NOISE MARGIN, CRITICAL CHARGE AND POWER-DELAY TRADEOFFS FOR SRAM DESIGN SPACE EXPLORATION

by

ARAVIND RAJENDRAN

Submitted in partial fulfillment of the requirements

for the degree of Master of Science

Thesis Adviser: Dr. Christos Papachristou

Department of Electrical Engineering and Computer Science

CASE WESTERN RESERVE UNIVERSITY

August, 2011

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis/dissertation of

Aravind Rajendran

candidate for the Master of Science degree *.

(signed) Dr. Chriostos Papachristou
(chair of the committee)

Dr. Frank Merat

Dr. Daniel Saab

(date) June 2, 2011

*We also certify that written approval has been obtained for any proprietary material contained therein.

To my parents and brothers

Contents

Contents iv

List of Tables vi

List of Figures vii

Nomenclature viii

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Contribution . . . 4
- 1.3 Outline . . . 5

2 Background 6
- 2.1 SRAM Architecture . . . 7
  - 2.1.1 SRAM Cell . . . 8
    - 2.1.1.1 Read Operation . . . 9
    - 2.1.1.2 Write Operation . . . 10
    - 2.1.1.3 Hold Operation . . . 11
  - 2.1.2 Address Decoders . . . 11
  - 2.1.3 Precharge Circuit . . . 14
  - 2.1.4 Sense Amplifiers . . . 14
  - 2.1.5 Write Drivers . . . 17
- 2.2 SRAM Stability . . . 18
  - 2.2.1 Static Noise Margin . . . 19
    - 2.2.1.1 Measurement of SNM . . . 22
  - 2.2.2 Soft Errors . . . 28

3 Test Methodology 32
- 3.1 Stability Improvement - Approach . . . 32
- 3.2 Test circuits . . . 35
  - 3.2.1 6T-SRAM . . . 35
  - 3.2.2 SRAM-C . . . 35

3.2.3 SRAM-T 37
3.2.4 SRAM-NSP 37
3.3 Transistor Sizing 38
3.4 Simulation Setup 40

4 Results **44**
4.1 6T-SRAM 44
4.2 SRAM-C 46
4.3 SRAM-T 47
4.4 SRAM-NSP 48

5 Summary **50**

Bibliography **51**

List of Tables

3.1 Characteristics of a 6T-SRAM cell with a variation in β-ratio in 32 nm Technology 33
3.2 Widths of transistors for minimum size SRAM cells 39

4.1 Performance characteristics of 6T-SRAM cell in 32 nm technology . . 45
4.2 Performance characteristics of SRAM-C cell in 32 nm technology . . 46
4.3 Performance characteristics of SRAM-T cell in 32 nm technology . . 47
4.4 Performance characteristics of SRAM-NSP cell in 32 nm technology . 48

List of Figures

1.1 Memory Hierarchy 2
1.2 SRAM cell scaling trend 3
1.3 Characteristics of SRAM in 32 nm Technology 5

2.1 Standard SRAM Cell 8
2.2 Read Operation 10
2.3 Write Operation 11
2.4 Circuit diagrams of a two-input AND gate 12
2.5 Multi-stage row decoder architectures 13
2.6 Pre-charge Circuits 14
2.7 Sense Amplifier Circuits 15
2.8 Write Driver Circuits 17
2.9 Worst-case Noise sources in SRAM cell 19
2.10 Static Noise Margin 20
2.11 VTCs of SRAM cell in the read mode and in the standby mode . . . 22
2.12 Butterfly Curve [1] 23
2.13 Circuit implementation of Equations 2.6 and 2.9 for finding the diagonal of the maximum embedded square in the two lobes of the VTCs [1] 25
2.14 N-Curve Simulation Setup 26
2.15 N-Curve 27
2.16 Soft Error Mechanism 28

3.1 Standard six transistor SRAM cell (6T-SRAM) 36
3.2 Capacitor based SRAM cell(SRAM-C) 36
3.3 Tristate SRAM cell (SRAM-T) 38
3.4 Tristate SRAM cell (SRAM-NSP) with separate read port 39
3.5 Butterfly curve simulation set up 41
3.6 SEU simulation setup 42
3.7 Simulation Flow 43

Nomenclature

6T-SRAM	Six Transistor Static Random Access Memory
ASIC	Application Specific Integrated Circuits
BL	Bit-line
BPTM	Berkeley Predictive Technology Model
CMOS	Complementary Metal-Oxide Semiconductor
DRAM	Dynamic Random Access Memory
DWL	Divided Word-line
ECC	Error Correction Codes
HWD	Hierarchical Word Decoding
RAM	Read/Write Random Access memory
RBL	Read-Bit-line
ROM	Read-only Random Access memory
RWL	Read-Word-line
SER	Soft Error Rate
SEU	Single Event Upset
SINM	Static Current Noise Margin
SNM	Static Noise Margin
SRAM	Static Random Access Memory
SRAM-C	SRAM with Capacitor Buffer

SRAM-NSP	Tri-State SRAM with Capacitor Buffer and separate read port
SRAM-T	Tri-State SRAM
SVNM	Static Voltage Noise Margin
VTC	Voltage Transfer Characteristics
WE	Write Enable
WL	Word-line

Noise Margin, Critical Charge and Power-Delay Tradeoffs for SRAM Design Space Exploration

Abstract

by

ARAVIND RAJENDRAN

Aggressive technology scaling in semiconductor devices has resulted in stability reduction for classic SRAM designs. This is especially problematic for large integrated circuits. The stability of SRAM cells can be affected by noise during a read operation and radiation during the standby mode. In this work, an approach to address the gradual stability reduction in SRAM designs is presented. An SRAM design tradeoffs approach is presented which improves the characteristics of SRAM by modulating the transistor sizing ratio, β. This method was tested on various SRAM designs in 32 nm technology. SRAM designs were optimized for various constraints in power consumption, performance, radiation tolerance and data stability. Different design trends produced by the extensive approach analysis is discussed in this thesis.

Chapter 1

Introduction

Static random access memories (SRAM) have been used as on-chip memories in high performance integrated circuits, due to its high access speed and compatibility with process and supply voltage. The demand for high performance due to aggressive CMOS technology scaling has increased the amount of on-chip memory integrated into modern semiconductor devices. The total area occupied by these memories has been rapidly increasing and reached over 70% [2]. The continued scaling of CMOS technology has also resulted in problems which were less severe in earlier generations. These include process induced variations, soft errors, transistor degradation mechanisms etc.

1.1 Motivation

Modern semiconductor devices demand more on-chip memory to meet the performance needs. But, it is not possible to integrate all the needed memory into the chip. Moreover, the cost per bit of fast embedded memories is high and requires more area compared to other memories. The speed gap between processors and memories has

resulted in the introduction of memory hierarchy into the processor architectures.

A typical memory hierarchy of a modern computer system is shown in Figure 1.1. SRAMs dominate the memory hierarchy in performance but they are often integrated in a lesser capacity due to the area limitations and the high cost per bit. It is also evident that the memory hierarchy is a fundamental consequence of maintaining the random access memory abstraction and practical limits on the cost and the power consumption.

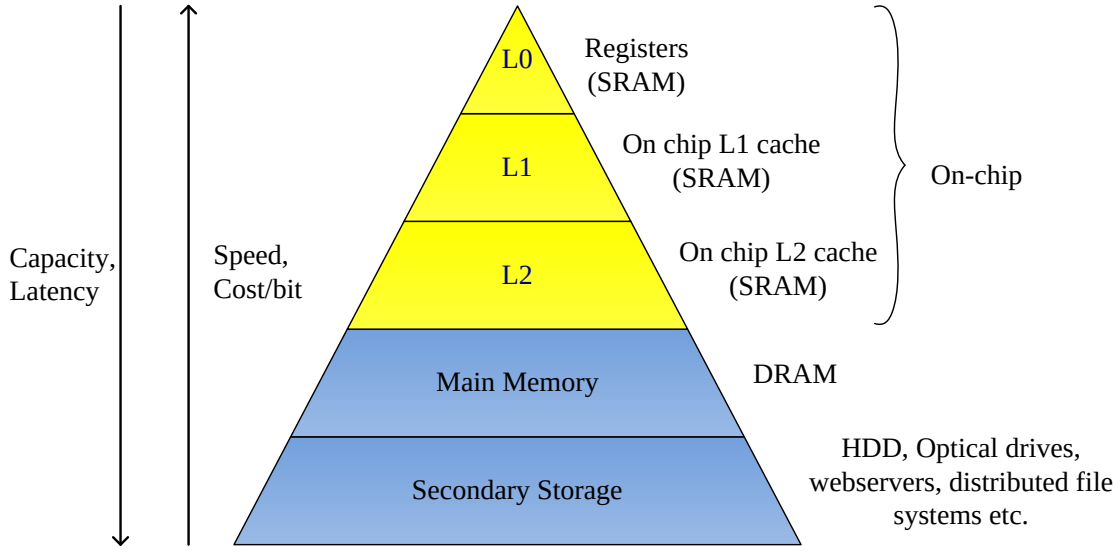


Figure 1.1: Memory Hierarchy

As the technology scales deeper into nanometer levels, the stability of SRAM to noise and radiation is reduced. It is becoming increasingly challenging to maintain an acceptable static noise margin (SNM) of SRAMs while scaling the minimum feature sizes and supply voltage [3]. Static noise margin (SNM) degradation, which characterizes the data integrity of SRAM during a read operation, has driven the development of SRAM cell design in to new direction as the supply voltage reaches near the threshold voltage. Moreover, the shrinking of the transistor dimensions has also increased the probability of radiation induced errors.

Figure 1.2 shows the SRAM scaling trend as the technology reaches deep into the nanometer level [2, 3]. It is evident that the cell area is ever shrinking at a rate of 0.5× per technology and reached 0.171 μm^2 at 32 nm node. At these dimensions SRAM is more susceptible to soft errors and even the ground level radiations can affect the data stored in the memory [4]. Also, SRAM is densely packed to achieve more memory per unit area efficiency. This increase the probability of the radiation induced error on an SRAM cell compared to any other logic inside the IC.

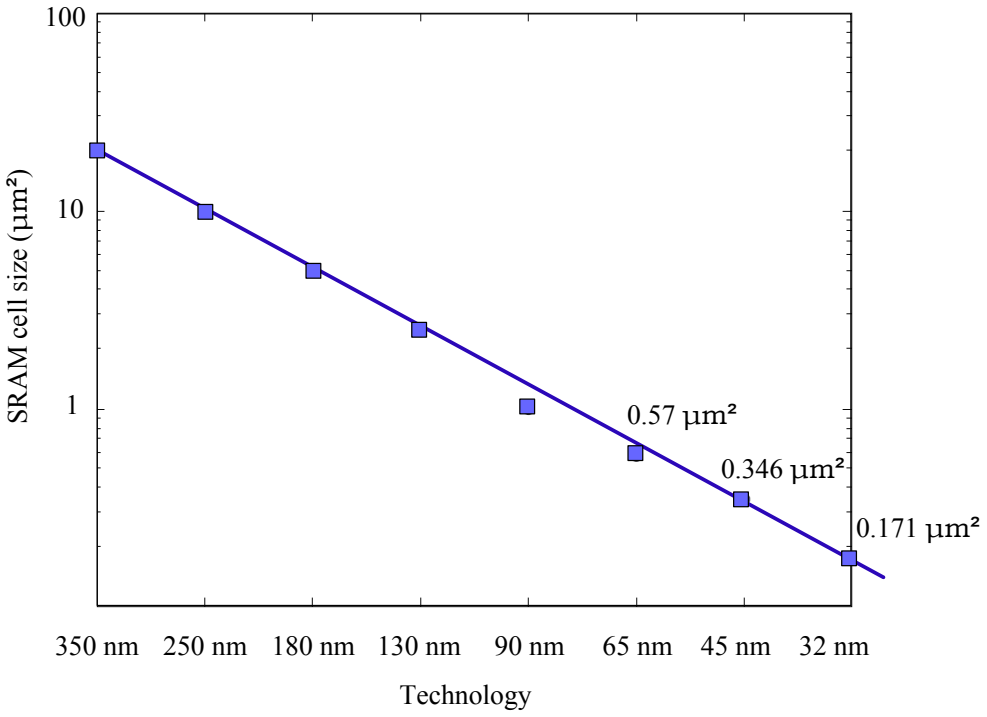


Figure 1.2: SRAM cell scaling trend

The continued scaling of supply voltage (V_{dd}) and minimum feature sizes have created challenges in designing stable and reliable embedded data storage. SRAM stability is projected to decrease by 4×, as CMOS technology scales from 250 nm to 45 nm. SRAM data stability depends on supply voltage, threshold voltage and SRAM sizing ratios. Process induced variations like random dopant fluctuations in

channel, variations in transistor dimensions induce variations in threshold voltage V_T in scaled down transistors [5, 6]. Such variations induce problems in designing identical devices particularly the SRAM cells with millions of bits integrated into the IC.

1.2 Contribution

In order to improve the overall performance of large systems, large arrays of minimum sized SRAMs are often integrated into the chip. However, such method will effect the reliable operation of the memory cells. These reliability issues have resulted in design constraint relaxation in terms of overall area. In this work, an SRAM design tradeoffs approach is presented to improve its performance characteristics by modulating the transistor sizing ratio, β. This approach is applied for different SRAM cells to produce good tradeoff driven by the following parameters: SNM, critical charge, write time delay and power consumption. Furthermore, the SNM is optimized while satisfying the other design constraints.

