64x32 SRAM in 90nm CMOS Technology

ENGR 848

Ali Attaran

Table of Contents

1 Introduction

Every digital system uses memory to store information. Semiconductor memory arrays are capable of storing large quantities of digital information which is required by digital systems. Static Random Access Memory (SRAM) is widely used in processor cache for data storage. The goal of the project is used to create fast, compact, reliable and low power consuming SRAM.

Design constraints: Minimum read SNM of 100 mVolts and maximum total power of 500 µWatts (leakage plus dynamic) both at nominal supply voltage of 1.2 Volts and worst case temperature of 110ºC.

1.1 Top level block diagram of the design showing inputs and outputs

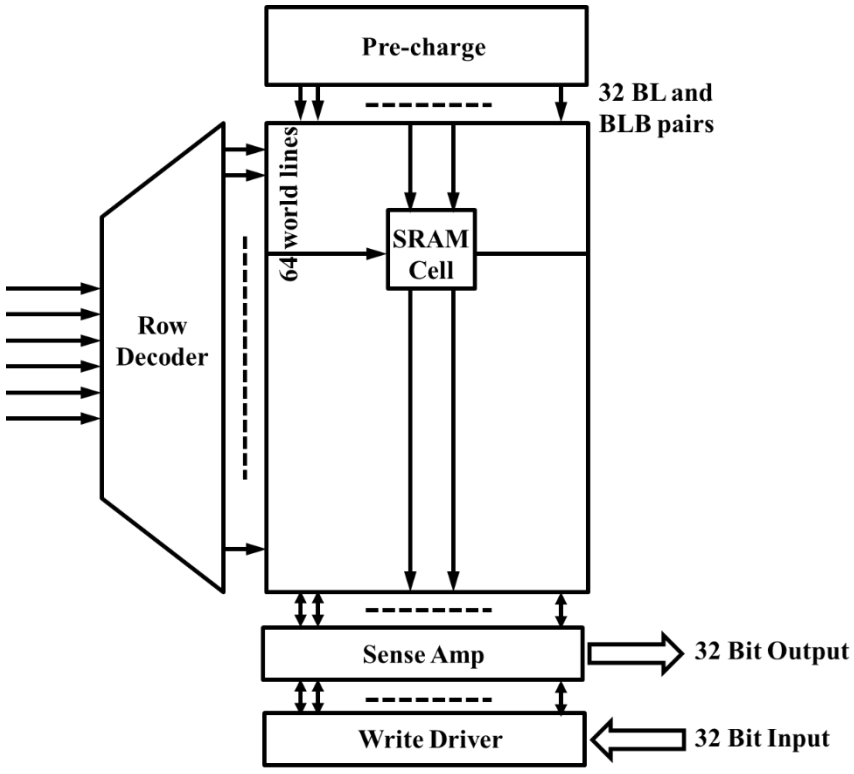


Figure 1. 64x32 SRAM design

1.2 Design specifications and read and write timing waveforms

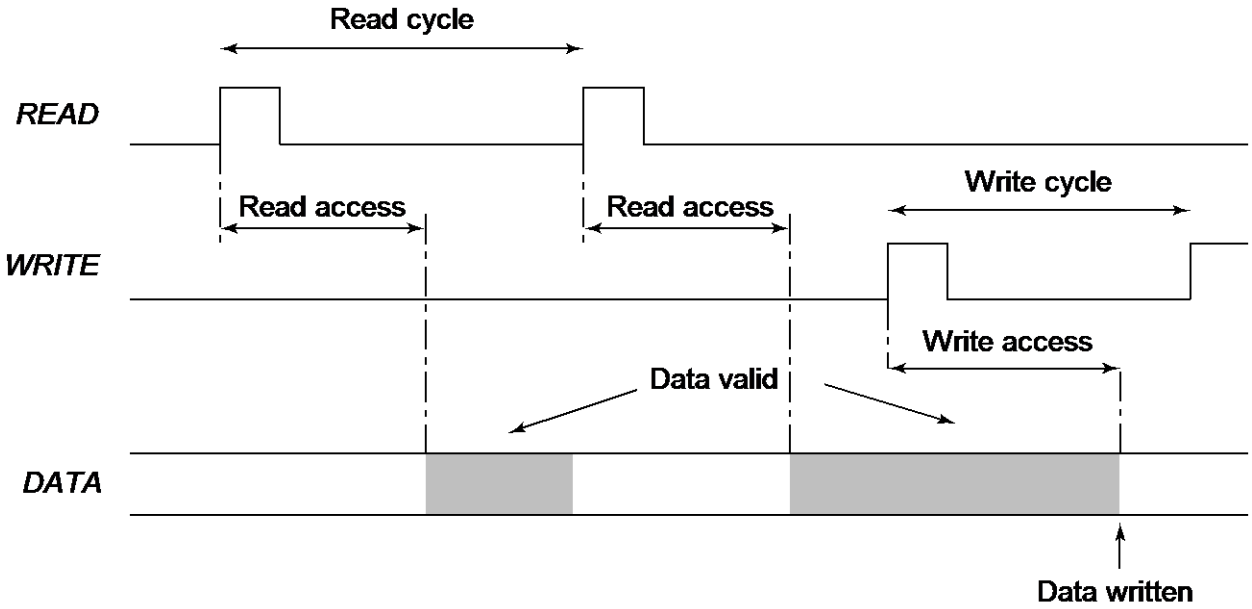


Figure 2. Timing scheme

2 Design partitioning

The Memory partitions were designed in different stages. The partition blocks are SRAM array, Row Decoder, Sense Amplifier, Controller, Precharge, write driver circuits and later integrated to build 64x32 SRAM.

2.1 SRAM Cell

Static random access memory (SRAM) is a type of volatile semiconductor memory meaning, it stores data as long as it is powered. SRAM uses bi-stable latching circuitry made of Transistors to store each bit. Unlike Dynamic RAM (DRAM), SRAM doesn't have a capacitor to store the data hence, SRAM works without refreshing. SRAM is often used as a memory cache. The most commonly used SRAM cell consists of 6 transistors and this configuration is called 6T Memory Cell. It consists of two cross-coupled inverters and two access transistors. The access transistors are connected to the word line (WL) at their respective gate terminals, and the bit lines (BL and BLbar) at their source/drain terminals. The word line is used to select the cell while the bit lines are used to perform read or write operations on the cell.

The access and pullup transistors were sized 0.24um and the pullup transistors were sized to 0.4um to maintain design requirements of SNM=150mV at operating temperature of 110 degree Celsius.

SRAM stand by power is estimated by providing VBL=VBLB=VDD=1.2V, VWL=0V. Standby power of single cell is estimated as 296nW.

SRAM cell layout: Metal 5 line was used for bitline and bitline bar Hence Metal 5 pin is used to label BL and BLB. Via 1,2,3 and 4 are used to connect metal 1 and metal 5.

Length of SRAM cell = 1.7

Height of SRAM cell = 3

Area of SRAM cell = 5.1

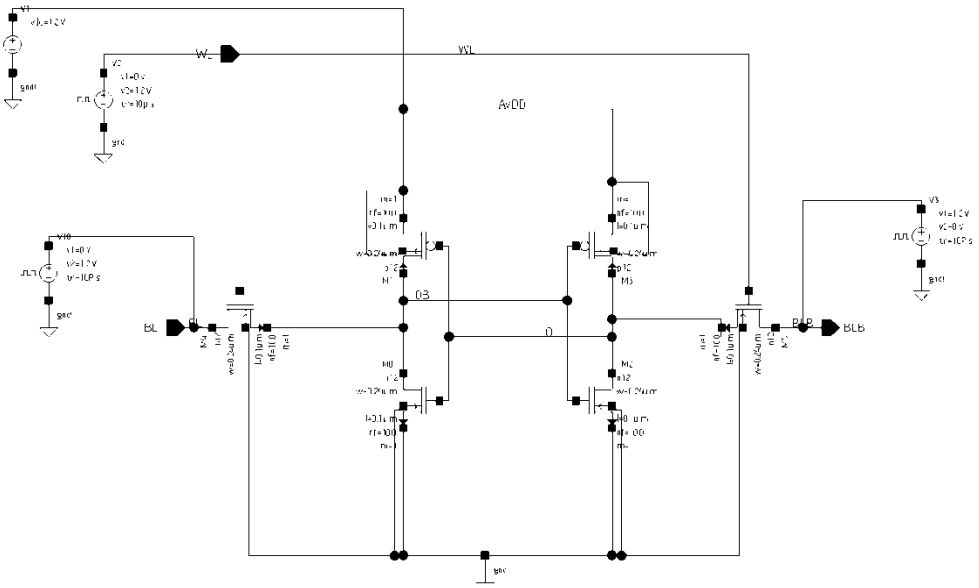


Figure 3. SRAM cell

Wn=0.4u,Wax=0.24u, wp=0.24u

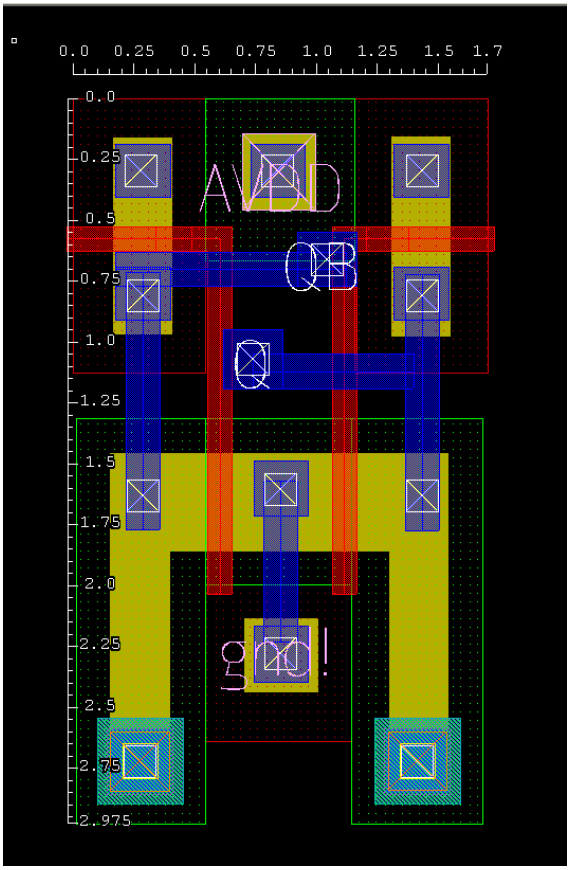
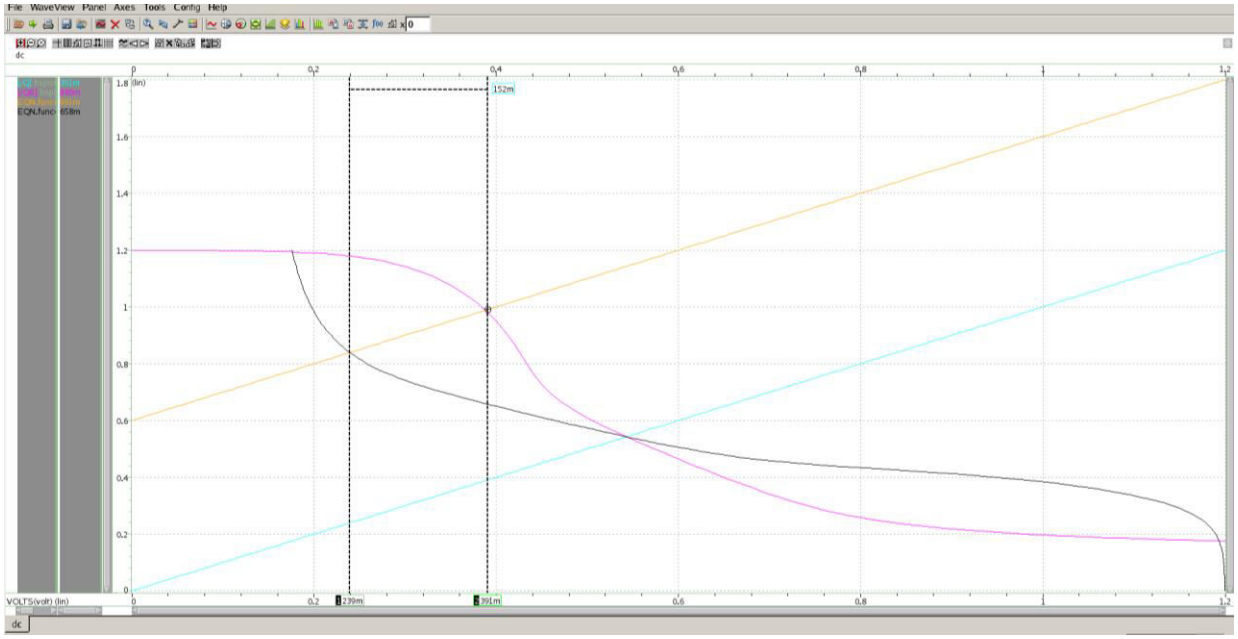


