DIGITAL LOGIC IMPLEMENTATION USING IN MEMORY BOOLEAN COMPUTATION

A DISSERTATION

Submitted in partial fulfillment of the requirements for the award of the degree

Of

MASTER OF TECHNOLOGY

In

ELECTRONICS AND COMMUNICATION ENGINEERING

(With Specialization in Microelectronics and VLSI)

By

SURBHI SHRIMALI

(Enrollment No. **17534010)**

Under the guidance of

Dr. ANAND BULUSU & Prof. SUDEB DASGUPTA

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY, ROORKEE

ROORKEE-247667

JUNE 2019

CANDIDATE’S DECLARATION

I hereby declare that the work presented in this seminar report with title “**Digital Logic Implementation using In-memory Boolean computation**" towards the fulfillment of the requirement for the award of the degree of **Master of Technology** in **Microelectronics and VLSI** submitted in the Dept. of Electronics & Communication Engineering, Indian Institute of Technology, Roorkee**,** India is an authentic record of my own work carried out during the period from **January 2019 to June 2019**under the supervision of **Dr. ANAND BULUSU**, Associate Professor and **Prof. SUDEB DASGUPTA**, Head of department, Dept. of ECE, IIT Roorkee. The content of this report has not been submitted by us for the award of any other degree of this or any other institute.

DATE: ....
PLACE: ....
ENROLL. NO. - 17534010

SIGNED: ....
(SURBHI SHRIMALI)

....

CERTIFICATE

This is to certify that the statement made by the candidate is correct to the best of our knowledge and belief.

DATE: ....

SIGNED: ....

(DR. ANAND BULUSU)
ASSOCIATE PROFESSOR
DEPT. OF ECE, IIT ROORKEE

SIGNED: ....

(PROF. SUDEB DASGUPTA)
HEAD OF DEPARTMENT
DEPT. OF ECE, IIT ROORKEE

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude towards my guide **Dr. ANAND BULUSU**, Associate Professor and my co-guide **Prof. SUDEB DASGUPTA**, Head of Department Dept. of Electronics and Communication Engineering, IIT Roorkee for their ideal guidance throughout the entire period. I want to thank them for the insightful discussions and constructive criticisms which certainly enhanced my knowledge as well as improved my skills. Their constant encouragement, support and motivation were key to overcome all the difficult and struggling phases.

Also, I would like to thank **Mr. Inder Choudhary and Mr. Dinesh Kushwaha**, Research Scholar, Microelectronics & VLSI, Dept. of Electronics and Communication Engineering, IIT Roorkee for guiding me as mentor whenever I faced any difficulties and helping me out in learning tool.

I would also like to thank **Department of Electronics and Communication Engineering, IIT Roorkee** for providing the lab and other resources for the project work.

My special thanks to my parents and all my friends for keeping me motivated and providing with valuable insights as a part of various healthy discussions.

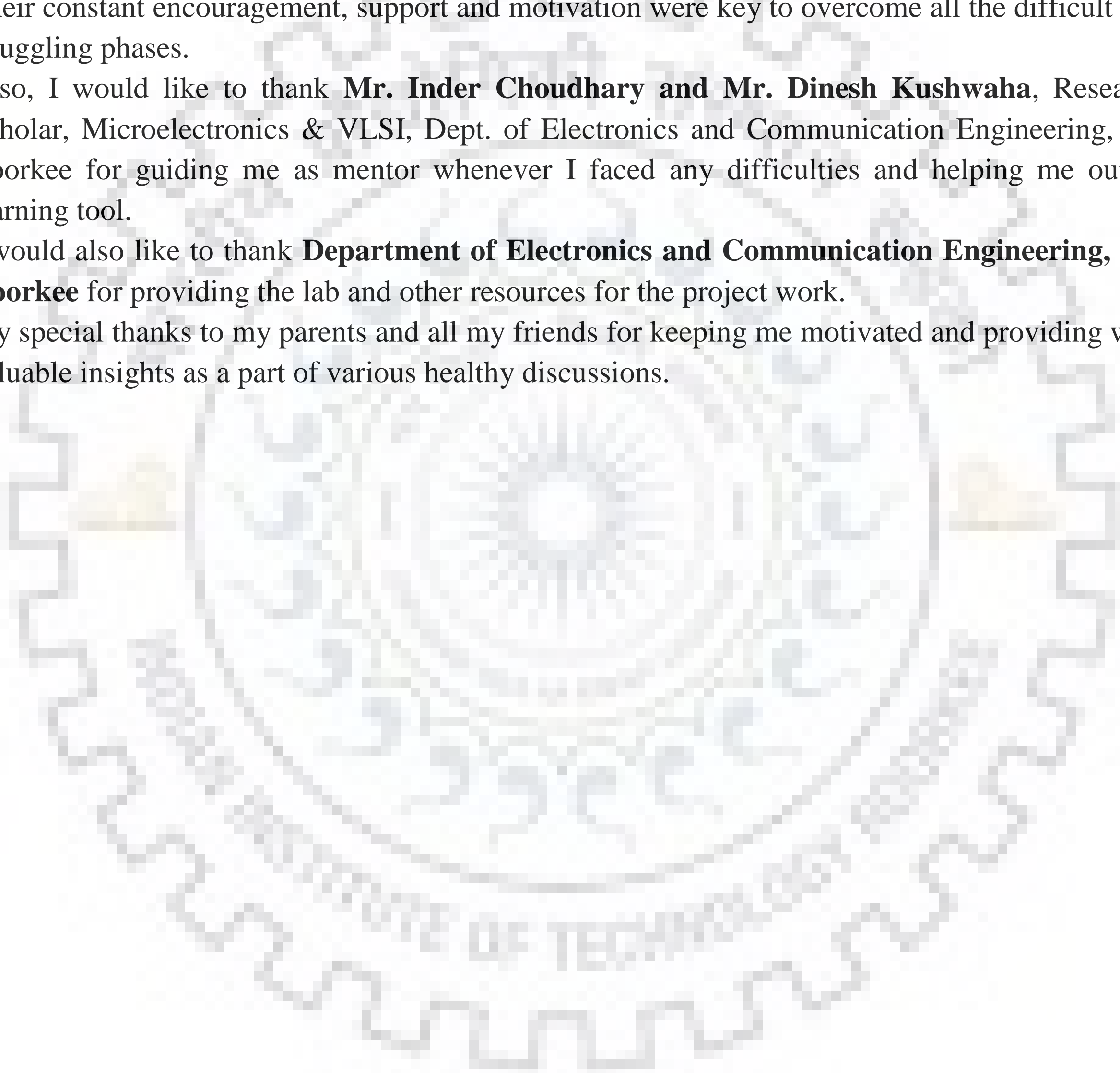

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Von-Neumann Architecture

1.2 State-of-the-art

1.3 Motivation behind In-Memory Computation

CHAPTER 2: IN-MEMORY COMPUTATION BLOCKS WITH CONVENTIONAL SENSE AMPLIFIERS

2.1 6T-SRAM cell

2.2 Precharge Circuit

2.3 Sense Amplifier

CHAPTER 3: PROPOSED SENSE AMPLIFIERS FOR IN-MEMORY COMPUTATION

3.1 NANDNOR Sense Amplifier

3.2 XOR Sense Amplifier

CHAPTER 4: SIMULATION METHODOLOGY, RESULTS AND COMPARISON

4.1 Computing using Conventional Design

4.2 Computing using proposed Design

CHAPTER 5: CONCLUSION AND FUTURE WORK

5.1 Conclusion

5.2 Future Work

REFERENCES

LIST OF FIGURES

Fig.1.1 Von-Neumann Architecture

Fig. 2.1 SRAM 6T cell Schematic and Layout

Fig. 2.2 Circuit for a) Gate capacitance calculation b) Parasitic capacitance calculation

Fig. 2.3 Precharge Circuit Schematic and layout using common centroid technique

Fig. 2.4 Asymmetric NOR Sense Amplifier circuit schematic and layout

Fig. 2.5 Asymmetric NAND Sense Amplifier circuit schematic and layout

Fig. 3.1 Proposed Nand-Nor Sense Amplifier circuit

Fig. 3.2 Proposed Nand-Nor Sense Amplifier layout

Fig 3.3 XOR Sense Amplifier Schematic and Layout

Fig.4.1 Schematic and Layout of the single column for a memory array using two 6T SRAM cells and conventional Sense Amplifiers for logic operation NAND-NOR-XOR

Fig.4.2 Timing diagram for any logic operation for in-memory computation

Fig.4.3 Schematic and Layout of the single column for a memory array using two 6T SRAM cells and Proposed Sense Amplifiers for logic operation NAND-NOR-XOR

Fig. 4.4 Post layout simulation results for proposed design

Fig.4.5 Simulation results for read operation using proposed sense amplifier

Fig. 4.6 Post layout corner analysis results for proposed design

Fig 4.7 MC simulation results with Vth variation a) for conventional NAND operation b) for proposed NAND operation

Fig 4.8 MC simulation results with Vth variation a) for conventional NOR operation b) for proposed NOR operation c)for conventional XOR operation d) for proposed XOR operation

Fig 4.9 Parametric simulation results with 30mV Vth variation at two temperatures (25°C and 80°C) and±10% variation in supply voltage for NAND operation

Fig 4.10 Parametric simulation results with 30mV Vth variation at two temperatures (25°C and 80°C) and ±10% variation in supply voltage a) NOR case 00 b) NOR case 10 c) XOR case 00 d) XOR case 10

Fig 4.11 Parametric simulation results with 30mV Vth variation at two temperatures (25°C and 80°C) and supply voltage variations from 1.2 to 1.8 V a) NAND b) NOR

Fig 4.12 Parametric simulation results with 30mV Vth variation at two temperatures (25°C and 80°C) and supply voltage variations from 1.2 to 1.8 V for XOR

Fig 5.1 Comparison of in-memory instruction with conventional processor instructions

LIST OF TABLES

Table 2.1 Sizing of transistors for SRAM cell

Table 2.2 Sizing of transistors for Precharge circuit

Table 2.3 Sizing of transistors for Sense Amplifier circuits

Table 3.1 Sizing of transistors for Proposed NANDNOR Sense Amplifier

Table 3.2 Sizing of transistors for Proposed XOR Sense Amplifier

Table 4.1 Settling time comparison for NAND, NOR and XOR logic operation

Table 4.2 A comparison table for proposed design and referenced design

CHAPTER 1

INTRODUCTION

Since the invention of Moore's law, a continuous downscaling of transistor sizing has become the main motivation to the growth of very large scale integration (VLSI) industry, but along with that, there is an increasing demand of speed and energy efficiency for computing applications. Over thirty years ago, Von Neumann architecture – a standard computer architecture - is being used for all the computing applications. [1]

1.1 Von-Neumann Architecture

Von-Neumann Architecture is a structure being followed by computers. It uses a single processor and a single memory for both data and instructions. It consist of following components:

a) **A Central Processing Unit (CPU) -**
The Central Processing Unit (CPU) executes the instructions of a computer program. It is also known as the processor. It has the Arithmetic and Logic Unit (ALU), Control Unit (CU) and a number of registers.

b) **A Storage Element -**
The storage element is a memory unit. It has a main memory that is called Random Access Memory (RAM). CPU can directly access this memory. RAM consists of an address part and a data part which are in binary form. The unique address will point out every location in the memory.

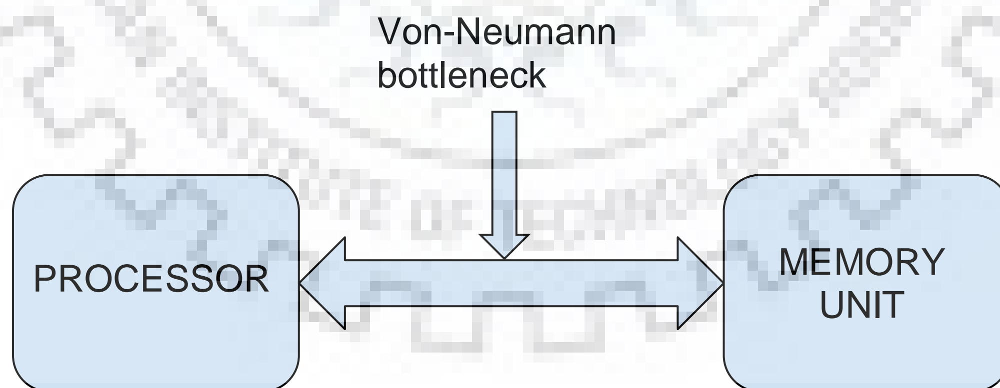

Fig.1.1 Von-Neumann Architecture

c) **A Von Neumann bottleneck –**

It is a connecting tube for data as well as address transfer between the CPU and the storage element. As the program executes, one instruction will be carried out at a time and the task of changing the contents of the storage element will be obtained by sequential transmission of instructions back and forth through the Von-Neumann bottleneck. [1]

The idea of Von Neumann architecture is that the instruction data and program data will be stored in the same memory. This design is still used in most computers produced today. [2]

The most considerable part of Von-Neumann Architecture is Von-Neumann bottleneck which has different meanings in terms of hardware (architecture) and software.

i) From a hardware point of view, as I mentioned previously, it is a connecting tube between the memory unit and processor for both data and address transfer.
ii) From a software point of view, the assignment statement is Von-Neumann bottleneck.

