64bit Memory System Design

Anthony Fisher, Nikunj Rajput, Vaishnavi Nandedkar, Yan Zhu

***Abstract*— Static Random Access Memory (SRAM) is a memory that uses bistable latching concept to store each bit. It is a component that must meet the requirements of working in Nano-technology ranges with stable performance and operation characteristics. The paper concentrates on techniques implemented for the designing of 6T SRAM. This involves decoders, SRAM cell, sense amplifier, data register and address registers using Cadence tool with 45nm CMOS technology. In this paper, the cell area of $0.580\mu m^2$ is achieved using circuit techniques and transistor scaling which gives low power dissipation and very less delay at a frequency of 1.89Ghz.**

***Index Terms*— 6T SRAM, 45 nm technology (FreePDK45), Cadence Virtuoso, Sense Amplifier, Noise Margin, Power**

I. Introduction

CMOS designs are implemented to achieve high speed, performance, and low power consumption. SRAM is an important memory device that has high read-write speed with meta-stability. The project implements 4*16 decoder with 4-bit memory cell CMOS technology. All the schematics have been designed in Cadence Virtuoso using FreePDK45 library with voltage restriction of 800mV. The objective of our project is to achieve target frequency of 1.5GHz and SRAM cell design of 0.8sq-µm with low power consumption. The main concentration of the design is to create low power and high performance memory using 6T SRAM that involves transistor scaling. The sizing should be done in an efficient manner to attain high density in memory design.

II. Design Overview

The high-level block diagram of 6T SRAM array with peripheral circuits is shown in Figure 1. The design is divided into four main blocks : 1. Registers 2. Decoder 3. SRAM Cell 4. Read, Write and PRE-charge circuitry.

A. Decoder and Registers

Registers used for storing address, input data and output data are each made up of four master-slave flip-flops. All of the transistors in address register are sized to minimize the clock-to-Q delay, thereby reducing the delay before the word-line goes high. Sizing of all the transistors in the registers is given in Table 4.

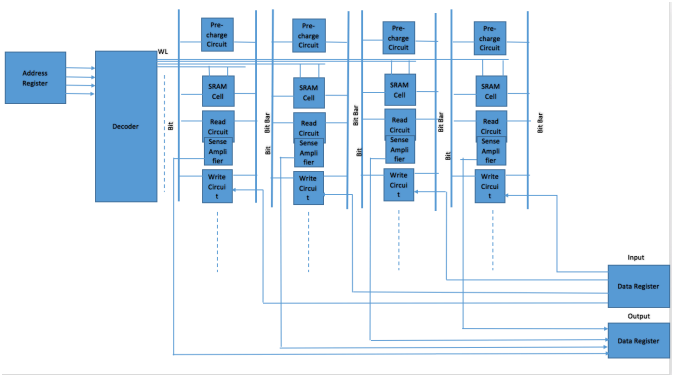

Figure 1. : SRAM array

Implementation of 4 to 16 decoder was done in two stages. One-stage decoder had larger fan in than the two stage decode, which made it slower as compared to the two stage decoder. First stage in the decoding is pre-decoding which was implemented using NAND gates. The second stage is post-decoding which was implemented using NOR gates.

Output of first stage of the decoder is all the possible combinations of the first two input bits and the possible combinations of the second two input bits. The second decoder stage gives the final decoder output based on the input combinations obtained from first stage. The implemented decoder uses AND2 gate with a buffered clock signal to keep the word-line zero initially, prevent glitches and also to synchronize the word-line with the clock signal. The 16 word-lines obtained from the decoder are then connected to the SRAM cell as shown in Figure 1. Sizing of all the transistors in the decoder is given in Table 4. The decoder transistors were sized using the logical effort method. Based on the fact that only one word-line is high at a time, to reduce the area of our design we have used a single NMOS transistor for the buffered clock input as a part of pull down network that is shared between all the 16 word lines.