Figure 4.SRAM cell layout

The noise margin is calculated by drawing the overlapped VTC (Butterfly diagram) for the cross-coupled inverters that form the memory cell. The largest square that can fit in the eyes of the butterfly diagram determines the noise margin. The butterfly diagram obtained is shown below.



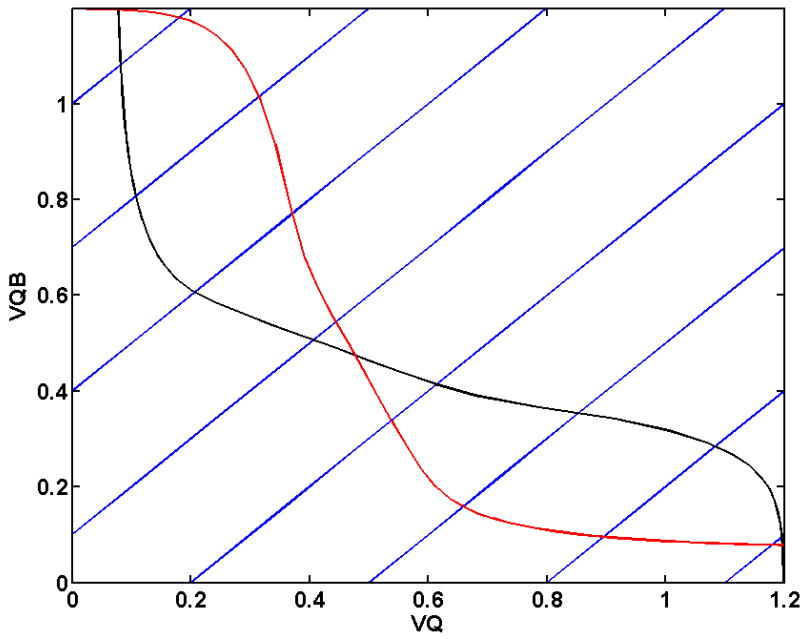


Figure 5. SRAM SNM butterfly figure

Figure 6. SRAM time domain simulation

Post-layout simulation:
Delay: 96.6 ps
Dynamic power:
  ptot= 658.3809n
Static power:
ptot= 3.6840n

2.2 Pre-charge

In both read and write operations, the bitlines are initially pulled up to high voltage.

This is done using a precharge circuit. The schematic of the circuit is as shown in Figure 9 below. A clock input is applied to the two pull-up transistors, called the balance transistors, connected between the two bitlines. When the wordline (WL) signal goes high, one bitline remains high and the other falls until WL goes low. The layout of the precharge circuit is as shown in Figure 8.

Precharge circuitry is used to precharge the bitline and bitline bar to predefined voltages. 3 PMOS transistors are used and their sizing is adjusted such that precharge delay is less than decoder delay. If the precharge circuitry is weak, bitline and bitline bar voltages don't reach full VDD value. Hence precharge size was sized to be 1 µm.

When precharge signal is off, the circuitry is turned off and it's in evaluation mode. Precharge width is limited by write driver width (3.1 µm). Here size of precharge width is 2.8 µm which is less than 3.1 µm.

Area = 3.7 * 3.4 = 12.58

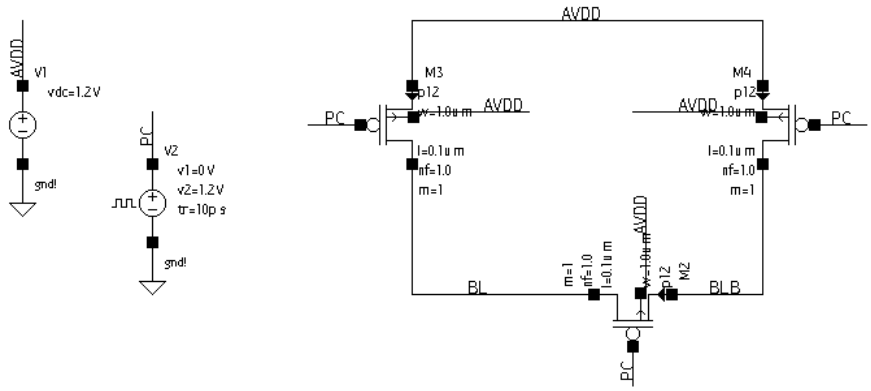


Figure 7. Pre-charge schematic

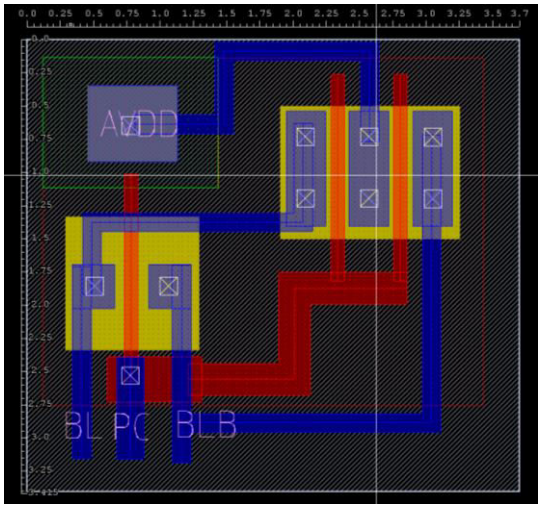


Figure 8. Prechare cell layout

32 cells of prechareg circuits:

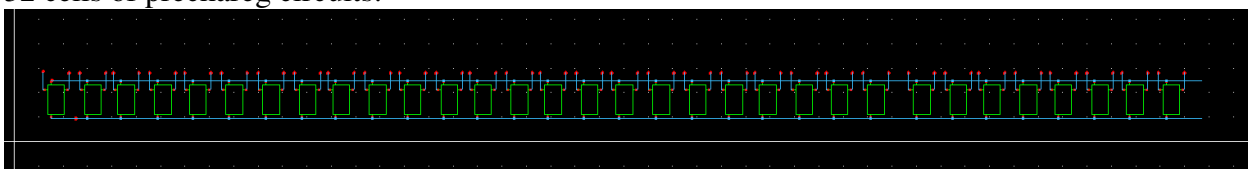


Figure 9. 32 precharge cells

Each cell has 3 PMOS. So it has 96 transistors.

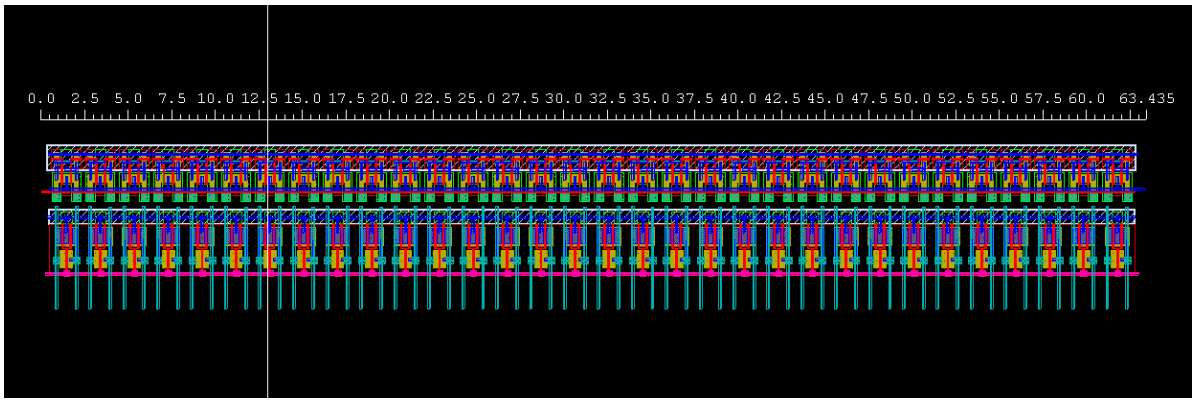


Figure 10. Layout of 32 precharge cells

2.3 Decoder

2.3.1 3 to 8 Decoder

Access time and power consumption of memories may be largely determined by decoder design. Row decoders take an n-bit address and produce 2^n outputs. Row decoders are used to select the required row in the memory array. The required wordline is activated based on the address given to the decoder. In our design we have 64 rows, hence n=5 address bits are used to select a row. Since the row decoder is used to activate one of the 64 wordlines, it has to be sized suitably using logical effort based on the capacitance of the wordline. The gate level schematic of one stage of row decoder is as shown in Figure.

Decoder is used to select one SRAM cell in an array of SRAM cells. Decoding is required in applications such as memory address decoding, seven segment display and data multiplexing. Binary decoder is made up of n inputs and 2n outputs. Only one output is active at any given time, corresponding to the input value.

CMOS Decoder can be designed with various logic like static, dynamic and domino. Since we wanted the SRAM to consume less power, CMOS Static decoder is used in the project.

Decoding is done in 2 stages: First stage consists of 3:8 Pre Decoder and the second stage is built using 6:64 decoder.

3:8 Pre Decoder is set up first using 3 Inverters and 8 Input NAND Gate Static Logic. By using 2 stages of pre-decoder, 6:64 decoder is constructed.

