Performance Analysis of FinFET based SRAM at Nano-scaled Technology Nodes

A Thesis Submitted in Fulfilment of the Requirement for the Award of the Degree of

Master of Engineering

In

Electronics and Communication Engineering

Submitted By

Simranjit Singh Mehra

Roll No: 801661024

Under Supervision of

Dr. Karamjit Singh Sandha

Assistant Professor,ECED

THAPAR INSTITUTE

OF ENGINEERING & TECHNOLOGY

(Deemed to be University)

ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT

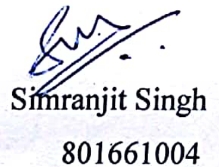
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY), PATIALA, PUNJAB

JULY, 2018

DECLARATION

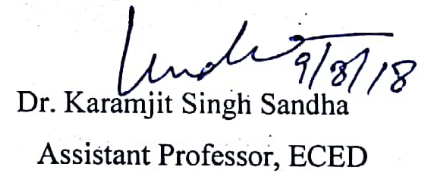
I hereby declare that the thesis entitled "Performance Analysis of FinFET based SRAM at Nano-scaled Technology Nodes" is an authentic record of my study carried out as per requirement for the award of degree of M.E. (Electronic and communication) of Thapar Institute of Engineering and Technology, Patiala, under the supervision of Dr. Karamjit Singh Sandha (Assistant Professor, ECED)



Date: 9/8/2018

Simranjit Singh
801661004

It is certified that above statement made by the student is correct to best of my knowledge and brief.



Date: 9/8/2018

Dr. Karamjit Singh Sandha
Assistant Professor, ECED

ACKNOWLEDGEMENT

I wish to express my deep gratitude and sincere thanks to my supervisor, **Dr. Karamjit Singh Sandha**, Assistant Professor, Electronics and Communication Department (ECED), Thapar Institute of Engineering and Technology, Patiala, for his invaluable guidance, constant encouragement, constructive comments, sympathetic attitude, and immense motivation, which has sustained my efforts at all stages of this work. His valuable advice and suggestions for the corrections, modifications and improvement did enhance my work.

I would like to express my gratitude to **Dr. Alpana Agarwal**, The Head of Electronics and Communication Department (ECED), Thapar Institute of Engineering and Technology, Patiala, for providing me with adequate environment in carrying out the work.

Finally, I want to extend my gratitude to all those persons who directly or indirectly helped me in carrying out this work in right direction.

Simranjit Singh
801661024
ME- ECE

ABSTRACT

SRAM is widely used type of memories in low powered consumer electronics. In fact, the majority of transistors found in many integrated circuits are those utilized in the SRAM bit cells with the percentage die occupied by this type of memory approaching 85%-90% integrated circuits. With the rise in demand for the low power high-speed devices, like smartphones and tablets, and the emergence of IoT devices, the need for scaling down SRAM has become a necessity. Since SRAM is typically constructed from traditional CMOS devices, all of the issues associated with MOSFET scaling are applicable to scaling of SRAM.

This work focuses on the study of the 6T FinFET SRAM, its advantages and disadvantages over MOSFET based 6T SRAM and different performance metrics such as delay time, power delay product (PDP) and static noise margin (SNM). A standard 6T SRAM cell is realized using predictive technology models (PTM) at 7nm, 10nm, 14nm ,16nm and 20nm technology nodes. These are evaluated and compared for previously mentioned performance metrics. Furthermore, the SRAM cell was power gated using fine grain gating technique and then compared with standard cell SRAM. The goal is to identify the trends for different parameters at different technology nodes. The models were then simulated in a variable temperature environment. The results conclude that the 7nm FinFET SRAM cells performs better at all aspects and the most resilient under variable temperature, as suggested theoretically, followed by SRAM cells at 10nm ,14nm, 16nm and 20nm nodes. It was also concluded that the gated versions of SRAM cells perform poorly then their standard counter parts but have improved PDP and lower power dissipation.

CONTENTS

DECLARATION..i

ACKNOWLEDGEMENT...ii

ABSTRACT..iii

List of Figures...vi

List of Tables……………………………………………………………………………………………ix

List of Abbreviations..v

1. INTRODUCTION ***1***

1.1 OVERVIEW *1*

1.2 TYPES OF MEMORIES *1*

1.2.1. Secondary Memory *3*

1.2.1.1. Sequential *3*

1.2.1.2 Random *3*

1.2.2 Primary Memory *4*

1.2.2.1 Read Only Memory *4*

1.2.2.2 Random Access Memory (RAM) *5*

1.3 MEMORY HIERARCHY *6*

1.4 SRAM *8*

1.5 SRAM WORKING *9*

1.6 TYPES OF SRAM *11*

1.7 FINFET SRAM: THE NEED AND ADVANTAGES *13*

1.8 DIFFICULTIES IN THE DESIGN OF MOSFET BASED SRAM *13*

1.9 FINFET *14*

1.10 SRAM DESIGN BASED ON FINFET *15*

1.11 PERFORMANCE MEASURES OF FINFET SRAM *16*

1.8.1. Static Noise Margin (SNM) *16*

1.8.2 Read Noise Margin (RNM) *17*

1.8.3 Write Noise Margin (WNM) *17*

1.8.4 Power And Delay *17*

2. LITERATURE SURVEY ***19***

3. STATEMENT OF PROBLEM BASED ON IDENTIFIED RESEARCH GAP ***24***

3.1 OVERVIEW *24*

3.2 STUDY GAPS *24*

3.3 OBJECTIVES 24

4. WORK DONE, RESULTS AND DISCUSSION 25

4.1 METHODOLOGY 25

4.2 SIMULATION PARAMETERS 26

4.3 DELAY AND POWER 26

4.4 STATIC NOISE MARGIN 35

4.5 EFFECT OF TEMPERATURE 44

5. CONCLUDING REMARKS AND FUTURE SCOPE ***54***

REFERENCES 56

List of Figures

Figure	*Figure Name*	*Page No.*
Figure 1.1	*Classification of primary memory based on data retention capabilities*	*2*
Figure 1.2	*Classification of memory based on usage*	*2*
Figure 1.3	*A typical DRAM cell. The transistor is used to access the capacitor by making the address line high '1'*	*5*
Figure 1.4	*Memory Hierarchy*	*7*
Figure 1.5	*An example of a typical 6T SRAM cell*	*8*
Figure 1.6	*Corresponding schematic diagram of read operation*	*9*
Figure 1.7	*Equivalent schematic diagram for write operation*	*10*
Figure 1.8	*4T SRAM cell*	*11*
Figure 1.9	*A 6T SRAM cell*	*12*
Figure 1.10.	*A TFT SRAM cell*	*12*
Figure 1.11	*A double gate FinFET*	*13*
Figure 1.12	*A: DELTA MOSFET, B: FinFET*	*15*
Figure 1.13	*6T FinFET SRAM cell*	*15*
Figure 1.14	*Static noise margin (SNM) for an SRAM cell. The square inside the "butterfly" curve shows the resilience of the cell against the DC noise. SNM is described as the length of the side of the square*	*16*
Figure 2.1	*Integration of fine grain power gating technique with a 6T SRAM cell using tied gate mode of FinFET*	*22*
Figure 2.2	*A pass gate feedback with fine grain power gating*	*22*
Figure 2.3	*Independent gate FinFET with power gating*	*23*
Figure 4.1	*Design of the 6T SRAM cell in S-edit*	*25*
Figure 4.2	*Design of SRAM cell with fine grain power gating.*	*26*
Figure 4.3	*Timing Diagram for 7nm FinFET*	*27*
Figure 4.4	*Timing Diagram for 10nm FinFET*	*27*

Figure 4.5	*Timing Diagram for 14nm FinFET*	*28*
Figure 4.6	*Timing Diagram for 16nm FinFET*	*28*
Figure 4.7	*Timing Diagram for 20nm FinFET*	*29*
Figure 4.8	*Comparison of delay times for different technology nodes*	*30*
Figure 4.9	*Timing diagram for 7nm finegrain*	*30*
Figure 4.10	*Timing diagram for 10nm finegrain*	*31*
Figure 4.11	*Timing diagram for 14nm finegrain*	*31*
Figure 4.12	*Timing diagram for 16nm finegrain*	*32*
Figure 4.13	*Timing diagram for 20nm finegrain*	*32*
Figure 4.14	*Comparisons of delay time and average power dissipation for fine grain power gating*	*33*
Figure 4.15	*Comparison between standard cell and gated SRAM cell for each technology node*	*34*
Figure 4.16	*Calculation Of The Snm For A Sram Cell*	*36*
Figure 4.17	*Determination Of The Snm By Rotating The Butterfly Curve*	*36*
Figure 4.18	*Butterfly Curve For 7nm Finfet SRAM Cell*	*37*
Figure 4.19	*Butterfly Curve For 10nm Finfet SRAM Cell*	*37*
Figure 4.20	*Butterfly Curve For 14nm Finfet SRAM Cell*	*38*
Figure 4.21	*Butterfly Curve For 16nm Finfet SRAM Cell*	*38*
Figure 4.22	*Butterfly Curve For 20nm Finfet SRAM Cell*	*39*
Figure 4.23	*Butterfly Curve For Fine Grain 7nm Finfet SRAM Cell*	*39*
Figure 4.24	*Butterfly Curve For Fine Grain 7nm Finfet SRAM Cell*	*40*
Figure 4.25	*Butterfly Curve For Fine Grain 7nm Finfet SRAM Cell*	*40*
Figure 4.26	*Butterfly Curve For Fine Grain 7nm Finfet SRAM Cell*	*41*
Figure 4.27	*Butterfly Curve For Fine Grain 7nm Finfet SRAM Cell*	*41*
Figure 4.28	*PDP trend with increasing technology nodes*	*42*
Figure 4.29	*SNM trend with increasing technology nodes*	*43*

Figure 4.30 *The delay curves under temperature variations for 7nm standard cell* *44*

Figure 4.31 *The delay curves under temperature variations for 10nm standard cell* *45*

Figure 4.32 *The delay curves under temperature variations for 14nm standard cell* *45*

Figure 4.33 *The delay curves under temperature variations for 16nm standard cell* *46*

Figure 4.34 *The delay curves under temperature variations for 20nm standard cell* *46*

Figure 4.35 *The delay curves under temperature variations for 7nm fine grain cell* *47*

Figure 4.36 *The delay curves under temperature variations for 10nm fine grain cell* *47*

Figure 4.37 *The delay curves under temperature variations for 14nm fine grain cell* *48*

Figure 4.38 *The delay curves under temperature variations for 16nm fine grain cell* *48*

Figure 4.39 *The delay curves under temperature variations for 20nm fine grain cell* *49*

Figure 4.40 *Average dynamic power comparison at 7nm node for different cell structure at varying temperature* *50*

Figure 4.41 *Average dynamic power comparison at 10nm node for different cell structure at varying temperature* *50*

Figure 4.42 *Average dynamic power comparison at 14nm node for different cell structure at varying temperature* *51*

Figure 4.43 *Average dynamic power comparison at 16nm node for different cell structure at varying temperature* *51*

Figure 4.44 *Average dynamic power comparison at 20nm node for different cell structure at varying temperature* *52*

Figure 4.45 *Effect of temperature on the SNM of SRAM cell for different technology nodes* *53*

List of Tables

Table	***Table Name***	***Page No.***
Table 4.1	*Simulation parameters for model files*	*26*
Table 4.2	*Times and average dynamic powers for different technology nodes*	*29*
Table 4.3	*Delay Times and average dynamic powers for fine grain cell at different technology nodes*	*33*
Table 4.4	*Comparison of average power and PDP for standard cell and fine grain cell at various technology nodes*	*35*
Table 4.5	*SNM and PDP data for different SRAM cell. The highest and second highest SNM, and the lowest and the second lowest PDP is highlighted for reference.*	*42*
Table 4.6	*Average change in delay time and dynamic power per degree Celsius increase in temperature*	*49*

List of Abbreviations

BL	Bit Line
CMOS	Complementary Metal Oxide Semiconductor
CR	Cell Ratio
DGFET	Dual Gate Field-Effect Transistor
DRAM	Dynamic Random Access Memory
FinFET	Fin Field-Effect Transistor
MOSFET	Metal Oxide Semiconductor Field-Effect Transistor
NMOS	n-type Metal Oxide Semiconductor
NVSRAM	Nonvolatile Static Random Access Memory
PMOS	p-type Metal Oxide Semiconductor
PTM	Predictive Technology Model
PDP	Power Delay Product
RAM	Random Access Memory
RNM	Read Noise Margin
SCE	Short Channel Effects
SNM	Static Noise Margin
SRAM	Static Random Access Memory
WL	Word Line
WNM	Write Noise Margin

CHAPTER 1
INTRODUCTION

1.1. OVERVIEW

The rising popularity for the on-the-go computing power and internet of things (IoT), in the modern technological era, has created a demand for low power electronics devices with greater memory capabilities to cater to various processing needs. The sizes of these memories are constantly shrinking in accordance to the Moore's law, however, this trend cannot continue past a certain threshold due to the physical limitations and drawbacks of traditional complementary metal oxide semiconductor (CMOS) devices. Before the CMOS logic, the widely used logic devices were transistor to transistor logic (TTL) and emitter coupled logic (ECL), however, they weren't particularly good for the low power applications because of their higher supply voltages and threshold voltages as compared to the CMOS. [1-5]

Initially the CMOS was a promising candidate towards the new trend of ultra-low power electronic devices. It remained like that for decades before reaching its limit. It turned out that with each scaling iteration of the CMOS manufacturing process, the parasitic passive elements began to surface. The reduction in the supply voltages and threshold voltage lead to the increasing subthreshold leakage. To tackle this issue various novel devices have been suggested. One such device is FinFET. [6-8]

FinFETs have been known to be successfully scaled to 7nm process by TSMC limited (Taiwan Semiconductor Manufacturing Company) and are being used in several consumer electronics such as ARM processors, GPUs, ASIC cards, etc. This has been possible owing to dramatic decrease in short channel effects and lower power dissipation of the FinFET technology. Since every digital system out there uses some sort of memory and memories occupy a large area on the die. Hence changes in the memory in terms of power dissipation or speed can affect the whole systems. For these reasons the memories have been an active field of study from the day of their perception. Today, a system can have two or more different types of memories. In a typical embedded or IoT system, we have a at least two different type of memory. The type of memory is generally decided by the embedded or IoT system and its envisioned application. To design an embedded or IoT system, the programmer has to keep in the memory footprint of the program as the memory comes at a premium. [9-11]

1.2. TYPES OF MEMORIES

Computer memories can be broadly categorised, based on their data retention capabilities, into volatile, memories that loose data when power supply to the memory is disconnected, and non-volatile, memories that retain data without an external power source. Non-volatile memory can be

classified as a secondary memory. Some example of secondary memories are hard disk drives, USB pen drives, solid state drives, etc. They are mostly used as a long-term data storage utility. Volatile memories in contrast are referred to as the primary memory or the main memory of a computing system. These memories are mainly used as RAM (Random Access Memory) and cache memory due to their higher speed of operation i.e. the read and write operations [12-13].

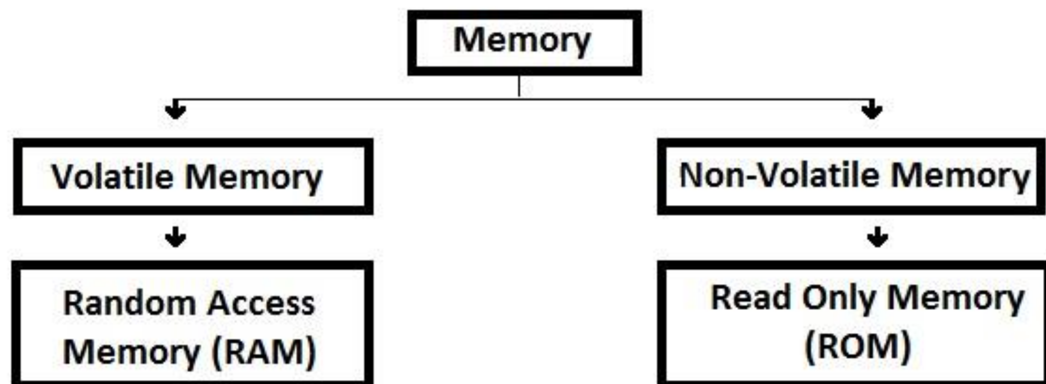


Figure 1.1 Classification of primary memory based on data retention capabilities [1]

Based on the usage patterns the memory is classified as shown in Figure. 1.2. Some of them are discussed in brief in the following section. The main scope of this thesis focuses mainly on the Random Access Memories (RAM) and more specifically on Static Random Access Memories (SRAM) which are discussed in detail in later sections. [3]

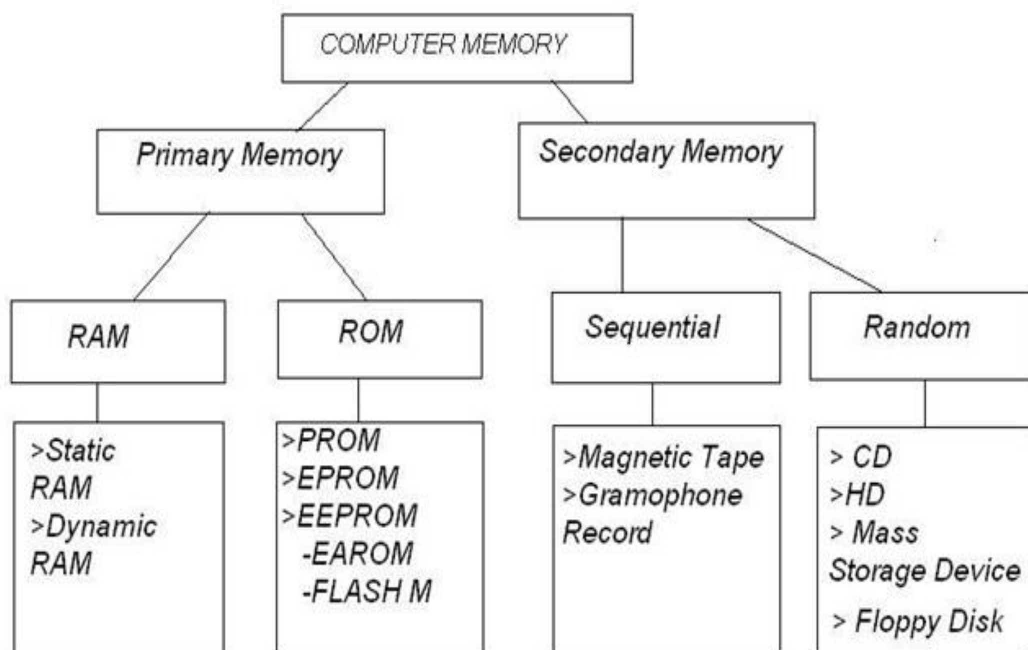


Figure 1.2 Classification of memory based on usage [1]

1.2.1. Secondary Memory

The secondary memory is used to the store the data for long term usage. They are usually cheaper to manufacture and have high storage density and are the slowest form of computer memory. These can be further classified based on the how the data the stored and accessed from the memory as follow. [13]

1.2.1.1. Sequential

Data can only be accessed or written in a sequential manner i.e. each block of data can be written or read to or from sequential memory locations.

- Magnetic Tapes: As the name suggests, they use a magnetic tape or reel. The magnetic material on the tapes can be magnetically manipulated by a tape head, a type of transducer and a motor to move the reel. They are also known as Tape Drives. The tape head would rest on top on the moving reel, picking up fluctuating magnetic fields and translating them into electrical signals. Since the tape requires a fluctuating magnetic field that makes data analogue in nature hence random access of the data is not possible as for fluctuating magnetic field requires a constant movement of the reel.

1.2.1.2. Random

These memories are mostly digital in nature and the data is stored in the form of blocks or sectors. Each block has its own physical address which can be used to retrieve or write data individually without disturbing any other data and thus the name, random memories. There are variety of random memories available today some of them are: [10-13]

- Floppy Disk Drives: The basic working principle is same that of the magnetic tape, however, instead of plastic reel coated with magnetic materials it uses a disk. The disk is divided into sectors and each sectors is further divided into blocks. The read/write head also moves in tandem with the disk unlike the magnetic tape where it was stationary. This is one of the slowest memory and has some of the lowest capacities (in several MBs) and hence has no practical use today's high speed computing environment. [10]

- Hard Disk Drives: The basic principle of working is same as that of a Floppy Disk Driver. It uses a magnetised metal disk instead of a plastic one which makes it durable and reliable that corresponds to higher RPM drives ranging from 5400 to 10000 rpm. The data capacity is also very high (in Terabytes). These are the most widely adopted type of secondary memory in use today.

- Compact Disks: These are light weight, low cost optical data storage devices. The data is stored on sectors on plastic disc with a reflective coating of aluminium or gold.

It uses a laser to read/write data to and from the disk. These are easy and economical to manufacture and can hold decent amounts of data (800MB – 25 GB) depending upon the type. Compact Discs are faster than a floppy disk but are way slower compared to a HDD and hence are mostly used to share data or temporary data backups. Also, these are prone to scratches and corrosion and hence are not very reliable. Although very popular during the 90s and early 2000s, their popularity has declined due to advent of cheaper and faster mass storage devices. [11]

- Mass Storage Devices: These very high speed digital data storage memories. They have the highest density of all the secondary type memories. As the name suggests they are used for large data storages. They are mostly based on NAND flash logic and are able to sustain the data transfer rates (mostly read is higher than write speeds) unlike other memories. The common examples of such memories are USB thumb drives, solid state drive, SD cards, etc. [13]

1.2.2. Primary Memory

Primary memory memories are the which are internally used by a computing system to perform high speed memory task. Such memories are usually faster than the secondary memories with transfer rates in several gigabits per second.

1.2.2.1. Read Only Memory

As the name suggests, it allows only data read and data cannot be written or further modified. However, certain ROMs allow for this while other strictly not at all. It is mostly non-volatile in nature that means it can retain data even after the supply is turned off. Although fast, manufacturing them in larger capacity reaches point of diminishing returns economically hence ROMs are primarily used for storing small program software such as computer BIOS or firmware in embedded systems. [13]

- PROM: Programmable ROM, is essentially a ROM which can be programmed just once. The data cannot be altered once programmed.

- EPROM: Erasable and Programmable ROM, as the name suggests, are programmable and can be re-programmed multiple times. However, the process of reprogramming is slow and complex. It requires an ultraviolet light to be flashed upon a small window on the EPROM chip to erase the data after which it is ready to be re-programmed.

- EEPROM: Electrically Erasable PROM, also known as E2PROM, works similarly as EPROM except that it doesn’t require methods like shining ultraviolet on it to erase it. It can be reprogrammed on-board using external or on-board programming circuits, which is way faster than taking off the chip from the board. For this reason, it has

gained a lot of popularity in embedded applications where rapid prototyping is a necessity.

1.2.2.2. Random Access Memory (RAM)

Random Access memory provides the ability for the data to be written or read from any random memory location. This means that the highest possible time (worst case scenario) is same for every bit of data regardless of its location on the memory. [14]

- DRAM: In consumer electronics industry DRAM is generally marketed simply as the RAM or the main memory of a device or a product. A typical DRAM memory cell consists of only a one transistor which translates to lower die area and higher memory density. This in turn makes DRAM inexpensive and widely used as the main memory for the computing devices. Also, from a marketing perspective, it is beneficial for the manufacturers to advertise the higher capacity of the DRAM (in Giga Bytes) rather than the lower capacity of SRAM (in Kilo Bytes or Mega Bytes depending upon the use case), albeit faster. However, DRAM require complex sensing and refresh circuitry which are costly and is topic of research. The data bits stored in DRAM cell using a capacitive structure. Hence data corruption due to charge leakage from the capacitors is a major concern. However, this problem is addressed by previously mentioned, complex refreshing circuit, which continuously refreshes the DRAM cells [14-18].

  The inherent problem of DRAM of charge leakage and continuous refreshing causes DRAM memories to consume a lot of power when idle and hence are not

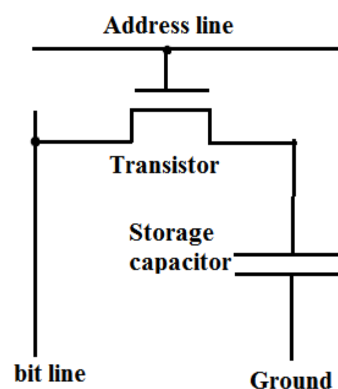


Figure 1.3 A typical DRAM cell. The transistor is used to access the capacitor by making the address line high ‘1’ [12].

suitable for embedded system applications that require to be operated in low power conditions. In addition, there exists a significant charging and discharging time associated with the capacitors that limits the speed of DRAM memory and hence are generally slower than SRAMs. SRAM eliminates the use of capacitors in favour of cross-coupled inverters or a latch circuitry for data storage, which eliminates the

disadvantages of the DRAM. DRAM also exhibits the property of data remanence i.e. data is retained by the DRAM cells even when it is powered down especially at lower temperatures. This property can be used for cold-boot hacking which is a security concern [12-14].

- SRAM: Static Random Access Memory, Unlike the DRAM, it uses a bi-stable multi-vibrator circuitry to store the data. The data is stored as long as it is powered on while DRAM needs to be refreshed after repeated intervals. SRAMs are inherently faster and are used in cache memories which needs to be in sync with the speed of the processor. SRAM is discussed in great detail in later sections.

1.3. MEMORY HEIRARCHY

Typically, a system uses not just one but several different type of memories. More often than not, an embedded system has to be both fast and economical. In such scenarios, choice and amount of a several type of memory is the key. As we have known by now that different memories work with different speeds and have varying cost and die area penalty. Typically, the cost of high speed memories is high and hence they are incorporated in a very small percentage for specific high priority task. The cheaper and slower memories are mostly used in abundant which are to be used for generic purposes. [14-18]

The memory hierarchy generally follows four key levels.

1. Internal memory – includes CPU Registers and cache memory.
2. Main memory – includes System RAM (DRAM)
3. On-system mass storage – constitutes the Secondary storage (HDD, SD cards, etc.)
4. Off-system mass storage – as called as Tertiary or removable storage.

To understand the flow of data we can think of an example where a set of instructions stored in a hard drive needs to be processed. The instructions in questions are stored in hard drive are already on the third stage of the hierarchy. From here the operating system of the system (say Windows) loads on the second stage that is the main memory. They stay there until the processor is ready to process the instructions. Once ready, few instructions from the set of instructions are moved on the cache memory. Depending upon the application and cost of system there can be several levels of cache memories. The instructions are further broken down to several moved to the processor registers from where they will be directly processed. The generated output will follow the same hierarchy back to the hard drive. [15]

One might argue that instead of following the hierarchy, data can be directly sent to the processor. The only problem here is that the processors are generally much faster than the secondary memories.

So, once the processor has finished executing one instructions it has to wait being idle until next instructions is fetched, effectively bottlenecking the system.

The counter argument to this would be to use only the cache memory. Theoretically, it would make the system faster in more ways than one but the fundamental problem with this is that the cache memories are volatile. Also, the cost and area penalty are too high to even build such a system. In an embedded system, miniaturisation is the keyword. So to keep the die sizes and cost minimum, smaller cache memories are built into the chips.

Figure 1.4 Memory Hierarchy

The cache memory which is the fastest memory of any system is basically an SRAM. There can be various levels of cache memory depending upon the use case of the system in question. In a modern household computer system there are three tiers of cache memory inside of a CPU, namely L1, L2 and L3 level caches. L1 being the fastest and L3 being the slowest yet significantly faster than the DRAM/main memory. The level 1 or L1 cache typically comes with a capacity of several kilobytes while the L2 and L3 are in the range of 2-16 megabytes depending upon the need of the system. Cache memories also helps to implement the concept of pipelining and it can be seen in large amounts of cache memories being used in multi-threaded applications in server systems and even in consumer grade hardware with multiple CPU cores. [14-18]

In small scale embedded applications like temperature sensing or switching applications, where too much information is not processsed, we only see basic implementation of cache memories limited to only L1 cache. This hugely brings down the cost and the complesxity of the embedded devices. In some advanced embedded applications that requires onboard data processing there can be upto two levels of cache memory. However, with the advent of the smartphone era and the increasing ddemand for the handheld processing power, we are beginning to see three levels of cacche memory in this segment of technology too. Hence the need to miniaturise the foot print and power consumption of SRAM is a major topic of research. It is interesting to notice that SRAM takes upto 90% of the die area in an SoC or CPU [12].

