

# **A Read-Decoupled Gated-Ground SRAM Architecture for Low-Power Embedded Memories**

**Wasim Hussain**

A Thesis  
In  
The Department of  
Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements for the Degree of  
Master of Applied Science (Electrical and Computer Engineering) at  
Concordia University  
Montreal, Quebec, Canada

October 2011

©Wasim Hussain, 2011

**CONCORDIA UNIVERSITY  
SCHOOL OF GRADUATE STUDIES**

This is to certify that the thesis prepared

By: Wasim Hussain

Entitled: “A Read-Decoupled Gated-Ground SRAM Architecture for Low-Power Embedded Memories”

and submitted in partial fulfillment of the requirements for the degree of

**Master of Applied Science**

Complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

|                       |                                      |
|-----------------------|--------------------------------------|
| _____                 | Chair                                |
| Dr. R. Raut           |                                      |
| _____                 | Examiner, External<br>To the Program |
| Dr. M. Mannan (CIISE) |                                      |
| _____                 | Examiner                             |
| Dr. G. Cowan          |                                      |
| _____                 | Supervisor                           |
| Dr. S. Jahinuzzaman   |                                      |

Approved by: \_\_\_\_\_  
Dr. W. E. Lynch, Chair  
Department of Electrical and Computer Engineering

\_\_\_\_\_ 20\_\_\_\_\_

\_\_\_\_\_

Dr. Robin A. L. Drew  
Dean, Faculty of Engineering and  
Computer Science

# **Abstract**

## **A Read-Decoupled Gated-Ground SRAM**

### **Architecture for Low-Power Embedded Memories**

**Wasim Hussain**

In order to meet the incessantly growing demand of performance, the amount of embedded or on-chip memory in microprocessors and systems-on-chip (SOC) is increasing. As much as 70% of the chip area is now dedicated to the embedded memory, which is primarily realized by the static random access memory (SRAM). Because of the large size of the SRAM, its yield and leakage power consumption dominate the overall yield and leakage power consumption of the chip. However, as the CMOS technology continues to scale in the sub-65 nanometer regime to reduce the transistor cost and the dynamic power, it poses a number of challenges on the SRAM design. In this thesis, we address these challenges and propose cell-level and architecture level solutions to increase the yield and reduce the leakage power consumption of the SRAM in nanoscale CMOS technologies.

The conventional six transistor (6T) SRAM cell inherently suffers from a trade-off between the read stability and write-ability because of using the same bit line pair for both the read and write operations. An optimum design at a given process and voltage condition is a key to ensuring the yield and reliability of the SRAM. However, with technology scaling, process-induced variations in the transistor dimensions and

electrical parameters coupled with variation in the operating conditions make it difficult to achieve a reasonably high yield. In this work, a gated SRAM architecture based on a seven transistor (7T) SRAM bit-cell is proposed to address these concerns. The proposed cell decouples the read bit line from the write bit lines. As a result, the storage node is not affected by any read induced noise during the read operation. Consequently, the proposed cell shows higher data stability and yield under varying process, voltage, and temperature (PVT) conditions. A single-ended sense amplifier is also presented to read from the proposed 7T cell while a unique write mechanism is used to reduce the write power to less than half of the write power of the conventional 6T cell. The proposed cell consumes similar silicon area and leakage power as the 6T cell when laid out and simulated using a commercial 65-nm CMOS technology. However, as much as 77% reduction in leakage power can be achieved by coupling the 7T cell with the column virtual grounding (CVG) technique, where a non-zero voltage is applied to the source terminals of driver NMOS transistors in the cell. The CVG technique also enables implementing multiple words per row, which is a key requirement for memories to avoid multiple-bit data upset in the event of radiation induced single event upset or soft error. In addition, the proposed cell inherently has a 30% larger soft error critical charge, making its soft error rate (SER) less than the half of that of the 6T cell.

# Acknowledgements

This thesis would not have been possible without the constant guidance and encouragement by my supervisor, Dr. Shah M. Jahinuzzaman. I owe my deepest gratitude to him for his relentless support, both professionally and personally, during my research at Concordia University. He has been a constant source of inspiration and has provided consistent succors and valuable suggestions throughout this project.

