An Area-Delay Optimized 4×4-Bit Carry-Save Array Multiplier In 180nm CMOS Technology

Chenchen Ding *IEEE Student Member*

Southern University of Science and Technology
School of Microelectronics
Shenzhen, China
11912401@mail.sustech.edu.cn

Shuo Feng *IEEE Student Member*

Southern University of Science and Technology
School of Microelectronics
Shenzhen, China
11911736@mail.sustech.edu.cn

***Abstract*—This report presents a 4x4-bits array multiplier, which has two 4-bit unsigned binary numbers as inputs, and a 8-bit binary product of these two numbers as the output. In this design, we implement the circuit using the 0.18-µm tsmc technology. The supply voltage is set to 1.8V, and the load of the 8-bit bus output are 50-fF capacitors for each bit. The objective of the project is to realize the 4x4-bit multiplier with small area and low latency, and the post-layout delay of the circuit is 1.0085ns and the overall area is 945.4 μm^2 The procedure of the circuit and layout design and the demonstrations of the post-layout simulation results are reflected in the report.**

***Index Terms*—Low latency circuits, Cadence, IC design, digital circuit, CMOS technology, multiplier, parasitic effect.**

I. Introduction

MULTIPLIERS are crucial components of the digital integrated circuit system. Since multiplication dominates the execution time of most digital signal processing (DSP) algorithms, there is a need of the high-speed multipliers. The application of this module is wide, such as for convolution, Fast Fourier Transform (FFT), filtering and in ALU of microprocessors.

Past work has made it possible to implement arbitrary bit multipliers, using circuit simulation as well as physical testing of the chips. [1] suggested a fast multiplier which decreases the latency by reorganizing the additions. [2] mentioned a modified booth's multiplier design which can provide higher speed performance of multiplier than the traditional **carry-propagate array multiplier** which is shown in Figure 1. Complementary Metal Oxide Semiconductor (CMOS) technology which is widely used in integrated digital circuit design enjoys its efficient usage of electrical power.

Based on the covered knowledge in the course SME 306 and the literature review, the delay of the multipliers, is one of the most critical performances. Thus, we focus on the optimization of the delay time. In addition, the small size is also important because it makes the system much more portable. Therefore, the motivation of this project is to design a widely used digital circuit cell, a 4x4-bit array multiplier. Specifically, our team aims to realize a small-area and low-latency 4x4-bit array multiplier. In this work, the design of **carry-save multiplier** describe in [3] is applied and our design is formed by the Full-Adder, Half-Adder and And logical circuit in complementary design (figure 2).

The Section II will give the objectives of our design; the advantages of the adopted structure will be given in Section III; the Section IV is the analysis for each cell; and Section V presents the layout and the simulation results of the overall structure.

This paper was produced by the students from the cource SME 306. They are in School of Microelectronics, SUSTech, SZ.

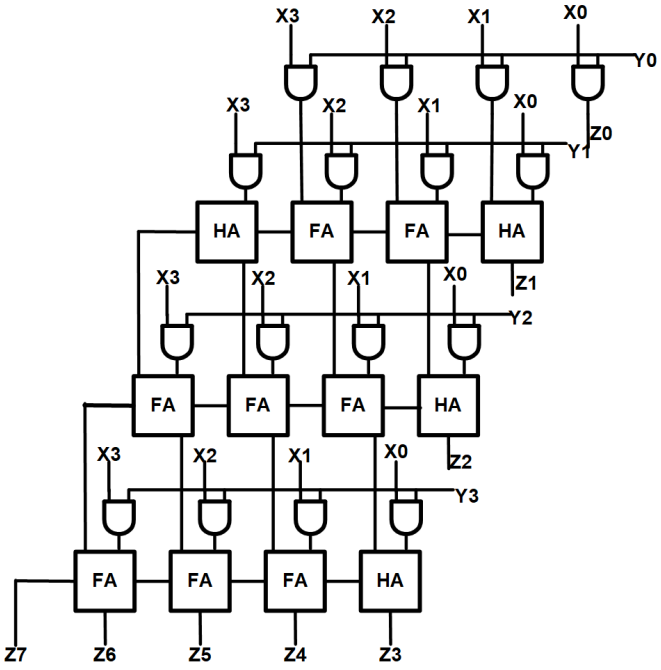


Fig. 1. Design of carry-propagate array multiplier

II. Problem Formulation

The project provides a schematic and layout design for a 4×4-bit array multiplier. The two 4-bit unsigned input numbers $\boldsymbol{X}$ and $\boldsymbol{Y}$ are set to X[3:0] and Y[3:0], respectively. 8-bit

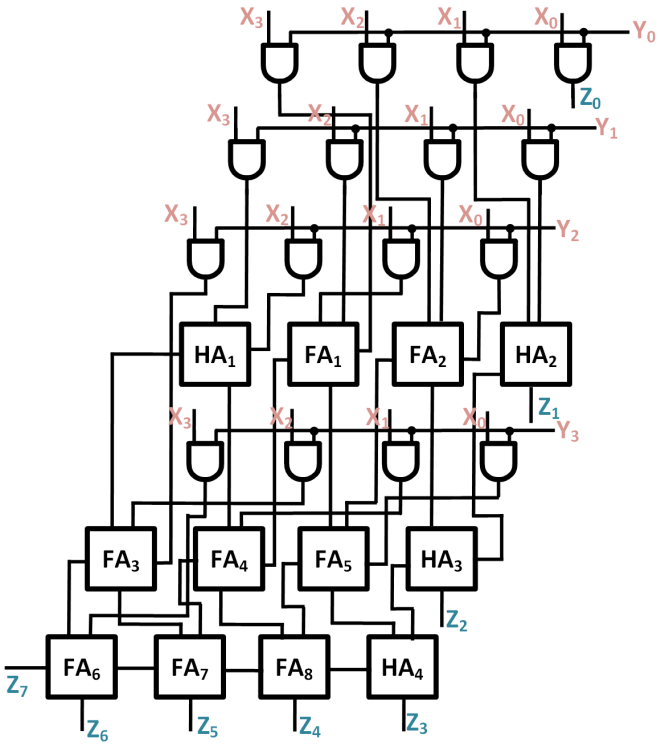


Fig. 2. Design of carry-save array multiplier (the adopted overall circuit structure).

output is set to Z[7:0]. A number with a subscript 0 means the least significant bit (LSB).