1.4. SRAM

SRAM is a type of semiconductor memory that incorporates the use of a flip-flop or a latch circuit to store data bits. A bi-stable multi-vibrator is a circuit, which is stable in either of the two states. Most flip-flops are bi-stable in nature i.e. they will either store a '1' or a '0' [15-18].

A typical SRAM cell counters the drawbacks of the DRAM. It does not use capacitors and hence there is no need of a complicated refresh circuit and lower power dissipation. A typical SRAM cell comprises of six transistors, of which two are access transistors and rest make up the latch circuitry. This, however, results in increased footprint of the memory cell, reducing the memory density and making it costly. Hence, manufacturers use SRAM conservatively across different computing devices depending upon the application. In most modern day computer, SRAM is fabricated alongside CPU on a single IC. Mainly used as a bridge between low speed main memory (DRAM) and faster CPU clocks [14].

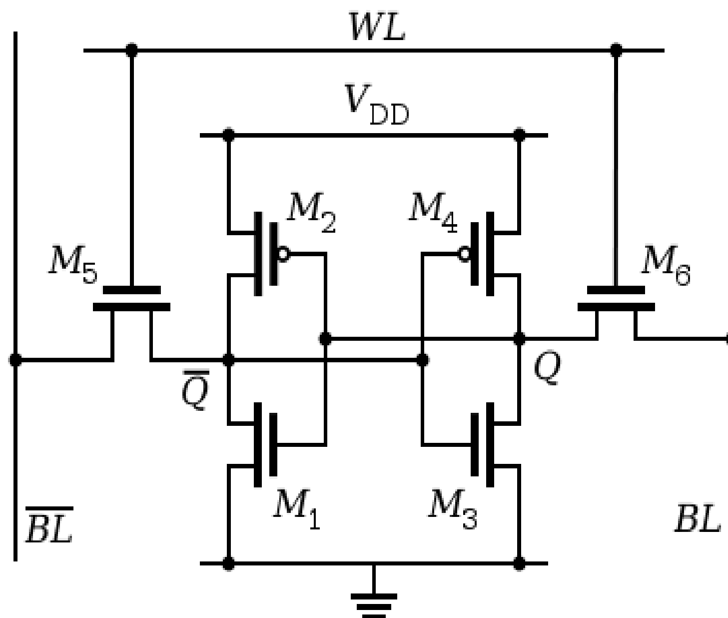


Figure 1.5. An example of a typical 6T SRAM cell [18]

Advantages of SRAM over DRAM:

1) Lower power consumption
2) No refresh circuitry is required
3) Reliable (data corruption while reading)

Even though SRAM counter all the drawbacks of the DRAM, but it also exhibits data remenance, which is a major security concern. There are certain software techniques developed to counter this problem, which flushes data on the RAM and is still a topic of research [13-19].

1.5. SRAM WORKING

Typically, a 6T SRAM cell has two control lines, one BIT line and one WORD line denoted as (BL) and (WL) respectively as shown in Figure 1.2. The transistors M5 and M6 are called the access transistors for the SRAM cell and transistors M1, M2, M3 and M4 form the bi-stable multi-vibrator or cross-coupled inverters, which stores a single bit of data. An SRAM cell operates in three modes

1) Stand-By mode (the circuit is idle)
2) Read Mode
3) Write Mode

In the stand-by mode the word line is set to '0' (WL ='0'), this turns off the transistors M6 and M5 which in turn cause the memory cell to be disconnected from the bit lines. The transistors M1 through M4, together, form two inverters connected in a feedback loop (cross-coupled inverters) as long as they are attached to the supply V_{DD}, the data stored stays in the latch until the supply is turned on. [14-20]

In Read mode, the word line is set to '1' (WL = '1'), turning on the transistors M6 and M5 connecting (access transistors) the SRAM cell to the bit lines BL and BLbar. The stored values at the nodes, namely, Q and Qbar gets transmitted to the bit lines, that is BL and BLbar. When the read operation is performed with Q = Vdd, the SRAM cell effectively corresponds to the schematic diagram shown in the figure.

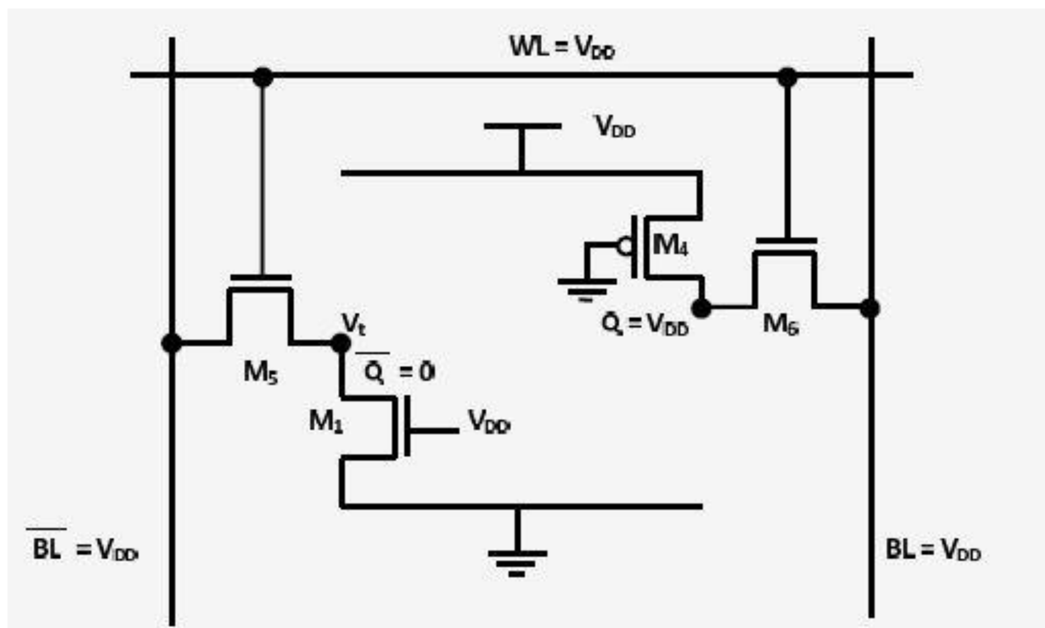


Figure 1.6 Corresponding schematic diagram of read operation [7]

When the node Q has reached the threshold voltage of the transistor, M3 (refer to the complete SRAM schematic), node Q will experience a drop in the node voltage level but renewing nature of the latch (cross-coupled inverters) will try to force the bit in the cell to be flipped. This will corrupt

the stored data bit. To prevent this from happening, the transistor M1 should be made strong enough so as to avoid the threshold level of the transistor M3 is never achieved by the node voltage at Q.

During Write mode, for writing ‘0’, the bit line is first lowered to ‘0’ (BL = ‘0’) and bit bar line is raised to ‘1’ or V_{DD} (BLbar= ‘1’). The memory cell to be written is picked by pulling-up the world line to V_{DD} or asserting logic ‘1’ to it. This will write logic ‘0’ to the cell. For writing logic ‘1’ the process is repeated but with values of it lines reversed i.e. BLbar is lowered to logic ‘0’ and BLbar is pulled up to logic ‘1’. The cell is selected by asserting the word line and logic ‘1’ is stored in the latch. [12]

This works because the access NMOS transistor M5 and M6 are generally made stronger than the inner transistors of the latch to easily override the previous state of the latch.

The schematic diagram shown in Fig. 1.7 is the equivalent circuit for a write ‘0’cycle, where voltage at the node is initially is set to Vdd. When the voltage at node Q starts to go down, at a certain voltage level, the PMOS, M2 (refer to the complete schematic for the SRAM cell) gets turned ON. This will lead to rise in the node voltage at Qbar and the cross-coupling action of the inverters will force the node Q to zero voltage level, that is the write ‘0’ action is completed.

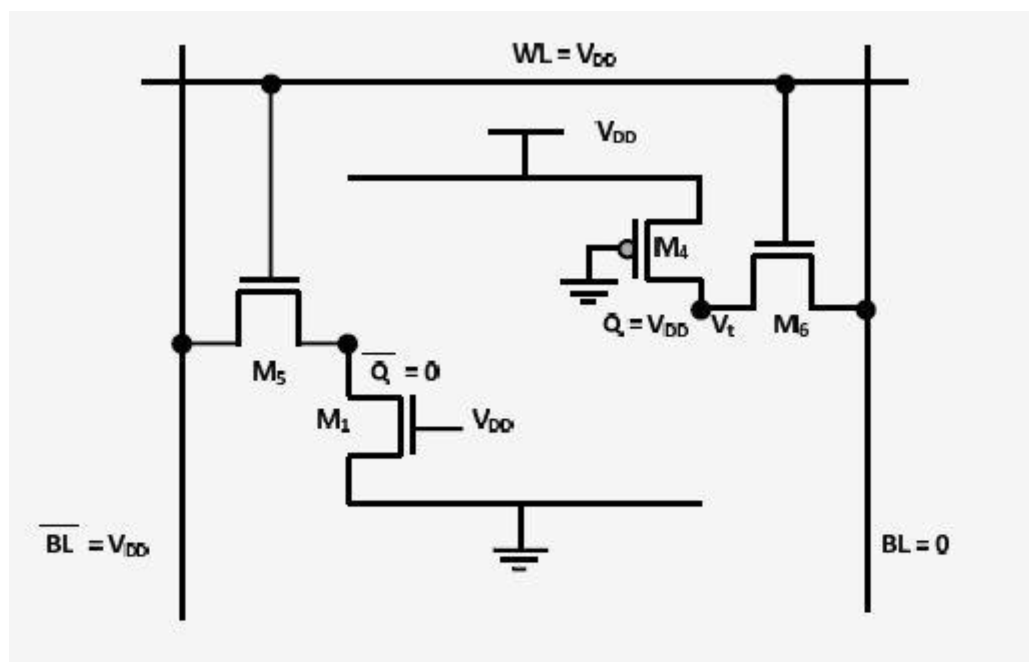


Figure 1.7 Equivalent schematic diagram for write operation [7]

To ensure that the above series of event happen successfully, preventing a write failure, the node voltage at Q should never go below the threshold level of the NMOS M4, which can be achieved by making transistor M6 is strong.

1.6. TYPES OF SRAM

Based on data retention, SRAM can be:

1) Volatile SRAM, a typical SRAM cell that losses data once the power is turned-down.
2) Non-Volatile SRAM or scram performs all the tasks of a typical SRAM with added functionality of retaining data when power is switched off. It is used in special applications where holding the data is vital, for example, data centres, aerospace, etc. [12-19].

Based on functionality:

1) Asynchronous SRAM, it is unconstrained of clock frequency and has additional control signals, excluding bit lines and word lines, to handle the memory. They are, Chip Select, Output enable and Write Enable denoted by CS, OE and WE respectively. CS selects or deselects the SRAM chip, when deselected the chip remains in the stand-by mode. OE operates the output and WE decides on the read and write cycles.
2) Synchronous SRAM, have their read-write cycles orchestrated with CPU cycles for high-speed purposes by using clocks as opposed to the control signals found in asynchronous SRAM.

Based on memory cell:

1) 4T (four transistor) SRAM: It has two NMOS pass (access) transistors and other two NMOS transistors connected to polysilicon resistor to form inverters. The size of this is lesser than the established 6T transistor but is still four measures larger than a conventional DRAM cell.

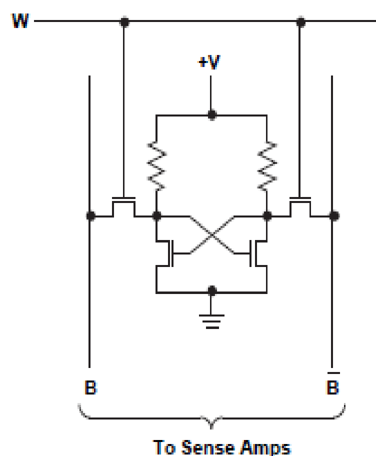


Figure 1.8. 4T SRAM cell [12]

2) 6T (six transistor) SRAM: Even though the 4T cell is smaller, it has several confines including high stand-by current owing to resistors, cell is susceptible to noise and soft errors since resistance is high and cell is not as fast as a 6T cell. A 6T cell discards all these limits by using PMOS in preference to a polysilicon resistor at the load end of the cell. That makes it a CMOS flip-flop. This improves speed, noise insusceptibility and lowers stand-by current from a 4T cell.

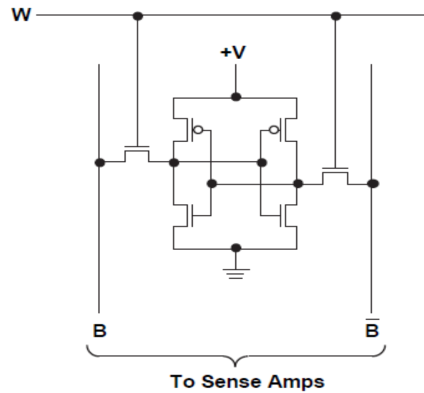


Figure 1.9. A 6T SRAM cell [12]

3) TFT (Thin Film Transistor) cell SRAM: To improve the 4T cell structure and condense the current flow from the resistor, TFT was developed. It is a resistor constituted as a PMOS transistor. The electrical properties alter by controlling the channel of the transistor. The performance is, however, is not analogous to a 6T cell but is better than a 4T cell [12-20].

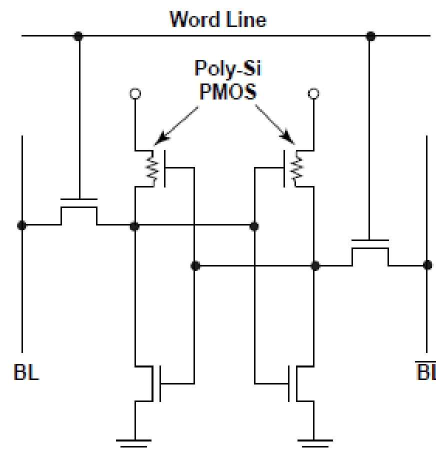


Figure 1.10. A TFT SRAM cell [12]

1.7. FINFET SRAM: THE NEED AND ADVANTAGES

The major challenge encountered by the semiconductor industry today is curtailing the footprint of the SRAM without negotiating on the performance to fabricate better and faster ICs since memory populates approximately 94% chip area [19-25]. Diminution in size can occur in two ways. One is cell contraction or device modelling and the other is interconnect scaling. Device scaling at nanoscale roots problems that are susceptible to the process variation such as variations of doping concentrations. The structures like DG-SOI (Dual Gate Semiconductor on Insulator) or FinFET can successfully replace the bulk transistors and are scalable without short channel effects (SCE). [14-15]

The leakage current in FinFET is stereotypically less than that of MOSFET and has loftier scalability for a specified gate insulator thickness [16-30].

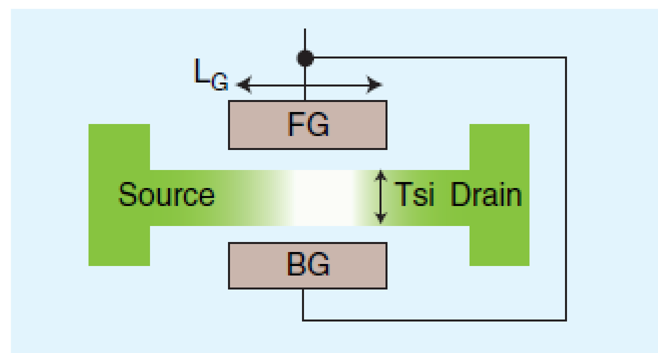


Figure 1.11. A double gate FinFET [18]

FinFet uses an intrinsic body that subdues variability in the performance of the device which are caused dopant ions concentration. Whereas, in planar bulk MOSFET, there is a stark process variability owing to the severely doped channel.

1.8. DIFFICULTIES IN THE DESIGN OF MOSFET BASED SRAM

The short channel effects (SCE) that occur in MOSFET structures ascending into the sub-50 nm system necessitate heavy doping [18]. Short Channel Effects in a MOSFET are associated with the channel width and source/drain depletion layer width. When they are of the same order, the SCEs come into play. This causes MOSFET to behave inversely. There are five distinguishable Short Channel Effects, they are:

1) Punch through and Drain induced barrier lowering (DIBL)
2) surface scattering
3) velocity saturation
4) impact ionization
5) hot electron effect

The heavy doping roots the mobility of the charge carriers to worsen due to impurity scattering. Worsens the sub-threshold swing owing to transverse electric field and upsurges junction capacitance.

The inconsistency in Threshold voltage (V_{th} triggered by random dopant variabilities is another troubling factor for bulk MOSFET at nanoscale regime. It becomes problematic to warranty the SRAM stability in large arrays with these deviations, which is a shortcoming for low powered embedded applications. However, this can be resolved by just increasing the transistor size, which is undesirable for the essential reason of scaling down the device in the first place in order to increase the density.

The huge standby power in SRAM arrays based on bulk MOSFET is the direct result of exponential increase in leakage current. Additionally, there can be parametric malfunction caused by the inconsistencies in the strength of different devices owing to process variations. These variations in device strengths and huge leakage currents makes is excruciating challenge to design a low power SRAM. Moreover, scaling down the single gate bulk CMOS devices further than sub-50nm nodes is challenging owing of heightened short channel effect (SCE). Because of the greater invulnerability to SCE, researches have established that the ultra-thin body double-gate MOSFET (DGFET) devices are apt for use in sub-50nm technologies, they also have improved scalability and improved on current paralleled to single gate devices. Moreover, DGFETs have insignificant junction capacitance, which interprets to lessened circuit delays. Also, they are lightly doped which eradicates threshold voltage (V_{th}) variations due to random dopant fluctuation.

1.9. FINFET

After the publication of the famous paper in 1965 by Gordon Moore on the evolution of density of transistors on an integrated chip, which prophesied the quadrupling of transistors per chip every three or so years. The semiconductor industry was able to stay on it quite remarkably for forty years. In 1990's, the International Technology Roadmap for Semiconductors (ITRS) organization was established by the joint endeavour of semiconductor companies and academic world to better forecast the future of the industry. Manufactures, however, efficaciously following the road map started to happenstance problems as they scaled down the manufacturing process for shorter channel MOSFETs. This set in motion for the efforts in developing devices which can surmount these challenges, mainly the short channel effects (SCE). The one elucidation to this was to use double-gate MOS (DGMOS) transistors, which attained a significant reduction in SCEs. This being the first step on the way to the development of FinFETs. [26-31]

The FinFET structure is more or less comparable to a DELTA transistor (fully Depleted Lean-channel TrAnsistor). The only modification is that the FinFET structure has a dielectric layer called the "hard mask" deposited between the gate and the silicon fin as shown in the figure.

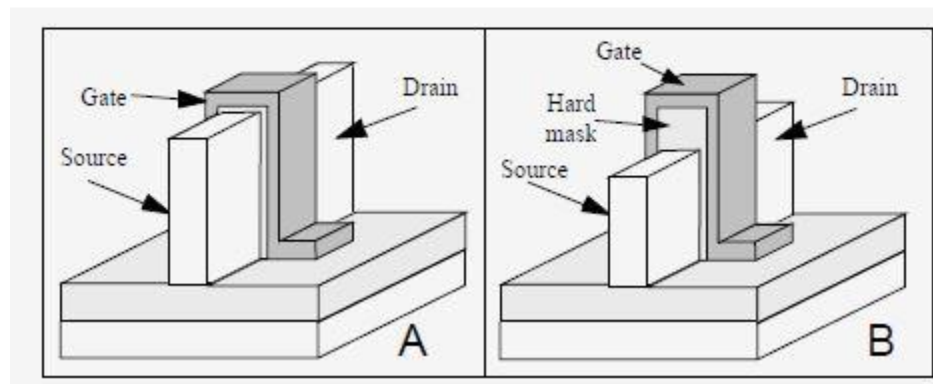


Figure 1.12 A: DELTA MOSFET, B: FinFET [19]

1.10. SRAM DESIGN BASED ON FINFET

FinFET has been accepted as the seemly contender for DGFET structure as discussed in the previous section. (Figure. 1.6) [20]. Like most MOSFET devices, appropriate optimisation of FinFET devices is obligatory to lower the leakage current and upsurge the stability. For example, the leakage current in FinFET SRAMs can be reduced by optimising the supply voltage (V_D), Fin height (H_{fin}) and threshold voltage (V_{th}). This can be achieved by enlarging Fin-height, which also allows for decrease in V_D. [21]. But dropping V_D can cause stability problems and hence both parameter must be judiciously optimised. Hence, there is a compromise between standby leakage current and stability.

FinFET SRAMs are used to realise memories applications that have a need of rapid retrieval times, low power consumption and forbearance to environmental elements. Additionally, they have lowermost static power dissipation and are well-matched with prevailing logic process, which makes them quite accepted. Associated to CMOS based SRAMs, FinFET based SRAMs have higher noise margins and switching speeds [22-32].

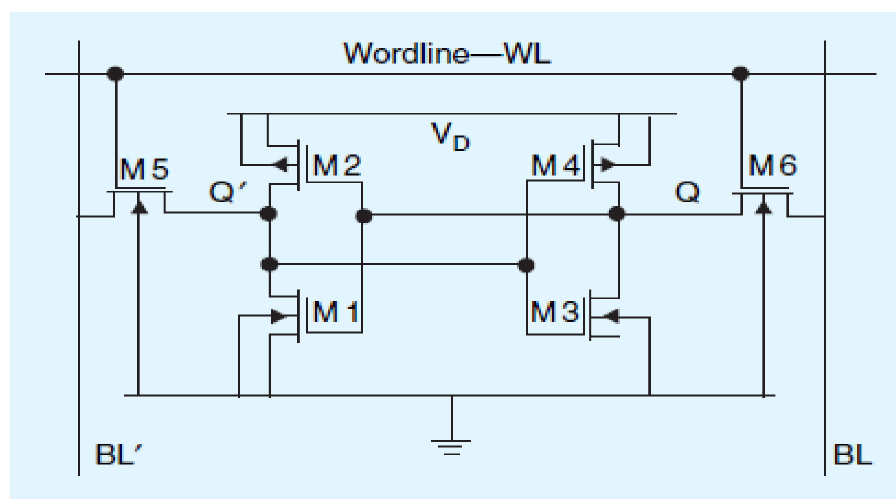


Figure 1.13. 6T FinFET SRAM cell [18]

The circuit design for a simple 6T SRAM cell is put forward in the Figure. 1.13. For high density memories, the cell size must essentially be small. However, the precise read process of the FinFET based SRAM cell is reliant on meticulous sizing of access transistor M5 and the pull-down transistor M1. The accurate write operation is reliant on the cautious sizing of access transistor M6 and pull-up transistor M4 as shown in the Figure 1.13. As explicated in [21], the most critical operation in terms of complexity is the read operation from the cell. If the access transistor, M5, is miniaturised in size, then the pull-down transistor, M1, has to be fashioned big enough so that the inverter, formed by the transistors M3-M4, doesn't accidently flip its output when the voltage rises on the Qbar node which inadvertently changes the bit inside the cell to '1'.

As explicated in [21], after the careful selection of transistor sizes for the inverters formed by the transistors M1-M2 and M3-M4 the sizing of the access transistors M5 and M6 becomes precarious for correct operation. The threshold at which the rationed inverter (M5-M6)-M2 switches must be kept below the threshold at which the M3-M4 inverter switches so that the flip-flop formed by the inverters can switch states from Q = '0' to Q = '1'. It has been established that the performance, noise margins, and power are affected significantly by the sizing of the transistors [33-37]. Consequently, to optimise the trade-off between power consumption, performance and reliability, sizes for n-channel and p-channel FinFETs must be selected carefully.

1.11. PERFORMANCE MEASURES OF FINFET SRAM

1.11.1. Static Noise Margin (SNM)

It is defined as the resistance of a cell to flip through a read operation. SNM is a standard quantity of the stability of a SRAM cell.

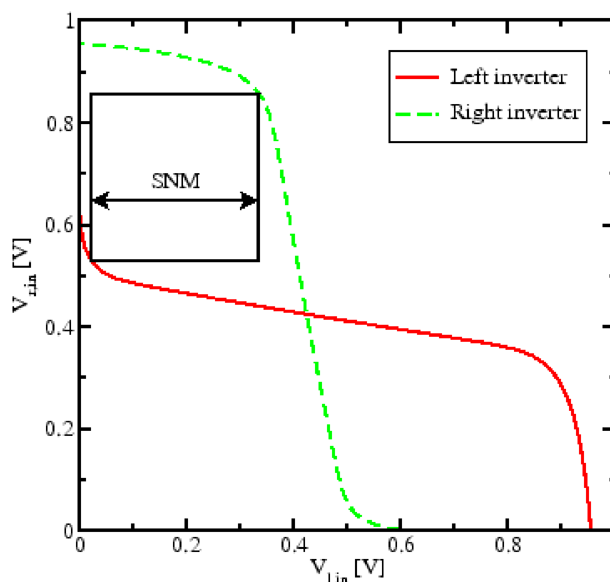


Figure 1.14 Static noise margin (SNM) for an SRAM cell. The square inside the "butterfly" curve shows the resilience of the cell against the DC noise. SNM is described as the length of the side of the square. [12].

SNM is computed by the side of the biggest square inside the butterfly inverter characteristics of a SRAM cell during a read cycle. Static Noise Margin rests on the threshold voltage of the FinFET used in the SRAM cell. A greater threshold voltage (Vth) will lead to a lower drive current which will make the writing operation tough and as a consequence, increasing the SNM. Henceforth, a higher Vth will result in low power cells with improved stability but at the price of performance. FinFETs deliver higher drive currents than MOSFETs can at greater threshold voltages allowing them to achieve higher noise margins alongside better write stability [24].

1.11.2. Read Noise Margin (RNM)

It is a quantity of sturdiness against reversing of stored data bits in the cell through a read operation [25-38]. A high RNM parallels to a high read stability. To achieve this the pull down transistor, M1, has to be made stronger than the access transistor, M5. RNM can also be improved by upscaling the pull-down transistor, M1, i.e. n-channel FinFET which inadvertently leads to lesser density. Alternatively, RNM can be improved by increasing the gate length of the access n-channel FinFET. However, this results in increased delay in the word line, WL, impacting the Write Noise Margins. This in turn harms the write performance of the SRAM cell. Hence a careful trade-off has to be made to optimise the performance of the SRAM.

The cell ratio (CR) of a SRAM cell governs the rise in the node voltage that stores the logic '0' during a read operation [38-44]. The Cell Ratio (CR) = (W1/L1)/ (W5/L5) is as shown in Figure 1.13. Cell ratios can determine the voltage drop across the pull-down transistor, M1. A smaller CR results in a higher voltage drop resulting in a smaller noise margin at node Q that is able to flip the cell.

1.11.3. Write Noise Margin (WNM)

It is the performance metric of the maximum bit line voltage (BL) that is sufficient enough to cause a change of state, that is flip the stored bit inside the FinFET SRAM cell when the bit line is pulled-up to logic '1'. A greater Write Noise Margin (WNM) equates to a higher stability of the cell [24]. For a faster write operation for bit '0'), a stronger access transistor (nFinFET) and a weaker pull up transistor (pFinFET enables the node storing logic '1' to discharge faster, thereby decreasing the time to write logic '0'. Alternatively, WNM is measured as the highest BLbar voltage that can flip the status of the cell when BL is pulled-up high. This leads us to the conclusions that WNM improves with a weaker pull-up transistor and a stronger access transistor which leads to penalty in overall cell area and a reduced read noise margin.

1.11.4. Power and Delay

Power consumption of a SRAM determines the application of the device it will be used in. The major advantage of FinFET SRAM over the CMOS SRAM is the lower power dissipation with a higher drive current owing of lower leakage currents and SCE's in FinFET [45-46].