The performance characteristics of a regular SRAM cell in 32 nm technology are presented in Figure 1.3. The plot shows the dependence of SNM to the β-ratio (pull-down to access transistor sizing ratio) of the SRAM. Although the SRAM cell performs well in the minimum dimensions, it has disadvantages in terms of SNM and critical charge, Q_{crit}. The SRAM cells shows improved SNM and Q_{crit} at a higher β-ratio.

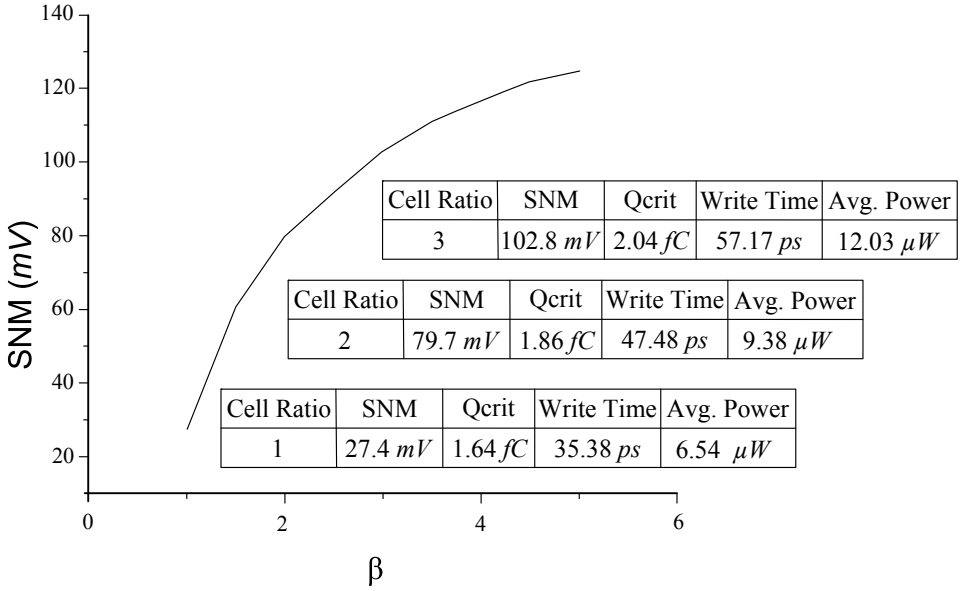


Figure 1.3: Characteristics of SRAM in 32 nm Technology

1.3 Outline

The thesis is organized as follows. In chapter 2. background information on SRAM architecture and data stability of SRAM cells are introduced. Stability improvement approach and simulation setup are presented in chapter 3. In chapter 4. the stability improvement approach is verified from the generated performance characteristics of SRAM cells. Finally in chapter 5. the thesis is concluded with a summary and a discussion of possible future work.

Chapter 2

Background

The stability of an SRAM cell is an important functional constraint in nanometer technologies as it determines the ability to retain stored information. This chapter presents the data stability and an introduction to memory structures with a focus on embedded SRAM. Next section describes the various methods to estimate the stability of the memory cell during a read access and the factors that affect the stability. Final section describes about the radiation induced soft errors in semiconductor memory devices.

A random-access memory is a class of semiconductor memory in which the stored data can be accessed in any order and at uniform time regardless of the physical location. Random-access memory is commonly classified as read-only memory (ROM) and read/write memory. Read/write random-access memories are generally referred to as RAM. Random-access memory is also classified based on the storage mode of the memory: volatile and nonvolatile memory. Volatile memory retains its data as long as power is applied, while nonvolatile memory will hold data indefinitely. RAM is synonymous with volatile memory, while ROM is synonymous with nonvolatile memory.

Memory cells used in volatile memories can be further divided into static structures or dynamic structures. Static RAM (SRAM) cells use feedback mechanism to maintain their state, while dynamic RAM (DRAM) cells use charge stored on a floating capacitor to hold the data. The charged stored in the capacitor is leaky, so dynamic cells must be refreshed periodically. The positive feedback between two complimentary inverters in SRAM provides a stable data and facilitates high speed read and write operations. Although SRAMs are faster, it requires more area per bit than DRAMs.

2.1 SRAM Architecture

An SRAM consists of an array of memory cells along with peripheral circuits, which enable reading from and writing into the array. A typical SRAM memory architecture is shown in Figure 2.1. The memory array consists of 2^n words of 2^m bits each. Each bit is stored in one memory cell. They share a common word-line (WL) in each row and a bit-line pairs ($BL, \overline{BL}$) in each column. The dimension of each SRAM array is limited by its electrical characteristics such as capacitances and resistances of the lines used to access cells in the array. Therefore, larger memories may be folded into multiple blocks with fewer rows and columns. After folding, each row of the memory contains 2^k words, so the array is physically organized as 2^{n-k} rows and 2^{m+k} columns. Every cell can be randomly addressed by selecting the appropriate word-line (WL) and bit-line pairs ($BL, \overline{BL}$), respectively, activated by the row and the column decoders.

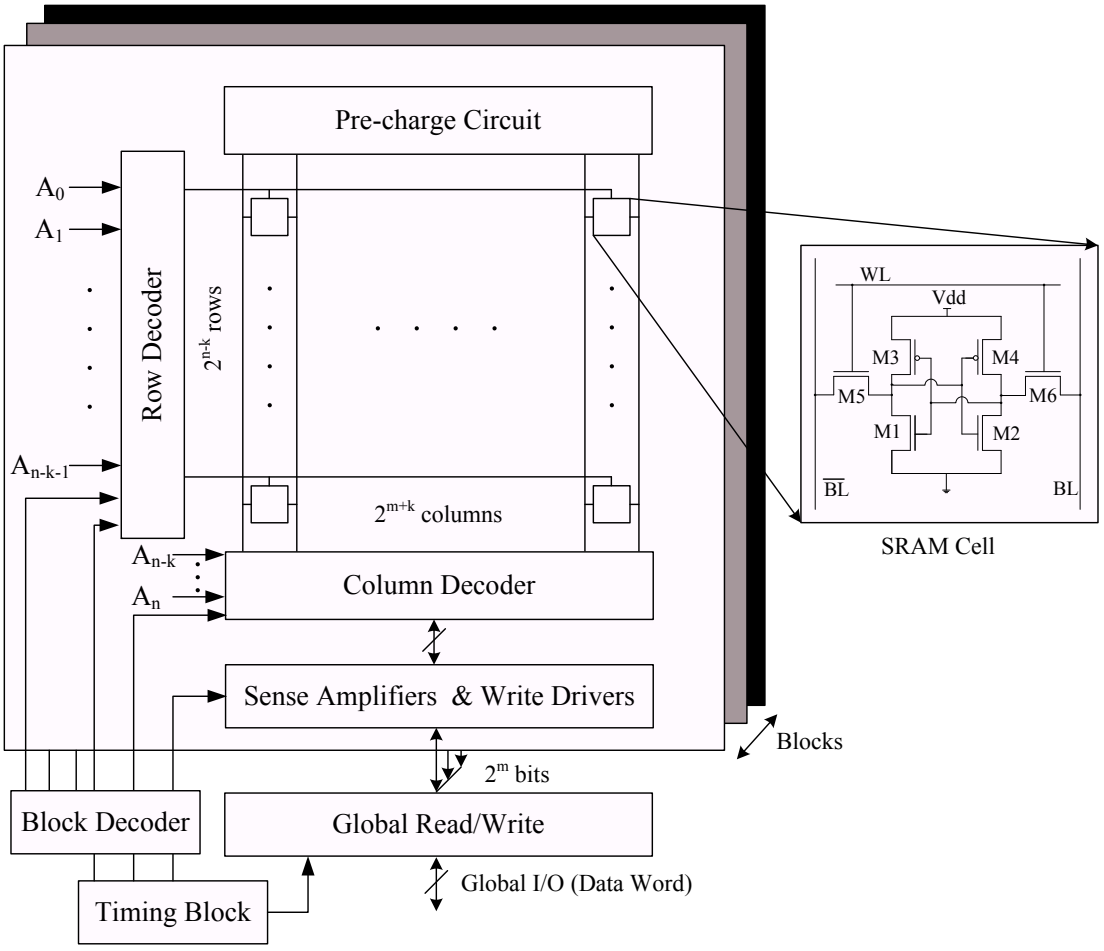


Figure 2.1: Standard SRAM Cell

2.1.1 SRAM Cell

An SRAM cell is the fundamental building block of the SRAM. Each cell holds one bit of information. It provides non-destructive reading, write capability and data storage as long as the SRAM cell is powered up. The main constraints in designing an SRAM cell are cell area, robustness, speed, power consumption and yield. The overall area of the SRAM cell has been a main constraint for designing the cell, as it helps to improve the performance and power consumption due to the reduction in cell capacitance. Moreover, reducing the size of the SRAM devices yields more bits

per unit area.

A regular six transistor SRAM (6T-SRAM) cell consists of two cross coupled inverters and access transistors on both storage nodes. The inverters (*M1, M3* and *M2, M4*) form a latch and holds the binary information. True and complimentary versions of the binary data are stored in storage nodes. The access transistors (*M5*, *M6*) allow access to the binary information during read and write operations and also provides isolation from the other circuits during hold state. The cells are accessed by asserting the word-line (WL) during a read or write operation.

An SRAM cell has three modes of operation: read, write and standby. In other words, it can be in three different states namely reading, writing or data retention. The minimum requirements of SRAM cell in different modes is presented below.

2.1.1.1 Read Operation

Figure 2.2 shows the 6T-SRAM during a read operation, where bit-lines are pre-charged to V_{dd} before the read operation by the bit-line load transistors. The read operation is initiated by enabling the word-line (WL) and thereby connecting the internal nodes of the SRAM cell to bit-lines (BL and $\overline{BL}$). The bit line voltage is pulled down by the NMOS transistor at the '0' storage node and this is detected by the sense amplifier.

Assume, the nodes *N1* and *N2* initially store values *'1'* and *'0'* respectively. When the word-line (WL) is asserted, the bit-line (BL) is pulled down through transistors *M1* and *M5*, forming a voltage divider. The node *VR* will no longer be in zero potential, due to the current flowing through *M5*, and it goes above *0 V*. The node potential should stay below the switching threshold of the inverter to avoid a destructive read. The rise in node potential depends on the sizing of the access transistor and the

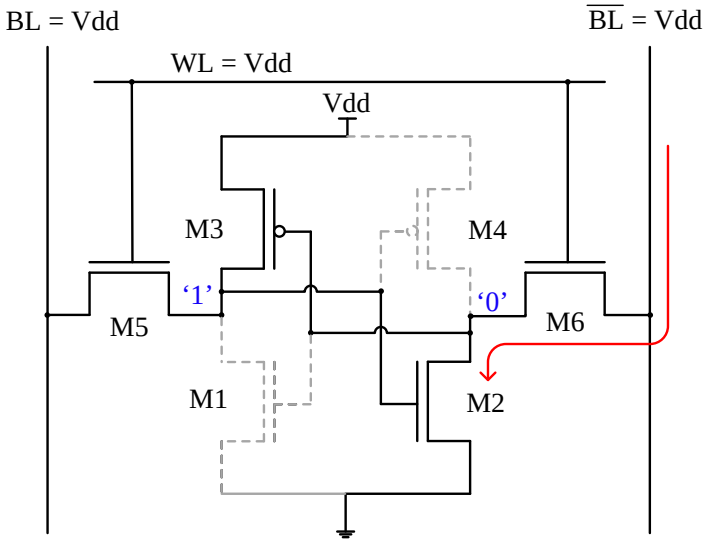


Figure 2.2: Read Operation

pull-down transistor defined by Cell Ratio as

$$CellRatio = \beta = \frac{W_2/L_2}{W_6/L_6} \tag{2.1}$$

2.1.1.2 Write Operation

The write cycle begins by forcing a differential voltage (V_{dd} , GND) at the BL pairs. This differential voltage corresponds to the data to be written at the storage nodes and it is controlled by the write drivers. The WL is then activated to store the information from the bit-line pairs to corresponding storage nodes.

Assume, the nodes VL and VR initially store values *'1'* and *'0'* respectively. When the WL is asserted the access transistor connected to BL is turned on, a current flows from V_{dd} to BL through $M3$ and $M5$. This current flow lowers the potential at VL. The potential at the node VL has to go below the trip point of the inverter for a successful write operation and this depends on the ratio of pull-up transistor ($M3$) and the access transistor ($M5$). This ratio is referred to as the γ - ratio. Since

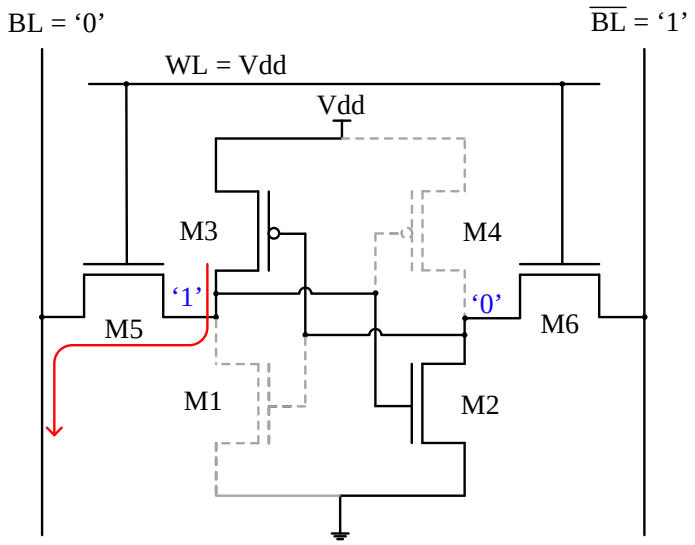


Figure 2.3: Write Operation

the pull-down transistor ($M2$) is designed to prevent data from flipping during a read access, the data has to be written through the access transistors and pull-up transistor.

