transmission gate logic to build XNOR gate can also avoid the problem of strength imbalance, as shown in Fig. 3 (b). Moreover, the strength is increased more effectively by connecting minimum sized transistors in parallel. This is because an increase in width leads to a threshold increase due to the RSCE. As shown in Fig. 3 (a), PMOS and NMOS are equivalent to a larger width by parallel connection of small size transistors.

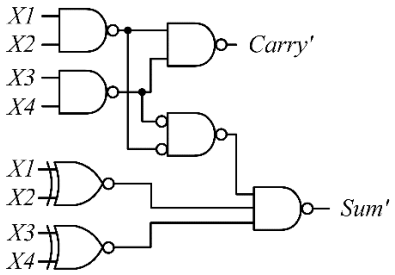

Fig. 2. Hardware Implementation of proposed approximate compressor.

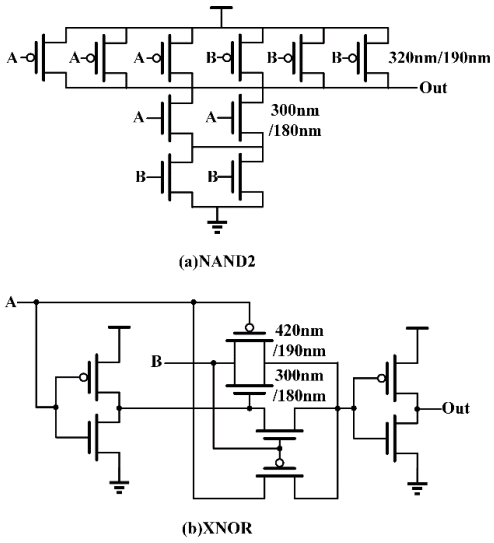

Fig. 3. Transistor level circuit optimization: (a)NAND2 gate (b)XNOR gate.

C. Simulation and Analysis

To evaluate the performance of the proposed circuits, HSPCIE simulations are performed with a supply voltage of 1.0V in a 0.18-μm standard CMOS technology and compares it with several competitive 4-2 compressors, including exact 4-2 compressor [2], design2 of simplified approximate 4-2 compressors (ACD2) [3] and design3 of the approximate compressors with C_{in} and C_{out} ignored (ACCI3) [5]. All these 4-2 compressors are implemented according to the circuit diagram mentioned in the respective references. The logic gates used in these 4-2 compressors are redesigned according to the above methods. The same logic gates in different 4-2 compressors have the same circuit structure and transistor size. To provide a realistic simulation environment reflecting the compressor operation in actual applications, each input is driven by buffered signals and each output is loaded with buffers. For each circuit, a large amount randomly data feed into the circuits, and all circuits operate at a clock rate of 10MHz. The power consumption is obtained through the product of the average supply current and the supply voltage of the circuit under test. The results are summarized in Table II. As indicated, the error rate is also included, which is defined as follows according to [5]:

$$ER = \frac{\sum_{t=1}^{T} D_t}{T}, where \;\; D_t = \begin{cases} 0, & if \;\; a_t = b_t \\ 1, & if \;\; a_t \neq b_t \end{cases} \qquad (8)$$

where T is the number of input values and a_t , b_t are the exact and approximate outputs respectively for a given input t.

TABLE II. COMPARISONS OF APPROXIMATE 4-2 COMPRESSORS

4-2 Compressor	Delay (*ns*)	Power (*uW*)	Energy (*fJ*)	EDP (*fJ·ns*)	Error rate
Exact[2]	0.988	0.317	0.313	0.309	0
ACD2[3]	0.605	0.276	0.167	0.101	100/256
ACCI3[5]	0.769	0.331	0.255	0.196	16/256
This work	**0.412**	**0.232**	**0.096**	**0.040**	**37/256**

As shown in Table II, our proposed approximate 4-2 compressor achieves the least delay, power dissipation, and energy per operation with a moderate error rate. It exhibits the highest energy efficiency due to its minimal EDP. On the other hand, exact 4-2 compressor is the most accurate with zero error rate but has the worst delay and energy efficiency. ACD2 has relatively lower delay and power, but its error rate is as high as 100/256. This implies a poor computation accuracy, which limits its application in the design of multiplication. Among all approximate compressors, the error rate of ACCI3 is the least, that indicates a best calculation accuracy. However, ACCI3 adopts transistors with large gate widths to improve its speed at low supply voltage so that the complexity of the compressor is increased as well as power consumption.