Propagation delay is dependent upon the SRAM array column height and delays caused by the interconnect wires. The concept of subdivision is applied to the SRAM array to lower the delay of the SRAM at the cost of increased power dissipation. Since the Power-Delay-Product of a device is a constant quantity, henceforth decreasing one increases the other and the other way around. Delay can be decreased by upsizing the FinFET but also increasing the power dissipation in the process. Nevertheless, reduction in the power dissipation requires longer channel length which will resulting in increased propagation delay [47-51].

CHAPTER 2

LITERATURE SURVEY/REVIEW

An extensive amount of literature has been studied thoroughly during the course of this thesis. The following sections provides with a brief account of the studied material which lead to the discovery of the gaps and shortcomings.

Bin Yu, (2001) et al. [14] in their paper discuss about the challenges in the deep-sub-100nm scaling of the conventional planar CMOS transistor. In this paper, they have reported results on a bulk-silicon planar transistor with scaled down physical length to 15 nm. Significant reductions in the energy-delay product and gate delay were achieved experimentally whilst scaling down the transistor aggressively. Gate delays of up to 0.29ps and up to 0.68ps for n-channel FET and p-channel FET, respectively, were exhibited by the purposed 15 nm CMOS devices at a supply voltage of 0.8V. They have achieved a record achievement in terms of gate-delay and energy-delay product.

Digh Hisamoto, (2000) et al. [15] put forward "a novel self-aligned double gate Silicon On Insulator structure (FinFET) as a Nano CMOS device". They have highlighted the fact that scaling down the conventional bulk MOSFETs down to a sub 50 nm regime has a severe drivability and leakage degradation. The proposed FinFET device can be fabricated with the already existing planar MOSFET process technologies due to the quasi-planar nature of the double-gate MOSFETs. They achieved the desired threshold voltage by fabricating the gate material from boron-doped $Si_{0.4}Ge_{0.6}$ for the ultra-thin body device. Their results established that the sub-threshold leakage was well contained despite the low channel impurity concentration. No kink effects were observed in the saturation region because of the floating body effect. A small saturation current was observed since after the dry etching; the sacrificial gate oxidation was not used. The DIBL (Drain Induced Bias Lowering) effect was also studied. The experimental results showed a stronger DIBL effect. Hence, to sufficiently small to suppress DIBL, a smaller Si-fin width than the gate length is sufficient. Their paper demonstrated following key features.

1) Even with 17nm gate length, the self-aligned double gate commendably subdues short channel effects.
2) Threshold voltage can be achieved accurately by using $Si_{0.4}Ge_{0.6}$ gate for ultrathin body MOSFET.
3) The Gate is "self-aligned" to the Source/Drain, which is also raised up to moderate parasitic resistance.

Azeez J. Bhavnagarvala, (2001) et al. [26] have discussed the influence of the intrinsic device variations on the CMOS SRAM cell steadiness. They studied the effects of the intrinsic threshold voltage variations on reductions in the static noise margin (SNM) in homogeneously doped

MOSFETs. Compacted physical and stochastic models were used to achieve the purposed study. They found that for sub-100nm CMOS process, six sigma deviations in SNM due to intrinsic oscillations were enough to exceed the nominal SNM. These big departures cause serious hindrances to the scaling of the channel length, transistor density and supply voltage scaling for the standard 6T CMOS SRAM cells. Their study concluded with new stochastic and physical models for 6T CMOS SRAM cell SNM, that are able to correctly assess the influence of stochastic variations in device threshold voltage, due to random dopant atoms placements, on the cell stability.

Byung-Do Yang, (2005) et al. [27] put forth a design for a low powered SRAM using local sense amplifiers and a hierarchical bit line (HBLSA-SRAM). Their proposed SRAM model has both reduced write swing voltages and capacitance of the bit lines. They were able to achieve lower write power consumption without affecting of degrading the noise margin. This was achieved by applying a low swing and full swing signal to high capacitive and low capacitive bit lines, respectively. By lowering the bit line swing voltages to the tenth of V_{DD}, they were able to achieve savings of 34% in the write power command compared to the conventional SRAM. An 8-K 32 bits SRAM chip was fabricated using a 0.25um CMOS process clocked at 200Mhz at 2.5V which consumes 26 mW read power and 28 mW write power.

Huifang Qin, (2004) et al. [28] focuses on supressing the leakage current, a critical aspect, in the low power memory designed. They proposed lowering the standby supply voltage (VDD) to limit the leakage current, which is also the Data Retention Voltage (DRV). They explored effects of DRV on a standard low leakage SRAM module and effects on DRV itself by process variations, transistor sizing and chip temperature. They developed an investigative model for the DRV with process and design parameters as varying parameters. The model is verified using simulations and experimental data collected from a 4KB SRAM chip fabricated at 0.13µm node. Under this low standby VDD, they were able to achieve over 90% savings in leakage power savings by using dual rail-standby scheme. The research also concluded that the DRV is strongly affected by SRAM cell sizing and process variations.

Jung Hwan Choi, (2006) et al. [29] proposed a methodology to "self-consistently solve leakage power with temperature to predict thermal runaway". They point that dynamic power and static power, together, make up the power dissipation of a logic. They also point that the dynamic power consumption is weakly coupled with temperature variation than the static power consumption. They targeted circuits based on 28 nm FinFETs, as they are more susceptible to thermal runaway than conventional bulk-MOSFETs. In this paper, they employed thermal models at gate level to produce thermal maps of a circuit block, analysed the static temperature rise and activity reliant dynamic power. The simulation results showed that the maximum affordable subthreshold leakage is dependent on the worst-case input activity of the circuit block. Which suggests that the thermal runaway depends on the circuit level parameters such as input activity and transistor level parameters such as sub-threshold level leakage current. This paper also unravels the fact that

FinFETs suffer more from self-heating and less efficient heat dissipation compared to bulk-MOSFETs. They also exhibited that for a minimal activity of 0.5 and a typical package thermal resistance, the thermal runaway can occur even at the ITRS postulated sub-threshold leakage (150nA/um, high-performance)

T. Miwa, (2006) et al. [30] described two new circuit techniques for “non-volatile SRAMs with back-up ferroelectric capacitors” (NVSRAMs). They were able to overcome the problems faced the original NVSRAM, size and low-voltage reliability, with these circuits. A new 0.25-pm-design-rule was purposed for the four-metal-layer NV-SRAM cell with an area occupancy of 9.7 pm^2 which is comparable to area of a 0.25-pm three-metal layer SRAM cell. Low voltage operation is carried out using a high-voltage negative-voltage plate line driver which improves NV-SRAM array’s non-volatile retention characteristics.

KYOO ITOH, (1995) et al. [31] dissert about the trends in the circuit technologies for the low-powered CMOS RAMs. The trends are reviewed in terms of three main key parameters which include DC current, charging capacitance and operating voltage. They provided a general overview about the RAM chip power supplies. The overview includes both types of RAMs that is SRAMs and DRAMs. They suggested that the progressive advancements in DRAM circuits has resulted in reduced power consumption that hasn't been achieved in the last decade and is twice or thrice in the order of magnitude for a fixed capacity memory chip. Combined with the low power perks of complementary MOS technology, the following two technologies have become the main reasons for the reduction in power. Firstly, the partial activation of multi divided arrays using the data and word lines multi-divisions that results in lower charging capacitance. Secondly, the reduction in external power supply like down converting scheme for on-chip voltage and half-V/sub DD pre-charging causing lower operating voltages. Partially activating the word line, which is multi-divisional drastically, reduces the data-line dc current. An auto-power down scheme further reduces the dc current which uses the address transition detection for word driver and column circuitry along with the improvements in sense amplifier circuitry. They also discussed how the subthreshold current reduction will become crucial in the future.

Abhisek Dixit, (2005) et al. [32] insinuated that the multi-gate field-effect transistor (FET) is an encouraging device architecture for the CMOS at 45nm technology node. They discussed that these devices because of their slender drain/source profile, go through a high parasitic resistance. A geometry based analytical model for source and drain was used to analyse the parasitic source-drain resistance, which was validated using experimental and 3-D device simulations results. The model is capable of predicting the limits to the scaling of parasitic Source-Drain resistance, the model also revealed that behaviour of parasitic source-drain resistance is a function of the contact resistance between the Source-Drain silicide and Si-fin. It was observed that the discriminatory epitaxial growth of Si on Source/Drain regions only may be inadequate to meet the semiconductor roadmap target for parasitic Source/Drain resistance at the 45-nm CMOS technology node.

Deeksha Anandani. (2015**)** et al. [33] proposed a model of integrating the fine grain and coarse grain power gating techniques with an SRAM cell and arrays of SRAM cells, respectively. The paper discusses the effects on gating on both MOSFET based SRAM and FinFET based SRAM. Later, the power gating is applied to the standard SRAM cell using the different operating modes of FinFET namely, tied gate and independent gate. The different configurations are tested for leakage power, static noise margins and delay. The simulations are performed using the PTM models at 32nm node.

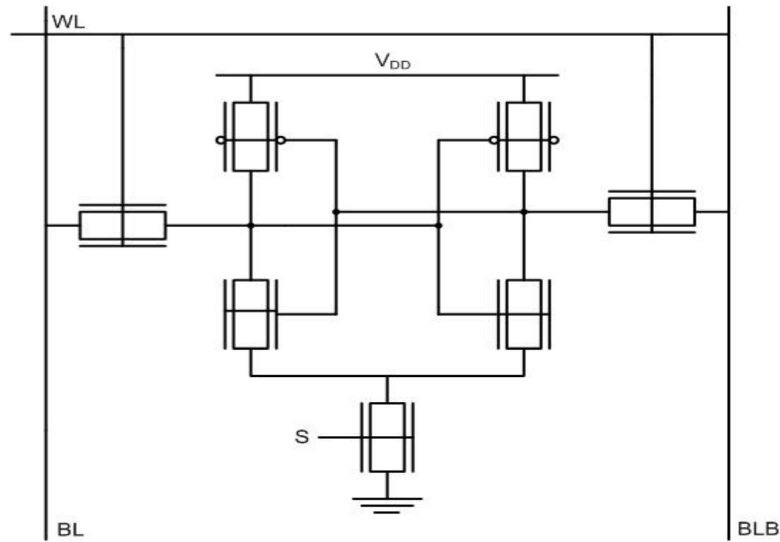


Figure 2.1 Integration of fine grain power gating technique with a 6T SRAM cell using tied gate mode of FinFET [33]

The same configuration is used as a reference for the purpose of this thesis at a different technology node, discussed later in detail. The advantage of applying power gating were verifiable by the simulations carried out the authors.

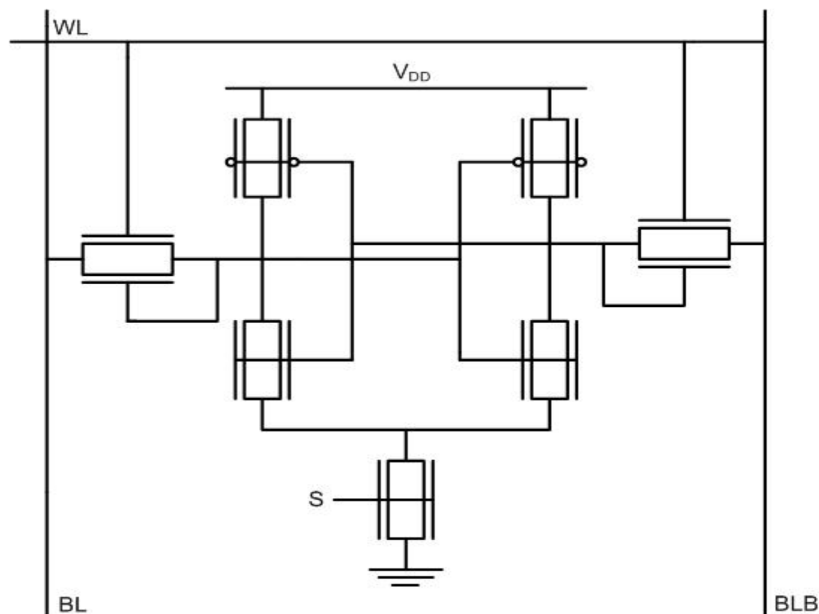


Figure 2.2 a pass gate feedback with fine grain power gating [33]

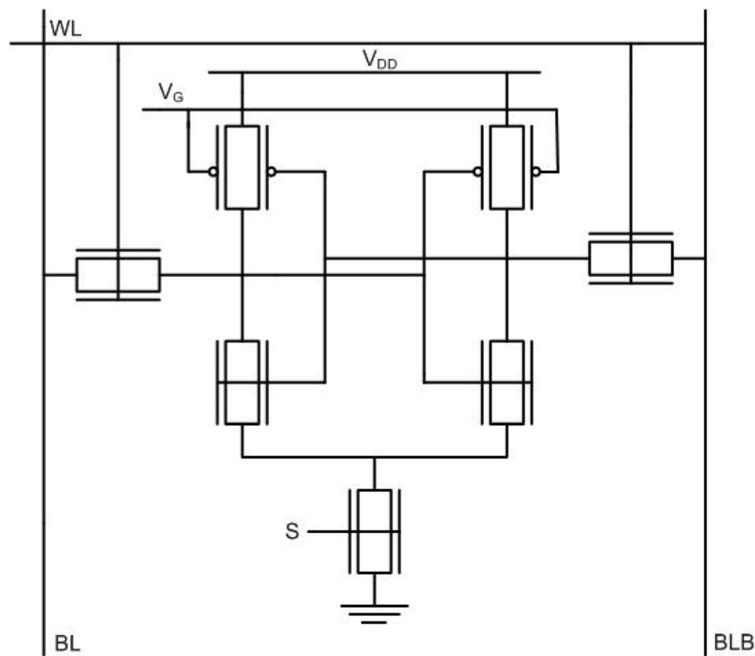


Figure 2.3 Independent gate FinFET with power gating [33]

By employing the purposed schematics, they were able to enhance the power dissipation and stability characteristics of the SRAM array. The implemented pass-gate feedback was successfully able avoid the read-write conflict that occurs in an SRAM cell.

Power gating is implemented to minimise the power consumption by cutting off the part of the circuit which are not in active use. This is cuts-off the current to the that part of the circuit in-turn reducing the leakage power. However, it comes at the cost of increased delay in the circuit. Hence, there's always a trade-off between performance and power savings.

One such method of power gating is Fine-Grain power gating. It involves adding a transistor to every block or part of the circuit that requires gating. In case of a 6T SRAM cell the cross-coupled inverters form one unit which are responsible for the major current leakage while in stand-by mode. The additional transistor is added between the source of the two nFINFET transistors and the ground. The gating transistor is activated each time for a read or write cycle. As it can be observed that for each memory cell an additional transistor is added to the cell. This incurs a huge area penalty in already die area sensitive application. Regardless of this the advantages in the form appreciable power savings cannot be neglected.

CHAPTER 3
STATEMENT OF PROBLEM BASED ON IDENTIFIED RESEARCH GAPS

3.1. OVERVIEW

SRAM is the most widely used and accepted form of embedded memory found in today's integrated circuits. As a fact, SRAM cells dominates the area on a die with a staggering 85%-90% of the transistors on ICs constitutes only the SRAM. Since this memory has been typically fabricated from conventional CMOS devices, all of the inherent issues accompanying MOSFET scaling are applicable to SRAM scaling. Issues such as gate oxide leakage, dopant fluctuations, contact resistance, control of short channel effects, abrupt and low spreading resistance junction technology must be fixed for sustained scaling of the traditional SRAM bit cell.

3.2. STUDY GAPS

Even with the advancements and the optimisation in scaling of the FinFET based SRAMs devices, the area where extensive work is needed to evaluate the effect of temperature on its performance for the better understanding of the overall memory design performance. As described in [7, 17], FinFETs are prone to thermal runaway and self-heating. It is also evident that SNM and RNM are negative functions of temperature while WNM increases with increasing temperature. Hence, temperature effects have to be taken into account while modelling memory circuits to better appreciate the actual performance of memory chip design under thermal variable environment conditions.

By making careful trade-offs on circuit level parameters and transistor level parameters, thermal runaway can be controlled. Another way to lower the effects of temperature is by refining the IC package quality i.e. adding a uniform heat spreader and efficient heat sink.

3.3. OBJECTIVES

The following objectives are proposed on the basis of the research gaps identified in the previous section.

1) To evaluate the performance of the FinFET based SRAM in terms of delay, power and PDP for different technology nodes.
2) To evaluate the performance of FinFET based SRAM in terms of SNM for different technology nodes.
3) To compare the performance of the FinFET SRAM with power gating and without power gating.
4) To compare the performance of the FinFET SRAM under variable temperature environment with performance of FinFET SRAM for steady temperature environment (room temperature).

CHAPTER 4

WORK DONE, RESULTS AND DISCUSSION

4.1. METHODOLOGY

In accordance with the objective defined in the previous chapter, the model of a 6T SRAM cell was realised in the Tanner EDA tools as shown in Figure. 4.1. The resulting SPICE netlist was exported to HSPICE to assess the basic read/write working of the SRAM cell. The schematic of the cell was realised using MOSFET models which were later replaced with equivalent FinFET models from Predictive Technology model (PTM). The transistors PMOS1 and NMOS1 form one inverter and PMOS2 and NMOS2 the other inverter which are cross-coupled which can be seen in the schematic. The access transistors NMOS3 and NMOS4 can be turned on or off by the write line 'WL'. These transistors connect the inner cross-coupled inverters formed by PMOS1-NMOS1 and PMOS2-NMOS2 to the bit lines 'BL' and 'BLbar'.

The evaluations were performed on the 20nm, 16nm, 14nm, 10nm and 7nm FinFET technology nodes from the Predictive Technology Model (PTM) library. All models used are low standby power (LTSP) models. The bit lines and word lines are both fed signals by pulse voltage sources. The nominal source voltage for the SRAM cell was kept at a nominal voltage of 0.9V for fair comparisons. The bit line pulse period was 2ns and write line pulse was set at a period of 20ns throughout for each technology nodes.

The layout of the SRAM cell is designed in S-Edit tool based on the structure provided in [7]. The designed SRAM cell is as follows:

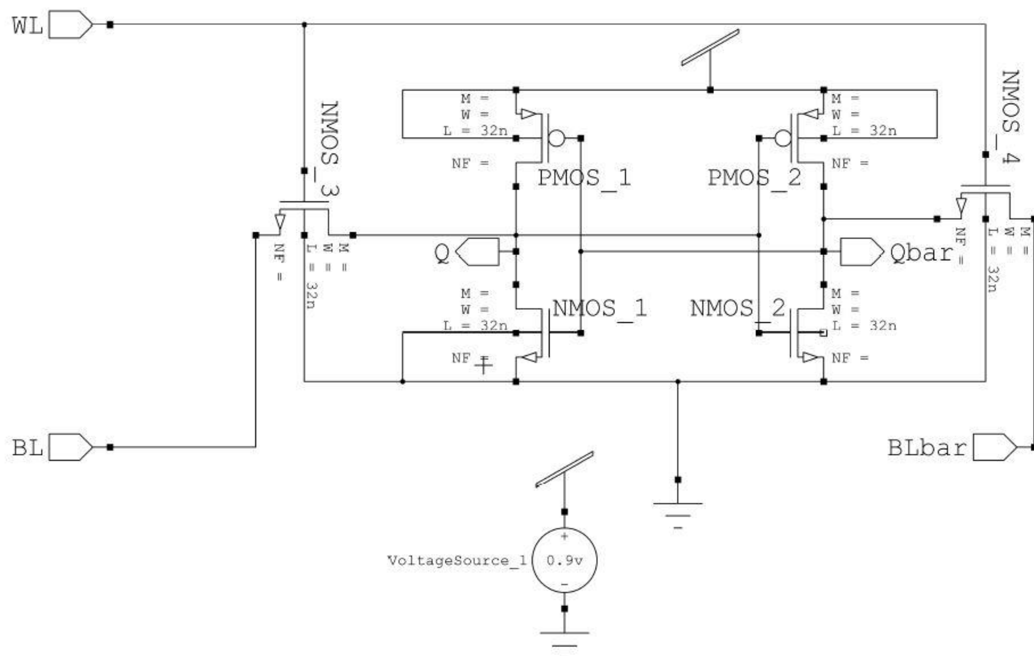


Figure 4.1 Design of the 6T SRAM cell in S-edit

The second circuit schematic using fine grain power gating was also modelled in S-edit tool in Tanner EDA tools, and the generated spice netlist was exported to HSPICE.

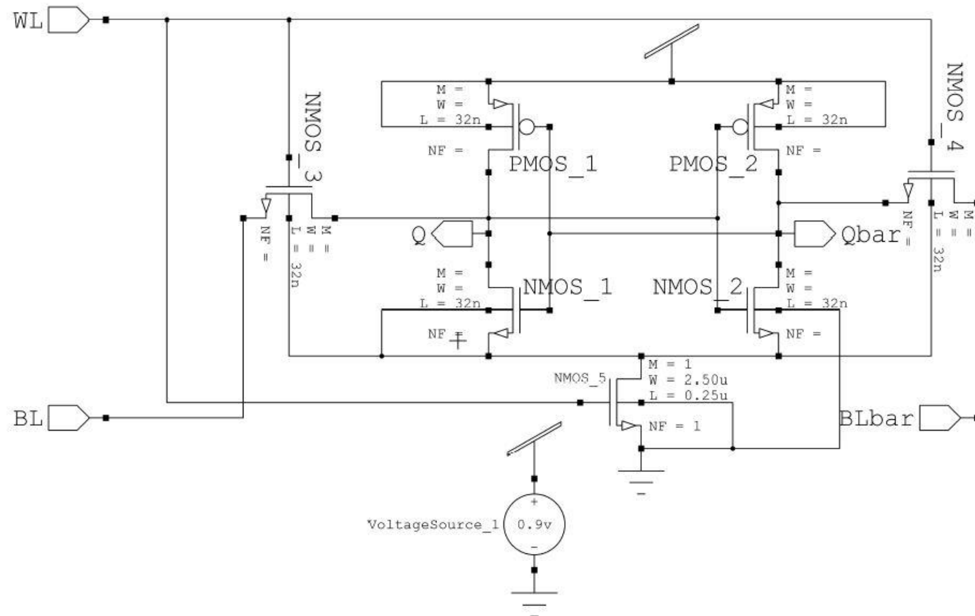


Figure 4.2 Design of SRAM cell with fine grain power gating.

4.2. SIMULATIONS PARAMETERS

The various simulations parameters that had been explicitly defined in the models are provided in table 4.1. All simulations were carried out with a nominal power supply, Vdd = 0.9V to keep the variables to a minimum.

Table 4.1 Simulation parameters for model files

	Fin height (nm)	Fin width (nm)	Effective length (nm)
7nm FinFET model	18	7	11
10nm FinFET model	21	9	14
14nm FinFET model	23	10	18
16nm FinFET model	26	12	20
20nm FinFET model	28	15	24

4.3. DELAY AND POWER

The designed SRAM cell is evaluated for dynamic power dissipation and the delay it takes to flip a bit inside the cell. The delay for writing ‘1’ and ‘0’ may differ either marginally or greatly depending upon the cell ratios, technology nodes, process variations and many other factors. The

timing diagrams for a standard SRAM cell under different technology nodes are shown followed by a comparative analysis of delay times in a tabular form.

