64-BIT MEMORY SYSTEM DESIGN

I-Ting Chen*, Anurag Marwah†, Hard Patel‡, and William Xia§
ECE-GY 6473: Introduction to VLSI System Design, Group 5
Electrical and Computer Engineering, New York University
Email: *itc233@nyu.edu, †am8482@nyu.edu, ‡hap338@nyu.edu, §wx312@nyu.edu

***Abstract*—This document is the final report for Group 5. It introduces design concepts and optimizations of five components of SRAM memory system - registers, decoder, write-circuit, read-circuit, and SRAM cells. Besides, the layout of 64-bit SRAM is also included in this report. Finally, we show our comprehensive testing results to demonstrate the correct functionality of combined circuit.**

***Index Terms*—SRAM, Decoder, Register, Write-circuit, Read-circuit**

I. Introduction

This section introduces an overview of memory system operation. The 64 bit SRAM memory system implemented in this project, operates in either of 3 modes at any time. Read, Write or Hold.

- Read: The data input to Address Register propagates to row decoder at rising edge of CLK, which decodes and turns the correct WL line ON. Based on the data in each SRAM cell, one of the bit line of each bit start discharging which is amplified by sense amplifier and stabilized by latch. The stabilized signal goes into Read Data Register and stores the read word. Write Column circuit is cut off during this cycle.
- Write: One word address is given to Address Register which enables the correct WL line, just like in read cycle. The data in Write Data Register gets propagated to Write Driver Circuitry and overpowers the feedback of WL selected memory cells to store data in memory. Read Column Circuit is cut off during this cycle.
- Hold: None of the WL turns ON, rendering SRAM array disconnected from column circuits. The data stored in array is retained.

II. System Block Diagram

A. Delay Path

Fig. 1 shows the delay paths of read operation and write operation. We can see the delay of read operation is around 210ps, while the delay of write operation is around 128ps.

B. Critical Components

1) Read operation: For read operation, the read circuit cannot read the cell data until the WL signal arrives. Therefore, the total delay of read operation is the accumulation of register delay, SRAM delay, and read-circuit delay.

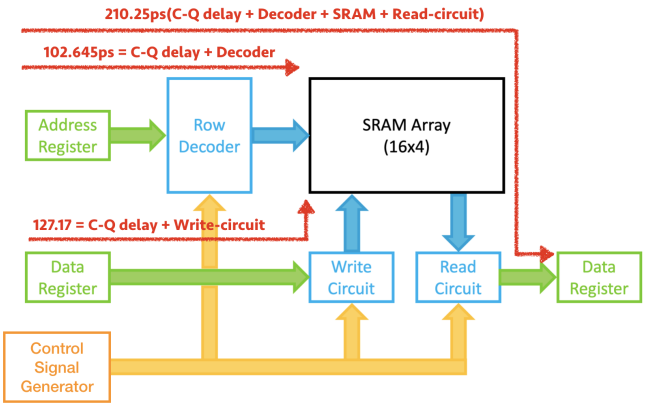

Fig. 1. Delay path of system block diagram

2) Write operation: On the other hand, decoder and write-circuit are working in parallel. Therefore, the delay of write operation is the maximum of these two components, which is smaller than the one of read operation. In our design, the write-circuit is the critical component of write operation.

III. Key Design Concept

A. Register

1) Low power consumption: Since flip-flop is a high power consuming device, it is not sized up unnecessarily. As a result, the PMOS and NMOS in the D Latch are of minimum balanced NAND2 gate size.

2) Balanced high-to-low delay and low-to-high delay: However, the testing result showed that this size did not provide a balanced C-Q delay, so the farthest NMOS is slightly sized up to achieve truly balanced C-Q delays. The size ratio of these two NMOS is $W_{near} : W_{far} = 180nm : 320nm$.

