PERFORMANCE EVALUATION OF LOW POWER SRAM CELL

MAJOR PROJECT

REPORT IN THE

FULLFILLMENT OF REWARD

FOR DEGREE

MASTER OF TECHNOLOGY

IN

VLSI DESIGN & EMBEDDED SYSTEM

Submitted by

Mayank Kaushik (2K22/VLS/07)

Under supervision of

Dr. N . Jayanthi

(Assistant Professor)

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(FORMERLY DELHI COLLEGE OF ENGINEERING)

BAWANA ROAD, DELHI -110042

CANDIDATE'S DECLARATION

I Mayank Kaushik , Roll No. 2K22/VLS/07, student of M.Tech (VLSI Design & Embedded System), hereby declare that the Major project Report titled "**PERFORMANCE EVALUATION OF LOW POWER SRAM CELL**" which is submitted by me to the Department of Electronics & Communication Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not derived from any source without appropriate citation. This work has not previously formed the basis for the award of any Degree, Diploma Associate-ship, Fellowship or other similar title or recognition.

Place : Delhi
Date : 31st May, 2024

Name : Mayank Kaushik
Roll no : 2K22/VLS/07

CERTIFICATE

I hereby certify that the Seminar Report titled "**PERFORMANCE EVALUATION OF LOW POWER SRAM CELL**," submitted by Mayank Kaushik, Roll No. 2K22/VLS/07, of the Electronics & Communication Engineering Department at Delhi Technological University, Delhi, is a record of the project work completed by the student under my supervision. To the best of my knowledge, this work has not been submitted, either partially or fully, for any degree or diploma at this university or any other institution..

\

Place: Delhi

Date: 31st May, 2024

DR. N. JAYANTHI

(ASSISTANT PROFESSOR)

ACKNOWLEDGEMENT

I would like to express my deep sense of gratitude and indebtedness to my high respected and esteemed guide DR N.JAYANTHI (Associate Professor, ECE) for having suggested the topic of my Seminar Report and for giving me complete freedom and flexibility to work on this topic. They have been very encouraging and motivating and the intensity of encouragement has always increased with time. Without their constant support and guidance, I would not have been able to attempt this project. I extend my sincere thanks to all my friends who have been patiently helped me directly or indirectly in accomplishing this project successfully.

Place: Delhi

Name: Mayank Kaushik

Date: 31st May, 2024

Roll No: 2K22/VLS/07

ABSTRACT

This study undertakes a comprehensive assessment of the performance of two pivotal static random access memory configurations: the traditional 6T and innovative 7T SRAM cells. Key parameters, encompassing read and write access times, power consumption, leakage current, area efficiency, stability, noise margins, process variability sensitivity, temperature and voltage sensitivity, technology node compatibility, dynamic power dissipation, and reliability, undergo a systematic analysis. The research strives to offer profound insights into the merits and demerits of each configuration. This knowledge is intended to empower designers and engineers, facilitating well-informed decisions tailored to specific application requirements. The outcomes of this investigation contribute significantly to the ongoing dialogue surrounding the optimization of SRAM design. Ultimately, the research aims to advance the efficiency and reliability of SRAM in a diverse range of semiconductor applications, fostering progress in memory technology and design

LIST OF FIGURES

TABLE OF CONTENTS

CONTENTS

CHAPTER 1

INTRODUCTION AND MOTIVATION

1.1 INTRODUCTION

In current time the demand of high-speed processors which are operated in very less power and required lesser area, is increasing exponentially with time. A perfect memory is reliable, space efficient, and quick with consuming minimum power. Almost all VLSI chips now include fast low power SRAM as a vital component, and it's particularly true for all the processors, where cache memory sizes are increasing on the SOCs with generation to overcome the performance gap between the memory and the processors[1][2]. Furthermore due to greater integration, running speed and the exponential expansion of battery operated product, power dissipation has become an essential problem [3]. The design of SRAM investigated in this thesis, with an emphasis on optimizing latency and power, but scaling of supply [6][11] and the process [4][5] will always be most important factor. This thesis looks at certain circuit approaches that may be utilised in with combination with scaling to produce fast low power operation.

As the demand in battery-operated portable devices increasing, low power and compact size design becoming a crucial factor and most favourable area of research for researchers. To design a fast processor, fast sampling of data is one of the major constraints, and SRAM plays this crucial role. SRAM works as mediator between processor and main memory which are interfaced with other slower peripheral. RAM work as cache memory in a processor. SRAM or cache memory consists of instruction data temporarily, which are frequently in use by the processor. SRAM is 2nd fastest memory in memory organization. It comes after the register type memory. With passing of time as the demand of high-speed operation is increasing, size of cache memory is also increasing proportionally. Since larger size of SRAM memory provides a wider operational bandwidth, which implies larger size of data can be received or can send to destination. Bandwidth is directly related to the speed of operation of processor. i.e., as bandwidth increases speed of operation will also improve. However, as the size of SRAM on System on Chip increases, it becomes a greater area consumer on SOC board. On an average it

consumes about 50% of the total area on a SOC board. This is a major concern. In this project, we will try to optimize our SRAM cell design at block level so that the requirement of area can be minimized. The next work will upon low power techniques at block level and at transistor level to minimize the power dissipation.

Some features of SRAM are like it does not require data refreshing after certain period of interval. These properties of SRAM cell eliminate the requirement of complex and area taking peripherals. It implies that SRAM can retain the data until unless power is not remove, So SRAM need to be connected to the Power Supply . As SRAM is a mediator between processor and other peripherals and memories, it has wide application in many areas for example in wireless communication, in DSPs, in portable devices etc.

SRAM cell generally designed using only MOSFETs. 6T SRAM is a most used design in SRAM cell design. Many other designs is also proposed by the researchers, which have an edge in terms operational speed, data stability, reduced read and write cycle time, but still 6T-wSRAM cell is the first choice of memory manufacturing industry. Since SRAM cell already consumes about 50% area of the total SOC board which is a major constraint. With passing of time much advancement is also proposes by the researcher in 6T SRAM cell design and it is highly compatible with other SRAM design. In this project the work will upon designing of a Low power 6T SRAM cell and using this cell which We design a SRAM memory using the concept of the memory banking. The design should be low power and high-speed peripherals of SRAM memory.

1.2 MOTIVATION

In modern era the demand of battery operated and a portable mobile device is increasing explosively. Surge in battery operated devices gives birth for the low power design and motivates researchers to research and design low power devices with optimizing area and speed trade-off as well. Since SRAM is a one of the most essential part of a SOC and it consumes larger area of a SOC. This constraint motivates me to work upon this topic and try to find an optimum solution which can minimize the speed, power, and area trade off.

SRAM is one of the most essential part of a SOC, it has a wider range of application. It is using in field of wireless communication, in DSPs, in field of bio medical equipment and devices, in portable devices, in multimedia devices, in smart phones and in various high-speed processor

This is also one the reason to decide to work upon this topic. As the size and demand of LP and high speed processors is increasing. So, find out an optimum solution at transistor level, at block level, and at architectural level, so that the power dissipation, area requirement can be reduced, and operational speed can be increased is the main motivation behind doing this project. In emerging sub-100 nm design technologies, leakage power is a major concern. Because most SRAM cells inactive, not accessible, and have little Dynamic activity , it becomes more critical for on-chip SRAM. In terms of leakage, they are on and use static electricity. With the widespread use of battery-powered electronic devices, integrated circuit power consumption is becoming increasingly essential, and because if number of transistor will be more then the leakage power will be more. So it can be said that leakage power is directly related to the number of transistors. SRAM can be a significant sources of leakage power in the Design. Through this work , I tried look at different ways to reduce SRAM's leakage (static) power and develop SRAM cells that can do so without compromising the reliability of the cell and performance.

1.3 OBJECTIVE

To work on this thesis various investigation done and investigate several strategies and different techniques for lowering the SRAM's leakage power design using CMOS technology. Stacked SRAM are proposed here, in order to lowering the leakage power w/o compromising the performance of the SRAM. It is also investigated the effect of different temperature on SRAM circuit

CHAPTER 2

LITERATURE REVIEW

- Shokoufeh Naghizadeh, Mohammad Gholami**[1**], Explained about the importance of SRAM. In present days there is increase in demand of portable device like mobile laptops, tablets and many more. These device provide limitless functionality, to provide limitless functionality they limitless power. But the battery technology is not developed at that speed. Due to this reason the portable device runs on battery, the battery should not drain too fast. So, for lowering the leakage Power consumption of the battery operated device. They show different techniques. SRAM is most important device for the electronic circuit. It is used in the many integrated chip or SOCs due to its speed. It is used for the fetching the data and instruction from the main memory.
- Debasis Mukherjee,Hemanta Kr.Mondal [2], Explained how to compute the static noise margin, read static noise margin, and write static noise margin of a 6T SRAM cell using the butterfly approach. Also explained the pull up ratio and cell ratio, how is it calculated and how it effect the stability of memory cells through his paper. In the conventional 6T SRAM cell for for stability during reading process the size of pull down (NMOS) transistor must be greater than the pull up transistor and the access transistor. The size of Access transistors must be lesser to reduce the bit line capacitances. In other hand for write operation the size of access transistor should be high, and it must have good current capability. The read stability and the write ability of the cells have conflicting design requirement. To solve this problem, separate path for read and write.
- Jan M. Rabaey, Anantha Chandrakasan B. Nikolic [3], Proposed SRAM design and also explained the design of memory using the SRAM cells, also explain about the other peripheral like row and tree decoder, pre-charge, sense amplifier and the connection between them for memory creation.
- Sung M. Kang and Y. Leblebici [4], explained SRAM cell design and it's reading writing and hold operation. Also explain that how a cell is connected to create memory cell, all the cell are connected together in array fashion. That the alternation of stored data is not

permitted in reading process. In his book DRAM is also introduced. ITRS predicted that the SRAM take 90% area of the chip and consume massive power it's about 50% of total power. So, we require the SRAM which will take less power and give high performance. Most of time SRAM is idle and it take power to store data. In other hand the passive does not required power to store the data. But we use SRAM because of its high speed. It helps to enhance the speed of the operation of the processor.

- K. Yamaguchi, H. Nambu [5], Proposed the 64KB CMOS In which ECL are the word line drivers. He uses combination of ECL work line, cell array of CMOS SRAM And some write circuit. The ECL Word line drivers and write circuit drives the CMOS SRAM Cell arrays w/o the use of an intermediate voltage level converter.
- Rakesh Dayaramji Chandankhede [6], Proposed a decoupled latch sense amplifier which is a current control that reduces power consumption while improving performance. When the enable signal is logic low, the bit line logic and Bit-line Bar logic grows on the latch output, i.e. differential voltage. Whenever the pull down NMOS transistor in SRAM is off, the logic on bit - line and logic on bit-line bar does not expand, and the enable signal remains high. When the EN signal is high, the low-voltage line will go to ground.
- Sreerama Reddy G M, P Chnadrasekhara Reddy,[7], Proposed a 8KB SRAM which consume less power than congenital 6T SRAM, also introduce memory banking approach which works at 800MHz. This 8KB low power SRAM was based on 180 nm technology. They also addressed the reduction in power dissipation and clock latency here.
- Harekrishna Kumar, V. K. Tomar, [8], proposed the various type of leakage power in SRAM through this paper and explain it. In present use of the portable devices has increased exponentially. And it is increasing day by day. SRAM is major part of embedded memory and SOC. As technology size decrease the leakage power of devices has become the major issue. It also degrades the power supply. We need to develop the device that consume less power, and gives high performance. Here a comparison has been done and the analysis of 6T, 7T, 8T, 9T, and 10T SRAM. Then give the different idea to reduce it, introduced many schematic diagrams to reduces the leakage power and current[4].
- Pavan kumar Bikki and PitchaiKaruppanan, [9], present a paper, in this paper they give a detail review of SRAM also give information about the different leakage power explain it and at the end gave the idea for reduction for those leakage power, In present scenario the

devices used in the world are small and almost each one is operated on the battery. SRAM is heart of chip, to make single chip millions of SRAM are required. So, designing the circuit that get embedded into these devices in such a way that it will Consume less power or energy is major challenge in present day. Because the leakage power is one third of total power consumption. It can affect the stored data or information in memory.

.

CHAPTER 3

OVERVIEW OF SRAM

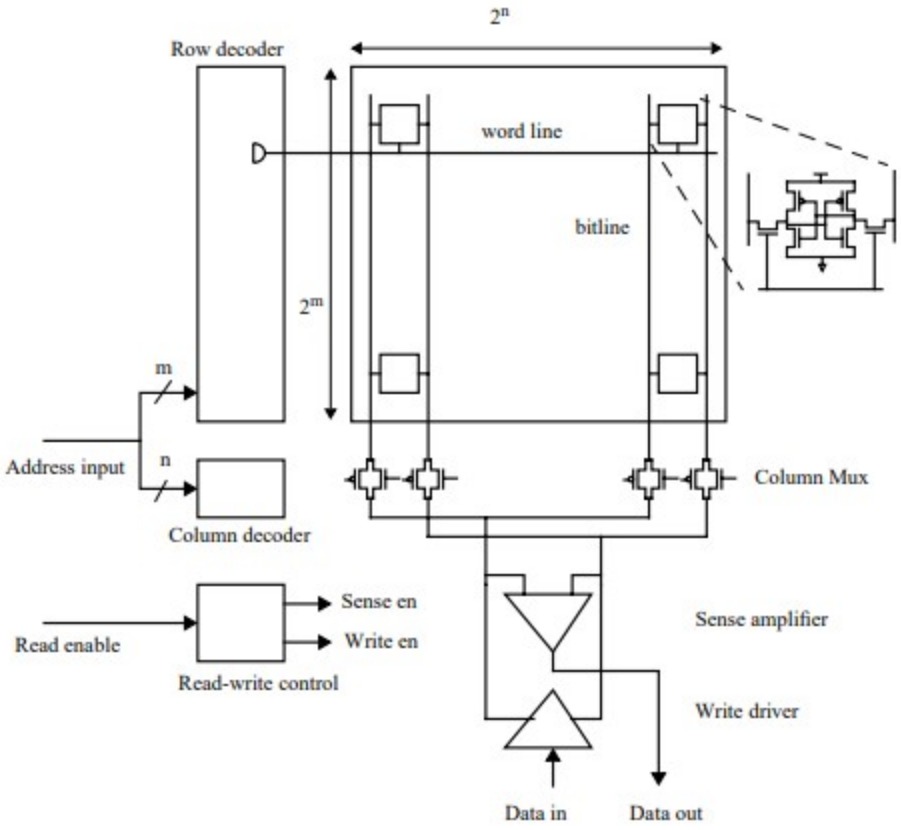
Over time, improvements are happened in SRAM array organization in memory and circuit design that leads to lower the delay and power of practical SRAMs. This chapter is going to explores about both of these subjects as well as the concerns that this thesis addresses. In Section 3.1, The different strategy like partitioning strategies before highlighting the key circuit solutions that has been published in the literature to improves speed and Power.

3.1 PARTITIONING OF SRAM

In large-scale SRAMs, dividing the cell array into smaller sub-arrays, rather than maintaining a single monolithic array as depicted in Figure 3.1, can yield significant improvements in delay and power efficiency. Typically, a large array is segmented into several identically sized sub-arrays, each holding a portion of the accessible word, termed the sub-word. These sub-arrays are all activated simultaneously to access the entire word [10][12]. While low-power SRAMs typically feature only one macro, high-performance SRAMs may incorporate up to 16 macros. Apart from sharing sections of the decoder, the macros can be considered independent RAMs.

Every macro in the SRAM adheres to the basic structure depicted in Fig 3.1. When accessing a row, the word line activates all cells within that row, while the column multiplexers retrieve the requested sub-word. However, macros with a large number of columns encounter two limitations with this setup: the RC delay on the word line increases exponentially with the number of cells in the row, and the power consumption of the bit-lines escalates linearly with the number of columns. Both of these challenges can be mitigated by dividing macros into smaller cell units using the Divided Word Line (DWL) approach, initially proposed by Yoshimoto et al. [13]. The DWL approach segments the lengthy word line of a typical array into k pieces, each independently triggered. This division reduces the word line length by k, consequently lowering the RC delay by K^2. As depicted in Figure 3.2, the DWL design divides a 256-column macro into four blocks, each comprising 64 columns. Row selection occurs in two stages: first, a global word line is activated, followed by the transmission of a block select signal to the appropriate blocks to enable the specified local word line. The local word line experiences reduced RC delay

since it spans just 64 columns. Despite the initial global activation, the DWL approach allows for more efficient word line



operation.

Fig. 3.1: Cell design framework using SRAM

The word line, nearly as long as the macro's width, experiences reduced RC delay compared to a full-length word line due to its lower capacitive loading. Despite being nearly as long as the macro's width, the global word line exhibits a smaller RC latency than a full-length word line because of its relatively low capacitive loading. Instead of encountering the loading of all 256 cells, it only deals with the input loading of the four word line drivers. Additionally, utilizing wider wires on an advanced-level metal layer can further reduce its resistance. By siting the word drivers at the midpoint of the word line segments and halving the length of each section, the word line's RC delay is decreased by another factor of four. The column current is also reduced by a factor of four because only 64 cells in the block are activated rather than all 256 cells in the undivided array. The Hierarchical Word Decoding (HWD) method [14] is based on partitioning

the word line recursively on the global word line (and the blocking select line) for large RAMs. Partitioning can also be used to minimize bit-line heights, as explained in the following section..

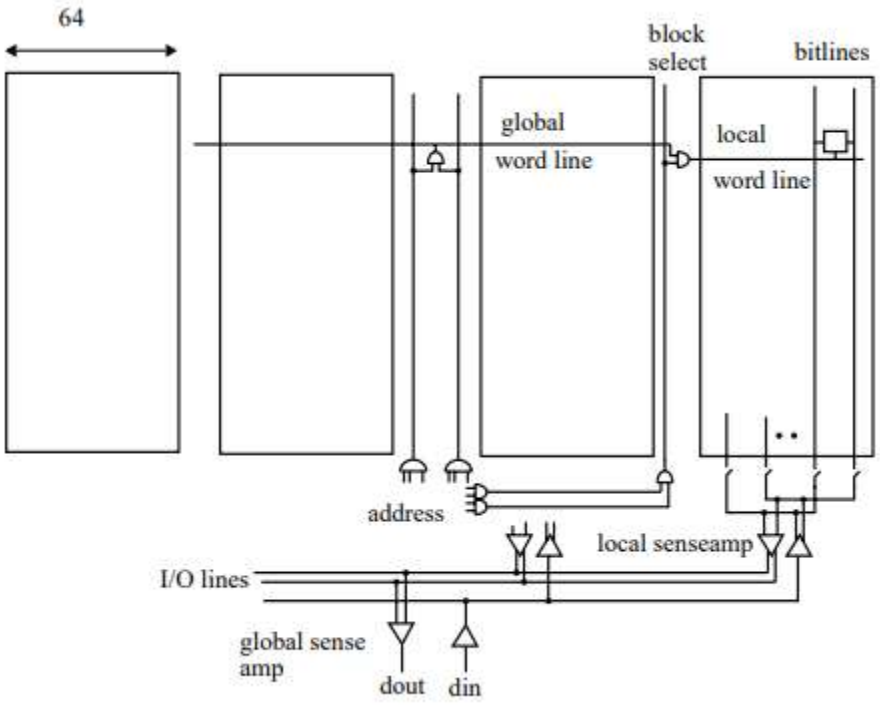


Fig. 3.2: Architecture of Divided Word Line (DWL)

RAM partitioning results in area overhead at the partition boundaries. A partition that deconstructs the word lines, for illustration, necessitates the deployment of word lines drivers at the boundaries. Because the RAM area defines the lengths of the decoder's global wires and the data channel, it has a direct impact on their latency and energy. We'll look at the trade offs in latency, energy and area achieved through partitioning.