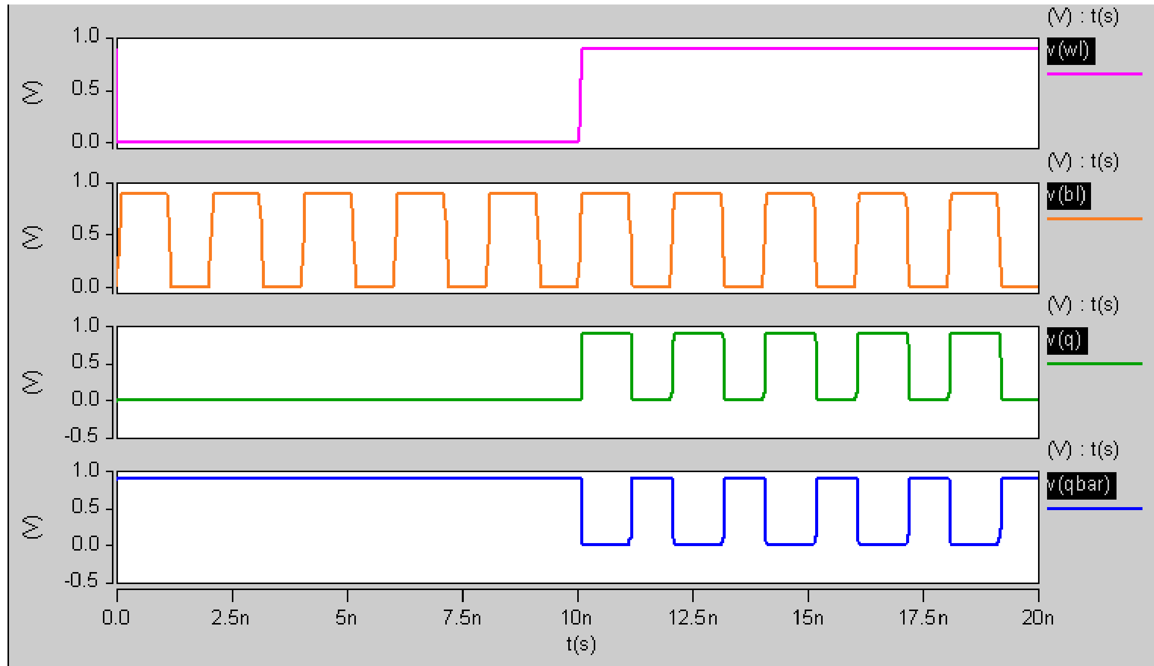


Figure 4.3 Timing Diagram for 7nm FinFET

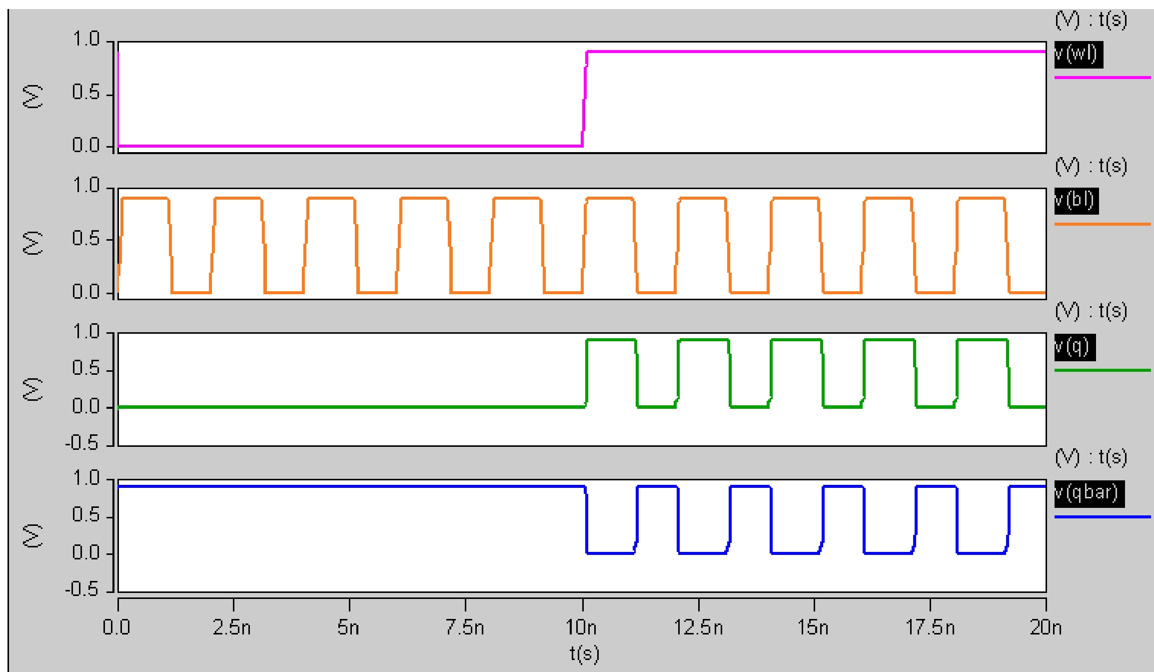


Figure 4.4 Timing Graph for 10nm FinFET

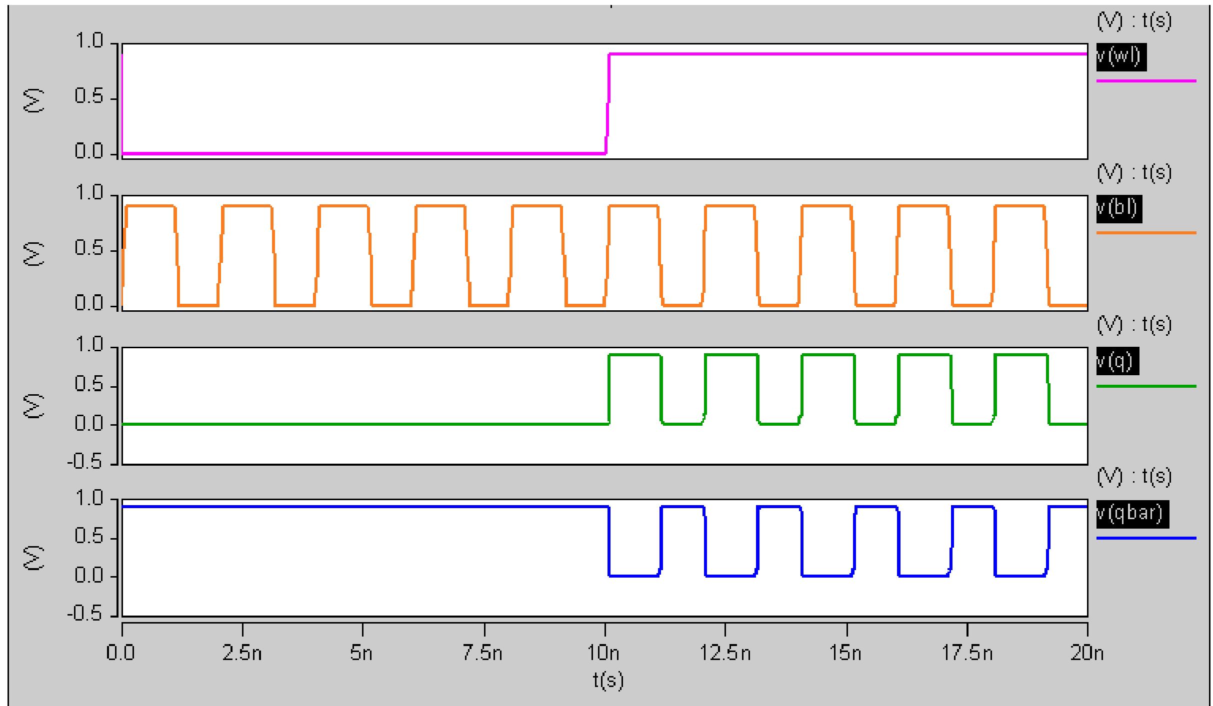


Figure 4.5 Timing Diagram for 14nm FinFET

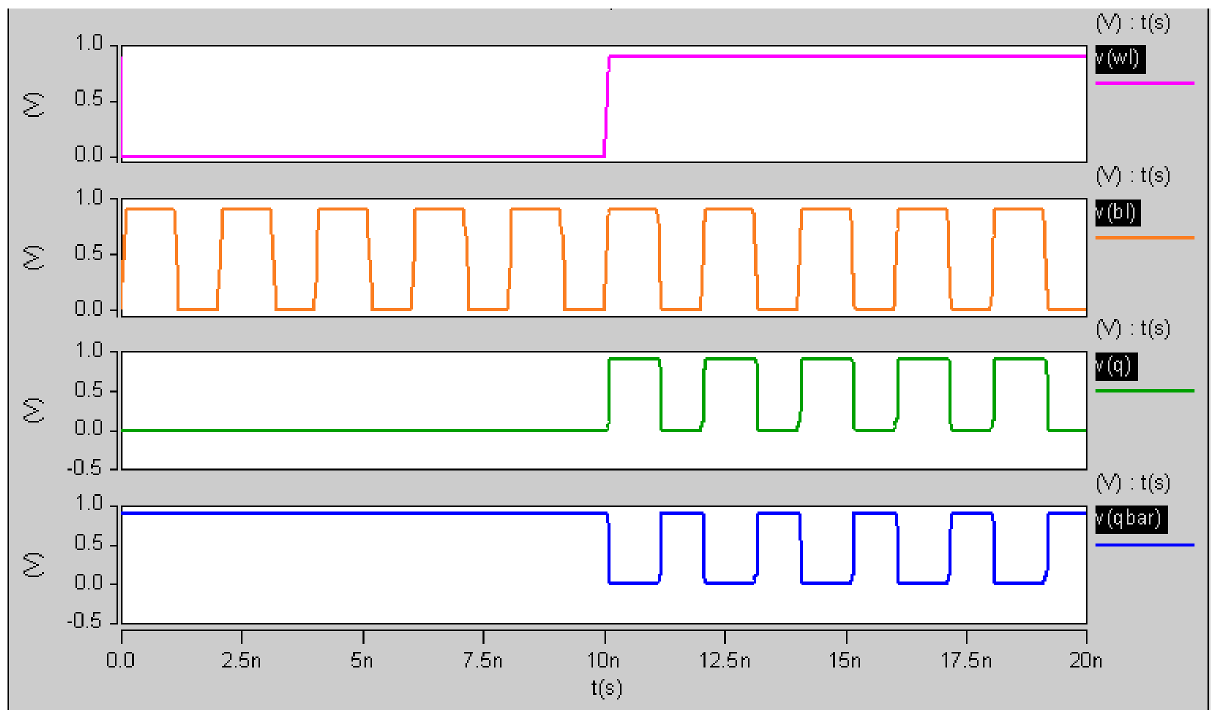


Figure 4.6 Timing Diagram for 16nm FinFET

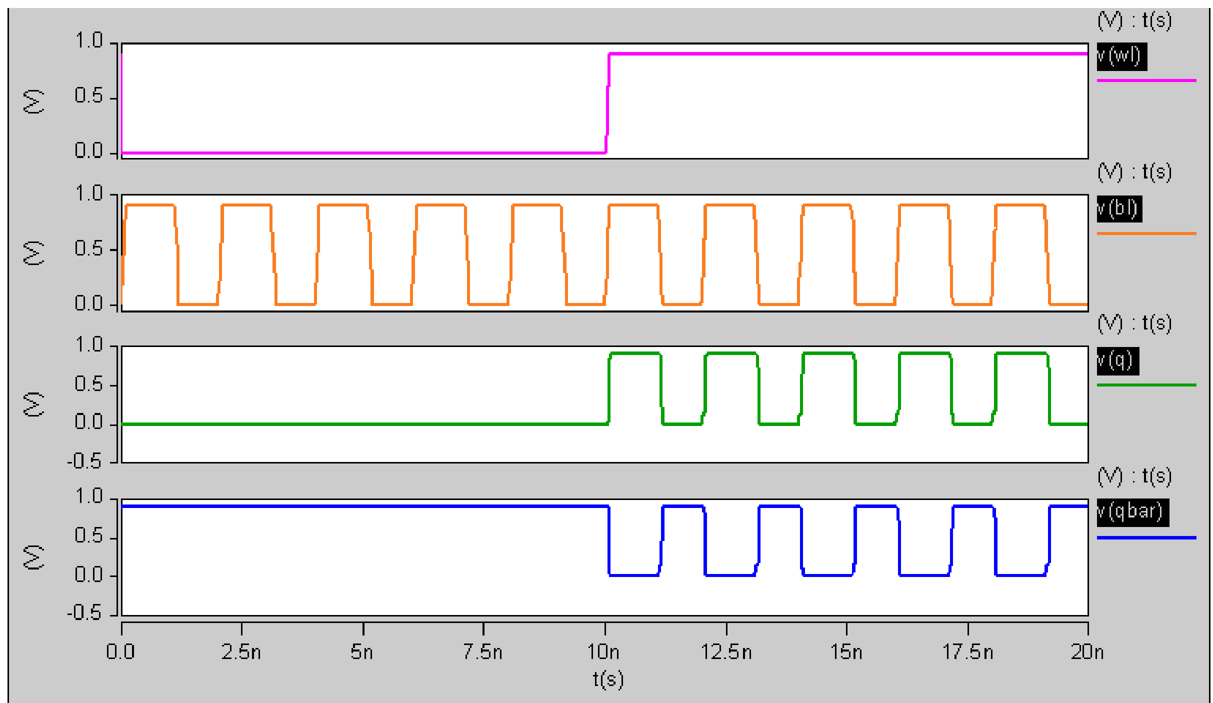


Figure 4.7 Timing Diagram for 20nm FinFET

The bit line (bl) is initialised with a pulsating voltage source with a period of 2ns. The rise time and fall time has been kept at a realistic low value of 0.1ps so as to simulate real world delay. The output nodes ‘q’ and ‘qbar’ are initialised with initial conditions of ‘0’ and ‘1’ respectively at the start of the simulation. The word line was set with pulse period of 20ns which results in a ten write cycles, corresponding to flipping the stored data bit five times. The same conditions are used to calculate the average dynamic power over the interval from 10ns to 20ns, when the cell is active. Table 4.1 shows the delay and power results.

Table 4.2 Delay Times and average dynamic powers for different technology nodes

	write '1' delay (ps)	write '0' delay (ps)	Average Dynamic Power (nW)
7nm standard SRAM cell	4.107	6.137	346.9
10nm standard SRAM cell	5.126	5.749	450.8
14nm standard SRAM cell	6.155	6.443	452.4
16nm standard SRAM cell	5.064	8.323	472
20nm standard SRAM cell	8.648	12.2	303.2

The above data is compiled in a form of a bar graph to better understand the trend of the above measured metrics.

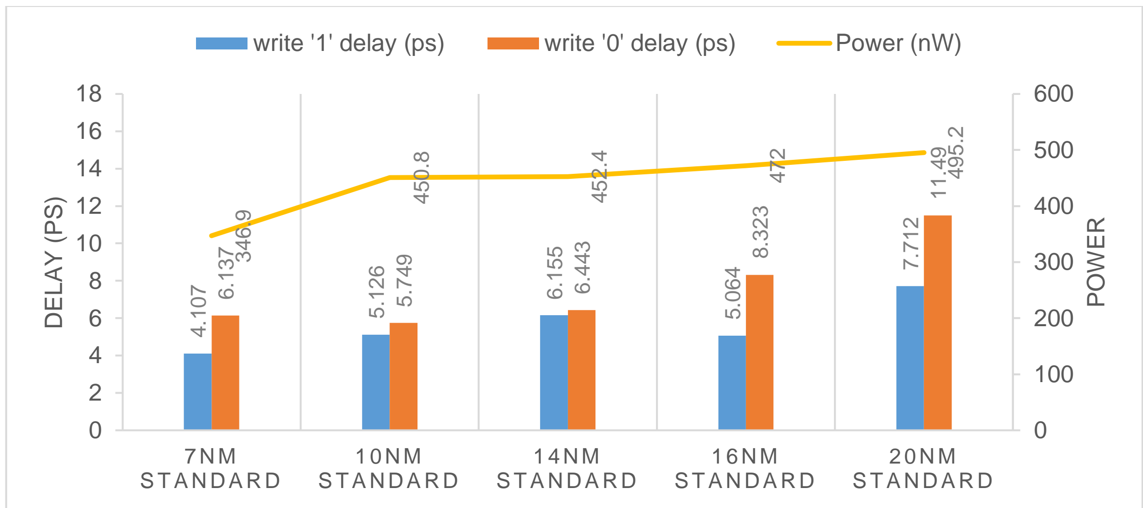


Figure 4.8 Comparison of delay times for different technology nodes

The graphical representation of the data gives us a clear picture of an increasing trend in the delay times as the technology node is scaled upwards. However, it is interesting to observe that write ‘1’ delay time is always smaller than the write ‘0’ delay. This can be attributed to the fact that the stored bit ‘1’ has to discharge through the access transistor and word line. The cell ratio of the cell also plays a big role in defining the delay times. A lower write delay can affect the read SNM of the cell making susceptible to the data corruption while reading. The whole process was repeated with the second schematics shown in the Figure 4.2.

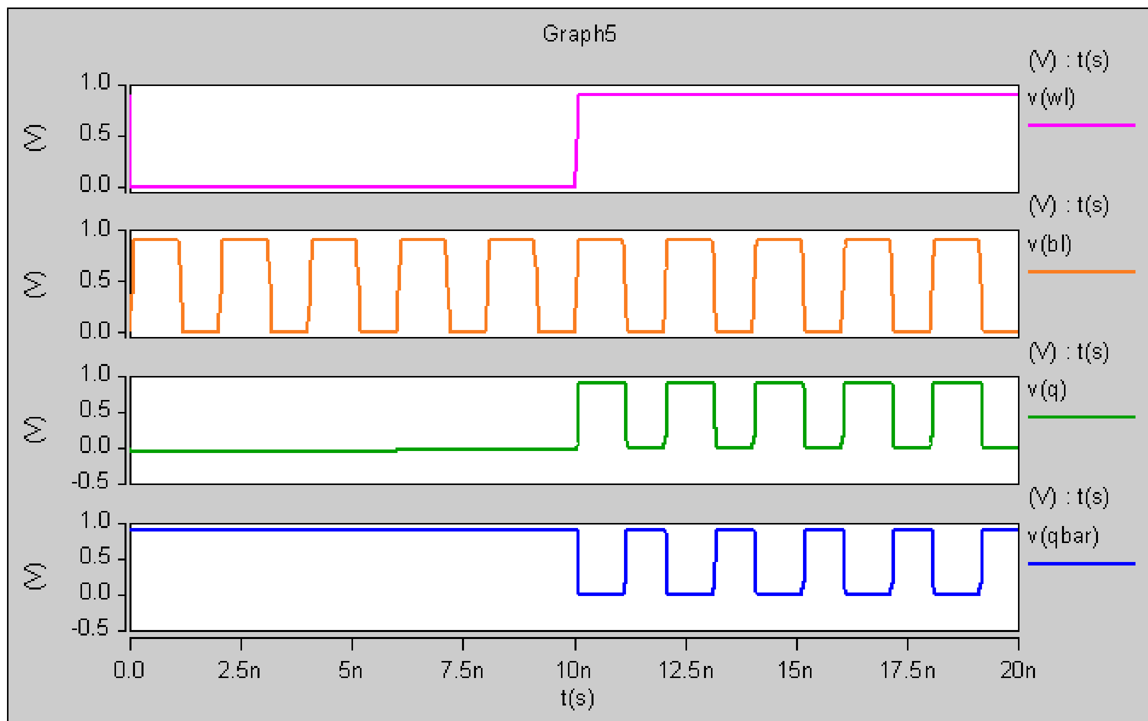


Figure 4.9 Timing diagram for 7nm finegrain

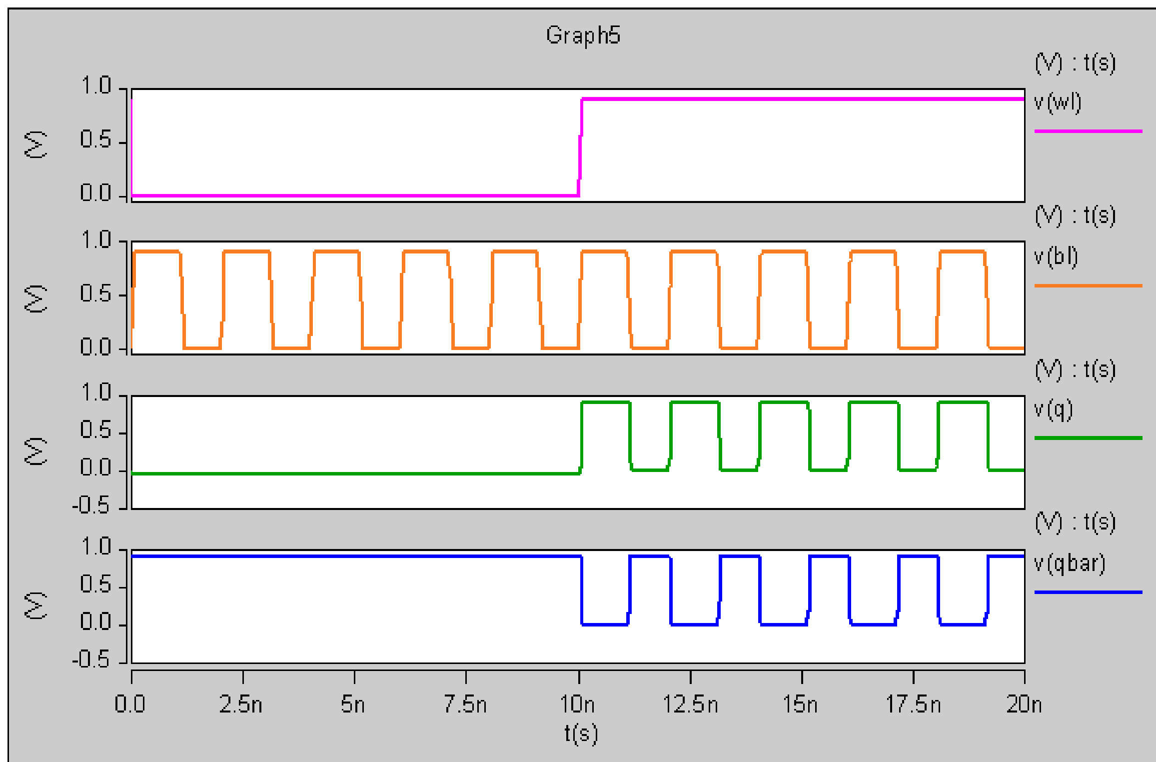


Figure 4.10 Timing diagram for 10nm finegrain

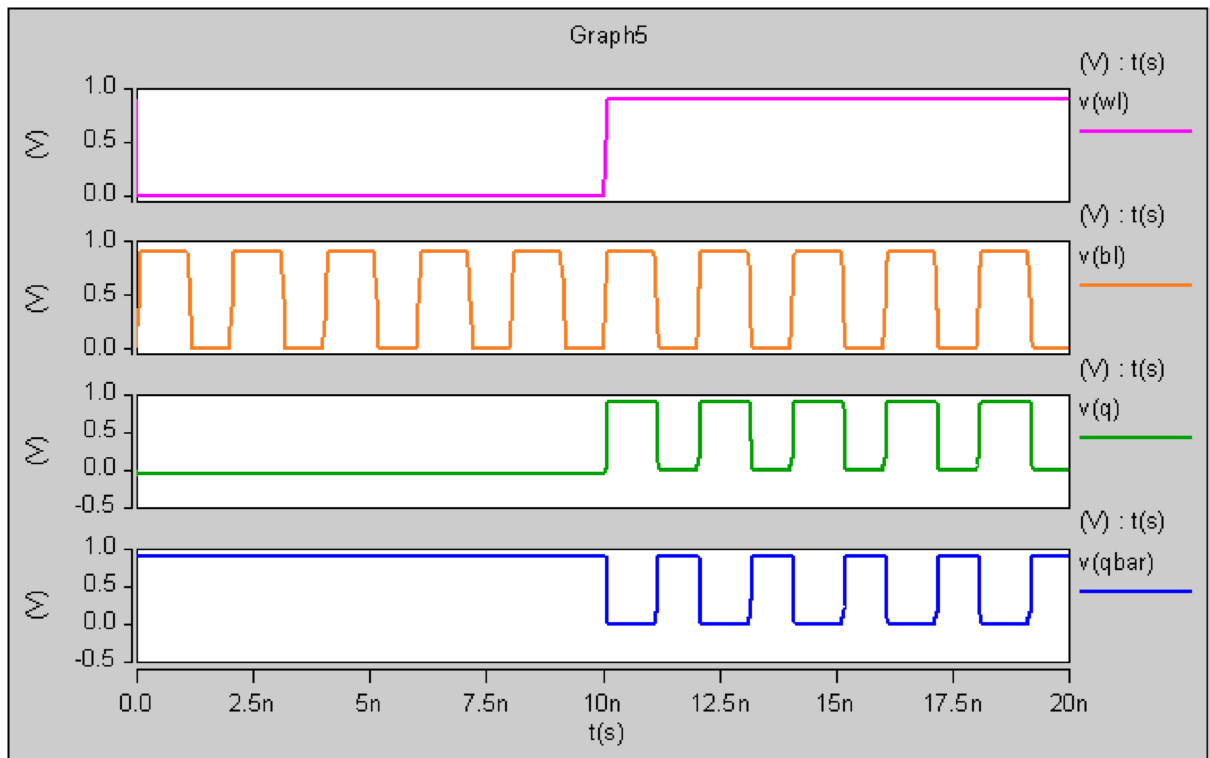


Figure 4.11 Timing diagram for 14nm finegrain

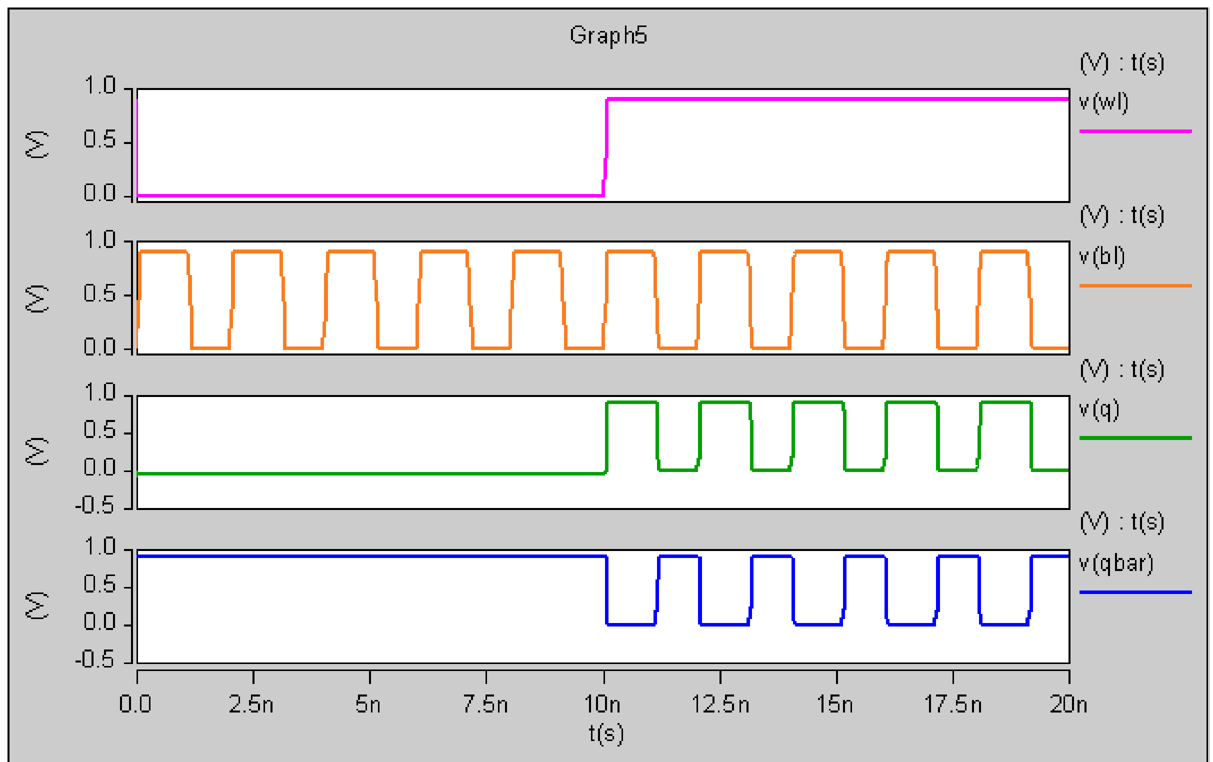


Figure 4.12 Timing diagram for 16nm finegrain

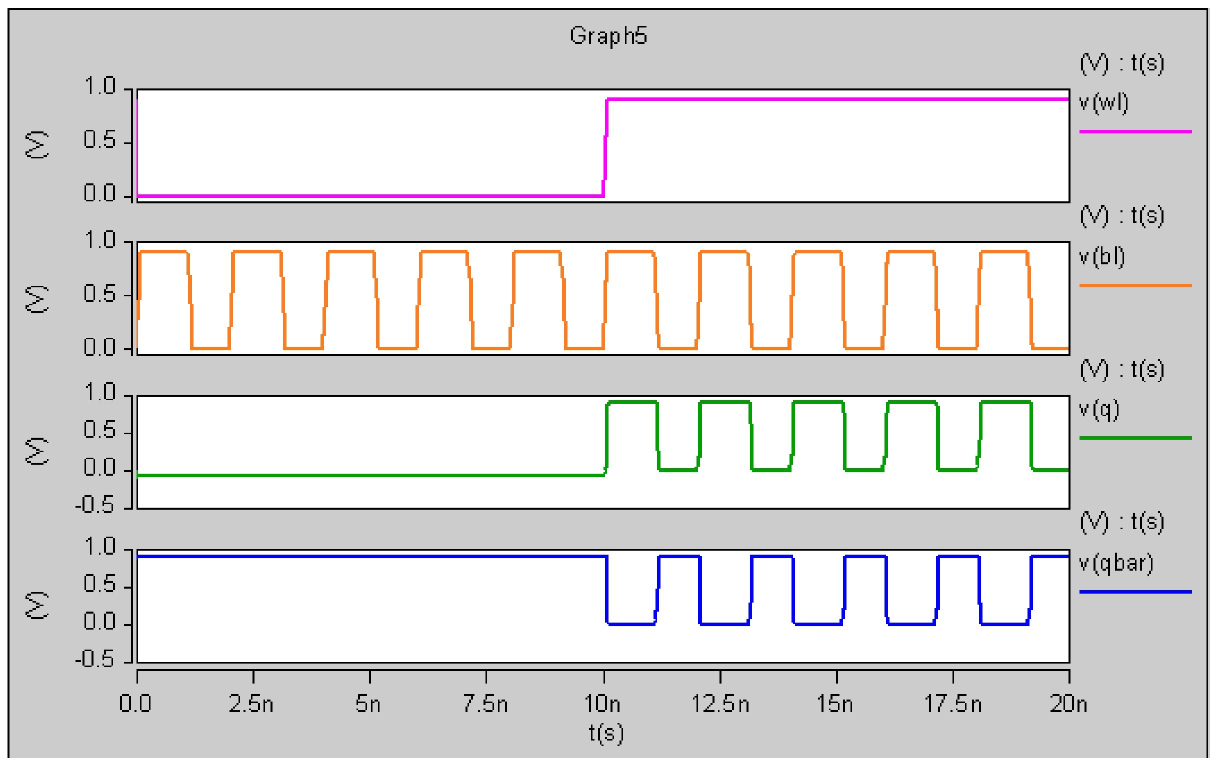


Figure 4.13 Timing diagram for 7nm finegrain

The corresponding data gathered from the above simulations is curated in a tabular form to better understand the varying trends in delay times and power dissipation.

Table 4.3 Delay Times and average dynamic powers for fine grain cell at different technology nodes

	write '1' delay (ps)	write '0' delay (ps)	Average Dynamic Power (nW)
7nm FineGrain	6.134	5.756	262.2
10nm FineGrain	4.895	6.314	297.6
14nm FineGrain	5.749	7.719	347.3
16nm FineGrain	5.673	9.122	375.2
20nm FineGrain	8.621	13.26	386.4

As expected the same trend follows in terms of power dissipation and delay time, however, at 7nm node the trend slightly changes with the write '1' delay being greater than the write '0' delay. Other than that the trend seems to be as predicted as shown in the graph.