I owe my deepest gratitude to my beloved parents. Their continuous encouragement made it possible for me to pursue a successful study and happy life in Montreal.

Last but not the least I would like to thank my colleagues in my lab. Whether it was regarding my research or my course work or my personal problems, they have always extended their supporting hands.

# Table of Contents

|   |     |
|---|-----|
| Table of Contents.....                        | VI  |
| List of Figures.....                          | X   |
| List of Tables.....                           | XIV |
| 1. Introduction.....                          | 1   |
| 1.1 Memory Hierarchy in Computer Systems..... | 2   |
| 1.2 SRAM Design Challenges.....               | 5   |
| 1.2.1 Process Variations.....                 | 6   |
| 1.2.2 Leakage Power Consumption.....          | 7   |
| 1.2.3 Single Event Upset (SEU).....           | 8   |
| 1.3 Motivation and Thesis Outline.....        | 8   |
| 2. SRAM Architecture and Operation.....       | 10  |
| 2.1 Basic SRAM Architecture.....              | 10  |
| 2.2 6T SRAM Cell.....                         | 12  |
| 2.2.1 Read Operation.....                     | 14  |
| 2.2.2 Write Operation.....                    | 16  |
| 2.3 Row Decoder.....                          | 19  |
| 2.4 Column Decoder or Multiplexer.....        | 22  |

|   |    |
|---|----|
| 2.5 Sense Amplifier .....   | 24 |
| 2.6 Write Drivers .....   | 28 |
| 2.7 Timing and Control Circuits .....   | 29 |
| 3. Impact of Process Variation on SRAMs .....                                 | 31 |
| 3.1 Process Variation .....   | 31 |
| 3.1.1 Impact of Intra-die Process Variation on Memory Cells .....             | 34 |
| 3.1.2 Impact of Process Variation on Read Stability .....                     | 35 |
| 3.1.3 Impact of Process Variation on Write Margin .....                       | 36 |
| 3.2 Existing SRAM Designs for Limiting the Impact of Process Variations ..... | 37 |
| 3.2.1 7T SRAM Cell .....  | 37 |
| 3.2.2 8T SRAM Cell .....  | 38 |
| 3.2.3 9T SRAM Cell .....  | 39 |
| 3.2.4 Performance Comparison of the Existing SRAM Design .....                | 40 |
| 4. Proposed 7T SRAM Cell and Sense-Amplifier .....                            | 43 |
| 4.1 Cell Design .....   | 43 |
| 4.2 Principle of Operation of the Proposed 7T Cell .....                      | 46 |
| 4.2.1 Cell Operation .....  | 46 |
| 4.2.2 Array Operation .....   | 47 |
| 4.3 Theoretical Analysis of the Proposed Cell .....                           | 49 |

|   |    |
|---|----|
| 4.4 The Proposed Single Ended Sense-Amplifier .....                                 | 52 |
| 4.4.1 The Principle of Operation of the Proposed Single Ended Sense-Amplifier ..... | 55 |
| 5. Validation and Comparison of the Proposed SRAM Cell.....                         | 56 |
| 5.1 Simulation Setup.....   | 56 |
| 5.2 Write Performance.....  | 61 |
| 5.3 Read Performance.....   | 65 |
| 5.4 Leakage Power.....  | 67 |
| 5.5 Soft Error Tolerance .....  | 67 |
| 5.6 Cell Area.....  | 72 |
| 5.7 Performance of the Sense Amplifier.....   | 74 |
| 6. A Low-Leakage Array Architecture with Column Virtual Grounding.....              | 76 |
| 6.1 Array Implementation with CVG .....   | 77 |
| 6.2 Performance Results .....   | 79 |
| 7. Conclusion.....  | 83 |
| 7.1 Contribution to the Field.....  | 83 |
| 7.1.1 The Proposed 7T SRAM Cell.....  | 83 |
| 7.1.2 The Proposed Single-Ended Sense Amplifier .....                               | 84 |
| 7.1.3 A Low-Leakage Array with Multiple Words in a Row .....                        | 84 |
| 7.2 Future Works .....  | 85 |



|                 |    |
|-----------------|----|
| References..... | 87 |
| Glossary.....   | 92 |