Delay= 164Ps

Decoder Area = 146.6*35 = 5131

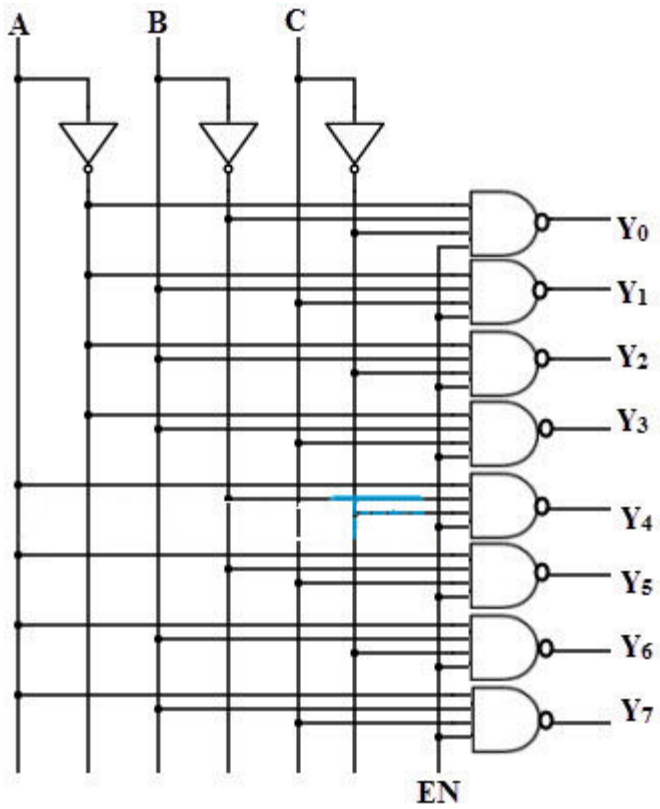


Figure 11. 3:8 Pre Decoder

2.3.2 6 to 64 decoder

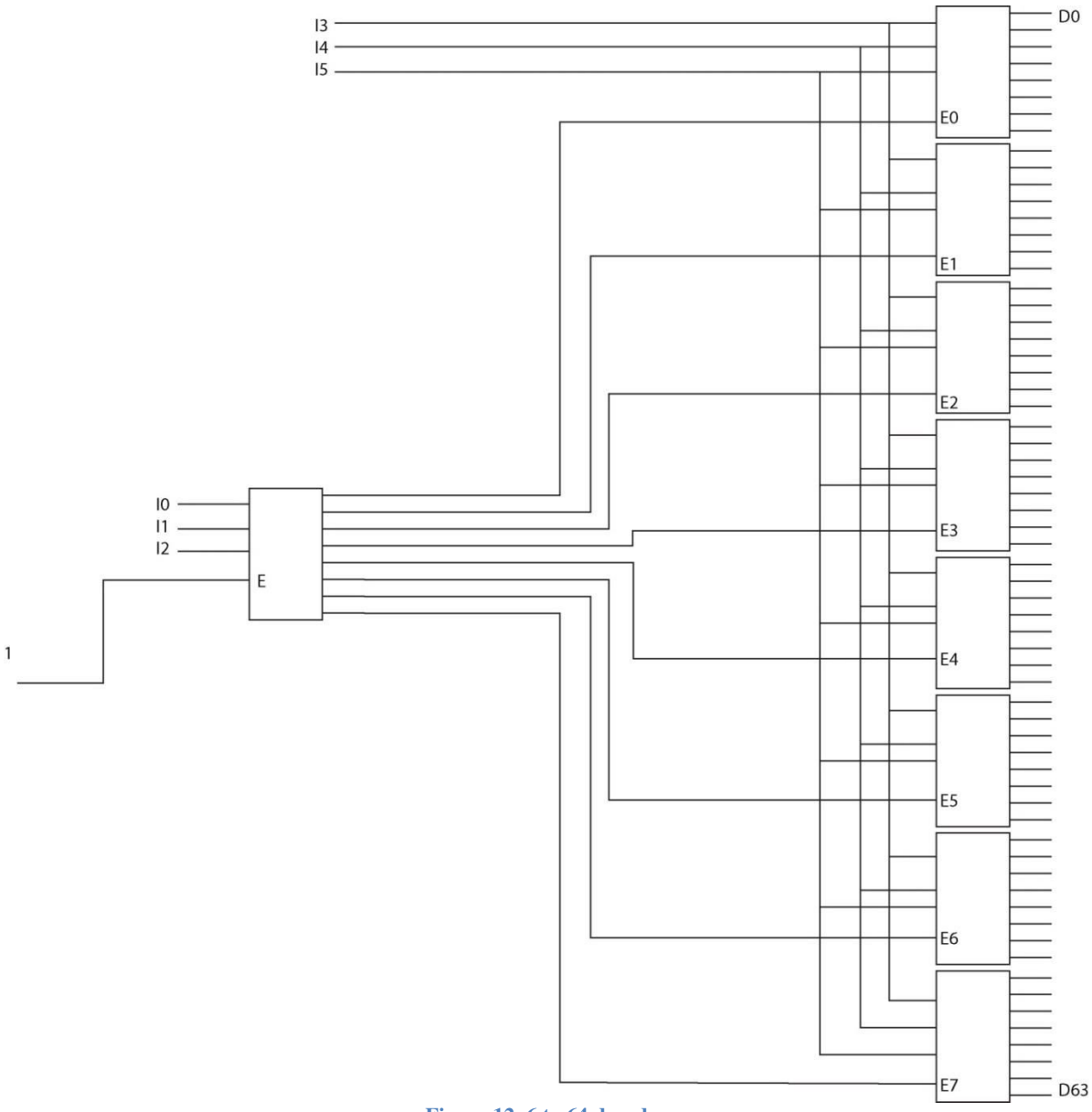


Figure 12. 6 to 64 decoder

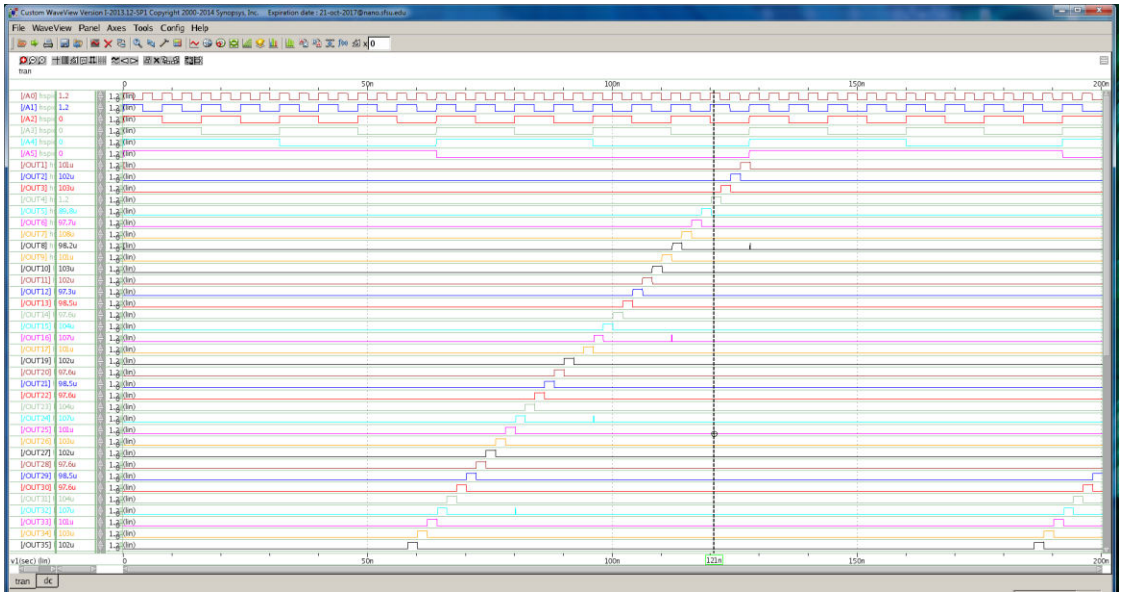


Figure 13. 6 to 64 decoder simulation

Post-layout simulation

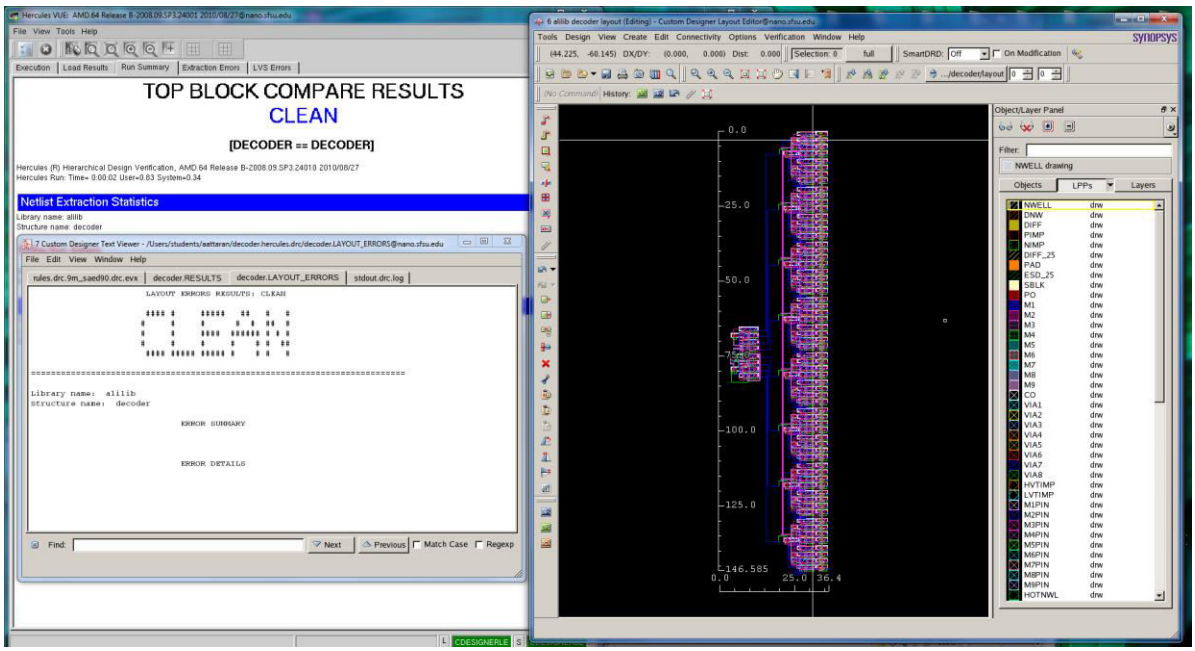


Figure 14. Post-layout simulation

Delay: 164Ps
Dynamic power:
ptot= 685.2530u
Static power:
ptot= 962.0136u

2.4 Sense Amplifier

The designed SRAM uses ten identical sense amplifiers to provide simultaneous output of ten data bits. In our design, we have used a current mode differential input single ended sense amplifier in order to attenuate the common mode noise and amplify the differential mode signals. The main reason for using this type of sense amplifier is to improve the noise immunity and speed of the read circuit. The differential signal that changes between the two bit lines during read operation is amplified by the differential pair current mode sense amplifier. The transistors are sized such that the differential voltage is amplified suitably for read operation. The output of the sense amplifier is then given to a pair of inverters in order to have a digital output. Inverted Write (WR) signal is given to the gate of the current source transistor in order to enable the sense amplifier only during read operation. The schematic and layout of the sense amplifier are as shown in the figures below.

Sense amplifier is one among the most important circuits in the periphery of CMOS memories. They affect both memory access time and overall memory power dissipation predominantly by their performance. As memory capacity is increased, it increases the bit line capacitances. This increased bit line capacitances in turn slows down voltage sensing resulting in slow and high energy consuming memories due to high energy bit line voltage swings.

Types of Sense Amplifiers:

1. Voltage Sensing Amplifiers
2. Current Sensing Amplifiers

In our design we preferred to use Current Sensing Amplifiers, considering the following points:

Significant reductions in bit-line voltage swing and major reduction in sensing delays and Lower dynamic power consumption and increased sensing speed

The key to these advantages lies in low input resistance of Current Sensing Amplifiers.