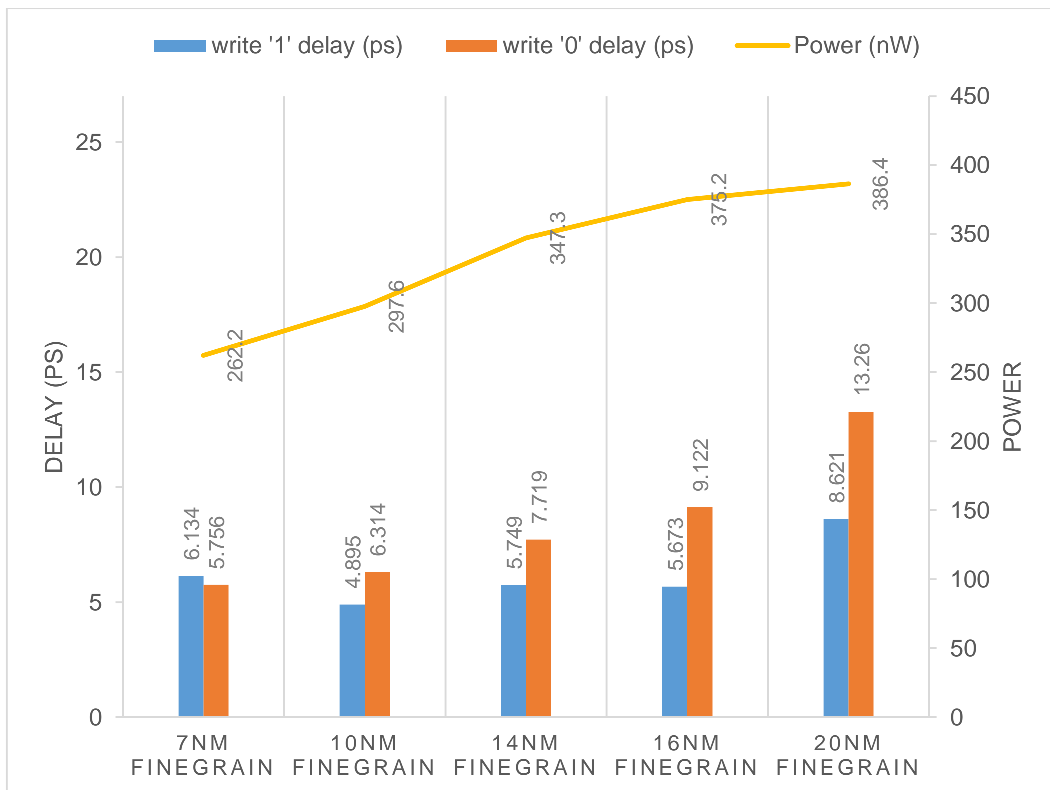


Figure 4.14 Comparisons of delay time and average power dissipation for fine grain power gating

The average power dissipation is found to be lesser with implementation of power gating when compared to its respective technology node. The improvements in power dissipation for cells at bigger node are comparable to that of the cells at lower nodes without gating. This is one place where gating can be implemented to reduce the power dissipation without opting for lower node transistors, the fact that manufacturers have to invest a huge amounts of capital while upgrading from one technology node to other. This can prove to be a kind of retro fitting

option for low cost, low power applications. The delay times, on the other hand, has an increasing trend as shown in the graph.

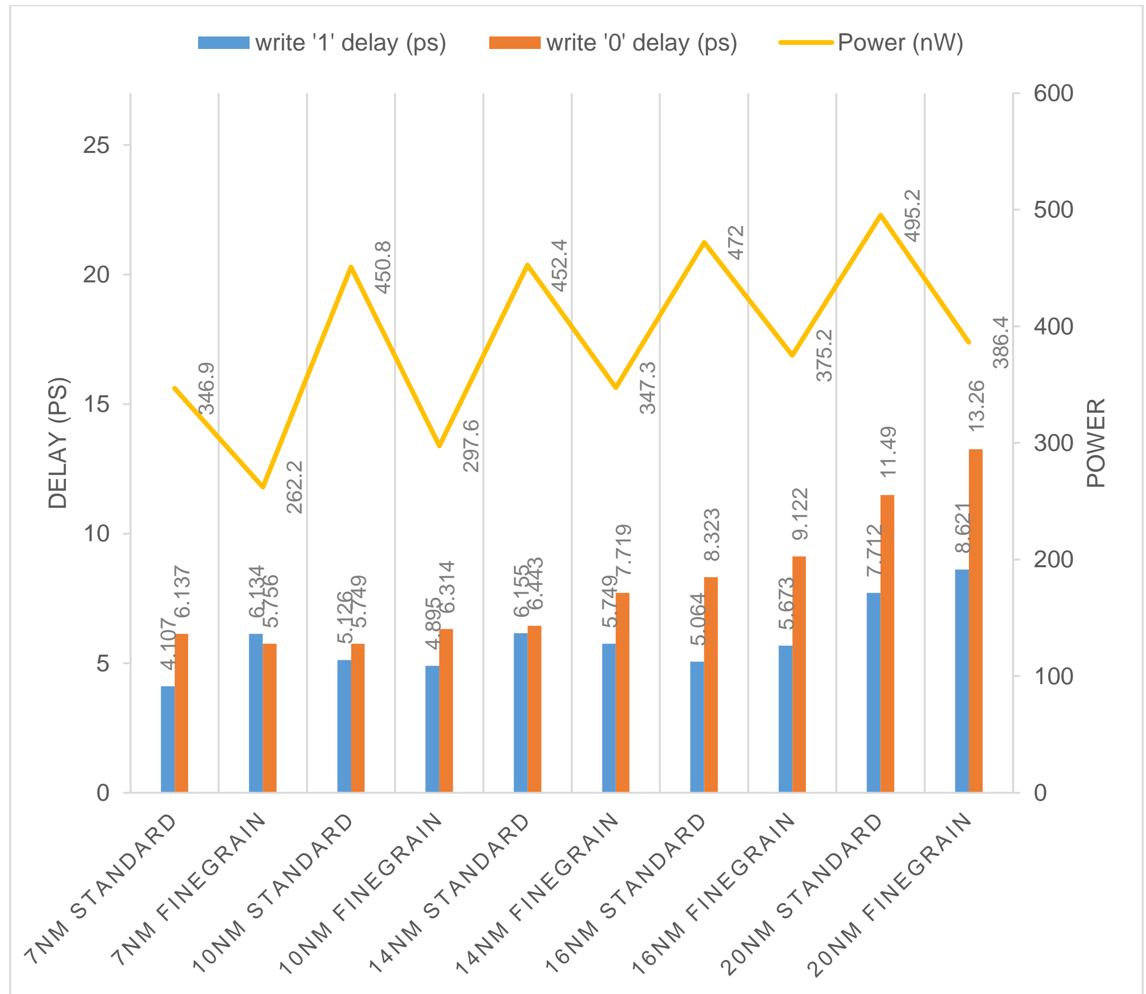


Figure 4.15 Comparison between standard cell and gated SRAM cell for each technology node

It can be observed that the power dissipation of 20nm fine grain cell is less than that of the standard cell SRAM at 16nm node. In fact, it is lower than all excluding the standard cell SRAM at 7nm node. It is found out that on average the fine grain cell has 24.82% less power dissipation than its equivalent standard cell counterpart.

The delay time trends aren't as straight forward as the power dissipations and may seem somewhat random. However, on a closer look it is evident that delay time is affected significantly. The average percentage increase in write '1' delay is found out to be 12.4129%. The highest reported increase in the 7nm node was with a staggering 49.35% increase in delay time. On the flip side, the lowest witnessed increase is actually an improvement with -4.51% and -6.59% decrease in delay time for 10nm and 14nm technology nodes.

The same trend follows for the write ‘0’ delay time, with an average increase in time of 9.68% with the highest being 19.08% for 14nm node and lowest -6.20% in 7nm node, which is improvement over the standard cell. Therefore, it makes much more sense to compare the average delays rather the individual write delays. The calculated average delay times along with the power delay product (PDP) are listed in the table. The power delay profile or switching energy is correlated with energy efficiency of a logic circuit.

Table 4.4 Comparison of average power and PDP for standard cell and fine grain cell at various technology nodes

	Average Power (nW)	average delay	PDP
7nm standard	346.9	5.122	1776.8218
7nm FineGrain	262.2	5.945	1558.779
10nm standard	450.8	5.4375	2451.225
10nm FineGrain	297.6	5.6045	1667.8992
14nm standard	452.4	6.299	2849.6676
14nm FineGrain	347.3	6.734	2338.7182
16nm standard	472	6.6935	3159.332
16nm FineGrain	375.2	7.3975	2775.542
20nm standard	495.2	9.601	4754.4152
20nm FineGrain	386.4	10.9405	4227.4092

The average percentage change observed in the average delay time comes out to be 10.10%. This means that cells with power gating can be expected to be 10% slower than without the power gating. However, when power delay product is taken into consideration, the gated cells have better PDP than the cells at same technology node. The power delay should be minimum for an energy efficient logic circuit, satisfying the definition, we have the cell with power gating at 7nm node with the lowest PDP. However, PDP only tells us about the efficiency of the circuit and not its stability. For that we need to calculate the static noise margin (SNM) for each circuit.

4.4. STATIC NOISE MARGIN

Another critical major metric to determine the performance of a SRAM is SNM, higher the SNM higher is the resistance of the SRAM to changes due to noise that may occur from the BIT line voltages or the voltage sources itself. The SNM was calculated graphically using the HSPICE code.

By definition, SNM is the minimum voltage required at each of the cell storage nodes to change the state of the cell. Graphically, the SNM was obtained by plotting the inverter characteristics of the two inverters on top of each other, creating a butterfly pattern or the eye diagram. The side of the maximum square that can be drawn between the eyes or the wings of the butterfly curve determines the SNM of the that SRAM cell as shown in the Figure. 4.16.

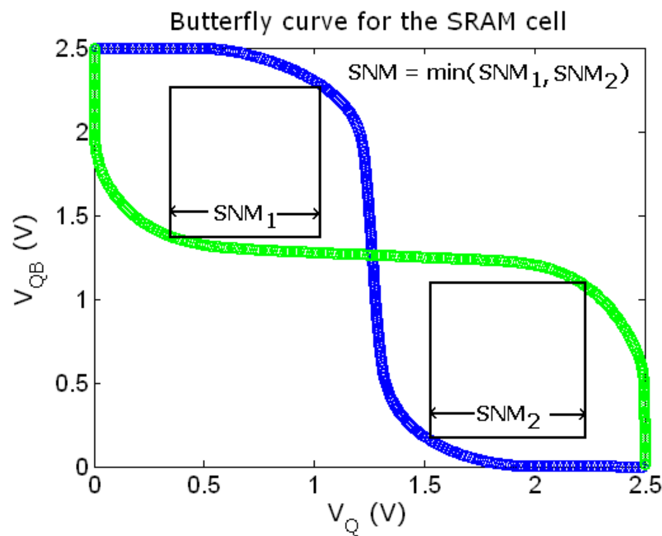


Figure 4.16 calculation of the SNM for a SRAM cell [23]

To implement the above mentioned cell in the HSPICE simulation, we require voltage controlled voltage sources (VCVS). [23]. The butterfly curve is rotated 45 degrees to calculate the maximum and minimum points on the inverter characteristic curves. The straight line joining the two points (maxima and minima) is the diagonal of the largest square that can fit inside this curve (see fig 4.17). The SNM then can be found simply by applying Pythagoras theorem.

The SNM is found when initiating a read process, hence the SNM found corresponds to the read SNM. To find the read SNM, the word line WL is set to high which causes the inverter characteristics to change, making them susceptible to noise, the minimum of this noise is our required noise margin for the SRAM cell in question.

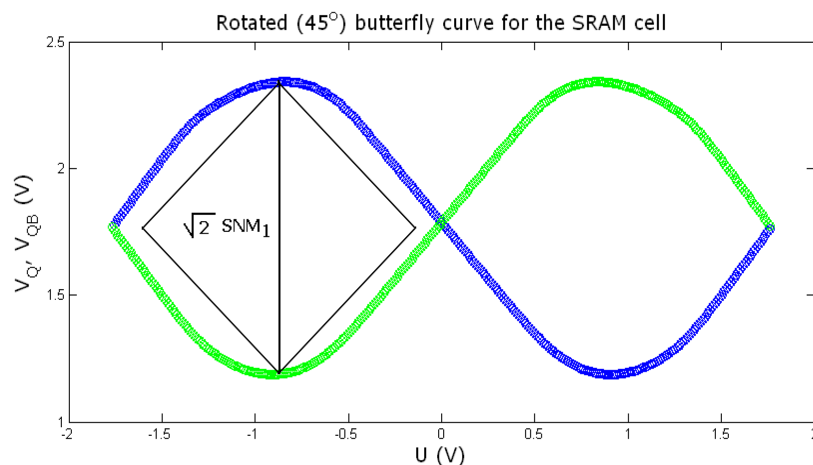


Figure 4.17 determination of the SNM by rotating the butterfly curve [23]

Following are the SNM curves of the SRAM cells mentioned in the previous section. Only the, butterfly curve is shown, as the 45 degrees rotated version is essentially the same. The SNM reading

are then presented in tabular form to have better comparison of the SRAM cell at different technology nodes with and without the fine grain gating method.