# List of Figures

|   |    |
|---|----|
| Figure 1.1: (a) Comparison of area of logic and memory in a SOC [1]. (b) Die photo of 1.5GHz Third Generation Itanium® 2 Processor [2]. ..... | 2  |
| Figure 1.2: Memory hierarchy of a modern personal computer. ....  | 3  |
| Figure 1.3: Schematic of a conventional six-transistor SRAM cell. ....  | 4  |
| Figure 1.4: Scaling of transistor gate length according to Moore’s Law. Adapted from [6]. .   | 5  |
| Figure 1.5: Scaling trend of SRAM bit-cell size [7]. ....   | 6  |
| Figure 1.6: Leakage power and total power consumption of microprocessors with technology scaling [9]. ....                                    | 7  |
| Figure 2.1: A typical SRAM architecture. ....   | 11 |
| Figure 2.2: Conventional 6T SRAM cell. ....   | 12 |
| Figure 2.3: The VTCs of two cross-couple inverters forming the butterfly curve of the SRAM cell. ....   | 13 |
| Figure 2.4: 6T SRAM cell during a read operation (The transistors in grayscale are OFF).  | 15 |
| Figure 2.5: 6T SRAM cell during a write operation (The transistors in grayscale are OFF).   | 17 |
| Figure 2.6: Segmented decoding of address bits in a row decoder .....   | 21 |
| Figure 2.7: A word line driver circuit to reduce PMOS leakage current. ....   | 22 |
| Figure 2.8: An SRAM array with: (a) single word per row and (b) multiple words per row.   | 23 |
| Figure 2.9: 4-to-1 column MUX: a) pre-decoder based and b) tree based. ....   | 24 |

|  |    |
|--|----|
| Figure 2.10: (a) SRAM column with the sense amplifier and precharge circuits and (b)<br>Basic differential sense amplifier with current mirror load. ....  | 25 |
| Figure 2.11: (a) A latch-type sense amplifier in an SRAM column. ....  | 27 |
| Figure 2.12: (a) A typical write driver used for conventional 6T SRAM cell. (b) A write<br>driver for SRAM cells with distinct write bit lines. ....   | 29 |
| Figure 2.13: Functional diagram of delay-line based clocked timing block. ....   | 30 |
| Figure 3.1: Types of process variation. Due to the variation, threshold voltage (or any other<br>property) of any two (or three) transistors selected from different (or same) dies<br>will be different. .... | 32 |
| Figure 3.2: An example of process-induced threshold voltage variation affecting read<br>stability .....  | 35 |
| Figure 3.3: An example of process-induced threshold voltage variation affecting the<br>writability to the cell.....  | 36 |
| Figure 3.4: 7T cell proposed in [11]. ....   | 37 |
| Figure 3.5: 8T SRAM cell proposed in [12]. ....  | 39 |
| Figure 3.6: 9T SRAM cell proposed in [13]. ....  | 40 |
| Figure 3.7: Comparison of leakage consumption of various SRAM designs.....   | 41 |
| Figure 3.8: Comparison of area of various SRAM designs. ....   | 42 |
| Figure 4.1: The proposed 7T SRAM cell. ....  | 44 |
| Figure 4.2: Worst-case static noise margin for 7T-SRAM and 6T-SRAM.....  | 45 |

|  |    |
|--|----|
| Figure 4.3: (a) Floor plan, where multiple words per row is implemented. (b) Floor plan, where one word per row is implemented. Sophisticated ECC codes are required for multiple bit corruption.. | 48 |
| Figure 4.4: (a) Inverter with an access transistor. (b) 6T SRAM cell.  | 49 |
| Figure 4.5: The Forward-VTC and the Inverse-VTC form the “butterfly” curve of two cross-coupled inverters.   | 50 |
| Figure 4.6: (a) A schematic of the “modified” inverter. (b) Two cross-coupled “modified” inverters constituting a memory cell named Portless SRAM Cell.  | 51 |
| Figure 4.7: The Butterfly curve of the cross-coupled “modified” inverter.  | 52 |
| Figure 4.8: A basic clocked sense amplifier.   | 53 |
| Figure 4.9: The proposed single-ended Sense-Amplifier.   | 54 |
| Figure 5.1: The proposed 7T SRAM cell with transistor sizing.  | 57 |
| Figure 5.2: The proposed single-ended Sense-Amplifier with transistor sizing.  | 57 |
| Figure 5.3: Schematic of a column of the 7T SRAM cell along with write driver and sense-amplifier circuitry used to perform read and write operations.   | 58 |
| Figure 5.4: Schematic of a column of the 6T SRAM cell along with write driver and sense-amplifier circuitry used to perform read and write operations.   | 59 |
| Figure 5.5: Simulating array behavior with peripherals.  | 60 |
| Figure 5.6: Energy consumption per column in a write operation.  | 62 |
| Figure 5.7: Transient waveform during write operation. (a) The write bit lines (BL and BLB). (b) The storage nodes of the cell.  | 64 |