Sense Amp Delay=96.6 ps

Area = 3.3*4.6

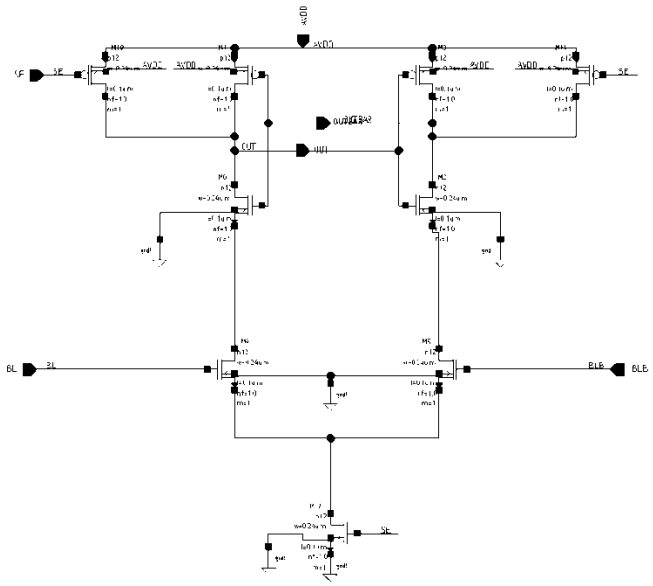


Figure 15. Sense amplifier schematic

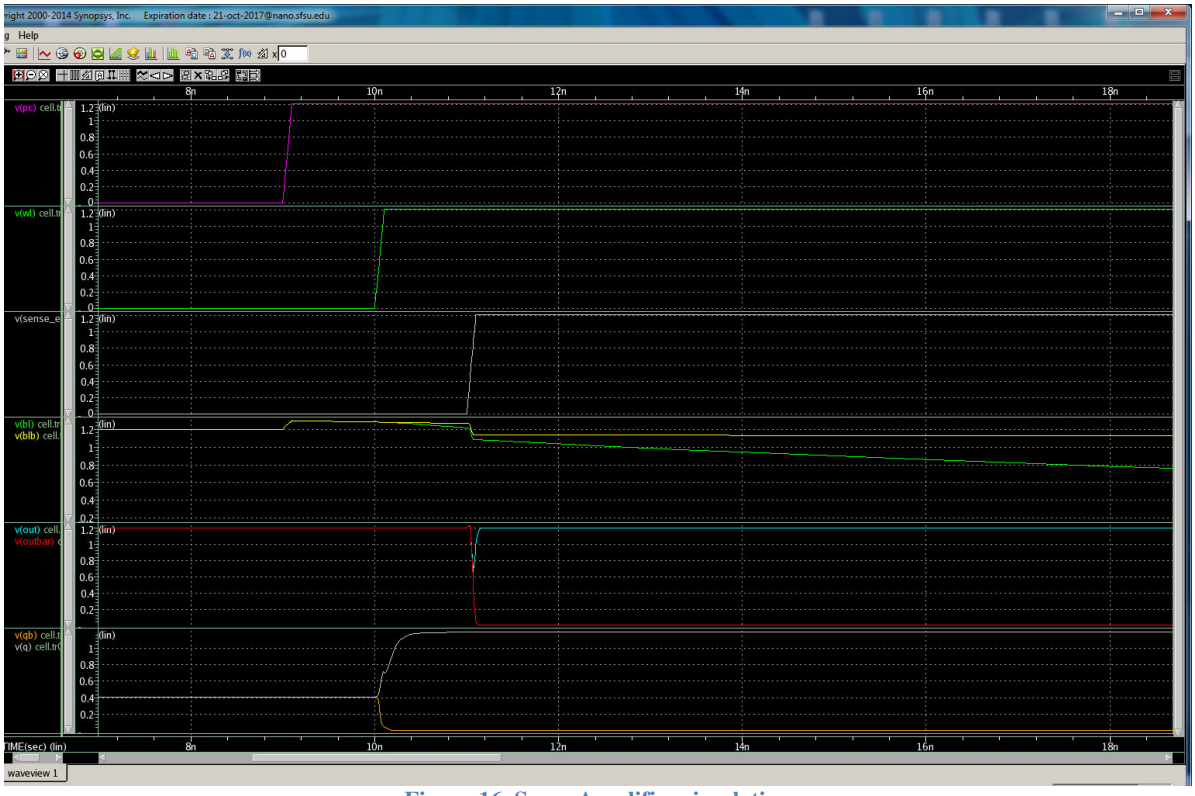


Figure 16. Sense Amplifier simulation

Post-layout simulation

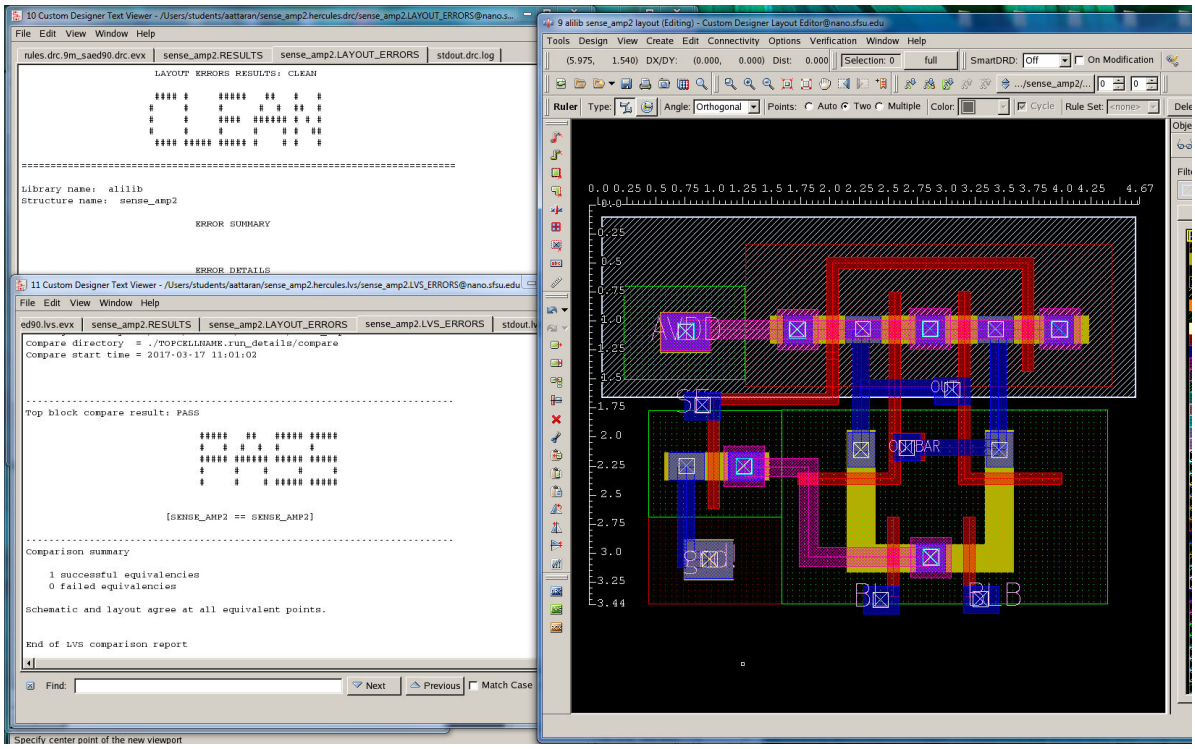


Figure 17. Sense amplifier layout

Delay: 96.6 ps
Dynamic power:

ptot= 658.3809n
Static power:
ptot= 3.6840n

2.5 Write Driver

During precharge both the BL and BLbar lines are charged to VDD. Before write operation, one of the bitlines must be driven high and the other low based on the data bit that is being written. The schematic of the write circuitry that we have used in our design is shown in Figure 19. During write operation, WR signal goes high and the 8 bit data can be written by giving required bit values to the corresponding input bits. These values are then passed through a set of pass transistors that are attached to the BL and BLbar lines so that the data bit will be written into the corresponding memory cell.

WD is designed using a tri-state buffer. The transistors were designed to achieve a delay less than the delay of decoder. In our design the decoder delay was 164 ps. The write driver has a delay of 96ps, when loaded with capacitance value of 64 (no of rows) nMOS transistors.

Write Driver Working:

When Enable is powered by a DC voltage and DIN is clocked high the bit line value goes low and it performs read operation of the memory cell. When DIN turns low the BL turns high resulting in writing data to the memory cell.

Write driver layout has a width of 6.2 and height of 3.6 micrometer. Each cell of write driver occupies an area of 22.32 micrometer. The array of 32 write drivers is built by shorting EN and ENB terminal of each cell.