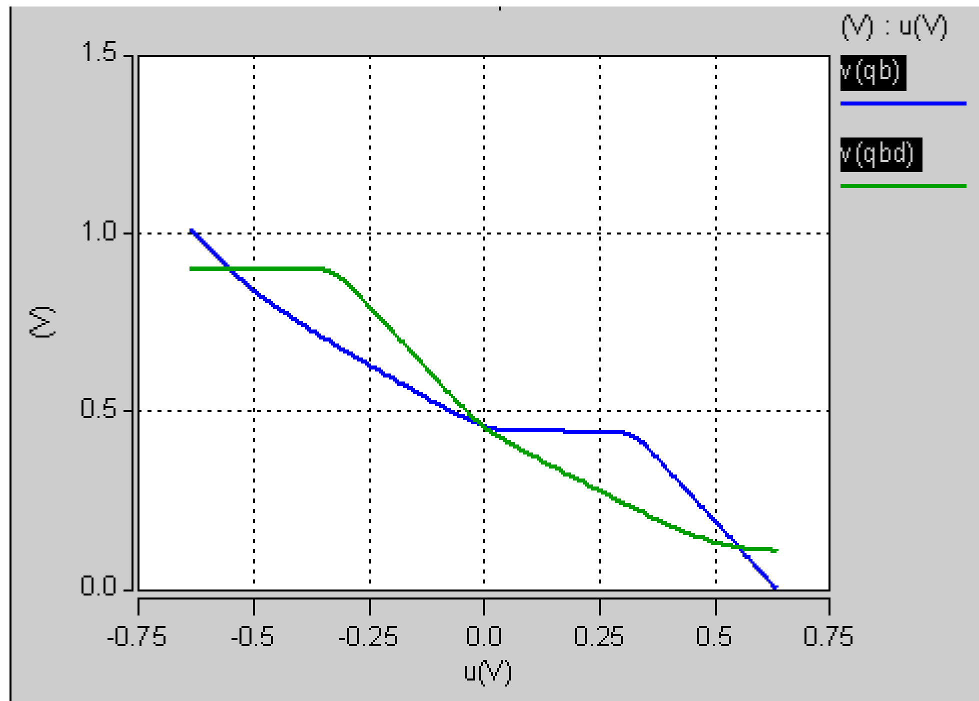


Figure 4.18 butterfly curve for 7nm FinFET SRAM cell

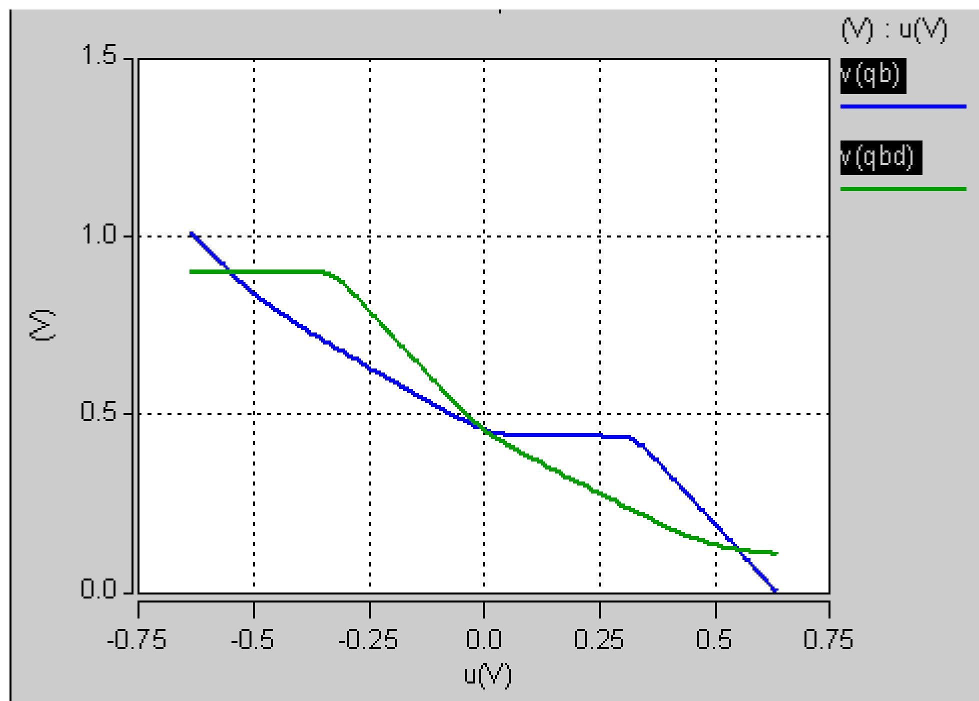


Figure 4.19 butterfly curve for 10nm FinFET SRAM cell

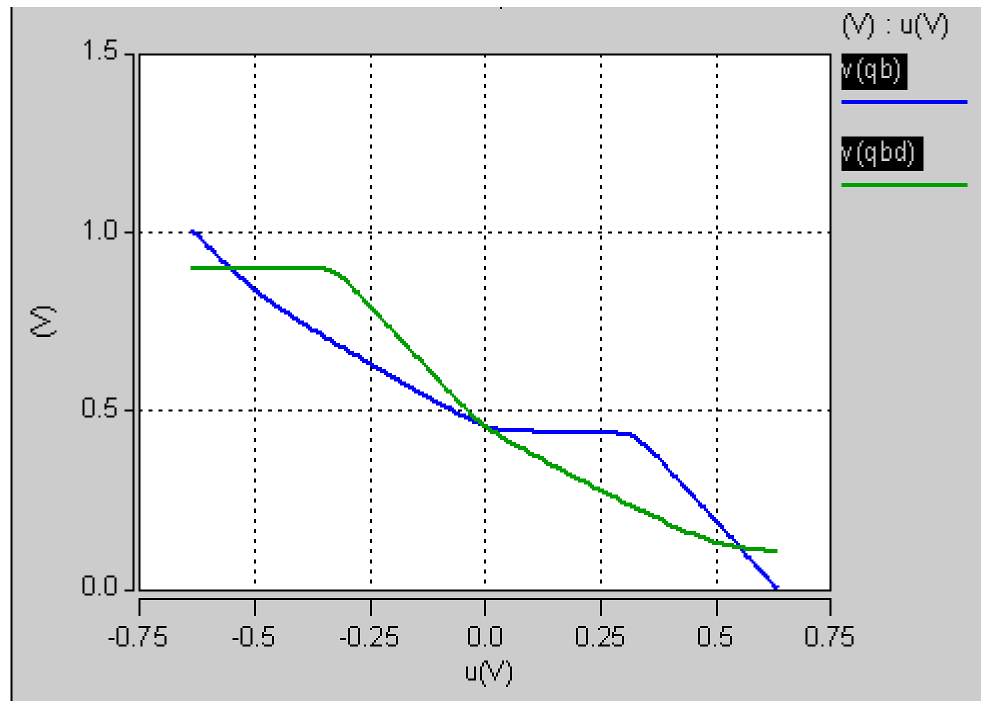


Figure 4.20 butterfly curve for 14nm FinFET SRAM cell

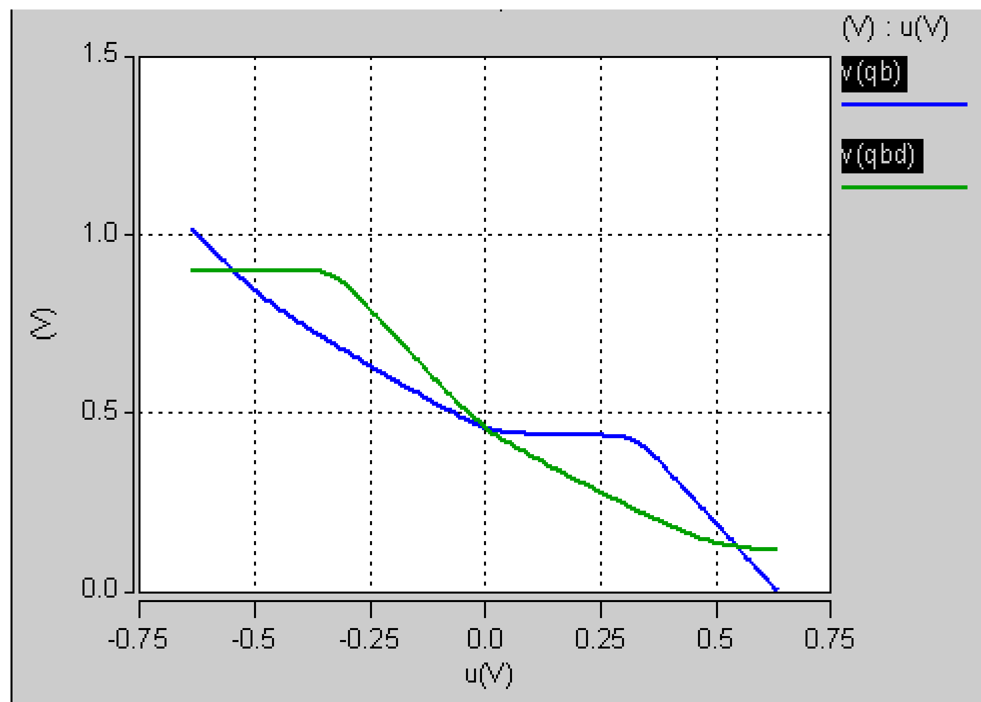


Figure 4.21 butterfly curve for 16nm FinFET SRAM cell

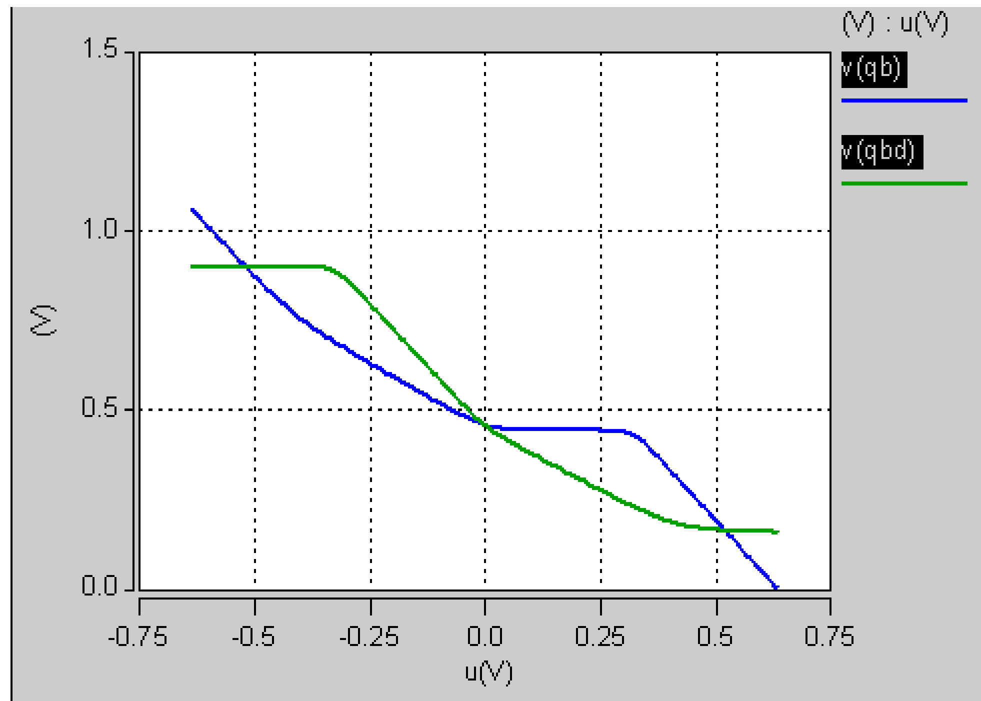


Figure 4.22 butterfly curve for 20nm FinFET SRAM cell

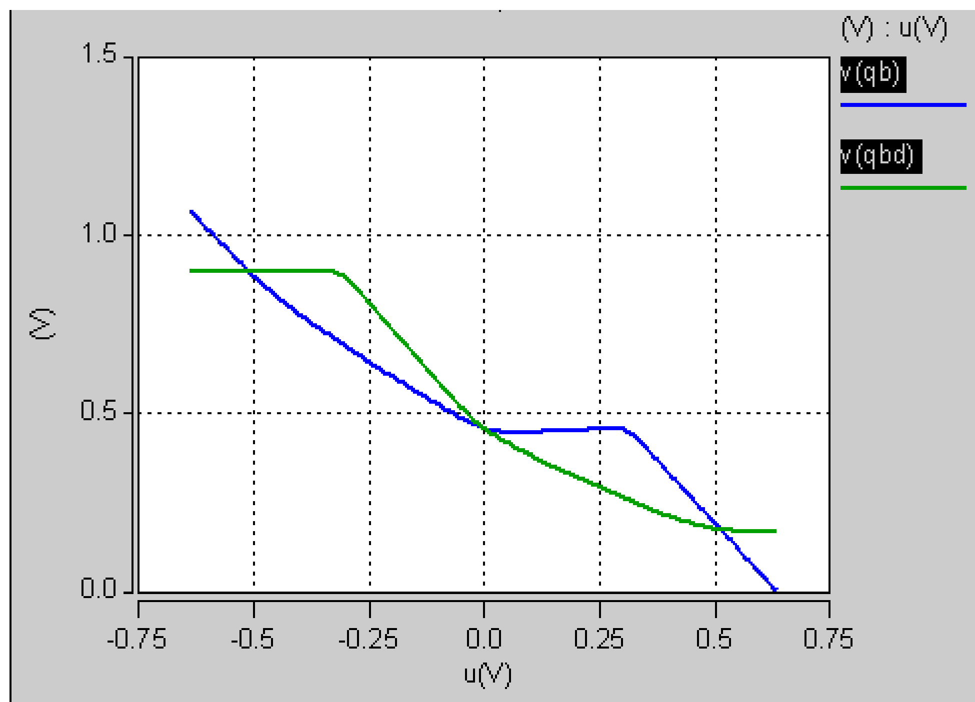


Figure 4.23 butterfly curve for fine grain 7nm FinFET SRAM cell

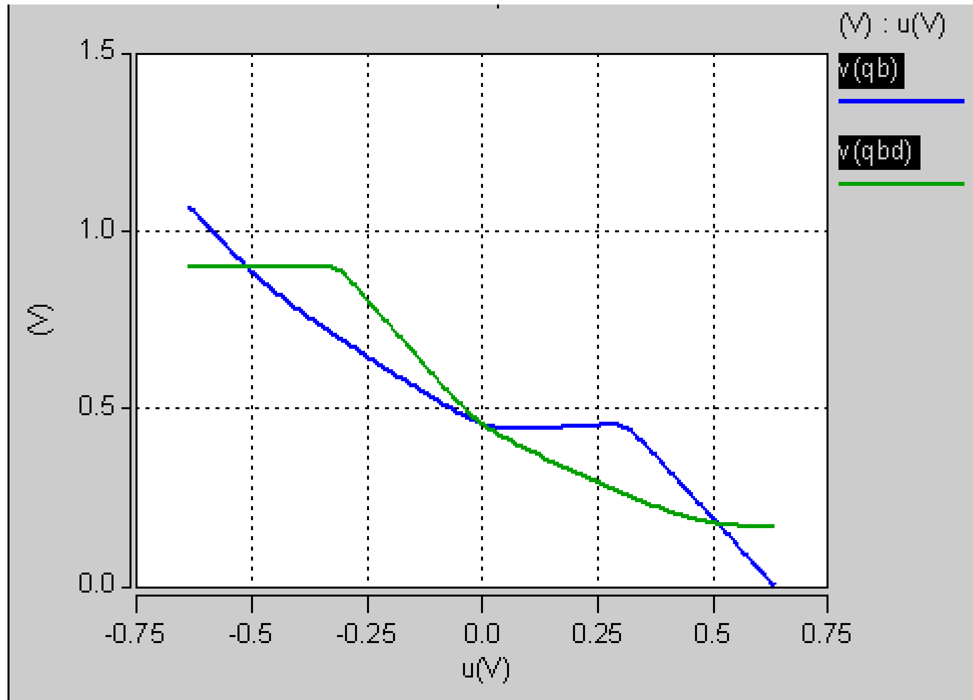


Figure 4.24 butterfly curve for fine grain 10nm FinFET SRAM cell

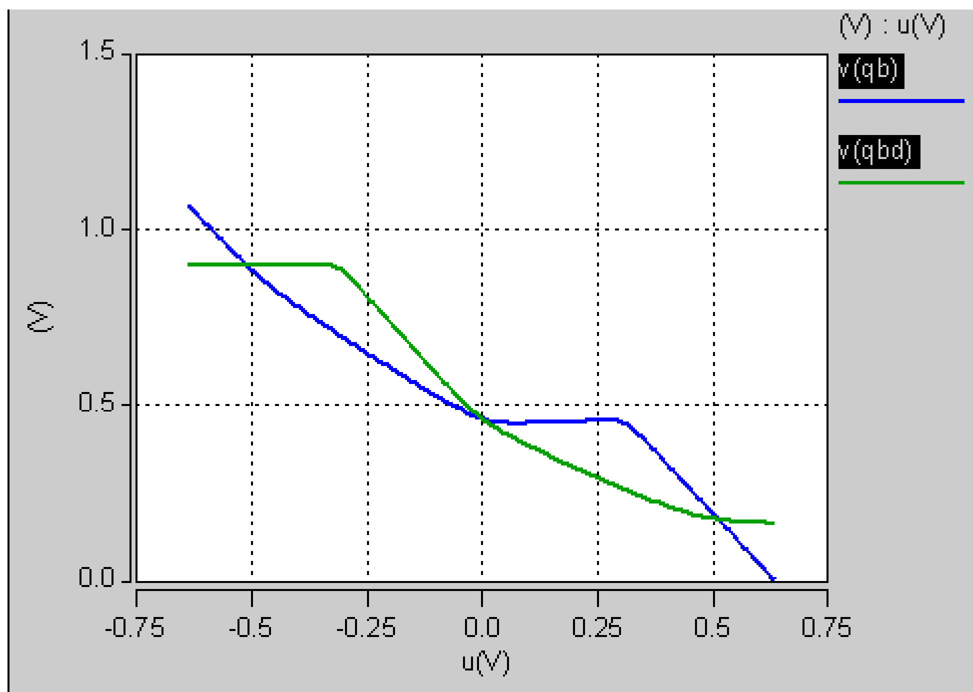


Figure 4.25 butterfly curve for fine grain 14nm FinFET SRAM cell

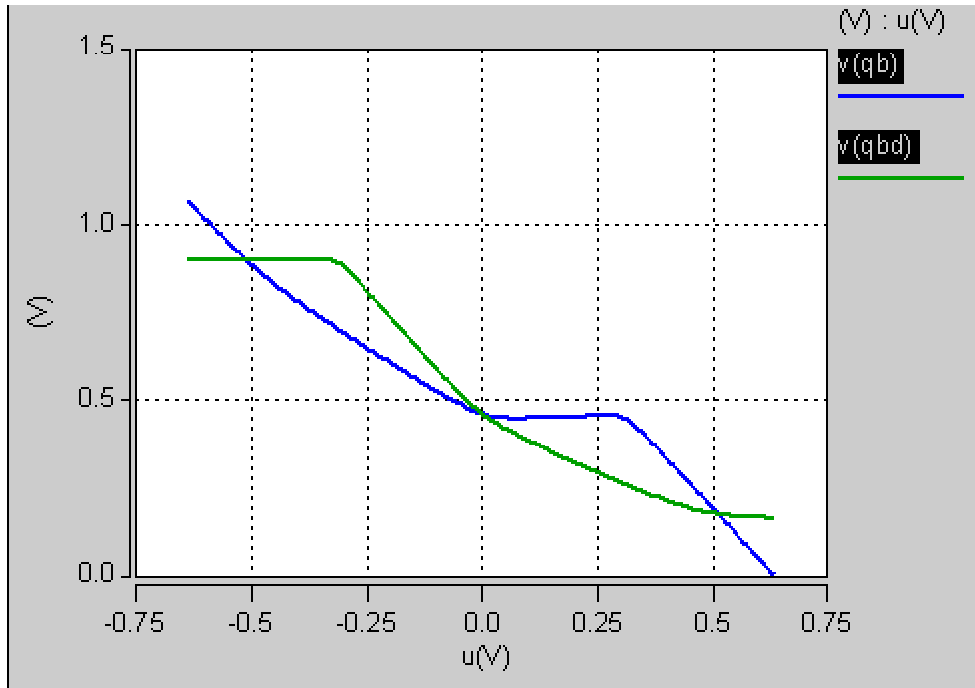


Figure 4.26 butterfly curve for fine grain 16nm FinFET SRAM cell

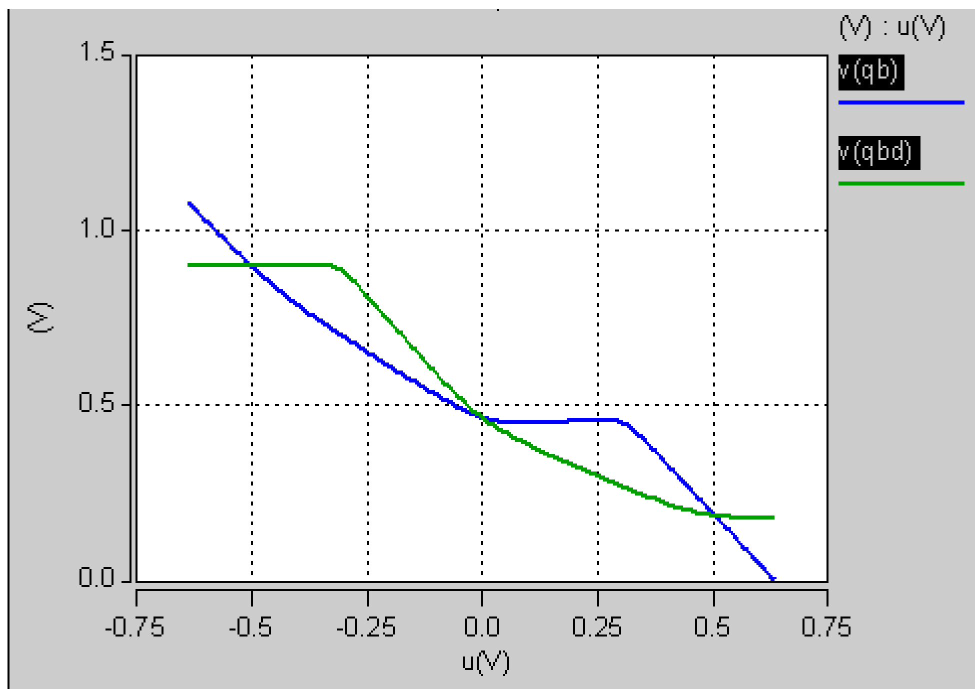


Figure 4.27 butterfly curve for fine grain 20nm FinFET SRAM cell

Table 4.5 SNM and PDP data for different SRAM cell. The highest and second highest SNM, and the lowest and the second lowest PDP is highlighted for reference.

	SNM (mV)	average delay	PDP
7nm standard	199.6	5.122	177.68218
7nm FineGrain	190.5	5.945	155.8779
10nm standard	196	5.4375	245.1225
10nm FineGrain	188.1	5.6045	166.78992
14nm standard	195.2	6.299	284.96676
14nm FineGrain	187.9	6.734	233.87182
16nm standard	190.5	6.6935	315.9332
16nm FineGrain	182.9	7.3975	277.5542
20nm standard	182	9.601	475.44152
20nm FineGrain	208.2	10.9405	422.74092

Observing the minor differences in the butterfly curve cannot be appreciated with naked eyes. However, the table gives a detailed summary of the simulation results so far. It is observed from the above data that the general trend for SNM for a standard cell SRAM cell follows a decreasing pattern as the technology node is increased. This follows the theoretical trend as explained in the previous chapters and hence confirms the theoretical aspect of the SRAM cells in general. The above data is plotted to better comprehend this trend.

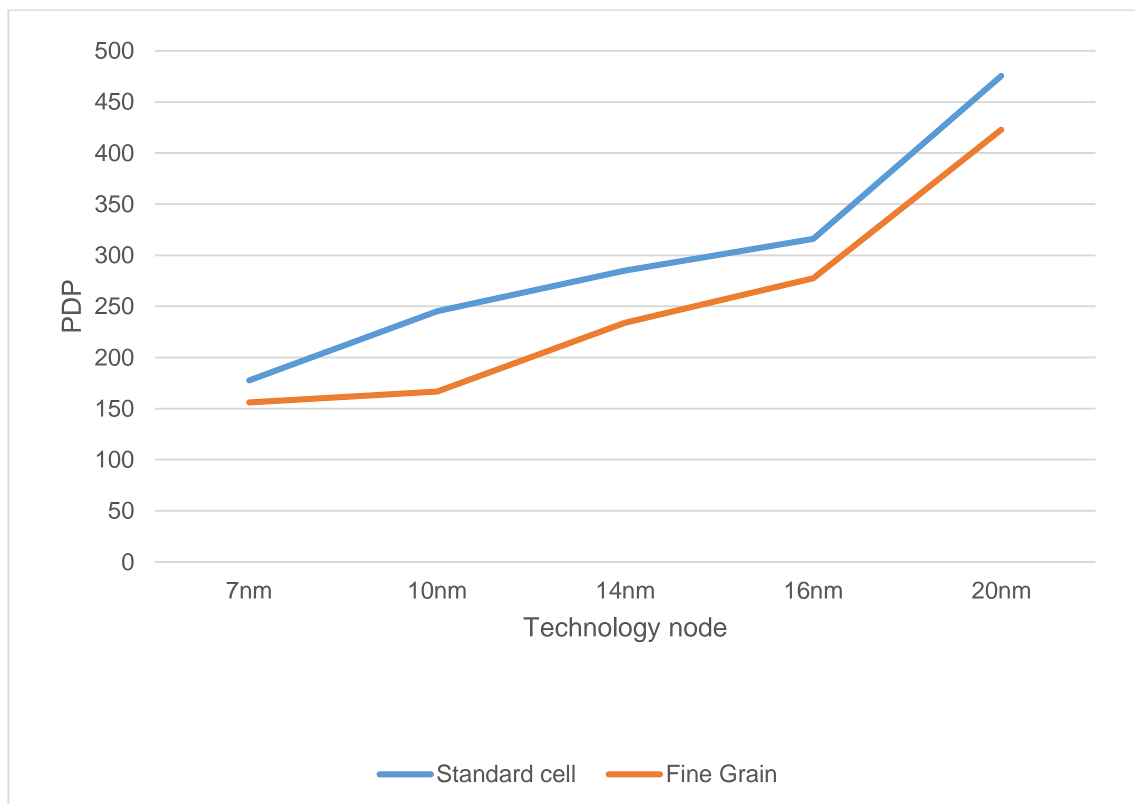


Figure 4.28 PDP trend with increasing technology nodes

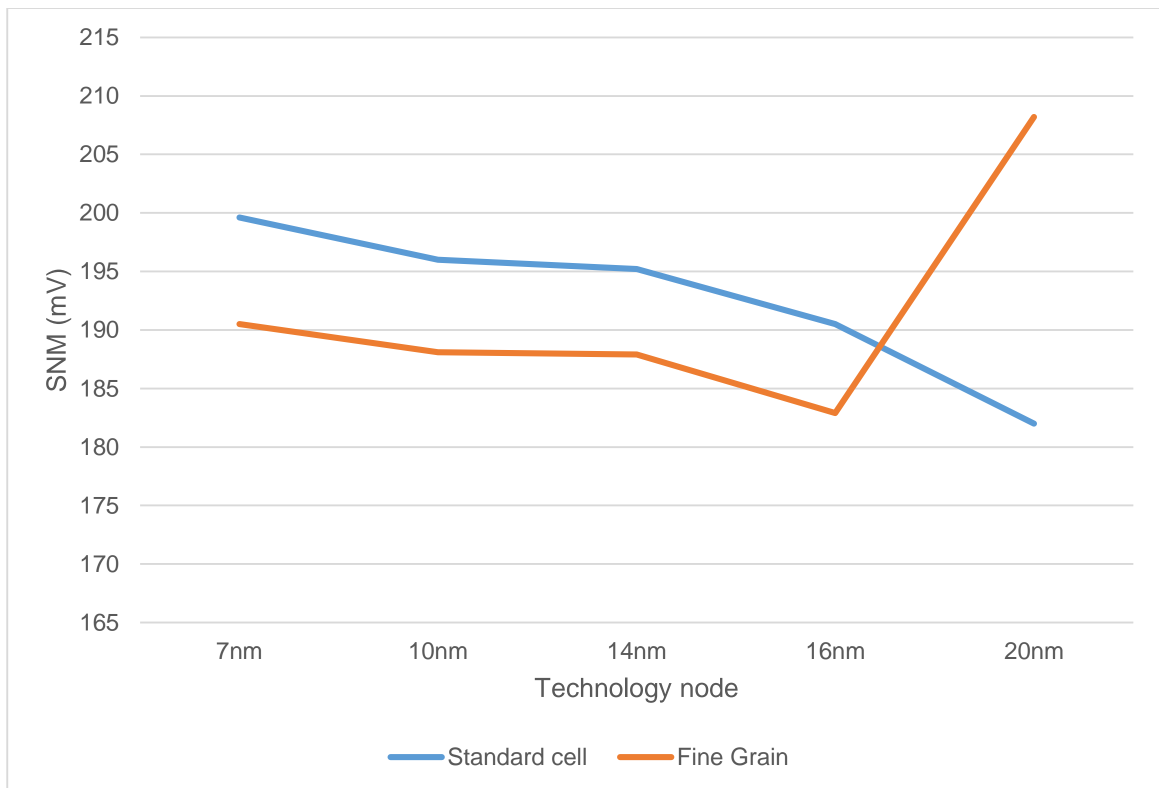


Figure 4.19 SNM trend with increasing technology nodes

Analysing the data gathered from the simulation, we can reach to a conclusion as to which can be a suitable candidate for a model SRAM cell, which is fast and reliable. The parameters to determine this are primarily the PDP and SNM. Firstly, considering the PDP, it has been the lowest of all the SRAM cells. It is found that the 7nm FineGrain cell has the lowest PDP followed by the 10nm FineGrain. However, this alone isn't sufficient to determine the better of them, as it can be the 10nm FineGrain despite having a larger PDP has a lower average delay time, at the same time having lesser SNM than 7nm FineGrain. Here, we can easily rule the 10nm FineGrain cell.

The second metric, SNM, when taken into account, revealed that the 20nm FineGrain cell has the maximum SNM which is a desirable property. However, the PDP is far worse than the 7nm FineGrain cell. Hence, ruling out the 20nm Fine Grain cell. The second highest SNM is for the 7nm standard cell, albeit having a slightly larger PDP than the 7nm Fine Grain cell, it also has a lower average delay time than the 7nm Fine Grain cell.

Since the simulation data confirms with the theoretical and mathematical models, it can be then safely concluded that the lower technology nodes, offer better stability, write speeds and power efficiency as compared to the higher technology nodes.

4.5. EFFECT OF TEMPERATURE

All the simulations till now were carried out under an ideal temperature of 25 degrees Celsius. As we know, devices don't normally operate in an ideal environment and has to stand to abuse of the elements. Hence, the same circuits models were simulated under a varying temperature environment ranging from -30 degree Celsius to 80 degrees Celsius in 1 degree Celsius steps. The SRAM parameters, namely, SNM, delay and average were calculated for each degree rise in the temperature. The resultant data set is very large (exactly, 110 simulations per SRAM model, equating to a total of 1100 simulations) and hence has been consolidated to a reasonable degree to make the comparison. The first order of measure is the write delay of the SRAM cells. The following graphs depicts write delays with increasing temperature. At first glance, the curves may seem random but upon closer inspection it can be noted that for the write '1' delay for the most part stays below the write '0' delay and the overall trend seems to indicate towards a rising delay time with increasing temperatures. The sharps dips can be attribute to the imperfections in the simulation environment and models files. It was also noticed that an SRAM cell at 20nm node seems to be stable with a steady increasing in both standard cell configuration and with fine grain power gating. This can be explained by the self-heating property of FinFETs, with larger surface area for a higher technology node, it is easier to loose heat quickly while for the lower nodes the heat is not efficiently dissipated causing further heating of the device causing a chain reaction.