B. Decoder

1) NOR-NAND Predecoder Logic: The decoder in first stage uses static CMOS NOR-NAND predecoder implementation. The WL generated by this stage is active low. NAND-NOR implementation could have also been used which gives active high WL, but is avoided because NOR is not a good pull up and it adds to transistor area in synchronization stage.

2) 2T Dynamic AND gate: To convert active low WL to active high WL, a reverse dynamic inverter (pull up network evaluates logic) with common pull up node is used at output of each WL. Without common pull up node, each WL inverter takes 3 transistors, but with common pull up it takes only 2.2 transistors each. This allows reduction of 13 transistors in comparison to regular dynamic design, and 61 transistors (16 WL * 4 for each NAND gate = 64 transistors - 3 transistors of common pull up network) in comparison to regular CMOS implementation.

3) Low Power Design: Despite using dynamic logic, low power consumption of $31.8\mu W$ is achieved for ascending WL truth table simulation, due to lower transistor count and optimal transistor sizing.

4) WL PRE sync: By using an inverted PRE signal instead of CLK in the 2T Dynamic AND gate, it not only inverts the WL but also synchronizes it with the PRE signal. The resultant WL is active high with WL high time equal to CLK high time.

C. Control Signal Generation

1) PRE: PRE is a time delayed version of CLK which is turned ON only when memory is either reading or writing. The PRE is deliberately delayed by around 100ps to match the C-Q delay + Decoder delay. In essence, PRE is

$$PRE = CLK_{delayed} \cap (RE \cup WE)$$

2) SAE: SAE is pulled high when CLK is high, PRE is high and RE is high. SAE turns ON after some time of PRE turning ON, and turns OFF before PRE turns OFF to save power & maintain data integrity.

$$SAE = CLK \cap RE \cap PRE_{delayed}$$

D. SRAM

1) Cell Type: The design of the SRAM cell is that of a six-transistor memory cell. Fig. 2 shows the schematic implemented: two cross-coupled inverters with access transistors to provide means of reading and writing to the cell.

2) Noise Margins: Testing performed for the noise margins showed that reducing the sizing of the transistors resulted in the write margin falling below the specification of 35% before the read margin falling before its specification of 25%. Thus, the pull-up, pull-down, and access transistors were sized to the smallest possible values while maintaining the write noise margin. This sizing turned out to be $W_{Access} : W_{PullUp} : W_{PullDown} = 112.5nm : 90nm : 112.5nm$.

3) Area Minimization: Apart from different layers brought as closer together as possible, the critical optimization was done using multi-finger layout. As part of it, access and pull down transistors were merged to save area.

E. Write-circuit

1) Inverter chain before transmission gate: To drive the heavy bit line capacitance, large inverter chain were put before the transmission gate. Originally it was put after transmission gate, but it was found that large inverter chain would keep pushing the bit line even it was not in write cycle. To prevent this condition, we put the inverter chain before the transmission gate, and sized up the transmission gate. Therefore, the transmission gate will cut off the connection to bit lines in read cycle. The size ratio of this two-stage inverter is $W_{PMOS,first} : W_{PMOS,second} = 385nm : 1773nm$

F. Read-circuit

1) Precharge Circuit: The precharge circuit uses an active-low PRE signal to charge the bit lines prior to reading from SRAM. Precharge circuit utilizes a multi-finger layout to conserve area and symmetry of the array. In addition, by using the global RE signal as an input to the transmission gate, the precharge circuit is disconnected from bit lines when RE is OFF. Since precharging is only needed when reading from cell, this optimization prevents unnecessary power loss in charging and does not hinder write cycle.

2) Sense Amplifier: The sense amplifier is a type of differential amplifier used here to detect the voltage differential between two bit lines of the same memory cell. The sense amplifier is also followed by a level-sensitive latch which efficiently pulls the output of the sense amplifier either to logic '1' or logic '0'. The operation of the sense amplifier is controlled by the SAE signal. This approach serves to conserve power. The SAE signal remains high only for approximately 100ps.