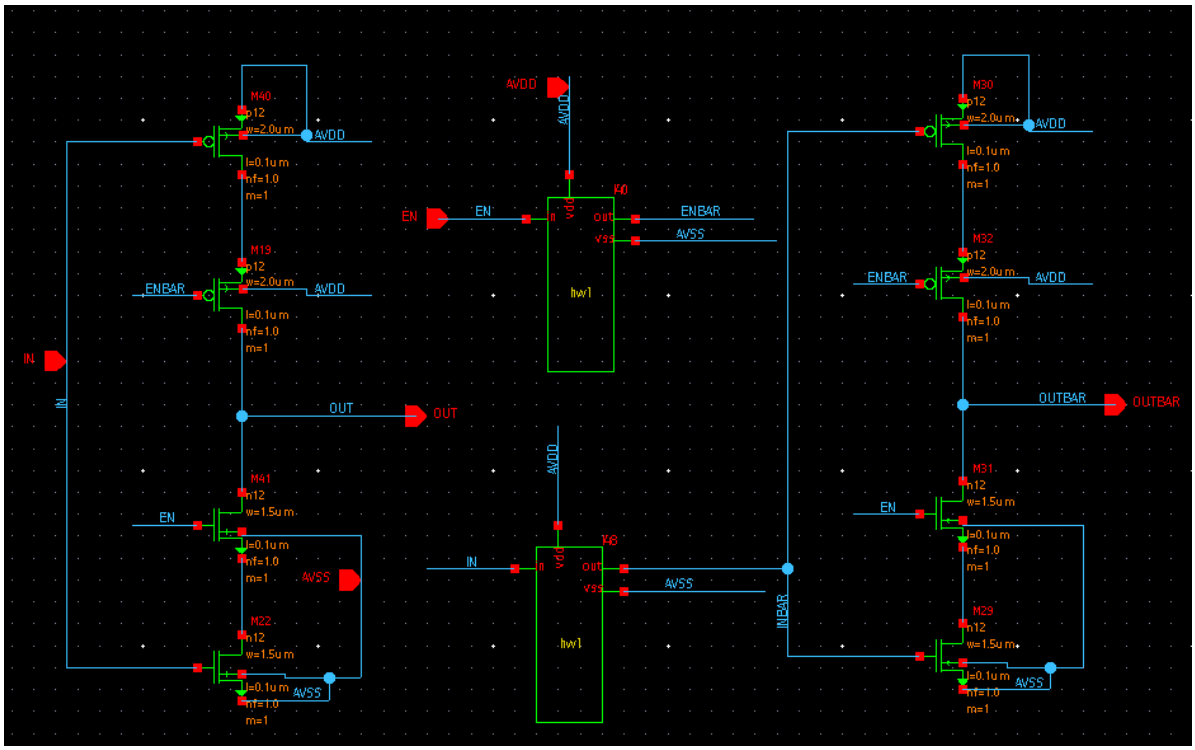


Figure 18. Write driver schematic

Figure 19. Write driver simulation

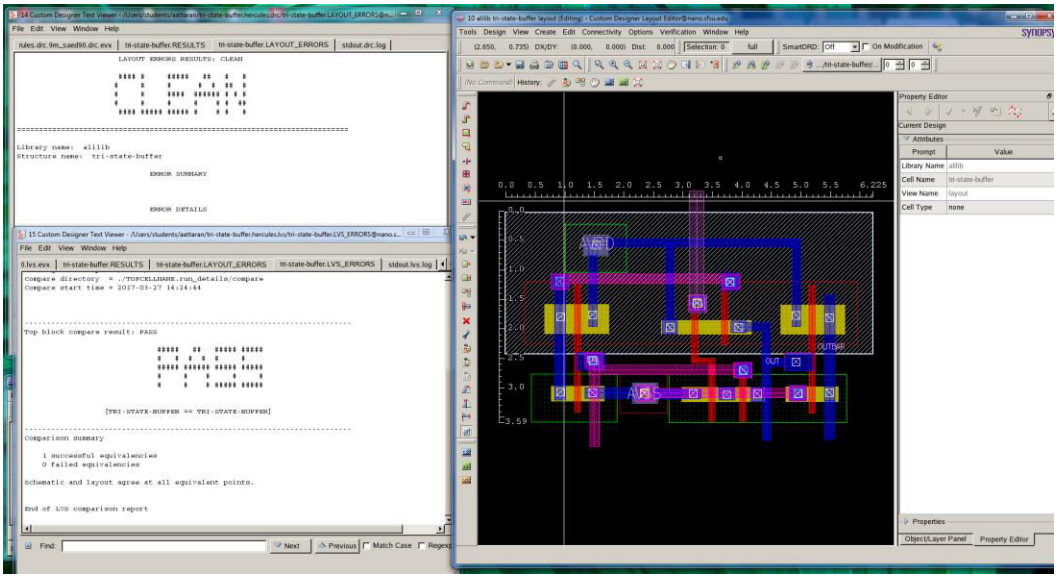


Figure 20. Layout of the write driver

The write driver needed upsizing because the write operation was unsuccessful in the complete design. So it's modified as below.

Post-layout simulation
Delay = 20.9 Ps

Dynamic power:
ptot= 773.1085n

Static power:

ptot= 705.7788n

2.6 Controller

Controller gates are designed to generate Sense Enable and precharge signals which are connected to the top level SRAM block. The transistors of the controller circuit is characterized to fine tune the output of top level SRAM. By tuning the transistors size in controller it is possible to enable the sense amplifier at the point of required 100mV bit-line differential voltage induced.

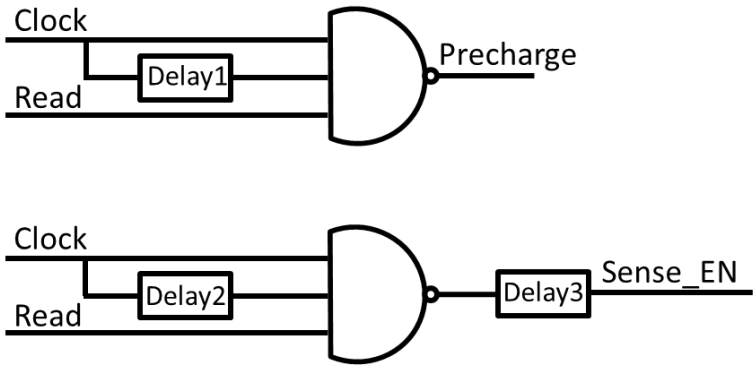


Figure 21. Controller signals

Delay 1 and 2 has 11 inverters and delay 3 has 21 inverters to generate the precharege and Sense enable signals at the right time.

Both precharge and sense_en signals need inverter chain buffers with progressive sizing to derive the next stages. The PC buffer is calculated to have 4 inverters with following sizes.

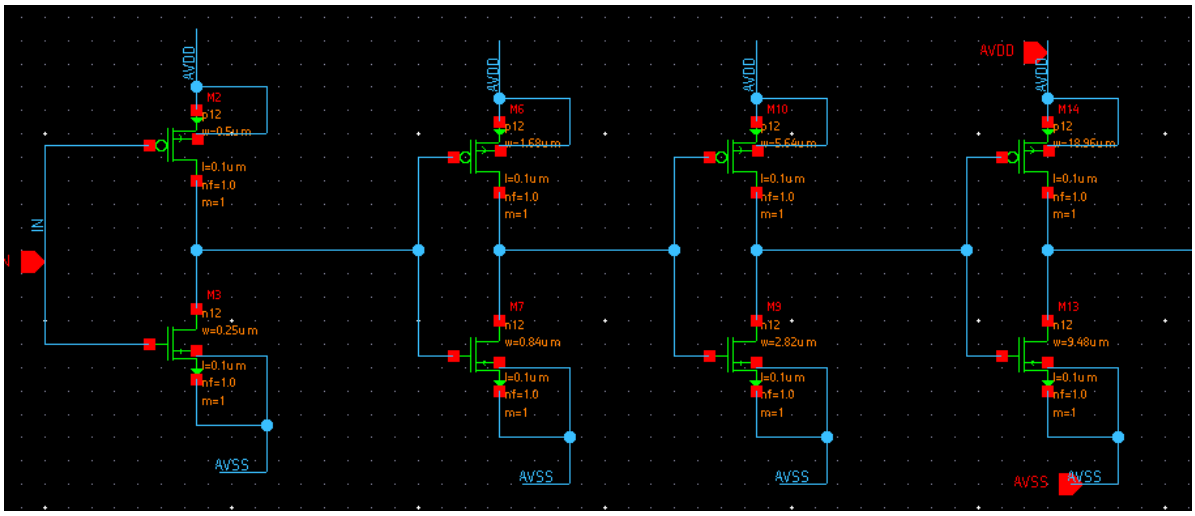


Figure 22. PC buffer

The sense_en buffer is calculated to have 4 inverters with following sizes.

Transistor size of the inverter at the output end is higher than that of input end to drive the output efficiently with load. Area of the layout=63*41=2583.

The cap ratio for PC is (32*3*1u)/(0.5u+0.25u) = 128

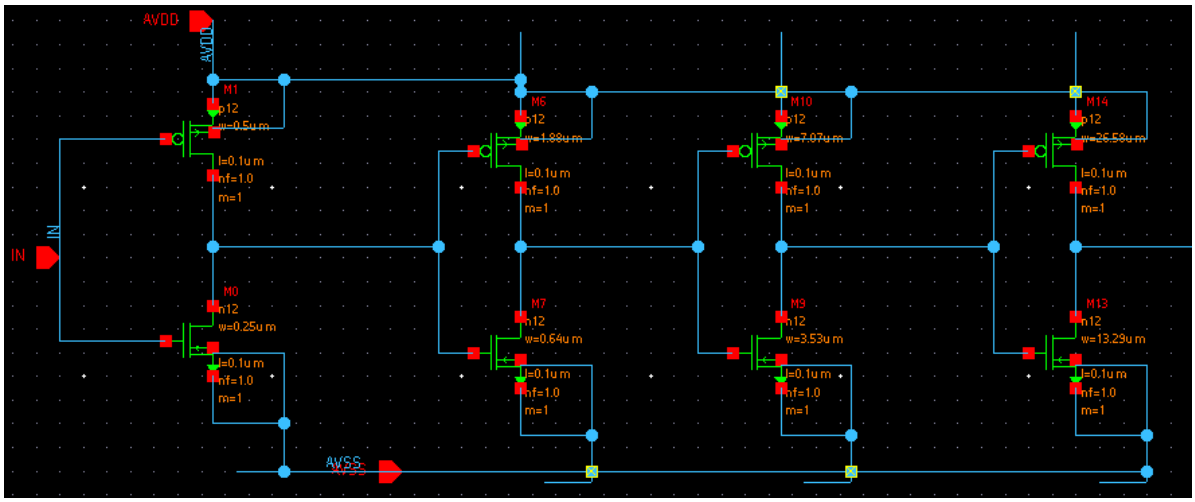


Figure 23. Sense_en buffer

The cap ratio for sense_en is (32*3*0.24u)/(0.5u+0.25u) = 75

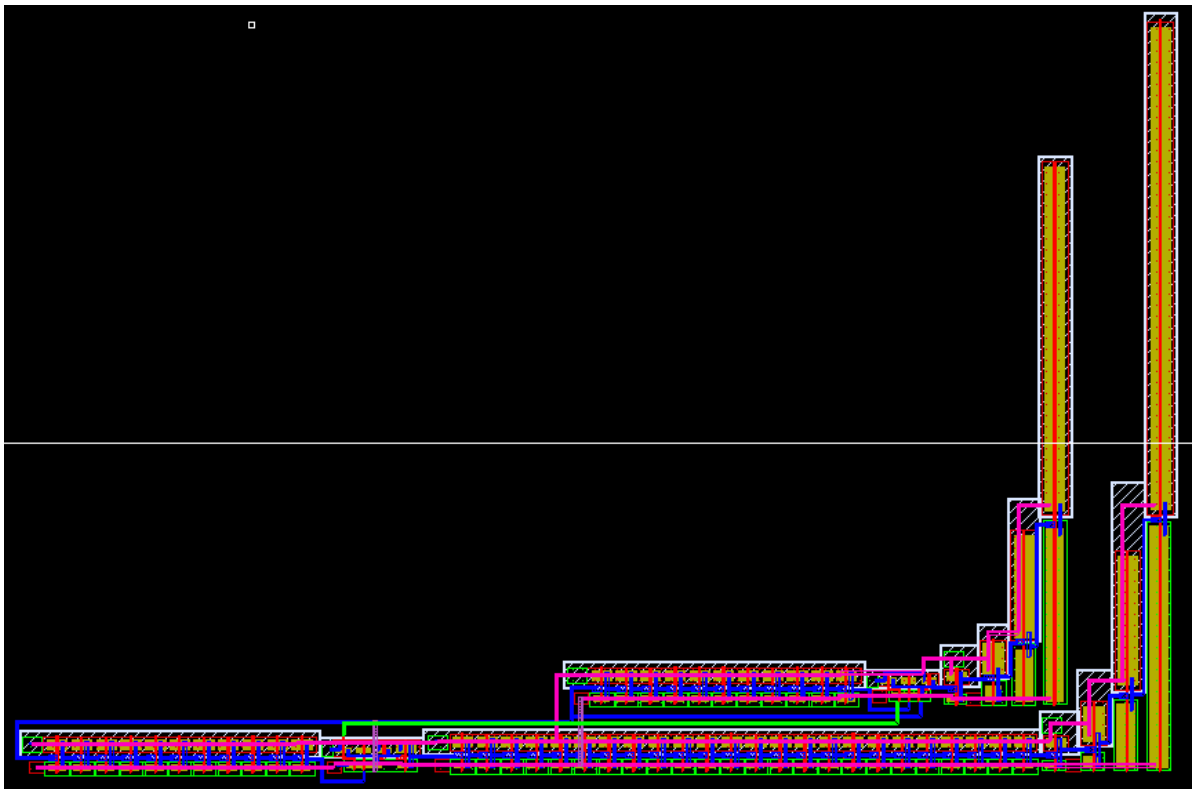


Figure 24. Controller layout

3 Complete Schematic and Layout

Once all the peripheral circuits are designed, all of the units are then integrated to the memory cell array. The complete SRAM schematic including precharge, clock buffer, row decoders, column decoders, sense amplifier and the write circuit is as shown in Figure 25.

