

Modeling and Optimization Techniques for Yield-Aware SRAM Pre-Silicon Tuning

Major Project Report

Submitted in partial fulfilment of the requirements

for the degree of

Master of Technology

in

Electronics & Communication Engineering

(VLSI Design)

By

Yash Nalaria

(16MECV15)



Electronics & Communication Engineering Program

Department of Electrical Engineering Institute of

Technology,

Nirma University, Ahmedabad – 382481

May-2018

Modeling and Optimization Techniques for Yield-Aware SRAM Pre-Silicon Tuning

Major Project Report

Submitted in partial fulfilment of the requirements

for the degree of

Master of Technology

in

Electronics & Communication Engineering
(VLSI Design)

By

Yash Nagaria

(16MECV15)

Under the guidance of

External Project Guide:

Mr. Shailendra Sharad

R&D Manager,

Synopsys India Pvt. Ltd.

Internal Project Guide:

DR. N. M. Devashrayee

(PG Coordinator (VLSI))

Nirma University.



Electronics & Communication Engineering Program

Department of Electrical Engineering Institute of

Technology,

Nirma University, Ahmedabad - 382481

May-2018

Declaration

This is to certify that

1. The thesis comprises my original work towards the degree of Master of Technology in VLSI Design at Nirma University and has not been submitted elsewhere for a degree.
2. Due acknowledgment has been made in the text to all other material used.

-Yash Nagaria
16MECV15



Certificate

This is to certify that the Major Project entitled “**Modeling and Optimization Techniques for Yield-Aware RAM Pre-Silicon Tuning**” submitted by **Yash Nagaria (16MECV15)**, towards the partial fulfilment of the requirements for the degree of Master of Technology in VLSI Design, Nirma University, Ahmedabad is the record of work carried out by him under our supervision and guidance. In our opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project, to the best of our knowledge, haven’t been submitted to any other university or institution for award of any degree or diploma.

Internal Guide:

Dr. N. M. Devashrayee

(PG Coordinator EC (VLSI))

Program Co-ordinator:

Dr. N. M. Devashrayee

(PG Coordinator EC (VLSI))

Dr. D. K. Kothari

Head, EC Dept.

Dr. Alka Mahajan

Director, IT - NU.

Date:

Place: Ahmedabad

Acknowledgement

First and foremost, sincere thanks to **Dr. N. M. Devashrayee** (P.G. Coordinator of VLSI design, Institute of technology, Nirma University, Ahmedabad). I enjoyed his vast knowledge and thank him a lot for giving valuable support for project work. I would like to thank my manager **Mr. Shailendra Sharad** (manager II, R&D SG, Solution Group) and to my team specially **Yogesh Bhai Patel** (R&D Engineer, II SG, Solution Group), **Jigar Gandhi** (R&D Engineer, II SG, Solution Group) and **Ruchin Jain** (R&D Engineer, II SG, Solution Group) for providing necessary information regarding the project and for their constant guidance, supervision, kind co-operation, and invaluable support in all aspects.

I would like to thank **Dr. Usha Mehta** (Professor (EC), Institute of Technology, Nirma University, Ahmedabad) for her guidance and valuable support. I would like to thank my all faculty members for providing encouragement, exchanging knowledge.

I also owe my colleagues in the Synopsys, special thanks for helping me on this path and for making project at Synopsys more enjoyable.

- **Yash Nagaria**
16MECV15

Abstract

Today's mobile devices, SoC or any smart electronic device demand large memories with stringent power and speed constraints. Cutting-edge process nodes (16nm,12nm,10nm,7nm) are needed to meet target bit densities; but these nodes have extreme statistical process variation, hurting yield on these high-volume chips. To design for yield, one must be able to measure it in the design phase. This in turn requires estimation of Bitcell and sense amp yields, for which a simple Monte Carlo approach would need billions of simulations.

As size of SRAM is getting smaller, so it is particularly valuable to parametric failure, which reduces yield. The problem in SRAM is that there is clash between stability during read operation and ability during write operation, as we optimize the array Bitcell for read stability, write ability degrades. So my work represents the techniques which increases the performance of read and write operation of SRAM.

Timing optimization for every memory instance generated from memory compiler using self-timing circuit. In the initial phase of the memory compiler development, various kind of analysis is done on the models provided by the foundry to decide the behaviour of the models in terms of timing, power, and leakage.

Bit cell and logic are the two kinds of models which foundry provide for memory design. Bit cell design is provided by foundry and memory designers can do very small changes in the Bit cell design. Most of the effort in the memory design is in the periphery design which is made up of logic models and this is the main reason to do the logic model analysis. My work represents the analyse various margins like read margin, write margin, logic margin. Also all margins in aging PVT.

Synopsys Inc. At A Glance



- A world leader in providing the semiconductor solutions that help our customers improve quality of life for everyone, both today and in the future
- Among the world's largest semiconductor companies
- A leading EDA (Electronic Design Automation) company serving all electronics segments
- Key strengths in Memory Compilers, semiconductor intellectual property (IP), Synopsys' comprehensive, integrated portfolio of system-level, IP, implementation, verification, manufacturing, optical and field-programmable gate array (FPGA) and EDA
- Chairman and CEO: Aart J. de Geus
- Approximately 9,000 employees
- Approximately 81 support & research & development centres around the globe
- Corporate Headquarters Mountain View, California
- Global presence with sales offices all around the world
- Public since 1992 - shares traded on NASDAQ Stock Market (Nasdaq: SNPS)
- Founded in 1986 by Dr. Aart de Geus and a team of engineers from General Electric's Microelectronics Centre in Research Triangle Park, North Carolina

Group Introduction

Synopsys provides a broad portfolio of high-quality, silicon-proven embedded memory and logic library solutions, enabling system-on-chip (SoC) designers to lower integration risk and speed time-to-market. Synopsys provides a broad portfolio of high-quality, silicon-proven embedded memory and logic library solutions, enabling system-on-chip (SoC) designers to lower integration risk and speed time-to-market. The DesignWare Duet Packages of Embedded Memories and Logic Libraries include memory compilers, ROMs, standard cells, Power Optimization Kits (POKs) and optional overdrive/low voltage PVTs that enable designers to achieve the maximum performance with the lowest possible power consumption for their specific application. The High Performance Core (HPC) Design Kit contains a suite of high-speed and high-density memory instances and logic cells specifically designed to enable SoC designers to optimize their CPU, GPU and DSP cores for maximum speed, smallest area, lowest power, or an optimum balance of all three. In addition, the DesignWare STAR Memory System provides an integrated built-in self-test (BIST) and repair solution that improves test quality and manufacturing yield, while the DesignWare STAR Hierarchical System automates hierarchical testing for analog/mixed-signal IP, digital logic blocks and interface IP on an SoC. Synopsys also provides a comprehensive family of multiple-time programmable (MTP) and few-time programmable (FTP) non-volatile memory (NVM) IP in standard CMOS process technologies.

- **Memory Compilers:** DesignWare Memory Compilers are optimized for high performance and high density with advanced power management features. Integrated STAR Memory System for detection and repair of manufacturing faults improves yield. The memory compilers are also a part of the Design Ware Duet Packages and HPC Design Kit.

Abbreviations

SRAM	Static Random Access Memory
6T	6 Transistor
SNM	Static Noise Margin
NBTI	Negative Bias Temperature Instability
PBTI	Positive Bias Temperature Instability
CHC	Channel Hot Carrier
TMI	TSMC Model Interface
RM	Read Margin
WM	Write Margin
BIST	Built in Self-Test

Contents

Acknowledgement	v
Abstract	vi
Synopsys Inc. At A Glance	vii
Group Introduction	viii
Abbreviation	xi
Chapter 1 Semiconductor Memory	1
1.1 Introduction	1
1.2 Memory Compiler	3
1.3 Memory Compiler Features	5
1.3 Role of Memory Compiler	6
1.5 Motivation	6
1.6 Problem statement	6
1.7 Summary	7
Chapter 2 SRAM Background and SRAM basic Architecture	9
2.1 Classification of Memory	9
2.2 Introduction	10
2.3 SRAM Basic Architecture	11
2.4 6T SRAM Cell	12
2.4.1 Read operation	13
2.4.2 Write operation:	14
2.5 Summary	16
Chapter 3 Bitcell Performance Evaluation	17
3.1 Static noise margin:	17
3.1.1 SNM Dependencies	17
3.1.2 SNM analysis:	18
3.1.3 Transistor Sizing effects on SNM	19

3.2	Read current	20
3.3	Leakage current:	21
3.4	Worst Leak current	22
3.2	Summary	23
Chapter 4	Read Margin and Write Margin	25
4.1	Read margin:	25
4.1.1	Read operation:.....	26
4.1.2	Read current Scaling:.....	27
4.1.3	Monte Carlo Simulation:.....	28
4.1.4	RM setting.....	28
4.1.5	Challenges faced and solution:	30
4.2	Write Margin.....	31
4.2.1	DC write margin.....	31
4.2.2	AC write Margin.....	33
4.2.2.1	AC write margin (Word Line driven)	33
4.3	Summary	37
Chapter 5	Margins analysis using Ageing Models	39
5.1	Introduction	39
5.2	Factors Affecting NBTI and PBTI.....	39
5.2.1	Operating voltage	40
5.2.2	Temperature	40
5.2.3	Stress Time.....	40
5.3	Flow (TMI ageing methodology).....	41
5.4	Effect of ageing models on signals(cycles).....	42
Chapter 6	Conclusion and future work.....	43
Bibliography	44

List of Figures

<i>Figure 1 data generated by Memory Compiler.....</i>	<i>3</i>
<i>Figure 2 Classification of Memories.....</i>	<i>9</i>
<i>Figure 3 DRAM (left) and SRAM (Right).....</i>	<i>10</i>
<i>Figure 4 SRAM basic Architecture.....</i>	<i>11</i>
<i>Figure 5 6T SRAM structure.....</i>	<i>12</i>
<i>Figure 6 Read current in 6T SRAM.....</i>	<i>14</i>
<i>Figure 7 Fundamental of Write operation.....</i>	<i>15</i>
<i>Figure 8 SNM Setup.....</i>	<i>18</i>
<i>Figure 9 Read Current.....</i>	<i>20</i>
<i>Figure 10 6T SRAM Bit cell.....</i>	<i>21</i>
<i>Figure 11 Column Bit cell and leakage current.....</i>	<i>22</i>
<i>Figure 12 SRAM Read Operation Process.....</i>	<i>26</i>
<i>Figure 13 Read current Scaling.....</i>	<i>27</i>
<i>Figure 14 Signals of Bitcell bit lines and Sense amp.....</i>	<i>28</i>
<i>Figure 15 RM setting block.....</i>	<i>29</i>
<i>Figure 16 RM Setting and bit line voltage.....</i>	<i>29</i>
<i>Figure 17 SRAM Write operation process.....</i>	<i>31</i>
<i>Figure 18 DC Write Margin setup.....</i>	<i>32</i>
<i>Figure 19 DC Write Margin Simulation Waveform.....</i>	<i>33</i>
<i>Figure 20 AC write Margin WL setup.....</i>	<i>34</i>
<i>Figure 21 AC write Margin WL driven Simulation waveform.....</i>	<i>35</i>
<i>Figure 22 AC write margin BL driven setup.....</i>	<i>36</i>
<i>Figure 23 AC write margin BL driven Simulation waveform.....</i>	<i>37</i>
<i>Figure 24 Ageing Methodology flow.....</i>	<i>41</i>
<i>Figure 255 waveforms of most important signals of SRAM (Read and write operation).....</i>	<i>42</i>

List of Tables

<i>Table 1 Memory Compiler Configuration parameters</i>	4
<i>Table 2 Bit line and sense amp value</i>	30

Chapter 1

Semiconductor Memory

In this chapter we will have some glimpse of basic understanding and design of semiconductor Memory. We will see about Random access memories (RAM), as we can access any bit of data from any location and at any time. We will also discuss on the basic of memory compiler and its features also.