B. 6T SRAM Design

6T SRAM implementation included sizing of the transistors based on write and read margins. The given cell is designed to achieve a read margin greater than 200mV, write margin greater than 280mV and total access time less than 333ps. The sizing of transistors is done using butterfly model. To calculate the read margin, VTC curve of schematic one (figure 2) with properly sized transistors is plotted in butterfly curve using MATLAB. Length of the maximum sized square that can fit in the curve, is the read margin for that sized SRAM.

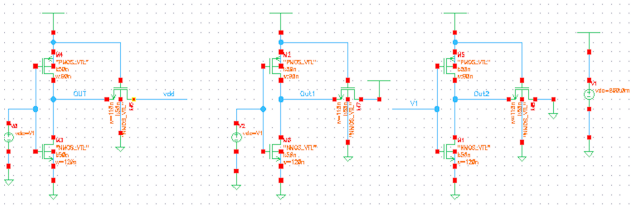

Figure 2. : (1)Read analysis schematic (2)Write analysis schematic

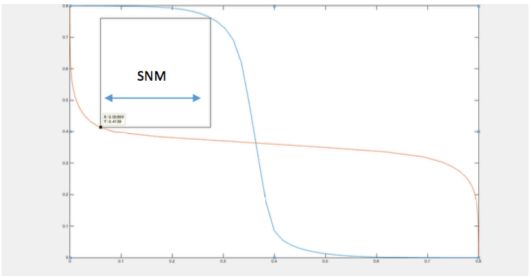

Figure 3. : Read margin

Similarly, the write margin is calculated by using the second schematic in figure 2. The sizes obtained using this method are verified using the conventional method. Read margin obtained using the final sizes is 213.3mV, write margin is 292.53mV. Access time of the cell is calculated using the following formula:

$$T_{access} = (C_{bit} * \Delta V)/(I_{read})$$

The bit-line capacitance is taken as 40fF. The bit-differential obtained for our sense-amplifier is 32mV. I_{read} for the given size is $5.26 * 10^{-05}$A. The access time was found to be 31.5ps. The final sizing of all the transistors in 6T SRAM is given in Table 4.

C. PRE-charge, Read and Write circuitry

A five PMOS transistors based design was adopted for the PRE-charge circuitry. PRE-charge circuit uses four PMOS transistors to charge the bit-line up to Vdd. The fifth PMOS transistor is used to avoid mis-match between the bit-line voltages in PRE-charge phase.

The read operation is done depending on the charging and discharging of the bit-lines based on the value stored in the SRAM. To discharge the bit-lines down to zero increases the delay of the design. Hence sense amplifier is used to increase the speed as well as reduce the power of the design.

Designed a "current mode latch sense amplifier" for our design as shown in figure 4. This sense amplifier uses pair of cross-coupled inverters to provide regenerative feedback. This feedback helps the circuit to give output quickly based on a small difference between the bit-lines thereby increasing speed of the design. With a 10% mismatch in the sizes of input transistors the bit differential obtained for our design is 32mV.

The write circuit is designed using two transmission gates and data drivers. The data driver is implemented using data buffers which are suitably sized to drive required data. The data to be written is given to one transmission gate and inverse of the data is given to another transmission gate. Based on the Write signal and the data to be written, one of the bit-lines discharges. The sizing of SRAM transistors is done such as based on the bit-line values the feedback loop in the SRAM is

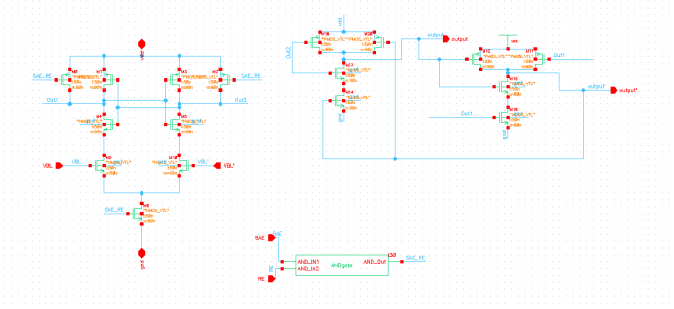

Figure 4. : Sense Amplifier

overpowered and new data is written inside the SRAM.