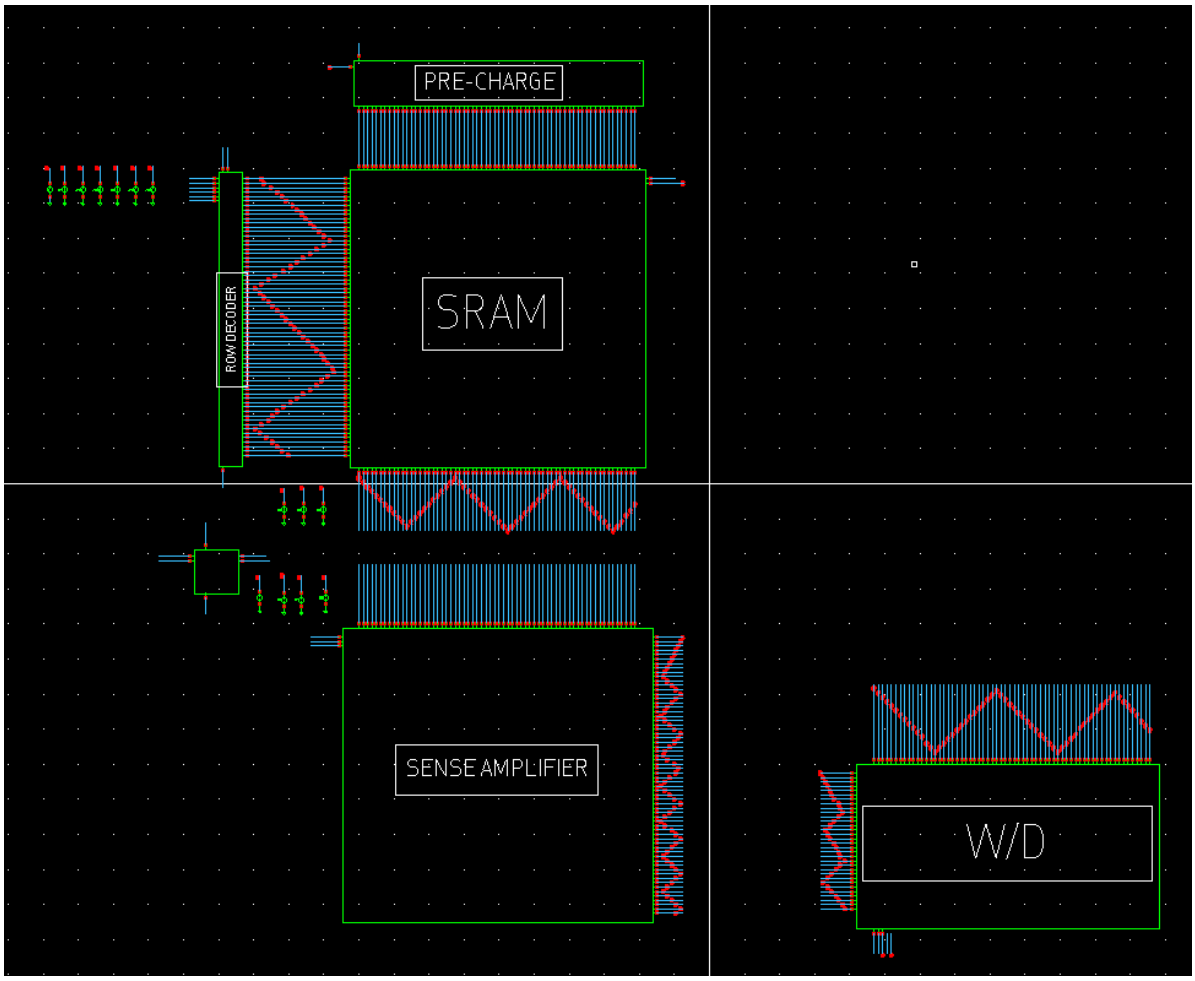


Figure 25. complete SRAM schematic

The corresponding layout of the design along with the rulers is given in Figure 26.

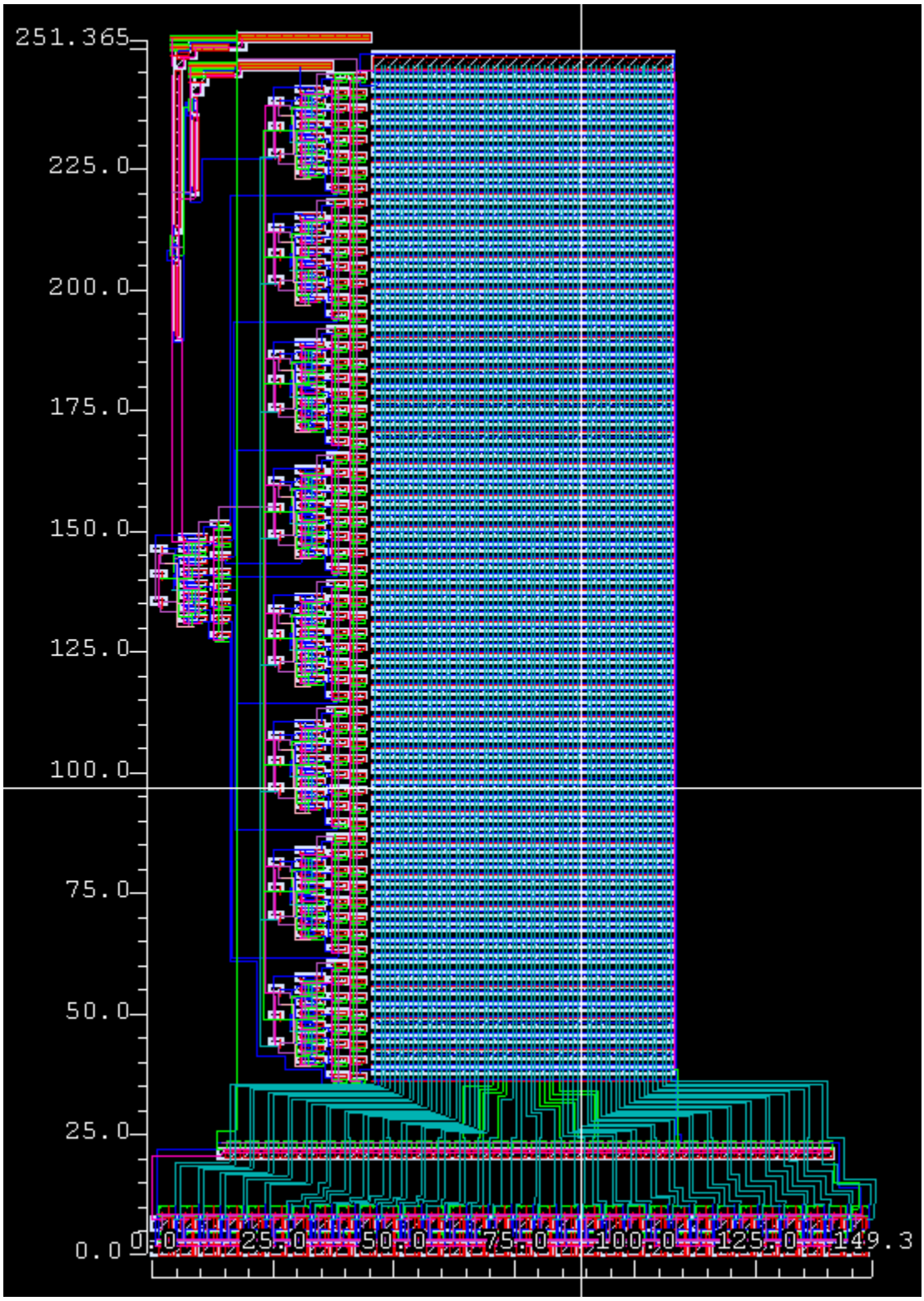


Figure 26. Area of the layout

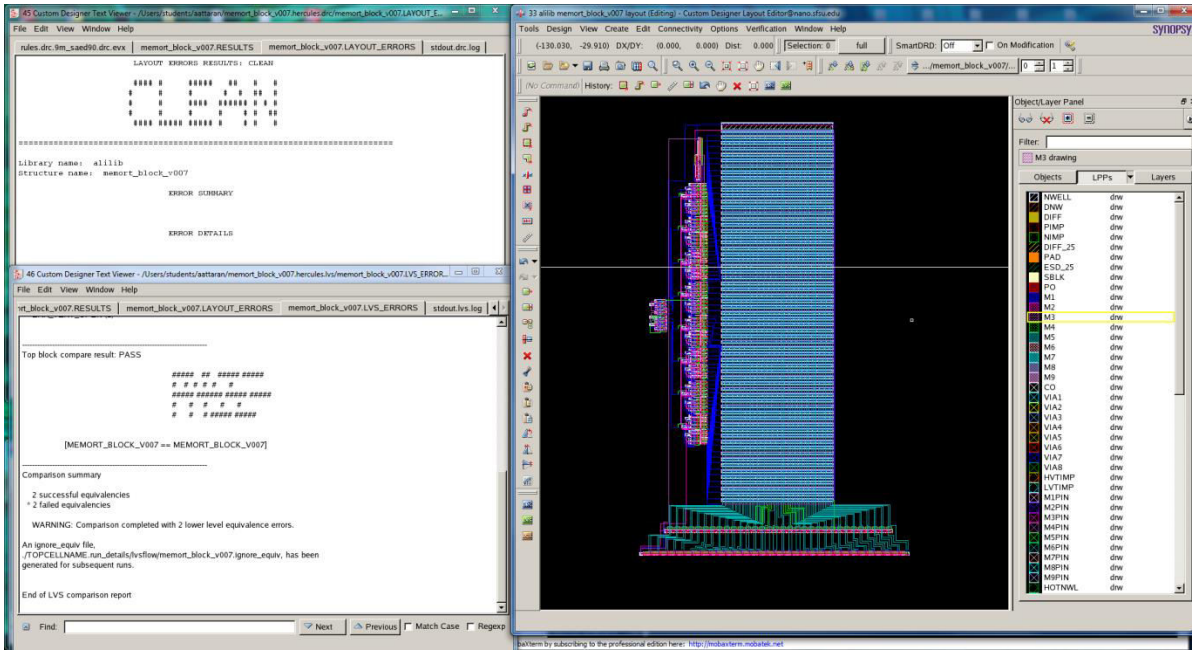


Figure 27. layout of the design

The total area of the design is 149.3*251.4= 37534.02

Therefore, total area that accounts for one bit is given by,

Area /bit = 18.32

The designed layout is checked in synopsis for design rule errors. There were no DRC errors in the layout. A snapshot of the DRC report is shown in Figure 26. The functionality is then tested by comparing the Layout versus Schematic (LVS). LVS matched and the report is shown in Figure 26.