1.1 Introduction

Now a days IoT is becoming the most executive source of growth for different industry and in human life. IOT applications would require devices that rely on long life batteries or self-harvesting i.e. ultra-low power devices. Also they should be small in size to be able to fit in the smart devices. With IOT, data generated from billions of devices will need to be process, so user will need higher storage capacity i.e. more memory.

As a technology node is getting smaller and performance is also getting better for electronics devices, so this approach will get over as there is limit in technology, so we are finding any alternative approach. Technology is shrinking very fast so there is increase in packaging density transistor, so power consumption has become an important factor for a SoC or any chip because of limited battery life.

Memory systems have evolved through a variety of devices that match this characteristic, from vacuum tubes, delay lines, relays, and ferrite cores to semiconductor materials. All microcomputers are using semiconductor memory which always consists of RAM and ROM, made in the form of LSI circuits. The principal features of these circuits are high density, low cost, and ease of use. Due to the wide range of manufacturing process available, considerable differences exist in the types of semiconductor memory. These differences manifest themselves in the form of following factors:

- Power consumption
- Architectures and Components
- GUI
- Packing density
- Speed of operation
- Methods of storage
- Cost

Having fast and accurate models at all stages of a design is essential if SoC designers have to succeed in designing chips with embedded memories. Therefore, embedded memory characterization is of increasing concern to design teams. However, the move to new process geometries is intensifying the challenge the number of memory instances per chip increases considerably at advanced process nodes. The parasitic are also becoming more significant in advance process geometries and have started impacting the timing performance of the device. To support the full range of process, voltage, and temperature corners (PVTs) and to cater the sensitivity of process variation, designers have to perform more and more memory characterization runs. On top of that, the data processing per characterization grows exponentially.

Computer aided design (CAD) tools are used for design automation and optimization. Computer simulation is, and will continue to be, an essential part of the design process, both for performance verification and for tuning of circuits. However, the emphasis on simulation must be well-balanced with the emphasis on hands-on-design and analytical estimates, so that the extensive use of computer-aided techniques does not overwhelm the significance of the latter. In addition to the transistor-level circuit design issues, the accurate prediction and reduction of interconnect parasitic has become a very significant topic in high performance digital integrated circuits, especially for deep sub-micron technologies. Digital systems require the capability of storing and retrieving large amounts of information at high speeds. Memories are circuits or systems that store digital information in large quantity, hence in today's SoC era, nearly 70% of the chip area are occupied by the memory itself. The semiconductor markets have embraced the fact that the

architecture of the memory structure has a considerable impact on the performance of the system and any yield loss of memory IP cause the failure of entire chip.

1.2 Memory Compiler

Memory compilers are typically intellectual property of memory vendors. The purpose of compilers is to automatically generate various kinds of memories depending on the customer order. These compilers support the generation of various memory capacities as well as static random-access memory (SRAM) types, e.g., single- and dual-port memories. Discrepancy between the customers' requirements and vendors' portfolio are typically left to be solved by the customer. This results in a remarkable extra engineering effort, if an optimized SRAM solution is desired. To reduce this engineering overhead and quality enhancement a more flexible memory compiler was developed.

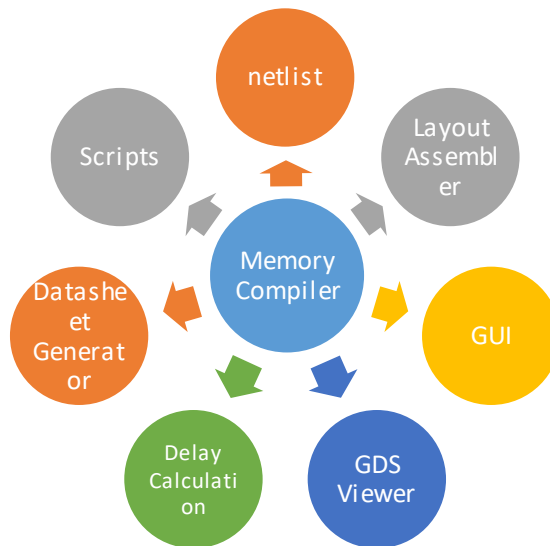


Figure 1 data generated by Memory Compiler

User (customer) input:

- no of words
- no of bits
- no of multiplexer
- no of bank

- no of column decoding

Mainly the compiler generates the following files:

- layout view (GDSII)
- netlist (SPICE)
- parasitic extraction netlist
- verilog test bench models
- timing and power models (NLDM and CCS)
- DRC /LVS verification reports
- datasheet

Types of Memory generated by memory compiler:

- Dual port SRAM
- Single port SRAM
- Low power dual port SRAM
- Low power single port SRAM
- High Speed Compilers
- High Density Compilers
- Low power Compilers
- Small Size compiler

	High-density (SP/DP SRAM)	High-density 1P RF	High-density 2P RF	High-density ROM	High-density STAR-16M	Ultra-high-density 2P RF	High-speed (SP/DP SRAM)
Total bits	256-1280K	128-128K	128-128K	256-1280K	256-16M	128-512K	256-1280K
World range	32-16K	16-1K	16-1K	64-64K	16-128K	16-2K	32-16K
I/O range	8-320	8-256	8-256	4-160	16-320	8-256	8-320
Column mux	4,8,16	2,4	1,2,4	8,16,32,64	4,8,16,32	2,4	4,8,16
Bank	1,2,4,8	1,2	1,2	1,2,4,8	1,2,4,8 VBK = 2,4,8	1,2	1,2,4,8
Redundancy	Column	Column	Column	None	Column	Column	Column
Periphery Vt	Standard/high	Standard/high	Standard/high	Standard/low	Standard/high	Standard/high	Standard/low

Table 1 Memory Compiler Configuration parameters

1.3 Memory Compiler Features

Following are the features:

1. Power management mode:

- a. **Mode-1:** This mode provides leakage reduction with fine-grained power gating and source biasing.
- b. **Mode-2:** This mode provides, integrated periphery power gating with data retention available and the memory outputs are held low.
- c. **Mode-3:** In this mode, there is a complete shutdown (both the periphery and array are power gated), with no data retention, and the memory outputs are held low.

2. Test mode (BIST Interface):

This option creates memory instances that include all the necessary logic to facilitate at-speed Built In Self-Test (BIST). When this feature is enabled, the generated memory instance includes multiplexers (MUXES) for all address, control and data signals as well as comparators and capture logic. All output signals are fully scan able and the data is synchronous with the external clock. Incorporating this logic into the memory instance reduces the critical path when BIST is enabled and reduces the number of wires that are required to route between the memory instance and the BIST engine. The integrated logic will also enable high performance testing of functional logic surrounding the memory in the designs, using ATPG scan tools. Synchronous Write-through is available when these options are enabled. This allows input data to flow to output pins synchronously with the clock.

3. Dual rail function:

This function enables the feature of dual power supply, for array and periphery (both individual). So for this implementation level shifter is added in between periphery and array.

4. Redundancy (Row, Column):

This feature enables the memory compiler to generate memory instances that include redundancy for repair. When this mode is activated, additional memory is added to the instance to be used when BIST diagnostics determine that a repair is necessary.

5. Assist circuit (Read, Write):

This feature is used to boost negative bit line. It can be done by Word line level Reduction.

1.3 Role of Memory Compiler

In System on Chip (SoC) design, i.e. customer required memory with different aspect ratio with different size. Memory compiler provide the features to generate the memory instances with different sizes with different features.

1.5 Motivation

As now a day's technology is shrinking, so it has become hard to design very less failure and error free (read and write operation must be successfully done) SRAM design. It means that yield must be increased. So thesis work represents the technique which increases yield. Project requires working with teams involved in major projects around the world, and implementation and debugging of SRAM memory compilers research problems.

1.6 Problem statement

Time to market as well as good design are the most important things in industry. So it means that we need most efficient mythology and flow starting from beginning to end level of design. Objective of this project is to provide very effective design of SRAM, so that there must be no failure in read and write operation, which indirectly increases yield.

1.7 Summary

In this chapter we have discussed the basic understanding of the semiconductor memory. We have also discussed the introduction to memory compiler, some of the memory compiler features and role of memory compiler. In next chapter we will discuss about basic memory architecture (SRAM) and read and write operation of SRAM.

Chapter 2

SRAM Background and SRAM basic Architecture

In this chapter we will discuss about classification of Memories, comparison of SRAM vs DRAM. This chapter also includes understanding of Basic SRAM Architecture, Read, and Write operation of 6T SRAM.

2.1 Classification of Memory

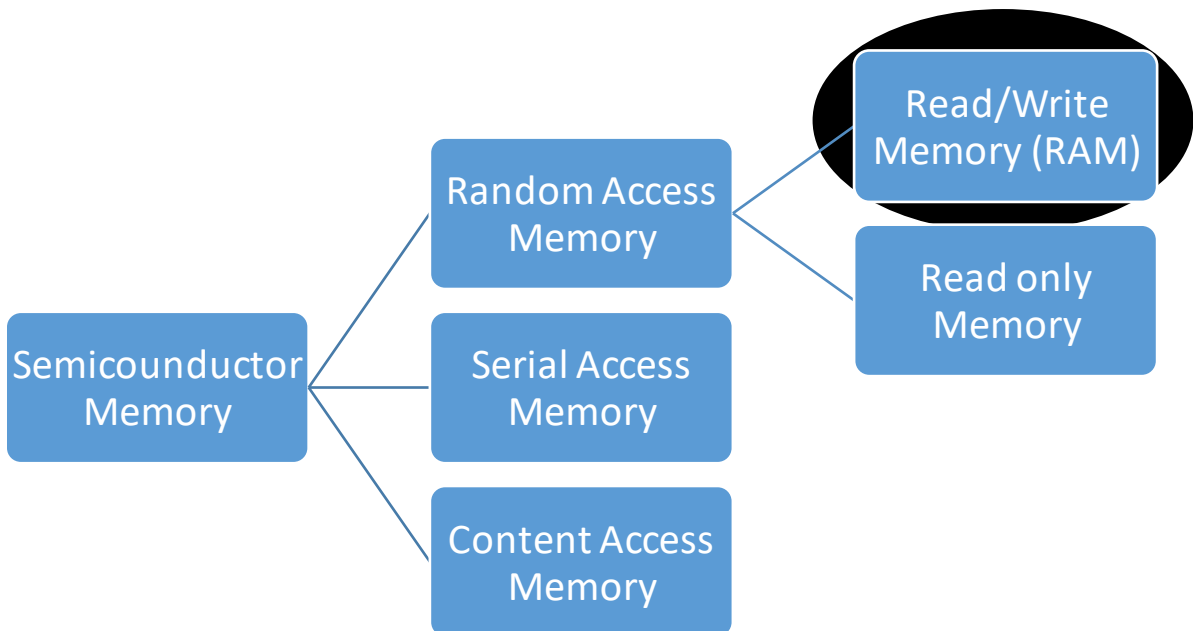


Figure 2 Classification of Memories

RAM can be classified into following categories Read/Write Memory and Read Only Memory. In a non-volatile Memory like Read Only Memory (ROM), the stored data is maintained indefinitely, even without power, and writing in to the

memory takes considerably more time (about milliseconds) than reading. Read/write memory (commonly called RAM) is data that is stored temporarily and the read and write time is approximately equal. RAM cells can be furthermore divided into static and dynamic memory cells. Static memory (SRAM) cells use a latch composed of cross-coupled inverters to store data for on chip application.