To prevent the write drivers being connected to the bit-lines when we are pre-charging them, we created an internal Write signal based on the global Write Enable and the clock. In this way, we open the write transmission gates when the clock goes high and close them again when the word-line goes low and pre-charging begins.

The SAE signal is generated such that the sense amplifier is enabled a short time after the word-line goes high. This allows time for a difference in voltage to be established between the bit-lines before sensing. The timing of these control signals can be seen in Figure 6.

D. Layout of 6T SRAM cell

We implemented two layout designs for SRAM cell to compare the area. We used our optimized transistor sizes of W_p = 90nm, W_{access} = 110nm, $W_{pulldown}$= 120nm. Option 1 with two ground rails, can be overlapped with neighboring cells both vertically and horizontally. The area of Option 1 is 0.602sq-µm. Option 2 with single ground rail, can be overlapped with neighboring cells vertically, but not horizontally. The area of Option 2 is 0.580sq-µm. Used option1 layout for the full array design.

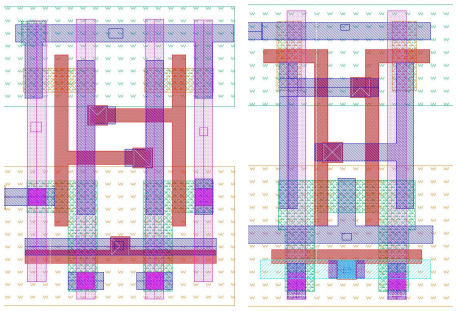

Figure 5. :Layout option 1(Left) Layout option 2(right)

III. SRAM INTEGRATION AND RESULTS

The test measurements were done with T_clock = 333ps, V_{supply} = 0.8V. The simulation results obtained for reading and writing of SRAM based on word-line 0 and word-line 7 are shown in figure 6.

Figure 6 shows a series of read and write operations. First, data of 1010 is written into word 0 (address 0000). This word is read in the following clock cycle and the output changes accordingly. In the third clock cycle, data of 0101 is written to word 7 (address 0111). In the fourth clock cycle, nothing is

done, neither read nor write. In the fifth and sixth clock cycles, the data is read from word 7 and word 0 respectively. The output transitions appropriately.

The layout for the complete array is shown in figure 7. Based on the layout the area of complete array is 43.95sq-um. Area of PRE-charge circuitry is 1.83sq-um.

The area efficiency is calculated using following formula :

$$Area\ efficiency = \frac{SRAM\ area}{Total\ area}$$

Area efficiency for our design is 0.96.

Timing of all the building blocks is given in Table 1. Based on Cadence simulation, the maximum frequency of the design is 1.89GHz. The bottleneck for the maximum frequency of the design appears to be pre-charging the bit-lines. At high frequency, following a write operation the bit-lines are not able to be charged back up to V_{DD} or equalized before the next clock cycle. If we needed to increase the frequency, increasing the size of the pre-charge transistors may help. Another approach may be to use dynamic logic in the row decoder to reduce the word-line delay and also synchronize the falling edge of the word line with the falling edge of the clock. This would allow more time for pre-charging.