4 Simulation and Results

The functionality of the SRAM is tested by writing a 8 bit data word 01011010 into the first and last row then reading the written value in the next clock cycle.

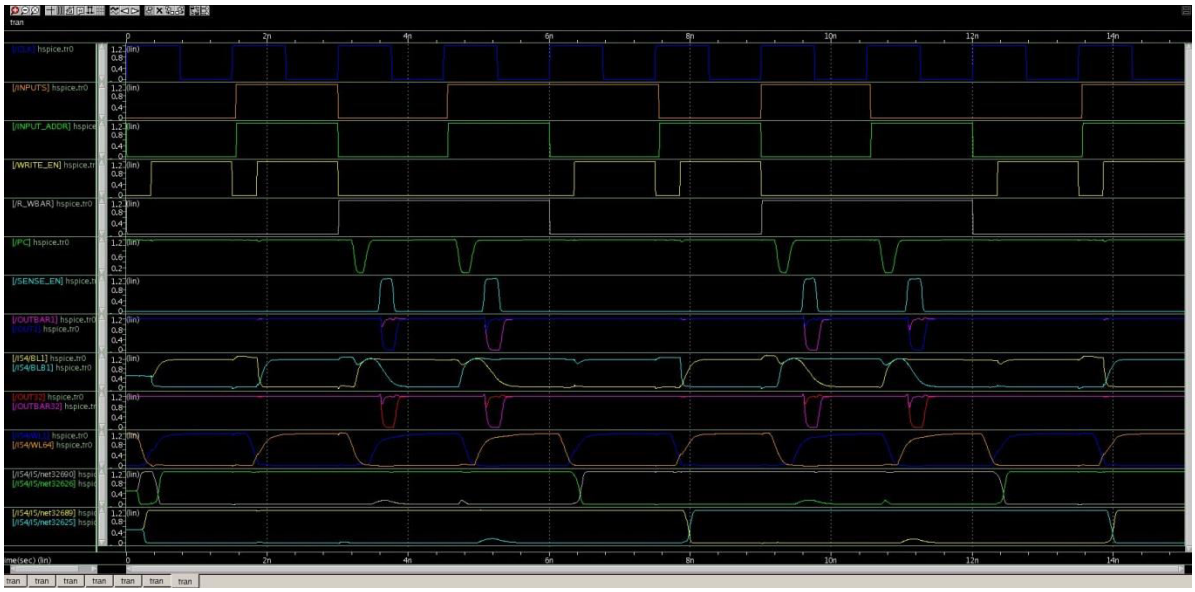


Figure 28. Write and read operations

The 100mV differential between BL and BLB is shown in figures below.

