

# Optimization design of 24 bit parallel MAC unit with saturation

Zhang Meng<sup>1</sup> Jia Junbo<sup>1,2</sup>

(<sup>1</sup> National ASIC System Engineering Research Center, Southeast University, Nanjing 210096, China)

(<sup>2</sup> Hisilicon Technologies Co. Ltd, Shenzhen 518129, China)

**Abstract:** An efficient design method for a 24 × 24 bit + 48 bit parallel saturating multiply-accumulate (MAC) unit is described. The augend in the MAC is merged as a partial product into Wallace tree array. The optimized saturation detection logic is proposed. The 679.2 μm × 132.5 μm area size has been achieved in 0.18 μm 1.8 V 1P6M CMOS technology by the full-custom circuit layout design. The simulation results show that the design way has significantly less area (about 23.52% reduction) and less delay than those of the common saturating MAC based on standard cell library.

**Key words:** multiply-accumulate; Booth encoding; Wallace tree; saturation detection; layout design

There are lots of multiply-accumulate (MAC) arithmetics in the data communication and digital signal process areas. The speed of the MAC unit is very important in improving performance of the system in these areas.

The MAC operation is common during the two clock cycles<sup>[1]</sup>. Because multiplication is critical for both speed and area, the research on constructing the high-speed multiplier is carried out, such as the rectangle-styled Wallace tree to eliminate the dead area, 4-2 compressor, redundant binary architecture, and so on<sup>[1-3]</sup>. Saturating arithmetic means that the number is clamped to the maximum possible value when an overflow occurs. Thus, it requires extra hardware to be added to binary adder and more arithmetic cycles. Previous studies on overflow detection and saturation have focused on fractional operands<sup>[4]</sup>, or operations other than multiplication<sup>[5]</sup>. In fact, with fractional two's complement multiplication, overflow is much easier to detect since it only occurs when the multiplication  $-1 \times -1$  is performed. Generally, parallel saturating MAC is suitable for high-performance digital signal processing systems since they have less delay than serial saturating MAC despite more area.

This paper describes an efficient design method in a 24 × 24 bit + 48 bit parallel saturating MAC unit. In the multiplier design, we applied the modified Booth encoding for high-speed operations to reduce the num-

ber of the partial products (PPs), and applied the Wallace tree composed of compressors to add the partial product. The augend is operated as a PP in the Wallace tree to deduce the critical path delay to the MAC operation. The logic of the saturation detection and the final value modification was added in the MAC design in saturating MAC operations. By means of ASIC custom circuit, the performance of the saturating MAC unit is improved in the area and in the speed of the critical path. The technique for two's complement MAC with overflow detection is to be used in our design of one kind of system on a chip (SoC) as intellectual property (IP) core.

## 1 High-Speed Parallel MAC's Structure

### 1.1 Modified Booth encoding

Because the second-order Booth encoding technique almost reduces the number of PPs by half and is good for signed (two's complement) operands, nearly 90% of the multipliers are designed using the Booth algorithm<sup>[1,3]</sup>. However, it requires the sign extension and one extra bit to handle unsigned operands.

An  $n$ -bit two's complement multiplier  $D$  can be finished by the modified Booth algorithm.

$$D = (-1)d_{n-1}2^{n-1} + \sum_{i=0}^{n-2} d_i 2^i = \sum_{i=0}^{n/2-1} (d_{2i-1} + d_{2i} + 2d_{2i+1})2^{2i} \quad (1)$$

where  $d_{-1} = 0$ .

### 1.2 Sign extension

When adding PPs in parallel using carry-save adders (CSA), each PP must be sign extended to the  $m + n$  binary position. Such a sign extension becomes very costly. In order to prevent that, we assume that all of

Received 2005-12-31.

**Foundation items:** The National Natural Science Foundation of China (No. 90407009), the National High Technology Research and Development Program of China (863 Program) (No. 2003AA1Z1340).

**Biography:** Zhang Meng (1964—), male, associate professor, zmeng@seu.edu.cn.

the PPs are negative. Therefore, the sum of all the sign extensions can be pre-calculated<sup>[1]</sup>, and the number of signs is

$$N = \sum_j^{M/2-1} ((-1)2^n)4^j = -2^n \left( \frac{2^M - 1}{3} \right) \quad (2)$$

### 1.3 Partial production tree and final adder

Because the Wallace-tree method adds the PPs in parallel at the same time, it is one of the best solutions to construct the high-speed multiplier (MPY). However, its physical design is difficult because of the complex interconnections and the problem of dead areas when the MPY is arranged in a rectangle. Concerning the complex interconnections, many methods have been studied<sup>[2]</sup>, such as the 4-2 compressor in replacing full-adders, redundant binary architecture, and so on. Meanwhile, we can apply a rectangular-styled construction method of a Wallace tree which reduces the dead area size with a simple layout.