IV. SRAM Cell

A. Schematic

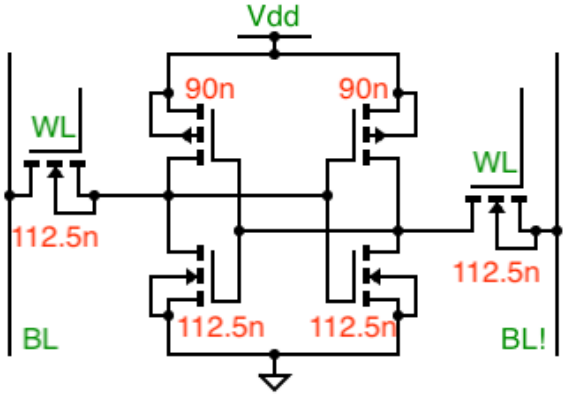

Fig. 2. Optimized SRAM Cell - Schematic

B. Single Cell Layout

Fig. 3 shows the layout of a single SRAM cell.

C. 16 x 4 Array Layout

Fig. 4 shows the layout of the complete SRAM array with 16 WL lines and 8 bitlines (BL and BL_BAR).

V. Key Design Properties

Table I shows all critical properties of five components.

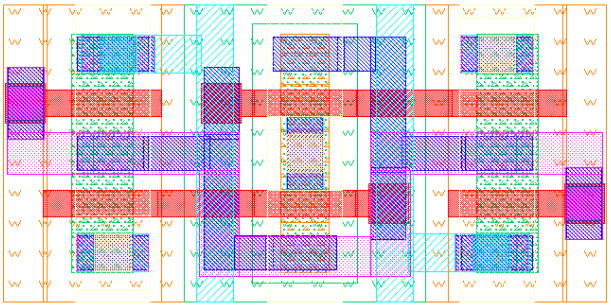

Fig. 3. Optimized SRAM Cell - Layout

TABLE I
KEY DESIGN PROPERTIES

Component	Specification	Value
Latch	C-Q Delay	28.5ps
	Set-Up Time	35ps
Decoder	Row Decoder Delay	53ps
SRAM Cell	Read Margin	26.64%
	Write Margin	35.20%
	Layout Area	0.6789μ
	Cell Access Time	11.53ps
Sense-amplifier	Worst Case Delay	38.18ps
	Offset Voltage	13mV
Write-circuit	Discharge bit line	71.81ps

VI. WAVEFORM FOR READ WRITE OPERATION

The waveform in Fig. 5 shows an alternating pattern for writing to, then reading from, the cell. Note that the SAE and PRE signals are omitted for clarity. W0 is an input to the write circuit and is written to the cell once WE goes high. OUT is the output at the sense amplifier latch (marks the delay of the read circuit), passed to the input of the read data register. This data is passed at the start of the following clock cycle as R0.

VII. TIMING

A. Read Delay

The read delay is taken from the rising edge of CLK to the change in the output of the sense amplifier latch. This delay 210.25ps, and includes the row circuit delay, sense amplifier delay, and latch delay.

B. Write Delay

The write delay is from the rising edge of CLK to the signal stored in the cell. The measured delay is 127.17ps, which includes C-Q delay and write-circuit delay.

C. Maximum Clock Frequency

The maximum operating frequency of this circuit is 2.86 GHz. This clock frequency is limited by the pulse width of SAE. Due to signal generation and power conservation methods, the SAE pulse width decreases as the CLK frequency increases. The sense amplifier is thus the limiting factor of this particular design.