Significance of Von-Neumann Bottleneck –

a) From **Hardware** point of view, frequent and heavy traffic of data between physically separated processor chip and memory chip give significance to this term "bottleneck". A large back and forth transfer of data between them causes a lot of energy consumption as well as limited throughput. Ironically, a large part of the traffic in the bottleneck is useless. [3]
b) From **Software** point of view, the assignment statements being used in the program. For the requirement of necessary repetitions, the programmer designs the nest of control statements. These repetition statements cause a large execution period. [1]

In recent years, the processors are improving their speed limits but memory units are not as faster as processor so the processor has to wait for the time data has to be fetched from memory unit. Thus it causes a limited transfer rate between processor and memory unit. Along with that, the fact is that instructions can only be done one at a time and can only be carried out sequentially, which cause weak performance and referred to as a bottleneck.

The Von-Neumann bottleneck problem can only be overcome by a different approach for processor architecture.

Approaches to overcome Von-Neumann Bottleneck –

a) Caching – Frequently used data can be stored in the cache memory (RAM only) which will be close to the processor and easily accessible than main memory.
b) Prefetching – The fetch operation will be performed before it is being requested to increase the fetching speed.
c) Multithreading – It allows to manage multiple requests at a time by same user.
d) New types of RAM such as DDR, SDRAM etc.
e) RAMBUS – It is a memory subsystem having RAM and its controller, and also the bus connecting RAM to the processor.
f) Processing in memory (PIM), which allows the memory to do computation as well inside it. So in this case, both the processor and memory unit will be integrated on a single chip.

1.2 State-of-the-art

Several researches has come up with the different approached to overcome Von-Neumann bottleneck. For example, an idea to integrate processing unit (ALU) near by the memory unit that is called near memory computing.[5] So beginning with near memory computation approach, now the idea of in-memory computation has become state-of-the-art. The digital domain of Computing-in-Memory (CiM) consists of a different processor architecture design while analog domain consists of machine learning classifier and neural network applications using dot product computation.[6] In addition to that, some papers are going beyond CMOS technology such as Magnetic RAMs [8], resistive RAMs [9] etc. but these emerging technologies needed a lot of satisfactory researches to get commercialized for on-chip memory.

1.3 Motivation behind In-Memory Computation

The biggest motivation behind in-memory computation is the fascinating idea of using memory not only as a storage unit but also as a processing unit. This idea has reduced large data transfer between memory and processor through Von-Neumann bottleneck in the standard computer architecture.

In Von-Neumann architecture, both CPU and memory were two separate chips and there was a lot of data transfer between them which caused less energy efficiency and throughput limitation. The idea of In-Memory Computation is to bypass the von-Neumann bottleneck by accomplishing computations right inside the memory array. In that case, there will not be frequent back and forth traffic between memory and processor thus we have expected improvements in both energy efficiency as well as throughput.

In-Memory Computation will allow memory array to work as a standard storing element and along with that, it will perform additional logical operations inside it. Computing within the memory array enhances the memory functionality thereby reducing the number of

unnecessary transfers of data for certain class of operations like vector bitwise Boolean logic etc.

In this project, I tried In-Memory boolean logic operations using conventional 6T SRAM memory cell block in order to understand the concept and future possibilities. In memory boolean computation is a possible approach to overcome the von-Neumann bottleneck. In the reference paper [3], NAND and NOR sense amplifiers were being proposed and along with that, XOR operation was obtained with the help of further NAND and NOR logic circuits. The proposed work is different from the work in [3] since I have given a separate combined NANDNOR sense amplifier and an XOR sense amplifier design. The simulation results were being compared with the reference paper and betterment of the proposed design has been proven successfully.

The aim of this work is to get in-memory computation in standard SRAM cells with minimum modification in the peripheral circuitry and also the focus was on the two crucial parameters of memory design that are chip area and power consumption. The proposed sense amplifier designs are the only change in the memory peripherals

CHAPTER 2

IN-MEMORY COMPUTATION BLOCKS WITH CONVENTIONAL SENSE AMPLIFIERS

The memory array used for Boolean computation is a bit different from the standard memory array as we need to perform logical operations in memory along with storage operation.
Now we will be discussing the in-memory computation blocks with optimum sizing tables and designed layout -

2.1 6T-SRAM Cell

The basic building block of any memory array is the cell which will be used to store a single bit. Conventional 6T SRAM cell consists of 6 transistors out of which a pair of NMOS transistors and a pair of PMOS transistors are used as back to back inverters and a pair of NMOS transistors will act as access transistors. To perform read and write operation, there will be one word line select (WL) to start either of the operations, by turning on access transistors and bit lines (BL and BLB).

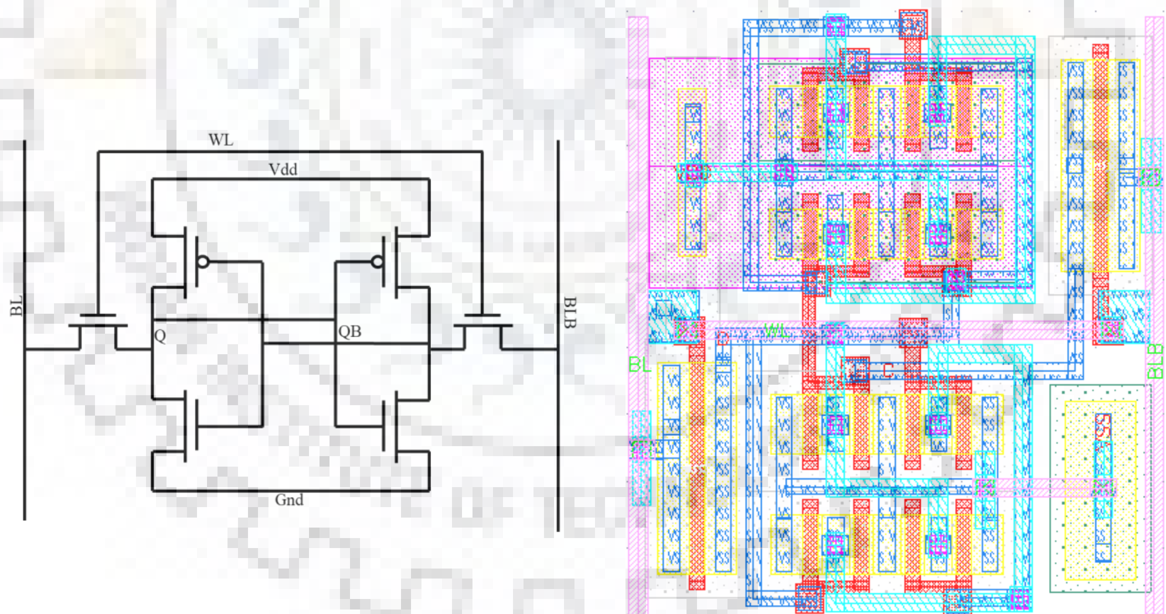

Fig. 2.1 SRAM 6T cell Schematic and Layout

There are three modes of operation in SRAM cell -

1. Hold Mode
2. Read Mode
3. Write Mode

To get the exact functionality, we need to size the transistors to obtain read and write stability. For that purpose, we need to check the configuration of each transistor in read and write mode and then to find out the optimum sizing of each transistor.

So from the CR ratio and PR ratio calculation, we obtain the sizing of SRAM cell. Here I I have kept sizing of the transistor as given below table:

Table 2.1 Sizing of transistors for SRAM cell

SRAM 6T	PMOS (Inverter)	NMOS(Inverter)	NMOS (Access)
$\frac{W}{L}$(um)	$\frac{2.97}{0.225}$	$\frac{3.53}{0.225}$	$\frac{3.18}{0.225}$

From the above sizing, SRAM schematic was designed in cadence tech 180nm scl library. The layout of the same was being drawn with common centroid technique (specifically latch part) to reduce mismatching as SRAM cells have higher chances of an erroneous response by flipping stored data.

There are different parameters which we need to take care of to reach minimum area and power requirements, for example, Read Access Time, Write Access Time, leakage current, static noise margin, etc.

One of the factors to affect all the above parameters are the Gate capacitance at WL line and parasitic capacitance at BL and BLB lines.

i) Gate Capacitance calculation

The wordline (WL) in the SRAM cell is connected to the gate terminal of two access transistors so the gate capacitance will be

$$\text{Cwl} = number\ of\ columns\ \times 2(Cgs + Cgd + Cgb + Cov)$$

Cg capacitance in the circuit has to be adjusted until the average delay from c to g equals the delay from c to d.

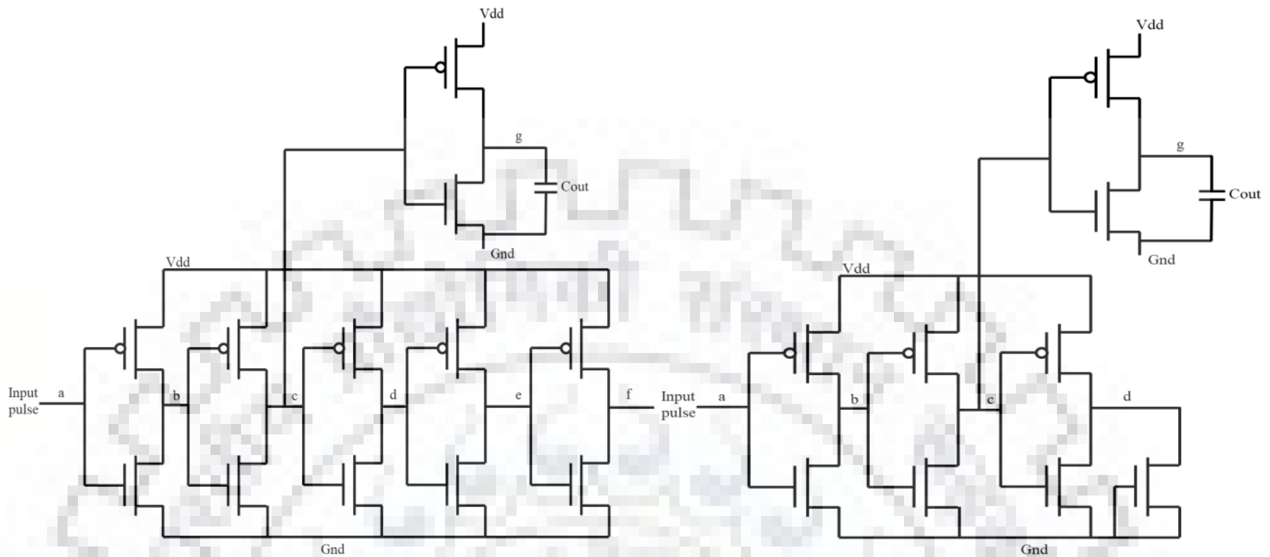

Fig. 2.2 Circuit for a) Gate capacitance calculation b) Parasitic capacitance calculation

ii) Parasitic Capacitance calculation :

The BL and BLB lines of SRAM cell are connected to the S/D terminal of access transistors so the total value of BL or BLB capacitance will be

$$Cbl = number\ of\ rows\ \times (Cdiff + Cwire)$$

To fulfill the objective of computing inside memory will be done by conventional 6T-SRAM because we do not want to make changes in the major part of memory array so that it will be easy to fabricate and easy to accept the smaller changes by industry.

From the above capacitance calculation, the obtained value of BL and BLB capacitance was

$$C_{BL} = C_{BLB} = 3.4207fF$$

For a memory array having 64 rows then total BL and BLB capacitance will be

$$C_{BL} = C_{BLB} = 218.57fF$$

2.2 Pre-charge Circuit

In a SRAM memory array, each bitcell's bit and bitbar lines has to be fully charged before any operation is going to be performed. The circuit required for this precharging operation of BL and BLB is called preharge circuit. The function of this circuit is that it will pull-up the bit lines of the column which is being selected to supply voltage level and perfectly equalized them before performing every read or write operation. The precharge circuit consists of a pair of PMOS transistors which will pull up bitline voltages to the supply voltage when they are turned on with Vpre signal and along with that, another PMOS transistor will be used to equalize voltages at both bitlines. The reason behind using a PMOS transistor is because of its property of passing good VDD.