Generally, the fast adder (carry-select or look-ahead) is used to produce the final answer. The advantage of the Wallace tree architecture is the optimized speed<sup>[6]</sup>. It can change the Wallace tree multiplier into a quite simple MPY or accumulator. One only needs to include all the PP bits in the same column in the inputs to the Wallace tree adders.

### 1.4 Saturation and overflow detection logic operation

In the general-purpose and application-specific computer or DSP systems, there are millions of saturating MAC operations<sup>[3,7]</sup>. For example, integer multiplication is often supported, where two  $n$ -bit integers (unsigned or two's complement) are multiplied to produce a  $2n$ -bit product. To prevent growth in word length, the above systems are desirable in order to return the  $n$  least significant bits of the product and a flag that indicates whether overflow has occurred or not.

## 2 VLSI Implementation and Simulation Results

### 2.1 Optimized architecture of saturating MAC

In the parallel saturating MAC unit, there are three units including the  $24 \times 24$  bit Booth multiplier, a 48 bit final adder and the unit dealing with saturation. Generally, the modified value of saturation is outputted after the multiplication and addition for saturating MAC. In order to reduce the critical path delay further, the saturation operation and MAC operations are carried out simultaneously as shown in Fig. 1.

The equation below is used to determine whether the overflow of MAC occurs or not:

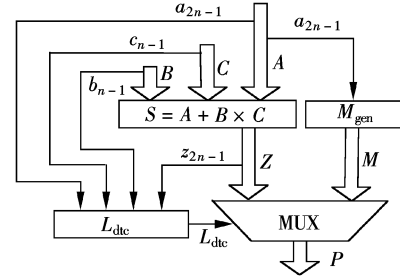


Fig. 1 Optimized saturating MAC unit

$$L_{dnc} = (b_{n-1} \oplus c_{n-1}) a_{2n-1} z_{2n-1} + b_{n-1} \oplus c_{n-1} a_{2n-1} z_{2n-1} \quad (3)$$

where  $(b_{n-1} \oplus c_{n-1})$  is the signed bit of  $B \times C$ . When the overflow occurs, 48 bit 2-1 MUX outputs the modified value of saturation.

### 2.2 Modified high-speed MAC

The above MAC unit is very important for the saturating MAC in speed and area. The MAC operations  $\langle S \rangle = \langle A + \langle B \times C \rangle \rangle$  include three operands: the augend  $A$ , multiplicand  $B$  and multiplier  $C$ . The conventional method of MAC is that multiplier and adder are serially arranged simply. This method causes more area and delay in its critical path since there are two serial adding stages. In our design, the augend  $A$  was treated as another PP in the Wallace tree adder as shown in Fig. 2.

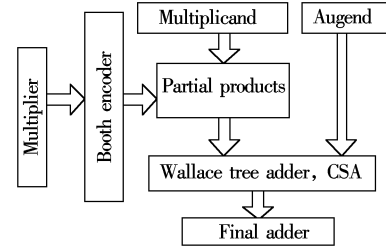


Fig. 2 Modified high-speed MAC block diagram

### 2.3 Layout design

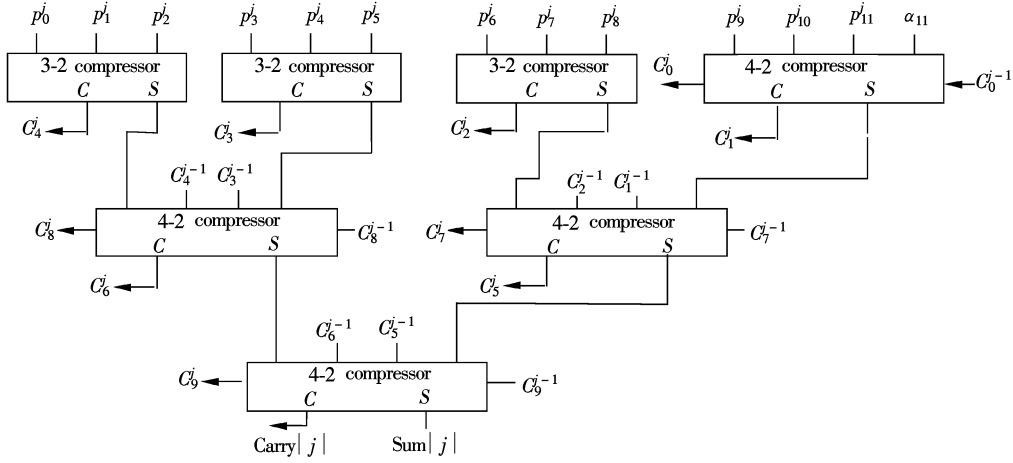
In our design of the  $24 \times 24$  bit + 48 bit parallel saturating MAC unit, for the multiplier design, we applied the radix 2 modified Booth encoding and 3-2 and 4-2 compressors applied to add the PPs in the Wallace tree adders. The pass-transistor was used as the full adder of the compressors. The overflow detection logic was added in the MAC design in order to perform saturating MAC operations.