IV. APPROXIMATE MULTIPLIER DESIGN WITH PROPOSED APPROXIMATE 4-2 COMPRESSOR

A. Approximate Multiplier Architecture

When the approximate compressor is used for partial products compression tree, the first thing to pay attention to is that the partial product closer to most significant bits have a great impact on the results and the partial product closer to least significant bits contribute little in calculation results. To make sure that accuracy loss will not be excessive, the partial products reduction tree is divided into two parts: accurate part

and approximate part. For an approximate $N\times N$ multiplier, the accurate part is that the most significant N bits still use the exact compression structure. The approximate part, which is in the least significant N bits, use approximate 4-2 compressors.

For a more intuitive explanation, we take an 8×8 unsigned Dadda-tree multiplier design as an example. In the first stage, two- input AND gates are used to generate all partial products. In the second stage, the approximate compressors proposed in the previous section are utilized in the partial products reduction tree. The last stage is an exact ripple carry adder to compute the final binary result. In Fig. 4, the first stage of partial products generation is not shown; only the second and third stages are illustrated and a dot represents a partial product.

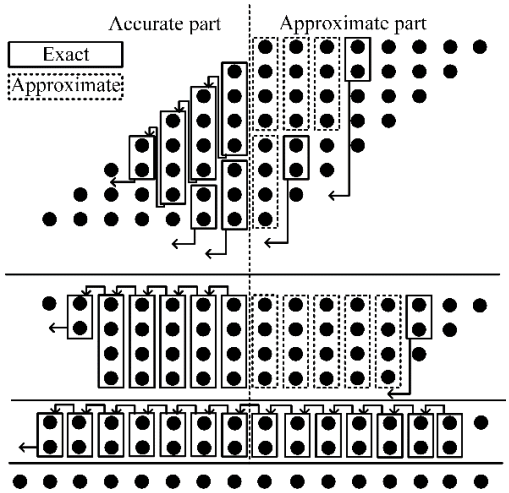

Fig. 4. Reduction circuitry of an 8×8 Dadda multiplier with approximate 4-2 compressor.

B. Circuit Performance Simulation

In this section, HSPICE simulations are performed in order to evaluate the performance of approximate multiplier with our proposed approximate 4-2 compressor. All approximate multipliers use the same architecture, as shown in Fig. 4, except that the 4-2 compressors in approximate part use ACD2 and ACCI3, respectively. The results are summarized in TABLE III, which includes the worst case delay, power consumption, energy per operation, and EDP.

TABLE III. CIRCUIT PERFORMANCE OF DIFFERENT MULTIPLIERS

Multiplier	Delay(*ns*)	Power(*uW*)	Energy(*fJ*)	EDP(*fJ·ns*)
Exact[2]	8.264	23.08	190.73	1576
ACD2[3]	7.533	21.71	163.54	1232
ACCI3[5]	7.533	23.52	177.18	1335
This work	**7.533**	**20.36**	**153.37**	**1155**

As shown in Table III, all approximate multipliers are faster than exact multiplier because the approximation on 4-2 compressors leads to faster propagation in most significant bit of approximate part at CPA calculation stage. Thus, the critical paths of all approximate multipliers are determined by the same accurate part so that all of them show the same delay. In terms of power consumption, the approximate multiplier with proposed approximate 4-2 compressor has the best performance, which is attributed to simplified complexity and circuit design optimization for low supply voltage. Therefore, the multiplier with proposed compressor shows lowest energy per operation and EDP as well as compressor itself. Compared with the accurate multiplier using exact compressor [2], the worst delay, average power consumption, energy per operation and EDP of the proposed compressor are reduced by 8.8%, 11.8% ,19.6% and 26.7%, respectively.