Table 2.2 Sizing of transistors for Precharge circuit

PRECHARGE Ckt	Mpre	Meq
$\frac{W}{L}$(um)	$\frac{2}{0.18}$	$\frac{1}{0.225}$

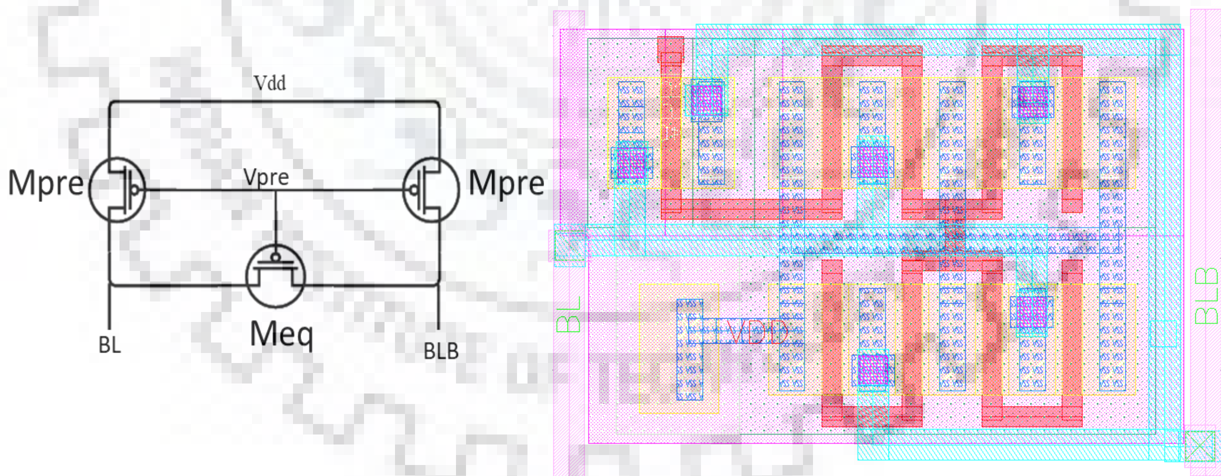

Fig. 2.3 Precharge Circuit Schematic and layout using common centroid technique

2.3 Sense Amplifier

Sense Amplifier (SA) is an important peripheral circuits for performing read operation in the CMOS Static Random Access Memories, and now for computing inside memory, the only changes are being done in this circuit to obtain Boolean logic operations. The primary job of a SA in memory is to sense the small voltage difference between bitlines and then amplifying that small signal variation to retrieve the stored data in a singlr cell. A small voltage difference is developed by discharging of either BL or BLB by the read access bitcell. More time will be required to perform read operation in memory due to small bitcell size and large bitlines capacitance. A fast, robust and low power sense amplifier is a biggest challenge because of a significantly large bitline capacitances.

In conventional latch-based sense amplifier, to obtain logic NAND/NOR/XOR operation, we need to design asymmetric latch-based sense amplifiers. We can skew any transistor in several ways and transistor sizing is one of them. The transistor sizing method is being used for skewing the transistors. For that the concept used is as below:

a) NOR Logic operation –

To obtain NOR operation using sense amplifier, skewing of transistors will be done for the pull-down transistors MBL and MBLB.

i) For the case of ‘AB’ = ‘00’ (‘11’) input, BL(BLB) discharges to 0V, but BLB (BL) remains in the same precharge state.
   1. For the case of ‘11’, BLB starts to discharge and the SA amplifies the voltage difference between BL and BLB, resulting in SAout=‘0’ and SAoutb = ‘1’.
   2. Whereas for the case ‘00’, BL starts to discharge, while RBL is at VDD, giving SAout=‘1’ and SAoutb = ‘0’.

ii) For the case of ‘AB’ = ‘01’ or ‘10’ both BL and BLB discharges to the same voltage so if the transistor MBL is deliberately sized bigger compared to MBLB, its current carrying capacity increases. Since the current carrying capability of MBL is more than MBLB, SAout node discharges faster, and the cross-coupled inverter pair of the SA stabilizes with SAout=‘0’ and SAoutb = ‘1’.

Thus it can be observed that SAout generates a NOR gate (thus, SAoutb outputs OR gate).

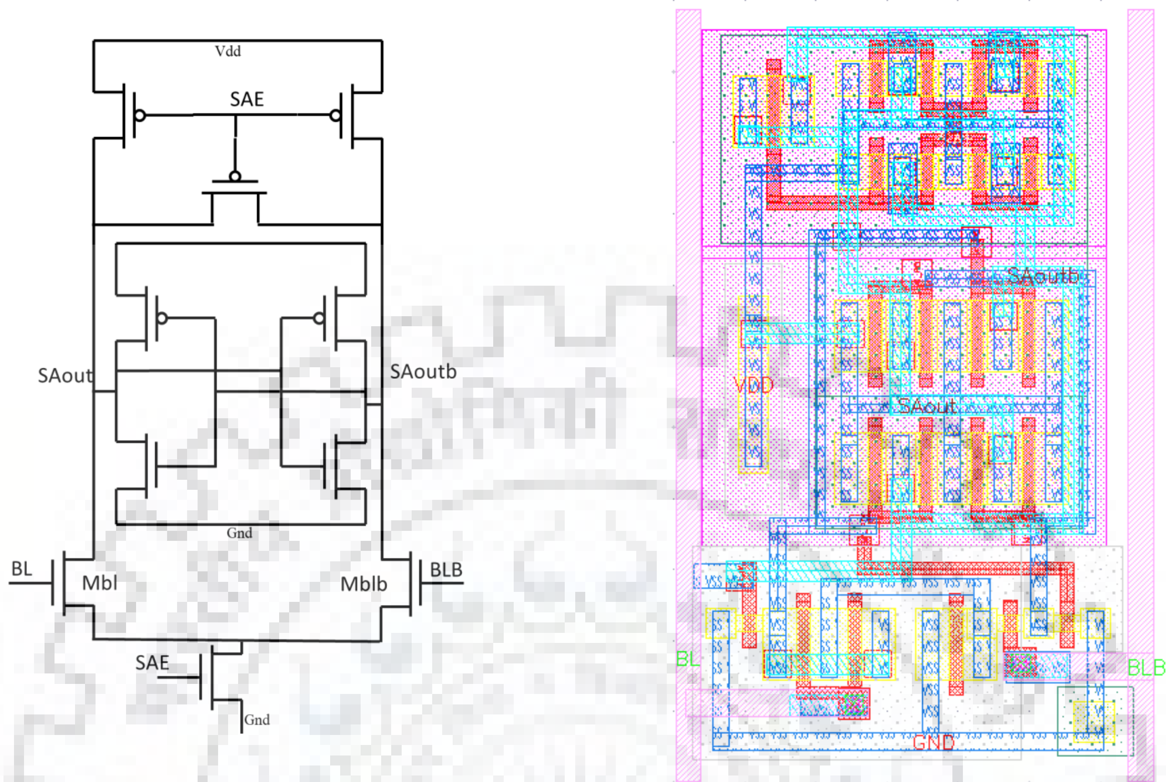

Fig. 2.4 Asymmetric NOR Sense Amplifier circuit schematic and layout

b) NAND Logic operation –

To obtain NAND operation, we need to do skewing of MBL and MBLB transistors but in the opposite way, as we have done in NOR sense amplifier. [3]

i) For the case of 'AB' = '00' ('11') input, BL(BLB) discharges to 0V, but BLB (BL) remains in the same precharge state.
   (1) For the case of '11', BLB starts to discharge and the SA amplifies the voltage difference between BL and BLB, resulting in SAout='0' and SAoutb = '1'.
   (2) Whereas for the case '00', BL starts to discharge, while RBL is at VDD, giving SAout='1' and SAoutb = '0'.

ii) For the case of 'AB' = '01' or '10' both BL and BLB discharges to the same voltage so if the transistor MBLB is deliberately sized bigger compared to MBL, its current carrying capacity increases. Since the current carrying capability of MBLB is more than MBL, SAoutb node discharges faster, and the cross-coupled inverter pair of the SA stabilizes with SAout='1' and SAoutb = '0'.

Thus it can be observed that SAout generates a NAND gate (thus, SAoutb outputs AND gate).

Table 2.3 Sizing of transistors for Sense Amplifier circuits

	NMOS(Inv)	PMOS(Inv)	MBL	MBLB	M_SAE
SANAND $\frac{W}{L}$(um)	$\frac{0.22}{0.18}$	$\frac{4}{0.18}$	$\frac{0.22}{0.18}$	$\frac{2}{0.18}$	$\frac{1}{0.18}$
SANOR $\frac{W}{L}$(um)	$\frac{0.22}{0.18}$	$\frac{4}{0.18}$	$\frac{2}{0.18}$	$\frac{0.22}{0.18}$	$\frac{1}{0.18}$

The schematic and layout drawn for NAND sense amplifier have been shown below:

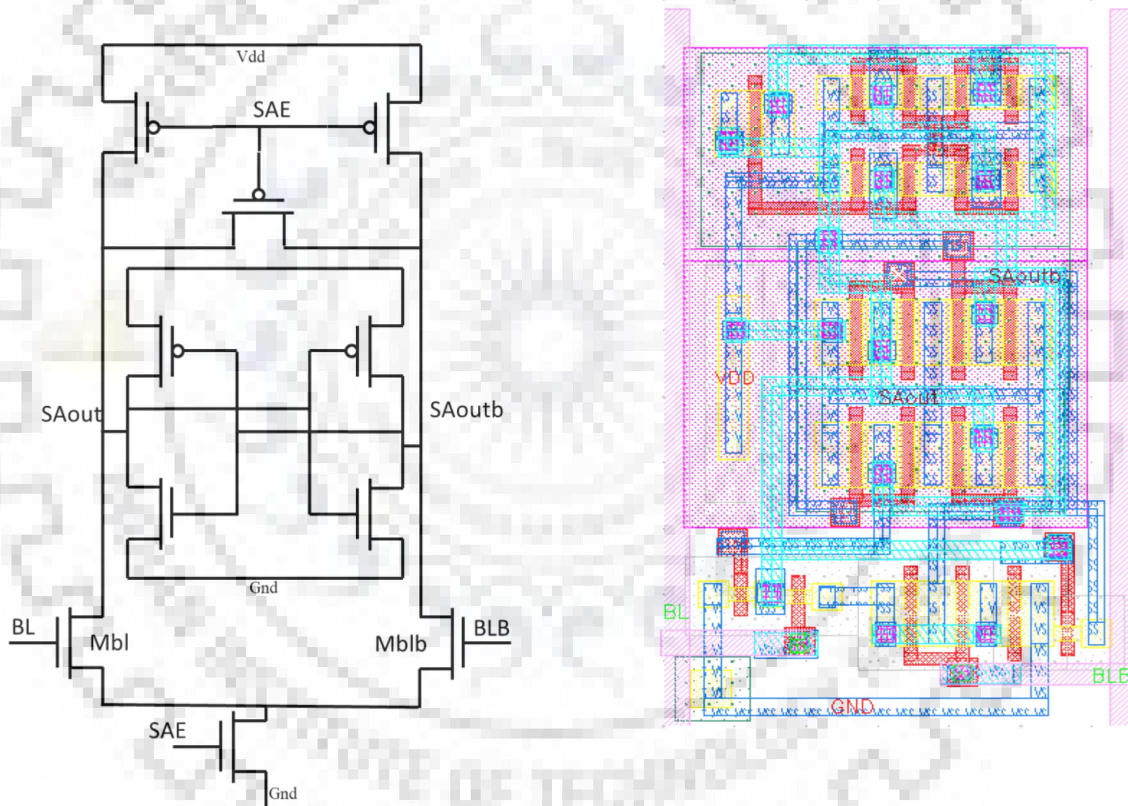

Fig. 2.5 Asymmetric NAND Sense Amplifier circuit schematic and layout

c) XOR Logic operation –

To obtain XOR operation, two Sense Amplifiers (one with NAND operation and another with NOR operation) will be used parallel. The SAout result of NAND sense amplifier and SAoutb result of NOR sense amplifier will be inputted to another NAND logic gate to get XNOR output. Similarly, the SAout result of NOR sense amplifier and SAoutb result of NAND sense amplifier will be inputted to NOR logic gate to obtain XOR output.

CHAPTER 3

PROPOSED SENSE AMPLIFIERS FOR IN-MEMORY COMPUTATION

The Sense amplifier circuit being proposed by reference [3] able to do AND, OR, NAND, NOR, XOR and XNOR operations. Reference [3] has given a different circuitry for NAND and NOR operations which makes the overall memory size bigger and thus more power consumption. In this work, a novel Sense Amplifier circuit is going to be proposed which can perform both NAND and NOR Boolean operations with a single circuitry. Along with that, a sense amplifier is also being proposed which can perform an independent XOR operation.

3.1 NANDNOR Sense Amplifier :

The idea for a combined Sense Amplifier was explored due to the problems being faced with the already proposed circuit as memory sizing, output delay, and power consumption, etc.