The performance, power and area (PPA) are the evaluation criteria of the circuit design. Pursuing higher rates means more complex logic, which may make the area increase; while pursuing smaller areas may bring poorer performance (too much latency), which is a trade-off. In this work, we aim to design a Performance-Area balanced 4×4 unsigned multiplier in a reasonable power consumption.

The upper bound of execution speed an the power consumption of multipliers are mainly limited by two factors: **semiconductor process**, and **multiplier architecture**. The area of a design is mainly influenced by sizes of transistors and the layout design. In this work, 0.18-um technology is used and the circuit structure is based on CMOS logic circuit, which is covered in the course of SME 306.

A. The determination of standard cells

The adopted multiplier consists of the blocks: **Full Adder**, the **Half Adder** and the **And Gate**. The overall structure contains these equivalent-height cells, which has been well placed and routed.

1) Schematic of And: The logical expression of And is

$$Y = \overline{\overline{A} + \overline{B}} = \overline{\overline{AB}} \tag{1}$$

Thus, we have the schematic structure of it as Figure 3

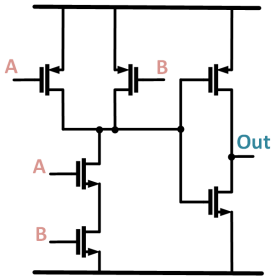


Fig. 3. The schematic of And Gate.

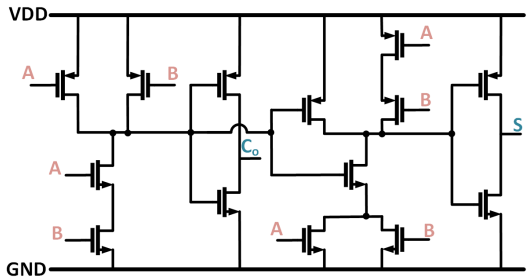


Fig. 4. The schematic of Half Adder.

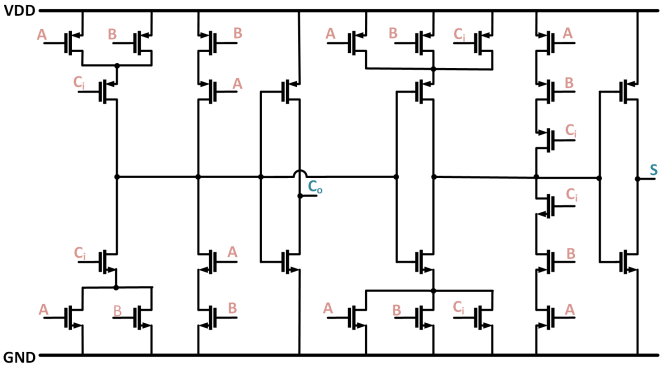


Fig. 5. The schematic of Full Adder.

2) Schematic of Half Adder: The logical expression of the Half Adder is

$$C_o = \overline{\overline{A} + \overline{B}} = \overline{\overline{AB}} \tag{2}$$

$$S = \overline{(\overline{A} + \overline{B}) + \overline{AB}} = \overline{\overline{AB}(A + B)} \tag{3}$$

Thus, we have the schematic structure of it as Figure 4

3) Schematic of Full Adder: In this project, the mirror full adder with 28-gate structure is chosen, thus, the logical expression of the Full Adder is

$$C_o = \overline{(\overline{A} + \overline{B})\overline{C_i} + (\overline{AB})} = \overline{(A + B)C_i + (AB)} \tag{4}$$

$$S = \overline{(\overline{A} + \overline{B})\overline{C_i} + (\overline{AB})(\overline{A} + \overline{B} + \overline{C_i}) + \overline{ABC_i}} \tag{5}$$

$$= \overline{(A + B)C_i + (AB)}(A + B + C_i) + (ABC_i)} \tag{6}$$

Thus, we have the schematic structure of it as Figure 5.

III. CMOS Design of 4x4-Bit Array Multiplier

This section first starts by describing the overall architecture of the circuit in gate-level. The comparison and the final selection of the used blocks and other designs in structure and size is also presented.

The overall structure of the multiplier that we finally adopted is shown in figure 2. This circuit structure consists of 12 two-input and gates, 4 half adders and 8 full adders. As analyzed above, this architecture can reduce the latency effectively as well as make sure that the area is small enough. Inputs are X_3 - X_0, and Y_3 - Y_0, and the outputs are the product of these two binary 4-bit number, and each bit is from Z_7 to Z_0, respectively.

A. The Analyze of Power Consumption

The power consumption is one of the most important factor of circuit design especially in integrated digital circuit design. The structure of the multiplier can be: **Pesudo-NMOS Structure**, **Transmission-Gate-Based Structure**, **CMOS Logic Gate Structure**, etc. In CMOS circuits, the static power consumption is mainly the power consumption caused by the leakage current. For a cmos circuit, there is no DC on-current at steady state, and ideally the static power consumption is zero. Therefore, the CMOS logic gate structure has both the lowest power consumption and the fast performance in the meantime. Hence, in this work the **CMOS Logic Gate** structure is chosen.

B. The Analyze of Performance (Delay)

The time-delay is one of the main factors to the performance of a digital circuit. In this part, we discuss the time-delay from two aspects: the logical design of the 4 ×4 multiplier and the parameter decision of the $\boldsymbol{W}$.

1) Logical Design: Besides the traditional array-multiplier in Figure1, [1]–[3] also provide several kinds of design of multiplier with smaller latency. In this work, the **carry-save multiplier** configuration is chosen for its efficiency of dealing with carry bits. Comparing the critical path of the design in Figure 6 with Figure 7, we can find the carry-save multiplier saves 2-full-adder-delay which means the latency of carry-save multiplier can be roughly estimated as 3/4 latency of the conventional carry-propagate multiplier.