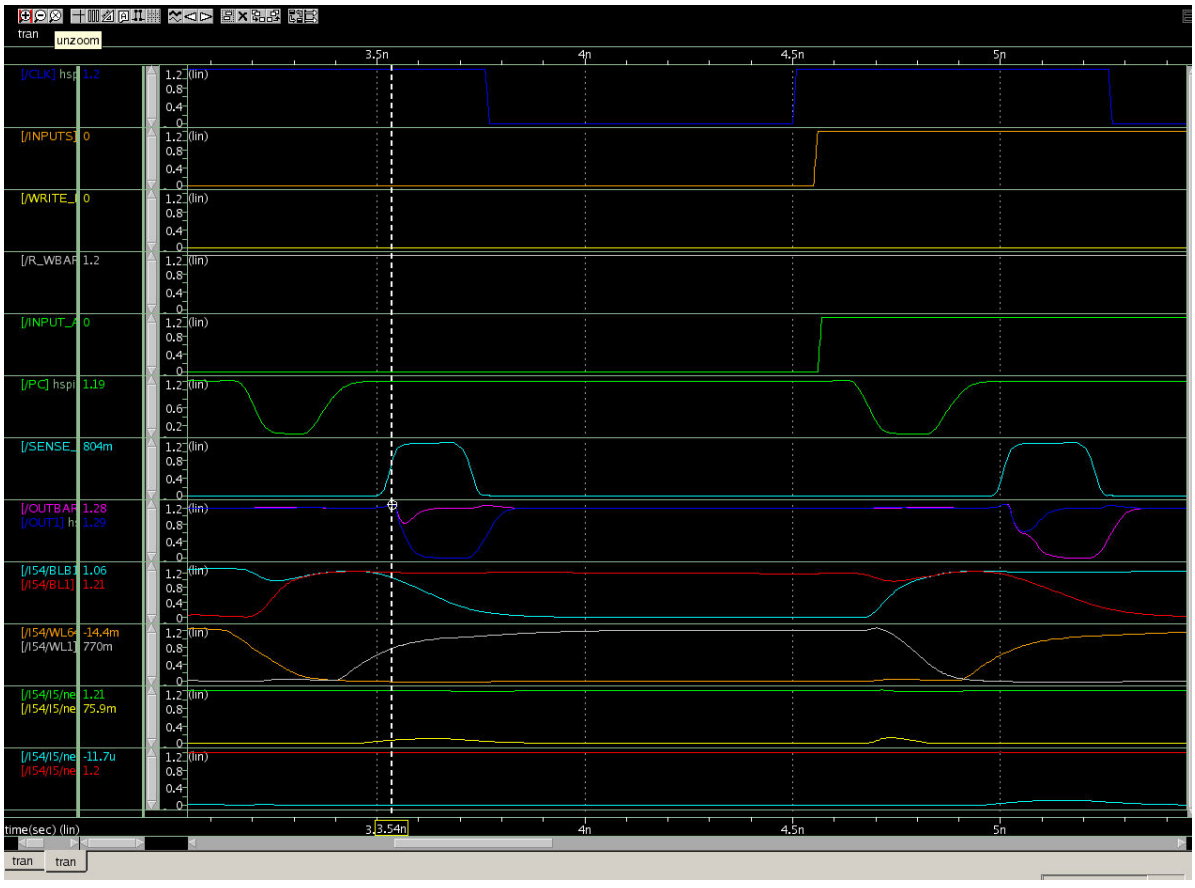


Figure 29. 100mV differential between BL and BLB

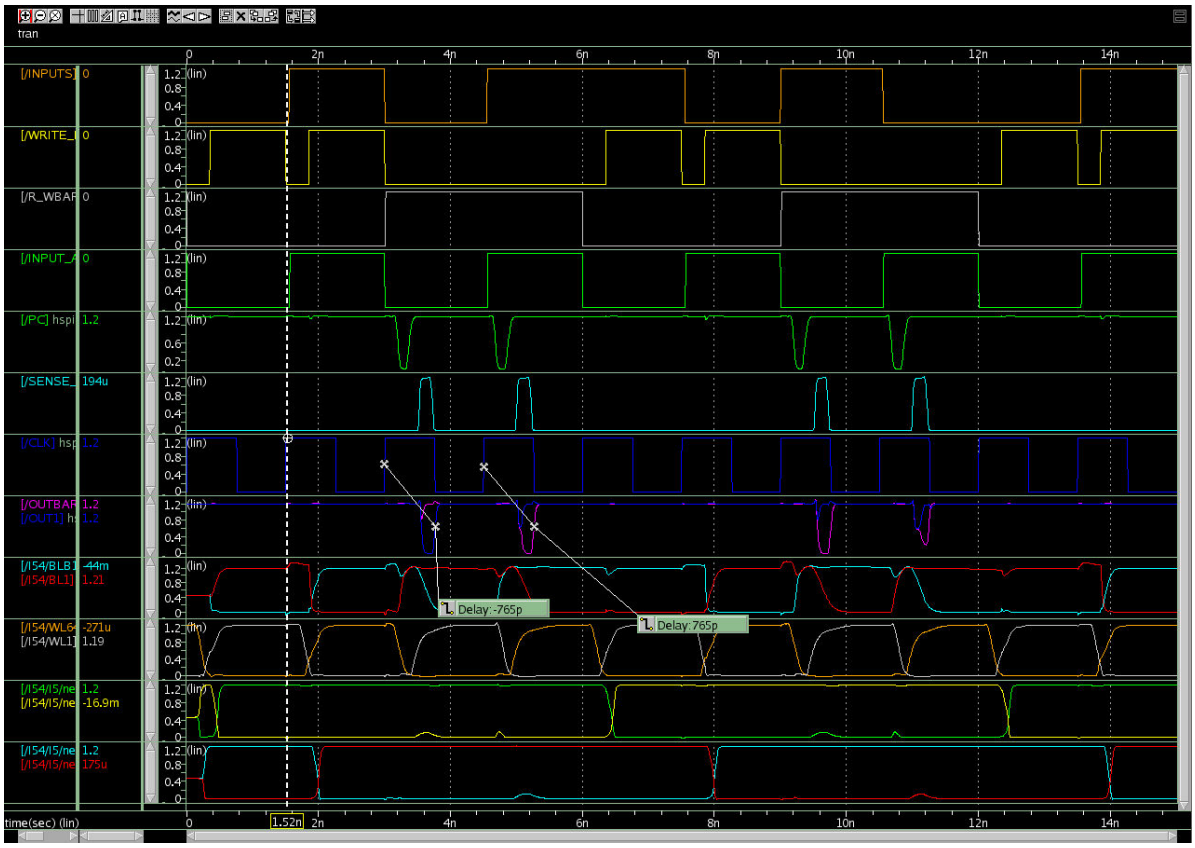


Figure 30. Read access time

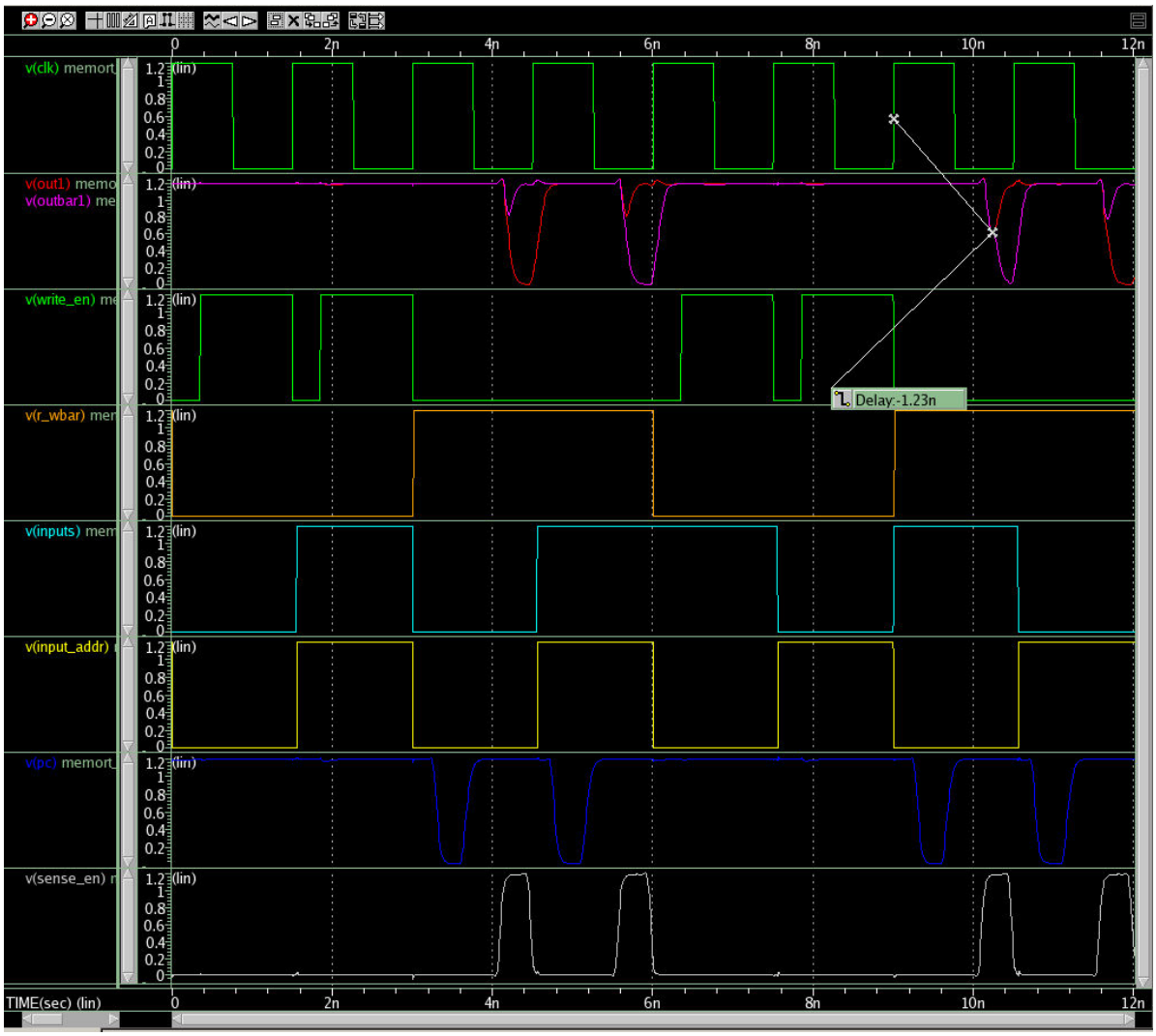


Figure 31. Read access time

When the clk is low, all the bitlines are precharged to VDD. During evaluation, the Write enable (WR) signal is activated. During this phase, the write operation take place and word bits are written into the corresponding memory cell depending on the row and column address.

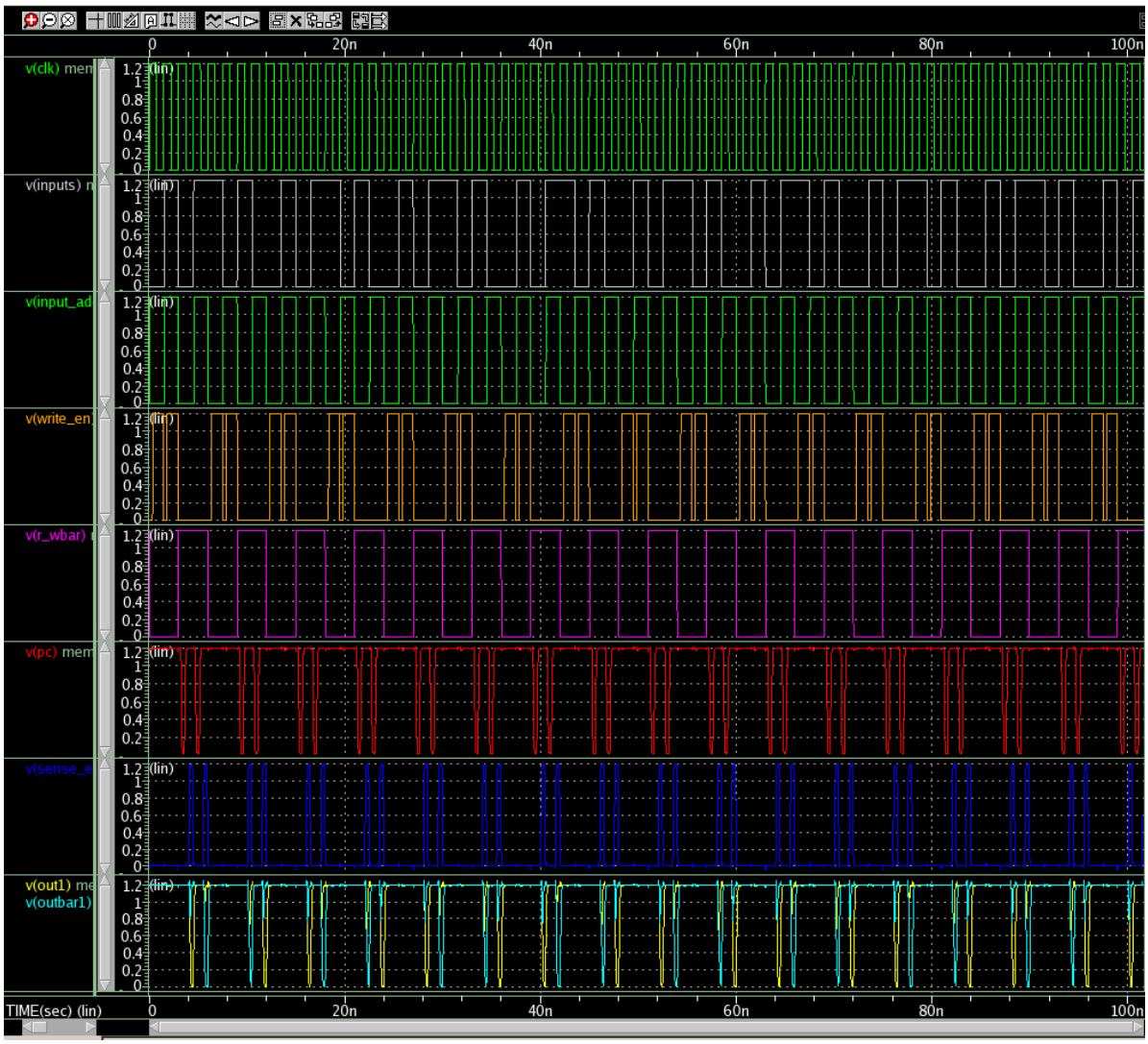


Figure 32. 80 clock cycle simulation

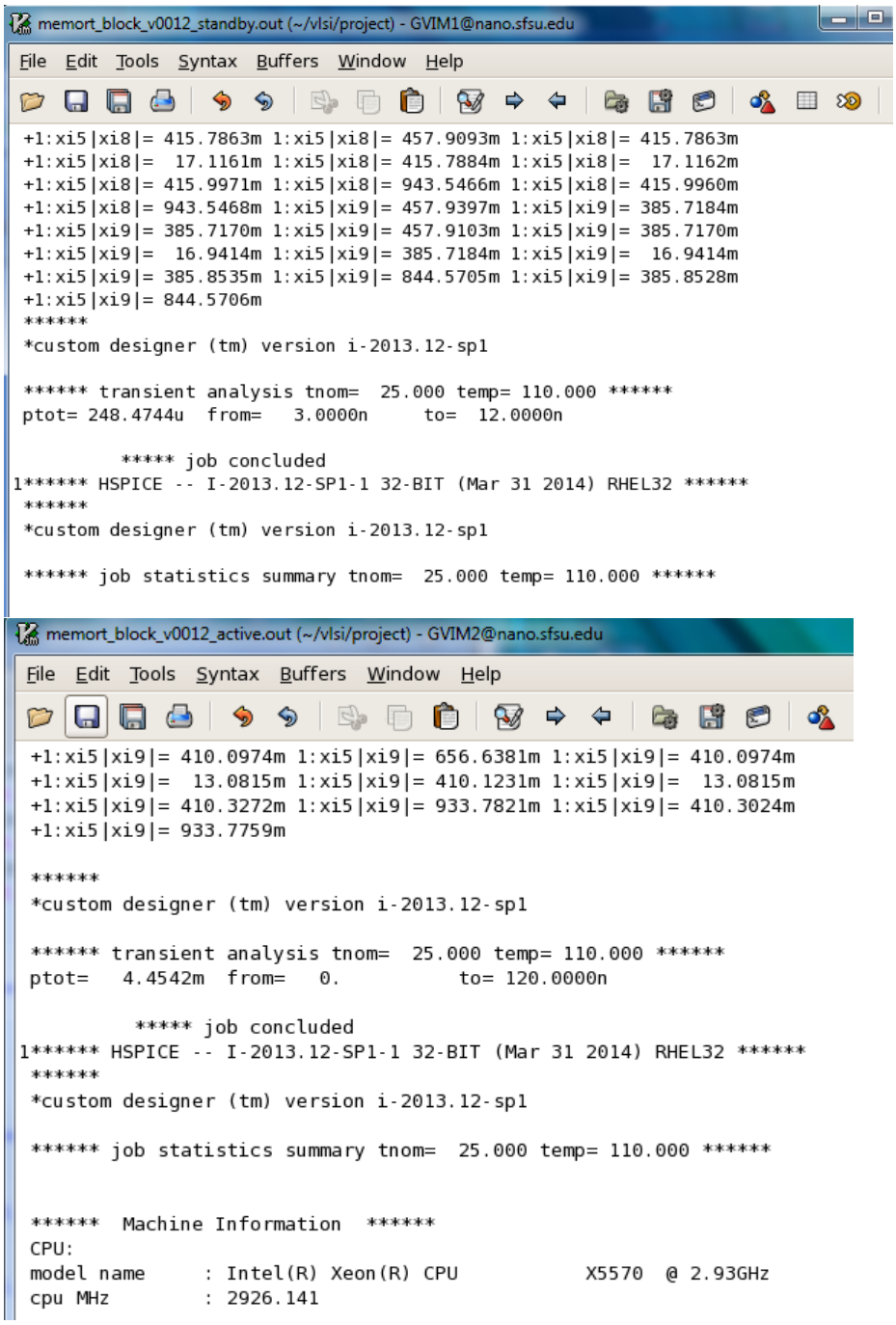


Figure 33. Dynamic and standby power Screenshot

5 Conclusion

The SRAM has comparatively low operating frequency but that is a trade-off for the low area/bit that we have tried to achieve. The memory cell has good noise margin and good control over read and write operations.

Input pattern for functional top-level verification and power/performance characterization:

An important final step in your projects is proper power/performance characterization. For the active power characterization, please apply the following pattern:

1. Write all 0 data to address 0
2. Write all 1 data to address 63
3. Read all 0 data from address 0
4. Read all 1 data from address 63
5. Write all 1 data to address 0
6. Write all 0 data to address 63
7. Read all 1 data from address 0
8. Read all 0 data from address 63

Run this 8-cycle sequence 10 times (i.e. 80 clock cycles total) and then measure the average power.

For the active power measurement, you do not have to run at max frequency because we will use the power per MHz as the metric for competition (i.e. we will divide your active power measurement, by your clock frequency to get average energy per cycle).

For the max frequency, you need to report the read access time, i.e. from 50% rise of the clock to 50% rise/fall of the output on the data bus (whichever longer), while meeting the 100mV bitline differential when the Sense Amplifier Enable (SE) signal is at 50% of its rise.

For standby power measurement, apply one write cycle to avoid metastable conditions for the SRAM cell, and then let all inputs and clock stay idle at zero and measure the steady state values of the supply current and multiply it by the supply voltage.

All the above measurements should be reported for post-layout of the top level.

Table 1. performance measurements

Active power (W)	Max clock frequency (Hz)	Active power/ clock frequency (energy/cell)	Read access time	Standby power (W)	Area
4.45m	666MHz (1.5ns period)	7.38pJ	1.23ps	248.4744u	149.3*251.4= 37534.02