3.2 CIRCUIT METHOD IN SRAM

The SRAM access pathway comprises two core elements: the decoder and the data path. The decoder encompasses the circuits from the address input to the word line, while the data path includes the circuits from the cells to the I/O ports.

The decoders logical function is analogous to 2N N-input AND gates, with a hierarchical implementation of the huge fan-in AND operation. Figure 3.3 depicts the architecture of a two-

level 8 * 256 converter. The pre decoder is the first level, where two groups of four address inputs and their complementing (A0, A0, A1, A1,...) are decoded first to activate one of its 16 pre-decoder output wires, resulting in partially decoded products (A0A1A2A3, A0A1A2A3, ...).

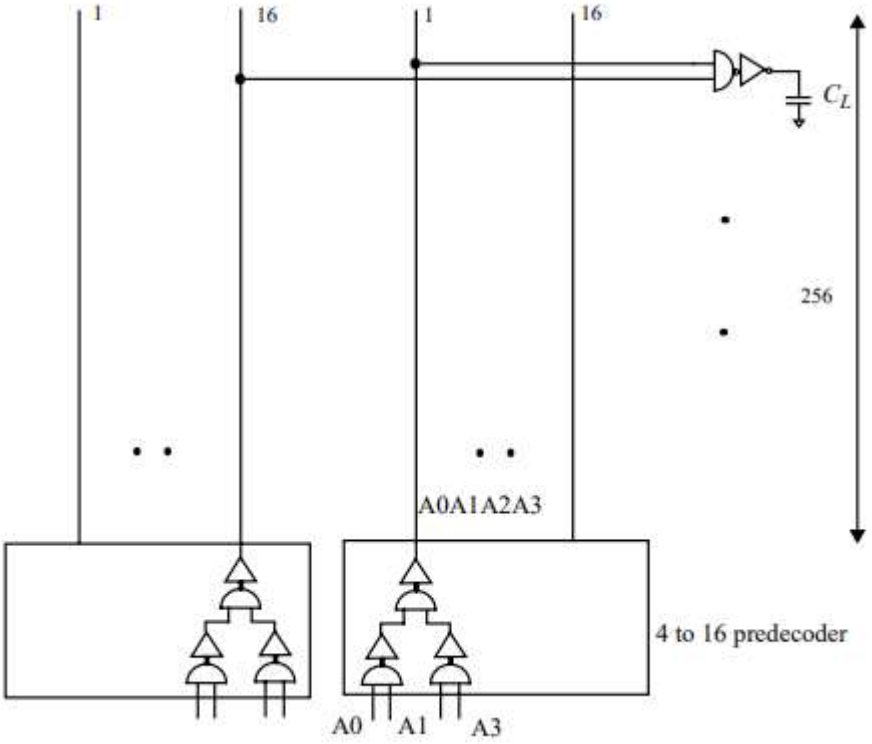
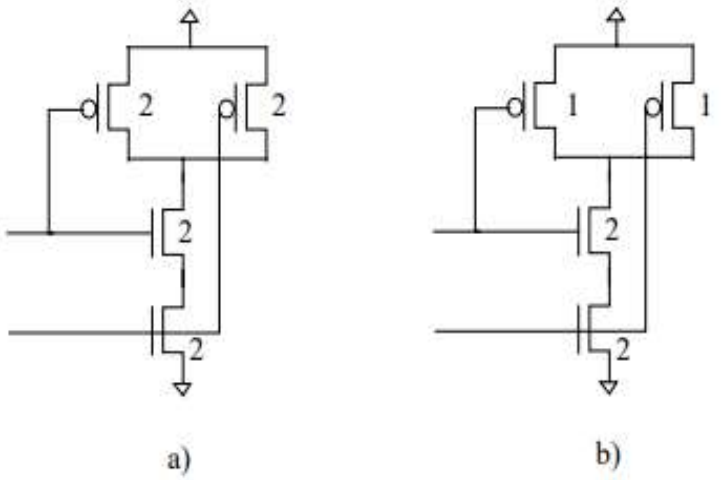


Fig. 3.3: Diagram of 8:256 decoders

To activate the word line, the pre-decoder outputs are merged at the next stage. The gate delays along with the crucial path, as well as the interconnection delays of the pre-decoder and word line wires, makes up the decoder delay.

In larger SRAMs, the wire latencies within the decoder structure, particularly those of the word line, become critical due to the wire RC delay increasing with the square of the wire length. The gate sizes in the decoder offer trade-offs between delay and power consumption. Various researchers have explored transistor sizes for both achieving fast speed [15][17] and minimizing power usage [18][19]. The presence of intermediary links from the pre-decode wires adds complexity to the decoder sizing problem. We address this challenge and propose minimum latency thresholds. Additionally, we investigate simple scaling strategies to achieve high speed and reduce power consumption. By optimizing the circuit design used for the decoder gates, significant reductions in decoder delay can be achieved. Previous designs employed a basic

combinational approach using static CMOS circuits to implement the decode logic function (Figure 3.4a) [20][22]. In this design, only one of its 2M word lines is active at any given time. The old word line is deactivated, and the new word line is activated if the new row address differs from the previous one in any



access.

Fig. 3.4: a) NAND gate (conventional) b) NAND gate by Namura

In such a setup, the delay for DE-asserting the old word line and asserting a new word line hinges on the decoder gate delay. This delay can be reduced by ensuring that each gate in the decode path exhibits equal rising and falling delays. By employing pulsed circuit techniques [23][25], where the word line functions as a pulse active for a minimum duration before deactivating, the decoder gate delay can be significantly diminished. This approach ensures all word lines are turned off before any access, simplifying the process for the decoder to enable the word line for the new row address. As the decoder logic chain only needs to transmit one type of transition, transistor sizes in the gates can be adjusted to hasten this transition and diminish the decode time delay. This method is illustrated in Figure 3.4(b) [26], where the PMOS in the NAND gates are half the size of a regular NAND structure. Decreasing the PMOS sizes by half in the pulsed design maintains the same rising latency, thus lowering the loading of the prior stage and subsequently reducing the entire decoder delay. Further refinement is discussed in [26], where the gate's DE-assertion is completely separated from its assertion. Figure 3.5 such a gate, with semiconductor transistor sizes in the logic chain skewed to accelerate the output assertion

once the inputs are engaged. Additional devices then reset the gate for the next access, refining the approach to reduce the decoder delay by separating the assert and DE-assertion

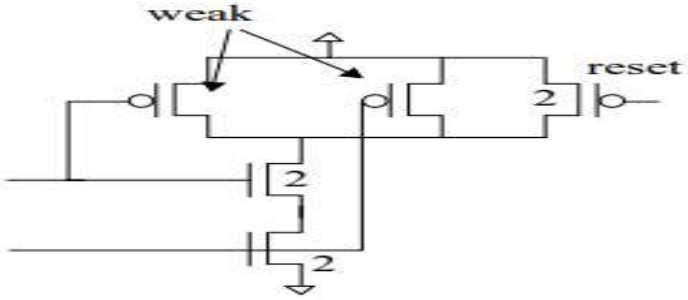


Fig. 3.5: Skew NAND gate

Another way a Pulsed decoder can help a low-power design is by lowering the power of the bit-line path, which we'll go into momentarily.

During read operations, the SRAM data path incorporates a multiplexer (and for write operations, a demultiplexer). The basic multiplexer setup comprises two phases: initially, memory cells in a column are all connected to a bit-line at the lowest level, and subsequently, a limited number of these bit-lines are multiplexed together by column pass transistors at the next level. In scenarios where the bit-line height is notably large, additional layers of metal may be employed to establish a multi-level bit-line hierarchy [27]. The multiplexer structure can be implemented in various configurations. Figure 3.6 displays two different designs for a 512-row block. Although the schematic depicts only the NMOS pass gates, the actual multiplexer would utilize Complementary MOS pass gates for differential bit lines to facilitate both reads and writes, whereas single-ended bit lines are employed in the figure to simplify the representation. Figure 3.6 (a) illustrates a single-level multiplexer architecture wherein two adjacent columns of 512 cells are multiplexed into a single sensing amplifier.

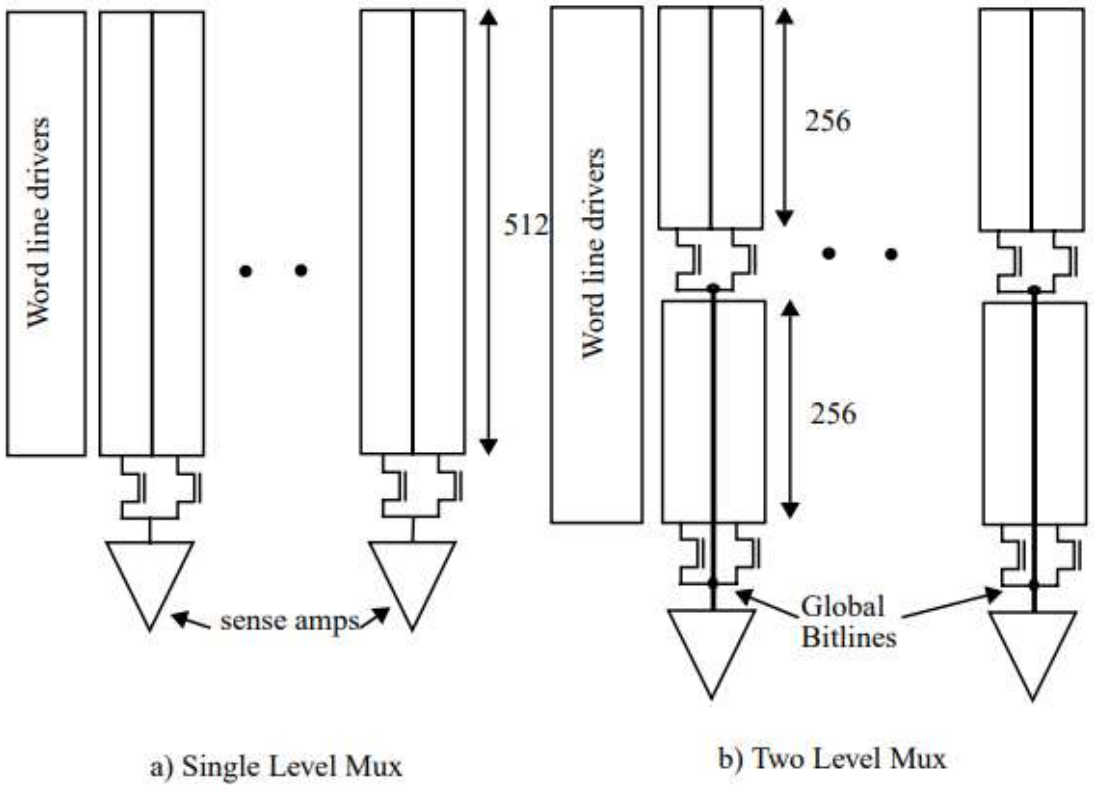
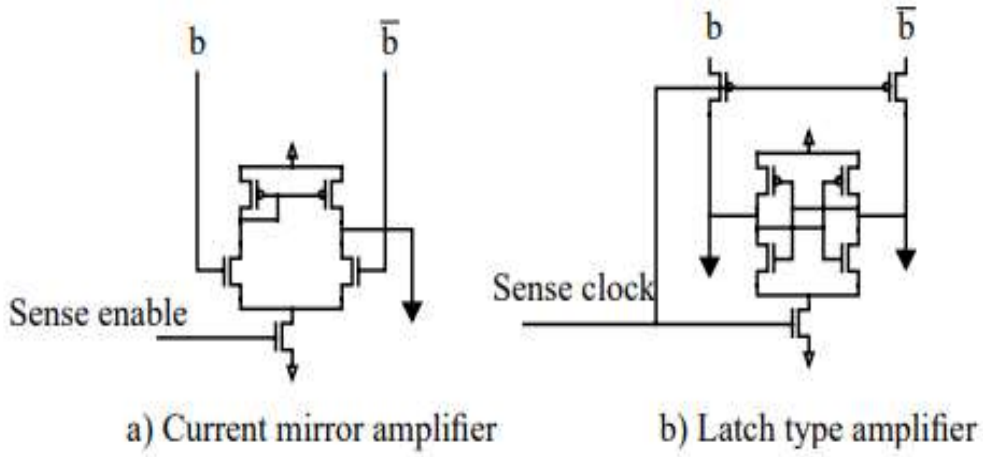


Fig. 3.6: Bit line Mux-architecture

Figure 3.6b presents a two-level structure wherein the first level multiplexes two 256-high columns, and the output is further multiplexed in the second level to generate the global bit lines feeding into the sense amplifier. The I/O lines connect the sense amplifiers' outputs to the I/O ports and can also support hierarchical multiplexing [28]. Due to their small size, memory cells exhibit high weakness, constraining the bit line slew rate during reads. Hence, sense amplifiers amplify the bit line signal to detect signals as low as 100mV. In traditional architecture,

sense amplifiers continue to slew after sensing the bit lines, resulting in a significant voltage differenceand power waste due to high bit line capacitance. To mitigate this, we can control the charge pushed down by the bit lines by restricting the word line pulse width [29][32]. This thesis proposes managing the word line pulse width just enough for the sense amplifiers to consistently

sense and prevent further slewing of the bit lines across various operating .



conditions.

Fig. 3.7: Two different type of Sense Amplifier a) current mirror type b) Latch type

Over time, various sense amplifier circuits have been proposed, generally classified into two main types: linear amplifiers [32] and latch amplifiers [10]. Figure 3.7 showcases a basic example from each class. The linear amplifier type (Figure 3.7a) requires a DC bias current to initialize the high gain region prior to the arrival of the bit line signal.

In high-end designs, multiple amplification stages are necessary to convert the small swing bit-line signal into a full swing Complementary MOS signal. These amplifiers are commonly utilized in such contexts. However, low-power and low-voltage designs are not recommended due to their consumption of biasing power and operation over a restricted supply voltage. In these scenarios, latch-type designs, as illustrated in Figure 3.7b, are employed. They consist of two interconnected amplification stages activated by a sense clock once an acceptable input differential is established. The positive feedback of the latch ensures full amplification of the input signal to a digital level. While this design consumes minimal power due to the absence of biasing power, it may be slower as the sense clock requires some timing margin. If the sense clock precedes the establishment of an adequate input differential, the output value may be inaccurate.

To accommodate extra timing margins, the timing of the sense clock is typically adjusted to account for worst-case operating and process conditions, bringing it below average situations. In this thesis, we investigate timing circuits that monitor bit line delays and are used to create a sense clock with reduced timing overhead.

In larger SRAM configurations, the Sense amplifier outputs are connected to the I/O lines, expanding the data flow hierarchy (as shown in Figure 3.2). These I/O lines serve as conduits for transmitting signals between the RAM's I/O ports and the memory blocks. Given that high access width SRAMs can entail substantial power consumption from these lines, signaling is managed with minimal swings. Employing low swing bit line techniques aims to diminish the power usage of the I/O lines.

CHAPTER 4

SRAM DESIGN PRINCIPAL AND PARAMETER `

A memory is one of the biggest inventions for the device which is used for storing the data or information. It is a very essential electronic component which is used in the computation field and many more. It is also used for transferring the data from one system to another system for example flash memory.

Similarly, a semiconductor memory is also very essential part for computer processing technology. Basically, it is the main memory element for the processor or any system on chip (SOC). We all know this is the era of system on chip, and this system of chip having to perform many complex tasks for many systems. For this it needs a high-speed memory inside it which can take the instruction and from the main memory for the system on chip (soc). If we talk about the semiconductor memory then it is very much capable to do this. It has very fast in compare to normal memory system.

Semiconductor memory are also classified same as ordinary memory. ROM, PROM, EPROM, SRAM and DRAM are the different type of semiconductor memory. The semiconductor memories are fabricated using the transistor. Mainly CMOS technology used to fabricate this kind storing element. Among the all the different types of semiconductors some of them is going to describe only RAM, because RAM is the most used semiconductor memory.

The semiconductor memory is mainly classified into two types that is RAM and ROM, and RAM further classified into two types

- SRAM
- DRAM

4.1 DRAM

DRAM is a type of memory which is frequently utilized in computer systems. The acronym DRAM refers to "dynamic random access memory.” Its cell structure is shown in the figure

given below. The DRAM cell diagram contains one bit line and one word line. One capacitor is connected followed by NMOS pass transistor. The information store in the DRAM in the form of charge and by the charging and discharging we can know that what kind of data stored in the DRAM cell. To store the data, we have to charge the capacitor and to read the data from the DRAM we have to discharge the capacitor.

For read or write we have to give address line high because here pass transistor is N type, and we all know that NMOS is on only when we give high to its input. So, for read and write NMOS must be on. And if NMOS is on then we should give the data to the bit line whatever data we want to store. And the data will store in the capacitor through the NMOS.

Similarly, if we want to read the data then at that time NMOS must be on then we can access the data. But in this case capacitor will discharge and it will lose the data.

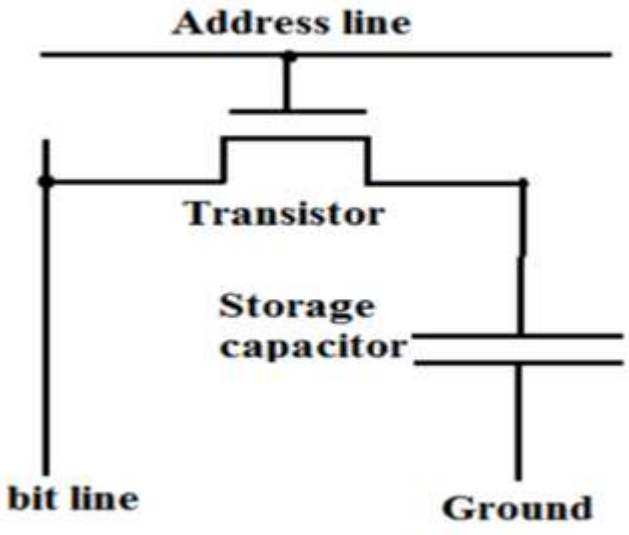


Fig. 4.1: schematic of dynamic RAM [34]

We have already discussed that when we read the data from the DRAM then the capacitor will lose its data, because during read operation it will discharge. That's why we need to refresh it periodically. Due to this reason, it is known as dynamic RAM.

4.2 SRAM

The SRAM contain two couple inverter, which is used as storing element. SRAM stands for static random-access memory. Before understand its work, we have to see its cell diagram. The cell diagrams of the SRAM is shown in the figure 2 given below. This is the conventional 6T SRAM, we can see that the SRAM contain only 6 transistors.