2) Parameter Decision: For each single transistor, the $\boldsymbol{W}$ and $\boldsymbol{L}$ are two main parameters which decide its electrical property. In this work, 180 nm process is used, to obtain the fastest circuit, the $\boldsymbol{L}$ of each transistor is 180 nm. Thus, the decision of $\boldsymbol{W}$ is vital to this design. To decide the $\boldsymbol{W}$, we use the parameter analysis tool to find the best choice. Considering the resistance matching, we firstly determine the ratio W_p/W_n. As the carrier mobility of NMOS is always larger than PMOS, the searching region is decided from 1.0 to 2.0. To simplify the question and to facilitate layout design, all the W_ns and W_ps are set to the same for all the transistors. Figure 8(a) shows the simulated timing waveform of output signals when sweeping parameter Wp/Wn. And in Figure 8(b) we can find that, when ratio become larger, from 1.0 to 2.0, the performance of Z_7 gets better but the thorn of Z_6 becomes larger. After scanning the parameters, it was decided that the ratio of 1.5 was the most suitable ratio. After deciding the ratio, we should chose a proper W_n for each transistor with

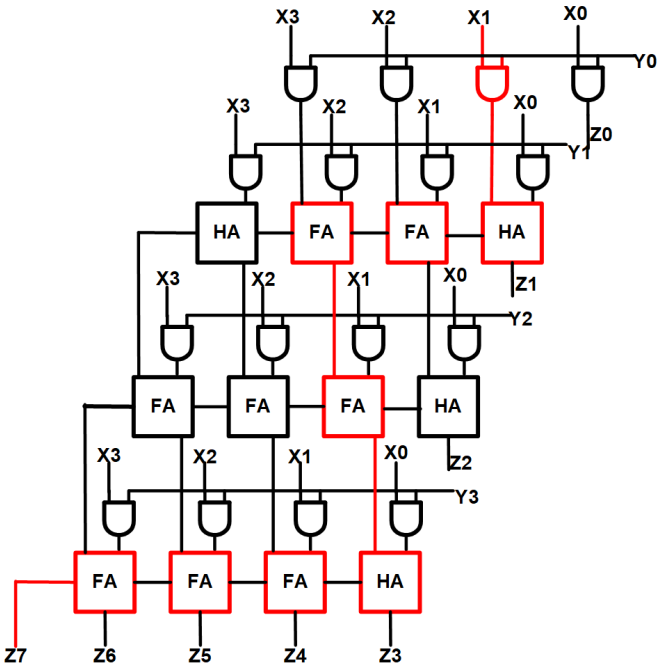


Fig. 6. Critical path of the conventional carry-propagate structure.

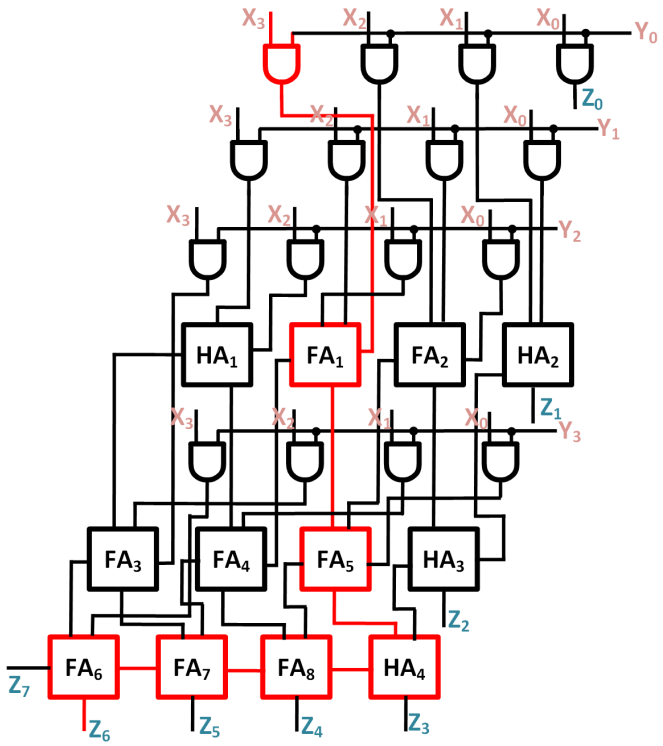
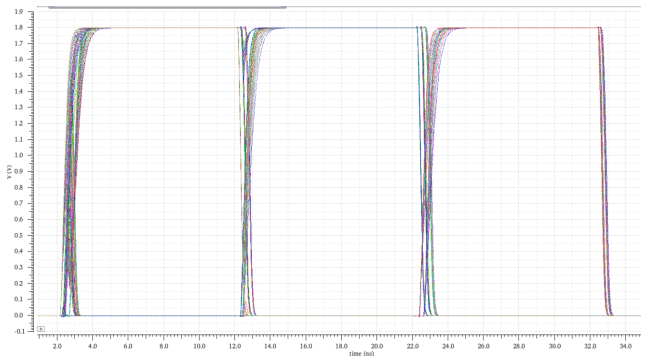
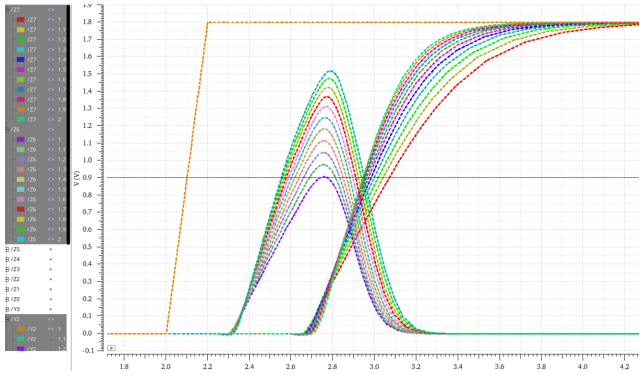


Fig. 7. Critical path of the proposed carry-save structure.



(a) Overall waveform of the outputs.



(b) Waveform of Z_6 and Z_7.

Fig. 8. The simulated timing waveform of output signals when sweeping parameter Wp/Wn.