This allows the value to be maintained in a cell if power is available. Data storage in dynamic memory cell (DRAM) is based on the dynamic storage of charge on a capacitor. Therefore, with dynamic memory cells, periodic refreshing is necessary to maintain the value. Transistor-level schematic of a SRAM and a DRAM cell can be found in Figure 2.2 Bit-lines form the data path to/from the cell, while word-lines select a cell to be accessed.

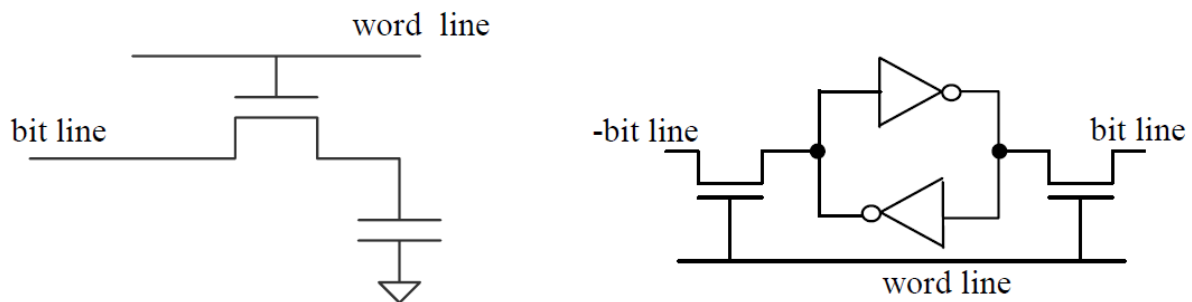


Figure 3 DRAM (left) and SRAM (Right)

2.2 Introduction

An SRAM is array of memory cells. To access particular Memory address or memory cell, address decoder is used for read and write operation. memory is divided into two parts array part and periphery part. Address decoder is a part of periphery. The basic block diagram of the SRAM contains arrays of memory cells, control block for address decoder for basic. We can say that performance of system is largely depends on the memory, therefore increase in speed and reduction in power leakage is important concern if performance of system is concern. In cases like this it is very important to reduce the leakage. This issues have been seen in many systems on chip memories. thus, it is necessary for designer to determine the cause of delays and leakage in memory blocks. Hence, it can be resolved or reduced,

and allowing to find a new better technique. which will improve performance of the system.

It is possible to reduce the power leakage by using different techniques. Such as bank structure, such as circuit partitioning, increase the gate oxide thickness thus gate oxide reduces, increase the threshold voltage. The bank structure technique increases the speed of memory. The control block, the address decoder and I/o ports are design by using low threshold transistor, while in the design of the bit-cell and sense amplifier high threshold voltage transistor are used.

2.3 SRAM Basic Architecture

SRAM architecture consists array which contains data storage Bitcell. There are two types of decoder used to read-write location in array.

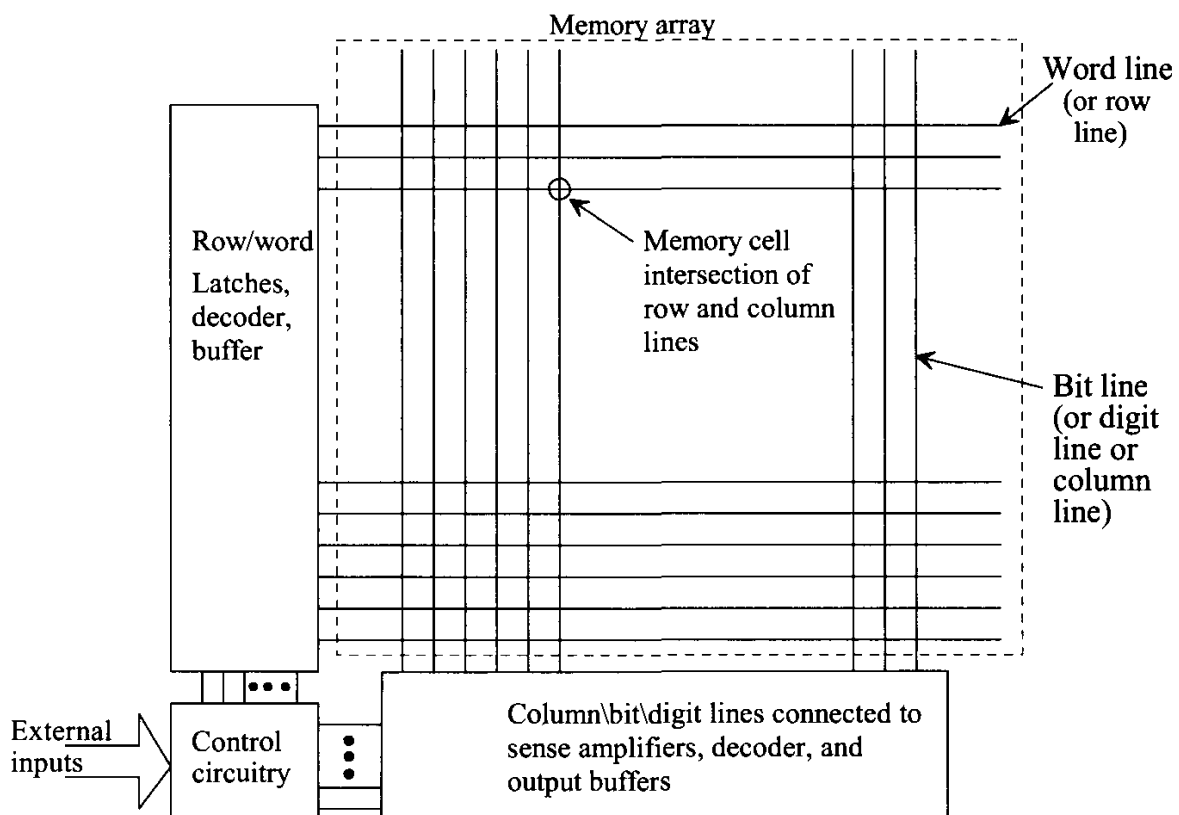


Figure 4 SRAM basic Architecture

The **RAM architecture** consists of the following structures:

SRAM Bitcell, used to store one data bit.

Bit Line Precharge Circuit, precharge bit lines to compensate for voltage drop across pass transistors.

Write Buffers, buffers write-data so that it can write on RAM cells.

Sense Amplifier, generate logic values based on difference on bit-line voltages.

Row & Column Decoders, for address generation to select Bitcell.

2.4 6T SRAM Cell

The structure of 6T SRAM cell is like two cross-coupled inverters. Here the Main efforts put to reduce this cell area such that millions of storage cells can fit on a chip. Steady state power dissipation is controlled by using larger threshold transistor in the array portion. This SRAM cell layout is highly optimized to reduce overall area in memory. For this sometimes M5 & M6 are replaced with undoped poly silicon. Such type of configuration called 4T SRAM because in cell it consists only four transistors.

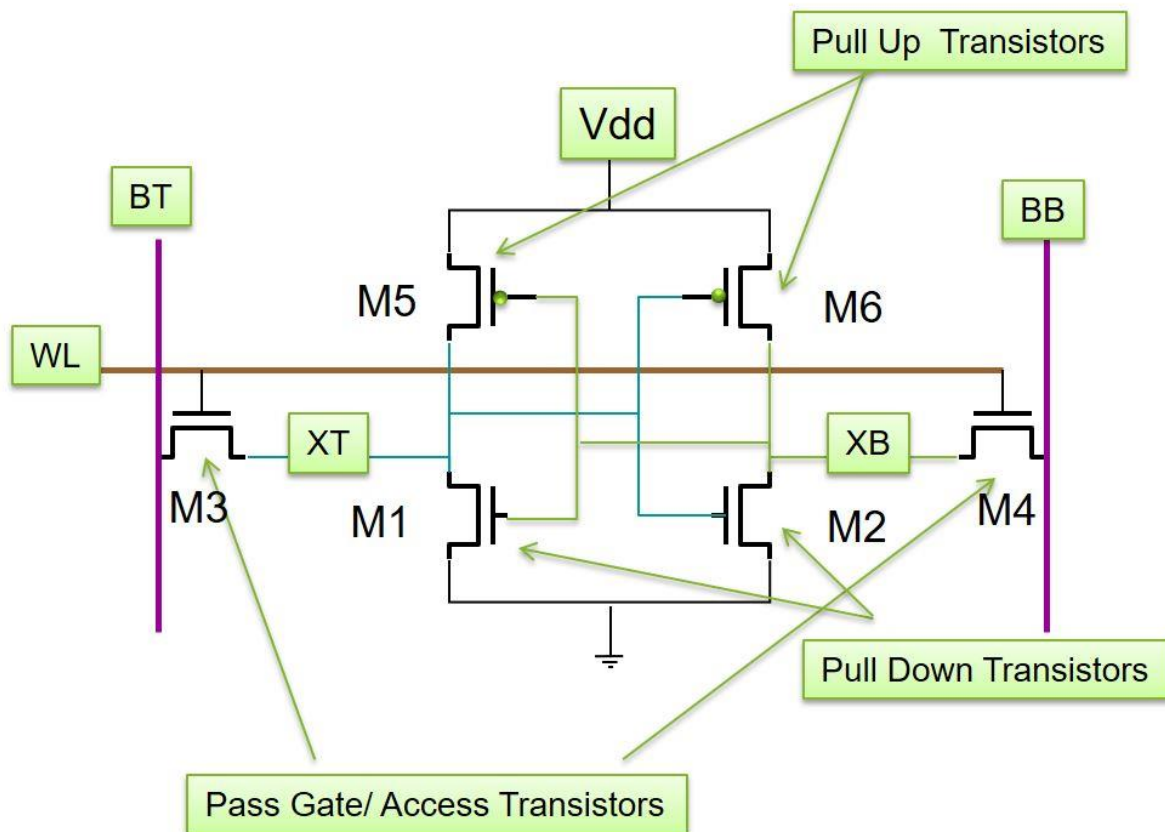


Figure 5 6T SRAM structure

For power reduction current through this pull up resistor can be reduced by using large size. So there is trade-off between area and power. So for this reason 6T is adopted by VLSI industry.

For a read operation, only one side of the cell draws current. thus, a small differential voltage develops between bit and bit bar column lines. The columns address decoder and multiplexer select the column lines to be accessed. The bit lines will get a voltage difference as the selected cells discharge one of the two bit lines. This difference is amplified and sent to output buffers. As there is large no of cell connected to Bit line so there will be large Capacitance also. The capacitance includes source /drain capacitance, wire capacitance and source /drain contact capacitance. A contact or via is connected between two cell.

For write operation, one of the bit line is pulled low if we want to store 0, while the other one is pulled low if we want to store 1. The requirement for a successful write operation is to swing the internal voltage of the cell past the switching threshold of the corresponding inverter after flipping word lines must go in reset to ensure single flip.

The design of the cell involves the selection of transistor sizes for all six transistors to guarantee proper read and write operations. Since the cell is symmetric, only three transistor sizes must be specified, either M1, M3, and M5 or M2, M4, and M6. Our main aim is to do good read stability which provides good read current and good write ability.