To make this conventional SRAM we need 4 N-type and 2 P-type transistors. Here N3 and N4 are the pass transistor which is N-type. Other 4 transistor are used for making cross couple inverter. Two-bit line is available in both right and left side, which is connected to NMOS from side. One word line is there.

4.2.1 IMPORTANCE OF SRAM

Various levels of memory hierarchy have been added in the memory architecture of a processor to narrow the Gap between the CPU's performance and the other storage speed [35]. This hierarchy ranges from on-chip, low-speed memories to external, high-speed memory like DRAM and hard drives. Processors aim to keep commonly

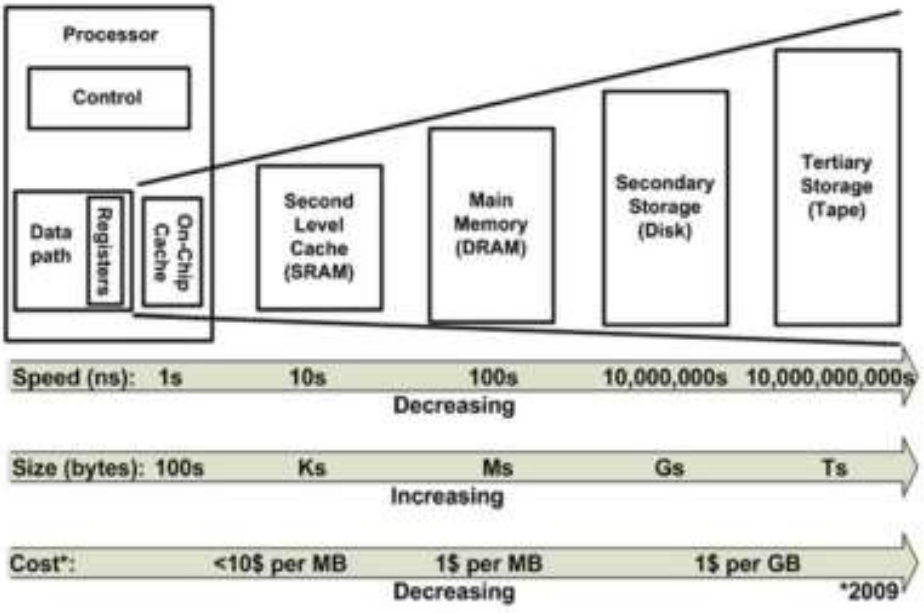


Fig. 4.2: hierarchy of memory

Used instructions and data close to themselves in fast on-chip memory known as cache or scratchpad (software controlled) memories. A typical memory hierarchy of a modern computer system is shown in Fig 4.2.

On-chip cache memory is classified into three levels: L1, L2, and L3 cache memories, each with a different capacity and speed. The storage capacity of a lower-level cache is lesser, but it is much faster than the other memories [35]. While the cache memories are managed by hardware and are hidden from view by the main processor, certain processors use memory allocator memories [36]. scratchpad memories are controlled by the CPU, and software can determine whether data or instructions require storage in scratchpad memories. SRAM is used to develop both scratchpad and cache memory. To satisfy the demands, system on chip applications and high performance CPUs are demanding additional on-chip memory, because it increases the operational speed. However, because to space and cost constraints, huge amounts of SRAM cannot be integrated into the chip. [37].

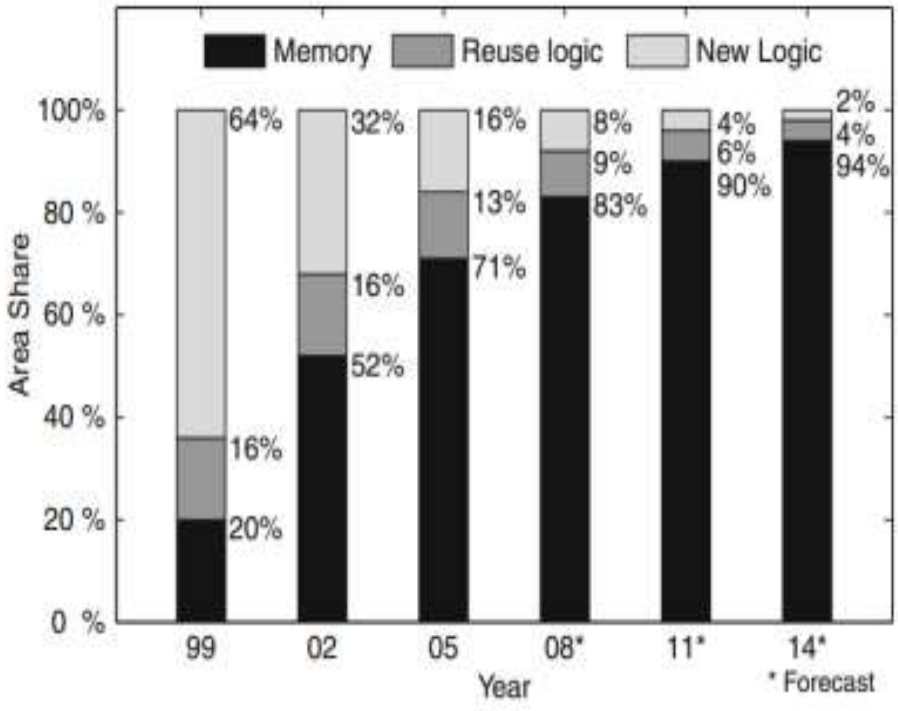


Fig. 4.3: Trend and area of memory on system on chip (SOC0

On a system-on-chip die, Figure 4.3 depicts the trend of embedded memory and logic area. This fig shows that the amount of SRAM on a chip is increased from 20% to 94% in about 15 years. The demand for high performance computation, which necessitates a large number of high-speed on chip memories, drove this massive development. This trend supports the massive efforts and ongoing research into SRAM, which aims to improve reliability while lowering cost and energy consumption, particularly leakage power. Because SRAM takes up a lot of space and leakage power proportional to the number of transistors in the design because if number of transistor will be more the leakage will be more or area, area is also responsible for leakage. Any attempt to reduce SRAM leakage will improves the entire design.

4.2.2. ARCHITECHTURE OF SRAM

A collection of SRAM bit-cells (or cells) and peripheral circuitry make up SRAM. Address decoders for all columns and rows, Sense-amplifiers, bit-line pre-charge circuits, write drivers and other peripheral circuitry facilitate writing, and reading into SRAM cells. Figure 4.4 depicts a traditional SRAM memory design with 2^n words of 2^{m+k}. SRAM is accessible using 2k words. Thousands or even millions of identical cells make up an SRAM array. for illustration, we can take a 4 Mb SRAM which demands 4,194,304 bit-cells, and even minor improvements in design requirements like Static power have a huge impact on the overall processor design or SOC design [37]. The major problems for high performance processors are speed and cell area, which designers strive to improve. energy consumption and dependability are the major or primary problems for energy-constrained applications and battery-powered devices [37]. In this project, I want to reduce the leakage of SRAM memory bit cell in order to lower system's leakage power.

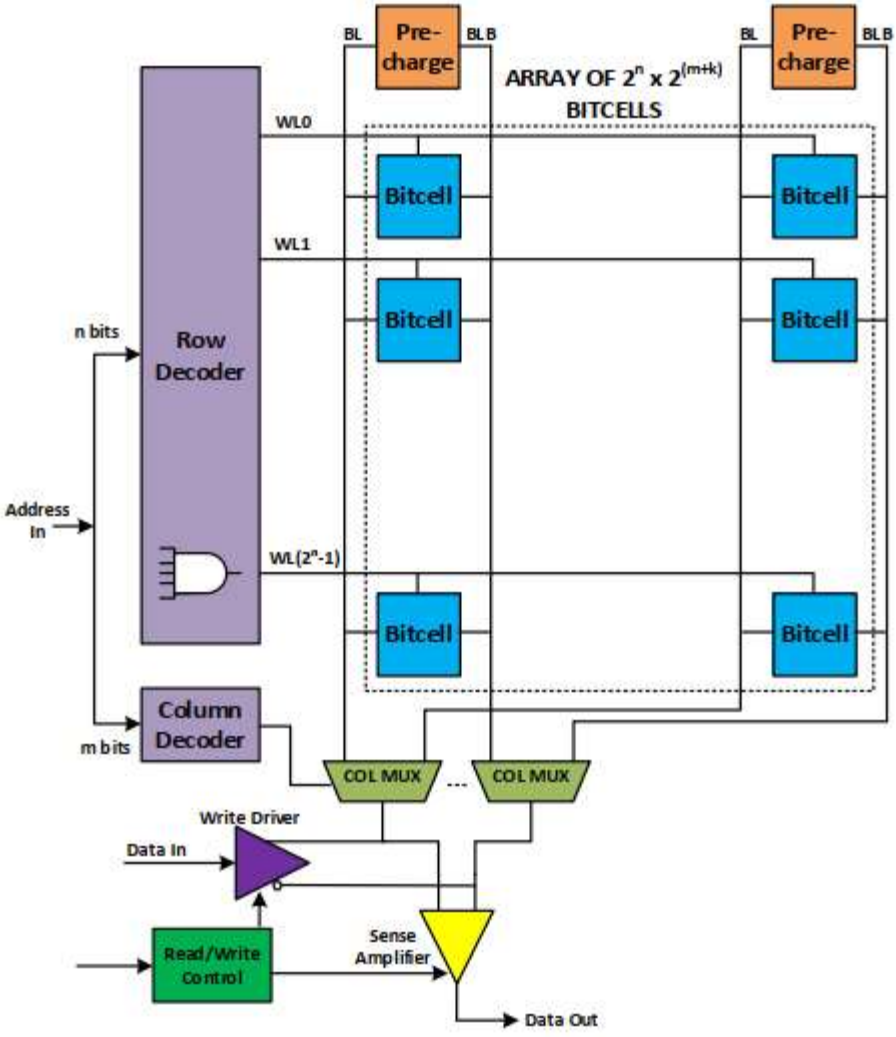


Fig. 4.4: Architecture of SRAM [37]

The memory bit-cell is the foundational component of an SRAM memory, responsible for storing individual bits of data. A single word line (WL) links memory bit-cells within a row, enabling access to those cells. In a column of memory cells, pairs of bit lines (bit-line BL and its complement BLB) are employed for both reading from and writing to the cells.

Large memories are arranged in fashion that has the same Horizontal and Vertical dimensions or order. As a result, a word line allows for many memory words, with column decoders used for picking the addressed memory word. Column multiplexers connect the selected columns' bit-

The bit line (BL) and its complement (BLB) are linked to sensing amplifiers for reading cell values and to write drivers for writing data into the memory cells. Row decoders decode the memory's address and activate the corresponding row in the memory array.

4.2.3 BITCELL OF SRAM

The fundamental component of an SRAM is the bit-cell, responsible for storing a single bit of data. A basic SRAM bit cell comprises six transistors, as illustrated in Figure 4.5, with four being NMOS and the remaining two being PMOS. Two pairs of cross-coupled inverters form a one-bit storage element. To create this storage element, four transistors are required, while the remaining two are utilized for reading and writing data. These two transistors, known as access transistors, are connected to the word line (WL). Since the access transistor is of N-type, it activates when the supply is at logic high. During reading and writing operations, the word line must be high as it is connected to the NMOS. When the word line is high, the NMOS provides a zero-resistance path between the drain and source, whereas it offers high resistance when the word line is at logic low.

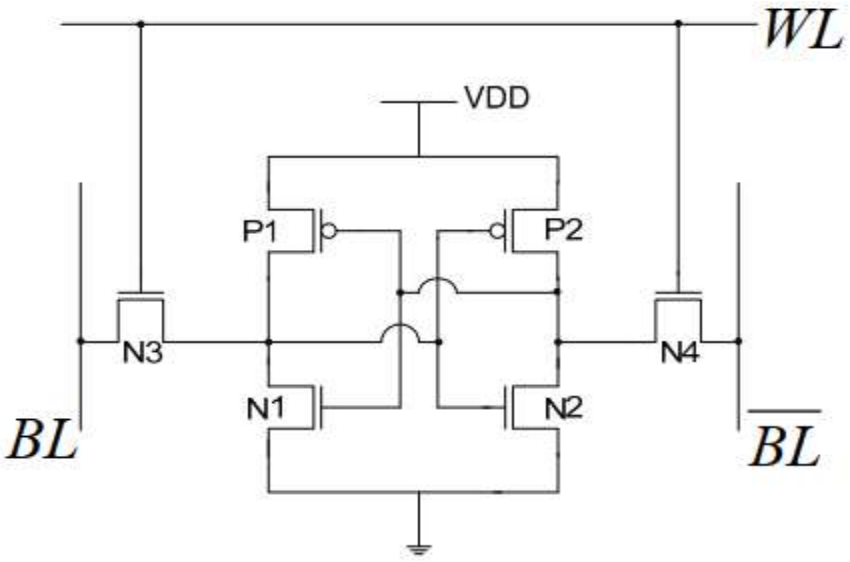


Fig. 4.5: Bit-cell of SRAM

4.2.4 SENSE AMPLIFIERS

Sense amplifiers are a very crucial SRAM peripheral among all the peripheral. The fundamental purpose of sensing amplifiers is to identify and transform the small differential voltage created during read from the bit-cells on BL and BLB to a logical value [37]. and it is also responsible for speed of read operation of SRAM Figure 4.6 depicts a typical current mirror sensing amplifier.

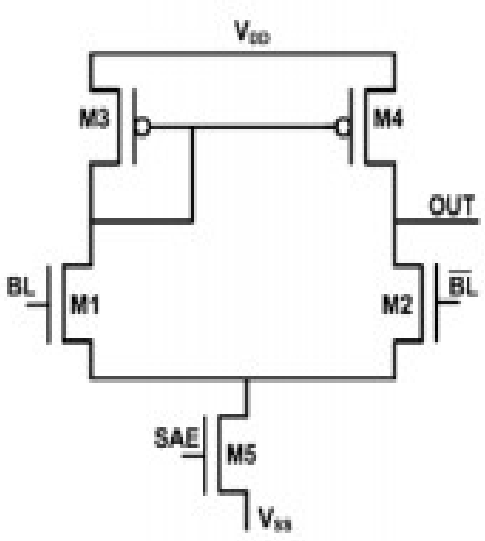


Fig. 4.6 sense amplifier of current mirror type

4.2.5 PRECHARGE CIRUIT

In SRAM, both the bit line (BL) and its complement (BLB) are connected to a pre-charge circuit for each bit-cell within a column. These pre-charge circuits are essential for enhancing the speed of read and write processes. Access transistors in SRAM bit-cells are NMOS transistors, which are proficient at conducting zero but less effective at conducting one. NMOS transistors tend to cease conducting effectively when the source voltage reaches Vdd-Vth. To expedite read and write operations in SRAM, a pre-charge cycle is initiated before any operation. Since PMOS transistors excel at conducting one, two PMOS transistors are incorporated into the circuit,

Attached to both the bit line (BL) and its complement (BLB), there exists a pre-charge circuit featuring two PMOS transistors. Control over the gate of these transistors is managed by a pre-charge circuit enabling signal. When the control signal is low, BL and BLB are linked to Vdd. Additionally, another PMOS transistor might be utilized to balance BL and BLB.

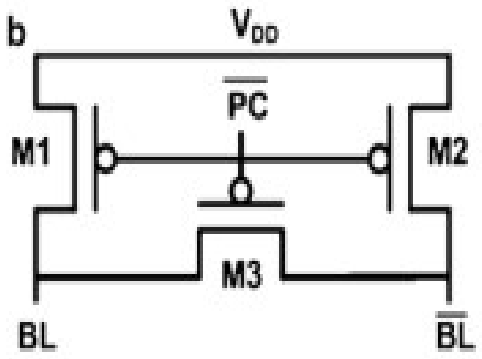


Fig. 4.7: pre-charge in SRAM

4.2.6 WRITE DRIVERS

When write operation is performing BL and their counterparts (BLB) are driven by write drivers. The write driver only needs to draw down the correct bit-line because BL and BLB are pre-charged to Vdd before every operation. Figure 4.8 [37] shows an examples circuit for a write driver utilising pass gates. The WE signal is being used for enabling SRAM writing.

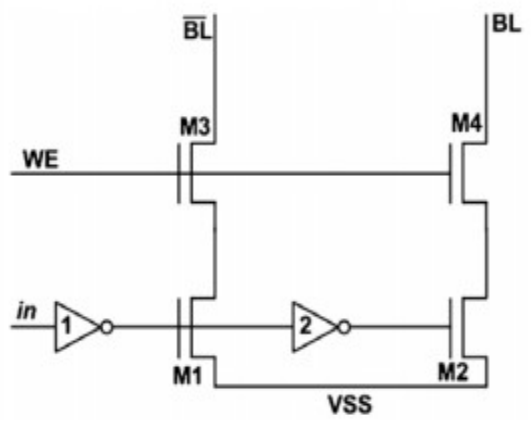


Fig. 4.8 pass gate are used in write driver

4.3 READ OPERATION AND WRITE OPERATION

To read the data or the information from the SRAM, the pass transistor must be on. Here in this figure N4 and N3 both is the access transistor. To on this transistor 1(high) should be given to the word line so that it can turn on the pass transistor. Then we need a sense amplifier for recognize the stored data. The pre-charge bit-line and bit-line bar both are connected to the sense amplifier. We can perform the read operation by using the single bit line but by using these two it can increase the speed of read operation.

To perform write operation pass transistor (N3 and N4) must be turn on like as read operation because the NMOS give direct access to the storing element or latch form element. Then we give the value to bit line whatever value we want to store in the cell.

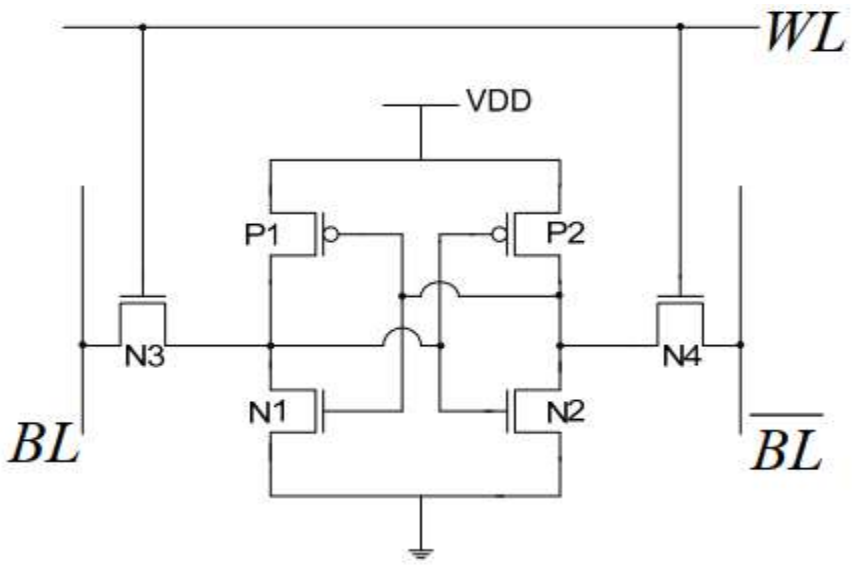


Fig. 4.9 cell diagram of conventional 6T SRAM

4.4 HOLD OPERATION

During the hold state, the word line is set to logic zero, causing the NMOS transistor connected to it to turn off since NMOS functions only when its gate receives a logic high signal. With the gate of the NMOS transistor set to logic zero, it forms a high-resistance path between the drain and source. As the word line is linked to the gate of the NMOS, both access transistors are deactivated, disconnecting the bit line and bit line bar from the storage cell since the access transistor presents a high resistance in its off state. In this state, the SRAM can retain stored data without enabling read or write operations..

