

Hybrid radix-4/-8 truncated multiplier for mobile GPU applications

Seongrim Choi, Gyeonghoon Kim, Hoi-Jun Yoo and Byeong-Gyu Nam

A hybrid radix-4/-8 truncated multiplier is proposed for mobile graphics processing unit (GPU) applications by combining the strong points of high-speed radix-4 and low-power radix-8 encoding schemes. A hybrid Booth encoder (HBE) for radix-4/-8 encoding is proposed by sharing the common logics between the two encodings. In addition, a hybrid radix-4/-8 truncation (HRT) scheme is proposed for more power reduction at a reasonable peak signal-to-noise ratio loss for mobile multimedia applications. Experimental results demonstrate 29.7 and 31.1% power savings from the HBE and the HRT schemes, respectively, resulting in a total of 60.7% power reduction from the previous work.

Introduction: Computer graphics has been moving into the mainstream of mobile multimedia computing these days. The graphics processing units (GPUs) incorporate shaders that contain floating-point multipliers and special function units (SFUs) such as powering, division and square root to compute a wide variety of graphics algorithms [1]. In these units, integer multipliers play a key role in implementing the mantissa part of the floating-point multipliers and the polynomial interpolation in the SFUs. Therefore, a low-power and high-speed integer multiplier becomes one of the most critical blocks in mobile GPU designs.

It is well known that the radix-4 encoding is good for high speed as its partial products can be made using just simple shift operations while radix-8 is good for low power as it produces a smaller number of partial products compared with radix-4 [2]. In this Letter, we propose a novel hybrid radix-4/-8 modified Booth encoding for low-power and high-speed multiplications by combining the strong points of high-speed radix-4 and low-power radix-8 encodings. There was a study on the hybrid radix-4/-8 encoding, but it was optimised for neither power nor speed as it aimed at a compromise between the two [2]. Moreover, it had a considerable power overhead in the Booth encoder due to the separate radix-4 and radix-8 encoders operating concurrently. Our proposed hybrid Booth encoder (HBE) mitigates this problem by sharing the major parts of the radix-4 and radix-8 encoders [3]. In addition, we propose a novel truncation scheme for more power reduction based on the usage pattern of partial products from the hybrid encoding. As a result, the proposed hybrid radix multiplier based on these two schemes demonstrates 60.7% power saving from the previous art [2].

Hybrid radix-4/-8 Booth encoder: Radix-4 and radix-8 are the most widely used encoding schemes for modified Booth multipliers. Radix-4 encodes every 2 bits of an input, whereas radix-8 takes every 3 bits for the encoding. Therefore, radix-8 produces fewer partial products and consumes less power than radix-4. However, radix-8 multiplication is slower than that of radix-4 due to the generation of $\pm 3B$ terms that do not have any corresponding shift operations and require additional carry propagation stages.

Table 1: Radix-8 partial product table ($A \times B$)

Inputs				Partial products				Booth selects			
x_{i+2}	x_{i+1}	x_i	x_{i-1}	PP_i	S_i	P_i	$2P_i$	$4P_i$			
0	0	0	0	0	0	0	0	0			
0	0	0	1	B	0	1	0	0			
0	0	1	0	B	0	1	0	0			
0	0	1	1	$2B$	0	0	1	0			
0	1	0	0	$2B$	0	0	1	0			
0	1	0	1	$3B$	—	—	—	—			
0	1	1	0	$3B$	—	—	—	—			
0	1	1	1	$4B$	0	0	0	1			
1	0	0	0	$-4B$	1	0	0	1			
1	0	0	1	$-3B$	—	—	—	—			
1	0	1	0	$-3B$	—	—	—	—			
1	0	1	1	$-2B$	1	0	1	0			
1	1	0	0	$-2B$	1	0	1	0			
1	1	0	1	$-B$	1	1	0	0			
1	1	1	0	$-B$	1	1	0	0			
1	1	1	1	0	1	0	0	0			

Our proposed multiplier operates normally in the radix-8 mode and it switches to the radix-4 if the radix-4/-8 HBE detects the production of $\pm 3B$ terms from its input patterns [3]. The $\pm 3B$ terms can be detected when $(x_{i+2} \oplus x_{i+1})$ and $(x_i \oplus x_{i-1})$ become true as shown in grey in Table 1. Therefore, just two XOR and one AND gates are used for the radix selection, as described in Fig. 1. In this way, the multiplier operates in the low-power radix-8 mode for 56% cases of inputs, and switches to the high-speed radix-4 mode for 44% of inputs [3].