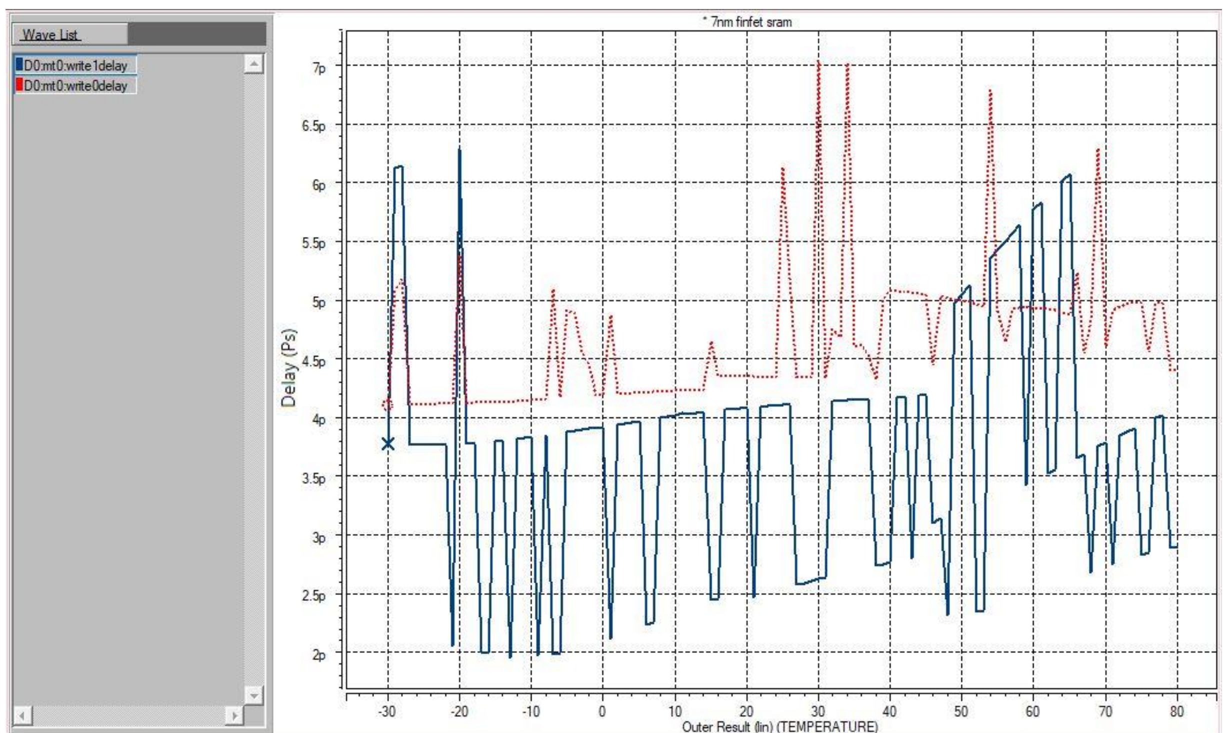


Figure 4.20 The delay curves under temperature variations for 7nm standard cell

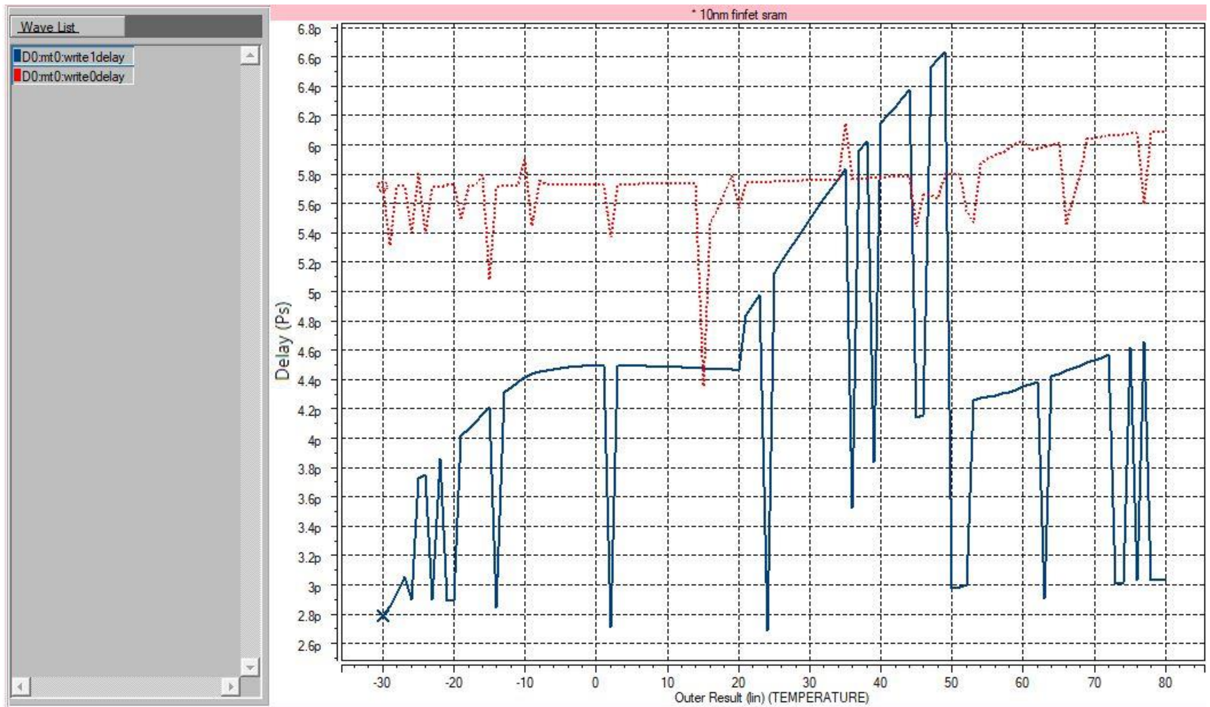


Figure 4.21 The delay curves under temperature variations for 10nm standard cell

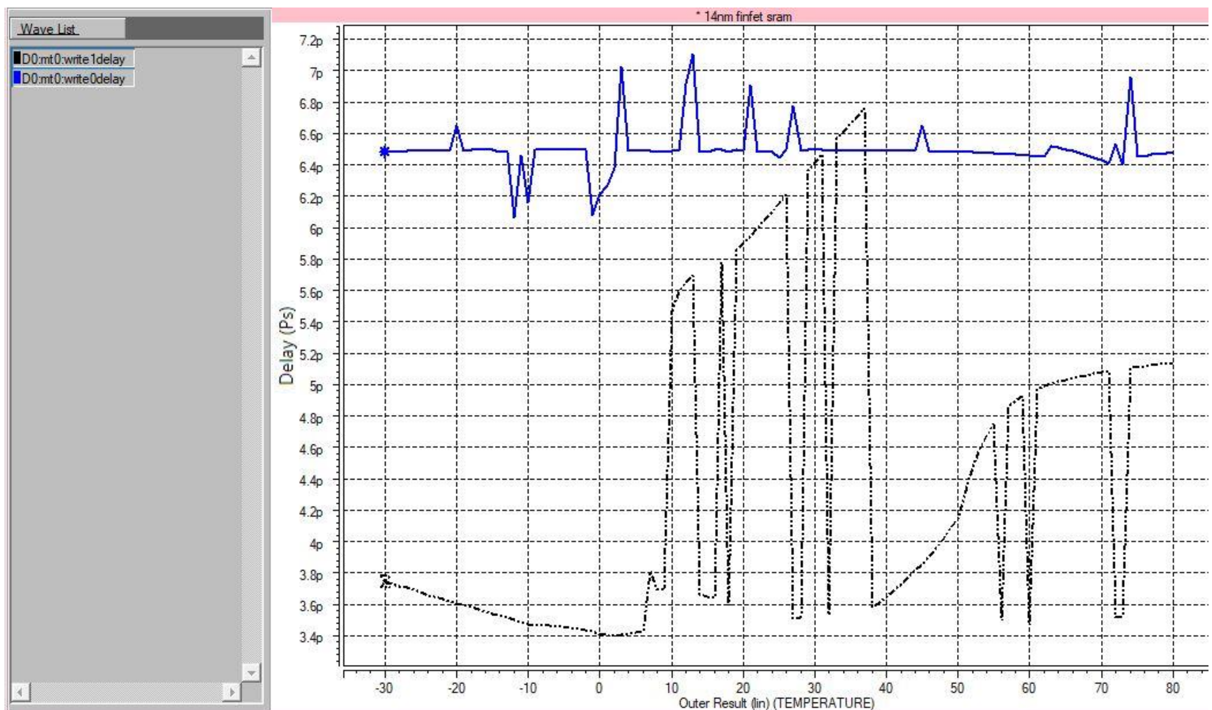


Figure 4.22 The delay curves under temperature variations for 14nm standard cell

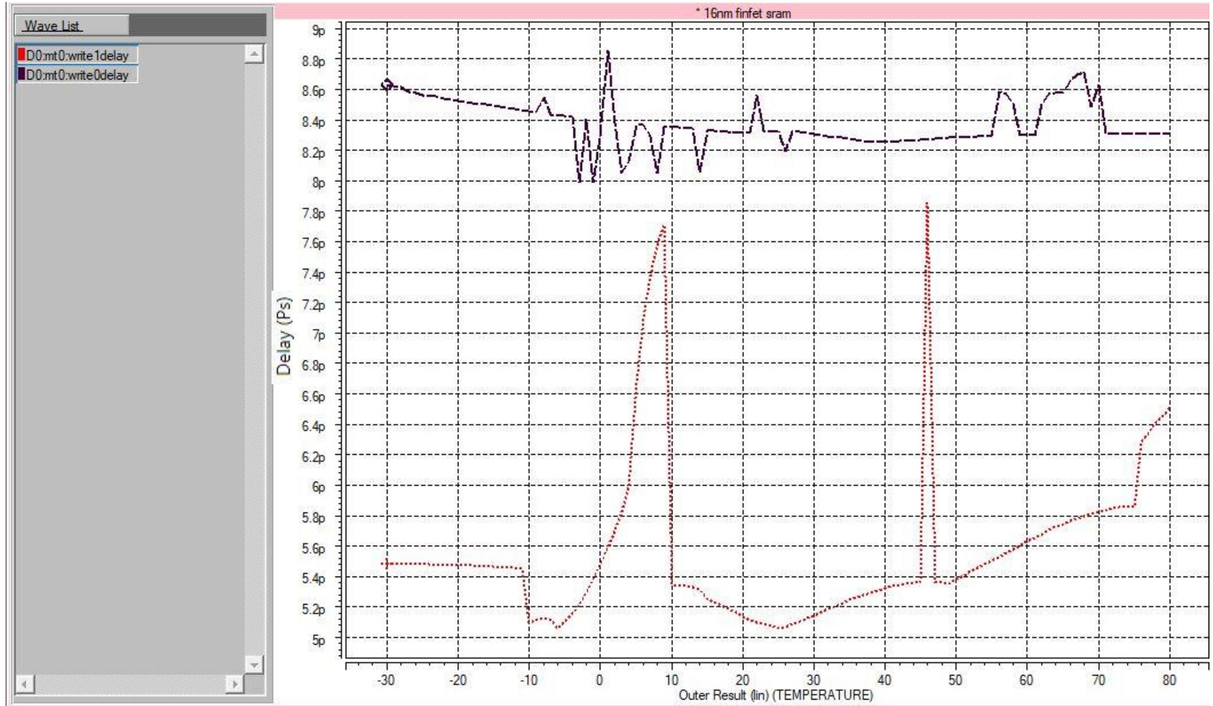


Figure 4.23 The delay curves under temperature variations for 16nm standard cell

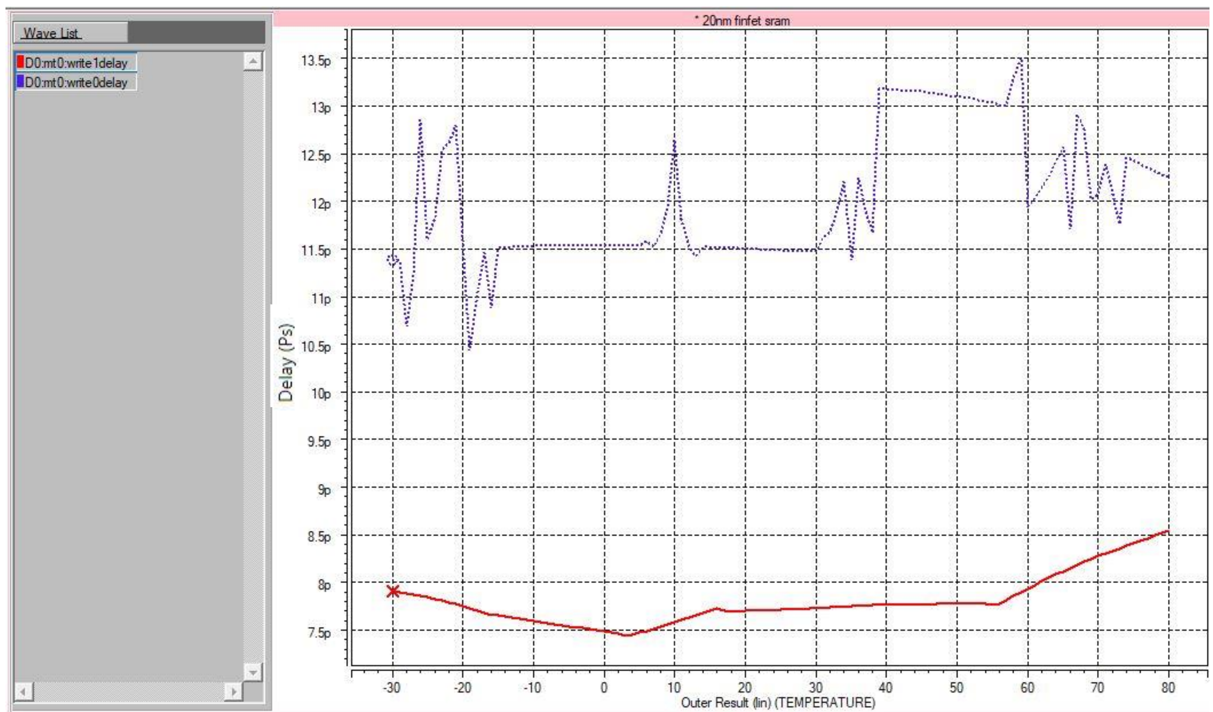


Figure 4.24 The delay curves under temperature variations for 20nm standard cell

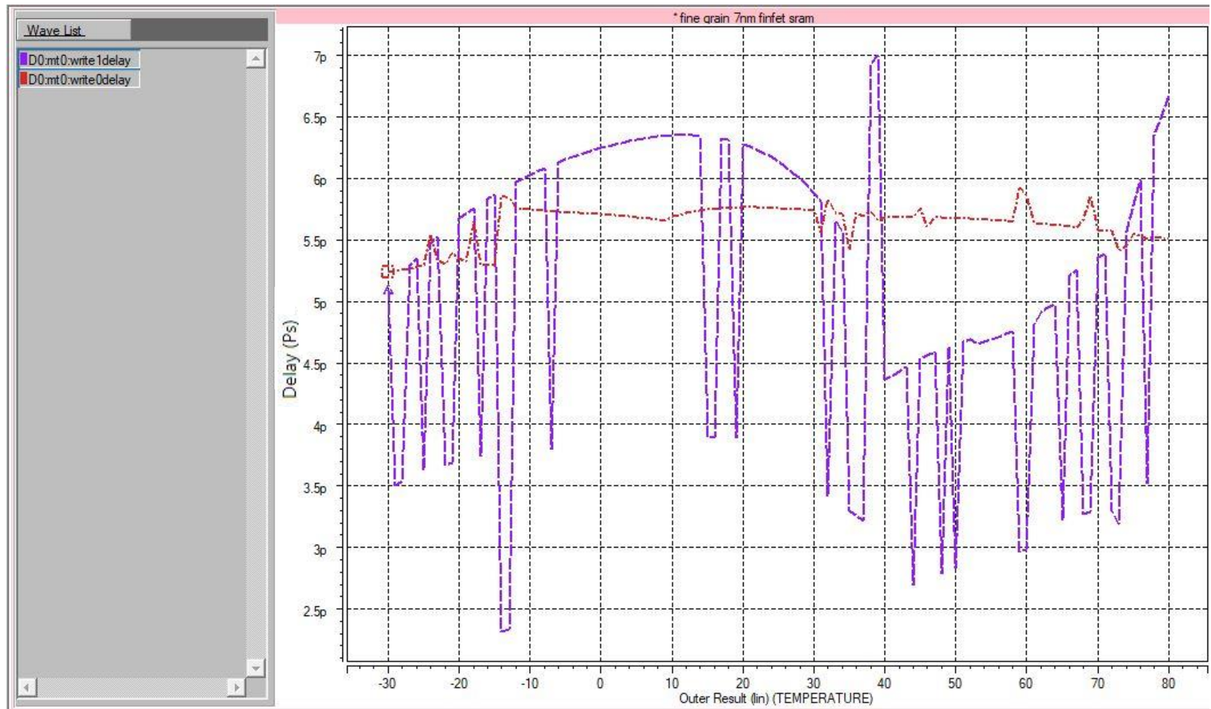


Figure 4.25 The delay curves under temperature variations for 7nm fine grain cell

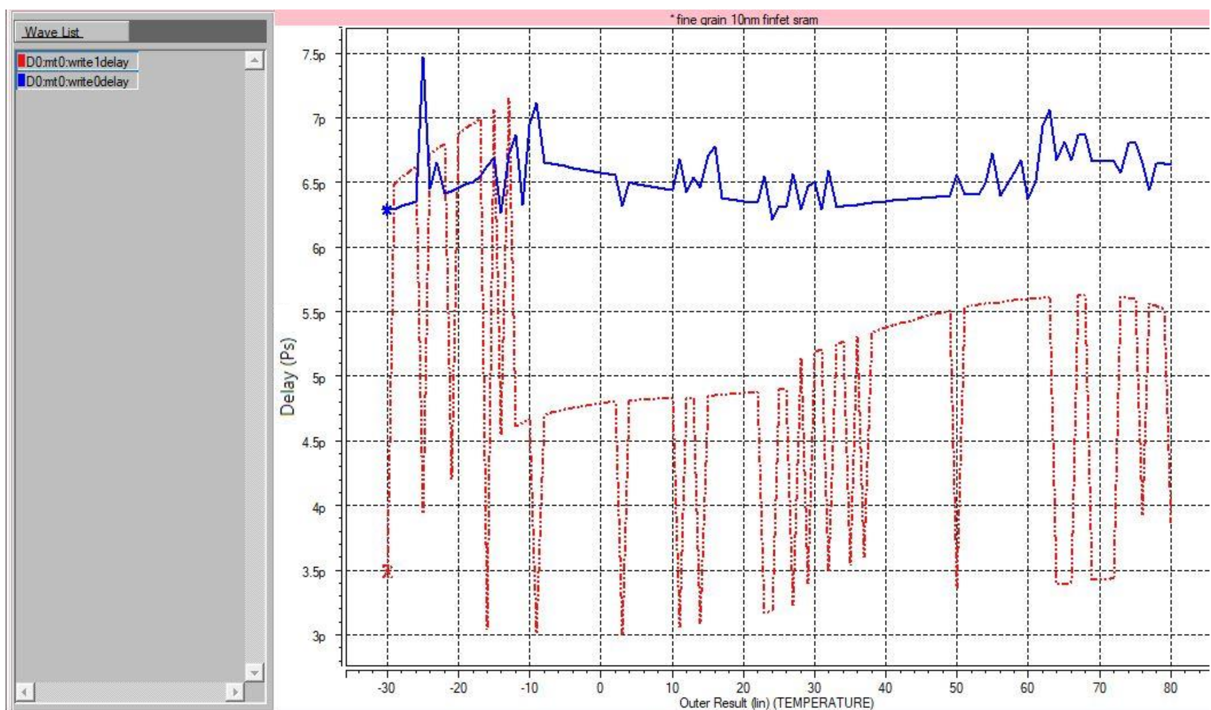


Figure 4.26 The delay curves under temperature variations for 10nm fine grain cell

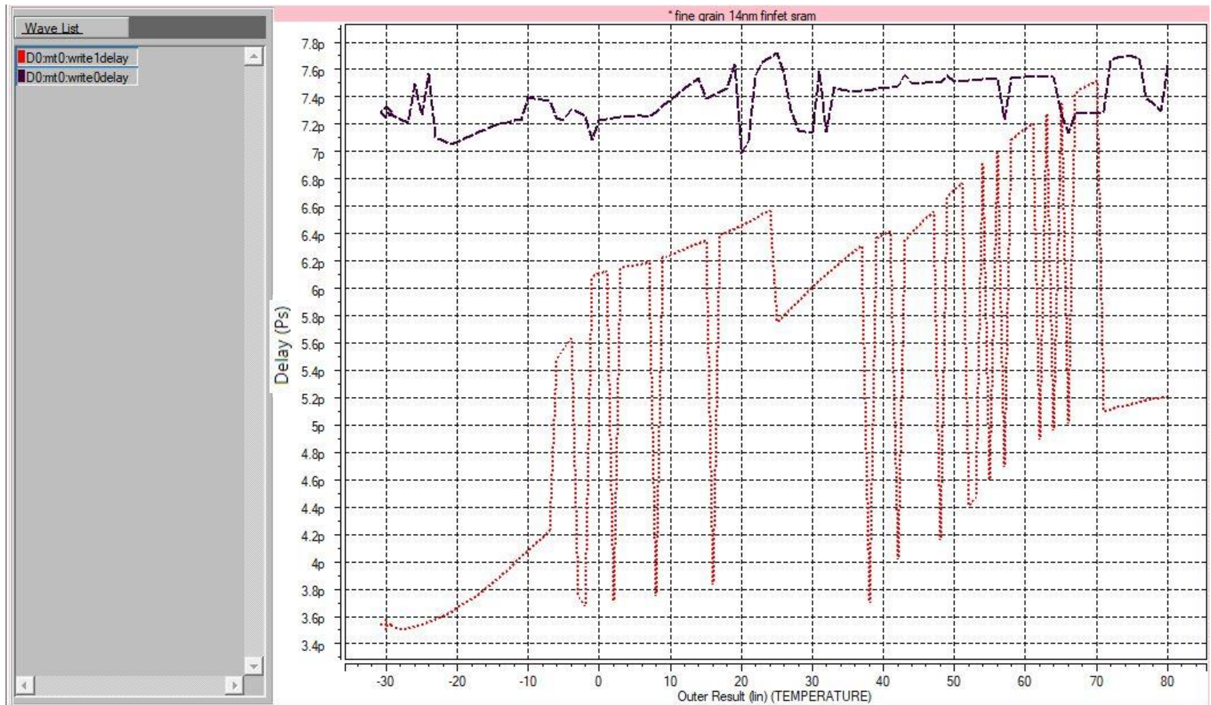


Figure 4.27 The delay curves under temperature variations for 14nm fine grain cell

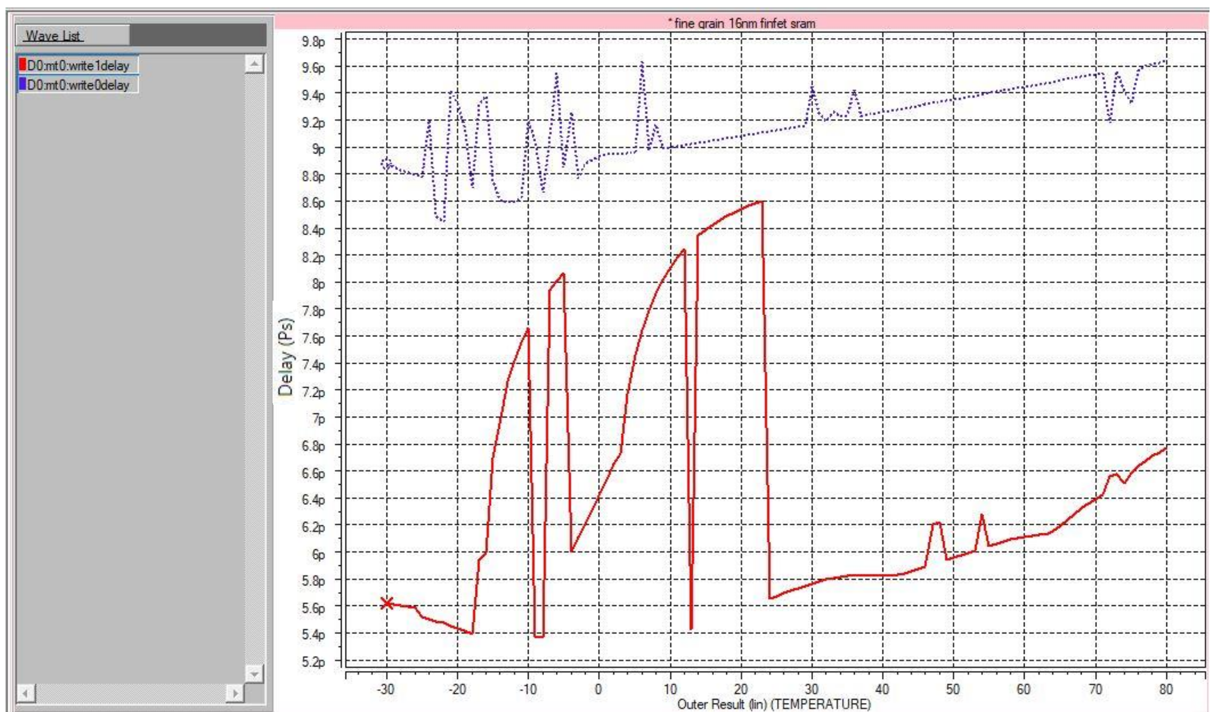


Figure 4.28 The delay curves under temperature variations for 16nm fine grain cell

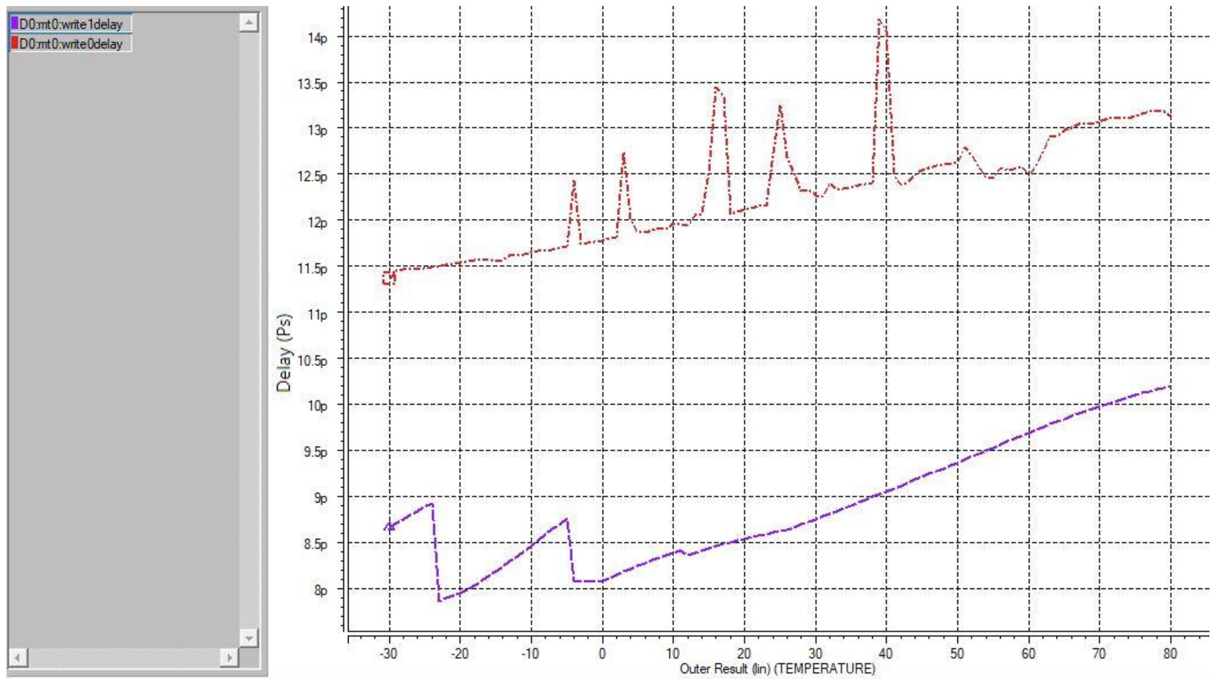


Figure 4.29 The delay curves under temperature variations for 20nm fine grain cell

To get better insights into the effects on delay caused by the temperature, average delay time at each operating temperature was calculated and difference between the consecutive average delays gave us the deviation at that point to the next. These deviations were then averaged to give us a parameter, i.e. average change in average delay per degree centigrade. This parameter can be used to compare the performances of all the SRAM models in the simulation.

Table 4.6 Average change in delay time and dynamic power per degree Celsius increase in temperature

	Average change in average delay per degree centigrade (ps)	Average change in dynamic power per degree centigrade (nW)
7nm standard	0.500827273	4.15182
7nm FineGrain	0.387318182	3.73727
10nm standard	0.259572727	5.88455
10nm FineGrain	0.390722727	4.19909
14nm standard	0.199254545	6.28000
14nm FineGrain	0.345895455	4.43727
16nm standard	0.080909091	3.69000
16nm FineGrain	0.145345455	2.95090
20nm standard	0.114040909	4.23455
20nm FineGrain	0.078095455	3.05272

It was observed that the change in average delay is more in the fine grain cell than their standard cell counter parts excepts for the 7nm and 20nm technology node where it is lesser than their standard cell counterpart. This, however doesn't not suggest that the fine grain cells are faster or slower. It just suggests the resilience of the cell with respect to the temperature change. In our simulation, it appears that 20nm Fine Grain cell is most resilient of the lot followed by the 16nm standard cell. Dynamic

power, however, follows the reverse trend throughout, that is change in average dynamic power is lesser for the fine grain cells than their counter parts. This is same as the case with no temperature fluctuations. The following graph very effectively shows the average power distribution comparisons between the standard cells and their fine grain counterparts.

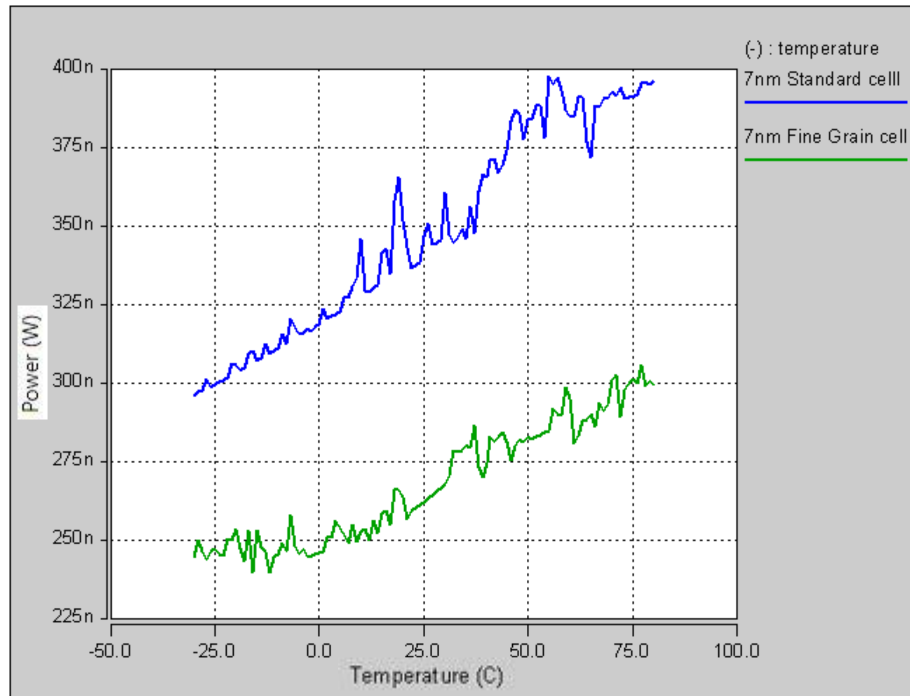


Figure 4.30 Average dynamic power comparison at 7nm node for different cell structure at varying temperature.

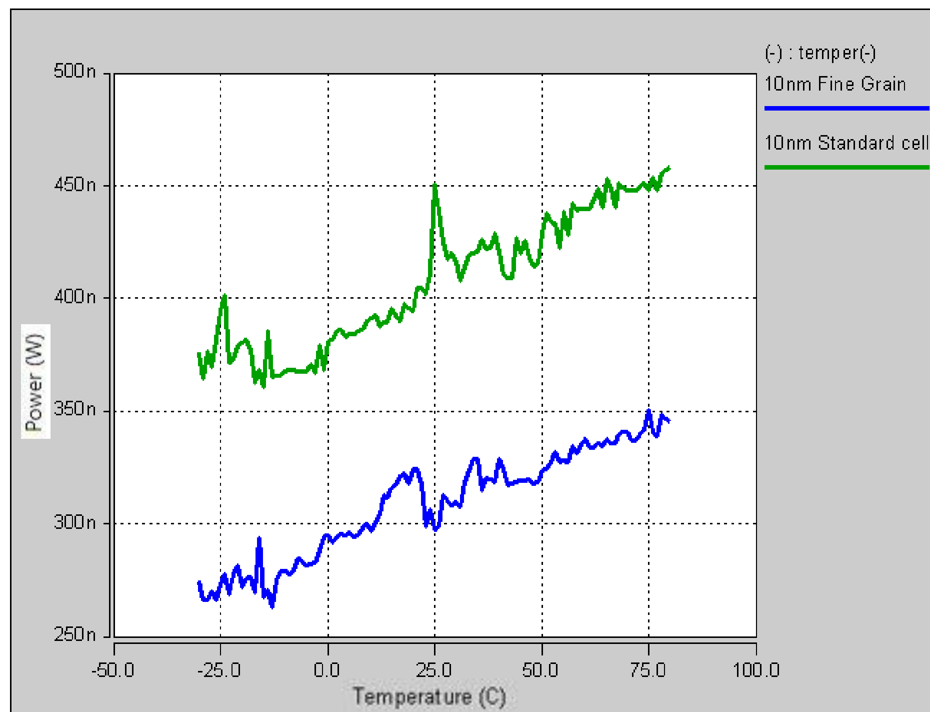


Figure 4.31 Average dynamic power comparison at 10nm node for different cell structure at varying temperature.

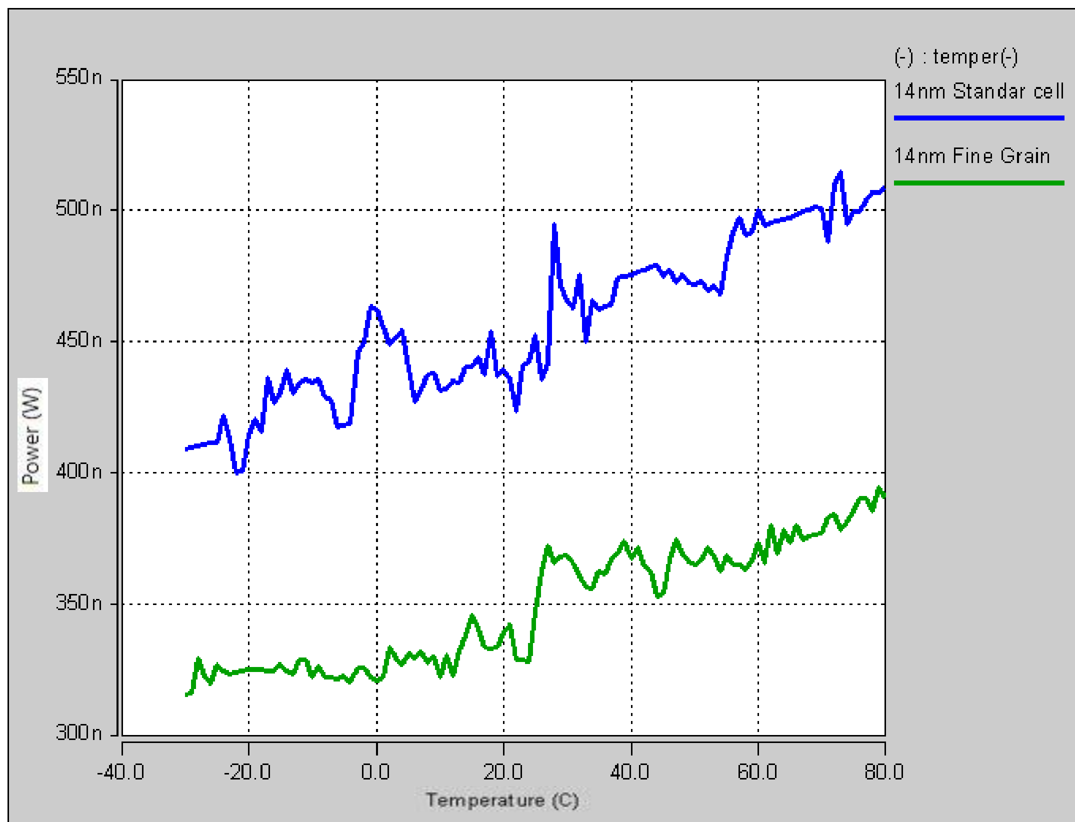


Figure 4.32 Average dynamic power comparison at 14nm node for different cell structure at varying temperature.