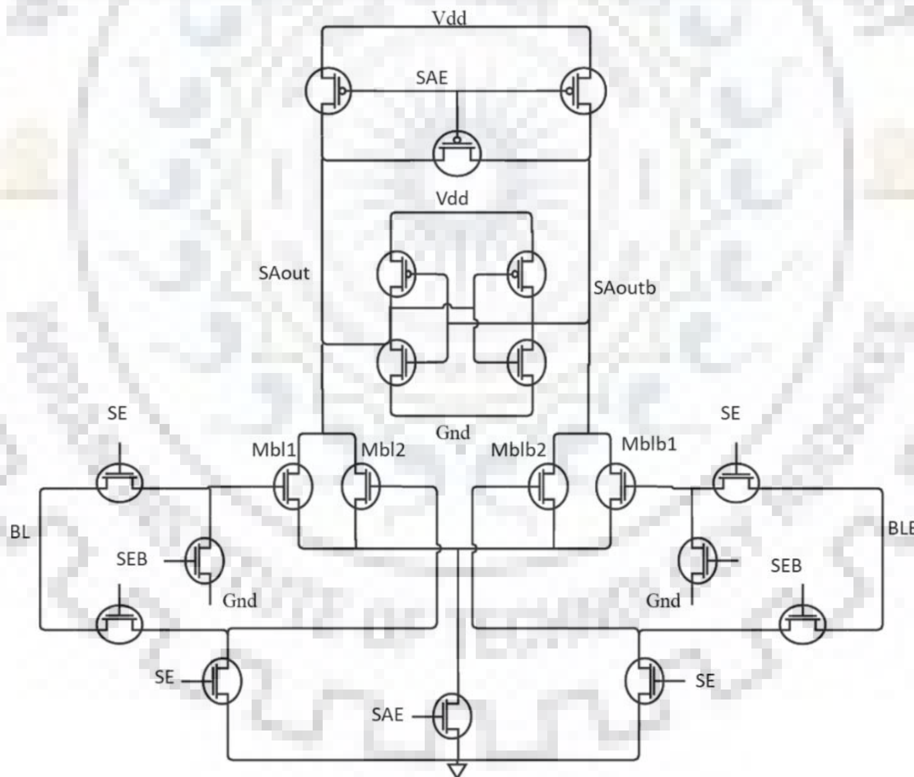

Fig. 3.1 Proposed NandNor Sense Amplifier circuit

In this circuit, a CMOS circuitry is particularly there to select the operation user wants to be done. In both bitline and bitlinebar side, we have symmetric multiplexer kind of circuitry which will select the operation going to be performed that is either NAND or NOR.

As shown in Fig 3.1, SE and SEbar signals are there to select the particular operation. Once the operation is being selected, the BL and BLB lines will get connected to the respective MBL and MBLB CMOS devices which are responsible for discharging of SAout/SAoutb.

The layout drawn for the same design is being shown below –

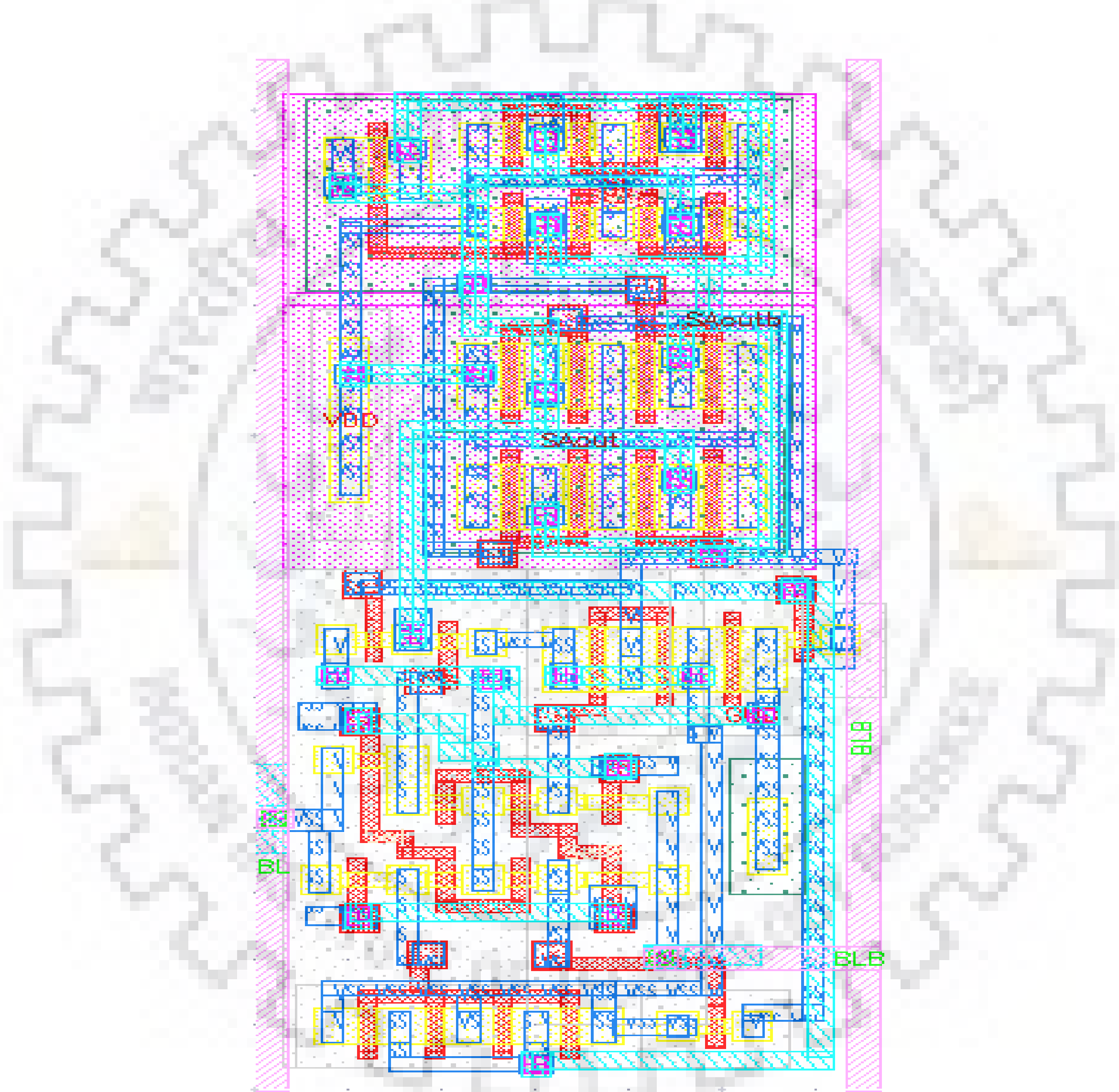

Fig. 3.2 Proposed NandNor Sense Amplifier layout

a) **NAND operation**

Initially, the BL/BLB are precharged to Vdd and similarly, SAout/SAoutb will also be precharged to supply voltage for SAE = gnd. Let us consider that two SRAM cells of memory array have stored A and B. For NAND case, in proposed sense amplifier, once SAE is at supply voltage, SE signal will be at supply voltage i.e. 1.8V and SEB will be opposite of that so MBL1 and MBLB1 transistors will get BL/BLB signal at their gate input while MBL2 and MBLB2 transistors will be connected to ground through transistors.

Different cases for NAND operation -

i) Case '00' – If AB = '00' then BL will discharge to almost ground voltage and BLB will be at VDD. Thus MBLB1 will get turned on and thus SAout will remain at VDD and SAoutb will get discharged to 0 volts.

ii) Case '01' – If AB = '01' then BL and BLB both will get discharged to almost half of supply voltage so both MBL1/MBLB1 will be turned ON but as the sizing of MBLB1 is greater than MBL1 so SAoutb will get discharged with higher current to near ground voltage and due to back to back circuitry in sense amplifier circuit, SAout will remain at Vdd voltage.

iii) Case '10' – If AB = '10' then BL and BLB both will get discharged to almost half of supply voltage so both MBL1/MBLB1 will be turned ON but as the sizing of MBLB1 is greater than MBL1 so SAoutb will get discharged with higher current to near ground voltage and due to back to back circuitry in sense amplifier circuit, SAout will remain at Vdd voltage.

iv) Case '11' – If AB = '11' then BLB will discharge to near ground voltage and BL will be at VDD. Thus MBL1 will get turned on and thus SAoutb will remain at VDD and SAout will get discharged to 0 volts.

b) **NOR operation**

Initially, the BL/BLB are precharged to Vdd and similarly, SAout/SAoutb will also be precharged to supply voltage for SAE = gnd. Let us consider that two SRAM cells of memory array have stored A and B. For NOR case, in proposed sense amplifier, once SAE is at supply voltage, SE signal will be at ground and SEB will be at supply voltage so MBL2 and MBLB2 transistors will get BL/BLB signal at their gate input while MBL1 and MBLB1 transistors will be connected to ground through transistors.

Different cases for NOR operation -

i) Case '00' – If AB = '00' then BL will discharge to almost ground voltage and BLB will be at VDD. Thus MBLB1 will get turned on and thus SAout will remain at VDD and SAoutb will get discharged to 0 volts.

ii) Case '01' – If AB = '01' then BL and BLB both will get discharged to almost half of supply voltage so both MBL1/MBLB1 will be turned ON but as the sizing of MBL1 is greater than MBLB1 so SAout will get discharged with higher current to near ground

voltage and due to back to back circuitry in sense amplifier circuit, SAoutb will remain at Vdd voltage.

iii) Case ‘10’ – If AB = ‘10’ then BL and BLB both will get discharged to almost half of supply voltage so both MBL1/MBLB1 will be turned ON but as the sizing of MBL1 is greater than MBLB1 so SAout will get discharged with higher current to near ground voltage and due to back to back circuitry in sense amplifier circuit, SAoutb will remain at Vdd voltage.

iv) Case ‘11’ – If AB = ‘11’ then BLB will discharge to near ground voltage and BL will be at VDD. Thus MBL1 will get turned on and thus SAoutb will remain at VDD and SAout will get discharged to 0 volts.

Table 3.1 Sizing of transistors for Proposed NANDNOR Sense Amplifier

SA_NAND NOR	NMOS (Inverter)	PMOS (Inverter)	M_SAE	MBL1 MBLB2	MBL2 MBLB1	All other transistor s
$\frac{W}{L}$(um)	$\frac{0.22}{0.18}$	$\frac{1}{0.18}$	$\frac{1.2}{0.18}$	$\frac{0.22}{0.18}$	$\frac{2}{0.18}$	$\frac{0.22}{0.18}$

3.2 XOR Sense Amplifier :

Initially, the BL/BLB are precharged to Vdd and similarly, SAout/SAoutb will also be precharged to supply voltage for SAE = gnd. Let us consider that two SRAM cells of memory array have stored A and B. For XOR case, in the proposed sense amplifier, once SAE is at supply voltage the sensing operation will be enabled.

Different cases for XOR logic operation -

a) Case ‘00’ – If AB = ‘00’ then BL will discharge to almost ground voltage and BLB will be at VDD. Because of BL and BLB inputs at both SAout and SAoutb discharging path, SAout and SAoutb will be stuck at the metastable state and reached to a voltage value approximately Vdd/2.

b) Case ‘01’ – If AB = ‘01’ then BL and BLB both will get discharged to almost half of supply voltage so all the four transistors in discharge path will be turned on but as one of the left side transistor is sized a bit higher than all other transistors so SAout will get discharged to ground and SAoutb will remain at Vdd.

c) Case ‘10’ – If AB = ‘10’ then BL and BLB both will get discharged to almost half of supply voltage so all the four transistors in discharge path will be turned on but as one of the left side transistor is sized a bit higher than all other transistors so SAout will get discharged to ground and SAoutb will remain at Vdd.

d) Case ‘11’ – If AB = ‘11’ then BLB will discharge to near ground voltage and BL will be at VDD. Because of BL and BLB inputs at both SAout and SAoutb discharging path, SAout and SAoutb will be stuck at the metastable state and reached to a voltage value approximately Vdd/2.

Now, to get XNOR and XOR results from SAout and SAoutb respectively, two inverters will be used at the output of Sense amplifier. A High-skewed inverter was being used at SAout port to get XNOR results. From the basics of High-skewed inverter, NMOS will be stronger then PMOS so discharging will be dominative thus for a large input voltage range, the output will be 0V. Thus for case 00/11, output will go to 0V. Similarly, at SAoutb port of XOR SA, a Low-skewed inverter will be used. As the PMOS will be dominating so charging will become fast and thus for a large range of input voltages, output voltage will get charged upto supply voltage. Thus we have obtained both XNOR and XOR outputs.

Table 3.2 Sizing of transistors for Proposed XOR Sense Amplifier

SA_XOR	PMOS (Inverter)	NMOS (Inverter)	M_SAE	MBL	All other transistors
$\frac{W}{L}$(um)	$\frac{4}{0.18}$	$\frac{0.22}{0.18}$	$\frac{1}{0.18}$	$\frac{4}{0.18}$	$\frac{3}{0.18}$

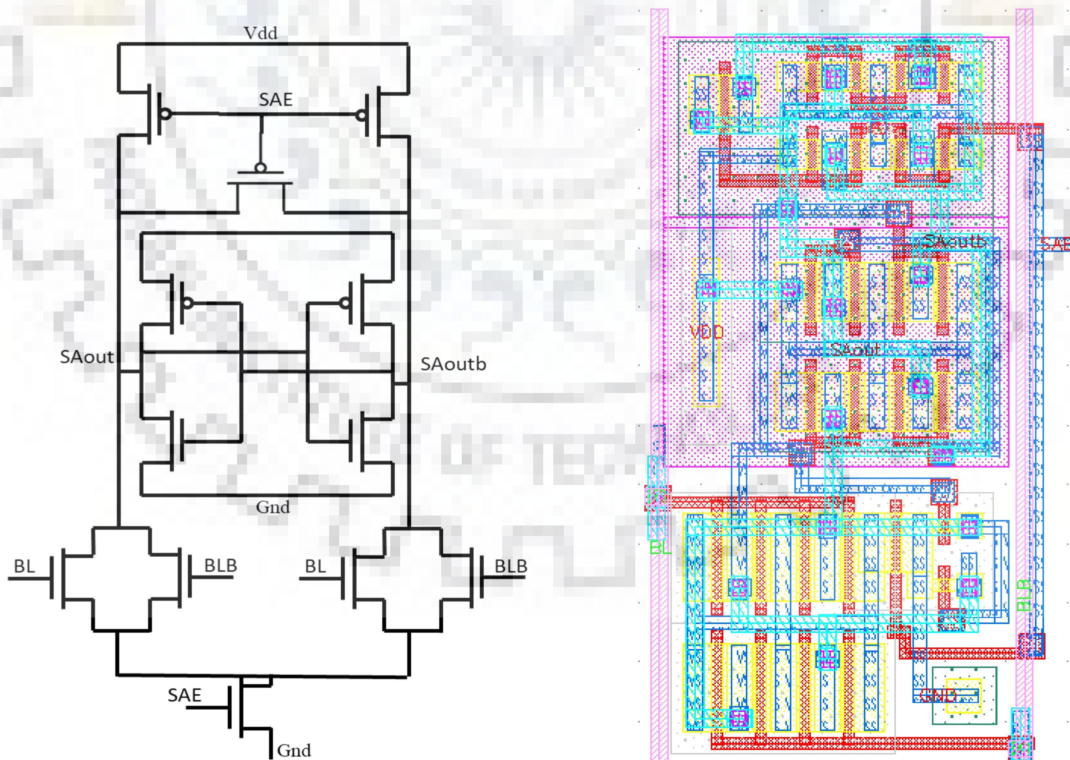

Fig 3.3 XOR Sense Amplifier Schematic and Layout

CHAPTER 4

SIMULATION METHODOLOGY, RESULTS AND COMPARISION

From the day memory came to picture, the main aim to be achieved was memory array area and power consumption over performance and speed. So for the work being proposed, we have taken our motivation as low power consumption requirement in computation in-memory.