4.5 COMPARISON BETWEEN SRAM AND DRAM

- Different memory used for different purpose like SRAM is used as a cache memory, whereas DRAM is used as main memory.
- SRAM is very fast comparing to DRAM.
- SRAM need 6 transistors to design whereas DRAM needs only one transistor along with one capacitor, so SRAM is costlier than the DRAM.
- Density of SRAM is low where density of DRAM is high.
- Power consumption in the SRAM is less compare to DRAM.

4.6 PERFORMANCE METRICS

Now it is going to discuss about some parameter according to that we can decide we can decide whether following RAM is good or bad.

4.7 CELL RATIO and PULL RATIO

For the satisfactory operation of SRAMs cell, the appropriate device sizing must be required. For that we must follow the design principal for the SRAM topology.

The design principle involving cell ratio and pull-up ratios plays a crucial role in improving SRAM stability. Cell ratio is determined by the width ratio of the pull-down transistor to the access transistor, while pull-up ratio represents the width of the pull-up transistor.

When the cell ratio surpasses 1, it enhances the read stability of SRAM. To bolster read stability, the pull-down transistor (NMOS) should be significantly stronger than both the access transistor and the pull-up transistor. Similarly, to ensure better write stability, the pull-up ratio should remain below 1.

4.8 STABILITY

It is a useful metrics used to design SRAM cell. It has been used for years as a useful metric for optimizing the design of SRAM cell. It can able to predict the effect of parameter variation. We can calculate the cell stability by calculating the noise margin for the memory. It is traditionally used for long time. The noise margin represents the cell ability to tolerate a certain presence of noise. Noise can be either current or voltage. The Noise margin is measured in terms of electric variable.

4.9 STATIC NOISE MARGIN (SNM)

In SRAM, the static noise margin (SNM) denotes the minimum voltage required to alter stored data. SNM is derived by plotting the voltage transfer characteristic curves of both inverters on a shared graph. Figure 4.10 depicts the SNM curve of the SRAM. The stability of the SRAM is determined by identifying the largest square that can be accommodated between the butterfly curve or the lobes.

Read SNM is defining when the read operation is performed. During the read operation the access transistor is one. In this case we draw the butterfly curve by considering the inverter and the access transistor. Then read SNM is decided by the largest square that can fit into the butterfly lobe.

Write SNM is little bit different from the read SNM. The both access transistor is on during the write operation. When we separate the both inverters to draw the VTC curve then it is not symmetric like Read SNM, its VTC curve cuts each other at Vdd which shows that it is mono stable. This is better for write stability in SRAM. Then the write SNM is also decided by the largest square that can fit into the butterfly lobe.

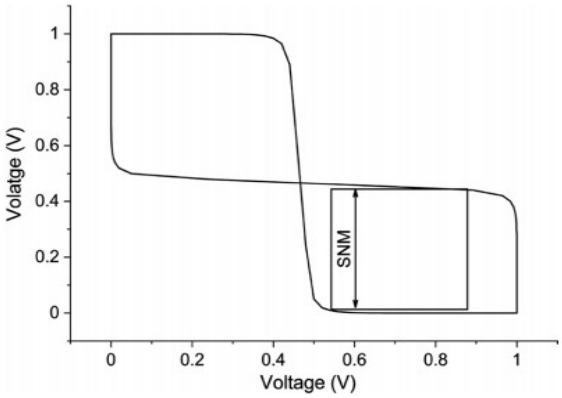


Fig. 4.10 Butterfly Curve of SRAM [6]

- N-curve method - this is the alternative method of butterfly cure. It is used to determined noise metrics. In this method, a voltage sweep, VIN is applied at the storage node Q from 0V to the maximum voltage Vdd. Then the relation between the current and voltage is plotted in a single plane. In this graph current is taken at Y-axis and voltage is taken to the X-axis.

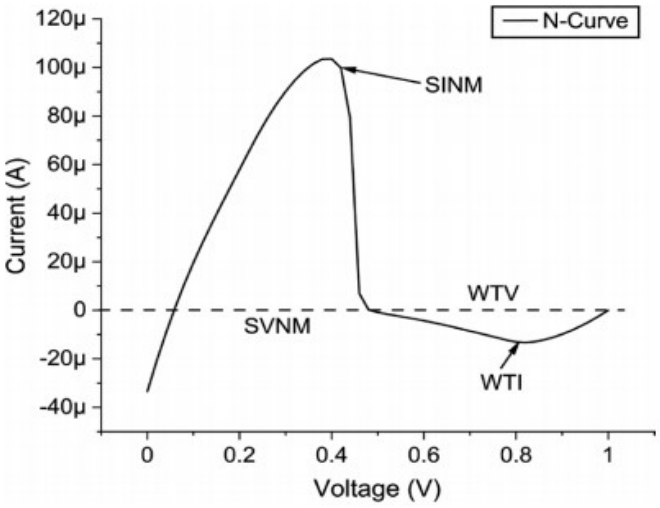


Fig. 4.11 Extracted N-curve [6]

Static voltage noise margin and static current noise margin are metrics used to measure read stability, while write stability is determined by write trip voltage and write trip current.

4.10 POWER DISSIPATION

Power dissipation in SRAM cell involves following two major components :

1. **Dynamic power consumption:** Power dissipation occurs during circuit operation,
   When output is changes with respect to change in input, in that case the load capacitor is charged or discharged. Due this reason we lose some power is known as dynamic power consumption.

$$E_c = \frac{1}{2}.C_L V_{DD}^2 \qquad (1)$$

2. **Static power loss:** The loss due to the leakage current in transistor in classified as the Static power loss. As technology is shrinking day by day leakage current is also increasing with it. The total leakage current can be given by following equation.

$$I_{total\ leak} = I_{(sub\_th)} + I_{(gate)} + I_{junction} \qquad (2)$$

CHAPTER 5

LEAKAGE POWER IN SRAM

In present day every electronic device which is used today are portable. Almost every portable device is runs on battery. So, designing the circuit which consume less energy and give high performance is very much important. It also affects the design process of the chip. So, we need different new technique which should not be power hungry. In order to obtain this kind of chip having low leakage current and high performance, complementary metal oxide semiconductor (CMOS) devices are scaling down in last 30 years. Area and delay were major concern in 70's and 80's. But complexity increased with the time and task of testing more difficult. After chip is manufactured then testing is done to find the defect in fabrication. Now size of chip is decrease to reduce the delay. But due to small size of the chip there is increase in leakage current. So, we need technique or we can say that we need a new design rule so that the product will consume less power. If the design circuit will consume less power can embedded with the device will increase the battery life of the device.

A scaled technology can have less area and can increase the speed but it comes with some issues, which are given below:

1. A device having the low threshold voltage can increase the sub-threshold leakage currents.
2. There is reduction of worst-case performance of a chip.

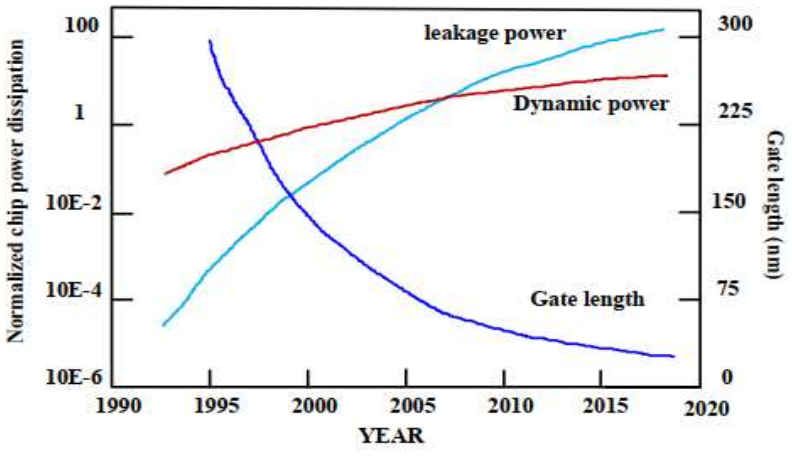


Fig. 5.1: leakage power in a chip [4]

The above fig 5.1 show the leakage power dissipation according to the international technology road map for the semiconductors(ITRS).

For the high performance of the chip transistors are scaling down. It increases the performance undoubtedly, but it becomes the reason for the more leakage current. Leakage current of a single transistor is very less but for making a chip there are millions of chips are required, and then this leakage current become countable. Figure 5.2 shows the components that affect the leakage current.

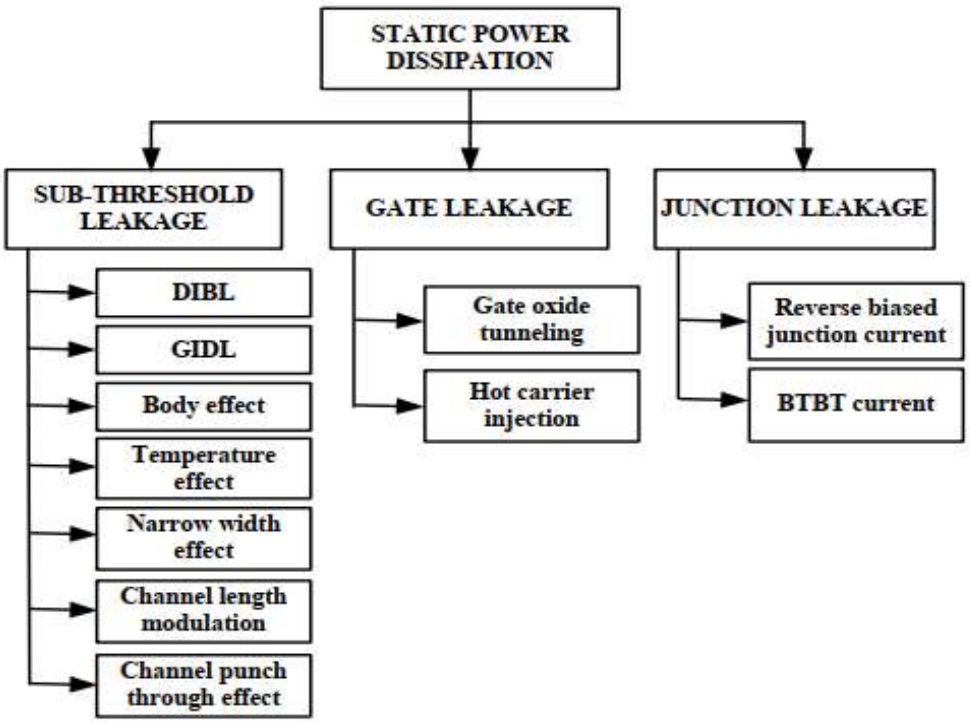


Fig. 5.2: leakage current in MOSFET [9]

There are some leakage currents depends when there is no channel is present. That is when Vgs smaller than the Vt (threshold voltage). This current is directly proportional to the exponent for Vgs-Vt. The DIBL current is depending upon the voltages between the drain and the source. Gate leakage current is eventually tunnelling through the oxide. In MOSFET oxide is resemble as the capacitor and it provide the very high resistance and practically there is no current is passing through this. But there some tunnelling and this tunnelling can be more if it gets thinner.

It is totally depending on the thickness of the oxide layer. The equation [4] represent total leakage current.

$$I_{total\ leak} = I_{(sub\_th)} + I_{(gate)} + I_{junction} \quad (3)$$

5.1 LEAKAGE CURRENT IN TRANSISTOR

Minimization of leakage current in transistor is very important to design low power SRAM. Some of the leakage current MOSFET is listed in below:

- Sub threshold leakage current
- Gate induced drain leakage current
- Punch through leakage current
- Gate tunneling leakage current

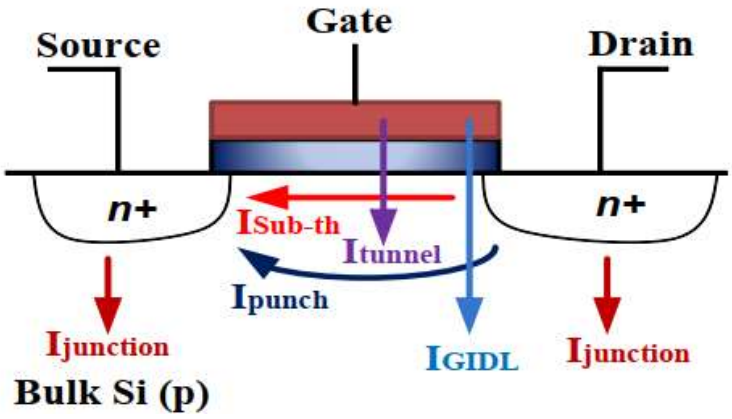


Fig. 5.3: Leakage currents in a NMOS transistor

5.1.1 SUB-THRESHOLD LEAKAGE CURRENT

Sub threshold leakage current occurs when Gate to Source Voltage (Vgs) is less than threshold voltage of the transistor (Vto). This leakage current flows between source and drain terminals of MOSFET. This current leads to Static power dissipation in transistor. Lesser the Leakage current

lesser the static power dissipation and more reliable the device is. Sub threshold current increases exponentially with respect to temperature.

$$I_{D(weak\ inversion)} = I_{on} \cdot e^{vgs/v_{th}} \quad (4)$$

5.1.2 GATE INDUCED DRAIN LEAKAGE CURRENT

Gate induced drain leakage develops due to band to band tunneling at gate and drain overlap region under the strong electric field. When the depletion layer at the interface decreases electric field increases. This type of current increases due to various factors such as reducing oxide thickness of gate, low threshold voltage devices and when there is more potential between Gate and Drain terminals.

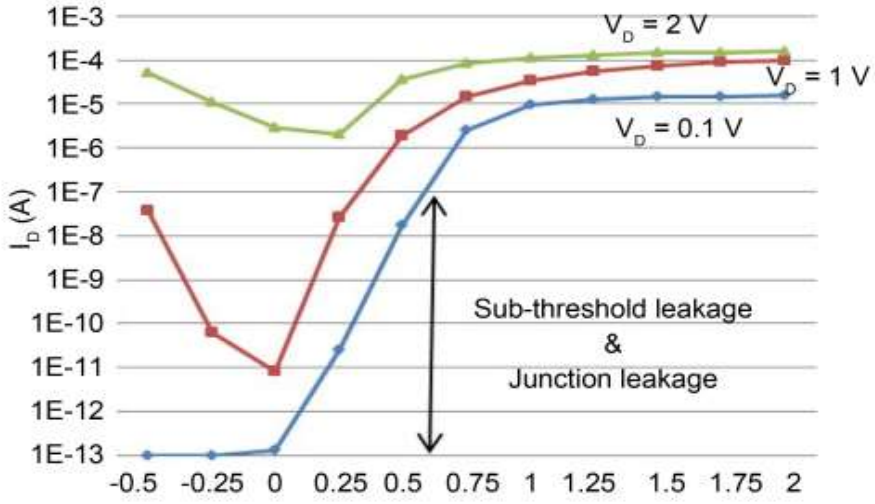


Fig. 5.4: Gate source voltage vs. drain current

5.1.3 PUNCH THROUGH LEAKAGE CURRENT

- Punch-through current is mainly observed in small-geometry MOS transistors, where the depletion regions of the source and drain are close together, extending into the short channel. To minimize this type of current, the following approaches can be employed:
- Substrate doping high
- Small Oxides

- Junctions be preferably long.

$$X_{dD} = \sqrt{\left[\frac{2\epsilon_{si}}{qN_A}\right](V_{DS} + \emptyset_{si} + V_{sb})}$$

$$X_{dD} = \sqrt{\left[\frac{2\epsilon_{si}}{qN_A}\right](\emptyset_{si} + V_{DS})}$$

The sub-threshold surface diffusion current (I_{Sdif}) for the short channel at its saturation level can be express as: -

$$I_{\$dif} \propto Dn_i^2 \,.\, e^{(\frac{q\Delta\emptyset_s}{kT})/L_{eff}} \qquad (5)$$

5.1.4 GATE TUNNELING LEAKAGE CURRENT

Gate tunneling leakage current is generated by the strong electric field across the thin gate oxide layer. This current's magnitude is predominantly influenced by the device's structural characteristics and the applied bias. Under conditions of high electric field, electrons tunnel back and forth between the gate and the bulk, traversing the gate oxide layer. The narrow potential barrier width facilitates the easy passage of highly charged electrons through the oxide layer, consequently elevating the gate current. When the gate voltage is negative (Vg<0), electrons from the n+ polysilicon can effortlessly penetrate the gate oxide and enter the gate, giving rise to what is known as Gate Current. Gate tunneling current can be succinctly described as follows:

$$I_g = I_{gc} + I_{gso} + I_{gdo} \quad (6)$$

CHAPTER 6

LEAKAGE POWER REDUCTION TECHNIQUES

We have already seen in previous chapter that we are losing some energy and power in SRAM just because of leakage current. We have also seen the different leakage current in transistor due to which we are losing the power and energy. We can reduce this kind of power lost by applying some design technique. These design techniques can not reduce the leakage power completely but it can save some energy. So, some design techniques are given below:

6.1 TRANSISTOR STACK

To Reduce leakage power in the SRAMs transistor stacking technique can be used. In this method or technique, we are connecting the transistors in series. Fig. 6.1 is representing the example of this technique. In each stage leakage current is reducing, because the resistance is increasing.

- In stacking technique NMOS transistor is used. In this configuration voltage from drain to source in decreases, and this lead to reduce in DIBL current and also the sub threshold leakage current.
- Here the Gate to source voltage is not more than zero, therefore it also helps in the reduction of sub threshold leakage current.
- In this configuration the voltage from Substrate to source voltage is -ve too. Therefore, due to body effect threshold voltage will increase. Hence sub-threshold leakage current is reduced.