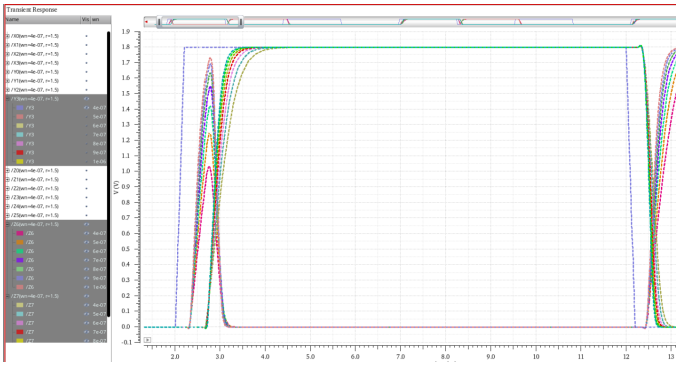


Fig. 9. Simulations of the output signals of Z_6 and Z_7.

parameter analysis. For real circuit, the considering factors varies, but in this project, only $Resistance$ is extracted, hence, this question is simplified to reduce the resistance. In previous study we know the equivalent resistance of the MOSFET in linear region is:

$$R_{dson} = \frac{\partial V_{ds}}{\partial I_{ds}} \approx \frac{1}{\mu C_{ox} \frac{W}{L}(V_{gs} - V_t)} \tag{7}$$

In this project, L is minimized to 180 nm thus, if we want to decrease the resistance, the W should be resonablly enlarged. In Figure 9, we can find that, if we ignore the resistance and the capacitance of the wires, when the size of W_n becomes larger, the delay of Z_7 decreases and the delay of Z_6 increases due to the larger thorn. Although, the total delay tends to decrease when the size gets larger, we finally choose W_n with 500 nm and W_p with 750 nm in consideration of the total area and delay simultaneously.

C. The Analysis of Area

Area is one of the most important factor in this project, which is mainly determined by the building-cells which is determined by the size of W & L. In addition, the placement and route skills are also essential for small area. In the previous parts the size of transistors are determined, and to minimize the layout area, the cell configuration is tightly placed based on the design rules. The layout will be shown in the later

Cell		AND					
	Load Capacitor	C1 = 15f		C2 = 8 * C1		C3 = 84* C1	
	Simulation Type	pre	post	pre	post	pre	post
Delay Time Data	Delay Time t_pLH	0.178ns	0.213ns	0.744ns	0.773ns	5.261ns	5.235ns
	Delay Time t_pHL	0.112ns	0.135ns	0.402ns	0.425ns	2.697ns	2.705ns
	Rise Time t_r	0.206ns	0.233ns	1.402ns	1.418ns	11.21ns	10.901ns
	Fall Time t_f	0.095ns	0.108ns	0.627ns	0.631ns	4.934ns	4.878ns
Transistor Size	Every NMOS W/L	500nm/180nm					
	Every PMOS W/L	750nm/180nm					

Fig. 10. Delay data and transistors size of AND Gate.

Cell		Half Adder					
	Load Capacitor	C1 = 15f		C2 = 8 * C1		C3 = 84* C1	
	Simulation Type	pre	post	pre	post	pre	post
Delay Time Data of Co	Delay Time t_pLH	0.193ns	0.222ns	0.752ns	0.782ns	5.195ns	5.261ns
	Delay Time t_pHL	0.162ns	0.217ns	0.455ns	0.515ns	1.726ns	2.781ns
	Rise Time t_r	0.207ns	0.233ns	1.403ns	1.429ns	10.934ns	10.874ns
	Fall Time t_f	0.106ns	0.127ns	0.627ns	0.641ns	4.861ns	4.908ns
	Load Capacitor	C1 = 15f		C2 = 8 * C1		C3 = 84* C1	
	Simulation Type	pre	post	pre	post	pre	post
Delay Time Data of S	Delay Time t_pLH	0.178ns	0.351ns	0.800ns	0.906ns	5.340ns	5.346ns
	Delay Time t_pHL	0.112ns	0.301ns	0.540ns	0.618ns	2.848ns	2.929ns
	Rise Time t_r	0.206ns	0.263ns	1.394ns	1.421ns	10.838ns	10.939ns
	Fall Time t_f	0.095ns	0.163ns	0.635ns	0.906ns	4.842ns	4.846ns
Transistor Size	Every NMOS W/L	500nm/180nm					
	Every PMOS W/L	750nm/180nm					

Fig. 11. Delay data and transistors size of Half Adder.

Cell		Full Adder					
	Load Capacitor	C1 = 15f		C2 = 8 * C1		C3 = 84* C1	
	Simulation Type	pre	post	pre	post	pre	post
Delay Time Data of Co	Delay Time t_pLH	0.155ns	0.182ns	0.744ns	0.737ns	5.170ns	5.225ns
	Delay Time t_pHL	0.160ns	0.191ns	0.452ns	0.488ns	2.719ns	2.752ns
	Rise Time t_r	0.201ns	0.227ns	1.387ns	1.338ns	11.016ns	10.857ns
	Fall Time t_f	0.105ns	0.125ns	0.622ns	0.646ns	4.904ns	4.869ns
	Load Capacitor	C1 = 15f		C2 = 8 * C1		C3 = 84* C1	
	Simulation Type	pre	post	pre	post	pre	post
Delay Time Data of S	Delay Time t_pLH	0.191ns	0.273ns	0.750ns	0.837ns	5.212ns	5.187ns
	Delay Time t_pHL	0.271ns	0.370ns	0.577ns	0.695ns	2.880ns	3.005ns
	Rise Time t_r	0.300ns	0.259ns	1.385ns	1.422ns	11.020ns	11.065ns
	Fall Time t_f	0.313ns	0.168ns	0.639ns	0.677ns	4.882ns	4.860ns
Transistor Size	Every NMOS W/L	500nm/180nm					
	Every PMOS W/L	750nm/180nm					

Fig. 12. Delay data and transistors size of Full Adder.