As we have used SRAM 6T cell to perform all logic operations on multiple bits, so simultaneous enabling will not be possible otherwise WL1 and WL2 will introduce read disturbs due to possible short-circuit paths. Hence, we employ a sequentially pulsed WL technique as a workaround.

4.1 Computation using Conventional design

To perform the bitwise logic operation, two asymmetric sense amplifiers will be used for both NAND and NOR operation separately and to obtain XOR operation, the AND/NAND and OR/NOR outputs will be combined by using additional NOR gate and NAND gate.

As shown in fig. 4.1, two SAs in parallel (one with MBL up-sized, SANOR, and one with MBLB up-sized, SANAND) enable bit-wise AND/NAND and OR/NOR logic gates. For XNOR operation, NAND gate will be used to combine 'SAout' results of SANOR and SANAND and for XOR operation, NOR gate will be used to combine 'SAout' result of SANOR and 'SAoutb' result of SANAND.

At first, the first SRAM cell will access bitlines by turning on WL1 line for a proper period of time and then the second SRAM cell will access bitlines by turning on WL2 line.

The WL pulse duration is chosen such that with the application of one WL pulse, BL/BLB drops to about ~VDD/2. If bits 'A' and 'B' both store '0' ('1'), BL (BLB) will finally discharge to 0V after the two consecutive pulses, whereas BLB (BL) remains at VDD as shown in fig 4.2. On the other hand, for cases where 'AB' = '10' and '01', the final voltages at BL and BLB would be the same (~VDD/2), approximately. Thus, for the cases '01' and '10' both BL and BLB would have a voltage ~VDD/2, while for '00' BL would be lower than BLB by ~VDD and for the case of '11,' BLB would be lower than BL by ~VDD.

Thus, using two asymmetric sense amplifiers and NAND and NOR logic gates, we can obtain all basic boolean logic operations.

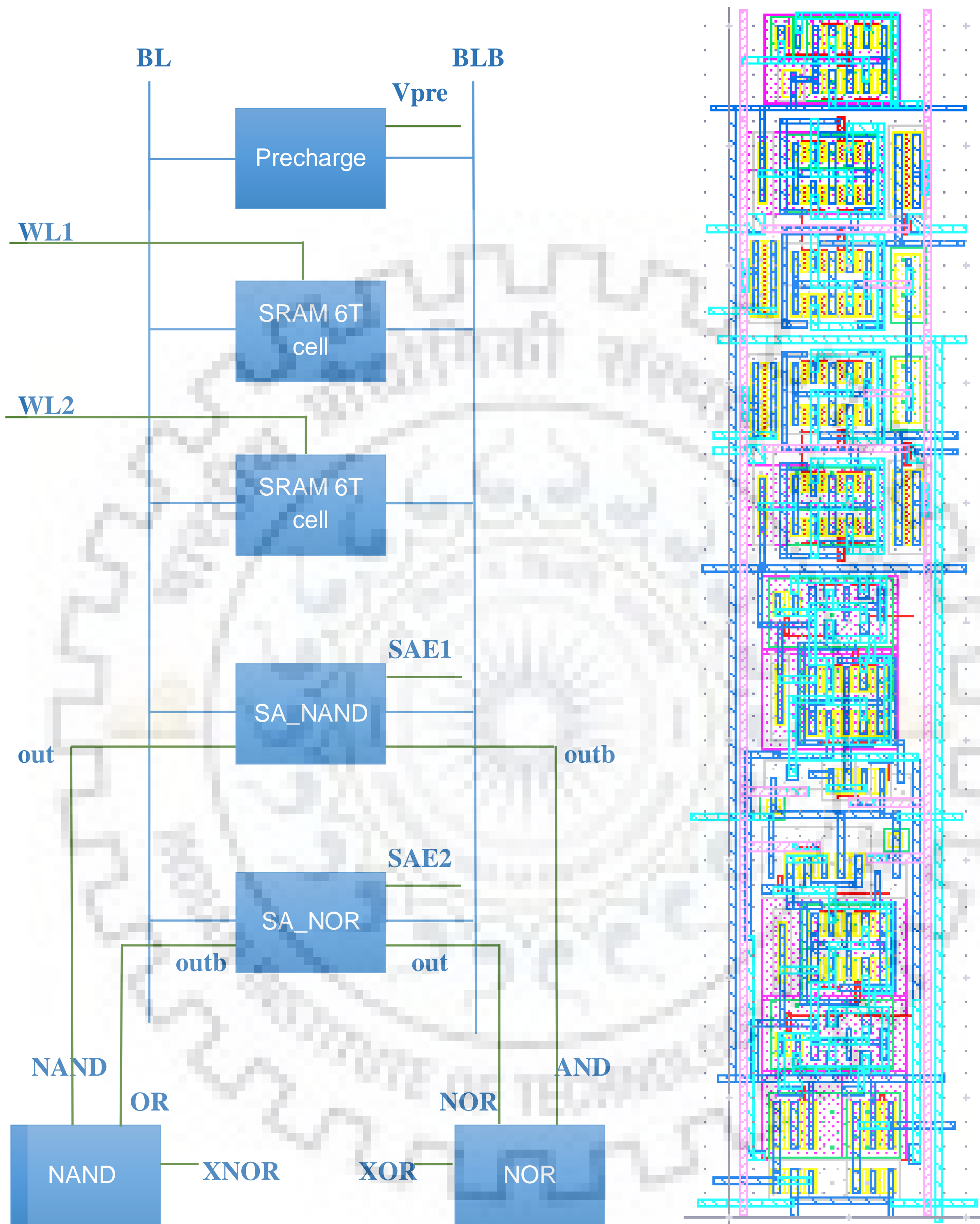

Fig.4.1 Schematic and Layout of the single column for a memory array using two 6T SRAM cells and conventional Sense Amplifiers for logic operation NAND-NOR-XOR

Simulation Results:

All the simulations are being done using 180nm scl technology. As shown in fig. 4.2, the bit and bit bar lines get discharged based on the time for which word line has been activated.

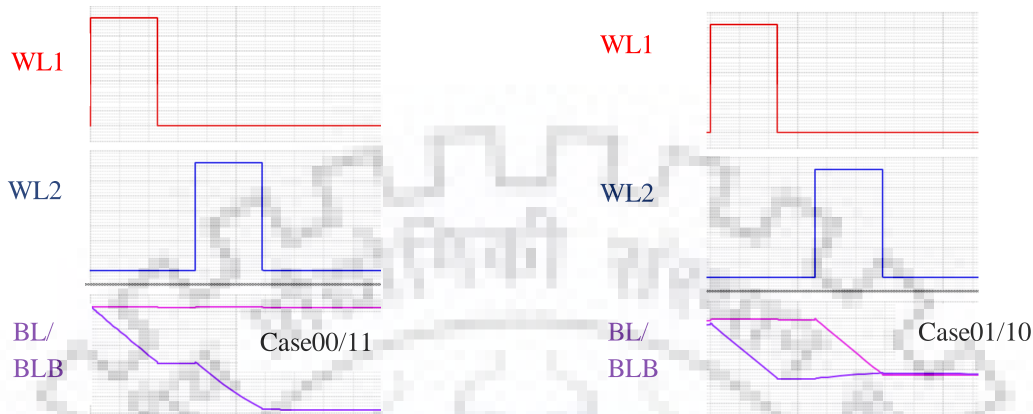

Fig.4.2 Timing diagram for any logic operation for in-memory computation

By using the above circuitry, post layout simulation has been done for bitwise logic operations. To study the effect of different process corners and to know the worst corner, a corner analysis was done. We also performed parametric simulations across all the corners under Vth, temperature and supply voltage variations for all possible cases i.e. 00, 01, 10 and 11 for NAND, NOR and XOR logic operation. Monte Carlo simulation was being performed for settling time of the design across worst process corners under Vth variations. The distribution of settling time is plotted under ±30mV threshold voltage variations at ss corner.

4.2 Computation using Proposed Design

In the proposed work, instead of two separate NAND and NOR sense amplifier, a single NANDNOR sense amplifier was being used and also an idea for XOR sense amplifier is being proposed. This idea has reduced both area as well as power consumption for computation in-memory. The schematic and layout for a memory design, performing logic operation using two SRAM 6T cells has been shown in fig. 4.3.

Based on the proposed design block diagram, the NAND or NOR operation will be performed based on the input signal given at Sel and Selb ports. The BL and BLB (dis)charging concept remain same as conventional design i.e. based on the pulse width of WL.

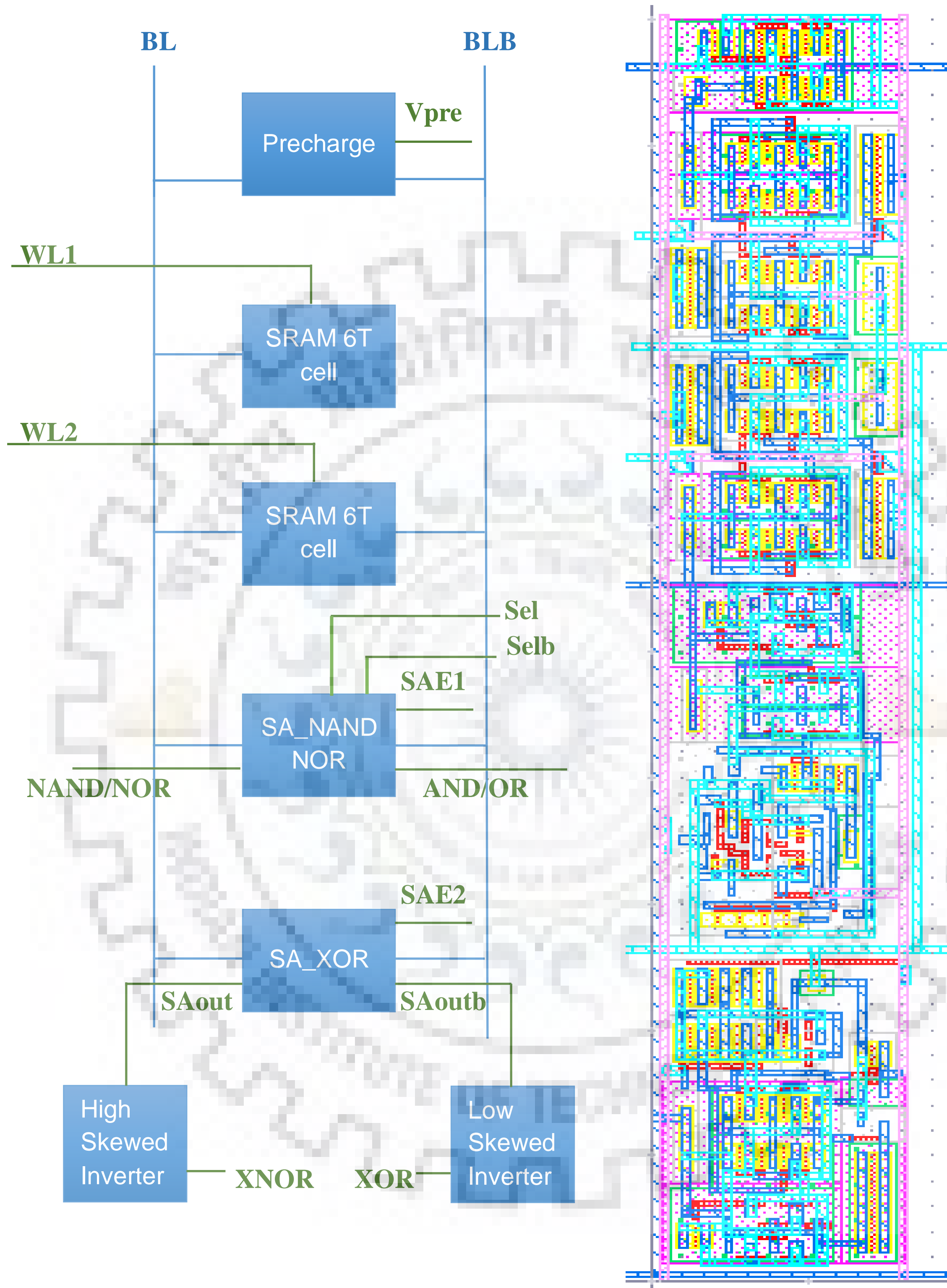

Fig.4.3 Schematic and Layout of the single column for a memory array using two 6T SRAM cells and **Proposed** Sense Amplifiers for logic operation NAND-NOR-XOR

Simulation Results:

By using the above circuitry, post layout simulation has been done for bitwise logic operations. To study the effect of different process corners and to know the worst corner, a corner analysis was done. We also performed parametric simulations across all the corners under Vth, temperature and supply voltage variations for all possible cases i.e. 00, 01, 10 and 11 for NAND, NOR and XOR logic operation. Monte Carlo simulation was being performed for settling time of the design across worst process corners under Vth variations. The distribution of settling time is plotted under ±30mV threshold voltage variations at ss corner.