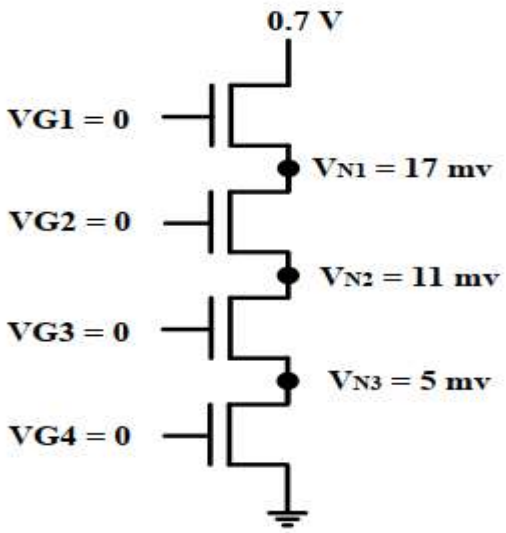


Fig. 6.1: Stacking Technique

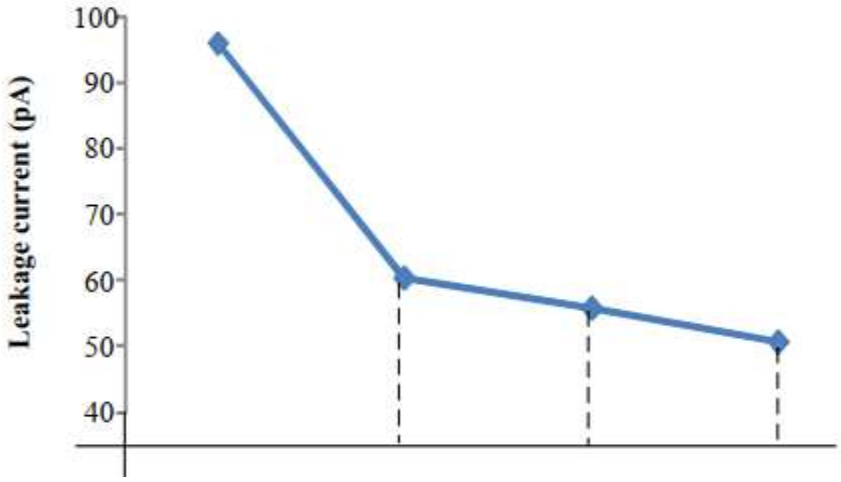


Fig. 6.2: curve of leakage current

A novel P4 SRAM cell has shown in the fig 6.3. There are 4 NMOS and 4 PMOS are present in this configuration. It is containing 2 more transistor than the conventional 6T SRAM. The 2 PMOS are stacked together and placed between the Vdd and ground. It works as a conventional 6T SRAM when WL (word line) is given low and these PMOS will be off when word line is given high. This P4-SRAM cell configuration is used for reducing the sub threshold leakage current.

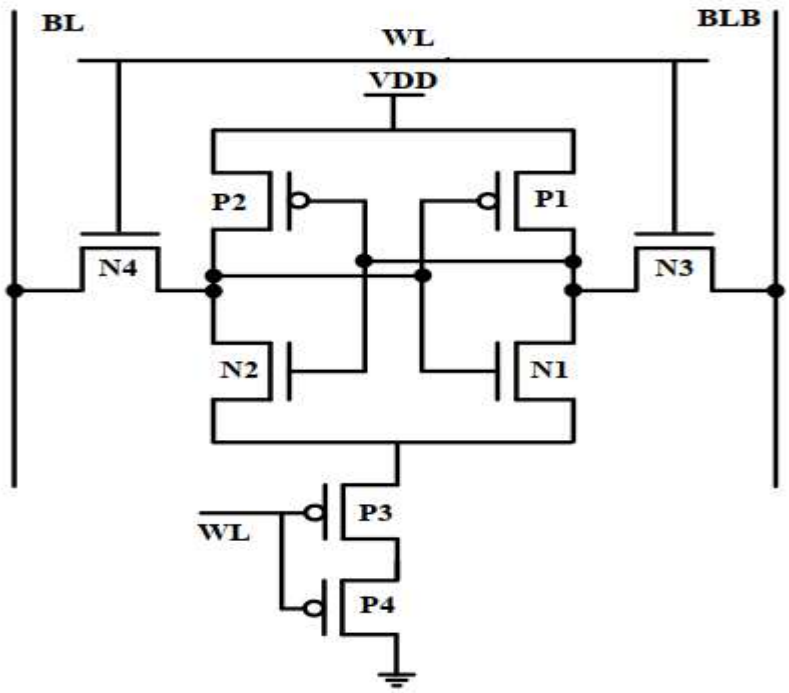


Fig. 6.3: P4-SRAM design [7]

6.1.1 AREA ISSUES IN STACKED TECHNIQUES

One of the biggest drawbacks of employing stacked transistors in SRAM cells could be area. Because in stacked technique more transistor are connected in series. Here attached one figure down below in order to compare the area. It is clearly seen that the area consume more in case of stacked transistor.

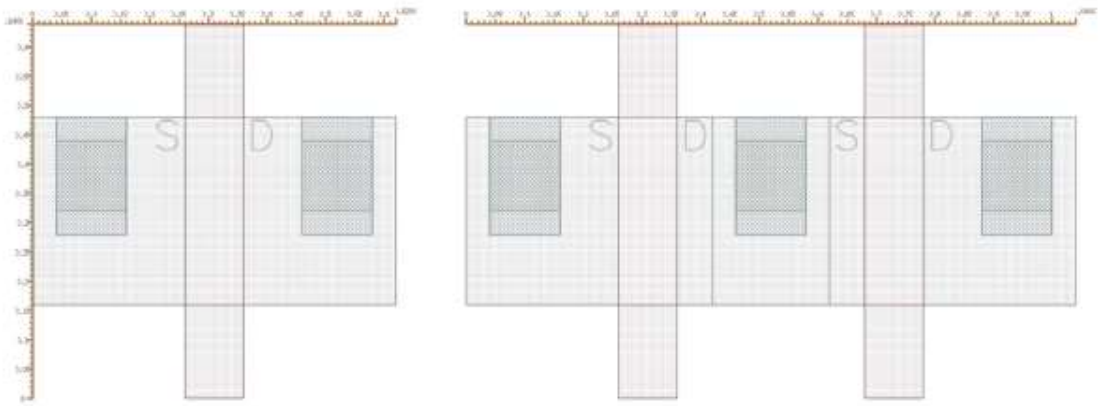


Fig. 6.4: Area comparison

6.2 GATING TECHNIQUE

In fig. 6.5 another technique is introduced for reduction in the leakage current. This method of power reduction is very much popular in the industry. This method is known as gated technique. To design the gated technique we required 7 transistors, that is one more transistor is required than the conventional 6T SRAM. The extra NMOS transistor is place between the Vdd and the ground. This configuration work as conventional 6T SRAM when control signal is given high, and when control signal is given low NMOS will be off. Thus, this configuration reduces the leakage current by eliminating the path between the Vdd and ground.

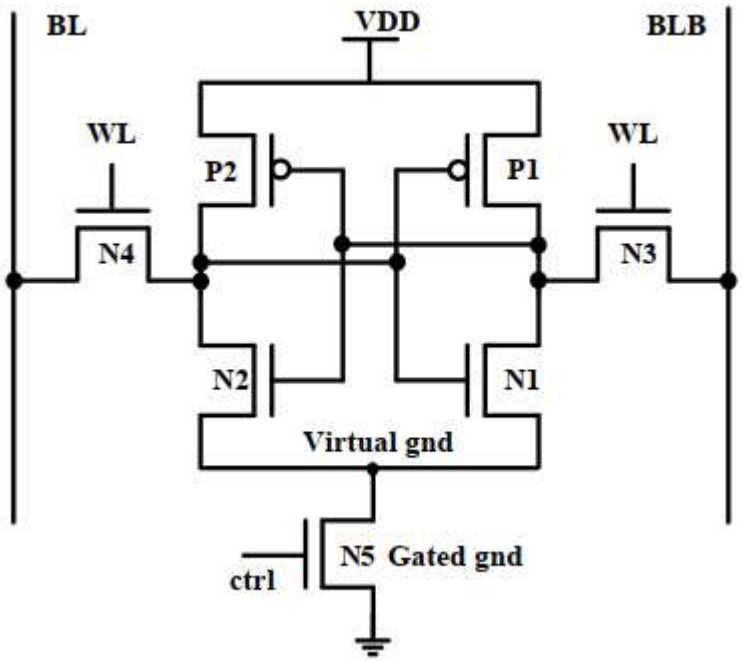


Fig. 6.5: Gated SRAM cell.[8]

6.3 GATED WITH N CONTROL SRAM

Figure 6.6 depicts the Gated with N control SRAM cell, functioning much like the Gated SRAM cell. It can interrupt the connection between Vdd and ground, aiming to minimize leakage current and refine delay performance. To fine-tune delay characteristics, a variety of transistors with distinct threshold voltages are employed. Those enclosed within a dotted circle feature higher threshold voltages, while those with lower threshold voltages are selected for swifter operation.

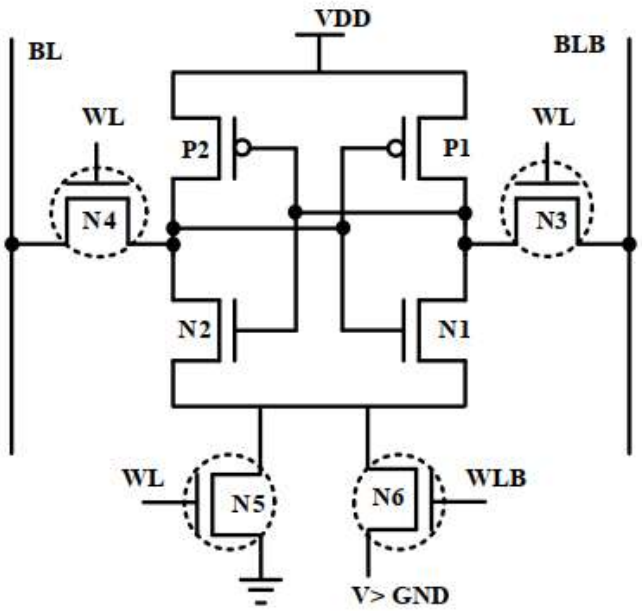


Fig. 6.6: NC SRAM cell [4]

CHAPTER 7

SIMULATION AND RESULT

In present era of microprocessor SRAM become the unavoidable part of it. But leakage current of the SRAM makes it more challenging. As size of the transistors are shrinking and leakage current also increasing with it. There is different type of leakage current are available (explained in previous chapter) which makes SRAM power hungry. Therefore, design engineer always tries to make SRAM with minimal power consumption.