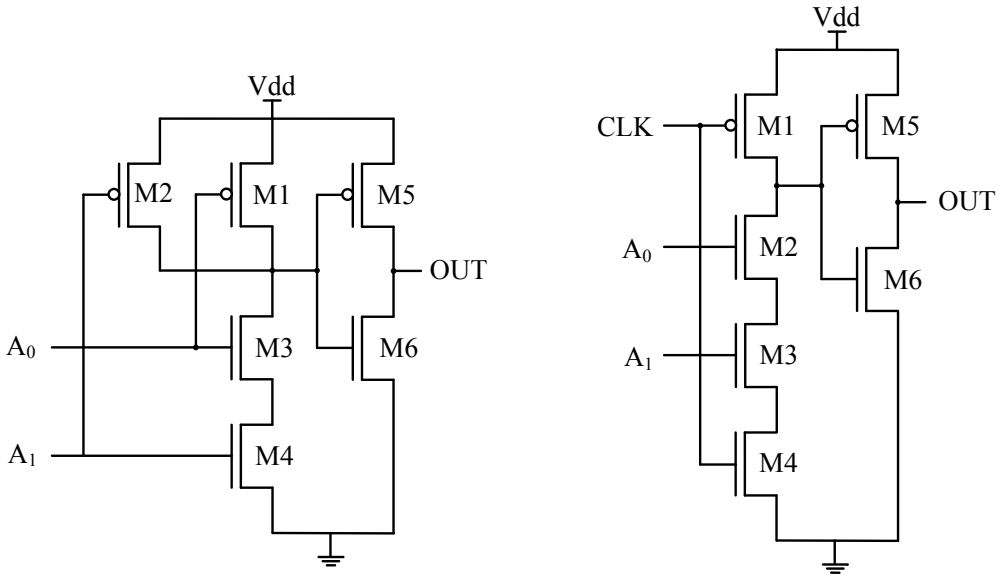
2.1.1.3 Hold Operation

When WL is not active, SRAM cell is in standby or data retention mode. The cross-coupled inverters will hold the data, through bistable action, as long as a sufficient power is applied to the cell. However, when V_{dd} gets lower than a certain voltage point the inverters will no longer be able to hold the data correctly. This voltage is called the data retention voltage (V_{hold}) of the SRAM cell.

2.1.2 Address Decoders

Address decoder is a matrix of logic elements that selects a row or a column of memory based on the input memory address. It also allows the number of interconnects in the SRAM to be reduced by a factor of $\log_2 N$, where N is the number of independent

addresses locations. There are two kinds of decoders used in SRAM, namely, row decoder and the column decoder. Row decoders are needed to select one row of word-lines out of a set of rows in the array. A fast decoder can be implemented by using AND/NAND and OR/NOR gates. Figure 2.4 shows the schematic diagrams of static and dynamic AND gate decoders. The static NAND-type structure is chosen due to its low power consumption during the decoded row transitions. The dynamic structure is chosen due to its speed and power improvement over conventional static NAND gates. Column decoders select the desired bit pairs out of the sets of bit pairs in the selected row. A typical dynamic AND gate decoder can be used for column decoding because it meets the delay requirements (column decode is not in the worst-case delay path) at a lower power consumption.



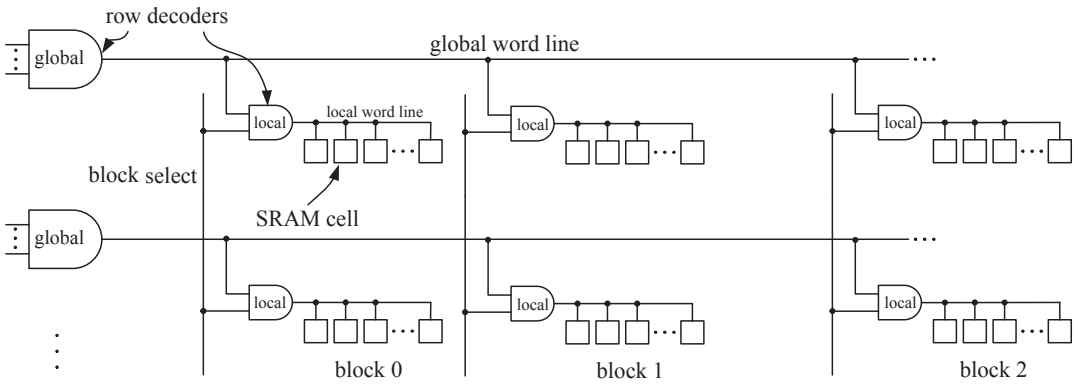
(a) Static CMOS

(b) Dynamic CMOS

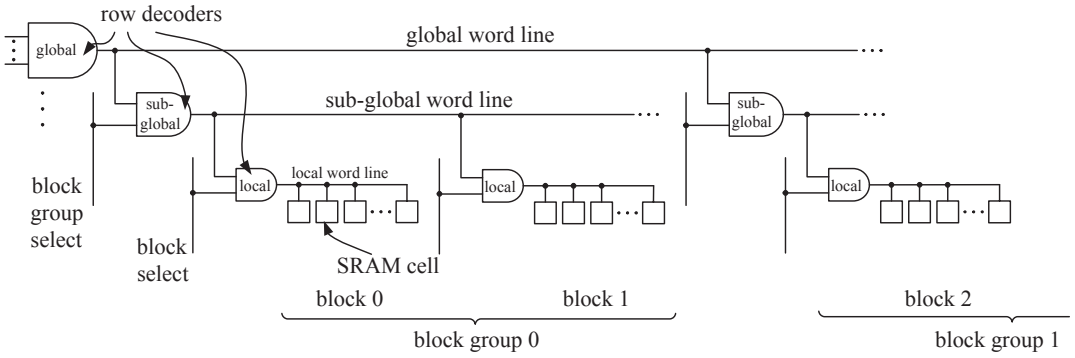
Figure 2.4: Circuit diagrams of a two-input AND gate

Single-stage decoders are attractive for small single-block memories. However, a highly integrated SRAM adopts a multi-divided memory cell array structure with a

multi-stage decoding scheme is used to achieve high-speed word decoding and reduce column power dissipation. The multi-stage decoder circuit has advantages over the one-stage decoder in reducing the number of transistors, fan-in and the loading on the address input buffers.



(a) Divided word-line (DWL) [7] structure



(b) Hierarchical Word Decoding (HWD) [8] structure

Figure 2.5: Multi-stage row decoder architectures

Figure 2.5(a) shows a typical partitioned memory array with divided word-line (DWL) decoder structure. A local word-line is activated when both the global word-line and block select are asserted. Since only one block is accessed at a time, the DWL structure reduces both the word-line delay and the power consumption [7]. To

reduce the capacitance of word-line in high density SRAM greater than 4Mb, the hierarchical word decoding (HWD) architecture was proposed [8]. The word-line is divided into multiple levels, determined by the total capacitance of the word select line to efficiently distribute it, to reduce delay and power.

2.1.3 Precharge Circuit

The primary function of the precharge circuit is to pull-up the bit-lines to V_{dd} levels before an SRAM operation (read or write). A simple precharge circuit consists of a pair of PMOS transistors, as shown in Figure 2.13(a). A clocked precharge circuit, shown in Figure 2.13(b) or 2.6(c), can be used to reduce the power consumption of the circuit.

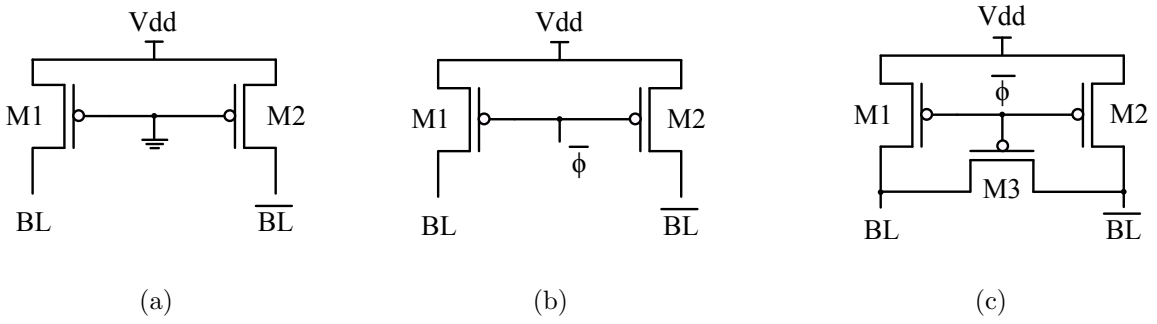


Figure 2.6: Pre-charge Circuits

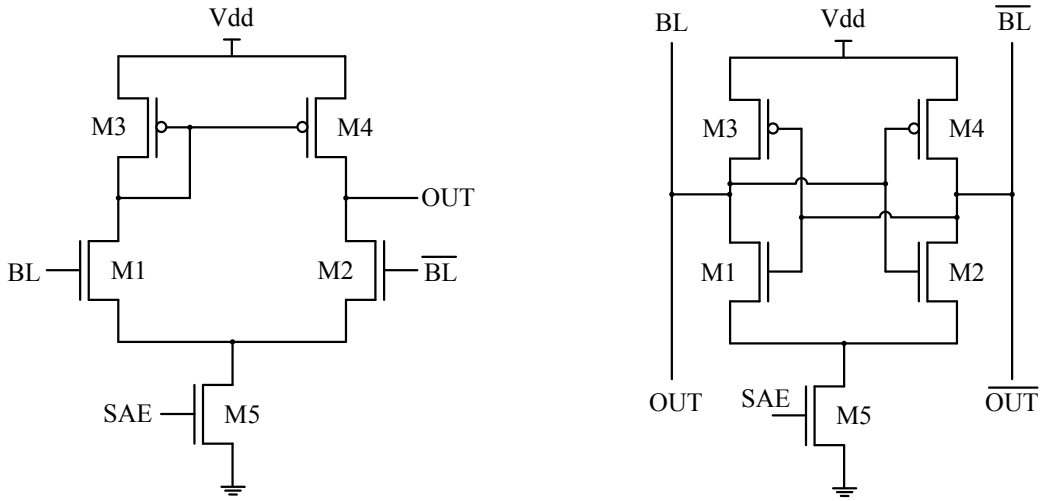
2.1.4 Sense Amplifiers

A sense amplifier detects the contents of the selected cell by amplifying a small analog differential voltage developed on the pre-charged bit-lines BL and $\overline{BL}$ during a read access. During the read cycle, one of the pre-charged bit-lines is pulled down by NMOS transistor of one of the inverters through the access transistor. The bit-line

pull down speed is slow due to the small cell size and large bit-line load capacitance. The high sensitivity of the sense amplifier (50 to 100mV) allows a faster data access.

The choice and design of a sense amplifier define the robustness of bit line sensing which impacts the read speed and power. Since SRAMs do not feature data refresh after sensing, the sensing operation must be nondestructive, as opposed to the destructive sensing of a DRAM cell. A sense amplifier also allows the storage cells to be small, since each individual cell need not fully discharge the bit line.

A sense amplifier is characterized by the parameters gain A, sensitivity S, offsets V_{off} and I_{off}, common mode rejection ratio $CMRR$, rise time t_{rise}, fall time t_{fall}, and sense delay t_{sense}. The design of a sense amplifier depends on the timing and layout constraints of the memory system. To alleviate the problems of process induced variations sense amplifiers often employ devices with non minimum length and width.



(a) Current-mirror type

(b) Latch-type

Figure 2.7: Sense Amplifier Circuits

A classical current-mirror differential sense amplifier is shown in Figure 2.7(a). The sensing operation starts with setting the operating point of the sense amplifier

by precharging both inputs to the identical voltage levels through the bit-lines. Once both bit-lines are precharged and equalized, the voltage levels are stored in the bit-line load capacitance. When the differential voltage at bit-lines exceeds the sensitivity of amplifier due to a read operation, it is activated by a sense amplifier enable (SAE) signal. Then, the sense amplifier amplifies the differential voltage to a logic voltage level. The gain, A, of the current-mirror sense amplifier is defined by

$$A = -g_{mM1}(r_{o2}||r_{o4}) \tag{2.2}$$

where g_{mM1} is the transconductance of transistor $M1$, and (r_{o2} and r_{o4}) are the small signal output resistance of $M2$ and $M4$, respectively. The gain can be increased by widening $M1$ and $M2$ or by increasing the biasing current.

A latch-type sense amplifier, shown in Figure 2.7(b), is formed by two cross coupled inverters. In this type of sense amplifier, the sense operation starts with biasing it in the high gain metastable state by precharging and equalizing its inputs. The transistor, $M5$, isolates the sense amplifier from bit-lines and prevents the full discharge of bit-line on the '0' storage node. A local precharge circuit may be used if the sense amplifier is completely isolated from the bit-lines by additional pass through transistors. After the differential voltage developed on the bit-lines exceeds the sensitivity of the sense amplifier, SAE signal is enabled and the bit-line isolation pass through transistors are turned off. Such separation helps to prevent the complete discharge of bit-line load capacitance. The feed back mechanism of the latch-type sense amplifier quickly picks up the differential voltage and drives the outputs to the full swing differential voltages.

2.1.5 Write Drivers

A write driver in SRAM quickly discharges one of the bit-lines from the precharge level to below the write margin of the cell. Normally, the write driver is enabled by the Write Enable (WE) signal and drives the bit line using full-swing discharge from the precharge level to ground. The order in which the word line is enabled and the write drivers are activated is not crucial for the correct write operation.