2.4.1 Read operation

We now describe the design details of the 6T RAM cell for the read operation using. Assume that logic '0' is stored in the cell. Therefore, M1 is on and M2 is off. Initially, bt and bb are precharge to a high voltage by a pair of column pull-up transistors. The row selection line, held low in the standby state, is raised to VDD which turns on access transistors M3 and M4. Current (I_{cell}) begins to flow through M3 and M1 to ground. The resulting cell current slowly discharges the capacitance C_{bit} . Meanwhile, on the other side of the cell, the voltage on remains high since there is no path to ground through M2.

The difference between bt and bb is fed to a sense amplifier to generate a valid low output, which is then stored in a data buffer. Upon completion of the read cycle, the word line is returned to zero and the column lines can be precharge back to a high value. When designing the transistor sizes for read stability, we must ensure that the stored values are not disturbed during the read cycle.

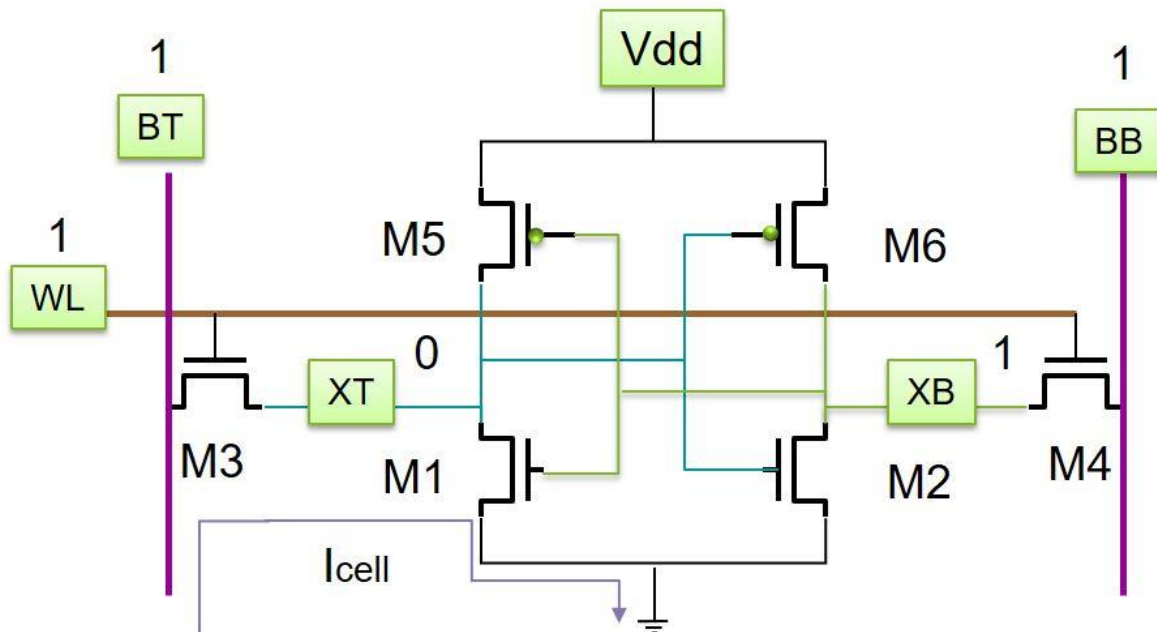


Figure 6 Read current in 6T SRAM

Sizing:

- Voltage at node XT should not exceed threshold voltage of MN2
- Pull down stronger than pass gate to meet above criteria
- Sizing can be decided by equating currents through pass gate MN3 and pull down MN1
- MN1: linear MN3: Saturation
- Ratio of pull down to pass gate ~ 1.5

2.4.2 Write operation:

The operation of writing 0 or 1 is accomplished by forcing one-bit line, either bt or bb, low while the other bit line remains at about VDD. To write 1, is forced low, and to write 0, bt is forced low. The cell must be designed such that the conductance

of M4 is several times larger than M6 so that the drain of M2 is pulled below V_S . This initiates a regenerative effect between the two inverters. Eventually, M1 turns off and its drain voltage rises to VDD due to the pull-up action of M5 and M3. At the same time, M2 turns on and assists M4 in pulling output to its intended low value. When the cell finally flips to the new state, the row line can be returned to its low standby level. The design of the SRAM cell for a proper write operation involves the transistor pair M6-M4. When the cell is first turned on for the write operation, they form a pseudo-NMOS inverter. Current flows through the two devices and lower the voltage at node from its starting value of VDD. The design of device sizes is based on pulling node below V_S to force the cell to switch via the regenerative action. Note that the bit line is pulled low before the word line goes up. This is to reduce the overall delay since the bit line will take some time to discharge due to its high capacitance.

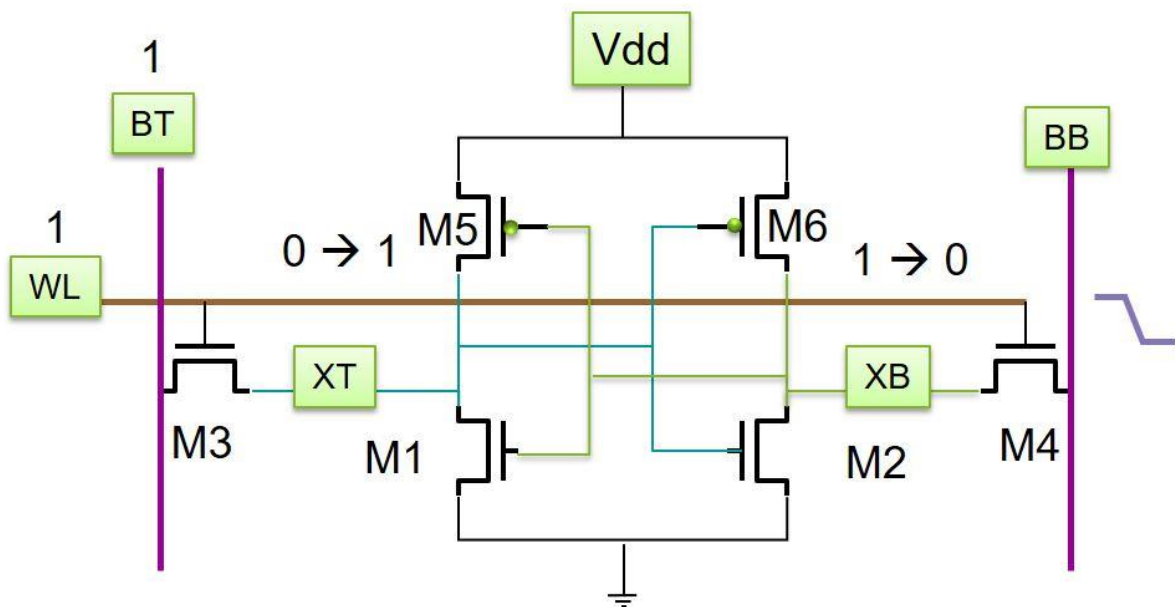


Figure 7 Fundamental of Write operation

Sizing:

- Conductance of pass gate must be larger than pull up so that drain of MN2 is pulled down below switching threshold voltage of inverter

- Size is determined by pulling drain of MN2 to V_T of transistor of MN1 and equating current through pull up MP2 and pass gate MN4
- MP2: saturation MN4: linear
- $(W/L)_{MN4} / (W/L)_{MP2} \sim 1.5$

2.5 Summary

In this chapter we discussed about SRAM architecture, various types of memories, 6T bit cell its sizing, read and write operation.

Chapter 3

Bitcell Performance Evaluation

In this chapter we will discuss that how Bitcell of SRAM is analysed.

The main parameters of Bitcell analysis are as below:

1. *Static noise margin*
 2. *Read current*
 3. *Leakage current*
-

3.1 Static noise margin:

Stability and Noise sustain is one of the most important aspect of SRAM. Process variation, worst condition and sensitivity towards noise is determined by cell stability. STATIC NOISE MARGIN determines stability of the cell.

3.1.1 SNM Dependencies

- Transistor width modulation
- Word line value
- Bit line value
- Power supply voltage
- Temperature

3.1.2 SNM analysis:

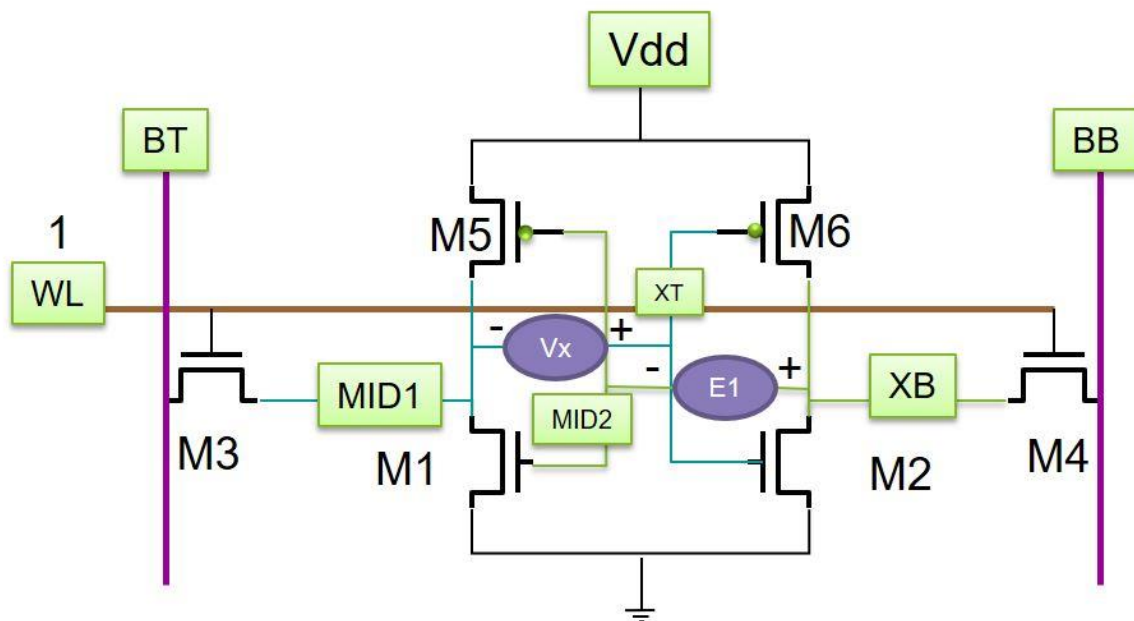


Figure 8 SNM Setup

Figure 3.1 shows setup for Static Noise Margin calculation. SNM is calculated by applying predetermined noise using voltage control voltage source. Initialize XT to 0 and XB to Vdd. Bit lines (BT, BB) and Word line (WL) are at Vdd. Introduce noise in the form of voltage sources Vx and E1 (voltage coupled to Vx). Slowly increase Vx from 0 and monitor the points MID1 and XB to see when the cell flips $V_{MID2} + V_{NOISE}$ less than V_t MN1 (Threshold Voltage of Pull down Transistor). Till this criterion is satisfied the memory cell will not flip. As soon as $V_a + V_{noise}$ becomes more than the threshold voltage of the pull down transistor, the memory cell gets flipped. Thus, care must be taken in deciding the transistor sizes while designing the memory cell.

3.1.3 Transistor Sizing effects on SNM

- **Pass Transistor sizing effects on SNM**

Pass transistor (W/L) ratio and Static Noise Margin are inversely proportional. As we increase the (W/L) ratio of pass transistor, Static Noise Margin decreases.

- **Pull-Up Transistor sizing effects on SNM**

Pull-Up transistor (W/L) and Static Noise Margin are directly proportional to each other. For SNM improvement we must increase (W/L) of Pull-Up transistor. Assuming the Bitcell as a resistive n/w we can see that resistance will decrease by increasing the (W/L) of Pull-Up transistor so voltage drop will be small and VDD will be maintained on B node.