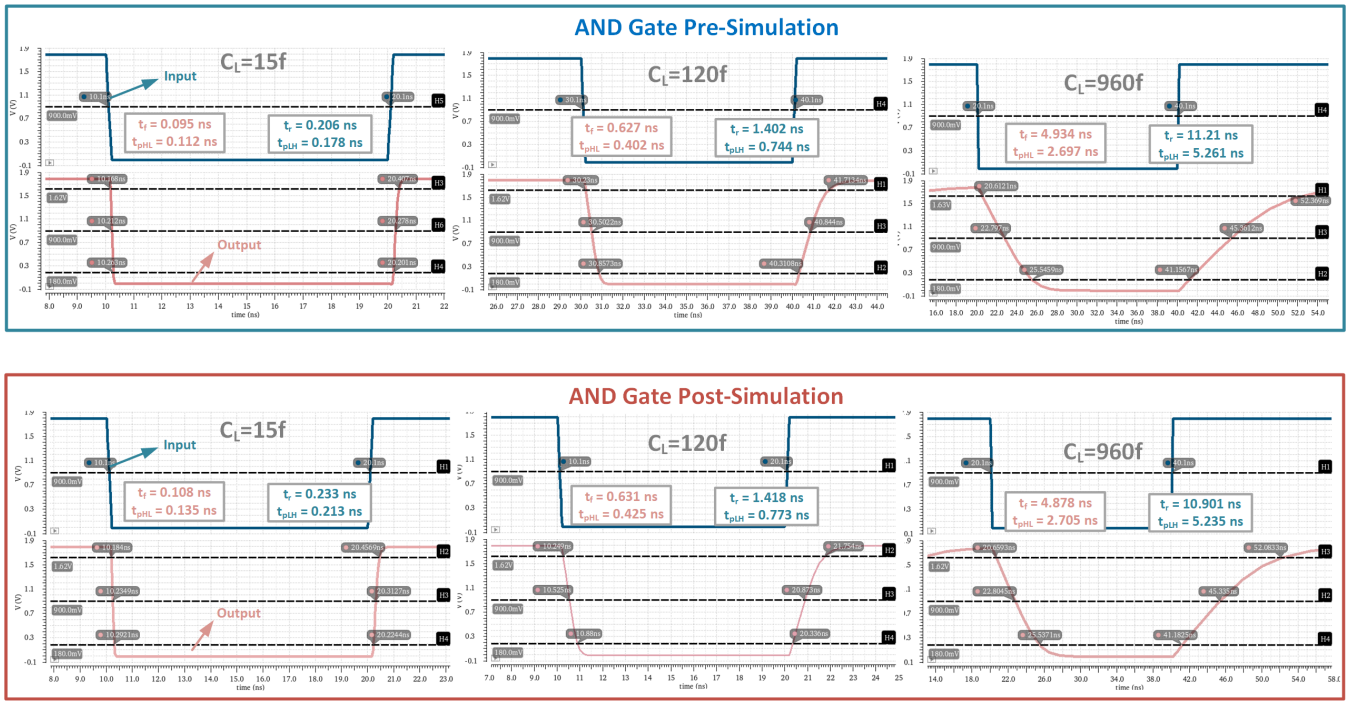


Fig. 13. The Pre and Post Simulation of single And Gate.

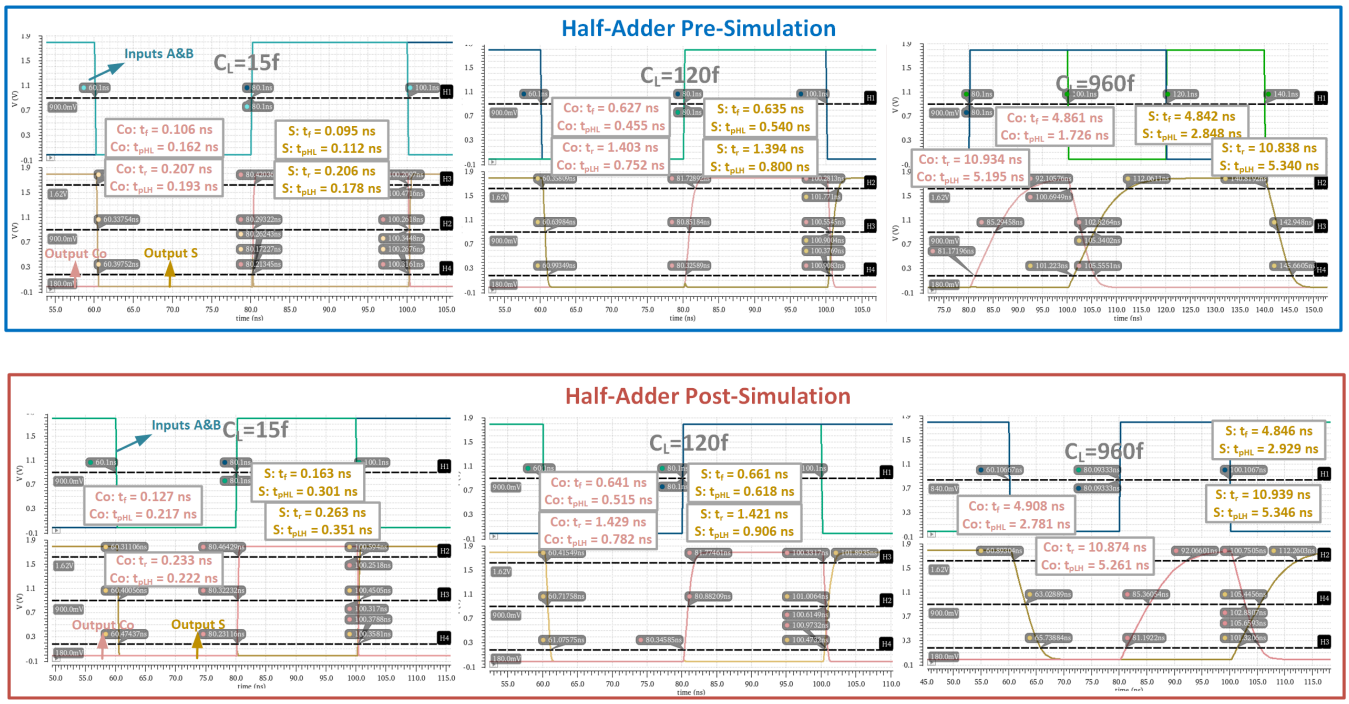


Fig. 14. The Pre and Post Simulation of single Half Adder.

chapters.