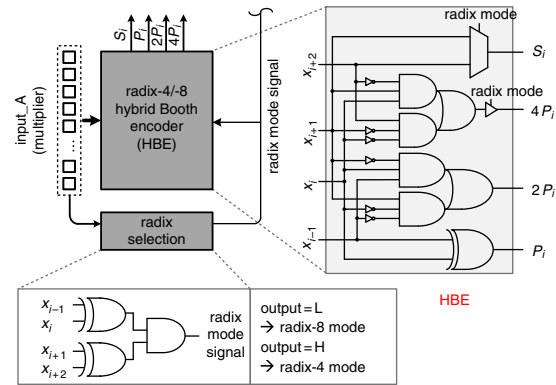


Fig. 1 Radix-4/-8 HBE

On the basis of this radix selection scheme, the radix-4/-8 HBE is proposed exploiting the similar structures of radix-4 and radix-8 encoders, and thus the major parts of the two encoders are shared for both radices. Fig. 1 shows the internal structure of the proposed HBE. It produces four encoding signals (i.e. S_i , P_i , $2P_i$ and $4P_i$) according to (1)–(4), for each 3 bit of the input. The P_i and $2P_i$ signals remain the same for both radix-4 and radix-8, and the S_i and $4P_i$ signals, shown in Fig. 1, just require an additional multiplexer and tri-state buffer according to the radix mode, as explained in (1) and (4). In this way, our HBE saves 29.7% of the power from the previous hybrid radix encoder [2]

$$S_i = \begin{cases} x_{i+2} & (\text{Radix mode} = 0) \\ x_{i+1} & (\text{Radix mode} = 1) \end{cases} \quad (1)$$

$$P_i = x_i \oplus x_{i+1} \quad (2)$$

$$2P_i = \overline{x_{i+1}}x_i x_{i-1} + x_{i+1}\overline{x_i}x_{i-1} \quad (3)$$

$$4P_i = \begin{cases} \overline{x_{i+2}}x_{i+1}x_i + x_{i+2}\overline{x_{i+1}}\overline{x_i} & (\text{Radix mode} = 0) \\ Z(\text{high-impedance}) & (\text{Radix mode} = 1) \end{cases} \quad (4)$$

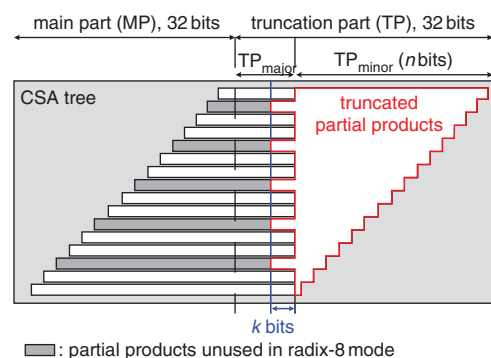


Fig. 2 Hybrid radix truncation

Hybrid radix-4/-8 truncation: In mobile multimedia applications, a reasonable amount of computing errors can be exploited for power efficiency [1, 4]. In this Letter, we investigate a novel truncation scheme for the hybrid radix-4/-8 encoding for more power reduction of the HBE multiplier at a tolerable degradation of peak signal-to-noise ratio (PSNR). As illustrated in Fig. 2, we divide the carry save adder tree of the multiplier into two parts, the truncation part (TP) and the main part and subdivide the TP into n bits of TP_{minor} and $(32 - n)$ bits of

TP_{major} . In conventional truncations, the TP_{minor} is normally truncated off but the TP_{major} is intact for minimal impacts on multiplication results. In this Letter, however, we make a truncation of k bits in the TP_{major} of the partial products unused in the radix-8 mode, as illustrated in Fig. 2. This k -bit truncation does not incur a significant increase in multiplication errors since the HBE scheme remains in the radix-8 mode with a high probability of 56% [3]. We chose the two truncation numbers, n and k , carefully to maximise the power efficiency of the multiplier. The n and k values for the best power efficiency are chosen to be 29 and 1, respectively, as depicted in Fig. 3. Thanks to this hybrid radix truncation (HRT) scheme combined with the proposed HBE, we achieved a 60.7% power reduction compared with the previous hybrid radix scheme [2].