At first the functionality of the proposed design was being checked by simulation results as shown in fig. 4.4. As shown in the figure, with proper BL discharging, NAND, NOR and XNOR results has been obtained.

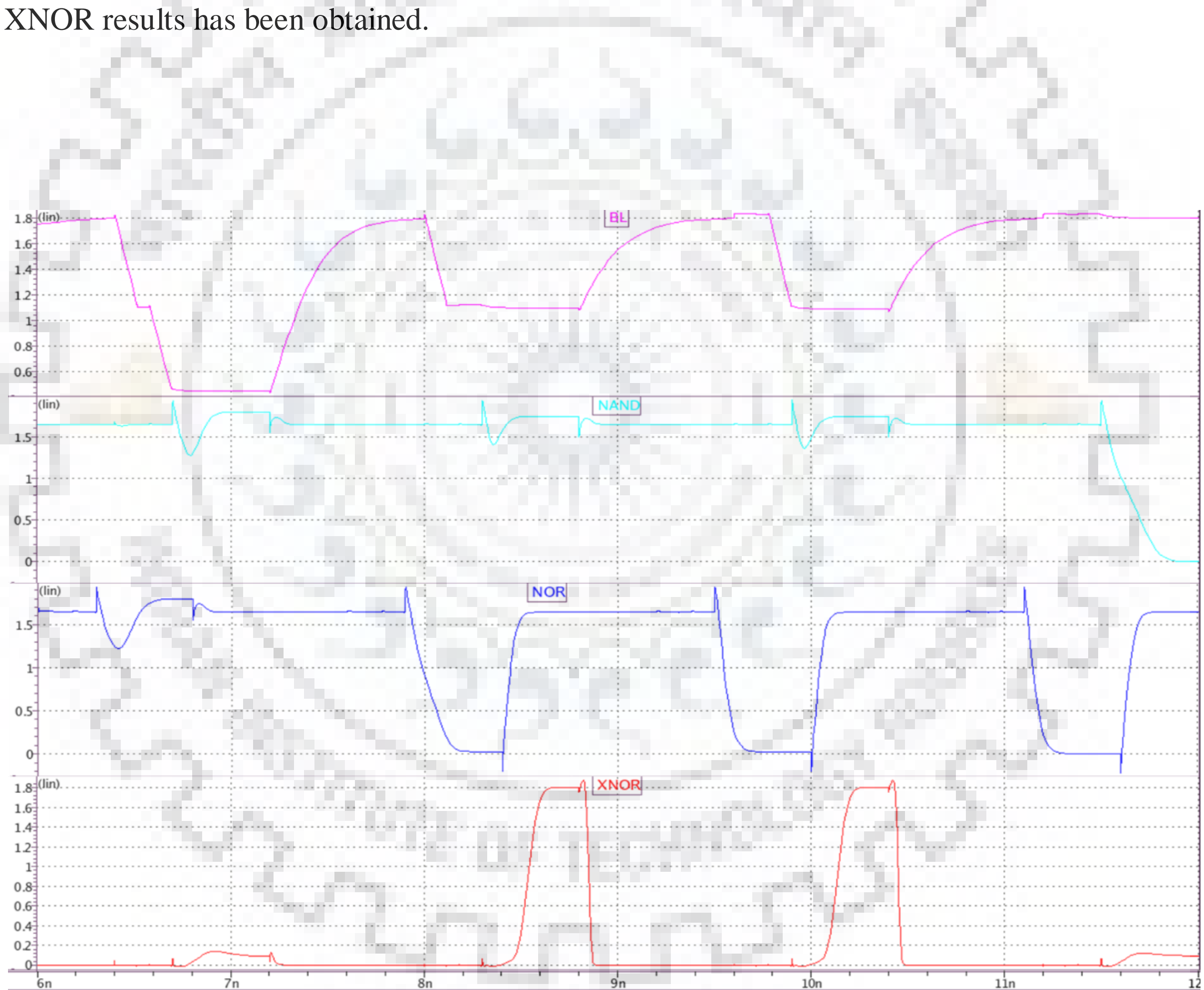

Fig. 4.4 Post layout simulation results for proposed design

As we know the basic operation of a memory is storage and for that sense amplifier is being used to read a bit in memory cell. So the proposed sense amplifier is even able to do read operation and the results for read operation has been shown in figure 4.5.

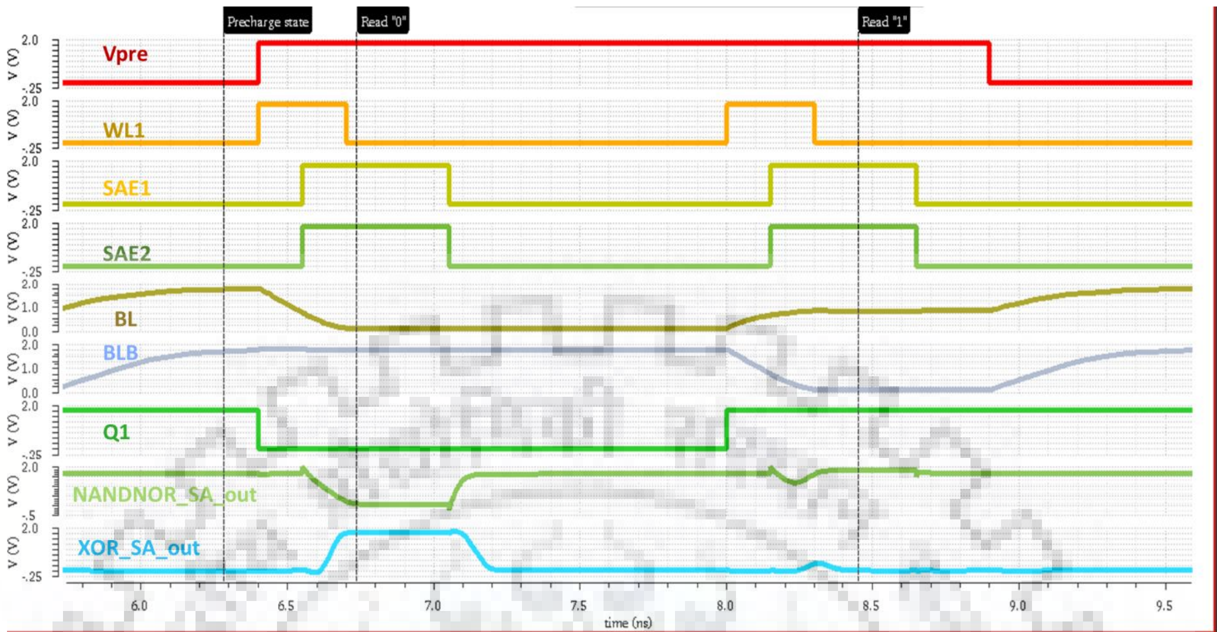

Fig.4.5 Simulation results for read operation using proposed sense amplifier

After getting satisfactory results for the optimum sizing of all transistors, we have done a corner analysis to check robustness of our design at all different corners. The results for corner analysis has been shown in fig. 4.6

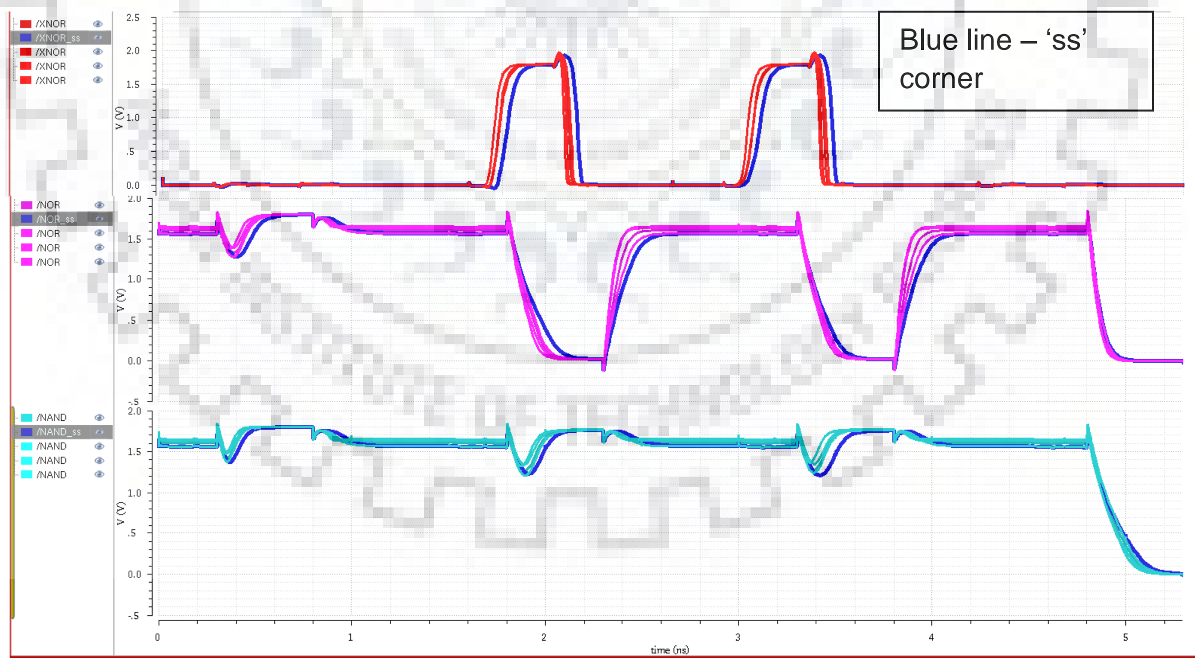

Fig. 4.6 Post layout corner analysis results for proposed design

From corner analysis results, we obtained the worst corner to be ‘ss’ corner for our design. To compare proposed design results with the reference design, Monte carlo simulations across all corners for threshold voltage variations has been done for the settling time of all possible cases (00/01/10/11) for NAND, NOR and XOR operation. Settling time is defined as the absolute time required for the response to reach and remain within a given tolerance band. For this case, 5% of tolerance was chosen.
From fig. 4.7, the mean latency of all the three operations has been compared. The results shows that the worst case latency of NAND and XOR operations for proposed design is even better than the reference design. But for NOR operation, the proposed design mean latency is better than the reference design but worst case delay is more for proposed design. As per the aim of our work, we focused on area and power consumption more than performance.