IV. Cells Description and Simulation Results

A. Standard Cell Library

Our standard cell liblary contains three cells: **AND Gate**, **Full-Adder**, **Half-Adder** cells. This section will presents a table with the transistors sizes (both width and length in micrometers), the delay data (output low-to-high propagation delay t_pLH, output high-to-low propagation delay t_pHL, rise time t_r and fall time t_f of each basic cell that is used in the multiplier design. The definitions of the four kinds of delay data are as follows:

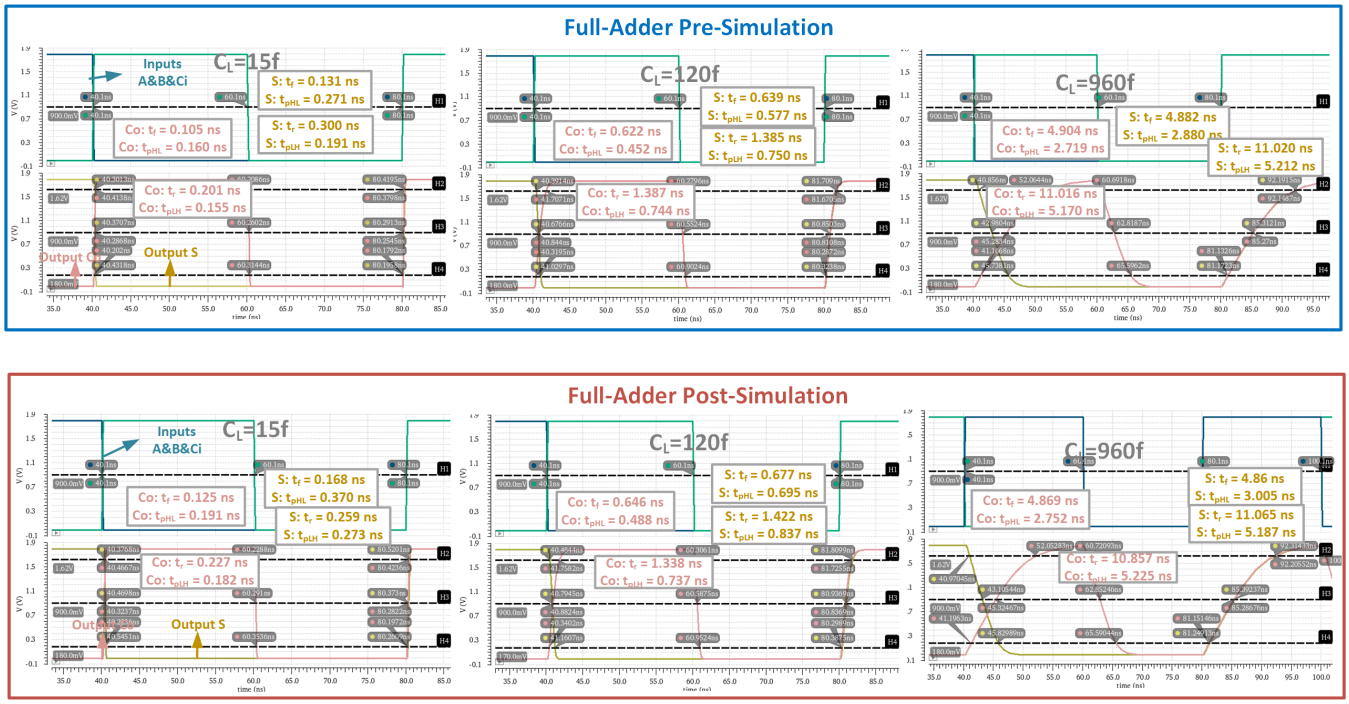


Fig. 15. The Pre and Post Simulation of single Full Adder.

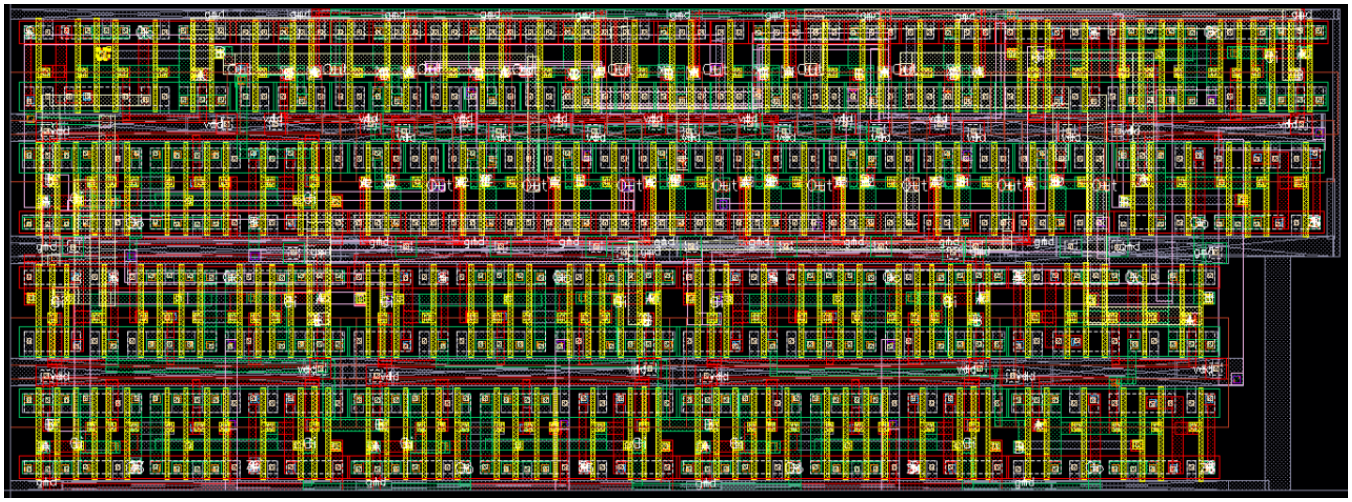


Fig. 16. The overall layout design of the multiplier.