We applied the fast carry look-ahead adder (CLA) as the 48 bit final adder of the multiplier. The 48 bit final adder array was divided into 12 groups, 4 bit per group.

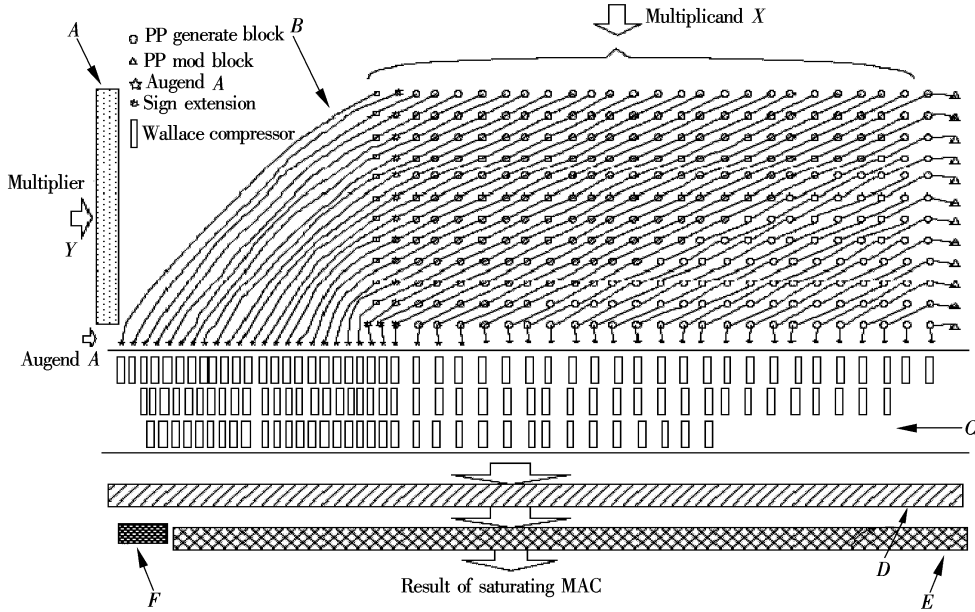
The largest Wallace tree section was composed of three 3-2 compressors and four 4-2 compressors, and the height of the tree was 3 stages. Other small trees used fewer compressors.

The structure of the largest Wallace tree in the 24 bit high-speed multiplier ( $\alpha_{11}$  is the corresponding bit of the augend) is shown in Fig. 3.

Finally, the  $24 \times 24$  bit + 48 bit layout method of MAC unit with saturation is shown in Fig. 4.



**Fig. 3** The largest Wallace tree in the 24 bit high-speed multiplier



**Fig. 4** The  $24 \times 24$  bit + 48 bit layout method of MAC unit with saturation

## 2.4 Simulation and results

The layout of the parallel saturating MAC unit was designed with SMIC 0.18  $\mu\text{m}$  1.8 V technology processing in N-well 6 metal layers. In order to reduce the area, we optimized the floorplan of the layout. The physical area of the saturation is  $679.2 \mu\text{m} \times 132.5 \mu\text{m}$  ( $0.0896 \text{ mm}^2$ ). Tab. 1 summarizes the area proportion of each part in the saturating MAC. It shows

**Tab. 1** Each part proportion in area

Parts	Area/ $\text{mm}^2$	Percent/%
Booth encoder and PPs	0.041 08	45.84
Wallace tree array	0.018 82	21.01
Final addition	0.008 75	9.77
Saturation detection and modifying logic	0.003 89	4.34
Blank	0.017 05	19.04

that the Booth encoder and PPs are the main factors in all parts of the saturating MAC.