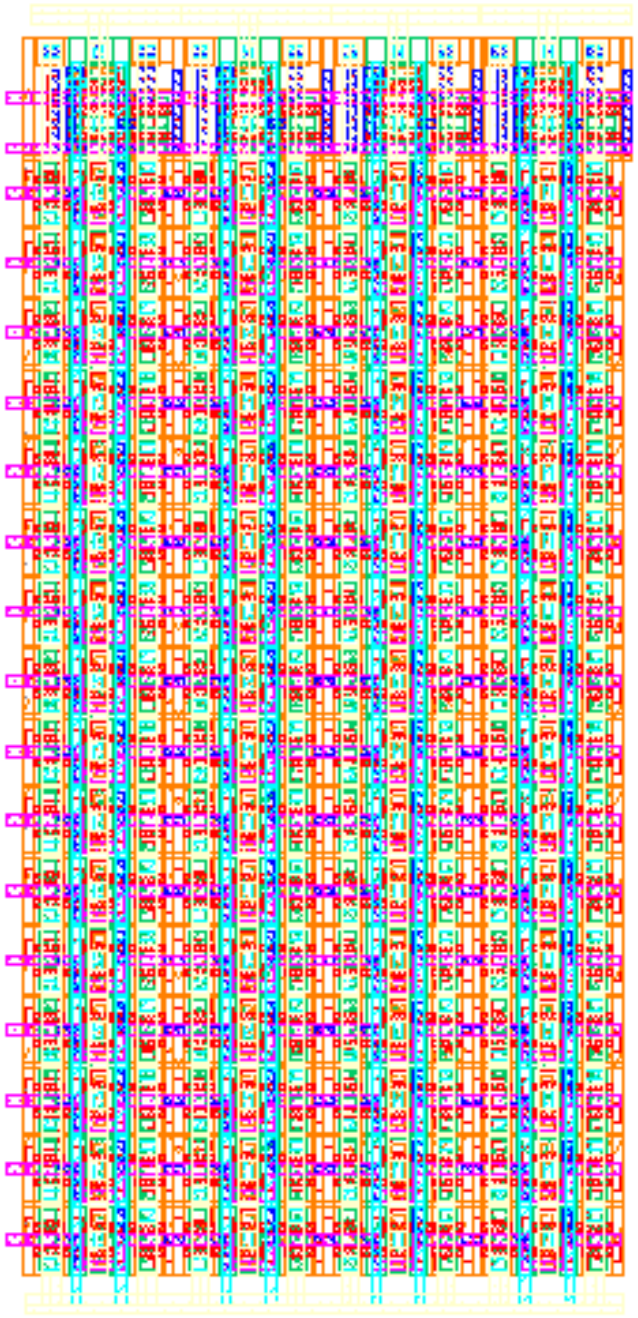

Fig. 4. SRAM Array - Layout

VIII. CONCLUSION

The design of this 64-bit SRAM memory system made clear the challenges of optimizing a complex circuit for multiple constraints. Although the implemented system meets all required specifications, power efficiency and area usage was favored over robustness of the circuit. This is clear from the signal generation and layout optimization.

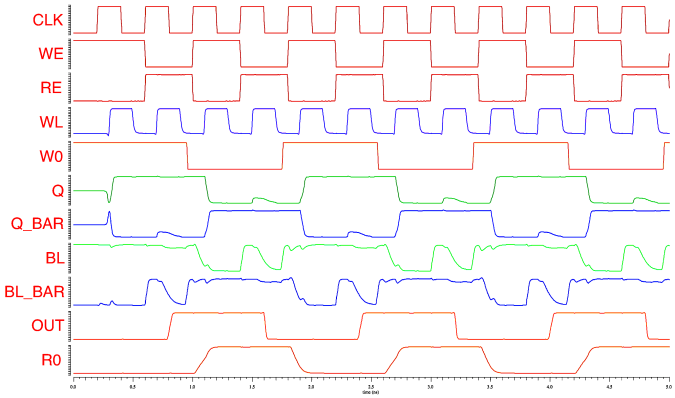

Fig. 5. Read and Write Operation - Waveform

IX. Cadence Directory

/projects/fall18/group5/freepdk45_fall18/Part3_submission is the Cadence directory which contains the SRAM memory system.