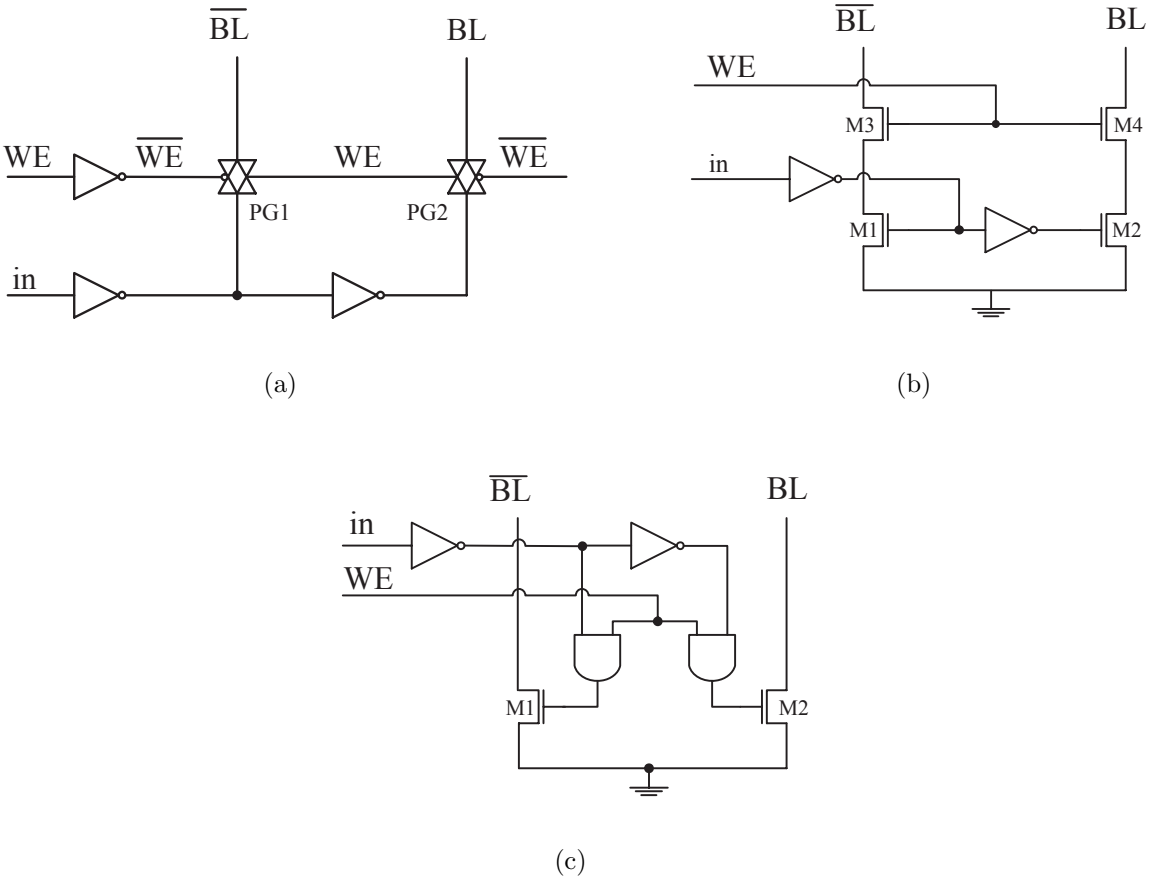


Figure 2.8: Write Driver Circuits

Some of the typical write driver circuits are presented in Figure 2.8. The circuit in Figure 2.8(a) uses two transmission gates *PG1* and *PG2* to write the input data, *in*, and its complement buffered by inverters to the bit lines BL and $\overline{BL}$. *PG1*

and *PG2* are activated by *WE* and its complementary $\overline{WE}$. BL or $\overline{BL}$ is discharged through the NMOS transistors in either inverter. The write driver presented in Figure 2.8(b) writes data through two stacked NMOS transistors, i.e., $M1, M3$ and $M2, M4$, which form two pass-transistor AND gates. The NMOS transistors *M3* and *M4* are activated by *WE* while the data *in* enables the transitor *M1* or *M2* through the inverters. When *WE* is enabled, BL or $\overline{BL}$ is discharged from the precharge level to the ground level through one of the transistors *M1* or *M2*. Another implementation of the write driver is presented in Figure 2.8(c). When *WE* is asserted, depending on the input data *in*, one of two AND gates is activated to turn on one of the pass-transistors *M1* or *M2*. Then, the corresponding bit-line discharges to the ground level through the pass-transistor.

Even though a greater discharge of the highly capacitive bit lines are required for a write operation, it can be carried out faster than a read operation. Only one write driver is needed for each SRAM column. Thus, the area impact of a larger write driver is not multiplied by the number of cells in the column and hence the write driver can be sized up if necessary.

2.2 SRAM Stability

Stability and robustness of an SRAM are characterized by its ability to retain stored data. The stability of the memory cell can be affected during read or standby mode. The disturbances produced during the read operation, read access disturbance, affects the cell stability during read mode. During the stand by mode, the stability of SRAM is mainly affected by radiation induced errors.

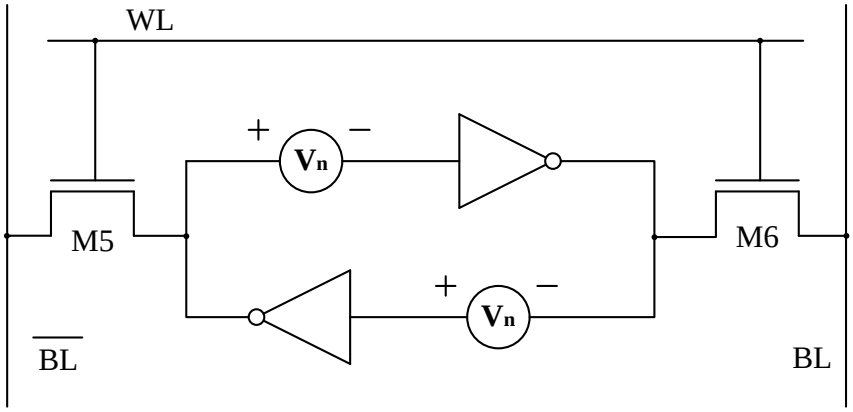


Figure 2.9: Worst-case Noise sources in SRAM cell

2.2.1 Static Noise Margin

The stability of an SRAM cell is a critical functional constraint in nanometer technologies as it determines the ability to retain stored information. The static noise margin (SNM) is a measure of the SRAM stability and it is defined as the maximum static noise voltage that can be tolerated by the SRAM without losing the stored information [9, 10]. In other words, SNM quantifies the amount of noise voltage V_n required to flip the cell data during a read access or standby mode. Figure 2.9 shows an SRAM cell presented as two equivalent inverters with the noise sources inserted between the corresponding inputs and outputs[11, 10]. Both series voltage noise sources (V_n) have the same value and act together to upset the state of the cell. This represents the worst-case SNM value for an SRAM cell [9].

The SNM of an SRAM cell can be represented graphically using the superimposed voltage transfer characteristics (VTC) of the inverters as shown in Figure 2.10. The resulting two-lobed curve is generally referred to as the '*butterfly curve*'. The thick lines in the curve refers to the DC characteristics of the cell for the condition where there is no noise (V_n = 0V). The two crossing points near the axes are the stable points, whereas the center crossing is a meta-stable point. Introduction of noise

sources causes the VTC of the inverters to shift, VTC of the inverter 1 to right and VTC of inverter 2 moves downward. The cell can hold the data as long as there are two lobes in the curve. Once the VTCs have moved away such that they only touch in two locations, one lobe disappears and any further increases in V_n will result in loss of ability to hold data. This value of V_n is the static noise margin. The thin lines on the plot illustrates the VTCs in this condition. They touch at the corner of the largest embedded square inside the lobe of the original butterfly plot.

The SNM is now defined as the length of the side of the largest embedded square inside the butterfly plot. In an ideal SRAM cell, the VTC of both inverters would be symmetrical. However, due to process variations changes in transistor attributes (length, width, oxide thickness, mobility etc.) could result in cell imbalance. If the inverters of cell are not identical, one lobe is smaller than the other. Then, the SNM of the cell is the length of the side of the largest square that fits inside the smallest of the two lobes. This indicates that the bit-cell is more susceptible to losing one particular data value.

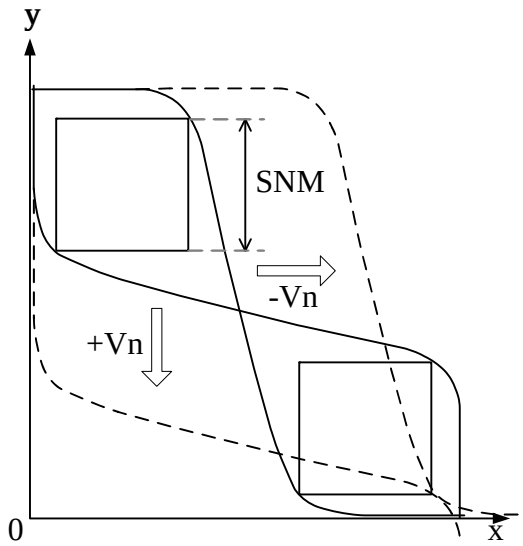


Figure 2.10: Static Noise Margin

Analytical expression of an SRAM cell using the basic MOS model equation with constant V_{th} is given by:

$$SNM_{6T} = V_{th} - \left(\frac{1}{k+1}\right) \times \left(\frac{V_{dd} - \frac{2r+1}{r+1} \times V_{th}}{1 + \frac{r}{k(r+1)}} - \frac{V_{dd} - 2V_{th}}{1 + k\frac{r}{q} + \sqrt{\frac{r}{q}\left(1 + 2k + \frac{r}{q}k^2\right)}} \right) \tag{2.3}$$

where V_{dd} is the supply voltage, q is the loop gain of pull-up transistor with respect to access transistor, r is the loop gain of pull-down transistor with respect to access transistor and k is a parameter depends on loop gain r. This SNM expression is derived without considering the second order effects effects such as mobility reduction and velocity saturation.

The memory cell is most vulnerable to noise during the read access than the hold state. The pre-charged bit-lines, connected to the storage nodes, discharges to ground through the access transistor and the pull-down transistor forming a voltage divider. The current flow during the read access elevates the potential at the '0' storages nodes. The rise in node potential, due to the voltage divider, depends on the strength of the transistors. Since the strength of the transistor is determined by its dimensions, the access transistor and pull-down transistor can be carefully sized to control the rise in the node voltage.

During standby mode the WL is in OFF state and the cell is disconnected from the external circuitry, and hence more immunity to noise. Thus, an SRAM cell has different noise immunity levels during read and hold operation, referred to as "read SNM" and "hold SNM", as shown in Figure 2.11. The dotted lines in the plot represents the DC characteristics of the cell during standby mode while the thick lines represents the VTCs during a read access. Figure 2.11 also shows the change in

the VTC during a read access and the decrease in SNM from standby mode to read access. SNM depends on the sizing ratio of transistors (β), threshold voltage (V_{th}) and supply voltage (V_{dd}). It can also be improved by decreasing the read time or modulating the word-line voltage [12].

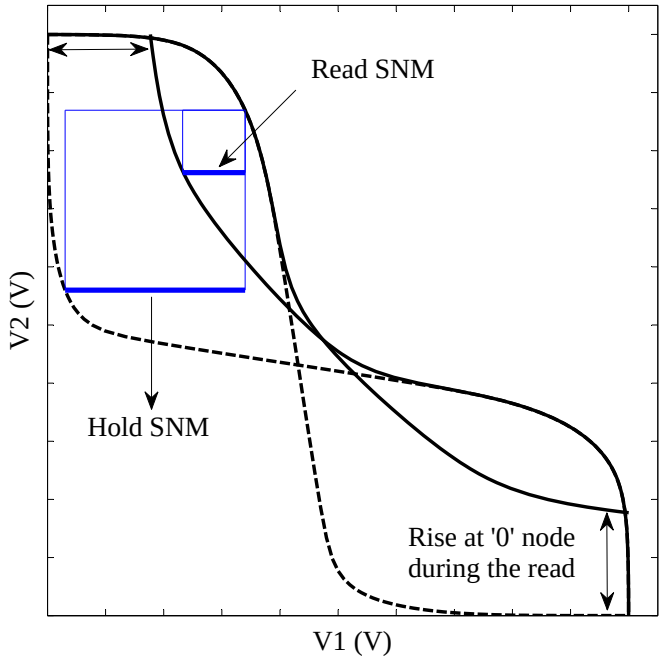


Figure 2.11: VTCs of SRAM cell in the read mode and in the standby mode

2.2.1.1 Measurement of SNM

There has been several methods proposed in the past to measure the SNM of an SRAM cell. The common method to determine the SNM is by extracting the voltage transfer characteristics of its inverters.

Maximum embedded square method

This technique was demonstrated by Seevinck et al. to graphically measure the SNM from the VTCs of the SRAM [10]. The two winged curve generated by this method

is generally referred to as the *'butterfly curve'* and the size of the eyes between the curves is a measure of the SNM [1]. This technique is very popular due to the fact that it can be easily automated using a circuit simulator like SPICE. Graphical method is only applicable to circuits with high input impedance ($R_{in} >> R_{out}$) and CMOS SRAM circuits satisfy this condition.

The VTC of the inverters during the read mode can be extracted by sweeping the voltage at the storage node with $BL, \overline{BL}, WL$ all held at V_{dd}. The butterfly curve is generated by superimposing the inverter VTCs that are inversed from each other as shown in Figure 2.12 . The side of the maximum possible nested square between the curves represents the SNM of that memory cell [10].