- **Pull-Down transistor sizing effects on SNM**

Pull-Down transistor (W/L) and Static Noise Margin are directly proportional to each other. By increasing (W/L) of Pull-up transistor, we can increase Static Noise Margins.

3.2 Read current

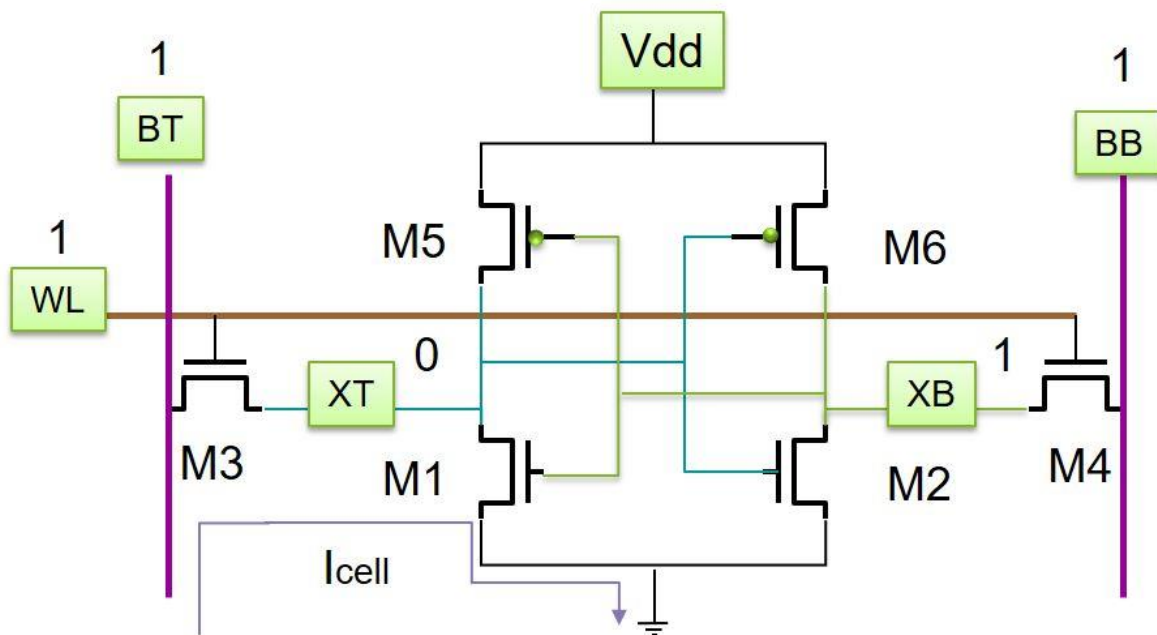


Figure 9 Read Current

Purpose:

The purpose of this simulation is to characterize DC read current

Method:

XT is initialized to 0

XB is initialized to Vdd

WL is at Vdd

BT is at Vdd

BB is at Vdd

The read current (I_{cell}) is defined as the current through the pass gate M3.

3.3 Leakage current:

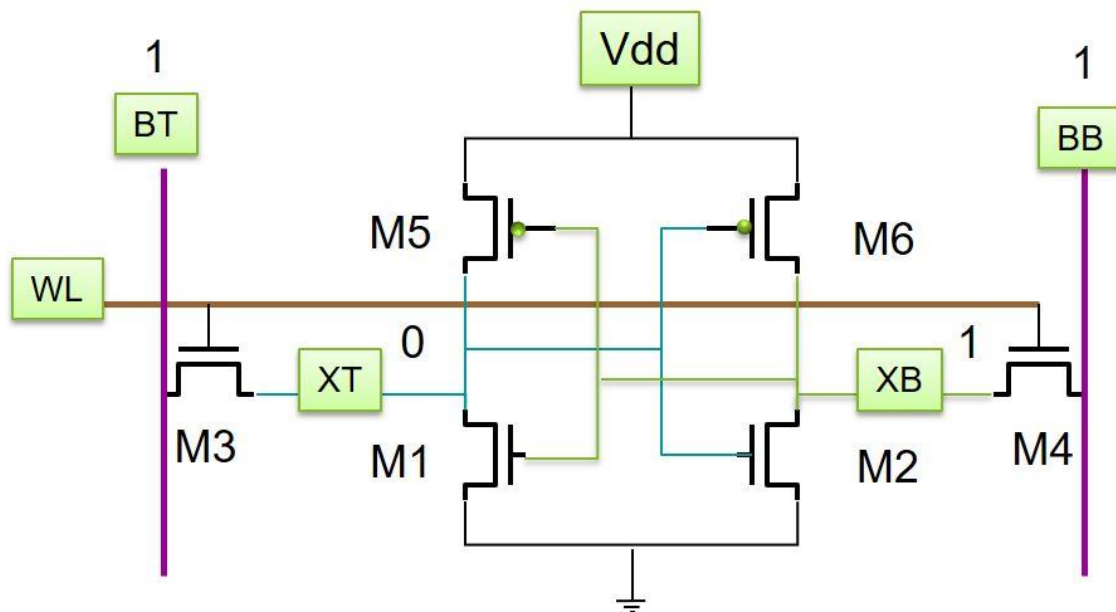


Figure 10 6T SRAM Bit cell

Purpose: The purpose of this analysis is to characterize the bit cell contribution to column leakage. the main purpose of this test is to see the margin available for the total cell leakage current in a long column (from unselected WLs) during a read operation.

Ileak is characterized to get maximum number of bit cells that can be supported per column. Since only one Bitcell will be activated during read operation in each column, so rest of the 'rows - 1' Bitcell will be in OFF state and deteriorating read operation by giving leak current.

- Ileak is determined by PG device.
- Method (DC characterization)
- XT is initialized to 0.
- XB is initialized to Vdd.
- WL is at 0 (OFF).

- BT is at Vdd.
- BB is at Vdd.
- I_{leak}= current through the pass gate next to node storing 0 (device M3 in figure)

3.4 Worst I_{leak} current

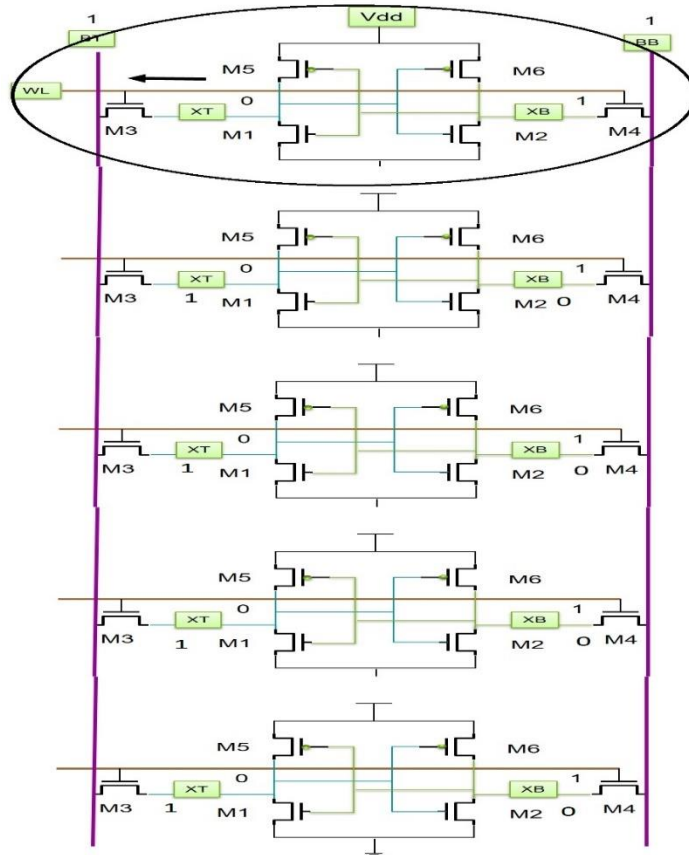


Figure 11 Column Bit cell and leakage current

Let circled Bitcell be activated.

In this worst case example, BT is being read and worst leakage current is fed to BB, which will tend to kill V_{diff} being fed to SA's internal nodes and thus Read operation will get slowed down.

$$I_{leak} < I_{read} = N$$

3.2 Summary

This chapter tells about SRAM Bitcell critical parameters and various setup to calculate it.

Chapter 4

Read Margin and Write Margin

In this chapter we will discuss about read and write margin and RM and WM setting is most important thing for SRAM. We will also look at how RM and WM is important for read and write operation.

4.1 Read margin:

Memory compiler is used to generate memory instances of different size and of different configuration i.e. different combination of various parameters like NW (No of Words), NB (No of Bits), CM (No of inputs to the Column Mux), and BK (No. of Bank) etc. so with different sizes and different configuration, word lines and bit lines will have different RC loads. For example, if instance generated is tall, bit lines have more load and word lines have fewer load. If instance generated is wide, bit lines have fewer loads and word line have more load. Because of these different loads for different instances, the read/write operation takes different amount of time. If we made our memory compiler by considering biggest instance (i.e. with max size) the smallest instance which is generated from the memory compiler has also timing as per the biggest instance and we will lose the speed advantage. So somehow we should design our memory compiler in such a way so that the time required to read/write from/to bit cell is optimum i.e. as per the size of instance generated.

4.1.1 Read operation:

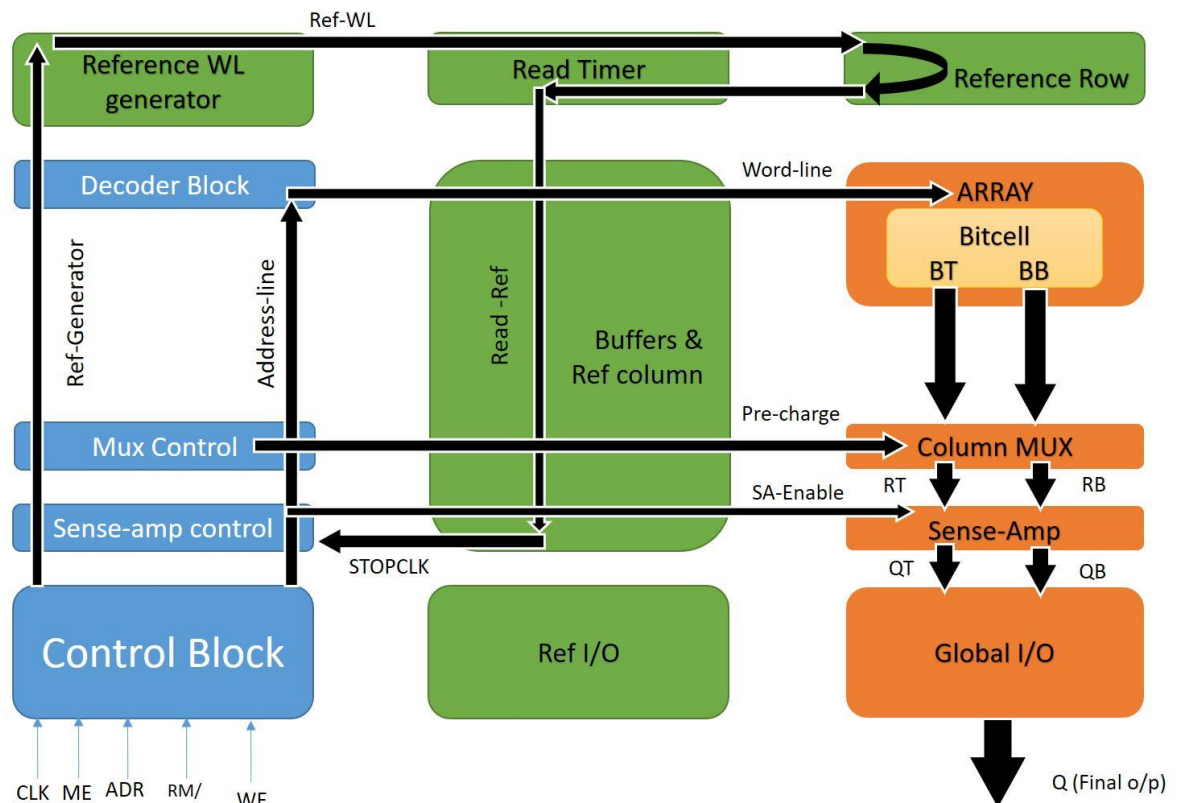


Figure 12 SRAM Read Operation Process

Fig 12 shows the read operation signals flow for SRAM. Read operation time basically made up of following different components:

- CLK (clock) to WL (Word Line) delay
- WL to Bit cell pass transistor delay (due to load on word lines)
- Bit line discharging delay (due to RC load on bit lines)
- Sense amplifier (SA) resolution time
- SA to output time

Out of all above components only word line delay and bit line delay are critical because they vary with size of memory component.