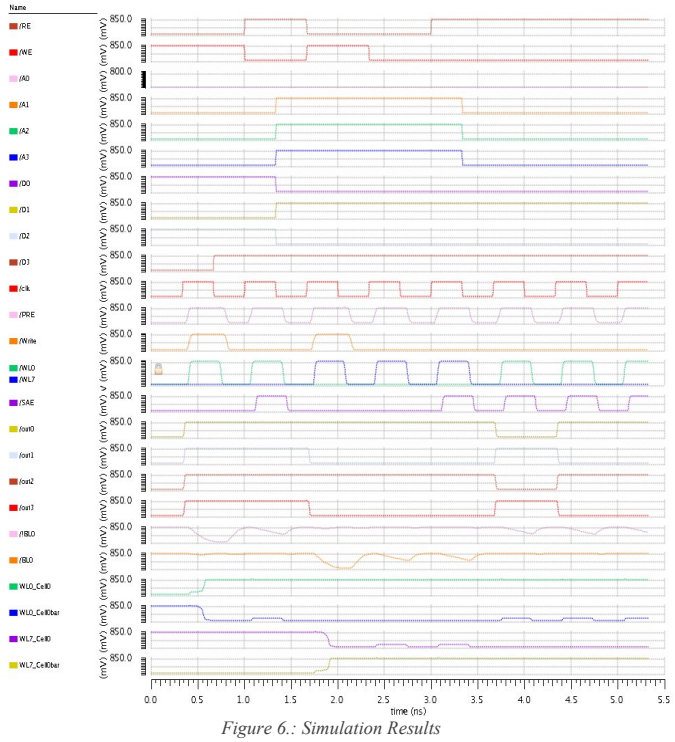

Figure 6.: Simulation Results

The lowest supply voltage at which our design can work is 695mV, after which the SRAM reading and writing fails. All the results for SRAM array are summarized in Table 3.

Decoder	
CLK to word-line rise delay	72.6856ps
CLK to word-line fall delay	84.0392ps
Registers	
Clock to Q(address)	13.9ps (rising Q), 20.5ps (falling Q)
Clock to Q(address)	19.0ps (rising Q), 25.6ps (falling Q)
SRAM array	
CLK to WL	71.3ps
Cell access time	31.7ps
Flip-flop clock-to-Q	19ps
Sense Amplifier Delay	54.3ps
Write Circuit(BL 90% to 10%)	199.7ps
Total read access time (clk to sense amp output)	191.9ps
Total write delay(clk to cell trip)	222.2ps

