A Low-Power 16 × 16-b Parallel Multiplier Utilizing Pass-Transistor Logic

C. F. Law, S. S. Rofail, and K. S. Yeo

Abstract—This paper describes a low-power 16 × 16-b parallel very large scale integration multiplier, designed and fabricated using a 0.8-μm double-metal double-poly BiCMOS process. In order to achieve low-power operation, the multiplier was designed utilizing mainly pass-transistor (PT) logic circuits. The inherent nonfull-swing nature of PT logic circuits were taken full advantage of, without significantly compromising the speed performance of the overall circuit implementation. New circuit implementations for the partial-product generator and the partial-product addition circuitry have been proposed, simulated, and fabricated. Experimental results showed that the worst case multiplication time of the test chip is 10.4 ns at a supply voltage of 3.3 V, and the average power dissipation is 38 mW at a frequency of 10 MHz.

Index Terms—Low-power VLSI design, parallel multipliers, pass-transistor logic.

I. INTRODUCTION

MOST advanced digital systems today incorporate a parallel multiplication unit to carry out high-speed mathematical operations. In many situations, the multiplier lies directly in the critical-path, resulting in an extremely high demand on its speed. In the past, considerable efforts were put into designing multipliers with higher speed and throughput, which resulted in fast multipliers which can operate with a delay time as low as 4.1 ns [1]. However, with the increasing importance of the power issue due to the portability and reliability concerns of electronic devices [2], recent work has started to look into circuit design techniques that will lower the power dissipation of multipliers [3]–[5].

This paper describes the design and fabrication of a 16 × 16-b parallel multiplier, based on a 0.8-μm BiCMOS process, for low-power applications. Pass-transistor (PT) logic is chosen to implement most of the logic functions within our multiplier. Emerging as an attractive replacement for the conventional static CMOS logic, especially in the design of arithmetic macros, PT logic requires fewer devices to implement basic logic functions in an arithmetic operation, such as the XOR function. This translates into lower input gate capacitance and power dissipation as compared to conventional static CMOS [2]. In the PT circuit implementations reported so far [6]–[9], transmission-gate (TG) design techniques which provide full voltage swings were widely adopted. In this paper, we present several circuits that fully exploit the inherent nonfull-swing (NFS) nature of PT logic. These circuits were used as basic building blocks within our multiplier to achieve low-power operation.

Various proposed and reported circuit implementations of the partial-product generator (PPG) and partial-product adder (PPA) are discussed in Sections II and III, respectively. Section IV presents the experimental measurements of the test chip. All circuit simulations are based on a 0.8-μm double-metal double-poly BiCMOS process, and carried out on the HSPICE simulator.

II. PARTIAL-PRODUCT GENERATOR (PPG)

To date, the most widely adopted technique for partial-product generation in large multipliers (16-b and above) is the modified Booth’s algorithm (MBA). The main attraction of MBA is that instead of generating n partial-products for an n-b multiplication, it only generates half of that. According to MBA, a signed binary number in its two-complement form can be partitioned into overlapping groups of three bits. By coding each of these groups, an n-b signed binary number can be represented as a sum of n/2 signed digits. As each signed digit takes the possible values of zero, ±1 and ±2, the required partial-products are all power-of-two multiples of the multiplicand (X), which are readily available.

The standard PPG circuit implementation requires five control bits, each representing a “0,” “±X,” “−X,” “+2X,” or “−2X” operation. The truth table for the control bits are shown in Table I. When implemented in full CMOS, the encoders only exhibit moderate performance [4]. To improve its performance, complementary PT logic (CPL) family cells have been used in [4]. Although significant improvement in power dissipation (30%) has been reported, the CPL encoder requires 122 transistors to implement, a 150% increase compared to the CMOS encoder (48 transistors), and provides only 6% improvement in speed. We present a PT Booth’s encoder (Fig. 1) which offers better performance over both the CMOS and CPL implementations in terms of power, speed and transistor count. From Table I, it is obvious that the control bit for “0” is high when $Y_{2i-1}$, $Y_{2i}$, and $Y_{2i+1}$ are the same.

Manuscript received November 17, 1998; revised March 20, 1999.

The authors are with the Division of Circuits and Systems, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: eksyeo@ntu.edu.sg).