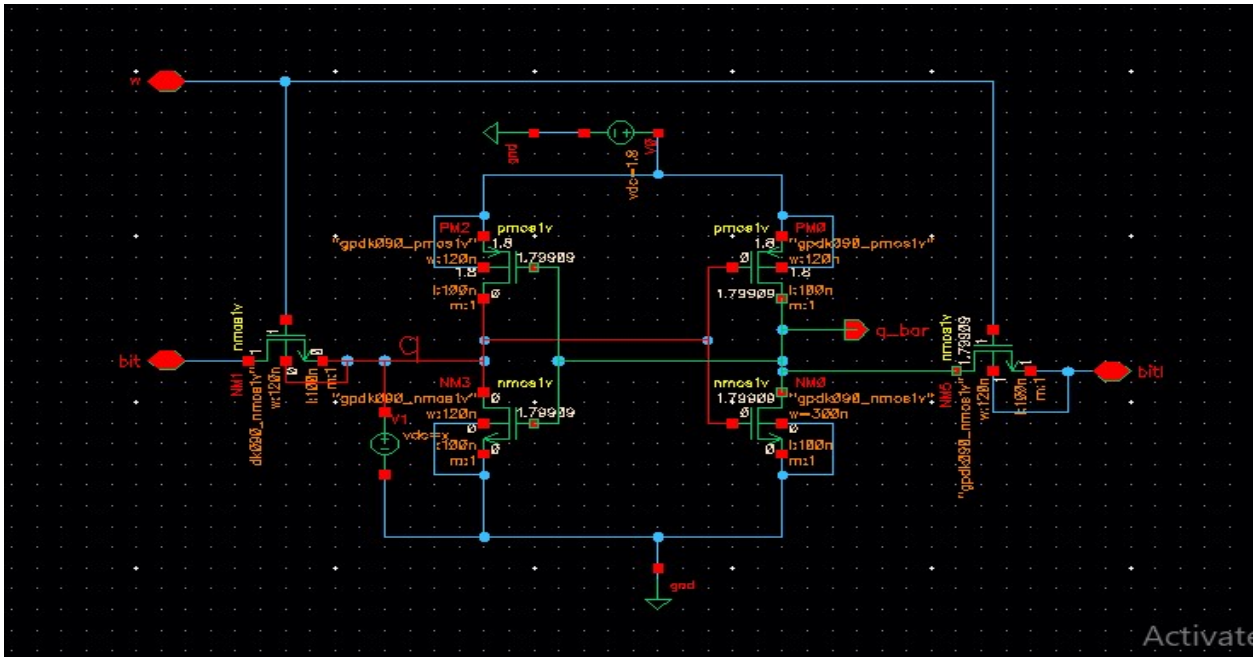


Fig. 7.1: Schematic of 6T SRAM

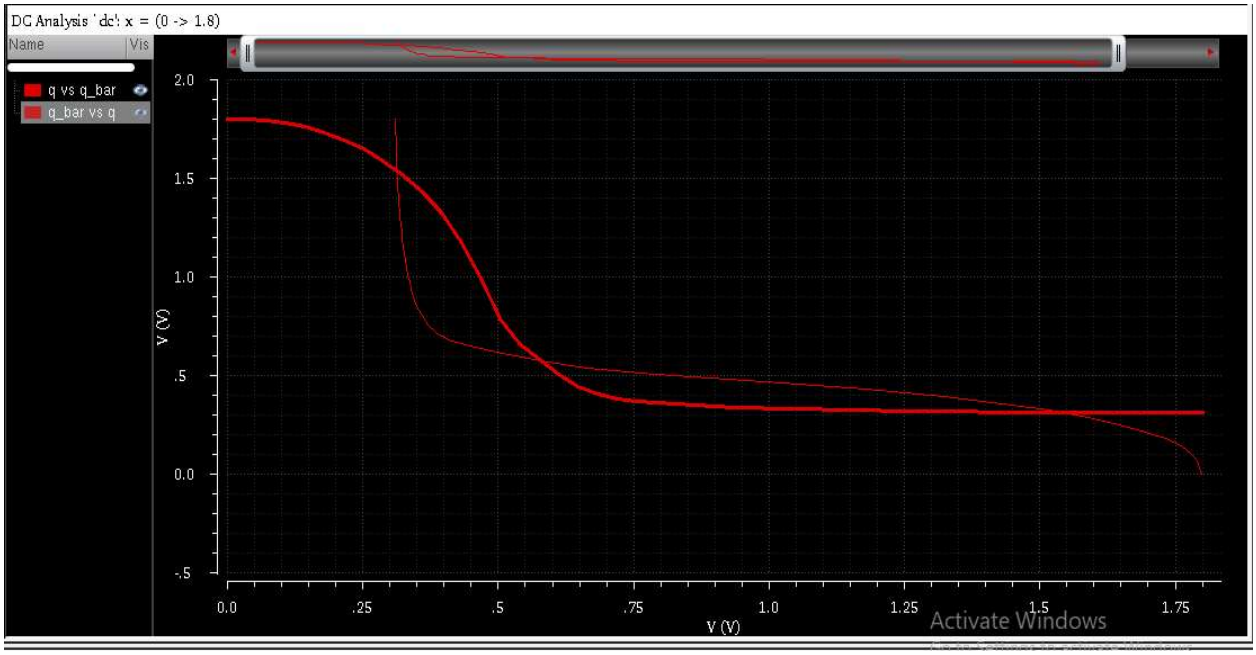


Fig. 7.2: Butterfly curve for 6T SRAM

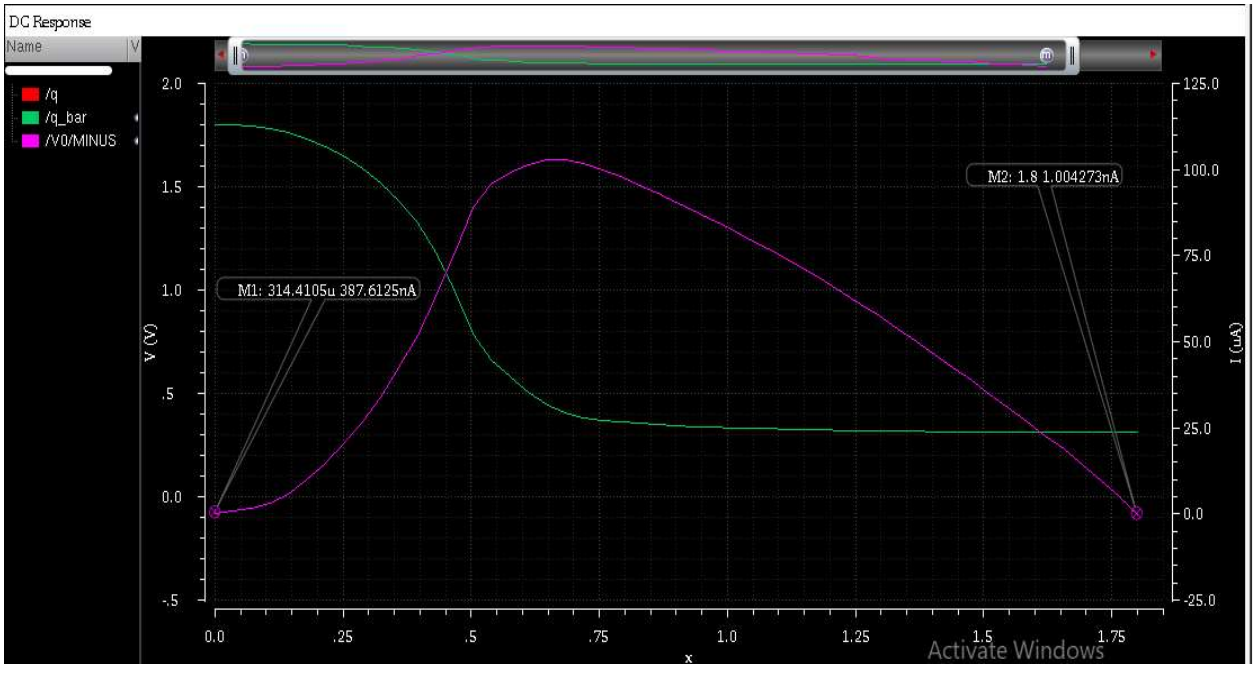


Fig. 7.3: leakage current of 6T SRAM

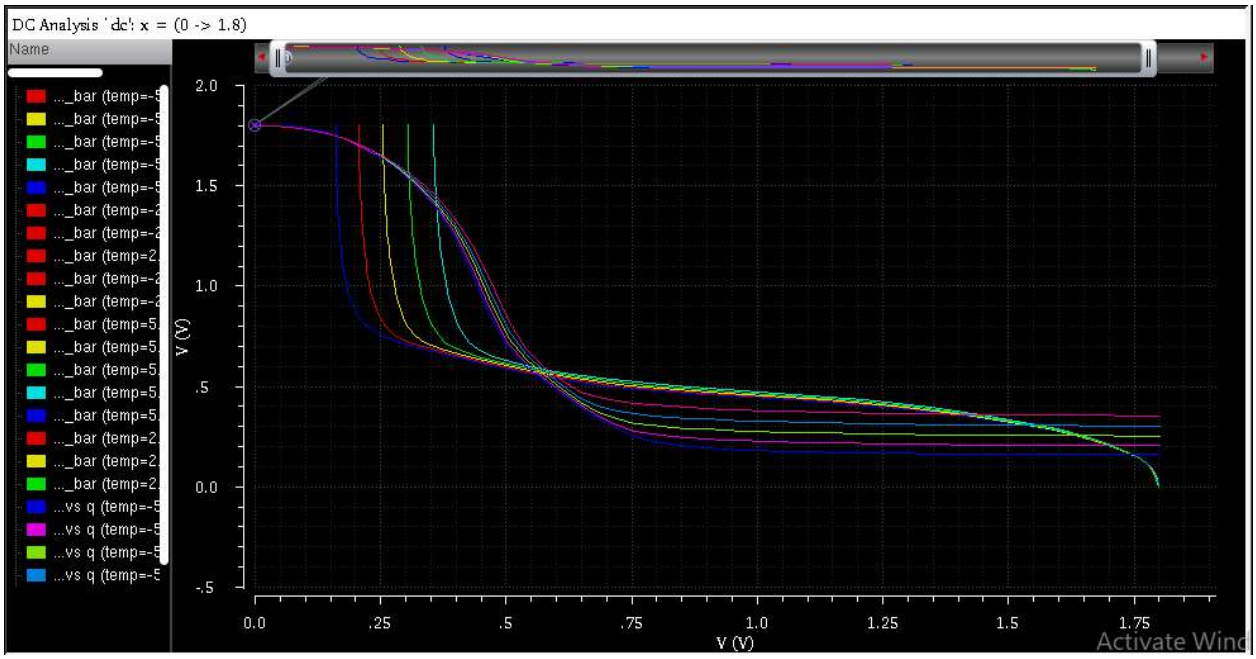


Fig. 7.4: Effect of temperature on SNM

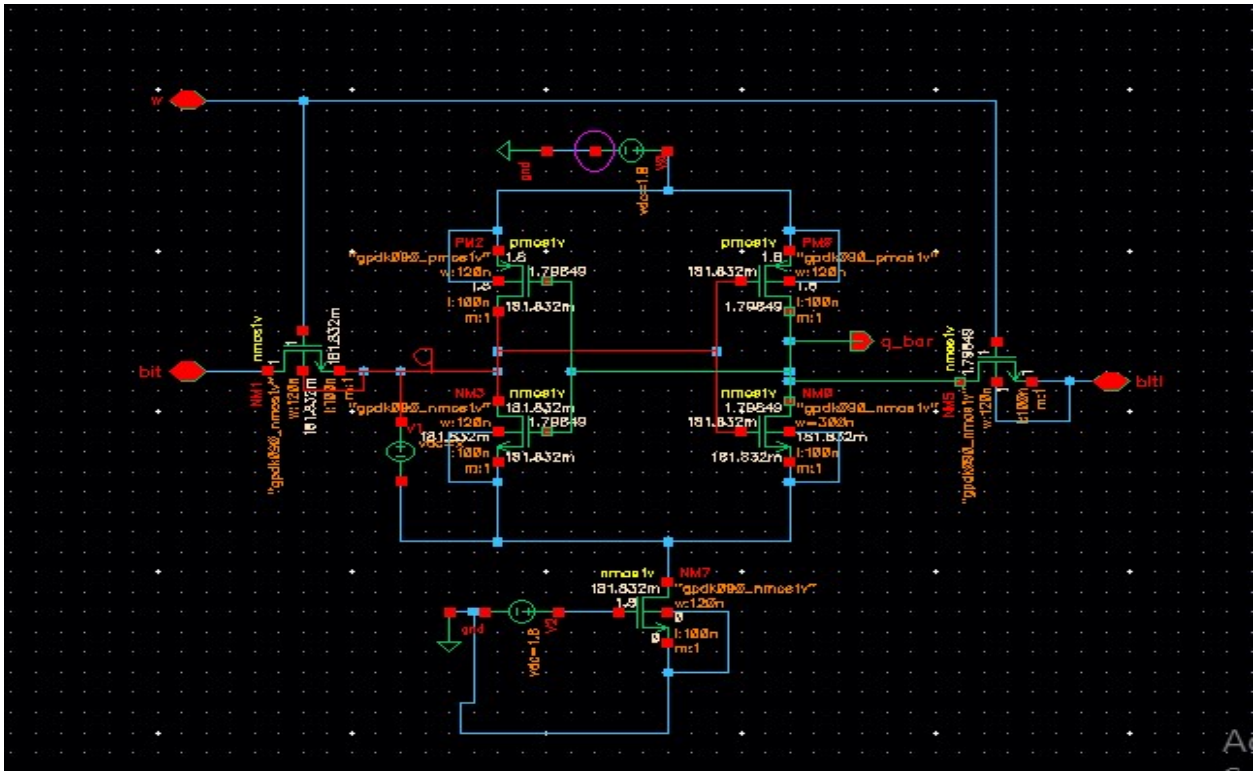


Fig. 7.5: Schematic of 2nd topology

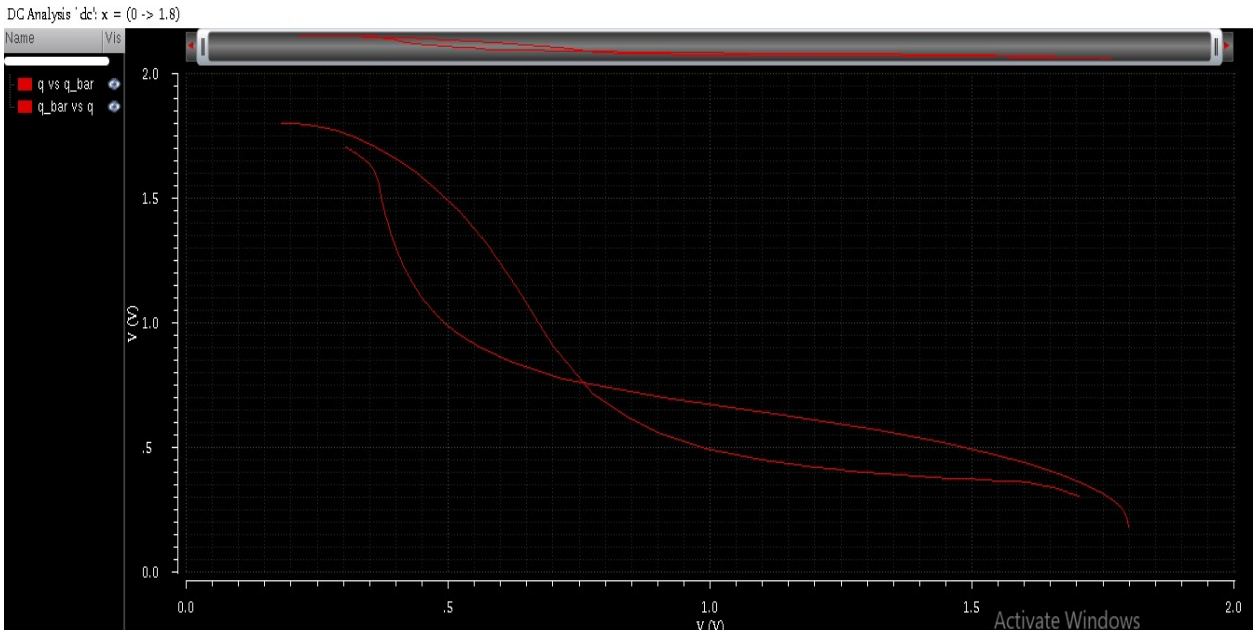


Fig. 7.6: Butterfly curve of 2nd topology

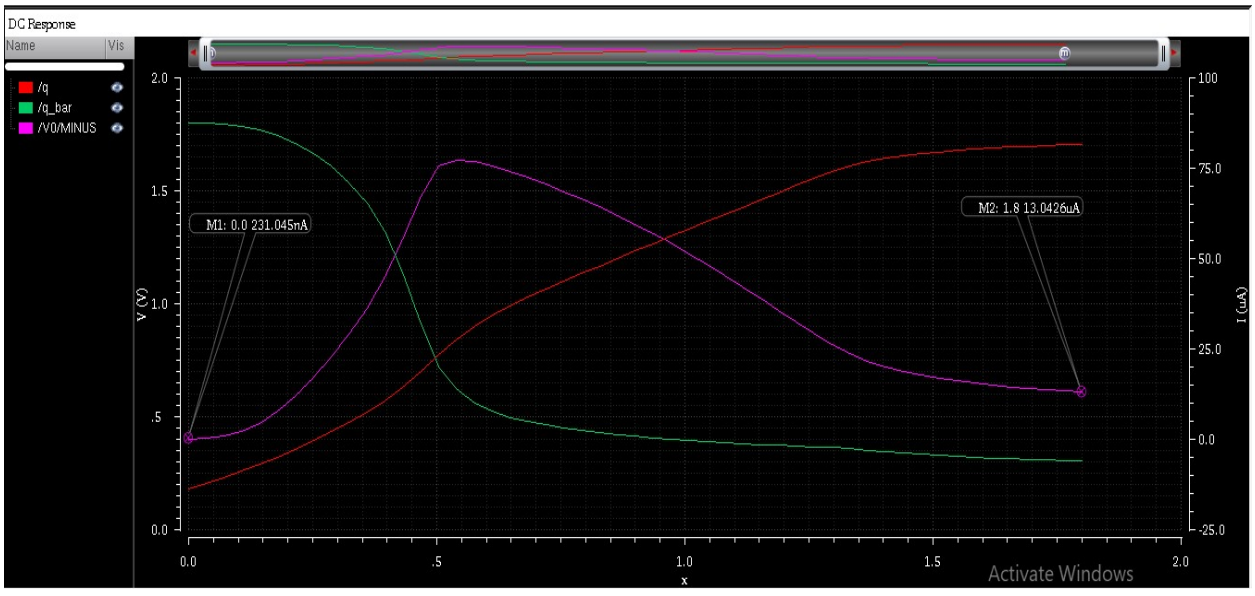


Fig. 7.7: leakage current of 2nd topology

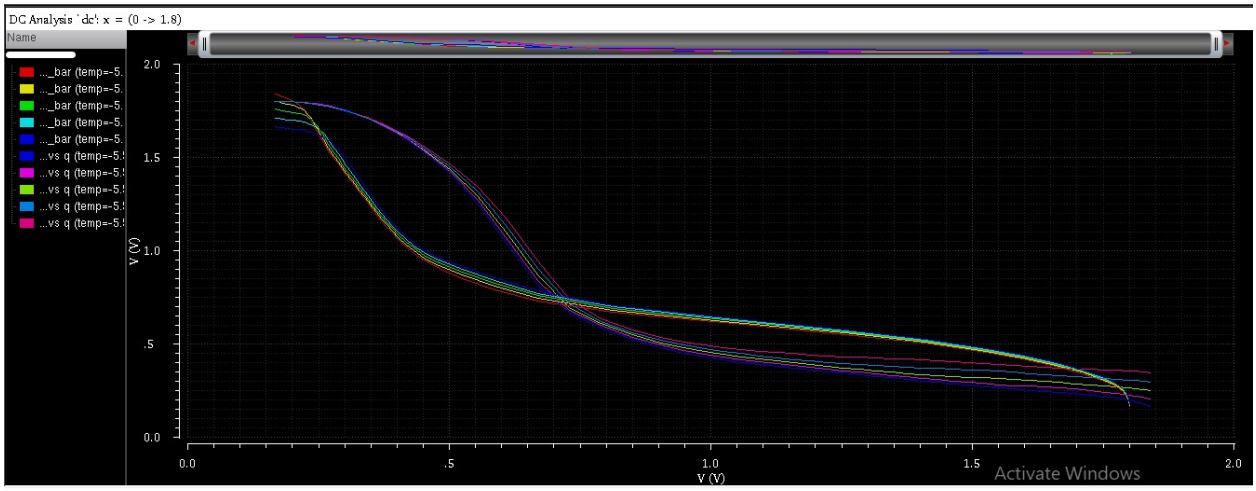


Fig. 7.8: temperature variation of 2nd topology

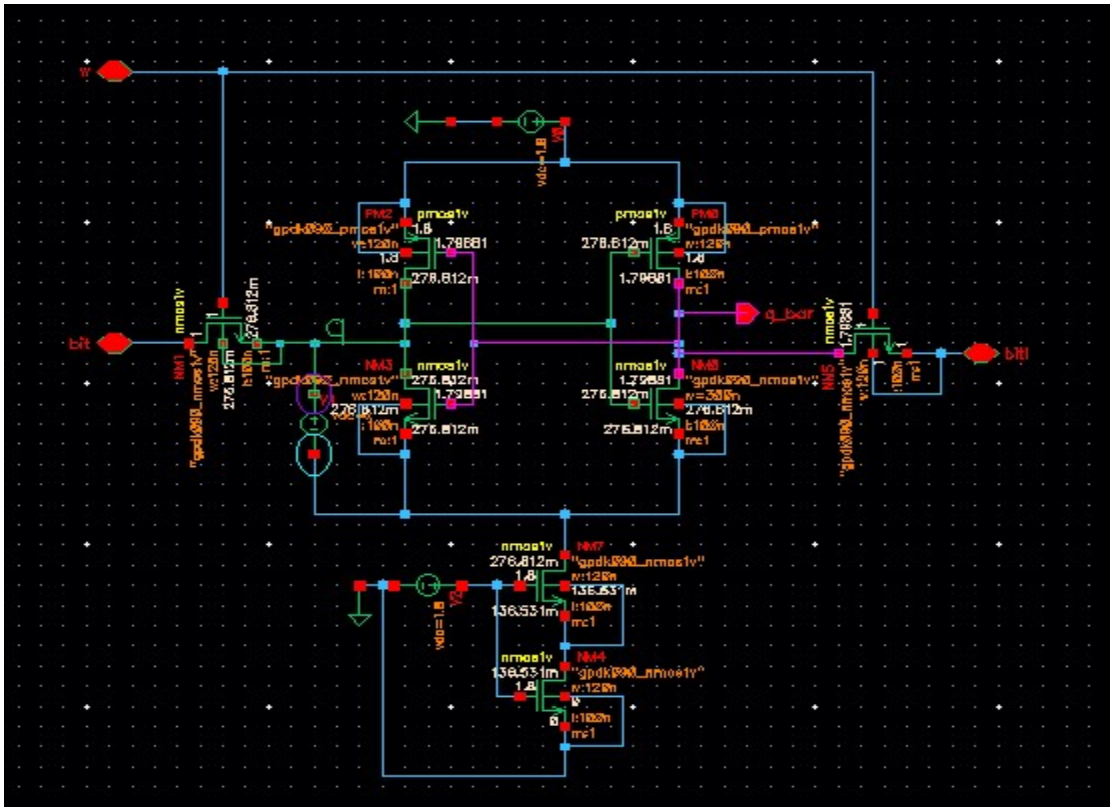


Fig. 7.9: Schematic of 3rd topology

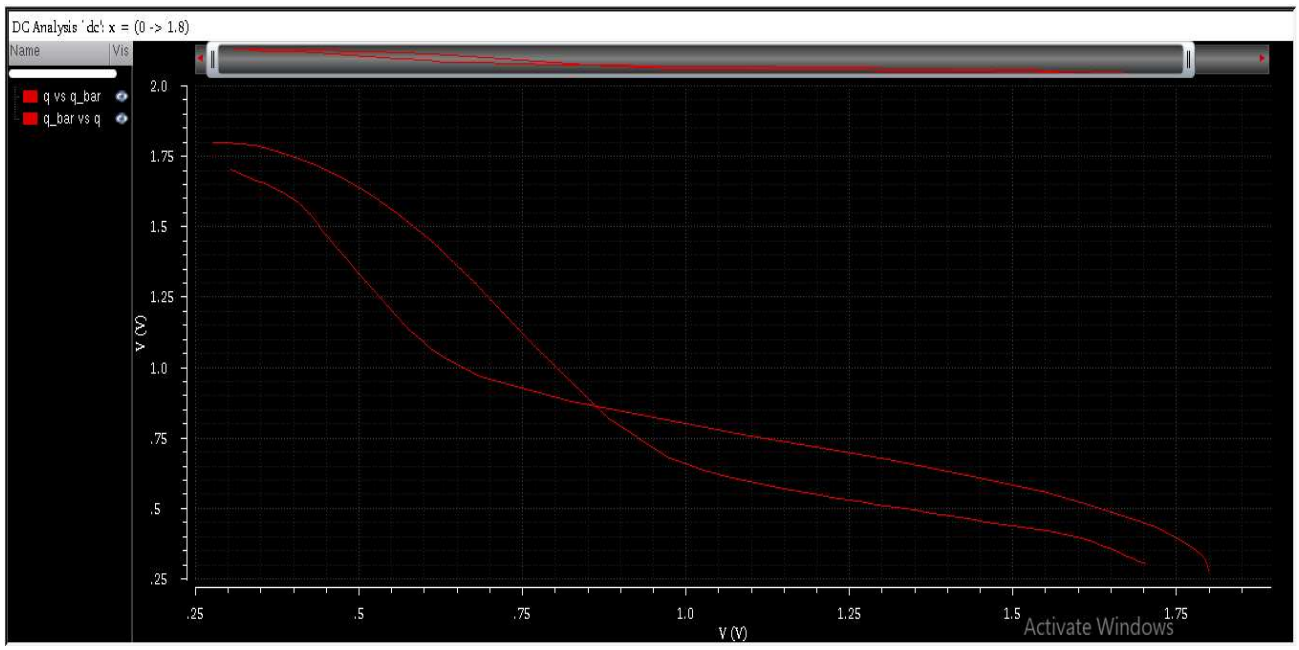


Fig. 7.10: Butterfly curve of 3rd topology

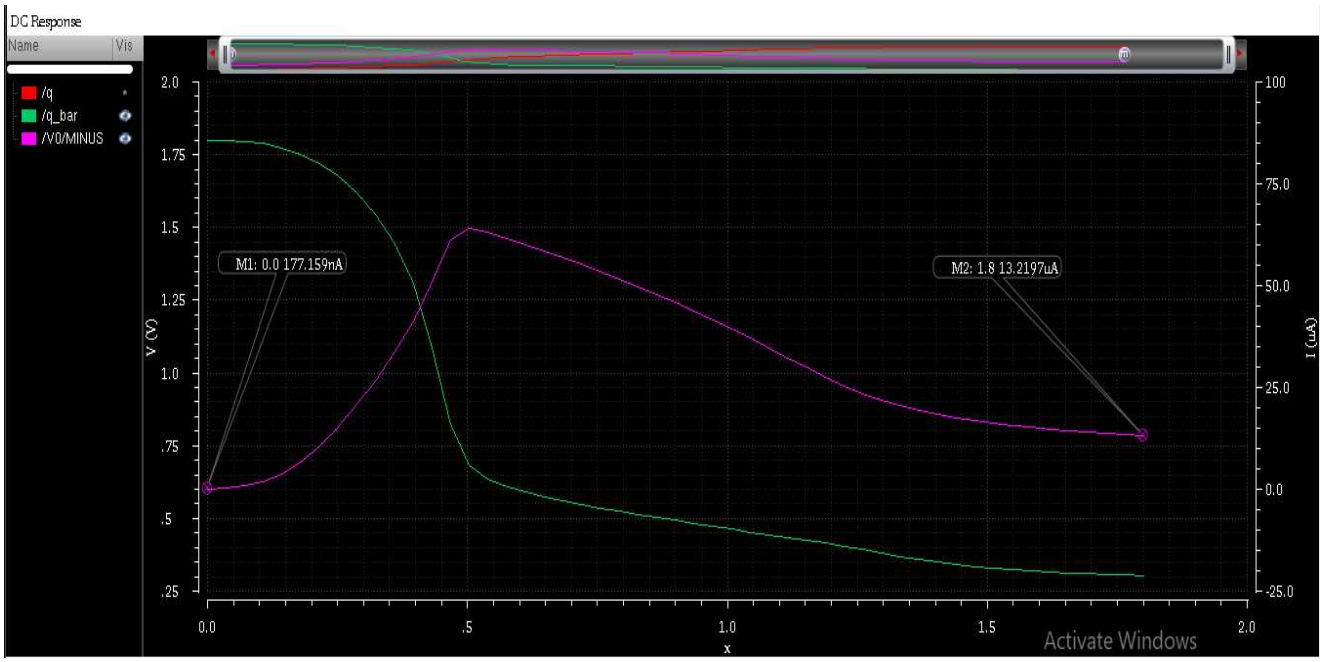


Fig. 7.11: leakage current of 3rd topology

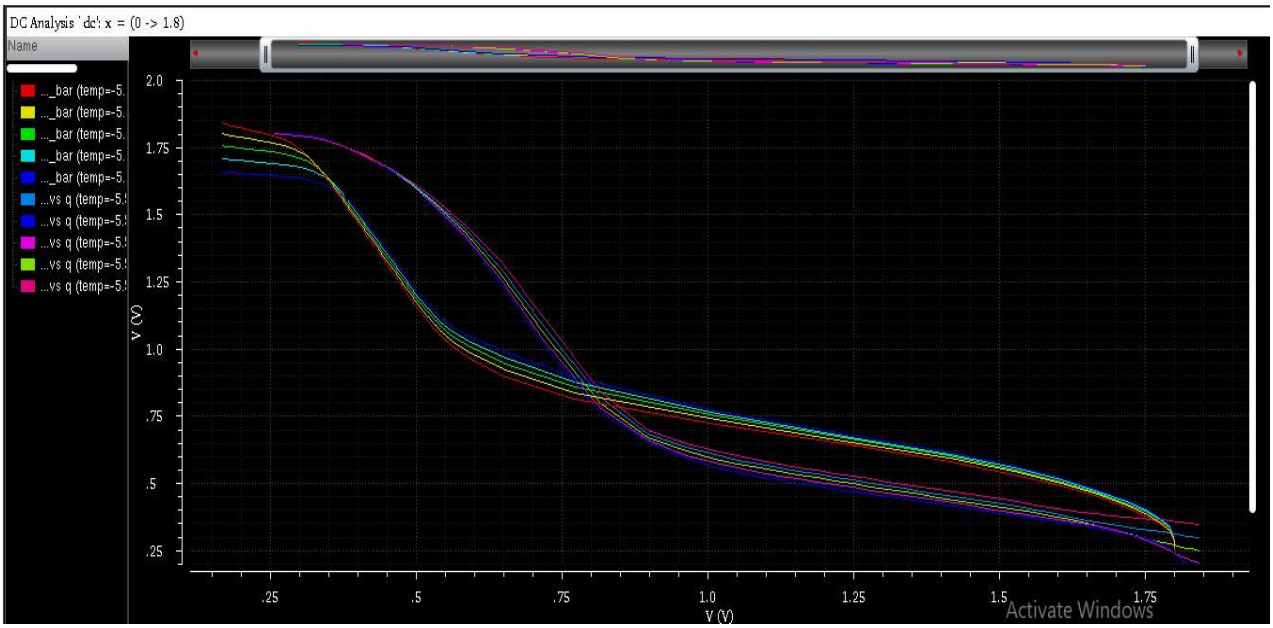


Fig. 7.12: temperature variation of 3rd topology

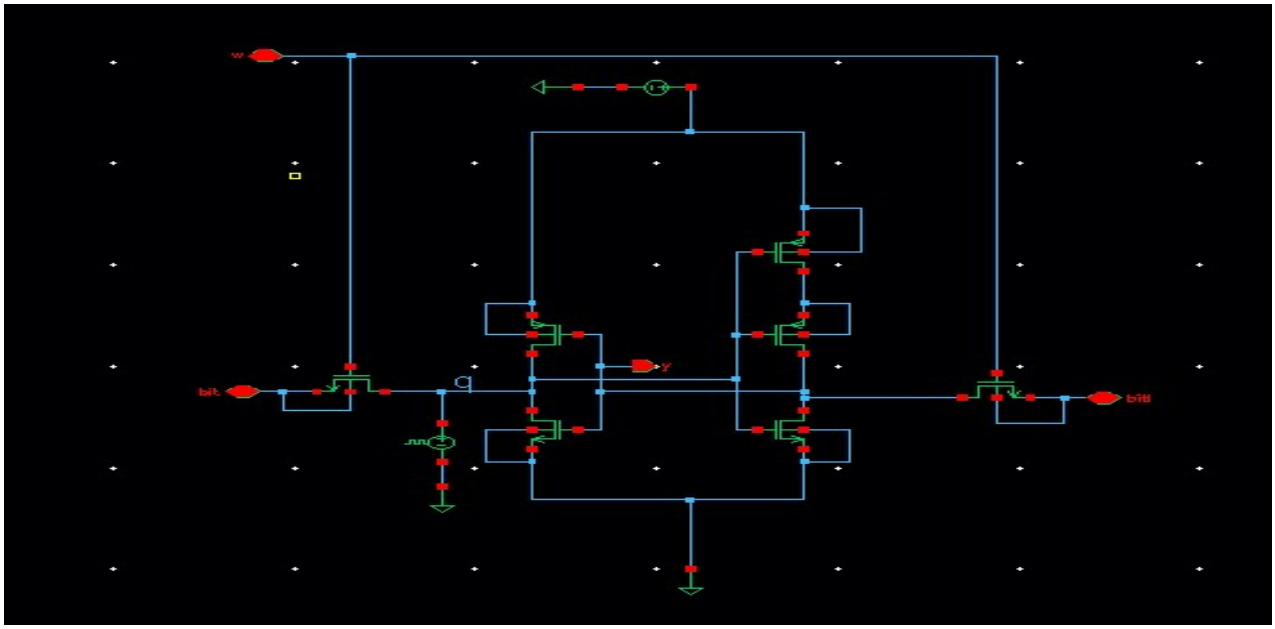


Fig 7.13: Schematic of 4th topology

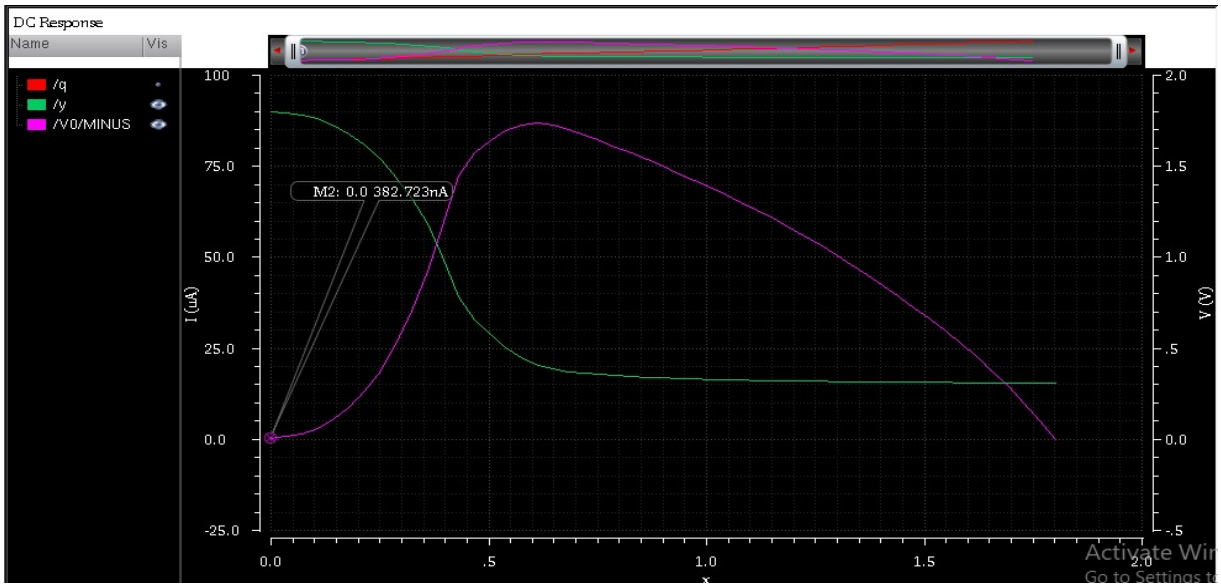


Fig 7.14: leakage current of 4th topology

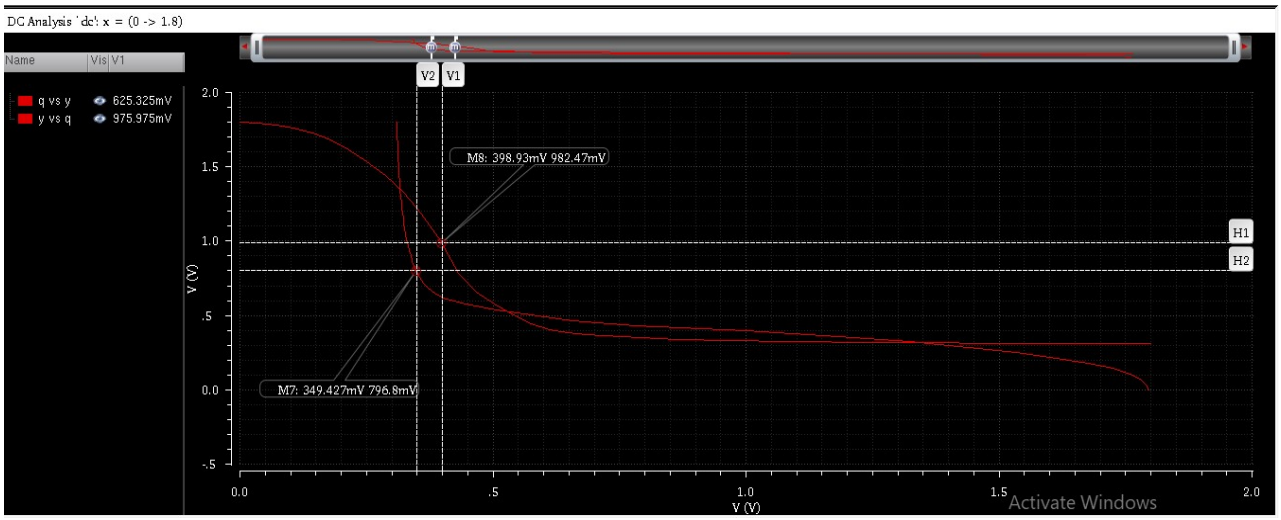


Fig. 7.15: SNM of 4th topology

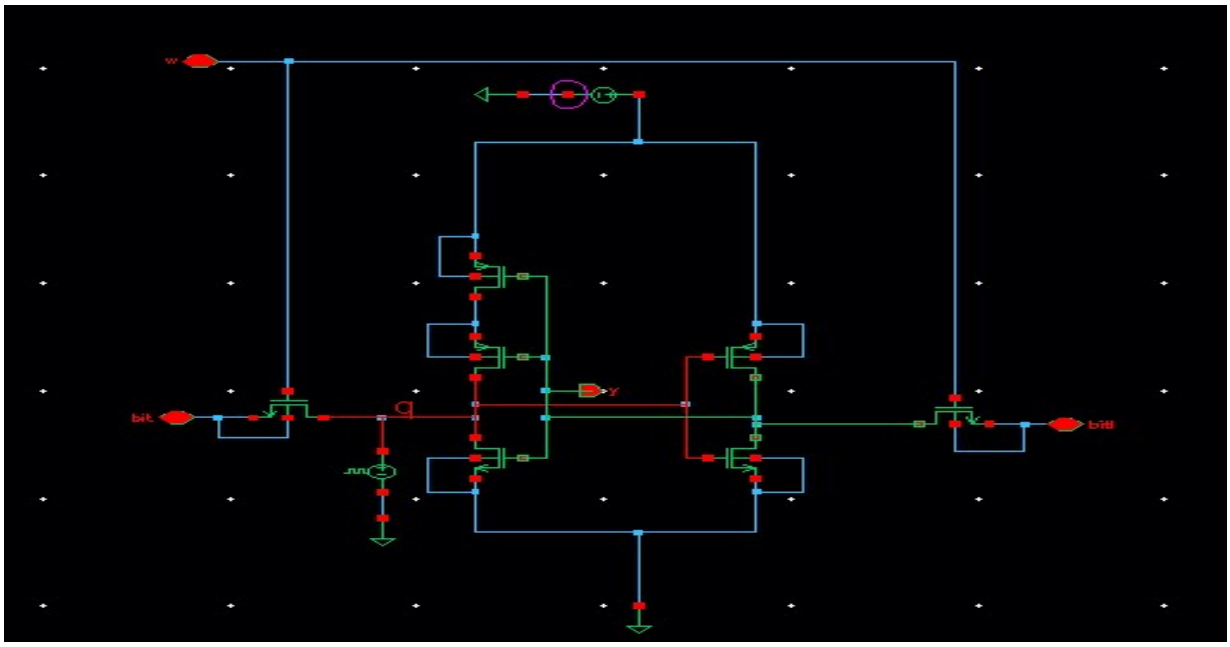


Fig 7.16: Schematic of 5th topology

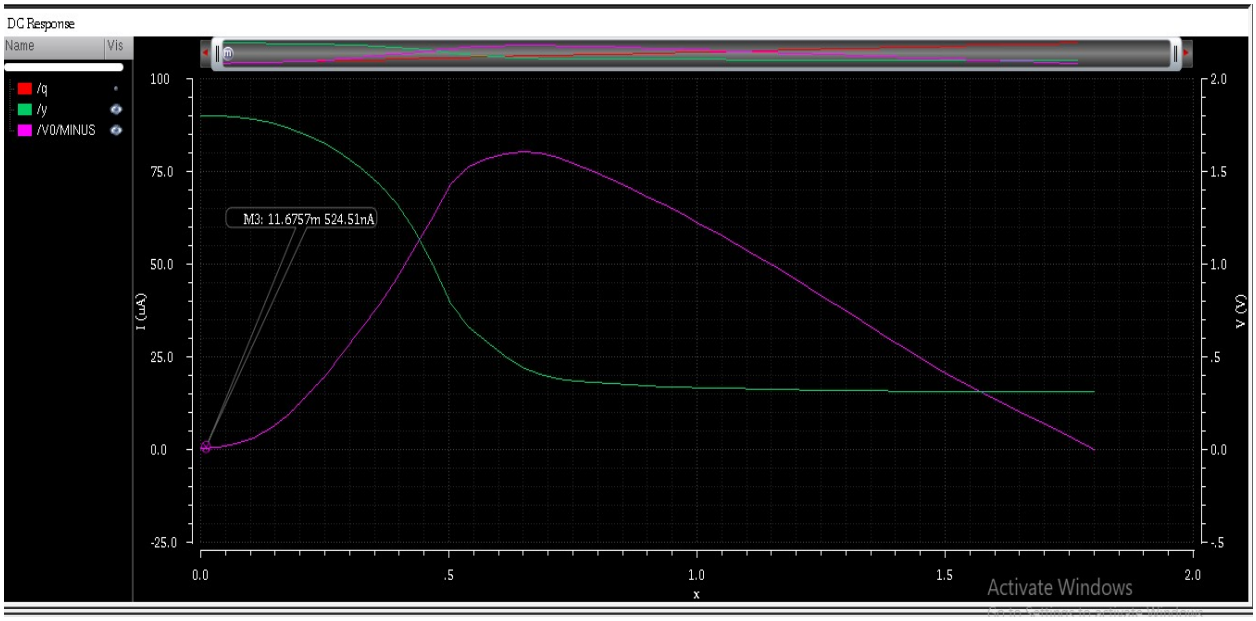


Fig 7.17: leakage current of 5th topology

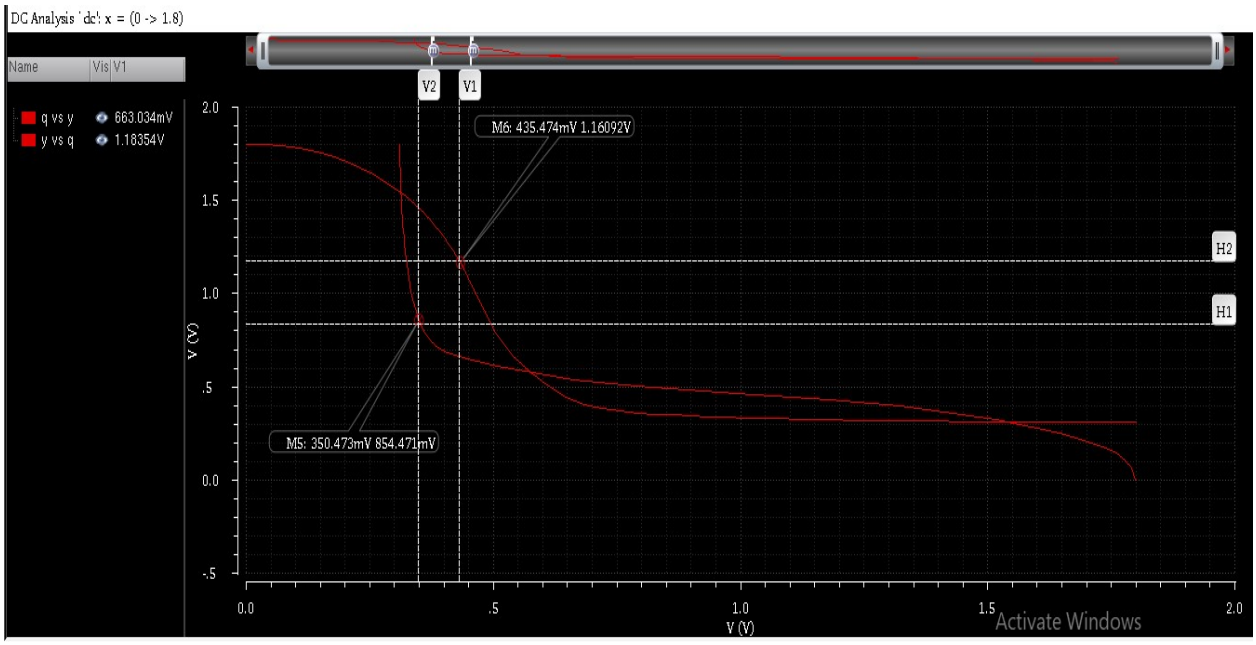


Fig. 7.18: SNM of 5th topology

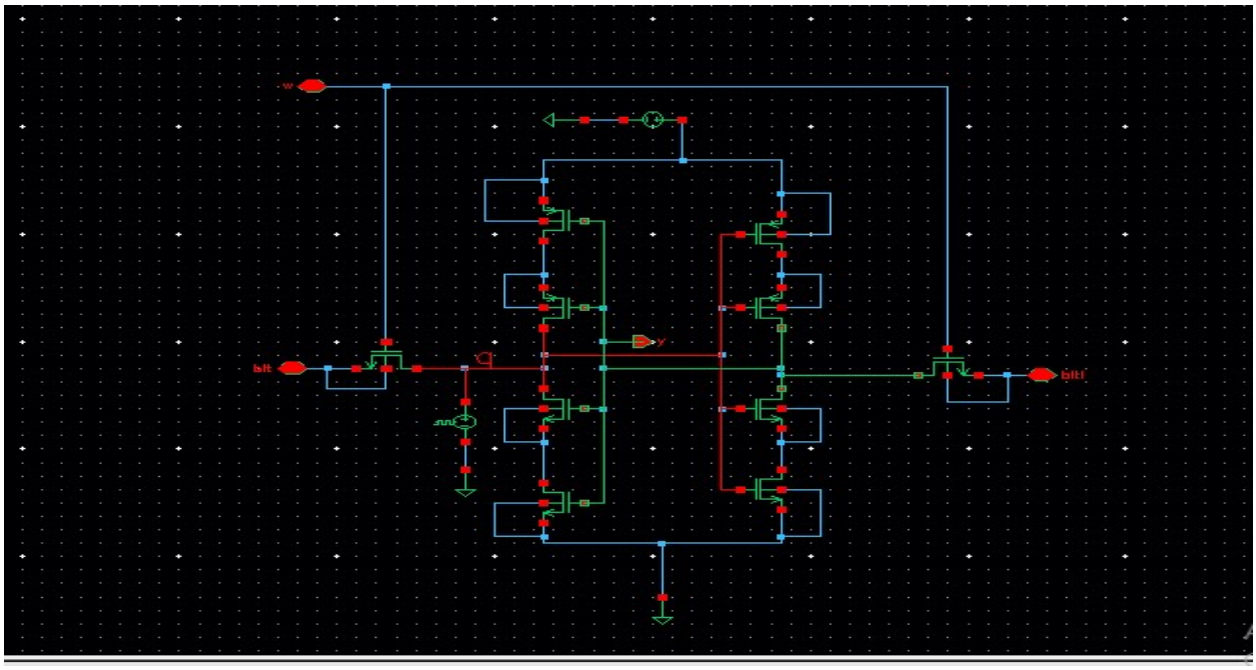


Fig 7.19: Schematic of 6th topology

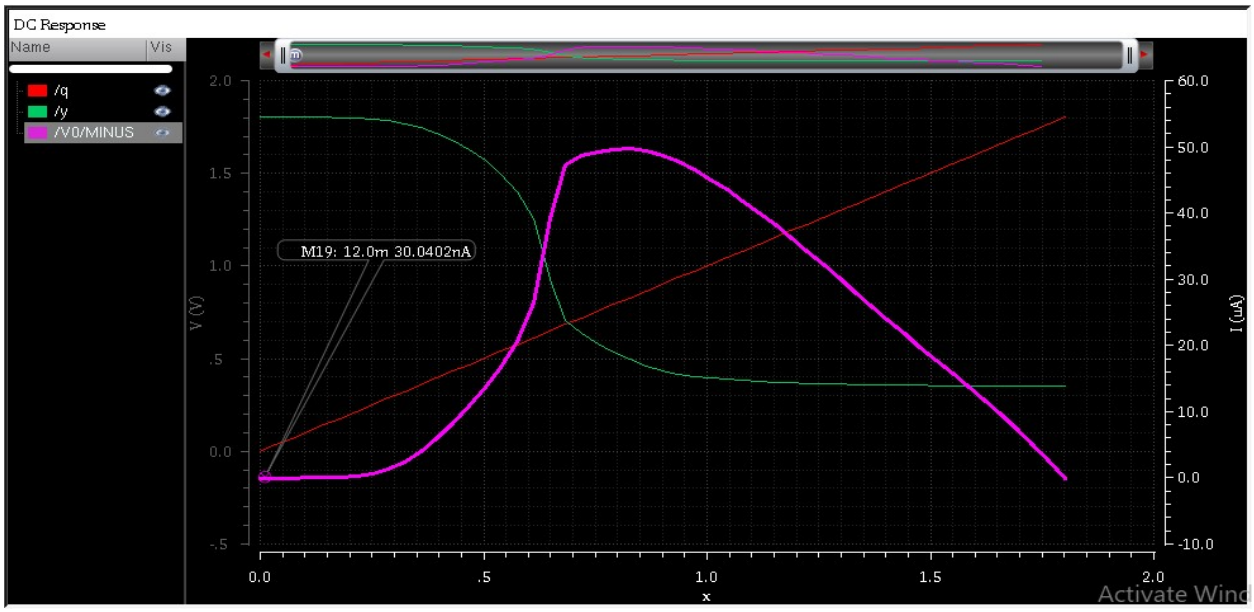


Fig 7.20: leakage current of 6th topology

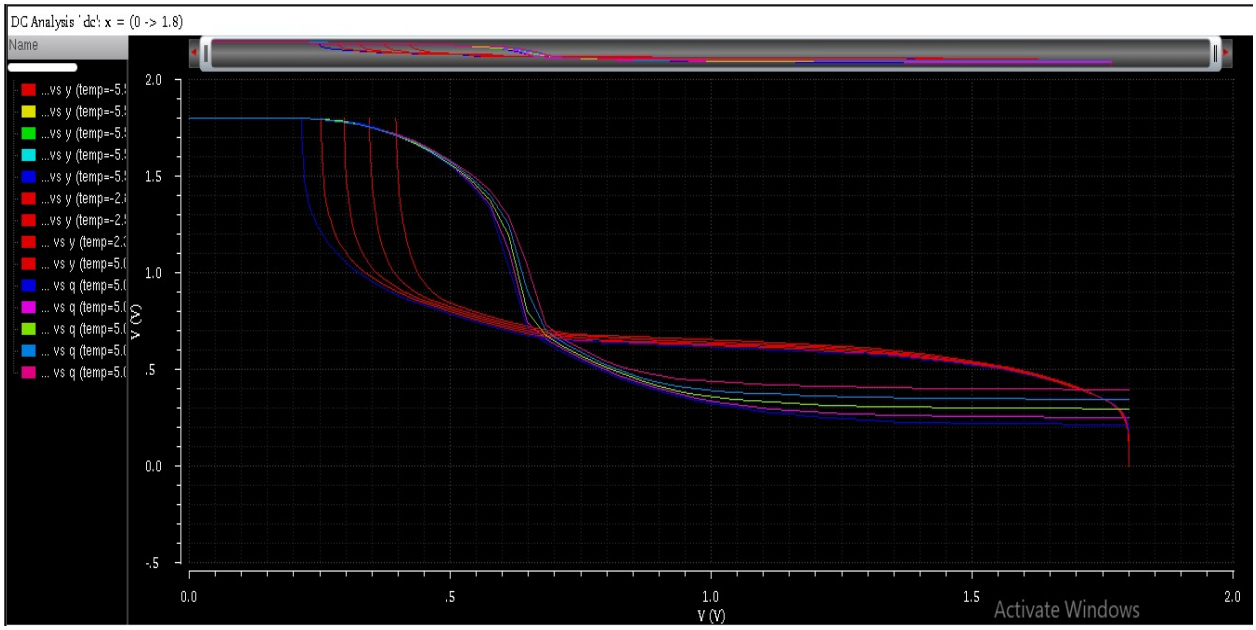


Fig. 7.21: temperature variation on 6th Topology

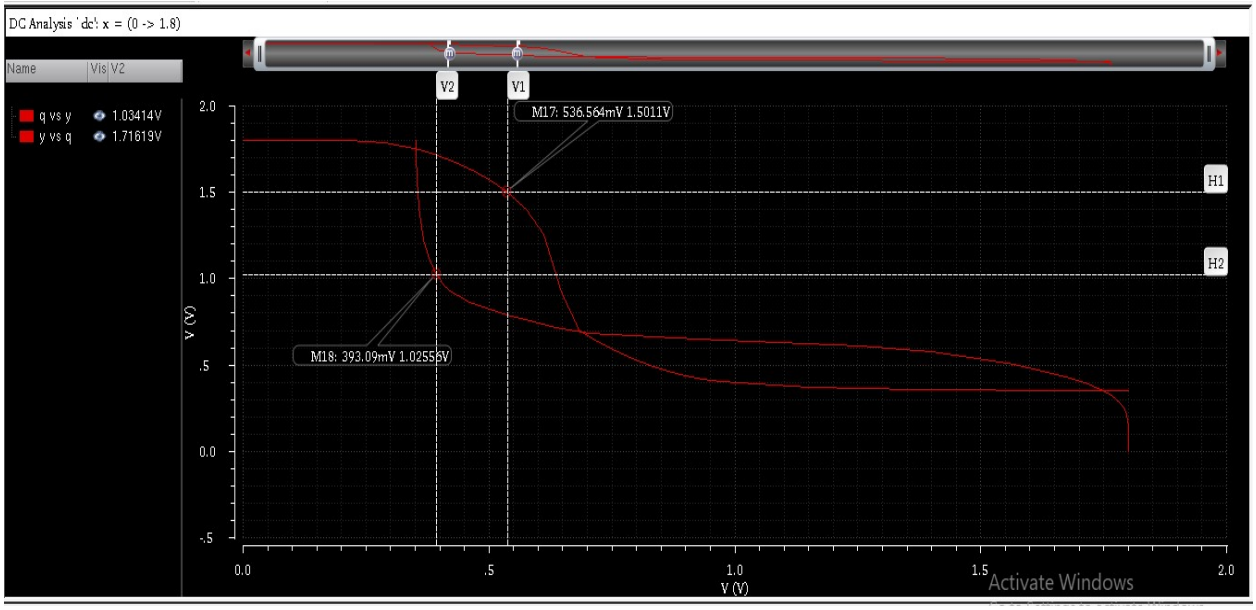


Fig. 7.22: SNM of 6th topology

COMPARISON

TOPOLOGIES	SNM	LEAKAGE CURRENT	NO. OF TRANSISTOR
6T SRAM	306.12	275.1	6
2nd Topology	700.13	230.2	7
3rd Topology	705.25	178.4	8
4th Topology	194.56	380.5	7
5th Topology	316.88	523.4	7
6th Topology	498.66	31.2	10

CONCLUSION

To reduce leakage current in traditional 6T SRAM cells, a modification has been made by incorporating extra transistors in series, known as the stacking technique. This approach, while increasing the overall area of the SRAM due to the additional transistors, effectively lowers power loss by reducing leakage current.

In 2nd topology one extra transistor is used tried to understand the stability of SRAM and also tried to analyse that when we make the schematic of SRAM then how it effect the stability of conventional 6T SRAM. It gives better static Noise margin than the 6T SRAM. The leakage current is less. It reduces the leakage current by 41.45%, than the 6T SRAM cell and its performance are better than the 6T SRAM with different temperature.

The 6T SRAM cell has been improved by incorporating two additional NMOS transistors between the storage cell and ground. This stacking technique not only enhances performance over the traditional 6T SRAM but also results in a better Static Noise Margin (SNM). The SNM of the second and third configurations are almost the same. Moreover, this modification significantly lowers power loss by reducing leakage current, achieving a 53.29% decrease Compared to the Conventional 6T SRAM cell. It also offers superior performance across different temperature ranges.

The 4th and 5th topology has higher leakage current than the 6T SRAM. These two methodologies require 7 transistors each. It neither reduces leakage current nor gives better noise margin.