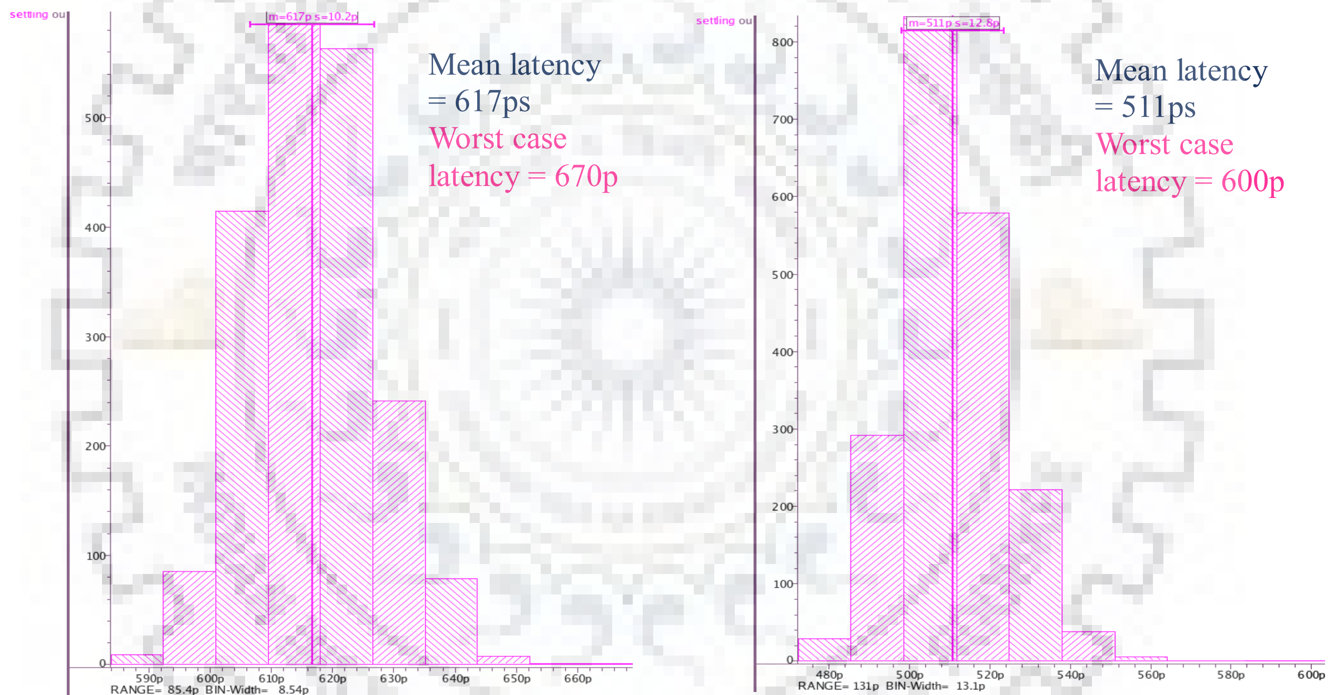

Fig 4.7 MC simulation results with Vth variation a) for conventional NAND operation b) for proposed NAND operation

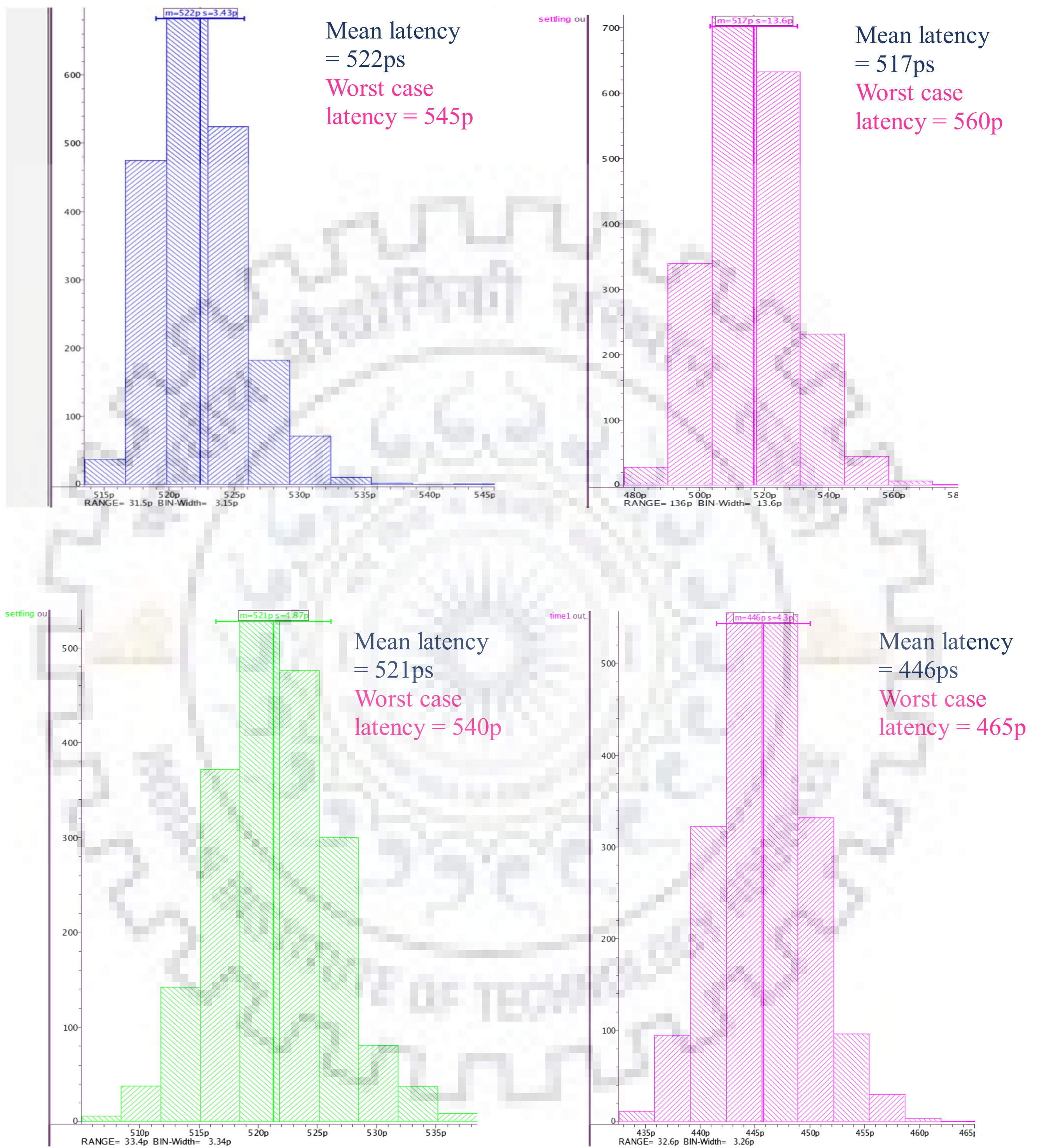

Fig 4.8 MC simulation results with Vth variation a) for conventional NOR operation b) for proposed NOR operation c)for conventional XOR operation d) for proposed XOR operation

Table 4.1 Settling time comparison for NAND, NOR and XOR logic operation

CORNER	DESIGN	NAND (ps)		NOR(ps)		XOR(ps)	
		Case 00/11	Case 01/10	Case 00/11	Case 01/10	Case 00/11	Case 01/10
tt	Conventional Sense Amplifier	597	559	522	530	532	521
	Proposed Sense Amplifier	483	538	517	489	454	446
ss	Conventional Sense Amplifier	627	577	620	570	580	652
	Proposed Sense Amplifier	517	565	638	573	471	668
sf	Conventional Sense Amplifier	597	566	543	537	527	594
	Proposed Sense Amplifier	488	529	547	546	455	608

The bottom-line of corner analysis was that we obtained “ss” corner as the worst corner and the second worst corner was “sf” corner thus the MC analysis was being done on the three corners “tt”, “ss” and “sf” and the timing results are given in a tabular form in table 4.1
From table 4.1, it is being concluded that NOR and XOR 00/11 cases at “ss” and “sf” coare better for the reference design while all other operations were good for proposed design. NAND, NOR and XOR results has been plotted under 30mV threshold voltage variation at two different temperatures and 10% supply voltage variations. The parametric analysis was done to check effect of threshold voltage variations on the proposed design. The parametric analysis result for NAND operation has been shown in fig. 4.9
Fig. 4.9 shows output voltage variation with five different voltages which are ±10% varying with the nominal supply voltage. We have obtained that the functionality of the proposed design does not get affected with Vth, temperature and supply voltage variations but it has affected performance of the design such as it has increased the delay.

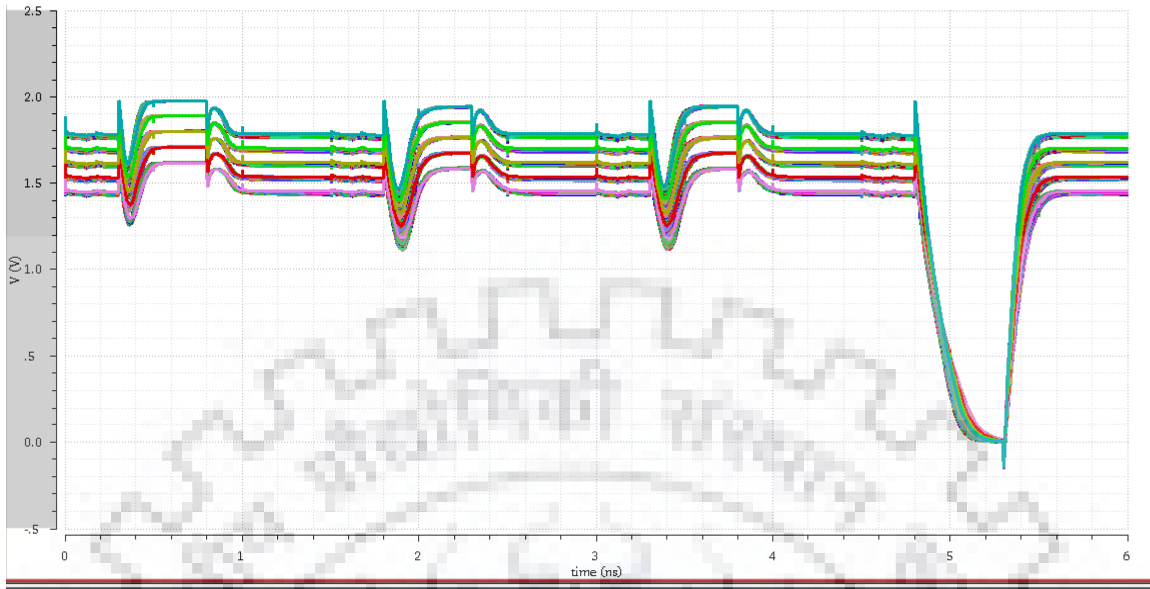

Fig 4.9 Parametric simulation results with 30mV Vth variation at two temperatures (25°C and 80°C) and ±10% variation in supply voltage for NAND operation

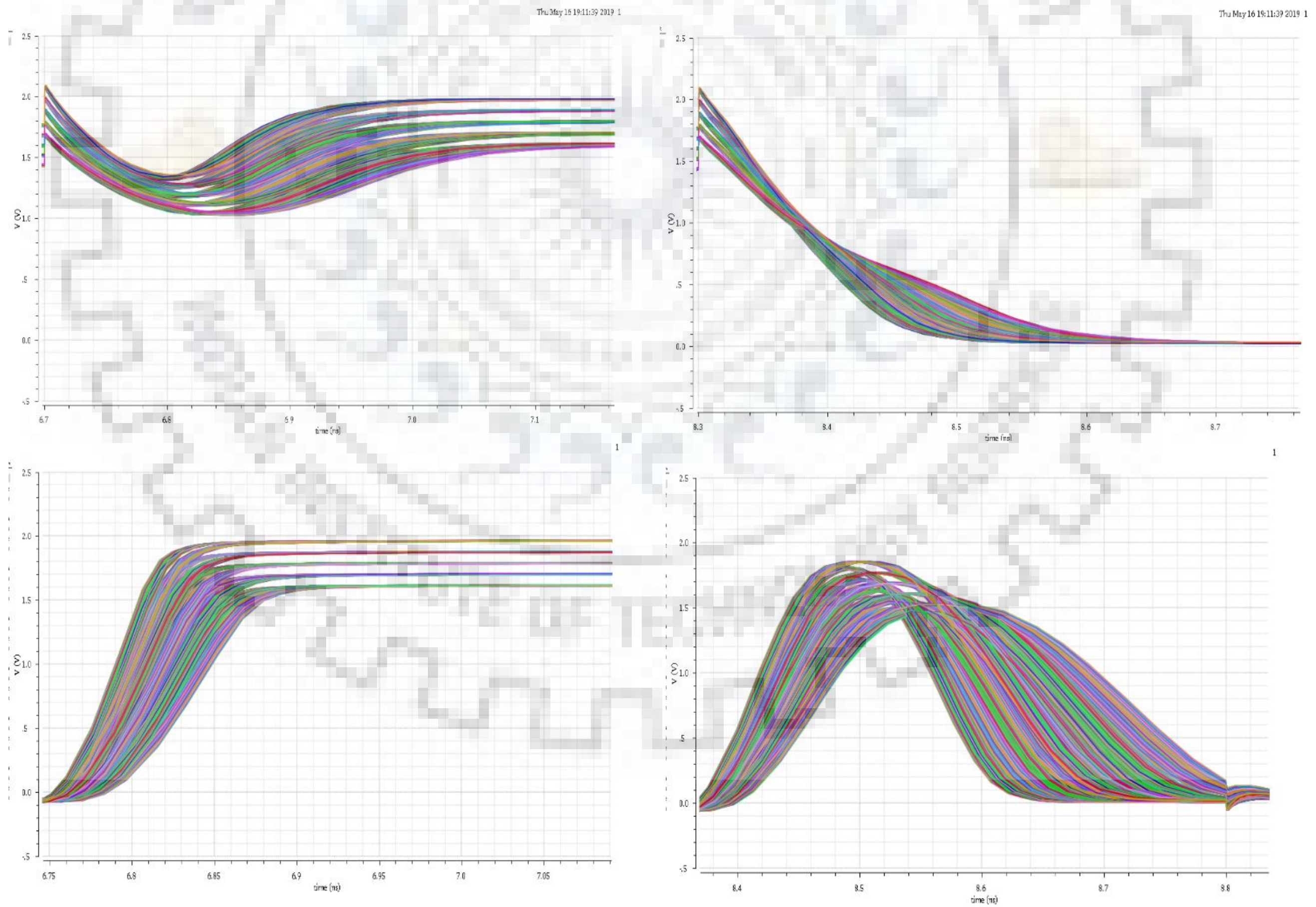

Fig 4.10 Parametric simulation results with 30mV Vth variation at two temperatures (25°C and 80°C) and ±10% variation in supply voltage a) NOR case 00 b) NOR case 10 c) XOR case 00 d) XOR case 10

The above results are zoomed out images of parametric simulation to see the clear difference with different variations. As the parametric simulation was being done with ±10% variation in supply voltage so we can see five different voltages near around 1.8V.