Publisher Item Identifier S 0018–9200(99)08225-6.
Therefore, it can be expressed as

\[ 0 = \left( Y_{2l-1} + Y_{2l} \right) \cdot \left( Y_{2l-1} + Y_{2l+1} \right) \tag{1} \]

The control bit for \( +X \) is high when \( Y_{2l-1} \) and \( Y_{2l} \) are different, provided \( Y_{2l+1} \) is low. The same is true for \( \bar{X} \), except that \( Y_{2l+1} \) must now be high. The expressions for these control bits are

\[ +X = \left( Y_{2l-1} + Y_{2l} \right) \cdot \bar{Y}_{2l+1} \tag{2} \]

\[ \bar{X} = \left( Y_{2l-1} + Y_{2l} \right) \cdot Y_{2l+1} \tag{3} \]

It is clear from (1)–(3) that an XOR operation between \( Y_{2l-1} \) and \( Y_{2l} \) should be performed to generate all three control bits. A PT XOR-XNOR pair carries out this operation and the results (XOR and XNOR) are fed simultaneously into three PT AND-NAND pairs to generate the respective control bits. The control bits for \( +2X \) and \( \bar{2}X \) are generated using conventional CPL logic style and therefore will not be discussed. The proposed circuit was compared with the CMOS and CPL circuits, and the results are shown in Table II.

<table>
<thead>
<tr>
<th>Improvement of Proposed Encoder over CMOS and CPL Encoders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Delay (ps)</td>
</tr>
<tr>
<td>Power (mW)</td>
</tr>
<tr>
<td>Power x Delay (ps)</td>
</tr>
<tr>
<td>Transistor Count</td>
</tr>
</tbody>
</table>

III. PARTIAL-PRODUCT ADDER

One common approach to partial-product addition is to use regular adder array, which has a regular structure and is easy to layout. However, it suffers from poor speed performance and power wastage due to spurious transitions. For large multipliers, another approach, the Wallace reduction technique, is usually used. This approach leads to much better speed performance due to the high-level of parallelism employed in the Wallace tree-adder, constructed using multiple-input compressors that can sum up several partial-products concurrently. The second approach was adopted in our circuit implementation to obtain the best speed performance possible, while PT logic circuits with NFS nodes were used to reduce the power of the Wallace tree-adder.

In a \( 16 \times 16 \)-b multiplier utilizing the MBA, there are eight partial-products to be added up. Thus, the 4-2 compressor was chosen as the basic building block of the PPA. It receives five input bits of the same weight \( (P_0, P_1, P_2, P_3, \text{ and } C_{in}) \), compresses them, and generates three output bits \( (S, C, \text{ and } C_{out}) \). Various circuit implementations of the 4-2 compressor have been reported. Full CMOS implementations usually suffer from high transistor count and input gate capacitance, leading to only moderate speed and power performance. The pseudo-CMOS implementation proposed in [10] (using a 0.5-\( \mu \)m
CMOS technology) utilized $n$-channel PT’s to reduce the transistor count of the basic building gate (an XOR gate), and an improvement of 12.5% in speed over the full CMOS circuit has been reported. To further simplify the design, a PT-multiplexer-based circuit comprising only of TG’s, using a 0.25-$\mu$m CMOS technology, was proposed in [6]. Using this technique, a multiplication time of 4.4 ns has been reported. Clearly, considerable efforts have been directed toward simplifying the design of the compressors and improving the speed of the Wallace tree-adder. Its power dissipation, however, was never a major consideration. We present a 4-2 compressor circuit design (Fig. 2) which is an improved version of the one proposed in [6], requiring fewer transistors to implement and consuming less power. The proposed design takes full advantage of the NFS nodes that are inherent in PT logic circuits. As shown in Fig. 2, it consists of two types of PT multiplexers, one providing NFS outputs (NFS MUX), and the other providing full-swing (FS) outputs using two PMOS pull-up devices (FS MUX). For each 4-2 compressor, only the internal nodes $n1$, $n2$, $n5$, and $n6$, and the output $S$, are pulled up to $V_{DD}$ for logic high, while the rest only reaches approximately $V_{DD} - V_T$. Among the NFS nodes are both the output carry signals ($C$ and $C_{out}$) and internal nodes $n3$ and $n4$. Special care is taken in routing the compressors to form the PPA. Since the outputs of the compressors in the first stage drive the inputs of those in the second stage, $C$ of the first stage, being a NFS node, must be used to drive $P_1$ or $P_3$ (which can accept NFS logic high) of the second stage. While $S$, being an FS node, can be used to drive any of the second stage compressor’s inputs. With this technique, about 50% of the nodes within the PPA are non full swing. Furthermore, as only two of the 4-2 compressor’s inputs require full voltage swing ($P_1$ and $P_2$), only four of the eight partial-products generated by PPG are required to achieve full swing. This leads to significant power reduction for the multiplier.