The 6th topology requires 10 transistors. It gives better noise marine as well as lesser leakage current than the 6T SRAM. It reduces the leakage current nearly 9 times than the 6T SRAM. The Noise Margin of 6th topology is far better than the 6T SRAM. So overall 6th topology is better than among all the above topologies, with one demerit it will take more are because of 10 transistors.

REFERENCES

1. Shokoufeh Naghizadeh1, · Mohammad Gholami2, “Two Novel Ultra-Low-Power SRAM Cells with Separate Read and Write Path” Springer Science Business Media, LLC, part of Springer Nature 2018.
2. Debasis Mukherjee,Hemanta Kr.Mondal,‖Static Noise Margin Analysis of SRAM Cell For High Speed Application‖, IJCSI International Journal of Computer Science Issues, Vol. 7,Issue 5,September 2010.
3. Jan M.Rabaey Anantha Chandrakasan Borivoje Nikolic ‖Digital Integrated Circuits‖, Pearson Education Electronics ,2003.
4. Sung-Mo Kang and Yusuf Leblebici ‖CMOS Digital Integrated Circuits‖, TATA McGRAW-HILL EDITION 2003.
5. Kunihiko Yamaguchi, Hiroaki Nambu, et. al. ―A 1.5-ns Access Time, 78-pm’ Memory-Cell Size, 64-kb ECL-CMOS SRAM‖, IEEE Journal of Solid-State Circuits, Vol. 27, No.2. February 1992.
6. Rakesh Dayaramji Chandankhede ―Design of High Speed Sense Amplifier for SRAM‖, 2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT).
7. Sreerama Reddy G M, P Chnadrasekhara Reddy, ‖Design and implementation of 8Kbits Low Power SRAM in 180nm Technology ‖,Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol II IMECS 2009,March 18-20,2009.
8. Harekrishna Kumar1 · V. K. Tomar,”A Review on Performance Evaluation of Different Low Power SRAM Cells in Nano- Scale Era”Springer Science Business Media, LLC, part of Springer Nature 2020.
9. PavankumarBikki, PitchaiKaruppanan, “SRAM Cell Leakage Control Techniques for Ultra Low Power Application: A Survey “scientific research publication,Motilal Nehru National Institute of Technology, Allahabad, India, February 2017.
10. T. Chappell, et. al., “A 2-ns Cycle, 3.8-ns Access 512-Kb CMOS ECL SRAM with a Fully Pipelined Architecture”, IEEE Journal of Solid State Circuits, vol. 26, no. 11, pp. 1577-1585, Nov. 1991.
11. H. Nambu, et. al., “A 1.8ns Access, 550MHz 4.5Mb CMOS SRAM”, 1998 IEEE International Solid State Circuits Conference, Digest of Technical Papers, pp. 360-361.
12. G. Braceras, et. al., “A 350MHz 3.3V 4Mb SRAM fabricated in a 0.3µm CMOS process”, 1997 IEEE International Solid State Circuits Conference, Digest of Technical Papers, pp. 404-405.
13. M. Yoshimoto, et. al., “A 64kb CMOS RAM with divided word line structure”, I983 IEEE International Solid State Circuits Conference, Digest of Technical Papers, pp. 58-59.

14. T. Hirose, et. al., “A 20nS 4Mb CMOS SRAM with Hierarchical Word Decoding Architecture”, 1990 IEEE International Solid State Circuits Conference, Digest of Technical Papers, pp.132-133.
15. R. C. Jaeger, “Comments on ‘An optimized output stage for MOS integrated circuits’,” IEEE Journal of Solid State Circuits, vol. SC-10, no. 3, pp. 185-186, June 1975.
16. C. Mead and L. Conway, Introduction to VLSI Systems, Reading, MA, Addison-Wesley, 1980.
17. N. C. Li, et. al., “CMOS tapered buffer”, IEEE Journal of Solid State Circuits, vol. 25, no. 4, pp. 1005-1008, August 1990.
18. J. Choi, et. al., “Design of CMOS tapered buffer for minimum power-delay product”, IEEE Journal of Solid State Circuits, vol. 29, no. 9, pp. 1142-1145, September 1994.