|  |    |
|--|----|
| Figure 5.8: Transient waveform of a cell where the write access transistor is OFF but one of the write bit line is discharged maximally. ....                          | 65 |
| Figure 5.9: A comparison of leakage currents of 6T cell and the proposed 7T cell as a function supply voltage. ....  | 68 |
| Figure 5.10: Time domain plots of cell node voltages (from Figure 2.2) for a state-flipping case. ....   | 69 |
| Figure 5.11: Comparison of critical charge between 6T and the proposed 7T SRAM cells.  | 71 |
| Figure 5.12: Comparison of SER between 6T and the proposed 7T SRAM cell. ....  | 72 |
| Figure 5.13: 7T cell Layout (The area inside the dotted boundary belongs to one cell). ....  | 73 |
| Figure 5.14: Waveform of the read bit line during read operation. ....   | 74 |
| Figure 5.15: Waveform of the two nodes of the latch inside the sense amplifier during read operation. (a) During '1' is being read. (b) During '0' is being read. .... | 75 |
| Figure 6.1: Memory array using Column Virtual Grounding (CVG). ....  | 77 |
| Figure 6.2: Array implementation of the proposed 7T SRAM cell with Column Virtual Grounding. ....  | 78 |
| Figure 6.3: Transient waveform of half-selected state. (a) When $V_{GND}=0V$ . (b) When $V_{GND}=300mV$ . ....   | 80 |
| Figure 6.4: A comparison of leakage currents of 6T cell and the proposed 7T cell as a function of rail-to-rail voltage. ....   | 81 |
| Figure 7.1: An enhanced version of the proposed cell. ....   | 85 |

## List of Tables

|  |    |
|--|----|
| Table 1: BL/BLB capacitance dependence to the stored data in the column .....  | 62 |
| Table 2: Energy consumption per column in a read operation. ....   | 65 |
| Table 3: Decoder energy consumption for asserting a word line during a read or write operation. ....   | 66 |
| Table 4: Total read delay. ....  | 66 |
| Table 5: Cell Leakage Current for $V_{DD}=1V$ . ....   | 67 |
| Table 6: Leakage comparison between with and without virtual grounding ( $V_{DD}=1V$ ). ....   | 81 |
| Table 7: The minimum average time between two consecutive access with CVG so that leakage power offsets the dynamic power needed for each access. .... | 82 |

# Chapter 1

## Introduction

The advancements of semiconductor technology have boosted the rapid growth of very large scale integrated (VLSI) systems in our day-to-day life. Microprocessors and systems-on-chip (SOCs) are now extensively used in a variety of applications ranging from smart phones to handheld computers, from entertainment systems to sophisticated automotive controllers, and from gaming devices to life-saving medical equipment. The processing speed or performance of these systems is primarily limited by the power budget, which is determined by the battery life for mobile devices. Since the performance demand of users is constantly increasing, it is critical to achieve as high performance as possible at the lowest possible power dissipation. An approach to meet this demand of performance is to increase the amount of memory embedded on the same chip with the microprocessor or the SOC. According to the Semiconductor Industry Association (SIA) International Technology Roadmap for Semiconductors (ITRS), more than half of the area of a typical IC design is occupied by embedded