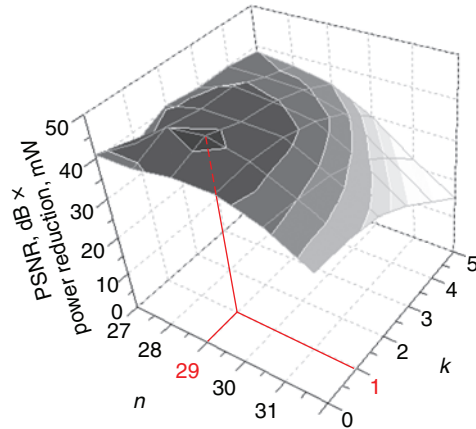


Fig. 3 Comparisons of PSNR and power reduction for HRT scheme

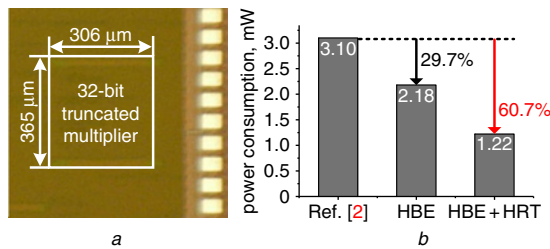


Fig. 4 Implementation results for proposed hybrid radix multiplier

a Die photograph

b Power reduction graph

Note: power consumptions of [2] are normalised to 0.18 μm technology node

Table 2: 32b \times 32b multiplier performance comparisons

	Hybrid radix [2]	HBE	This work
Technology	1.2 μm	0.18 μm	0.18 μm
$M \times N$	32 b \times 32 b	32 b \times 32 b	32 b \times 32 b
Nominal V_{dd} , V	5	1.8	1.8
Input frequency, MHz	10	10	10
Power, mW	3.1	2.18	1.22
Area, μm^2	N/A	211 600	111 690
Propagation delay, ns	3.51	4.12	3.91
PSNR, dB	63.69	63.69	47.44
PDP, nJ	10.9	8.98	4.77

Note: power and delay of [2] are normalised to 0.18 μm technology node.

Experimental results: The proposed 32 b \times 32 b hybrid radix-4/-8 truncated multiplier was fabricated using the 0.18 μm CMOS technology. Fig. 4a shows a die photograph of the chip with an active area of

111 690 μm^2 . Its propagation delay was measured to be 3.91 ns and the average power was 1.22 mW when measured at the 10 MHz input frequency. This 10 MHz is used for the power measurement in the previous hybrid radix scheme [2], and we used the same value for comparison purpose. As shown in Fig. 4b, the HBE scheme saves 29.7% of the power and the HRT scheme reduces an additional 31.1% of power, resulting in a total of 60.7% power reduction from the previous work [2]. The resulting PSNR of 47.4 dB produced by the HRT scheme still remains above 20 dB, the minimum required PSNR value for mobile multimedia applications [5]. Table 2 presents performance comparisons with previous works. The power and delay of the previous hybrid radix multiplier [2] are normalised to the 0.18 μm technology node for comparisons. Table 2 shows that our approach attains a 56% reduction in the power-and-delay product (PDP) from the previous art [2].

Conclusion: A novel hybrid radix-4/-8 truncated multiplier is proposed for mobile GPU applications. The proposed multiplier demonstrates 60.7% power reduction through the proposed HBE and the HRT schemes. In terms of PDP, it achieves a 56% reduction compared with the previous work.

Acknowledgments: This work was partly supported by the IT R&D programme of MKE/KEIT [10041664, The Development of Fusion Processor based on Multi-Shader GPU].

© The Institution of Engineering and Technology 2014

3 May 2014

doi: 10.1049/el.2014.1427

One or more of the Figures in this Letter are available in colour online.

Seongrim Choi and Byeong-Gyu Nam (*Department of Computer Science and Engineering, Chungnam National University, 99, Daehak-ro, Yuseong-gu, Daejeon 305-764, Republic of Korea*)

E-mail: bgnam@cnu.ac.kr

Gyeonghoon Kim and Hoi-Jun Yoo (*Department of Electrical Engineering, KAIST, 335, Gwahangno, Yuseong-gu, Daejeon 305-701, Republic of Korea*)

References

- Nam, B.-G., *et al.*: 'An embedded stream processor core based on logarithmic arithmetic for a low-power 3-D graphics SoC', *IEEE J. Solid-State Circuits*, 2009, **44**, (5), pp. 1554–1570
- Cherkauer, B.S., *et al.*: 'A hybrid radix-4/radix-8 low power signed multiplier architecture', *IEEE Trans. Circuits Syst. II*, 1997, **44**, (8), pp. 656–659
- Kim, G., *et al.*: 'A low-energy hybrid radix-4/-8 multiplier for portable multimedia applications'. *IEEE Int. Symp. Circuits and Systems*, 2011, pp. 1171–1174
- Wang, J.-P., *et al.*: 'High-accuracy fixed width modified booth multipliers for lossy applications', *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, 2011, **19**, (1), pp. 52–60
- Thomos, N., *et al.*: 'Optimized transmission of JPEG2000 streams over wireless channels', *IEEE Trans. Image Process.*, 2006, **15**, (1), pp. 54–67

Copyright of Electronics Letters is the property of Institution of Engineering & Technology and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.