- **Propagation delay time** t_{pLH}: this is the time difference between the input transition (50%) and the output level, specifically, the output changes from Low to High level.
- **Propagation delay time** t_{pHL}: this is the time difference between the input transition (50%) and the output level, specifically, the output changes from High to Low level.
- **Rise time** t_r: the time for a waveform to rise from 10% to 90% of its steady state value.
- **Fall time** t_f: the time for a waveform to fall from 90% to 10% of its steady state value.

The corresponding delay time data and the size of the transistors in each cell are shown in figure 10, 11 and 12, respectively.

B. And Logic Gate Circuit Design

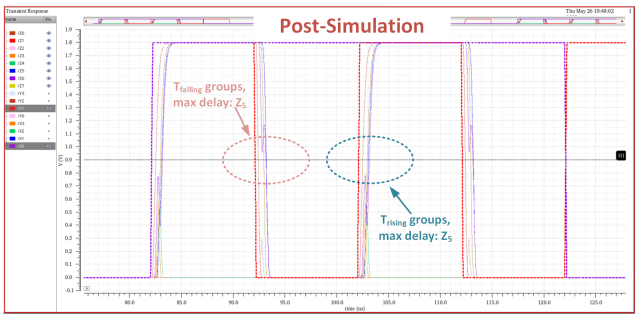
The circuit structure of And Gate is shown in figure 3, and the size of the transistor is shown in figure 10. The simulation of the single And Gate is shown in figure 13.

C. Half Adder Circuit Design

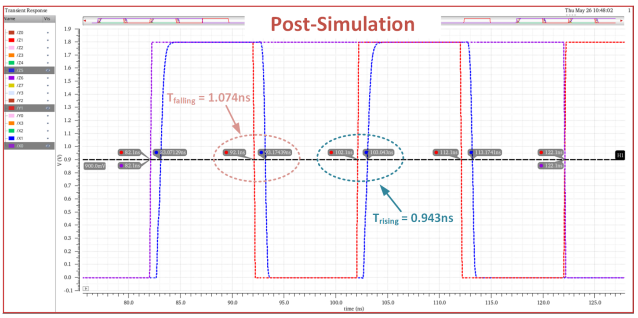
The circuit structure of Half Adder is shown in figure 4, and the size of the transistor is shown in figure 11. The simulation of the single And Gate is shown in figure 14.

D. Full Adder Circuit Design

The circuit structure of Full Adder is shown in figure 5, and the size of the transistor is shown in figure 12. The simulation

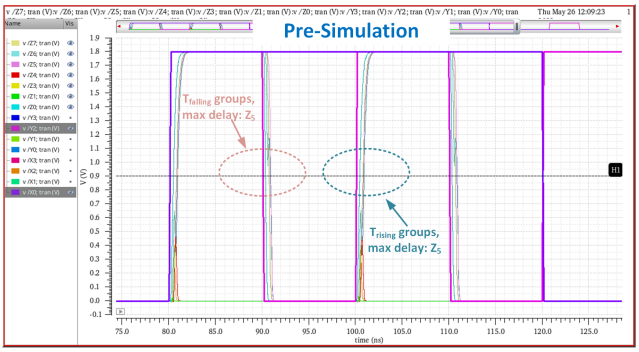


(a) Selection of the output signal with the maximum delay of post-simulation.

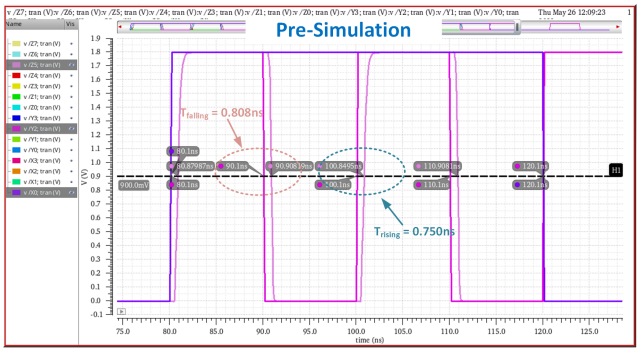


(b) The marked value of propagation delay of post-simulation.

Fig. 17. Post layout simulations propagation delay.



(a) Selection of the output signal with the maximum delay of pre-simulation.



(b) The marked value of propagation delay of pre-simulation.

Fig. 18. Pre-simulations propagation delay.

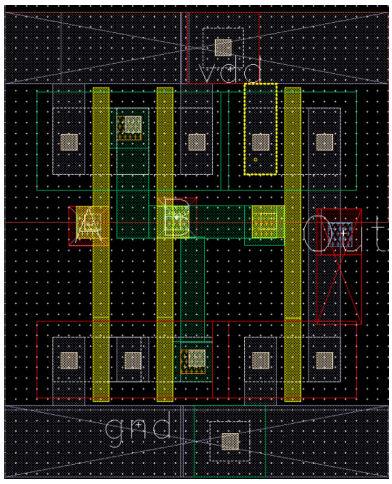


Fig. 19. The layout of And Gate.

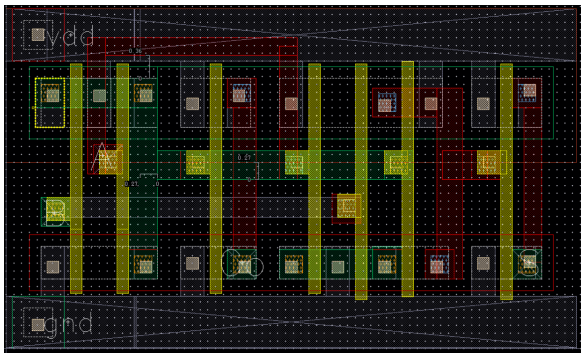


Fig. 20. The layout of Half Adder.