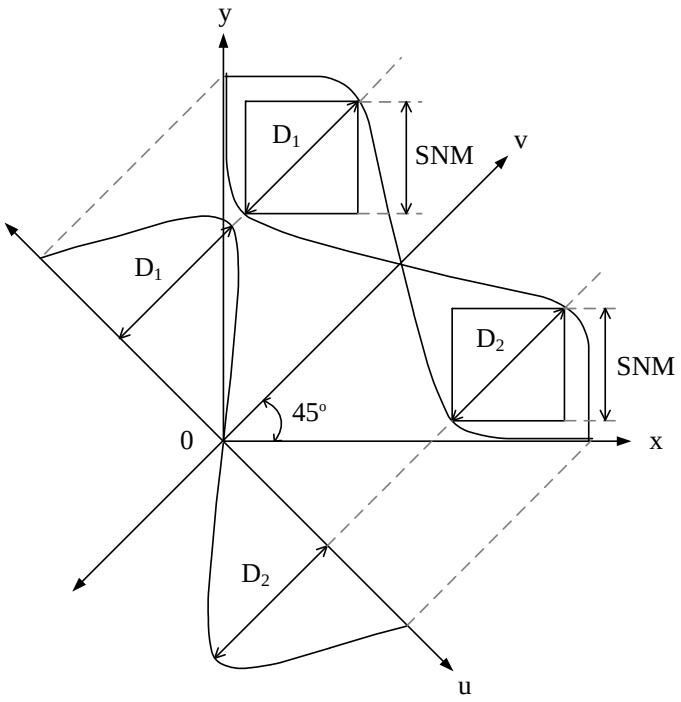


Figure 2.12: Butterfly Curve [1]

The procedure to determine the SNM is as follows. The butterfly cure is represented in x-y coordinate system. The u-v coordinate system is rotated 45° anti-

clockwise to the x-y coordinate system. In this arrangement, the v axis is parallel to the diagonal of the nested squares. Also, the distance between the two VTCs parallel to v axis represent the diagonal of the biggest nested square with respect to u axis. Thus, the peaks of the curve in the u-v coordinate system represent the diagonal of the the maximum possible nested squares in the corresponding "eyes". Due to process variations the inverters are not generally identical and hence slightly different VTCs and thus $D_1 \neq D_2$. Suppose $D_1 > D_2$, then $D_1/\sqrt{2}$ yields the SNM of the SRAM cell.

The mathematical representation of the above mentioned algorithm algorithm can be defined by the functions $y = F_1(x)$ and $y = F_2'(x)$, where $F_2'(x)$ is the mirrored $F_2(x)$. $F_1(x)$, in terms of u and v, can be found by using the equations

$$x = \frac{1}{\sqrt{2}}\, u \; + \; \frac{1}{\sqrt{2}}\, v \tag{2.4}$$

$$y = -\,\frac{1}{\sqrt{2}}\, u \; + \; \frac{1}{\sqrt{2}}\, v \tag{2.5}$$

substituting Equations 2.4 and 2.5 in $y = F_1(x)$ gives:

$$v = u \; + \; \sqrt{2}\, F_1\left(\frac{1}{\sqrt{2}}\, u \; + \; \frac{1}{\sqrt{2}}\, v\right) \tag{2.6}$$

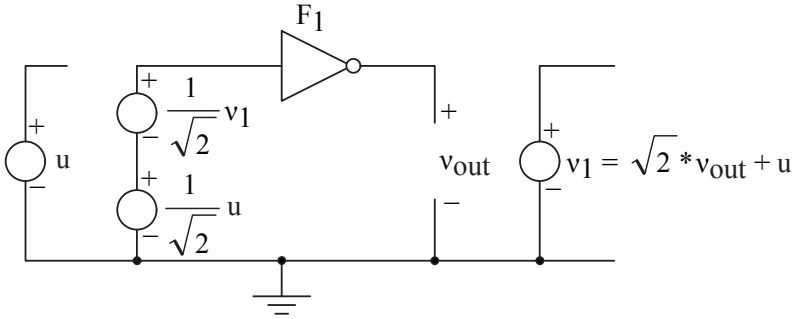
To find $F_2'(x)$, $F_2(x)$ is mirrored in x-y coordinate system with respect to v axis, which gives:

$$x = -\,\frac{1}{\sqrt{2}}\, u \; + \; \frac{1}{\sqrt{2}}\, v \tag{2.7}$$

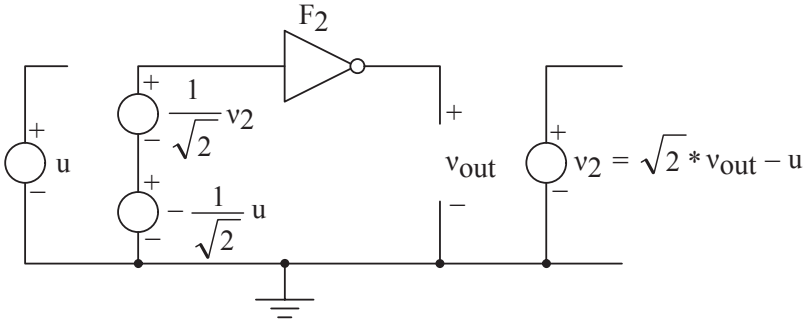
$$y = \frac{1}{\sqrt{2}}\, u \; + \; \frac{1}{\sqrt{2}}\, v \tag{2.8}$$

substituting Equations 2.7 and 2.8 in $y = F_1(x)$ gives:

$$v = -\,u\,+\,\sqrt{2}\,F_2\left(-\,\frac{1}{\sqrt{2}}u\,+\,\frac{1}{\sqrt{2}}\,v\right) \tag{2.9}$$



(a)



(b)

Figure 2.13: Circuit implementation of Equations 2.6 and 2.9 for finding the diagonal of the maximum embedded square in the two lobes of the VTCs [1]

Equations 2.6 and 2.9 express v as a function of u and it can be found using a circuit simulator like SPICE by implementing the equations into circuits as shown in Figure 2.13. The difference between the solution of the equations 2.6 and 2.9 is represented by the sinusoidal like curve in Figure 2.12. The absolute value of the peak amplitudes of this curve gives the diagonal of the maximum possible nested squares.

From this the worst-case SNM of the SRAM cell can be found by multiplying the smallest of the two amplitudes with $1/\sqrt{2}$.

N-Curve

Another method to determine the SNM is by using N-curve [13]. In this method SNM is characterized by both voltage and current. N-curve can be generated by sweeping the voltage at "0" storage node (V_1) and also measuring the current as shown if Figure 2.14. During this period the bit-lines ($BL, \overline{BL}$) and word-line (WL) are held at active high, V_{dd}.

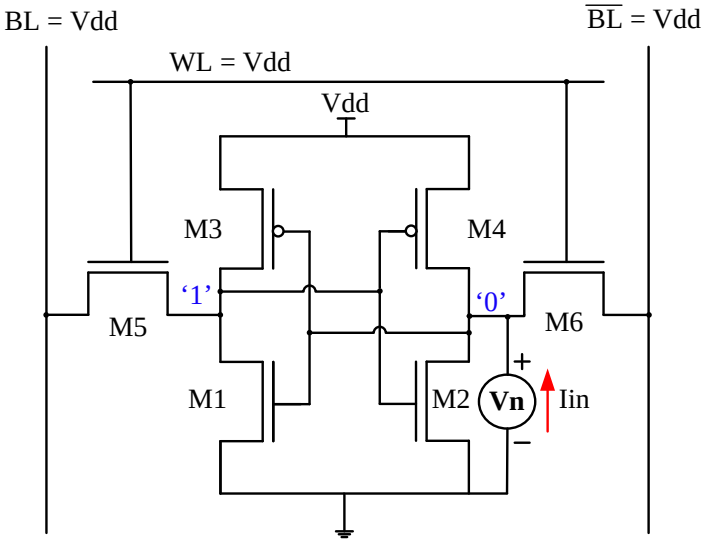


Figure 2.14: N-Curve Simulation Setup

Figure 2.15 shows the N-Curve of an SRAM cell, where x axis represents the node voltage at V_2 and y axis represents the injected current (I_{in}). At the three points A, B and C of the N-curve the injected current into the node (V_2) is zero. The two points A and C corresponds to the stable points of the butterfly curve and B corresponds to the metastable point. The voltage difference between the points A and B indicates the maximum DC noise voltage that can be tolerated at the node V_2 before the cell flips

the data. This metric is called the static voltage noise margin (SVNM) [14]. N-curve also gives current information that can be used to characterize the read stability. This metric is called static current noise margin (SINM). It is defined as the maximum DC current that can be injected into the SRAM cell without losing its data. SINM is represented in the curve as the peak current between the points A and B. The voltage in point A depends on cell ratio (β) while the voltage in point C depends on the γ-ratio of the cell.

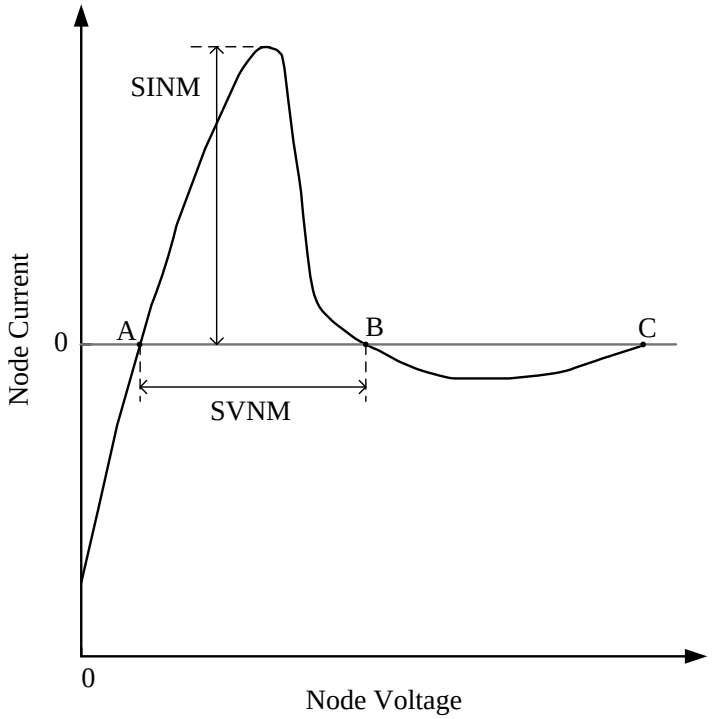


Figure 2.15: N-Curve
[13]

2.2.2 Soft Errors

The term soft error refers to an error caused by radiation or electromagnetic pulses that can be corrected by performing one or more normal functions of the device containing the latch or memory cell [15]. These charged particles that cause soft errors can originate directly from radioactive materials and cosmic rays or indirectly as a result of high-energy particle interaction with the semiconductor itself. High energy netrons from cosmic radiations Soft errors are random, usually not catastrophic, and normally do not destroy the device. Although the data can be corrected if the errors are detected, in complex systems the correction is highly unlikely and this data error can eventually lead to system failure.

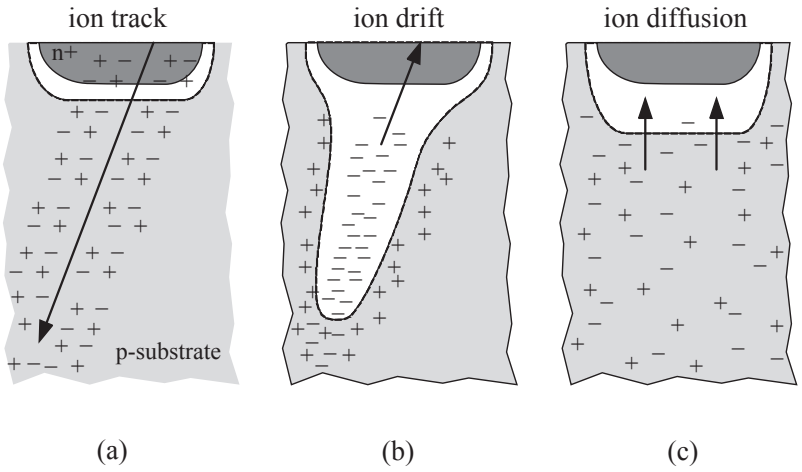


Figure 2.16: Soft Error Mechanism

A Single Event Upset (SEU) in an SRAM cell occurs when a charged particle strikes a sensitive node and flips the state of the SRAM cell, causing a soft error. In a 6T-SRAM cell the reverse-biased junctions between the drain and substrate are more sensitive to SEU, caused by ionizing particles, particularly at the node storing a logic high [16]. This is due to the fact that the storage node is supported by a relatively weak PMOS pull-up transistor compared to the strong NMOS pull-down transistor

at the other node [4].

The collection of the charge generated by an ionizing event occurs within a few micrometer of the p-n junction as shown in Figure 2.16. The collected charge, Q_{coll}, depends on the type of the ionizing particle, its trajectory, its energy value and its location of the impact. Q_{coll} ranges from 1 fC amounts to hundreds of fC. The minimum charge required to flip the bit stored in the cell is called critical charge, Q_{crit} [17]. If Q_{coll} is less than the Q_{crit} of the storage node, the current provided by the pull-up transistor will prevent the flipping of the bit. Thus, in an SRAM cell both the capacitance of the storage node and the restoring current provided by the pull-up transistors contribute to the cell critical charge, Q_{crit}. Q_{crit} can be defined as [18]:

$$Q_{crit} = \int_0^{\tau_{flip}} I_D \mathrm{d}t = (C_{node} \times V_{dd}) + (I_{restore} \times \tau_{flip}) \tag{2.10}$$

where C_{node} and V_{node} are the capacitance and the voltage of the affected storage node respectively, $I_{restore}$ is the restore current provided by the pull-up transistor, and τ_{flip} is the time required for the feedback mechanism to take over from the ions current and flip the cell.