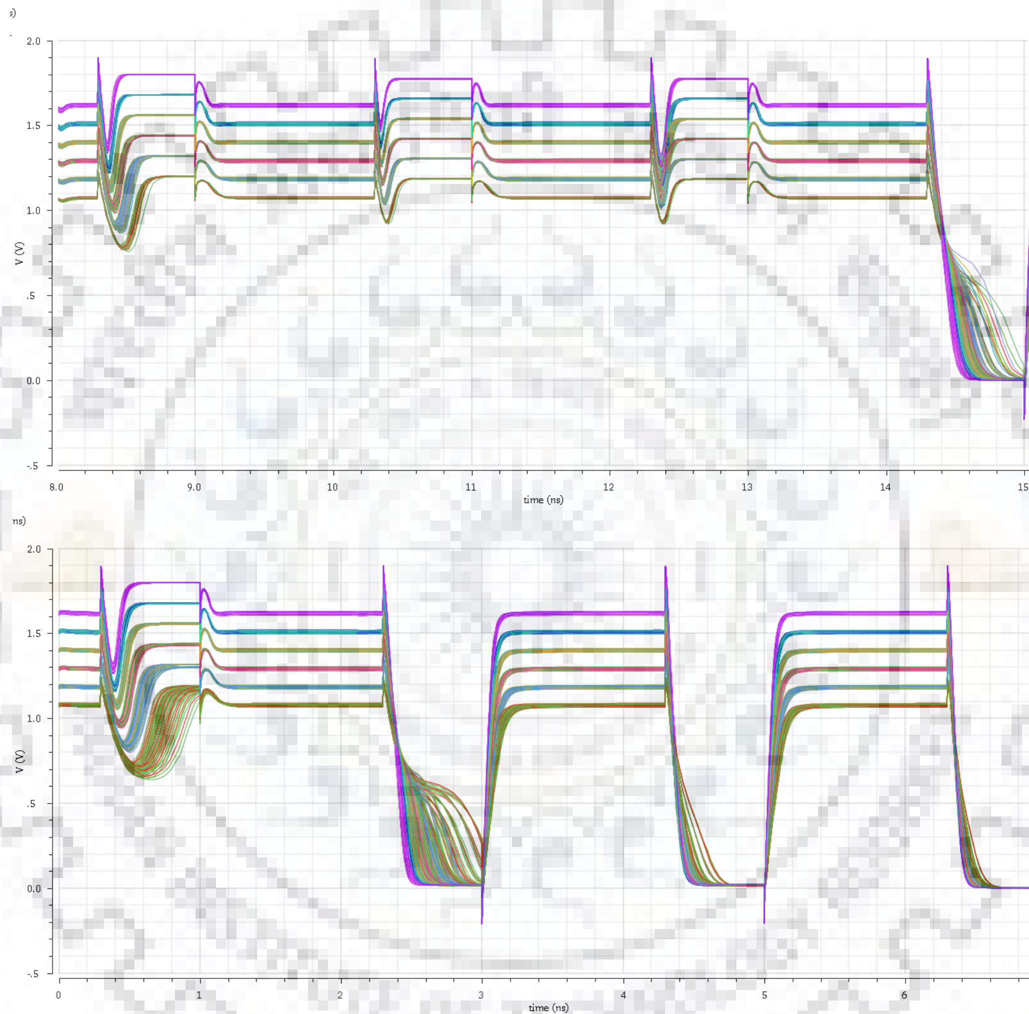

Fig 4.11 Parametric simulation results with 30mV Vth variation at two temperatures (25°C and 80°C) and supply voltage variations from 1.2 to 1.8 V a) NAND b) NOR

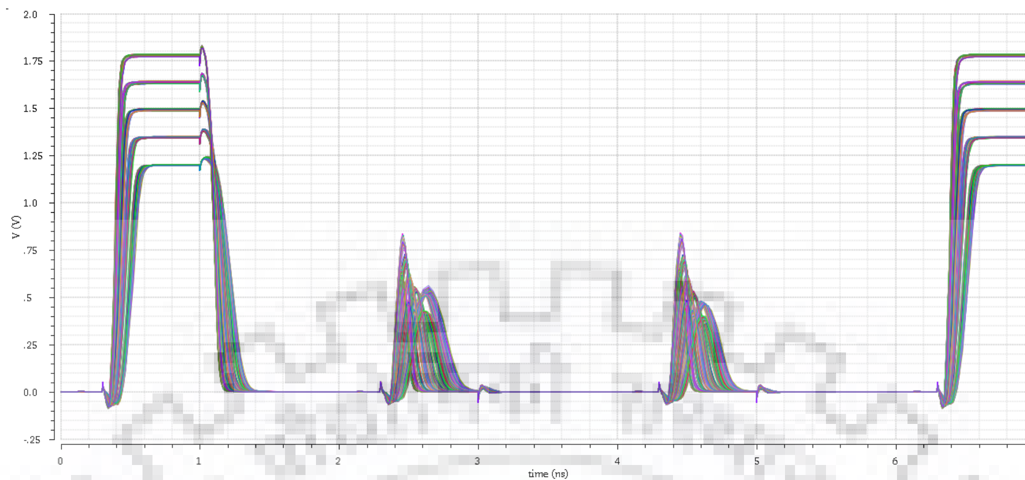

Fig 4.12 Parametric simulation results with 30mV Vth variation at two temperatures (25°C and 80°C) and supply voltage variations from 1.2 to 1.8 V for XOR

Along with the ±10% variation in supply voltage, another parametric simulation was being done to find out the minimum supply voltage at which our design can withstand. For that we have done parametric simulation similar as before just with a different range of supply voltage. We have obtained that the supply voltage range from 1.2 to 1.8 was working perfectly fine while we were doing pre layout simulations but it worked only for 1.35 to 1.8 volts for post layout simulations only for NOR operation otherwise it worked well for the other two operations.

It resembles that the transistors sizes has to be done in a different way to let the design work for even lower supply voltages. And as the results concluded that the reference design was also not working below 1.2 Volts of supply voltage.

Fig. 4.11 shows the parametric simulation results for all the logic operations for the proposed design.

Thus from the above simulation results, we concluded that 1.2 Volts is the minimum supply voltage at which NAND and XOR operations will be functionally correct but 1.35 Volts will be the minimum supply voltage for NOR operation to be functionality correct.

All the simulation results we have shown till now, shows a performance comparison of our design with the reference design but now we will see an area and power consumption comparison as well for both the design.

Table 4.2 A comparison table for proposed design and referenced design

DESIGN	OPERATIONS	AREA (um^2)	LATENCY (ns)	POWER CONSUMPTION (nW)
Conventional Sense Amplifier	NAND NOR	8.96×44.55 **399.168**	**0.617**	**0.8107**
	XOR	8.96×50.7 **454.272**	**0.616**	**1.1471**
Proposed Sense Amplifier	NAND NOR	7.975 × 37.18 **296.510**	**0.583**	**0.6325**
	XOR	7.975 × 35.45 **282.713**	**0.592**	**0.9950**

Table 4.2 compares the proposed design with the referenced design. At first it shows comparison of layout area for a single column for NAND, NOR and XOR operations. We can see that for both proposed NAND-NOR and XOR sense amplifier designs, we have got almost **34.622%** of higher area requirement for NAND-NOR and **60.683%** of area increment for XOR referenced design which seems to be lot more and will even increase for a full memory design. For the case of latency comparison, we can see that it does not get affected much i.e. just **4-6%** of betterment. For the case of power consumption, almost **28%** of more power consumption will be there for NAND-NOR operation and similarly almost 15% of more power consumption be there for XOR operation for two bits using reference design.

CHAPTER 5
CONCLUSION AND FUTURE WORK

Von-Neumann machines have been used as standard computer architecture. However, it has energy and throughput limitations. The recent phase of data-intensive applications like artificial intelligence, image recognition, cryptography, etc. requires novel computing architecture to fulfill efficient energy and throughput requirements. 'In-memory' computing is one of the best promising approaches that could sustain the throughput and energy requirements for future computing platforms.

5.1 Conclusion

We have seen that the logic operations such as NAND, NOR, and XOR can be obtained by the proposed sense amplifier circuitry. As we know, in case of memory array, area and power are two crucial parameters to be take care of. So we did propose a sense amplifier design separate for NAND-NOR operation and then another sense amplifier for XOR operation. So now unlike the reference design, the XOR operation will not be dependent on NAND and NOR operation. So we can have separate column for both NAND-NOR operations and XOR operation with comparatively lesser area and power consumption as well. Along with area and power consumption, our design is better from performance point of view as well. To check the robustness of sense amplifier design, simulation at all corners have been performed using ScL transistor models.

So, what we have concluded from our proposed idea is given below –

a) The proposed sense amplifier design is functionally correct.
b) All corner analysis results shows the robustness of the proposed design.
c) The better results of settling time i.e. latency for NAND, NOR and XOR operations was being proven with MC analysis with process variation. Even the worst case settling time is better than the reference design results for NAND and XOR operations.
d) The parametric analysis results with Vth and supply voltage variation at two different temperature shows the reliability of the design.
e) The optimum sizing of all the transistors were being done to be useful for supply voltage range of 1.2 – 1.8 volts.
f) There will be no dependency of XOR operation on NAND and NOR operation.
g) Our design has given better area and power dissipation results as well.

5.2 Future Work

To evaluate a system level implementations using described SRAM memory array instead of conventional SRAM memory array, any basic ALU operation through processor instruction can be compared with it. By utilizing such in-memory computations, we expect a reduction in energy expensive data movements over the bus between the processor and the memory blocks. On an architecture level, using a memory array having computation functionality provides us an idea to avoid the delay time required in "operation fetch" in typical von-Neumann architecture.

a) To perform any ALU function in the processor, there used to be three main operations that is "FETCH-DECODE-EXECUTE".
   For a single instruction, the cycle will be like the first processor will read data from a particular memory address by decoding that instruction and then that will be stored in the temporary registers of ALU for the operation to be performed. Again the data will be stored in a particular memory address.
   Now with the new nonstandard von-Neumann architecture, we do not need a transfer of data from memory to processor for computation which will save lots of our time. For example

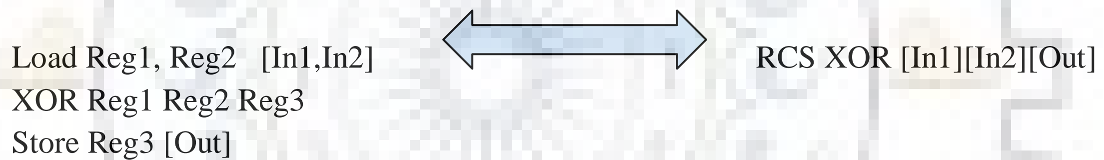

Fig 5.1 Comparison of in-memory instruction with conventional processor instructions

b) To go further, the most basic operation in ALU is an addition. Using In-Memory computation, we can design architecture to perform addition inside memory with lesser time and energy consumption.
   We have separately checked logic operations inside memory and we are being able to do write operation as well so now we can combine both functions to get complex circuits but care must be taken with timing constraints.
   So there will be a lot of future possibilities with the proposed design.

REFERENCES

[1] J. Backus, “Can programming be liberated from the von neumann style?: A functional style and its algebra of programs,” *Commun. ACM, vol. 21, no. 8, pp. 613–641, Aug. 1978.*

[2] “Von-Neumann-Architecture” ComputerScience.GSCE.Guru

[3] “X-SRAM: Enabling In-Memory Boolean Computations in CMOS Static Random Access Memories” *IEEE Transactions on Circuits and Systems (Volume: 65, Issue 12, June. 2018)*

[4] S. Li, C. Xu, Q. Zou, J. Zhao, Y. Lu, and Y. Xie, “Pinatubo,” in *Proceedings of the 53rd Annual Design Automation Conference on - DAC16*. ACM Press, 2016.

[5] W. M. Snelgrove, M. Stumm, D. Elliott, R. McKenzie, and C. Cojocaru, “Computational ram: Implementing processors in memory,” *IEEE Design & Test of Computers, vol. 16, pp. 32–41, 1999.*

[6] A. Jaiswal, I. Chakraborty, A. Agrawal, and K. Roy, “8t sram cell as a multi-bit dot product engine for beyond von-neumann computing,” *arXiv preprint arXiv:1802.08601, 2018.*

[7] J. Zhang, Z. Wang, and N. Verma, “In-memory computation of a machine-learning classifier in a standard 6t SRAM array,” *IEEE Journal of Solid-State Circuits, vol. 52, no. 4, pp. 915–924, apr 2017.*

[8] W. Kang, H. Wang, Z. Wang, Y. Zhang, and W. Zhao, “In-memory processing paradigm for bitwise logic operations in stt-mram,” *IEEE Transactions on Magnetics,* 2017.

[9] S. Shirinzadeh, M. Soeken, P.-E. Gaillardon, and R. Drechsler, “Fast logic synthesis for RRAM-based in-memory computing using majorityinverter graphs,” in *Proceedings of the 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE).* Research Publishing Services, 2016.