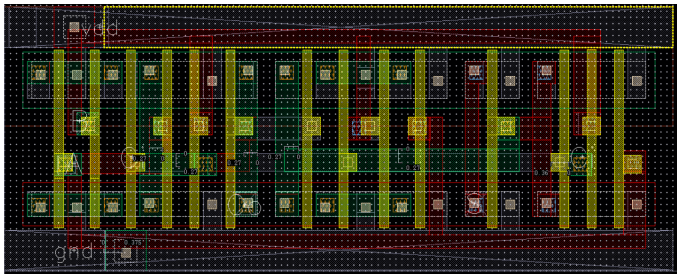


Fig. 21. The layout of Full Adder.

of the single And Gate is shown in figure 15.

V. 4x4-Bit Array Multiplier Layout and Delay

A. Layout Display

The standard cells are shown in figure 19, 20, 21. And the final layout of multiplier is shown in Figure 16.

TABLE I
THE SUMMARY TABLE OF THE PARAMETERS

Overall Parameters		
List	**Description**	**Value**
W_p	The width of PMOS transistors	750 nm
L_p	The length of PMOS transistors	180 nm
W_n	The width of NMOS transistors	500 nm
L_n	The length of NMOS transistors	180 nm
C_l	The load capacitance for all outputs	50 fF
V_{dd}	The supply voltage	1.8 V
Extraction Mode	The parasitic extraction mode	**R Only**

The lengths and widths of the layout are 51.16 um and 18.48 um, respectively. Therefore, the overall area of the layout is:

$$Area = 51.16 \times 18.48 \mu m^2 = 945.4 \mu m^2 \tag{8}$$

B. Delay Analysis

The selection of the maximum delay is based on the simulation results as shown in figure17(a). The maximum post-layout propagation delay is marked in figure 17(b), and the value is:

$$Delay_{post} = \frac{t_{pHL} + t_{pLH}}{2} = \frac{1.074ns + 0.943ns}{2} = 1.0085ns \tag{9}$$

Next we compare the post-layout propagation delay with the schematic level delay. The decision of the maximum delay in pre-simulation is similar with that in post-layout simulation (figure 18(a)). As shown in figure 18(b), the maximum pre-simulation propagation delay value is:

$$Delay_{pre} = \frac{t_{pHL} + t_{pLH}}{2} = \frac{0.808ns + 0.750ns}{2} = 0.779ns \tag{10}$$

It is obvious to see that the post-simulation is much slower than the pre-simulation, about 29%. Here we use the post-simulation with the parameter of extracting the parasitic resistor. The presence of the parasitic resistor makes the time constant RC increase, so the latency increases and the corresponding propagation delay increases. The schematic level circuit simulation tool does not take the parasitic effect of placement and route of layout into consideration, so such simulation results are not accurate.

According to [4], the parasitic resistance and capacitance extracted based on our layout design can be critical in affecting the actual performance of the circuit. In order to get an idea of how the design would work from the layout, we should perform a post-layout simulation from the extracted view. Finally, the summery table of the parameters is shown in Table I.

C. Some Drawbacks of the Project

Actually in project, we are asked to choose "R Only" when do the extract. In this situation, the parasitic capacitance caused by the layout was ignored and not extracted. Actually the capacitance would more severely cause the delay to this circuit. Hence, these these tests above can't represent the "real performance" of this circuit design. Furthermore, in this simulation, deviation caused by process is ignored and the input signals are pure and ideal which is impossible to get in real test situations. These would also leads to the huge difference of the performance between the real circuit and simulated circuit in this project.

VI. CONCLUSION

In the project, a 4x4-bit carry-save array multiplier with maximum propagation delay of 1.0085 ns and total layout area of 945.4 μm^2 is presented. We compared several structures of the multiplier and select the carry-save structure eventually. From this project, we learned the process of IC design: from schematic design to post layout extraction and simulation, and developed good design techniques. Through layout design, we learned the importance of post-layout simulation and how to optimize the actual layout for fabrication. We encountered the following problems. (1) the selection of the structure of the circuit: we started to use the traditional structure, but later found that the delay is large, so after the literature review, we optimized to a new structure, that is carry-save structure, although it took a lot of time, we also gained a lot. (2) transistor size selection: we needed to iterate to select the best size, so we chose to sweep the parameters and select the size at the top-level (i.e., in the big structure of the multiplier). In fact, this work is worth doing in every cell, and we will try more, and we believe that the size selection can be further optimized. (3) the layout can be further optimized, our current strategy is to ensure the smallest area of different modules placement, but did not take into account the alignment of each cell's IO port, which may lead to a larger parasitic situation, which requires comprehensive consideration. Therefore, we plan to optimize the circuit from two aspects: one is based on the circuit itself, from the transistor size and layout planning; on the other hand, adding more complex logic or other circuit components to improve the circuit performance.

REFERENCES

[1] C. S. Wallace, "A Suggestion for a Fast Multiplier," in IEEE Transactions on Electronic Computers, vol. EC-13, no. 1, pp. 14-17, Feb. 1964, doi: 10.1109/PGEC.1964.263830.

[2] N. R. Shanbhag and P. Juneja, "Parallel implementation of a 4*4-bit multiplier using a modified Booth's algorithm," in IEEE Journal of Solid-State Circuits, vol. 23, no. 4, pp. 1010-1013, Aug. 1988, doi: 10.1109/4.353.

[3] C. Vonk. (2022). A FASTER MULTIPLIER CIRCUIT. Available: https://coertvonk.com/hw/building-math-circuits/faster-parameterized-multiplier-in-verilog-30774

[4] V. G. Alper Meric, Sameer Sonkusale. (2007). Post Layout Simulation. Available: https://www.seas.upenn.edu/ ese570/manual/manual_18.htm

Chenchen Ding Chenchen Ding is a year-2019 undergraduate student. He is involved in the study and research of AI Accelerator Design with Prof. Hao Yu's group.

Office: 13357655698

Email: dingcc2019@mail.sustech.edu.cn

He enjoys 50% contribution in this project

Shuo Feng Shuo Feng is a year-2019 undergraduate student. She is involved in the study and research of ICs design with Prof. Quan Pan's group.

Office: 18306665189

Email: fengs2019@mail.sustech.edu.cn

She enjoys 50% contribution in this project