The PPA was implemented using the various 4-2 compressors discussed above and comparisons, in terms of speed, power, power-delay product, and transistor count, were made at 3.3 V. The simulation results are shown in Table III. When compared to the pseudo-CMOS implementation, the proposed implementation achieved significant improvements in delay, power, and transistor count. The shorter delay in the proposed implementation (2.2 ns) is due to its much shorter critical-path (three PT multiplexers) compared to the pseudo-CMOS implementation (two $n$-channel XOR gates and two CMOS complex gates). The presence of NFS nodes and 48% cut in the transistor count has led to an improvement in power dissipation by 44%. Significant improvements over the TG implementation in terms of power (62%) and transistor count (37.5%) were also obtained. Although the lower current drive capability of NFS nodes, as compared to FS-nodes, has caused the proposed implementation to suffer a 16% decrease in speed, the power-delay product still improved by over 50%. In conclusion, the low-power and low-transistor-count characteristics of the improved PT 4-2 compressor are very useful in the design of a low-power high-performance PPA, with relatively small circuit area.
**TABLE IV**

**CHARACTERISTICS OF FABRICATED 16 × 16-b MULTIPLIER**

<table>
<thead>
<tr>
<th>Process</th>
<th>0.8 μm double-metal double-poly BiCMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size of multiplier and multiplicand</td>
<td>16</td>
</tr>
<tr>
<td>Size of product</td>
<td>32</td>
</tr>
<tr>
<td>Multiplication time (3.3 V)</td>
<td>10.4 ns</td>
</tr>
<tr>
<td>Power dissipation (3.3 V, 10 MHz)</td>
<td>38 mW</td>
</tr>
<tr>
<td>Core Area</td>
<td>2.8 × 1.2 mm²</td>
</tr>
<tr>
<td>Transistor Count</td>
<td>5180</td>
</tr>
</tbody>
</table>

**IV. FABRICATION AND EXPERIMENTAL EVALUATION**

The multiplier was fabricated on a test chip using a 0.8-μm double-metal double-poly BiCMOS process. To measure its worst case multiplication time, input test patterns are applied to trigger its critical-path, which includes a Booth’s encoder, a control-line buffer, a partial-product selector, two 4-2 compressors, a half-adder, and the 32-b two-operand carry-select adder, with carry propagation from the fourteenth to the highest (thirty-first) bit-position. One such pattern is shown in Fig. 3. The worst case (rise) delay is measured to be 10.4 ns. The average power dissipation of the test chip, inclusive of the multiplier core, input/output pads, output multiplexers and testing circuitry, with no probes at the outputs, is 38 mW. The multiplication time and power dissipation of the fabricated device are measured for the supply range of 2.5 V to 5 V, and the results are compared with some of the reported multipliers of the same width, as shown in Fig. 4. At 3.3 V, the multipliers reported in [11] and [12], which used a 0.6- and 0.5-μm CMOS technology, respectively, achieved, as expected, better multiplication time compared to our work. Our multiplier, however, provides significant saving in power. At 10 MHz, it is less than half that of [11] and even less when compared to [12]. Similar observation is made at 4 V when our multiplier is compared to the one reported in [7] which is based on a 0.5-μm CMOS process, where over 50% reduction in power is obtained. The characteristics of the fabricated device are shown in Table IV.

**V. CONCLUSION**

We have presented several low-power PT circuit techniques for parallel multiplication. Taking full advantage of the low-transistor-count and NFS nature of PT logic, we have successfully implemented low-power circuit blocks which serve as basic building units within a 16 × 16-b multiplier, including a new Booth’s encoder and a modified 4-2 compressor. Experimental measurements on the fabricated multiplier and comparisons with other reported multipliers have verified its low-power characteristics. The total power dissipation of the test chip at 3.3 V is 38 mW at 10 MHz, with a worst case multiplication time of 10.4 ns.

**REFERENCES**