C. Error Performance Analysis

Except error rate (*ER*), three error metrics are introduced to evaluate the accuracy of various approximate multipliers [9]. They are normalized mean error (E_{mean}), normalized maximum absolute error (E_{max}) and normalized mean square error (E_{mse}), respectively. The E_{mean}, E_{max}, and E_{mse} are define as:

$$\begin{aligned} E_{mean} &= E\{P_A - P\} / 2^{2N}. \\ E_{max} &= max\{|P_A - P|\} / 2^{2N}. \\ E_{mse} &= E\{(P_A - P)^2\} / 2^{2N}. \end{aligned} \tag{9}$$

where P_A indicates accurate product while P indicates approximate products.

TABLE IV. ERROR PERFORMANCE OF DIFFERENT MULTIPLIERS

Multiplier	*ER*	E_{mean}	E_{mse}	E_{max}
Exact[2]	0%	0	0	0
ACD2[3]	92.85%	2.07×10^{-4}	2.74×10^{-1}	8.91×10^{-3}
ACCI3[5]	44.24%	1.70×10^{-4}	1.09×10^{-1}	8.17×10^{-3}
This work	**66.45%**	**1.34×10^{-3}**	**2.53×10^{-1}**	**7.93×10^{-3}**

As illustrated in Table IV, the approximate multiplier with proposed approximate 4-2 compressor has a moderate *ER*, which is consistent with analysis of approximate 4-2 compressors in Section III. The approximate multiplier with proposed approximate 4-2 compressors has worse E_{mean} because the *Diff*. in Table I are always negative values and this error will accumulate in multiplier. Our proposed approximate multiplier shows advantages compared with ACD2 on E_{mse}, which represents that the error distribution is more centralized and is beneficial in multimedia and DSP applications. Meanwhile, the multiplier with proposed approximate 4-2 compressors shows best E_{max}, which means high reliability and is the key parameter in safety critical applications such as autopilot for unmanned aircrafts and control of nuclear reactors [10].

V. CONCLUSION

In this paper, a new approximate 4-2 compressor is proposed, which achieves smallest delay and power consumption under low supply voltage and shows a high energy efficiency. Simulation results indicate that the multiplier employed with our proposed approximate 4-2 compressor has lowest power consumption with a moderate error performance. Compared with the exact compressor and exact multiplier, the EDP of the proposed design are reduced 87.1% and 26.7% respectively.

REFERENCES

[1] Q. Xu, T. Mytkowicz, and N. S. Kim, "Approximate computing: A survey," *IEEE Des. Test*, vol. 33, no. 1, pp. 8–22, Feb. 2016.

[2] C.-H. Chang, J. Gu, and M. Zhang, "Ultra low-voltage low- power CMOS 4-2 and 5-2 compressors for fast arithmetic circuits," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 10, pp. 1985–1997, Oct. 2004.

[3] 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.

[4] L. Qian, C. Wang, W. Liu, F. Lombardi and J. Han, "Design and evaluation of an approximate Wallace-Booth multiplier," 2016 *IEEE Int. Symp. Circuits Syst. (ISCAS)*, Montreal, QC, 2016, pp. 1974-1977.

[5] Z. Yang, J. Han and F. Lombardi. "Approximate compressor for error resilient multiplier design," *IEEE Int. Symp. On Defect Fault Tolerance in VLSI and Nano. Syst.*, MA, USA, Oct. 2015.

[6] S. Venkatachalam and S.-B. Ko, "Design of power and area efficient approximate multipliers," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 5, pp. 1782–1786, May 2017.

[7] J. Jun, J. Song and C. Kim, "A Near-Threshold Voltage Oriented Digital Cell Library for High-Energy Efficiency and Optimized Performance in 65nm CMOS Process," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 5, pp. 1567-1580, May 2018.

[8] M. Aliotto, "Ultra-Low Power VLSI Circuit Design Demystified and Explained: A Tutorial," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 1, pp. 3–29, January 2012.

[9] Z. Zhang, Y. He, "A Low Error Energy-Efficient Fixed-Width Booth Multiplier with Sign-Digit-Based Conditional Probability Estimation," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 65, no. 2, pp. 236–240, May. 2017.

[10] 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, Sept. 2011.