Sense amplifier starts detection of data when the SAE (Sense Amplifier Enable) signal is activated and SAE signal is generated from STOPCLK signal, which is generated from reference block. So we must control SAE signal as per the size of our memory.

4.1.2 Read current Scaling:

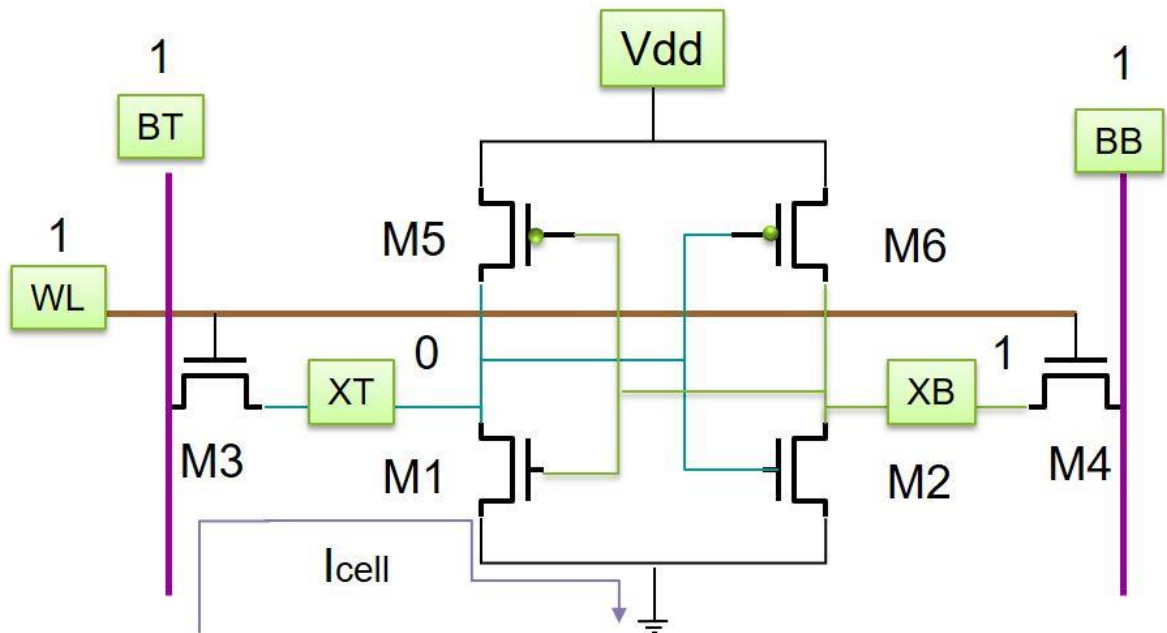


Figure 13 Read current Scaling

What is Read Margin?

Minimum differential voltage required at the input of a 5-sigma SA connected to a 5-sigma bit cell.

4.1.3 Monte Carlo Simulation:

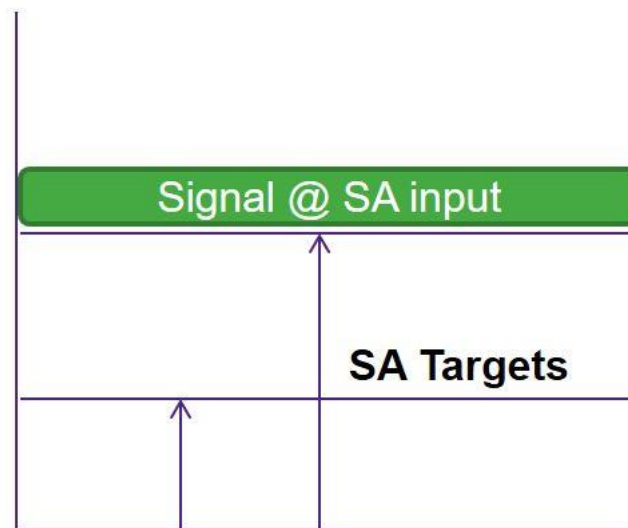


Figure 14 Signals of Bitcell bit lines and Sense amp

We do BCA and Sense Amplifier Analysis

For Simple example:

- Bitcell Mean Current = 20uA
- Bitcell Current Sigma (1 Sigma) = 2uA
- Bitcell Worst Read Current = $20 - 5 \times 2 = 10\text{uA}$
- Weak Current Ratio = $10/20 = 0.5$
- Worst case Read current is evaluated from **100M monte** Now let's task 10mV Sense Amplifier Sigma, 5 Sigma = 50mV
- **Signal Target** Consider Weak Bitcell = $50/0.5 = 100\text{mV}$
- **PVT: SSG_FF** → Bitcell Slow and Self time path fast

4.1.4 RM setting

Reference bit line is discharged through four pull down transistor connected to it. The gate node of each pull-down transistor is connected to RM pins RM [3:0] (fig). There are sixteen different combinations possible by toggling these pins to logic 1 or logic 0. So we have sixteen different possible rate at which the reference bit line can be discharged and can control the SAE generation time. RM setting is process

of selecting the one of these combinations so that we can make sure that enough differential voltage has been developed across the terminal of sense amplifier.

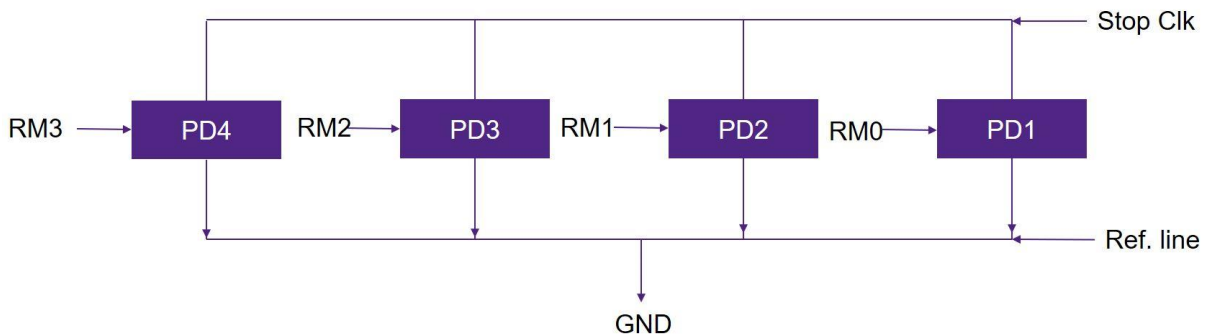
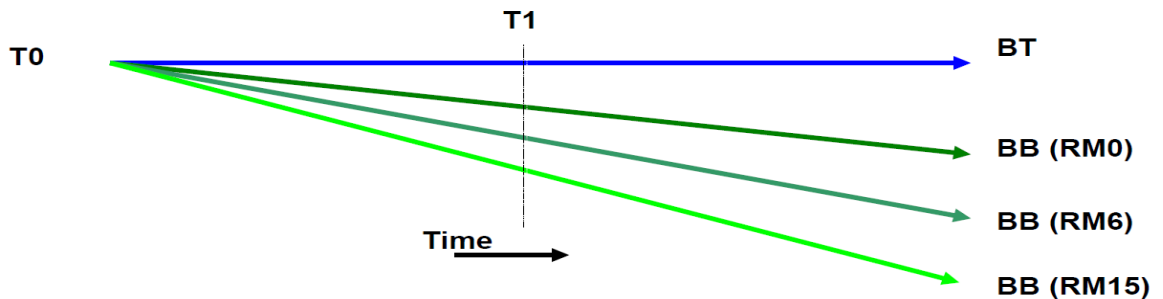


Figure 15 RM setting block

- Higher Input Voltage results in Greater Reliability of the sensed data
- Delay time → Longer cycle time → reduce operating speed → increase access time
- Trade-off between: memory speed → yield/reliability



RM	Differential @ T1
0	70 mV
6	150 mV
15	225 mV

Figure 16 RM Setting and bit line voltage

Here as shown in fig 4.5 RM setting is shown wrt its difference voltage between BT and BB at time T1.

4.1.5 Challenges faced and solution:

ROW	Column	Mode	VT	BC (mA)	SA(mA)	Delta	
256 (max)	128 (max)	Fast vdd	(0.81)	ULVT (200mv appx.)	0.0191	0.0192	-0.0001
256 (max)	192 (max)	Default vdd	(0.72)	ULVT (200mv appx.)	0.0229	0.0239	-0.0010

Table 2 Bit line and sense amp value

RM: Solution

Here in above reading for big Instance only we are getting Negative.

It's because of we are not getting required (95%) gate voltage at Pass transistor. So reduce target to 90% or increase Gate Voltage of Pass Transistor.

4.2 Write Margin

To successfully write the data on to the bit cell we do analysis of various write margins during the design process. Various margins for write operation are as follows.

- (1) DC write margin (Voltage domain)
- (2) AC write margin (Time domain)

Two types of AC (Time domain) margins are there.