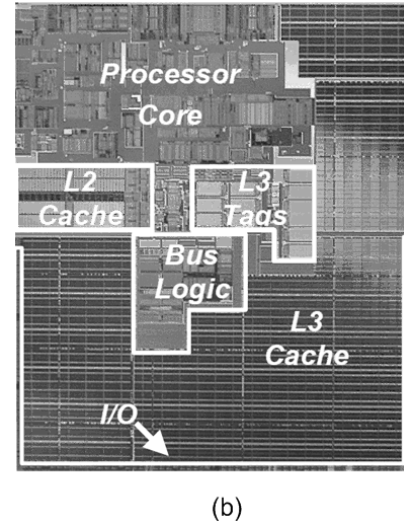
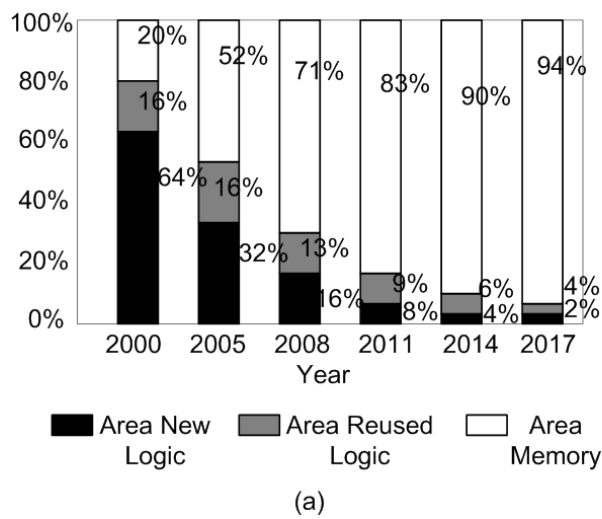


Figure 1.1: (a) Comparison of area of logic and memory in a SOC [1]. (b) Die photo of 1.5GHz Third Generation Itanium® 2 Processor [2].

memory (Figure 1.1(a)). Embedded memories are designed with rules more aggressive than the rest of the logic on a semiconductor chip. Accordingly as much as 70% chip area is dedicated to memories in present microprocessors and SOC's (Figure 1.1(b)). However, given the power constraint, increasing the size of the cache memory is very challenging and requires a bottom-up design approach from the bit-cell level to the architecture level.

## 1.1 Memory Hierarchy in Computer Systems

Ideally, a computer system will provide maximum performance when unlimited amount of fast memory is dedicated to itself [3]. However, implementing large-capacity memory with fast operation speed is not feasible due to the physical limitations of the electrical circuits. To circumvent this limitation, a computer system



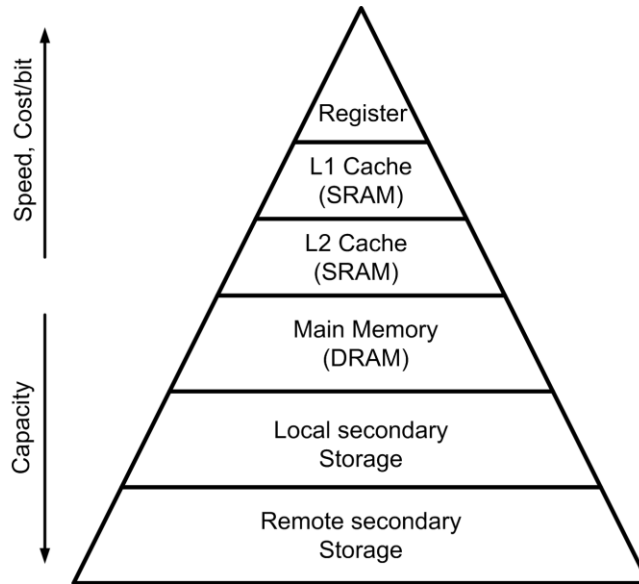


Figure 1.2: Memory hierarchy of a modern personal computer.

uses a variety of memory, which can be described through a memory hierarchy (shown in Figure 1.2). It is an arrangement of different types of memories with different capacities and operation speeds to approximate the desired unlimited memory capacity. At the top of the pyramid is the register, which is the closest to the processor core and is the fastest (typical cycle time is one CPU cycle  $\sim 0.25\text{ns}$ ) memory element. At the same time, it is the most expensive and hence the smallest memory. On the other hand, at the bottom of the pyramid is the slowest (cycle time  $\sim$  few seconds), largest, and cheapest memory element.