We used the Nanosim + VCS co-simulation to simulate the MAC circuit, and the worst delay of the saturating MAC is 3.01 ns. Tab. 2 gives the path delay of the parts of the saturating MAC.

As a comparison to the conventional standard cell synthesis, a saturating MAC was also designed in standard cell netlist generator of digital CMOS process shown in Tab. 3, according to multiplication and addition and overflow logic, which are serial. Compared to the conventional standard cell synthesis, our method requires 23.52% less area.

**Tab. 2** The path delay of parts of the saturating MAC

Parts	Delay/ns	Percent/%
Booth code and PP generation circuit	0.44	14.61
Wallace tree array	1.08	35.88
Final adder	1.40	46.51
Logic to detect saturation and to modify the final value	0.09	2.99
Total	3.01	100

**Tab. 3** Comparison of custom layout and standard netlist generating layout of the saturating MAC

Comparison	Area/mm <sup>2</sup>	Delay/ns	Power/mW
Proposed	0.0896	3.01	34.6
Standard synthesis	0.1106	3.349	38.30
Reduction/%	23.52	11.27	10.7

3 Conclusion

In this paper we present a full-custom circuit design and VLSI implementation of a parallel saturating MAC unit, which can perform  $24 \times 24$  bit signed multiplications and 48 bit signed additions, with a throughput of one cycle. In the saturating MAC unit, the overflow detection and saturation operation is parallel with MAC operation, so the delay of overflow detection and saturation is beyond the critical path delay. The augend is operated as a PP in the Wallace tree in order to reduce the delay to the MAC operation. Compared with the standard cell library design methods, our method has

significantly less area, critical path delay and power dissipation.

References

[1] Farooqui A A, Oklobdzija V G. General data-path organization of a MAC unit for VLSI implementation of DSP processors [A]. In: *Proc of the IEEE International Symposium on Circuits and Systems* [C]. Monterey, CA, USA, 1998. 260 – 263.

[2] Itoh Niichi, Naemura Yuka, Makino Hiroshi, et al. A 600 MHz  $54 \times 54$  bit multiplier with rectangular-styled Wallace tree[J]. *IEEE Journal of Solid-State Circuits*, 2001, **36**(2): 249 – 257.

[3] Mustafa Gok. Integer multiplier and squarer architectures with overflow detection [D]. Bethlehem: EECS Department of Lehigh University, 2003.

[4] Yadav N, Schulte M J, Glossner J. Parallel saturating fractional arithmetic units [A]. In: *Proc of the Ninth Great Lakes Symp on VLSI* [C]. Michigan, 1999. 214 – 217.

[5] Balzola P I, Schulte M J, Ruan, J, et al. Design alternatives for parallel saturating multioperand adders[A]. In: *Proc of the International Conference on Computer Design* [C]. Austin: VLSI in Computers & Processors, 2001. 172 – 177.

[6] Schulte M J, Balzola P I, Akkas A, et al. Integer multiplication with overflow detection or saturation [J]. *IEEE Transactions on Computers*, 2000, **49**(7): 681 – 691.

[7] Fadavi-Ardekani Jalil.  $M \times N$  Booth encoded multiplier generator using optimized Wallace tree[J]. *IEEE Transactions on VLSI Systems*, 1993, **1**(2): 120 – 125.

一种具有饱和处理功能的 24 位并行乘加单元优化设计

张 萌<sup>1</sup> 贾俊波<sup>1,2</sup>

(<sup>1</sup> 东南大学国家专用集成电路系统工程技术研究中心, 南京 210096)  
(<sup>2</sup> 深圳海思有限责任公司, 深圳 518129)

**摘要:** 阐述了一种  $24 \times 24$  bit + 48 bit 带饱和处理的乘加单元的优化设计. 在乘法器的设计中, 采用改进的 Booth 算法, 并将被加数作为乘法器的一个部分积参与到 Wallace 树阵列中来完成乘加运算, 大大提高了 MAC 的性能, 同时还设计出优化的饱和检测逻辑电路. 利用  $0.18 \mu\text{m}$  1.8 V 1P6M 标准 CMOS 工艺通过全定制方式实现了面积为  $679.2 \mu\text{m} \times 132.5 \mu\text{m}$  的带饱和处理的 MAC 单元, 仿真结果表明: 它与软件设计系统综合出的传统 MAC 单元相比, 性能上有很大的改善, 在节约 23.52% 的面积情况下速度也有一定的提高.

**关键词:** 乘加器; Booth 编码; Wallace 树; 饱和检测; 布局

**中图分类号:** TP302