- i. WL driven
- ii. BL driven

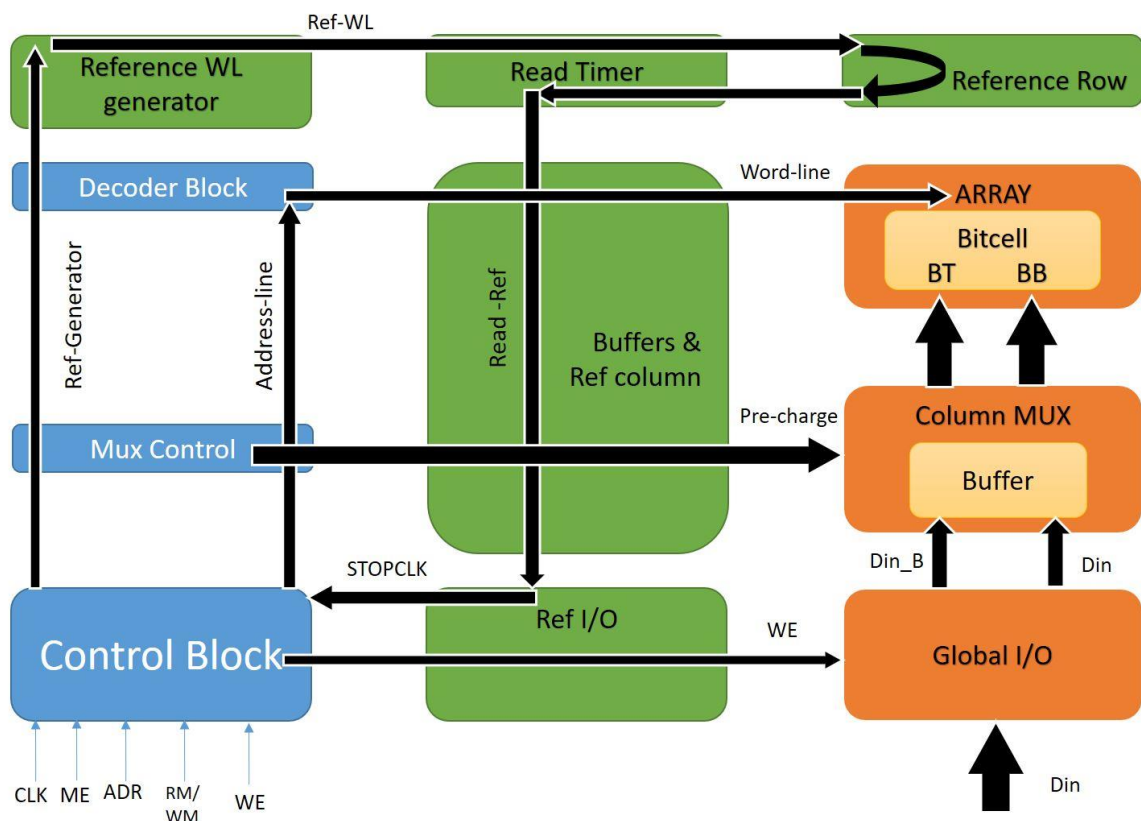


Figure 17 SRAM Write operation process

4.2.1 DC write margin

Purpose:

The purpose of this margin is to characterize the minimum requirement for the voltage condition on the critical nodes during the write cycle. This simulation is intended to look at how well the cell can be written to. This is the DC simulation so simple net list is needed (extracted net list may not be intended). Failure of this margin results into cell not

(6) Internal nodes XT and XB are monitored and when the cell flips. A cell is considered to be flipped when XT's and XB's values cross each other. The V_x voltage at which the cell flips are then subtracted from 5% of V_{dd} and that is defined to be the voltage domain write margin. (as shown in Figure 4.6 below)

(7) $WM = (BL \text{ voltage at which XT and XB meet}) - 0.05 \cdot v_{dd}$

(8) Here the smaller the write margin the harder it is to write the bit cell.

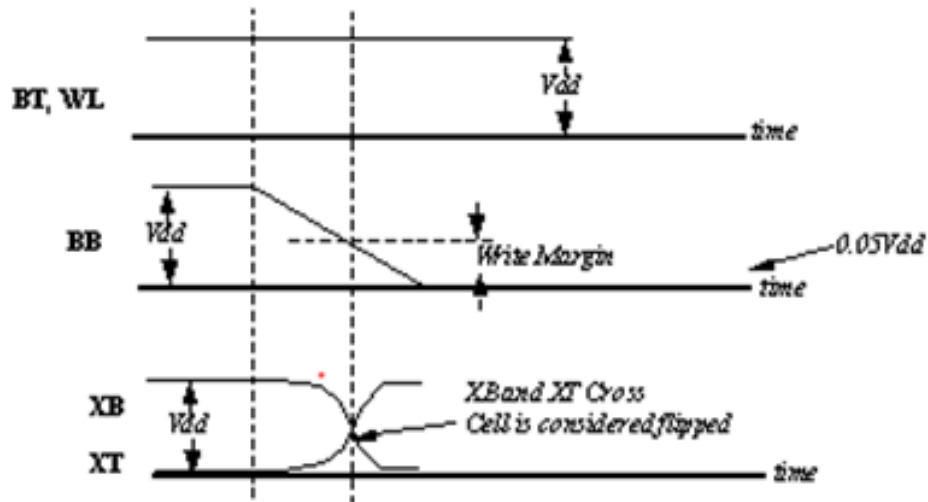


Figure 19 DC Write Margin Simulation Waveform

4.2.2 AC write Margin

4.2.2.1 AC write margin (Word Line driven)

Purpose:

The purpose of this simulation is to characterize the minimum requirement for the time on critical nodes during the write cycle. This situation is intended to look at how good cell can be written to. This is AC simulation and so extracted net list must be used.

This margin gives the minimum time required by bit cell to get flipped once the WL gets fired and BT is already at logic 0. In this setup we keep BT at logic high, BB at logic low and give the ramp at WL and measure the time in which the bit cell gets flipped. The value of time at which XT/XB lines reached at 80% of the final value from the 50% of WL, is called WL driven time domain write margin. This margins ensures write to weakest cell with respect to word-line (Sigma, based on yield aimed, determined statistically). Failure of this margin implies word line width is

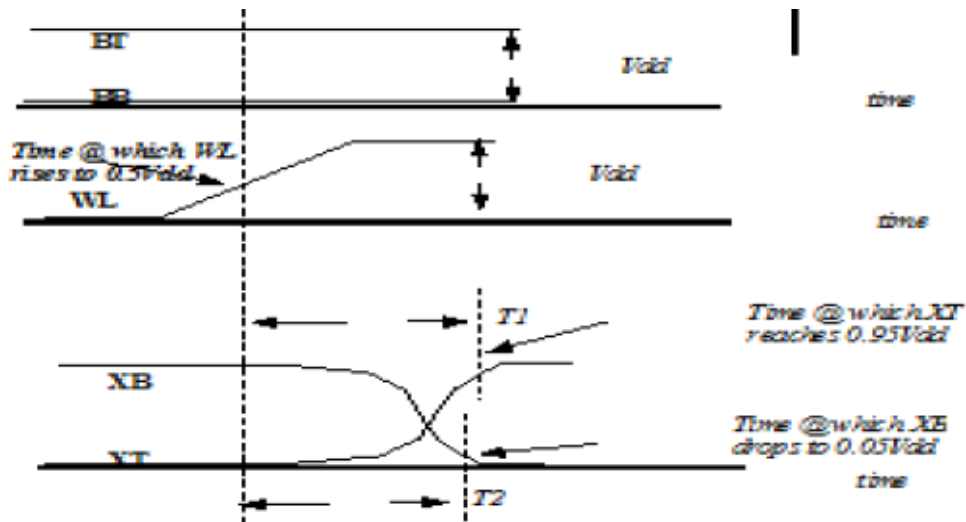


Figure 21 AC write Margin WL driven Simulation waveform

4.2.2.2 AC write margin (Bit line driven)

Purpose:

The purpose of this simulation is to characterize the minimum requirement for the time on critical nodes during the write cycle. This simulation is intended to look at the how well the cell can be written to. This is an AC simulation so extracted net list must be used. This margin gives the minimum time required for the bit cell to get flipped once the BL gets _red and WL is already at logic 1. In this setup we keep WL at logic high, BT at logic high and give the ramp at BB and measure the time in which the bit cell gets flipped. The value of time at which XT/XB nodes reaches to 80% of the final value from the 50% of BB, is called BB driven time domain write margin. This margins ensures write to weakest cell with respect to bit line (Sigma, based on yield aimed, determined statistically). Failure of this margin implies data time from the write driver is not sufficient for write. This margin should be checked at all the corners of the physical instance and for both write 0 and write 1.

Method:

The basic methodology here is to put the bit cell in a state, turn on the WL and then quickly drive the BL down to zero from Vdd. Two points are monitored: 1) where the initially low internal node rises to 95% Vdd 2) where the initially high internal

node drops to 5% of Vdd. The maximum of these two points are taken to be the write margin in seconds.

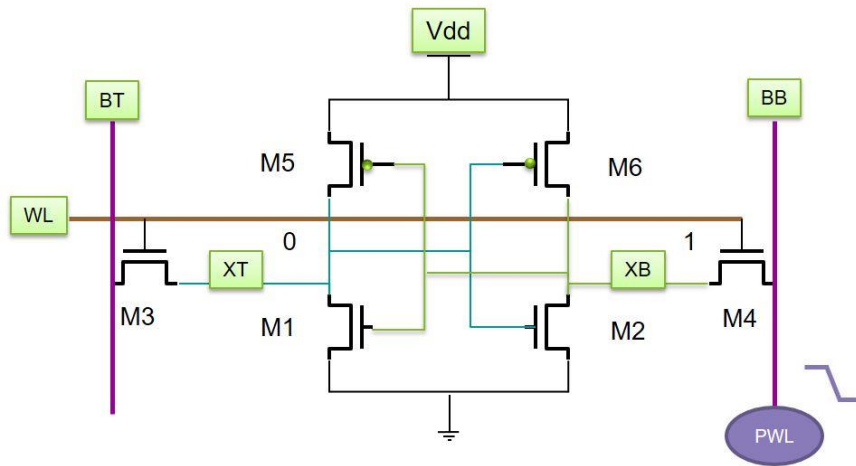


Figure 22 AC write margin BL driven setup

- (1) XT is initialized to 0.
- (2) XB is initialized to Vdd.
- (3) WL is turned on and therefore is at Vdd.
- (4) BT is at Vdd.
- (5) BB is a transient going from Vdd to 0 in 1ns using piece-wise-linear function.
- (6) Write margin (measured in time domain) is defined as the maximum of T1 (time between BB reaching 50% Vdd to XT rising to 95% Vdd) and T2 (time between BB reaching 50% Vdd to XB dropping to 5% Vdd)

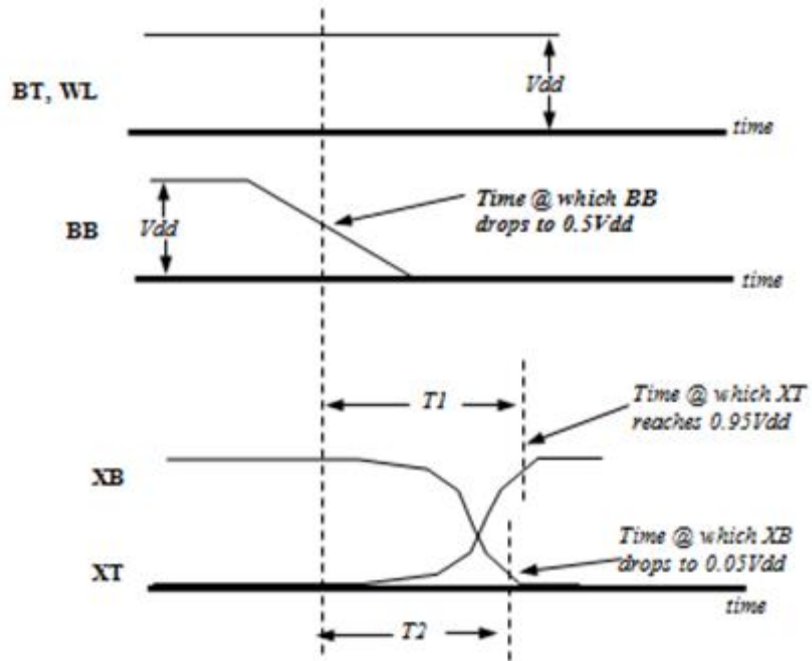


Figure 23 AC write margin BL driven Simulation waveform

4.3 Summary

In this chapter we discussed about Read margin and write margin and methods to measure margin. we also discussed about RM/WM settings.

Chapter 5

Margins analysis using Ageing Models

In this chapter we will discuss about TMI Ageing Models, its effects on memory architecture, different types of margin analysis using Ageing models and TMI (TSMC Model Interface) ageing methodology.