When an energetic charged particle strikes the sensitive node, electron-hole pairs are generated by the interaction of the charged particle with the semiconductor atoms. The electron-hole pairs are collected at the opposite potentials of the reverse biased junction causing a current pulse for a few hundred pico seconds [19, 20, 21, 22, 23]. An analytical model for the induced current waveform for ion track charge collection is generally approximated as a double exponential curve with a rapid rise and fall time:

$$I(t) = \frac{Q_{coll}}{\tau_\alpha - \tau_\beta} \left(e^{-\frac{t}{\tau_\alpha}} - e^{-\frac{t}{\tau_\beta}} \right) \tag{2.11}$$

where Q_{coll} is the collected charge at the in fC at the sensitive node, τ_α is a process dependent collection time constant of the junction, and τ_β is a technology independent ion track establishment time constant. Typical value of τ_α and τ_β are $1.64 \times 10^{10} sec$ and $5 \times 10^{11} sec$ respectively. If the collected charge is greater than the Q_{crit} the generated current will overcome the $I_{restore}$ causing a change of state of the SRAM cell.

The high energy neutrons from cosmic radiation are the primary source for soft errors in modern ICs. Neutrons are stable and have high flux due to their charge neutrality. Less than 1% of primary flux reach the Earth's surface and it is a function of altitude. Neutrons do not directly cause SEUs but induce soft errors primarily by the neutron-induced silicon recoils. Neutons interact with the silicon nuclei producing secondary ions which can cause soft errors. The charge density for neutron-generated silicon recoils is 25-150 $fC/\mu m$.

There are following methods to mitigate the soft error problem in SRAMs.

- Eliminating the sources of radiation or reducing their intensity
- Error Correction Codes (ECC)
- Radiation hardening by layout and circuit techniques
- Circuit and/or system redundancy

One method to reduce the soft error rate (SER) is by choosing the right semiconductor and right materials during manufacturing. Error correction codes (ECC) can eliminate the soft errors by adding redundant bits to the data word so that if any error occurs it could be corrected. Radiation hardening involves increasing the storage node capacitance in order to increase cell Q_{crit}. It is often accomplished by increasing the size of transistors particularly the width of the transistors. And finally,

adding the redundancy on the circuit and/or system level is another powerful option that helps to reduce the sensitivity to ionizing radiation.

Chapter 3

Test Methodology

This chapter discusses the approach taken to improve the stability of SRAM designs. The stability improvement approach is achieved through the transistor width modulation. Next section discusses about different SRAM designs used to verify our approach and final section discusses the simulation setup to generate the performance characteristics of SRAM cells.

3.1 Stability Improvement - Approach

In this work an SRAM design trade-offs approach to improve the characteristics of SRAM by modulating the transistor sizing ratio, β, is presented. A variation in β-ratio from 1 to 3 can significantly improve the SNM of the SRAM cell. This property of SRAM is utilized to improve its stability. Other SRAM characteristics such as the Q_{crit}, the write performance and the power consumption can also be improved with SRAM design trade-offs approach. However, the desired SRAM design is characterized by a set of quadruple values concerning the SNM, Q_{crit}, write time (W_{time}) and power consumption. Note the Q_{crit} and the power consumption increases

as the total area of the SRAM cell increases. At the same time, the SNM improves by increasing the ratios of the transistors, β, within the cell. Thus an increase in β-ratio results in an improved SNM and Q_{crit} at the expense of the write time and power consumption.

	β-ratio				
	1	1.5	2	2.5	3
SNM (mV)	27.4	60.7	79.7	91.9	102.8
$Q_{crit}(fC)$	1.64	1.75	1.86	1.95	2.04
$W_{time}(ps)$	35.38	42.25	47.48	52.17	57.17
Power (μW)	6.54	7.96	9.38	10.72	12.04

Table 3.1: Characteristics of a 6T-SRAM cell with a variation in β-ratio in 32 nm Technology

To motivate this point, consider Table 3.1 which shows a fragment of experimentally derived basic 6T-SRAM cells as the β-ratio varies from 1 to 3. It is clear from this table that the SNM improves more than 2x as the β-ratio of SRAM changes from 1 to 3. Every design version is associated by its β and characterized by the quadruple set of parameters, i.e. the SNM in mV, Q_{crit} in fC, W_{time} in ps and the power consumption in μW. In this work, the preferred parameter to optimize is the SNM. However, consideration should also be given to the other three parameters, Q_{crit}, W_{time}, and power consumption. This approach uses the last parameters as design constraints to be satisfied while SNM undergoing optimization improvements. For example, if $Q_{crit} \geq 1.7fC$ then all SNM with for $\beta \geq 1.5$ would satisfy the critical charge. Moreover, if $W_{time} \leq 48ps$ and Power $\leq 10\mu W$ then the corresponding SNM with $\beta \leq 2$ will satisfy these constraints. Overall for $1.5 \leq \beta \leq 2$ all constraints would be satisfied which means values between 60.7 to 79.2 mV optimize SNM under the designer constraints.

This constraint driven tradeoff process can be generalized as follows. Suppose

$\mathcal{B} = \{\beta_1, \beta_2, \beta_3, ...\}$ is the set of β-ratios for design versions 1, 2, 3, ..., respectively, derived by experimentation. Let Q_{min}, T_{max}, P_{max} be the critical charge, write time and power constraint values. That is, for indices q, t, p in $\{1, 2, 3, ...\}$ let β_q, β_t, β_p be corresponding subsets in $\mathcal{B}$, then we have the following constraint relations

$$Q(\beta_q) \geq Q_{min} \geq Q(\beta_{q-1})$$
$$T(\beta_t) \leq T_{max} \leq T(\beta_{t+1})$$
$$P(\beta_p) \leq P_{max} \leq P(\beta_{p+1})$$

where $Q(\beta_q)$, $T(\beta_t)$ and $P(\beta_p)$ are the critical charge, time delay and power for SRAM versions q, t and p, respectively. Suppose now the solution to the above constraints are the following *beta* subsets, respectively, $\mathcal{B}_Q = \{\beta_q, \beta_{q+1}, ...\}$, $\mathcal{B}_T = \{\beta_t, \beta_{t-1}, ...\}$ and $\mathcal{B}_P = \{\beta_p, \beta_{p-1}, ...\}$. Then the tradeoff solution satisfying all constraints can be expressed by the β subset intersections

$$\mathcal{B}_{all} = \mathcal{B}_Q \cap \mathcal{B}_T \cap \mathcal{B}_P$$

where the β elements of $\mathcal{B}_{all}$ provide the corresponding improved SNM tradeoff values (see Table 3.1) that satisfy the above constraints. However, since all parameter values in Table 3.1 grow monotonically, we can express our tradeoff solution in terms simpler than the subset intersections using the previous constraint values for β, i.e. β_q, β_t and β_p. Thus all β_i points satisfying the following relation

$$\beta_q \leq \beta_i \leq min\{\beta_t, \beta_p\}$$

satisfy the constraints. Note these β_i points are contiguous in the table such as Table 3.1, meaning that they lie within β_q and $min\{\beta_t, \beta_p\}$. However, if $\beta_q > min\{\beta_t, \beta_p\}$

then there is no solution satisfying the designer constraints. For the previous example, $\beta_q = 1.5$, $\beta_t = 2$, $\beta_p = 2$ and $\beta_q \leq \beta \leq min\{\beta_t, \beta_p\}$, which yields $1.5 \leq \beta \leq 2$, or $\beta = \{1.5, 2\}$.

3.2 Test circuits

This section discusses various SRAM designs across different architectures that are used to verify the stability improvement approach. Most of the designs are modification in SRAM cell to reduce the sensitivity to radiations. These cells are designed such that it can tolerate the current generated due to single event transients (SET) created by the impact of charged particles.

3.2.1 6T-SRAM

A classic six transistor SRAM (6T-SRAM) cell is used as the reference design to verify stability improvement approach. A 6T-SRAM cell is shown in Figure 3.1. As mentioned in chapter 2. 6T-SRAM cell uses a positive feedback between two cross coupled latches formed by transistors $M1$-$M4$ to store one bit of data. Data is accessed or written through the bit-lines by activating WL. The access transistors, $M5$-$M6$, isolates the cell from other circuitry during standby mode.

3.2.2 SRAM-C

The SRAM-C design, as shown in Figure 3.2, uses a capacitor between the nodes to increase the overall charge storage of the cell, thus enhancing radiation immunity. The capacitor based SEU protection method increases the area overhead. However, the problem of the area overhead is addressed in [29] by vertically stacking the capacitor

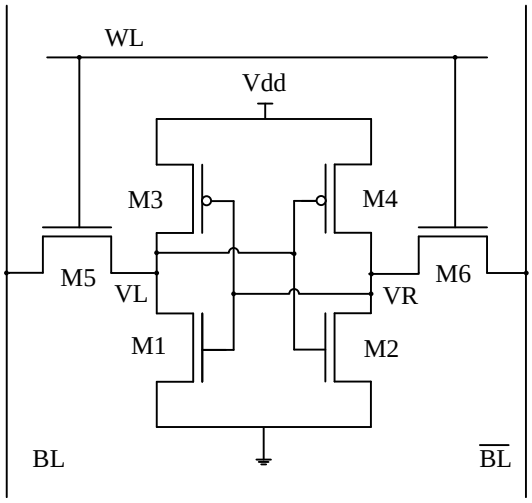


Figure 3.1: Standard six transistor SRAM cell (6T-SRAM)

to minimize the area footprint of the SRAM cell. The major drawback of the capacitor based models are the impact on the write performance and incompatibility with the low power CMOS technology scaling.

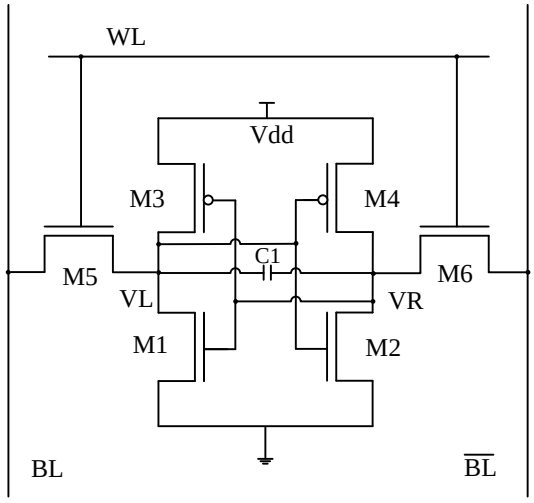


Figure 3.2: Capacitor based SRAM cell(SRAM-C)

3.2.3 SRAM-T

In SRAM-T design proposed in [30], two additional tristate inverters are added to 6T-SRAM cell to improve soft error immunity. These tri-state inverters, classified as outer core, are activated by $\overline{WL}$ during standby mode as shown in Figure 3.3. The addition of the outer core tri-state inverters, $M7 - M10$, has the effect of increasing the critical charge, Q_{crit}, at the nodes, $V1$ and $V2$, and thus the tolerance level of the SRAM cell to SEUs is greatly improved. The level of tolerance is dependent on the physical parameters and characteristics of the transistors in the outer core inverters. Moreover, the tristate inverters have minimal impact on the write performance as they are active only during the standby mode.

3.2.4 SRAM-NSP

In this SRAM-NSP design, additional two CMOS transistors: $M7 - M10$, along with two NMOS transistors: $M11 - M12$, and a vertically stacked capacitor, $C1$, are connected to the storage nodes to improve soft error protection, as shown in Figure 3.4. Furthermore, two additional transistors, *M13* and *M14*, are connected to the capacitor to isolate the cell during a read access. The CMOS transistors act as a switch and are activated only during the standby mode through $\overline{WL}$. The capacitor act as a charge buffer and helps to improve the Q_{crit} of the SRAM cell. During a read or write mode, NMOS transistors are turned *ON* and capacitor discharges through it. Once the SRAM cell goes back to the hold mode the capacitor is re-introduced into the system.