19. B. S. Cherkauer and E. G. Friedman, “A unified design methodology for CMOS tapered buffers”, IEEE Journal of Solid State Circuits, vol. 3, no. 1, pp. 99-110, March 1995.
20. M. Yoshimoto, et. al., “A 64kb CMOS RAM with divided word line structure”, I983 IEEE International Solid State Circuits Conference, Digest of Technical Papers, pp. 58-59.
21. O. Minato, et. al., “2Kx8 bit Hi-CMOS static RAMs”, IEEE Journal of Solid State Circuits, vol. SC-15, no. 4, pp. 656-660, August 1980.
22. Y. Kohno, et. al., “A 14-ns 1-Mbit CMOS SRAM with variable bit organization”, IEEE Journal of Solid State Circuits, vol. 29, no. 9, pp. 1060-1065, October 1988.
23. H. Nambu, et. al., “A 1.8ns Access, 550MHz 4.5Mb CMOS SRAM”, 1998 IEEE International Solid State Circuits Conference, Digest of Technical Papers, pp. 360-361.
24. G. Braceras, et. al., “A 350MHz 3.3V 4Mb SRAM fabricated in a 0.3µm CMOS process”, 1997 IEEE International Solid State Circuits Conference, Digest of Technical Papers, pp. 404-405.
25. K. Nakamura, et. al., “A 500MHz 4Mb CMOS pipeline-burst cache SRAM with point-to-point noise reduction coding I/O”, 1997 IEEE International Solid State Circuits Conference, Digest of Technical Papers, pp. 406-407.
26. K. Osada, et. al., “A 2ns Access, 285MHz, Two-Port Cache Macro using Double Global Bit-Line Pairs”, ISSCC Digest of Technical Papers, pp. 402-403, February 1997.
27. K. Seno, et. al., “A 9-ns 16-Mb CMOS SRAM with offset-compensated current sense amplifier”, IEEE Journal of Solid State Circuits, vol.28, no. 11, Nov. 1993.
28. O. Minato, et. al., “A 20ns 64K CMOS SRAM”, 1984 IEEE International Solid State Circuits Conference, Digest of Technical Papers, pp. 222-223.
29. S. Yamamoto, et. al., “256k CMOS SRAM With Variable Impedance Data-Line Loads”, IEEE Journal of Solid State Circuits, vol. SC-20, pp. 924-928, October 1985.
30. K. J. Schultz, et. al., “Low-Supply-Noise Low-Power Embedded Modular SRAM for Mixed Analog-Digital ICs”, Proceedings of 1992 IEEE Custom Integrated Circuits Conference, p. 7.1/1-4.

31. P. Reed, et. al., “A 250MHz 5W RISC Microprocessor with On-Chip L2 Cache Controller”, 1997 IEEE International Solid State Circuits Conference, Digest of Technical Papers, pp. 412-413.
32. K. Sasaki, et. al., “A 15-ns 1-Mbit CMOS SRAM”, IEEE Journal of Solid State Circuits, vol. 23, no. 5, pp. 1067-1071, October 1988.
33. M. Matsumiya, et. al., “A 15-ns 16-Mb CMOS SRAM with interdigitated bit-line architecture”, IEEE Journal of Solid State Circuits, vol. 27, no. 11, pp. 1497-1502, November 1992.
34. Anantha P. Chandrakasan and Jan M. Rabaey, “Digital Integrated Circuits: A Design Perspective”.
35. John L Hennessy and David A Patterson, Computer architecture: a quantitative approach, Elsevier, 2012.
36. Rajeshwari Banakar, Stefan Steinke, Bo-Sik Lee, M Balakrishnan, and Peter Marwedel, Scratchpad memory: design alternative for cache on-chip memory in embedded systems, Proceedings of the tenth international symposium on Hardware/software codesign, ACM, 2002, pp. 73–78.
37. Jawar Singh, Saraju P Mohanty, and Dhiraj Pradhan, Robust sram designs and analysis, Springer, 2012.
38. Yih Wang, Hong Jo Ahn, Uddalak Bhattacharya, Zhanping Chen, Tom, Coan, Fatih Hamzaoglu, Walid M Hafez, Chia-Hong Jan, Pramod Kolar, Sarvesh H Kulkarni, et al., A 1.1 ghz 12 μa/mb-leakage sram design in 65 nm ultra-low-power cmos technology with integrated leakage reduction for mobile applications, Solid-State Circuits, IEEE Journal of 43 (2008), no. 1, 172–179.