5.1 Introduction

There are mostly two factors which effect circuit lifetime which are Channel Hot Carrier (CHC) and Bias Temperature Instability (both NBTI and PBTI). Due to BTI threshold voltage gets shift and it is strong function of stress voltage and temperature. Due to operating temperature and stress time, NBTI and PBTI create a decrease in drain-to-source current and an increase in propagation delay. Due to exponential dependence on delay and leakage power threshold voltage is an important parameter. So the variation in the threshold voltage may affect the operating frequency. Further, we will discuss the factors affecting the NBTI and PBTI

5.2 Factors Affecting NBTI and PBTI

The following factors will affect NBTI and PBTI:

- Operating Voltage
- Temperature
- Stress Time

5.2.1 Operating voltage

As the operating voltage increases, the negative bias of the PMOS transistors increases, which in turn increases the NBTI degradation. Scaling of technology node leads to high electric fields at the gate, causing NBTI. Higher operating voltages result in higher electric fields across the device junction resulting in higher stress.

5.2.2 Temperature

The effect of NBTI worsens at an elevated temperature and shows exponential dependence due to the increasing dissociation of Si-H bonds at high temperature. Typical stress temperatures range from 100°C to 250°C, encountered during burn-in. During extremely high performance applications, possibly at the highest operating frequencies, the higher toggling activity of signal nets can result in formation of local hot-spots inside the chip's functioning major parts, resulting in an elevated temperature in some parts of the chips.

5.2.3 Stress Time

For transistors connected in stack, degradation due to NBTI and PBTI cannot be considered individually because it becomes a function of signal probabilities of transistors present up and down in the stack for PMOS and NMOS, respectively. When there are many transistors connected in series, the equivalent signal probability is taken into consideration for stress time. The variation in the threshold voltage is caused by variations in fixed charge and interface trapped charge density, also referred to as permanent traps. Due to the unsaturated valence electron at the SiO₂- Si interface, holes are attracted, weakening the Si-H bond, resulting in V_{th} degradation. Due to technology scaling down and the use of high k material, the PBTI creates a bulk trap similar to an interface trap of SiO₂ in NBTI.

5.3 Flow (TMI ageing methodology)

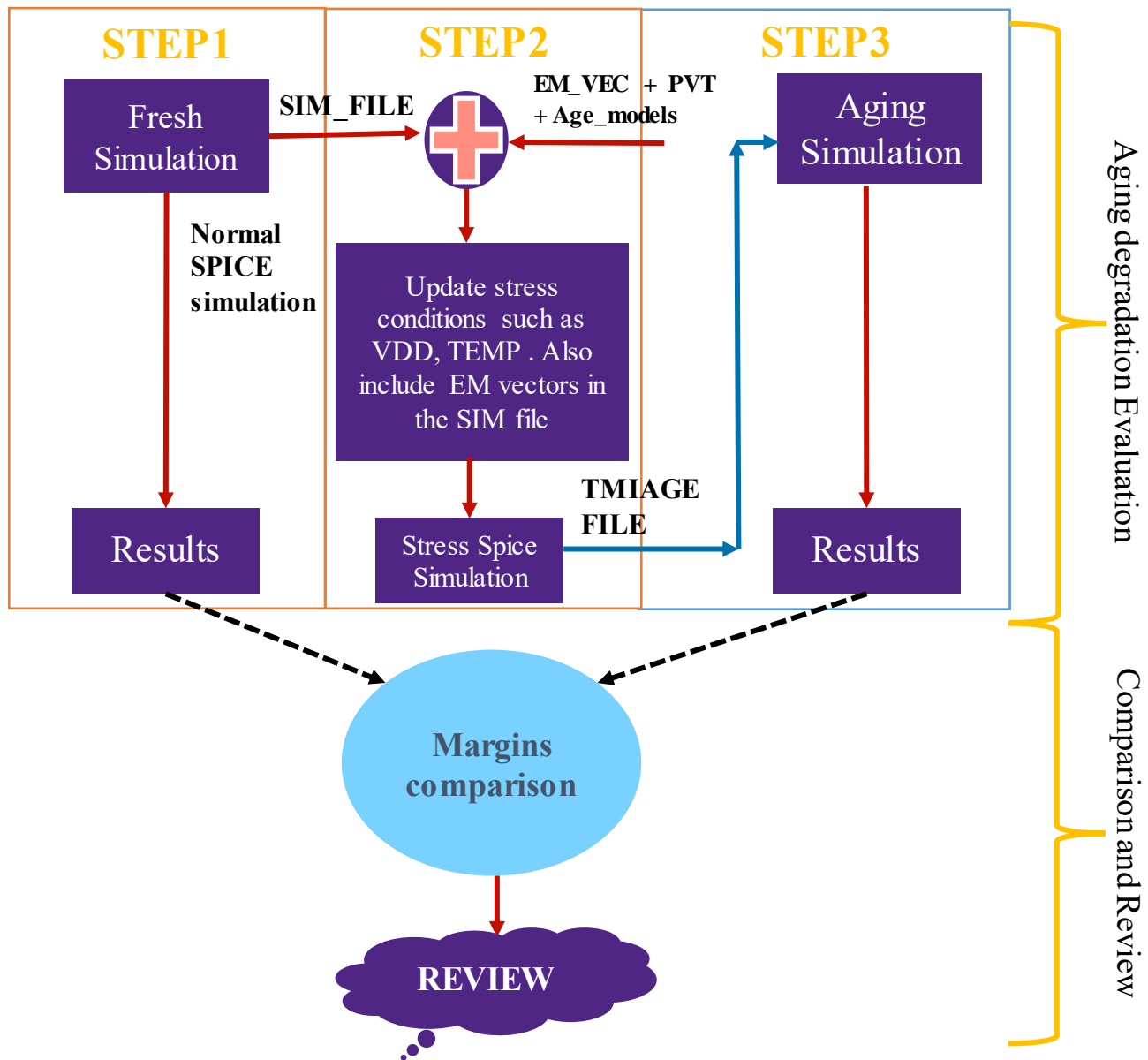


Figure 24 Ageing Methodology flow

Figure 24 indicates the flow for Ageing margin analysis. Following are the steps in details:

Step 1: Run fresh Simulation

Step 2: Stress Simulation (Following updates are done in sim file)

- Use ageing models

- Use max voltage (1.05) and temp(105)
- Exact Timing data from old measurements
- Disable all other function key, as we are interested in read and write operation.
- After simulation Read and write operation should be verified using waveform.

Step 3: Ageing Simulation: (similar to step 1 with following update)

- Add number of the devices in instance netlist in sim file
- Include tmiage file in sim file
- Add year number at which we have consider ageing effect (like 10 years)

5.4 Effect of ageing models on signals(cycles)

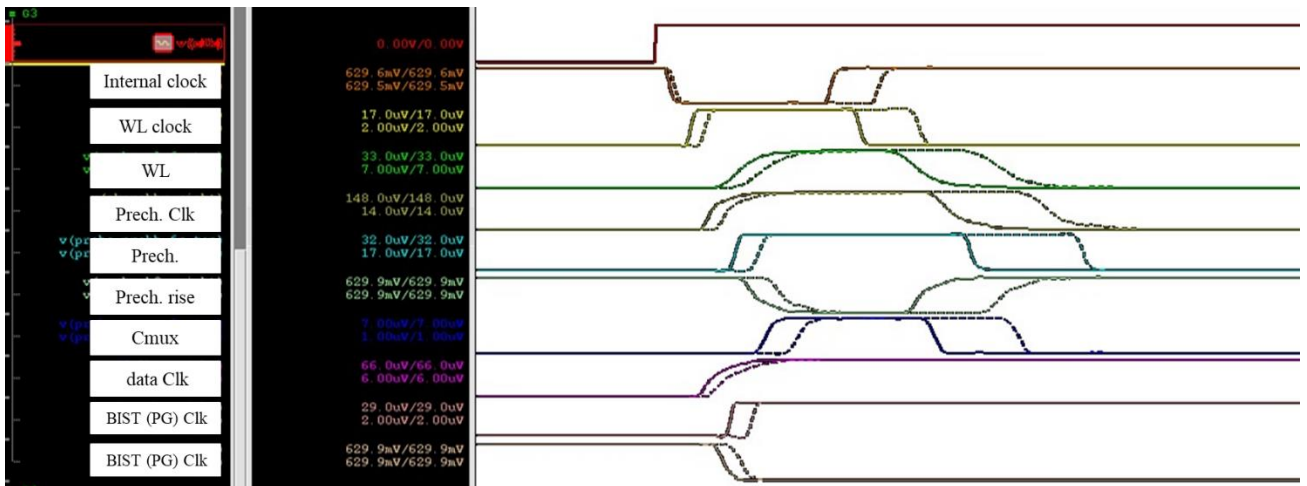


Figure 255 waveforms of most important signals of SRAM (Read and write operation)

From the waveform we can observe that non-uniform degradation at the starting and ending of the cycle. Less degradation at the START and more degradation at the END of the cycle

Chapter 6

Conclusion and future work

This project report explains the yield improvement techniques. As in digital design we have timing constraints for different logic paths. As an example one signal must return earlier than another selected signal to same logic for correct functionality. So for proper functionality we should pass this timing constraints by sufficient margin.

So margin analysis for improving read operation and write operation is most important task, which will further effect on post silicon so that yield is increased.

Till now in my project I have covered Read margin and write margin, which is the most important margin for read and write operation.

As we know the effect of ageing is very important in any VLSI design circuits so in my Project II (final) I did analysis in Ageing PVT and observed which part of memory is getting worst due to ageing effect and I have also discussed the flow of this analysis. Here the comparison of margins is done i.e. Normal simulation margins is compared with ageing PVT margins.

My future work will be on Logic margin/Design margin like timing margin, BIST margin, Power Gating margin. This margin tells us about all the signals which are set and in proper form due to which setup and hold violation doesn't occur, which will affect data during read and write operation and indirectly also affects yield by SRAM.

Bibliography

- [1] Sung-Mo-Kang and Yusuf Leblebici, “CMOS Digital Integrated Circuits Analysis and Design” *Reference Book, 4th Edition*
- [2] K. Bartleson, T. Wood, R. Goldman, “Synopsys’ Educational Generic Memory Compiler” *Microelectronics Education (EWME), 10th European Workshop, August 2014*
- [3] R. Ruchi , Sudeb Dasgupta, “Compact Analytical Model to extract Write Static Noise Margin (WSNM) for SRAM Cell at 45nm & 65nm nodes” *IEEE Transactions on Semiconductor Manufacturing Volume: PP, Issue: 99 , November 2017*
- [4] Ayon Manna, V S Kanchana Bhaaskaran “Improved read noise margin characteristics for single bit line SRAM cell using adiabatically operated word line”, *Nextgen Electronic Technologies: Silicon to Software (ICNETS2), 2017 International Conference, October 2017*
- [5] Nan Zheng, Pinaki Mazumder, “Modeling and Mitigation of Static Noise Margin Variation in Subthreshold SRAM Cells”, *IEEE Transactions on Circuits and Systems I: Regular Papers Volume: 64, Oct. 2017*
- [6] Meng-Fan Chang, Chien-Fu Chen, Ting-Hao Chang, “A Compact-Area Low-VDDmin 6T SRAM With Improvement in Cell Stability, Read Speed, and Write Margin Using a Dual-Split-Control-Assist Scheme”, *IEEE Journal of Solid-State Circuits Volume: 52, Issue: Sept. 2017*
- [7] Chunyu Peng, Songsong Xiao, Wenjuan Lu, “Average 7T1R Nonvolatile SRAM With R/W Margin Enhanced for Low-Power Application”, *IEEE*

*Transactions on Very Large Scale Integration (VLSI) Systems Volume: PP, Issue:
99, December 2017*

[8] Synopsys Internal Material