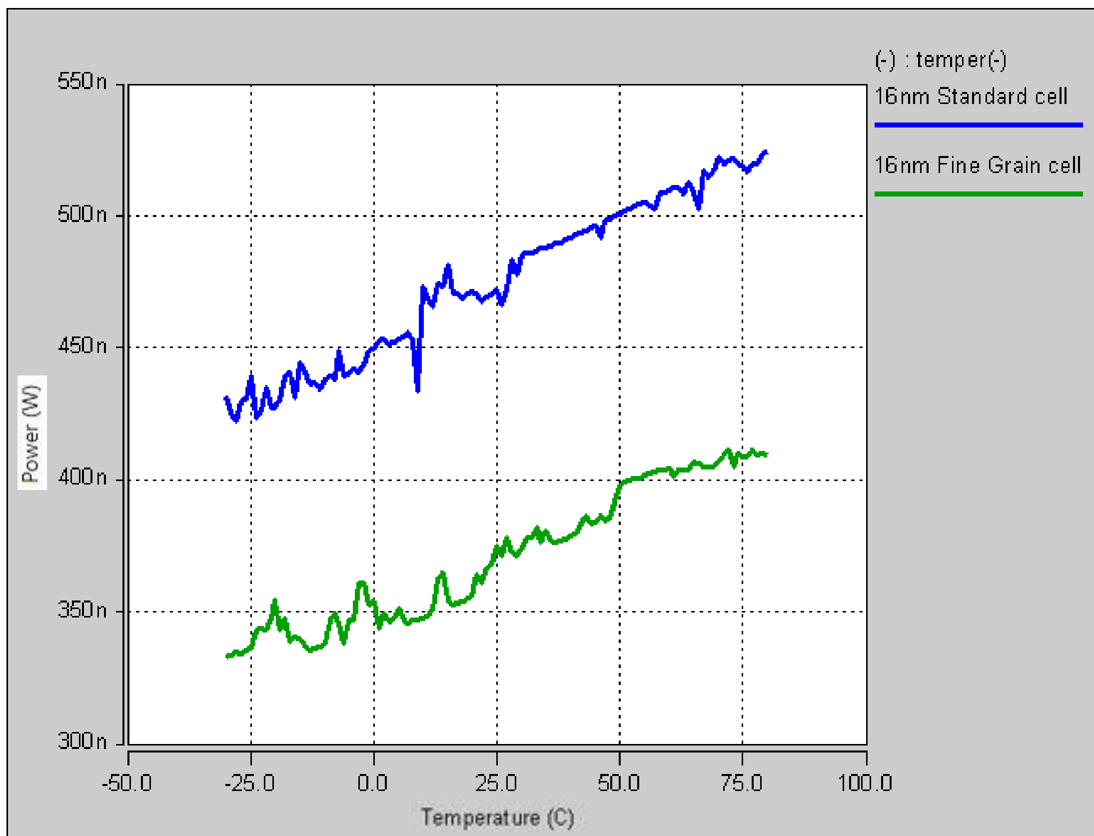


Figure 4.33 Average dynamic power comparison at 16nm node for different cell structure at varying temperature.

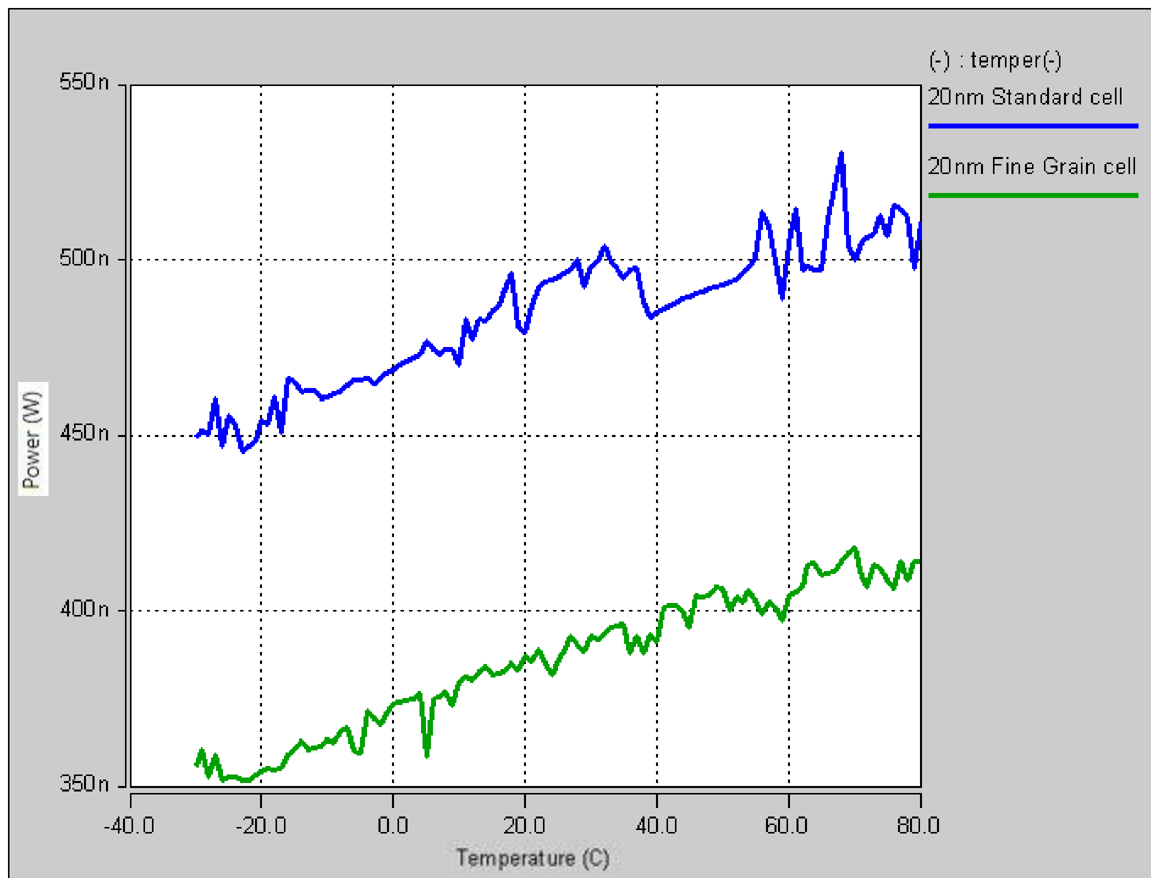


Figure 4.34 Average dynamic power comparison at 20nm node for different cell structure at varying temperature

The simulations carried out in the previous sections have already established the fact that the power gated version of the standard 6T SRAM cell consumes lesser power. The same fact can be seen throughout the temperature range for 5 different technology nodes. The average power curves never intersect each other proves the fact mentioned earlier.

Apart from the comparisons between cell structures and technology nodes, it is observed that the power dissipation follows a rising trend for each cell structure regardless of the technology nodes which isn't surprising.

Later on, the effect of temperature on the static noise margin (SNM) was simulated on the standard 6T SRAM cells for different technology nodes. The simulation results show a steady decline in the noise margin as the temperature increases, the highest SNM is recorded at the minimum temperature of the simulation test range. Regardless, of the technology node, the SNM seems to worsen with each degree increase in the temperature. This phenomenon can be observed in the real world also. While overclocking the CPUs, in certain applications, the CPUs can be overclocked to well beyond their physical limits of 4-5 GHz, at these speeds, these CPUs then generate too much heat and cannot provide stable outputs. To stabilise these CPUs, specialised cooling systems, like liquid nitrogen, are used to bring the temperature down to negative 100 degrees Celsius to achieve the desired performance.

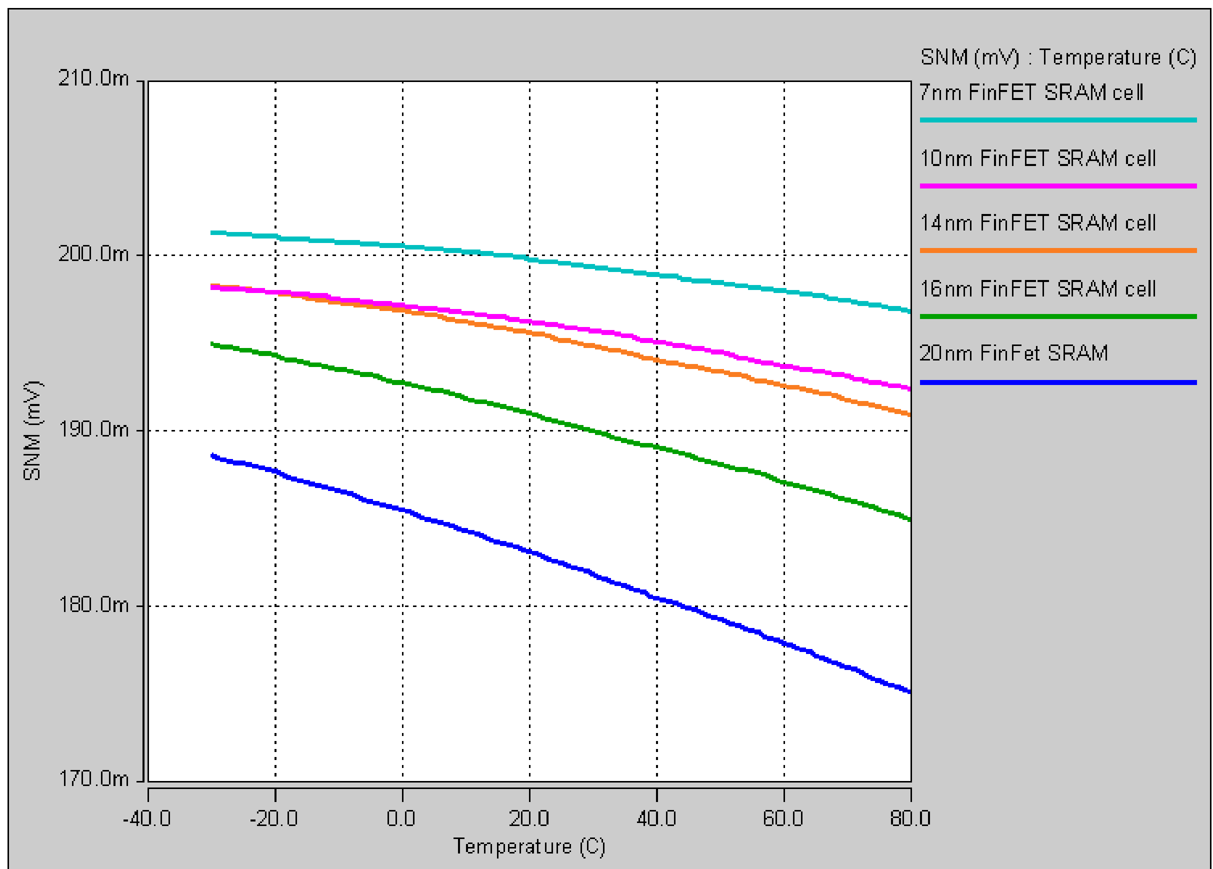


Figure 4.35 Effect of temperature on the SNM of SRAM cell for different technology nodes

We also suffer this in normal consumer electronics, where a device may heat up with excessive use and later starts to slow down and may even stop responding sometimes.

The results again show the experimental proof that the lower technology nodes have improved noise margins. It is also observed that the slope of the SNM curve is steepest for the 20nm FinFET and least for the 7nm FinFET which translates to better stability under variable temperature environment.

.

CHAPTER 5
CONCLUDING REMARKS AND FUTURE SCOPE

5.1. Conclusion

This thesis report discussed different static random access memories (SRAMs) and their performance metrics like Static Noise Margin (SNM), Read Noise Margin (RNM) and Write Noise Margin (WNM). The differences between conventional bulk-MOSFET based SRAM and DGFET or FinFET based SRAM have also been discussed in extensive detail. The conventional CMOS cells are not effortlessly scalable in sub 50 nm node due to numerous considerations like sub-threshold leakage current and process variation effects that harshly reduce cell stability. FinFET based SRAM design has great scalability with lower leakage current as equated to MOSFET based SRAM with trifling process variation effects.

This thesis focused on the simulation of a standard 6T FinFET SRAM cell at different technology nodes, comparing them on the basis of various parameters and successfully established an experimental proof to the theory discussed in the earlier chapters of the thesis. The SRAM cell is then power gated and compared for the same parameters.

The simulation data confirms that the needs for the scaling down of the transistor for SRAM uses both in terms of area and performance. Highest SNM and lowest PDP of 7nm FinFET SRAM is the proof for stability and efficiency combined with the lowest delay makes it ideal for embedded applications. The point to be noted is that the model used are predictive technology models (PTM) which are based on the IRTS roadmap, hence this simulation is a proof of a concept. Semiconductor industry is already its way past the 10nm FinFET process and we can expect 7nm FinFET devices by late 2018.

5.2. Future Scope

All the simulations were done taken into account a cell ratio, CR=1, which isn't all that great. Fine tuning the CR can yield us better results and possibly a better SRAM cell. As observed, the power dissipation can be dramatically reduced by employing gating techniques. A further in-depth study can improve the standard 6T SRAM cell.

SRAM chips are widely used in different consumer electronics. With increasing demand for low power mobile devices and emerging IoT devices, the need for scaling down the SRAM cells is a major area of development. With limitations in scaling conventional bulk-MOSFET devices, FinFET SRAM scaling has emerged as a promising field for further research and development. The emerging technology of internet of thing devices and wireless sensor networks pushes the operating environment for electronics devices to extreme. These extreme environment applications compel designer to research and develop devices, which can perform in such condition without performance degradation.

In the recent times, we have seen FinFET been scaled down to 10nm processes. A latest smartphone has a processors based on 10nm FinFETs and as discussed earlier, the production of next generation

7nm devices have already started, which are scheduled to reach the common consumer by late 2018. There have been talks and road maps to further scale down the FinFET to 5nm, keeping the Moore's law relevant. This will create a new generation of ultra-low power, high-performing devices with far greater transistor density than the current generation.

REFERENCES

[1] Y. Yang, H. Jeong, S. C. Song, J. Wang, G. Yeap and S. O. Jung, "Single Bit-Line 7T SRAM Cell for Near-Threshold Voltage Operation With Enhanced Performance and Energy in 14 nm FinFET Technology," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 7, pp. 1023-1032, July 2016.

[2] R. W. Mann *et al.*, "Array Termination Impacts in Advanced SRAM," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 9, pp. 2449-2457, Sept. 2017.

[3] T. W. Oh, H. Jeong, K. Kang, J. Park, Y. Yang and S. O. Jung, "Power-Gated 9T SRAM Cell for Low-Energy Operation," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 3, pp. 1183-1187, March 2017.

[4] S. Kurude, S. Mittal and U. Ganguly, "Statistical Variability Analysis of SRAM Cell for Emerging Transistor Technologies," in *IEEE Transactions on Electron Devices*, vol. 63, no. 9, pp. 3514-3520, Sept. 2016.

[5] A. G. Akkala, R. Venkatesan, A. Raghunathan and K. Roy, "Asymmetric Underlapped Sub-10-nm n-FinFETs for High-Speed and Low-Leakage 6T SRAMs," in *IEEE Transactions on Electron Devices*, vol. 63, no. 3, pp. 1034-1040, March 2016.

[6] N. Agrawal, H. Liu, R. Arghavani, V. Narayanan and S. Datta, "Impact of Variation in Nanoscale Silicon and Non-Silicon FinFETs and Tunnel FETs on Device and SRAM Performance," in *IEEE Transactions on Electron Devices*, vol. 62, no. 6, pp. 1691-1697, June 2015.

[7] Y. Yang, H. Jeong, S. C. Song, J. Wang, G. Yeap and S. O. Jung, "Single Bit-Line 7T SRAM Cell for Near-Threshold Voltage Operation With Enhanced Performance and Energy in 14 nm FinFET Technology," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 7, pp. 1023-1032, July 2016.

[8] D. Kraak *et al.*, "Degradation analysis of high performance 14nm FinFET SRAM," *2018 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Dresden, 2018, pp. 201-206.

[10] T. Song *et al.*, "A 7nm FinFET SRAM using EUV lithography with dual write-driver-assist circuitry for low-voltage applications," *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, San Francisco, CA, 2018, pp. 198-200.

[11] M. N. Kishor and S. S. Narkhede, "Design of a ternary FinFET SRAM cell," *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, Indore, 2016, pp. 1-5.

[12] Computer organization. (4th ed.). [S.l.]: *McGraw-Hill*. ISBN 0-07-114323-8.

[13] International Technology Roadmap for Semiconductors (ITRS). San Jose, CA: *Semiconductor Industry Association*, 2007.

[14] B. Yu, H. Wang, A. Joshi, Q. Xiang, E. Ibok, and M. R. Lin, “15 nm gate length planar CMOS transistor,” in *IEDM Tech. Dig., 2001*, pp. 937–939.

[15] D. Hisamoto, W. C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T. J. King, J. Bokor, and C. Hu, “FinFET—A self-aligned double-gate MOSFET scalable to 20 nm,” *IEEE Trans. Electron Devices*, vol. 47, pp. 2320–2325, 2000.

[16] J. Y. S. Balasubramanium, “Design of sub-50 nm FinFET based low power SRAMs,” *Semiconductor Science Technology*, vol. 23, p. 13, 2008.

[17] K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, D. Murray, N. Vallepalli, Y. Wang, B. Zheng, and M. Bohr, “A 3-GHz 70 MB SRAM in 65nm CMOS technology with integrated column-based dynamic power supply,” in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2005, pp. 474–476.

[18] B. Raj, A. K. Saxena and S. Dasgupta, "Nanoscale FinFET Based SRAM Cell Design: Analysis of Performance Metric, Process Variation, Underlapped FinFET, and Temperature Effect," in *IEEE Circuits and Systems Magazine*, vol. 11, no. 3, pp. 38-50, thirdquarter 2011.

[19] Colinge, J. (2011). *FinFETs and other multi-gate transistors*. New York: Springer.

[20] E. Chin, M. Dunga, and B. Nikolic, “Design trade-offs of a 6T FinFET SRAM cell in the presence of variations,” in *Proc. IEEE Symp. VLSI Circuits*, *2006*, pp. 445–449.

[21] F. Sheikh and V. Varadarajan, “The impact of device-width quantization on digital circuit design using FinFET structures,” in *Proc. EE241 Spring, 2004*, pp. 1–6.

[22] P. T. Su, C. H. Jin, C. J. Dong, H. S. Yeon, P. Donggun, K. Kinam, E. Yoon, and L. J. Ho, “Characteristics of the full CMOS SRAM cell using body tied TG MOSFETs (bulk FinFETs),” *IEEE Trans. Electron Devices*, vol. 53, pp. 481–487, 2006.

[23] K. Itoh, K. Sasaki, and Y. Nakagome, “Trends in low power RAM circuit technology,” in *Proc. IEEE IEDM Tech. Dig., Apr. 1995*, pp. 524–543.

[24] L. Bagheriye, R. Saeidi and S. Toofan, "Low power and roboust FinFET SRAM cell using independent gate control," *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, Montreal, QC, 2016, pp. 49-52.

[25] Rashmi, A. Kranti, and G. A. Armstrong, “6-T SRAM cell design with nanoscale double-gate SOI MOSFETs: Impact of source/drain engineering and circuit topology,” *Semiconductor Science Technology*, p. 13, 2008.

[26] A. J. Bhavnagarwala, T. Xinghai, and J. D. Meindl, “The impact of intrinsic device fl uctuations on CMOS SRAM cell stability*,” IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 658–665, Apr. 2001.

[27] B. D. Yang and L. S. Kim, “A low-power SRAM using hierarchical bit line and local sense amplifiers*,” IEEE Journal of Solid-State Circuits*, vol. 40, no. 6, pp. 1366–1376, 2005.

[28] H. Qin, Y. Cao, D. Markovic, A. Vladimirescu, and J. Rabaey, “SRAM leakage suppression by minimizing standby supply voltage,” in *Proc. 5th Int. Symp. Quality Electronic Design, 2004*, pp. 55–60.

[29] J. H. Choi, A. Bansal, M. Meterelliyoz, J. Murthy, and K. Roy, "Leakage power dependent temperature estimation to predict thermal runaway in FinFET circuits," in *IEEE Proc. Int. Conf. Computer Aided Design (ICCAD)*, Nov. 5–9, 2006, pp. 583–586.

[30] T. Miwa et al., "*A 512 Kbit low-voltage NV-SRAM with the size of a conventional SRAM*," 2001 Symposium on VLSI Circuits. Digest of Technical Papers (IEEE Cat. No.01CH37185), Kyoto, Japan, 2001, pp. 129-132.

[31] K. Itoh, K. Sasaki, and Y. Nakagome, "Trends in low power RAM circuit technology," in *Proc. IEEE IEDM Tech. Dig*., Apr. 1995, pp. 524–543.

[32] A. Dixit, A. Kottantharayil, N. Collaert, M. Goodwin, M. Jurczak, and K. D. Meyer, "Analysis of the parasitic S/D resistance in multiple-gate FET," *IEEE Trans. Electron Devices*, vol. 52, no. 6, pp. 1132–1139, June 2005.

[33] D. Anandani, A. Kumar and V. S. K. Bhaaskaran, "Gating techniques for 6T SRAM cell using different modes of FinFET," *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Kochi, 2015, pp. 483-487.

[34] Seevinck, E., List, F. J., and Lohstroh, J., "Statis-Noise Margin Analysis of MOS SRAM Cells,", *IEEE Journal of Solid-State Circuits*, Vol. SC-22, No. 5, October, 1987.

[35] R. Yarmand, B. Ebrahimi, H. Afzali-Kusha, A. Afzali-Kusha and M. Pedram, "High-performance and high-yield 5 nm underlapped FinFET SRAM design using P-type access transistors," *Sixteenth International Symposium on Quality Electronic Design*, Santa Clara, CA, 2015, pp. 10-17.

[36] X. Zhang *et al*., "Analysis of 7/8-nm Bulk-Si FinFET Technologies for 6T-SRAM Scaling," in *IEEE Transactions on Electron Devices*, vol. 63, no. 4, pp. 1502-1507, April 2016.

[37] M. Ichihashi, Y. Woo and S. Parihar, "SRAM cell performance analysis beyond 10-nm FinFET technology," *2016 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA)*, Hsinchu, 2016, pp. 1-2.

[38] A. Bhavnagarwala, I. Iqbal, An Nguyen, D. Ondricek, V. Chandra and R. Aitken, "A 400mV active VMIN, 200mV retention VMIN, 2.8 GHz 64Kb SRAM with a 0.09 um^2 6T bitcell in a 16nm FinFET CMOS process," *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, Honolulu, HI, 2016, pp. 1-2.

[39] X. Wang *et al*., "Process informed accurate compact modelling of 14-nm FinFET variability and application to statistical 6T-SRAM simulations," *2016 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Nuremberg, 2016, pp. 303-306.

[40] M. L. Fan, Y. S. Wu, V. P. H. Hu, C. Y. Hsieh, P. Su and C. T. Chuang, "Comparison of 4T and 6T FinFET SRAM Cells for Subthreshold Operation Considering Variability—A Model-Based Approach," in *IEEE Transactions on Electron Devices*, vol. 58, no. 3, pp. 609-616, March 2011.

[41] V. P. H. Hu, M. L. Fan, C. Y. Hsieh, P. Su and C. T. Chuang, "FinFET SRAM Cell Optimization Considering Temporal Variability Due to NBTI/PBTI, Surface Orientation and Various Gate Dielectrics," in *IEEE Transactions on Electron Devices*, vol. 58, no. 3, pp. 805-811, March 2011.

[42] D. D. Lu, C. H. Lin, A. M. Niknejad and C. Hu, "Compact Modeling of Variation in FinFET SRAM Cells," in *IEEE Design & Test of Computers*, vol. 27, no. 2, pp. 44-50, March-April 2010.

[43] P. Zheng, Y. B. Liao, N. Damrongplasit, M. H. Chiang and T. J. K. Liu, "Variation-Aware Comparative Study of 10-nm GAA Versus FinFET 6-T SRAM Performance and Yield," in *IEEE Transactions on Electron Devices*, vol. 61, no. 12, pp. 3949-3954, Dec. 2014.

[44] Y. Li, H. W. Cheng and M. H. Han, "Statistical Simulation of Static Noise Margin Variability in Static Random Access Memory," in *IEEE Transactions on Semiconductor Manufacturing*, vol. 23, no. 4, pp. 509-516, Nov. 2010.

[45] H. Villacorta, V. Champac, S. Bota and J. Segura, "FinFET SRAM hardening through design and technology parameters considering process variations," *2013 14th European Conference on Radiation and Its Effects on Components and Systems (RADECS)*, Oxford, 2013, pp. 1-7.

[46] H. Villacorta, J. Segura, S. Bota and V. Champac, "Analysis of fin height on FinFET SRAM cell hardening," *2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS)*, College Station, TX, 2014, pp. 671-674.

[47] K. Kang, H. Jeong, J. Lee and S. O. Jung, "Comparative analysis of 1:1:2 and 1:2:2 FinFET SRAM bit-cell using assist circuit," *2013 International SoC Design Conference (ISOCC)*, Busan, 2013, pp. 035-038.

[48] X. Wang *et al.*, "Impact of statistical variability and charge trapping on 14 nm SOI FinFET SRAM cell stability," *2013 Proceedings of the European Solid-State Device Research Conference (ESSDERC)*, Bucharest, 2013, pp. 234-237.

[49] S. M. Salahuddin, Hailong Jiao and V. Kursun, "Characterization of FinFET SRAM cells with asymmetrically gate underlapped bitline access transistors under process parameter fluctuations," *2013 IEEE International Conference of Electron Devices and Solid-state Circuits*, Hong Kong, 2013, pp. 1-2.

[50] M. A. Turi and J. G. Delgado-Frias, "Performance-power tradeoffs of 8T FinFET SRAM cells," *2011 IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Seoul, 2011, pp. 1-4.

[51] K. Endo, S. i. O'uchi, T. Matsukawa, Y. Liu and M. Masahara, "Independent double-gate FinFET SRAM technology," *The 4th IEEE International NanoElectronics Conference*, Tao-Yuan, 2011, pp. 1-2.

This is a preview of the print version of your report. Please click "print" to continue or "done" to close this window.

print print done

Turnitin Originality Report

Thesis by Simranjit Singh 801661024

From papers (kss)

- Processed on 09-Aug-2018 16:15 +0530
- ID: 988690434
- Word Count: 13616

Similarity Index
9%
Similarity by Source

Internet Sources:
4%
Publications:
7%
Student Papers:
2%

sources:
1

1% match (Internet from 26-Apr-2010)

http://www.cecs.uci.edu/~papers/iccad06/papers/7D_3.pdf
2

1% match (publications)

Raj, Balwinder, A. Saxena, and S. Dasgupta. "Nanoscale FinFET Based SRAM Cell Design: Analysis of Performance Metric, Process Variation, Underlapped FinFET, and Temperature Effect", IEEE Circuits and Systems Magazine, 2011.
3

1% match (Internet from 27-Sep-2016)

https://www.scribd.com/doc/139367653/Embedded-Systems-Theory-and-Design-Methodology
4

< 1% match (student papers from 05-Oct-2017)

Submitted to National Institute of Technology, Hamirpur on 2017-10-05
5

< 1% match (publications)

Byung-Do Yang, Lee-Sup Kim. "A low-power SRAM using hierarchical bit line and local sense amplifiers", IEEE Journal of Solid-State Circuits, 2005
6

< 1% match (publications)

A. Dixit. "Analysis of the Parasitic S/D Resistance in Multiple-Gate FETs", IEEE Transactions on Electron Devices, 6/2005

7

< 1% match (publications)

T. Miwa, J. Yamada, H. Koike, T. Nakura, T. Kobayashi, N. Kasai, H. Toyoshima. "A 512 Kbit low-voltage NV-SRAM with the size of a conventional SRAM", 2001 Symposium on VLSI Circuits. Digest of Technical Papers (IEEE Cat. No.01CH37185), 2001

8

< 1% match (publications)

Jain, S., K. Santhosh, M. Pattanaik, and B. Raj. "A 10-T SRAM cell with inbuilt charge sharing for dynamic power reduction", 2013 International Conference on Advances in Technology and Engineering (ICATE), 2013.

9

< 1% match (publications)

Chenming Hu. "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm", IEEE Transactions on Electron Devices, 2000

10

< 1% match (Internet from 07-Nov-2017)

http://digitalassets.lib.berkeley.edu/etd/ucb/text/Lu_berkeley_0028E_11629.pdf

11

< 1% match (publications)

A.J. Bhavnagarwala. "The impact of intrinsic device fluctuations on CMOS SRAM cell stability", IEEE Journal of Solid-State Circuits, 4/2001

12

< 1% match (Internet from 04-Nov-2017)

http://www.ijcaonline.org/archives/volume171/number6/hajare-2017-ijca-915005.pdf

13

< 1% match (Internet from 22-Dec-2012)

http://dspace.thapar.edu:8080/dspace/bitstream/10266/1816/1/Final_thesis.pdf

14

< 1% match (student papers from 22-Jul-2018)

Submitted to Symbiosis International University on 2018-07-22

15

< 1% match (publications)

"IEEE-ICDCS conference proceeding", 2012 International Conference on Devices Circuits and Systems (ICDCS), 03/2012

16

< 1% match (publications)

Kaushik Roy. "Thermal analysis of 8-T SRAM for nano-scaled technologies", Proceeding of the thirteenth international symposium on Low power electronics and design - ISLPED 08 ISLPED 08, 2008