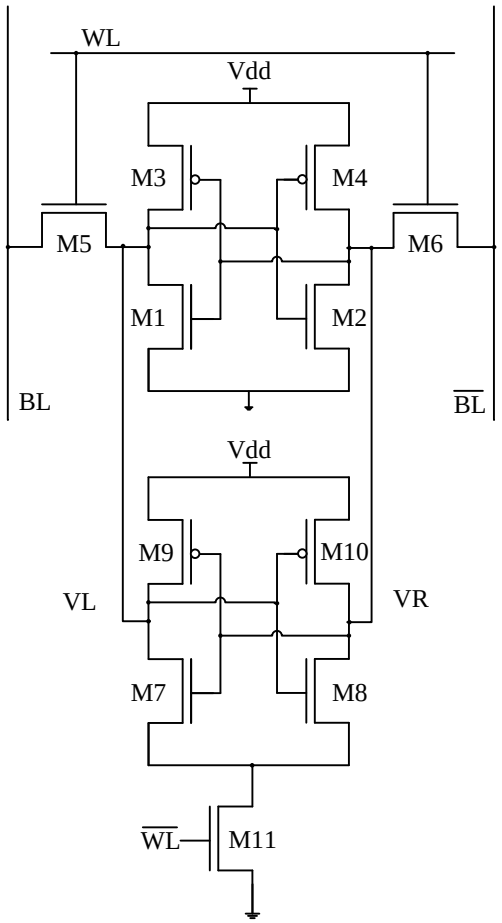


Figure 3.3: Tristate SRAM cell (SRAM-T)

3.3 Transistor Sizing

The different operating modes of an SRAM cell: read, write, standby, often impose contradicting requirements on transistor sizing. This is mainly due to the fact that the access transistor is used for both reading and writing data into the cell. For a successful read and write operation the β and γ ratio of the cell must be greater than 1. To achieve this, transistors with minimum length, $L_{min} = 2\lambda$, and minimum width, $W_{min} = 4\lambda$ were used, where λ is the minimum feature size of a particular technology node. Furthermore, the pull-down and pull-up transistors were used in such a way that they meet the condition $W_{PD} > W_{PG} > W_{PU}$, where W_{PD}, W_{PG}, W_{PU} are the

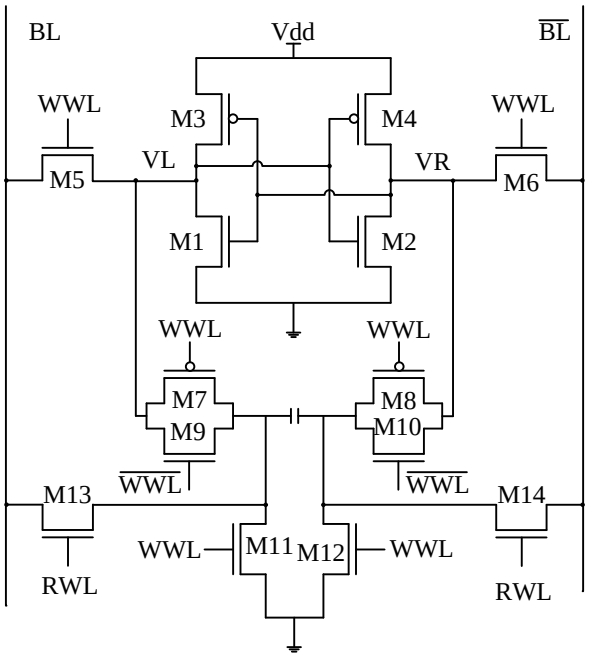


Figure 3.4: Tristate SRAM cell (SRAM-NSP) with separate read port

transistor widths of pull-down (PD), pass-gate (PG) and pull-up transistors (PU) respectively. Table 3.2 shows the transistor widths of the SRAM cells used in the simulation. For all the designs, the transistor sizing ratio used for the minimum sized SRAM cell were $\beta = 1$ and $\gamma = 1.5$ for .

Design	6T-SRAM	SRAM-C	SRAM-T	SRAM-NSP
$M1, M2$	6λ	6λ	6λ	6λ
$M3, M4$	4λ	4λ	4λ	4λ
$M5, M6$	6λ	6λ	6λ	6λ
$M7, M8$	–	–	6λ	6λ
$M9, M10$	–	–	4λ	6λ
$M11, M12$	–	–	6λ	6λ
$M13, M14$	–	–	6λ	6λ

Table 3.2: Widths of transistors for minimum size SRAM cells

3.4 Simulation Setup

In the simulations, the SNM of the 6T-SRAM cell is determined during a read access and then Q_{crit}, write time and power consumption are measured. The set of characteristics for the 6T-SRAM cell while varying transistor sizing ratio, β were calculated. Three additional SRAM designs: SRAM-C, SRAM-T, and SRAM-NSP were explored to find the good tradeoff points between SNM, Q_{crit}, Write time delay and power consumption. The 6T-SRAM, SRAM-C, SRAM-T, and SRAM-NSP cells were designed for 32 nm process technologies using the Berkeley Predictive Technology Model (BPTM) data for bulk CMOS [31]. The simulations are performed using HSPICE at a constant supply voltage (V_{dd}) of 1.0 V.

Seevinck's method [10] was used to calculate the static noise margin during a read access. It is a graphical method in which voltage transfer characteristics of the two bistable elements in the SRAM cell are plotted to determine the SNM. This method makes it easier to automate using a circuit simulator. According to this method, the area inside the two lobes is a measure of the sensitivity of SRAM cell to noise. The side of the maximum possible nested square inside the smallest lobe represents the SNM of that memory cell.

The procedure to plot the butterfly curve is as follows. First, the feedback of the bistable elements are broken. Next, the VTC of the inverter formed by half of the SRAM cell is estimated by sweeping V1, from 0 to VDD and measuring V2 as shown in Figure 3.5. This is performed again on the other side of the SRAM cell by sweeping the voltage V2 and measuring V1. These two curves are used to construct the butterfly curve of the cell. During the measurement, the WL and bitlines ($BL, \overline{BL}$) are held at V_{dd} to create a SRAM cell read operation. To calculate the SNM values graphically, DC analysis from HSPICE to MATLAB were exported

using Perl scripts. In MATLAB the data is analytically solved to find the side of the maximum embedded square.

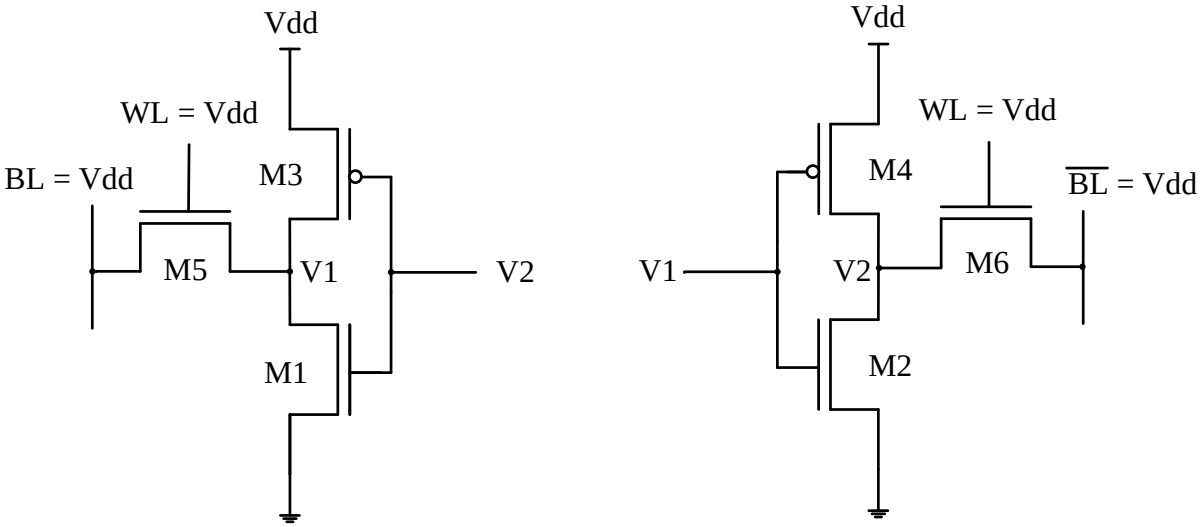


Figure 3.5: Butterfly curve simulation set up

As discussed in Chapter 2. soft error can be quantified by the amount of charge required to flip the bit stored in the cell. Particle strike in a memory cell results in a transient current flow and if the the charge created due to radiation is greater than the critical charge, Q_{crit} of the cell node results in a soft error. The critical charge of the node can be found by injecting a current pulse, enough to flip the data, as shown in Figure 3.6. The critical charge can be quantified by integrating the current pulse over time (area under the current pulse). Since the charge required for 1 - 0 transition is lesser than the 0 - 1 transition, the '1' storage node was considered, worst case scenario.

Figure 3.7 shows the simulation flow to calculate the SRAM performance characteristics that are used in the approach. HSPICE was used to simulate the operation of the SRAM cells. The nominal operating conditions and the cell parameters are specified in the SRAM netlist. For generating SNM values of the cell, DC analysis

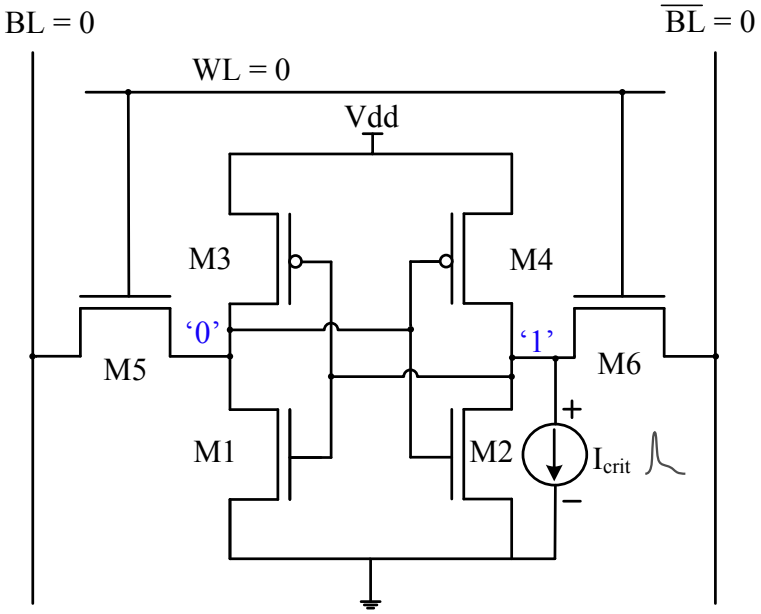


Figure 3.6: SEU simulation setup

of the cell was performed without the feedback. The resulting data is used to calculate the SNM values from MATLAB analytical simulation. For generating the power consumption (P_{avg}), write time delay (W_{time}) and critical charge (Q_{crit}), transient analysis of SRAM cells was performed. The P_{avg} and W_{time} of the cell were calculated during a write operation while the Q_{crit} is calculated during the standby mode of the cell.

Keeping the process technology and supply voltage constant, stability analysis of the four SRAM designs was performed. During this analysis the transistor sizing ratio, (β) was varied. Stability tests of the SRAM cells were performed for the variation in transistor sizing ratios in the same process technology and constant supply voltage. The reliability, write performance and power consumption of the SRAM cells for different β ratio were measured. The critical charge of the node was quantified as the measure of the reliability due to soft errors. The write performance is calculated as the time required to change from $10 - 90\%$ of the node voltage. For power consumption

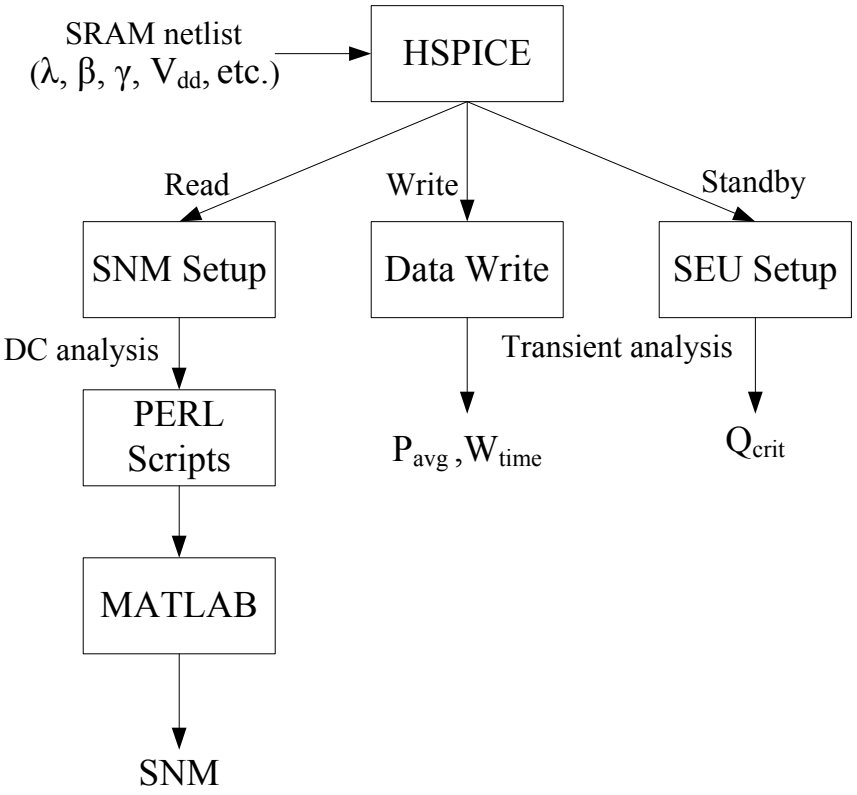


Figure 3.7: Simulation Flow

measurements, the average power consumed during a write operation, was taken into consideration.

Chapter 4

Results

In this chapter verification of the stability improvement approach mentioned in the previous chapter is discussed. The data generated for additional designs SRAM-C, SRAM-T, SRAM-NSP are used to show that this method is applicable to different SRAM cells and architectures.

The goal of this approach is to use tradeoffs between SRAM characteristics within performance constraints to enhance the particular SRAM design. The constraints considered in this approach are critical charge, write time delay and power consumption. The choice of the good tradeoff points is determined by the SNM level of the required application.