Table 1. : Timing

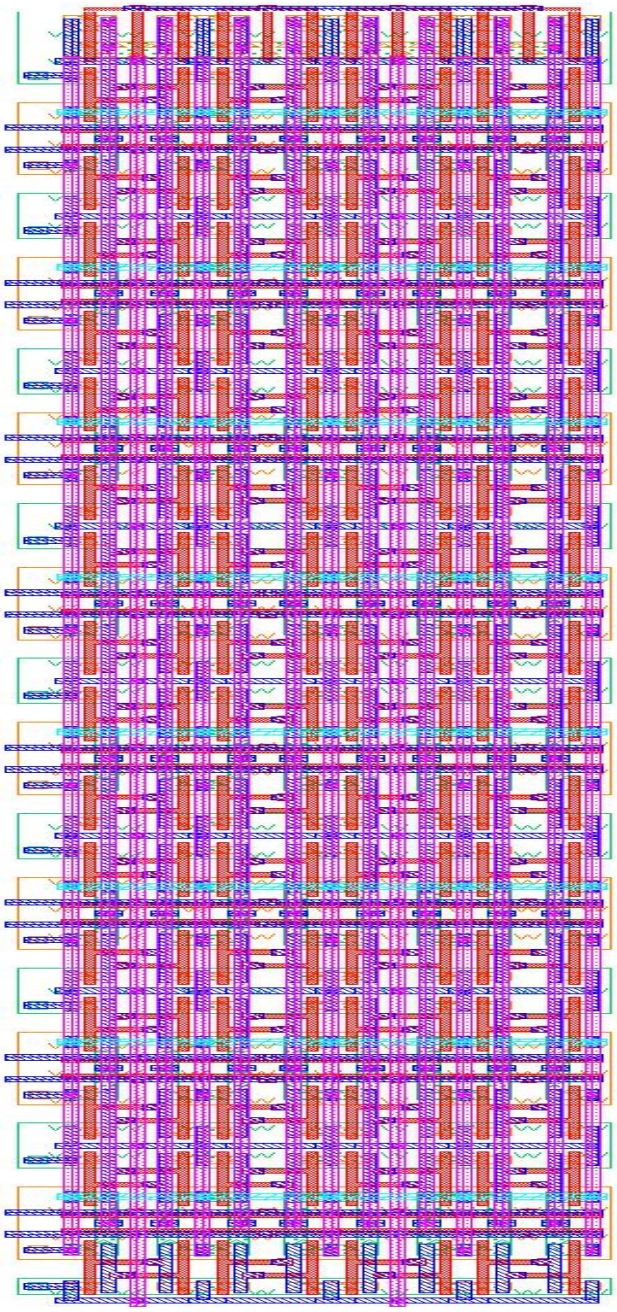

Figure 7. : Complete array layout

Decoding logic for one clock cycle	-22.3 μW
Address Register	9.95 µW
Data Register	7.302 µW

Table 2. : Average power dissipation

Array area	43.95 μm^2
PRE-charge Area	1.83 μm^2
SRAM area	0.602 μm^2
Max frequency	1.89 GHz
Min source voltage	695 mV

Table 3. : Array Results

Table 2 shows the average power dissipation of different building blocks. Power dissipation of decoder was calculated over one clock cycle using the calculator tool in cadence.

SRAM	
Pre- charge PMOS	L = 50nm, W = 180nm
Pull up transistors	L = 50nm, W = 90nm
Pull down transistors	L = 50nm, W = 120nm
Access transistors	L = 50nm, W = 110nm
Address Register	
Inverter in master stage of Flip Flop(PMOS)	L = 50nm, W = 242.5nm
Inverter in master stage of Flip Flop(NMOS)	L = 50nm, W = 152.5nm
Inverter in slave stage of Flip Flop(PMOS)	L = 50nm, W = 412.5nm
Inverter in slave stage of Flip Flop(NMOS)	L = 50nm, W = 260nm
Flip Flop transmission gate M1/M3(PMOS)	L = 50nm, W = 225nm
Flip Flop transmission gate M1/M3(NMOS)	L = 50nm, W = 225nm
Sense amplifier	
NMOS connected to VBL	L = 50nm, W = 99nm
NOT gate sizing to minimize word line delay	
Pull up transistors	L = 50nm, W = 1,195nm
Pull down transistors	L = 50nm, W = 775nm
Decoding (Buffer Logic)	
Pull up transistors	L = 50nm, W = 400nm
Pull down transistors	L = 50nm, W = 252.2nm

Table 4. : Sizing

For the registers, average power measurements were made by taking the average of the product of the current from V_{DD} and the voltage at V_{DD}, using the Cadence calculator. The measurement was made using an 11ns simulation in which the input data was switching with every clock cycle.

Path for the compete circuit is:
/home/FALL2016/apf312/cadence/AF_Project/final.

IV. Participation

Anthony Fisher: Worked on Decoder logic, Registers, Full array layout, Report.
Nikunj Rajput: Worked on Write circuitry, SRAM cell Layout, Report.
Vaishnavi Nandedkar: Worked on Read circuitry, Sense amplifier, Sizing of SRAM, Report.
Yan Zhu: Worked on integration Peripheral circuitry.

V. Conclusion

This paper presents design of 16 by 4 SRAM array in 45nm technology. The SRAM cell as well as the peripheral circuitry was designed to have minimum area and minimum access time. Implemented the memory to have low power dissipation. Simulation results for read, write functions, obtained using Cadence Virtuoso, show the exact functioning of SRAM.

Layout for the design was implemented to reduce the memory area by reusing the components of the layout. Obtained area efficiency of 96%.