4.1 6T-SRAM

In the case of a 6T-SRAM cell, the SNM levels are significantly degraded during a read access. Table 4.1 shows the characteristics of 6T-SRAM cell in 32 nm technology. A careful choice of β-ratio is important to achieve a better SNM levels for the cell. At the same time, the other performance characteristics of the cell will change with

Design	β	SNM (mV)	Q_{crit} (fC)	Delay (ps)	Power (μW)
6T-SRAM	1	27.40	1.64	35.38	6.54
	1.2	45.00	1.69	37.39	7.09
	1.4	54.40	1.70	40.54	7.67
	1.6	64.60	1.73	43.44	8.27
	1.8	72.40	1.81	45.37	8.83
	2	**79.70**	**1.86**	**47.48**	**9.38**
	2.2	84.10	1.89	49.32	9.92
	2.4	91.00	1.93	51.21	10.46
	2.6	94.60	1.97	53.14	10.99
	2.8	98.30	2.01	55.18	11.51
	3	102.80	2.04	57.17	12.03

Table 4.1: Performance characteristics of 6T-SRAM cell in 32 nm technology

an increase in β-ratio. The selection of the β-ratio is determined by its impacts on write time delay and power consumption versus area footprint of the cell. Suppose the characteristics of the desired design version is restricted by the following design constraints

$$Q_{min} = 1.7fC$$

$$T_{max} = 48ps$$

$$P_{max} = 10\mu W$$

The choice of the SNM levels should satisfy all of the constraints listed above. For power consumption, $P_{max} = 10\mu W$ all 6T-SRAM designs for $\beta = 1$ to 2 satisfy the power consumption requirements. For critical charge requirement of $Q_{min} = 1.7fC$ all 6T-SRAM designs for β above 1.6 satisfy the requirement. For write performance requirement of $T_{max} = 48ps$ all 6T-SRAM designs for $\beta = 1$ to 2.2 satisfy the require-

ment. Applying all the constraints requirements on the previously mentioned design versions, a subset of design versions for 6T-SRAM cell is attained with β between 1 to 2.

4.2 SRAM-C

Design	β	SNM (mV)	Q_{crit} (fC)	Delay (ps)	Power (μW)
SRAM-C	1	27.40	5.97	247.30	47.24
	1.2	45.00	5.95	250.11	47.84
	1.4	54.40	5.91	250.05	48.44
	1.6	64.60	5.92	254.04	49.01
	1.8	72.40	5.99	254.59	49.52
	2	79.70	6.02	256.00	50.02
	2.2	**84.10**	**6.05**	**257.97**	**50.49**
	2.4	91.00	6.08	262.39	50.94
	2.6	94.60	6.10	264.47	51.37
	2.8	98.30	6.12	265.04	51.92
	3	102.80	6.15	268.25	52.44

Table 4.2: Performance characteristics of SRAM-C cell in 32 nm technology

In an SRAM-C design, the capacitor is connected between the nodes causes a large delay during the write operation. The rest of the SRAM-C architecture is similar to the 6T-SRAM cell. It is interesting to note that the performance characteristics of this cell increase uniformly with an incremental β, as shown in Table 4.2. Suppose the choice of SNM is driven by the following design requirements

$$Q_{min} = 6fC$$

$$T_{max} = 260ps$$

$$P_{max} = 52\mu W$$

For performance requirements, $Q_{min} = 6fC, T_{max} = 260ps, P_{max} = 52\mu W$ all SRAM-C design versions for $\beta = 2$ to 2.2 satisfy the requirements.

4.3 SRAM-T

Design	β	SNM (mV)	Q_{crit} (fC)	Delay (ps)	Power (μW)
SRAM-T	1	177.40	3.16	38.38	12.33
	1.2	185.40	3.21	39.68	13.01
	1.4	192.70	3.26	41.31	13.67
	1.6	199.90	3.31	42.76	14.31
	1.8	**206.70**	**3.35**	**43.64**	**14.97**
	2	211.10	3.40	45.17	15.59
	2.2	214.70	3.44	46.74	16.24
	2.4	219.00	3.48	48.53	16.89
	2.6	223.50	3.41	49.95	17.52
	2.8	225.60	3.56	51.28	18.15
	3	229.00	3.60	52.66	18.78

Table 4.3: Performance characteristics of SRAM-T cell in 32 nm technology

The SNM of the SRAM-T cells can be optimized for the β-ratio between 1.6 and 1.8 for the performance constraints shown below.

$$Q_{min} = 3.3fC$$

$$T_{max} = 50ps$$

$$P_{max} = 15\mu W$$

Table 4.3 shows the SRAM-T design versions for the different transistor sizing ratio, β. It also follows the trend of incremental change in the design constraints as β increases.

4.4 SRAM-NSP

Design	β	SNM (mV)	Q_{crit} (fC)	Delay (ps)	Power (μW)
SRAM-NSP	1	112.10	7.36	51.04	10.29
	1.2	121.10	7.40	53.19	10.86
	1.4	129.60	7.45	55.11	11.32
	1.6	135.30	7.49	56.07	11.83
	1.8	140.70	7.53	58.05	12.54
	2	**143.10**	**7.56**	**60.62**	**13.08**
	2.2	147.60	7.60	63.48	13.62
	2.4	150.40	7.64	66.34	14.14
	2.6	152.30	7.68	67.76	14.30
	2.8	154.30	7.71	70.12	15.14
	3	156.50	7.75	71.86	15.71

Table 4.4: Performance characteristics of SRAM-NSP cell in 32 nm technology

The trend in the performance characteristics of the SRAM-NSP is presented in the Table 4.4. SRAM-NSP cells show better SNM levels for the minimum β due a different read mechanism. The design has relatively high initial Q_{crit} levels because of the extra capacitor connected between the storage nodes. The choice of β-ratio between 1 and 2 satisfy the design constraints shown below.

$$Q_{min} = 6fC$$

$$T_{max} = 60ps$$

$$P_{max} = 15\mu W$$

A similar trend of increase in Q_{crit}, write time and power consumption across all the cells is seen with an increase in the β-ratio. This monotonic nature of the

performance characteristics is a direct result of additional cell area, which influences the overall characteristics of the cell. The 6T-SRAM cell has the lowest SNM and Q_{crit} levels, in a region where it has better write time performance and power savings. These observations can be attributed to the small transistor dimensions compared to other tested SRAM designs. The additional components in the other SRAM cell architectures improve its Q_{crit} levels at the expense of write time delay and power consumption.

Chapter 5

Summary

This work presents an SRAM design tradeoffs approach to improve the characteristics of SRAM by modulating the transistor sizing ratio, β. The monotonic nature of the SRAM characteristics to improve the SNM were explored. The SRAM designs were optimized with β for various constraints in power consumption, performance, radiation tolerance, and data stability. Different design trends produced by the analysis of the tradeoff approach were discussed. This approach can be applied to different SRAM designs.

This work focuses on a small selection of tradeoffs for improving the SNM. This tradeoff approach can be extended to a more general design space exploration problem which encompasses leakage power, interconnect parasitics, layout area, degradation mechanisms, process parameter variations etc.

Bibliography

[1] F. J. List, "The static noise margin of sram cells," in *Solid-State Circuits Conference, 1986. ESSCIRC '86. Twelfth European*, 1986, pp. 16 –18.

[2] H. Yamauchi, "A discussion on sram circuit design trend in deeper nanometer-scale technologies," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 18, no. 5, pp. 763 –774, May 2010.

[3] ITRS. international technology roadmap for semiconductors - 2009. [Online]. Available: http://www.itrs.net/Links/2009ITRS/Home2009.htm

[4] R. Baumann, "Soft errors in advanced computer systems," *Design Test of Computers, IEEE*, vol. 22, no. 3, pp. 258 – 266, may-june 2005.

[5] A. Bhavnagarwala, X. Tang, and J. Meindl, "The impact of intrinsic device fluctuations on cmos sram cell stability," *Solid-State Circuits, IEEE Journal of*, vol. 36, no. 4, pp. 658 –665, Apr. 2001.

[6] F. shi Lai and C.-F. Lee, "On-chip voltage down converter to improve sram read/write margin and static power for sub-nano cmos technology," *Solid-State Circuits, IEEE Journal of*, vol. 42, no. 9, pp. 2061 –2070, 2007.

[7] M. Yoshimoto, K. Anami, H. Shinohara, T. Yoshihara, H. Takagi, S. Nagao, S. Kayano, and T. Nakano, "A divided word-line structure in the static ram and its application to a 64k full cmos ram," *Solid-State Circuits, IEEE Journal of*, vol. 18, no. 5, pp. 479 – 485, oct 1983.

[8] T. Hirose, H. Kuriyama, S. Murakami, K. Yuzuriha, T. Mukai, K. Tsutsumi, Y. Nishimura, Y. Kohno, and K. Anami, "A 20 ns 4 mb cmos sram with hierarchical word decoding architecture," in *Solid-State Circuits Conference, 1990. Digest of Technical Papers. 37th ISSCC., 1990 IEEE International*, feb. 1990, pp. 132 –133.

[9] J. Lohstroh, E. Seevinck, and J. de Groot, "Worst-case static noise margin criteria for logic circuits and their mathematical equivalence," *Solid-State Circuits, IEEE Journal of*, vol. 18, no. 6, pp. 803 – 807, Dec. 1983.

[10] E. Seevinck, F. List, and J. Lohstroh, "Static-noise margin analysis of mos sram cells," *Solid-State Circuits, IEEE Journal of*, vol. 22, no. 5, pp. 748 – 754, Oct. 1987.

[11] C. F. Hill, "Noise margin and noise immunity in logic circuits," *Microelectronics*, vol. 1, no. 4, pp. 16 – 21, 1968.

[12] B. Alorda, G. Torrens, S. Bota, and J. Segura, "Static and dynamic stability improvement strategies for 6t cmos low-power srams," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2010*, 2010, pp. 429 –434.

[13] C. Wann, R. Wong, D. Frank, R. Mann, S.-B. Ko, P. Croce, D. Lea, D. Hoyniak, Y.-M. Lee, J. Toomey, M. Weybright, and J. Sudijono, "Sram cell design for stability methodology," in *VLSI Technology, 2005. (VLSI-TSA-Tech). 2005 IEEE VLSI-TSA International Symposium on*, 2005, pp. 21 – 22.

[14] E. Grossar, M. Stucchi, K. Maex, and W. Dehaene, "Read stability and write-ability analysis of sram cells for nanometer technologies," *Solid-State Circuits, IEEE Journal of*, vol. 41, no. 11, pp. 2577 –2588, 2006.

[15] (2009, Dec) JEDEC dictionary of terms for solid state technology 5th edition: ESD88D. [Online]. Available: http://www.jedec.org/standards-documents/docs/jesd-88c

[16] V. Degalahal, N. Vijaykrishnan, and M. Irwin, "Analyzing soft errors in leakage optimized sram design," in *VLSI Design, 2003. Proceedings. 16th International Conference on*, 2003, pp. 227 – 233.

[17] F. Wang and V. Agrawal, "Single event upset: An embedded tutorial," in *VLSI Design, 2008. VLSID 2008. 21st International Conference on*, Jan. 2008, pp. 429–434.

[18] P. Roche, J. Palau, G. Bruguier, C. Tavernier, R. Ecoffet, and J. Gasiot, "Determination of key parameters for seu occurrence using 3-d full cell sram simulations," *Nuclear Science, IEEE Transactions on*, vol. 46, no. 6, pp. 1354 –1362, Dec. 1999.

[19] F. Vargas and M. Nicolaidis, "Seu-tolerant sram design based on current monitoring," in *Fault-Tolerant Computing, 1994. FTCS-24. Digest of Papers., Twenty-Fourth International Symposium on*, Jun 1994, pp. 106–115.

[20] P. C. Murley and G. R. Srinivasan, "Soft-error monte carlo modeling program, semm," *IBM J. Res. Dev.*, vol. 40, no. 1, pp. 109–118, 1996.

[21] P. Hazucha, K. Johansson, and C. Svensson, "Neutron induced soft errors in cmos memories under reduced bias," *Nuclear Science, IEEE Transactions on*, vol. 45, no. 6, pp. 2921–2928, Dec 1998.

[22] L. Anghel and M. Nicolaidis, "Cost reduction and evaluation of temporary faults detecting technique," in *DATE '00: Proceedings of the conference on Design, automation and test in Europe*. New York, NY, USA: ACM, 2000, pp. 591–598.

[23] G. C. Messenger, "Collection of charge on junction nodes from ion tracks," *Nuclear Science, IEEE Transactions on*, vol. 29, no. 6, pp. 2024–2031, Dec. 1982.

[24] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th ed. USA: Addison-Wesley Publishing Company, 2010.

[25] A. Pavlov and M. Sachdev, *CMOS SRAM Circuit Design and Parametric Test in Nano-Scaled Technologies: Process-Aware SRAM Design and Test*, 1st ed., V. D. Agrawal, Ed. Springer Publishing Company, Incorporated, 2008.

[26] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital integrated circuits- A design perspective*, 2nd ed. Prentice Hall, 2004.

[27] K. Zhang, *Embedded Memories for Nano-Scale VLSIs*, 4th ed. USA: Springer US, 2009.

[28] A. Strong, E. Wu, R. Vollertsen, J. Sune, G. Rosa, T. Sullivan, and S. Rauch, *Reliability Wearout Mechanisms in Advanced CMOS Technologies()*, 2009.

[29] M. Lysinger, F. Jacquet, M. Zamanian, D. McClure, P. Roche, N. Sahoo, and J. Russell, "A radiation hardened nano-power 8mb sram in 130nm cmos," *Quality Electronic Design, International Symposium on*, vol. 0, pp. 23–29, 2008.

[30] Y. Shiyanovskii, F. Wolff, and C. Papachristou, "Sram cell design using tri-state devices for seu protection," in *On-Line Testing Symposium, 2009. IOLTS 2009. 15th IEEE International*, 2009, pp. 114 –119.

[31] Y. Cao, T. Sato, M. Orshansky, D. Sylvester, and C. Hu, "New paradigm of predictive mosfet and interconnect modeling for early circuit simulation," in *Custom Integrated Circuits Conference, 2000. CICC. Proceedings of the IEEE 2000*, 2000, pp. 201–204.