The cache memory is less expensive than the registers and can operate at a speed as close as the CPU speed. As can be seen in Figure 1.2, more than one level of cache memory can be used. The higher level cache will be smaller in size but its speed will be near the CPU clock speed while the lower level cache will have larger capacity but

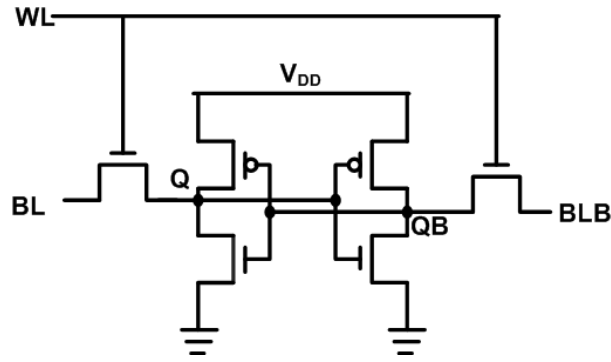


Figure 1.3: Schematic of a conventional six-transistor SRAM cell.

slower speed. Thus, fast cache access, entailed in a small sized cache, is provided while the larger (but slower) cache will provide data (and instructions) without requiring access to the off-chip main memory. Access to the off-chip main memory slows down the processing speed significantly because current high-end processors operate at 3-4GHz while even the fastest off-chip memory operates at 600MHz [4]. Primarily, the cache memory is realized by the static random access memory (SRAM) because of its compatibility with the standard logic process and the high operating speed. A typical SRAM consists of an array of cells that store the data bits and peripheral circuits that allow to access a cell in a given row and column. The cell consists of six transistors (6T) – four transistors form two complementary storage nodes (Q and QB) with a back-to-back inverter pair while the other two transistors allow access to the storage nodes (see Figure 1.3) [5]. The inverters continuously drive each other and the cell retains the data without any refresh mechanism as long as the power supply is provided. The cell is accessed for read or write operation by asserting the word line (WL). The functionality and power consumption of the cell depend on the proper sizing of the transistors, the operating voltage, and the fabrication process.

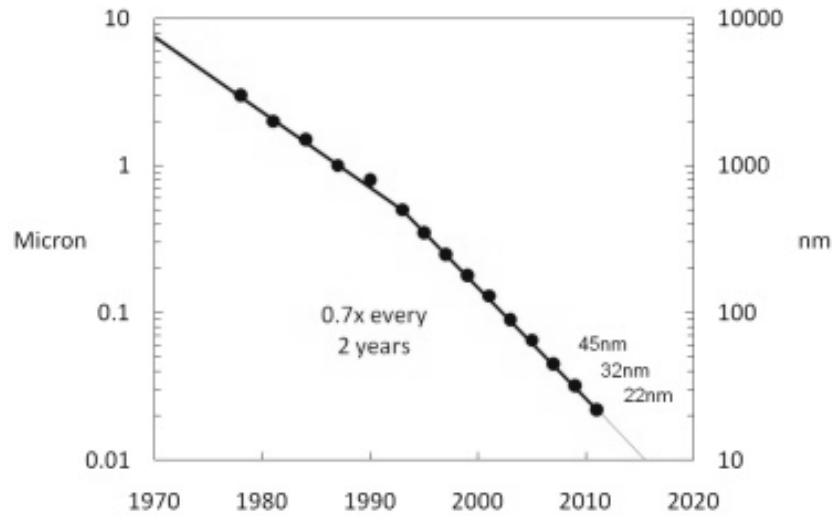


Figure 1.4: Scaling of transistor gate length according to Moore’s Law. Adapted from [6].

## 1.2 SRAM Design Challenges

The advancement in VLSI systems has primarily been achieved by the technology scaling where the transistor dimensions and operating voltage have been reduced. The scaling followed the famous Moore’s law, bringing the transistor gate length to as low as 22nm and the number of transistors per chip to as high as two billion (see Figure 1.4) [6]. As a result, the memory density is doubled in every process generation [7] as shown in Figure 1.5. However, scaling has brought in several challenges for the SRAM design. In particular, the increased process induced variations in transistor threshold voltage and dimensions, the higher leakage power consumption, and increased sensitivity to external noise sources, such as radiation induced single event voltage transients have become key concerns to address.

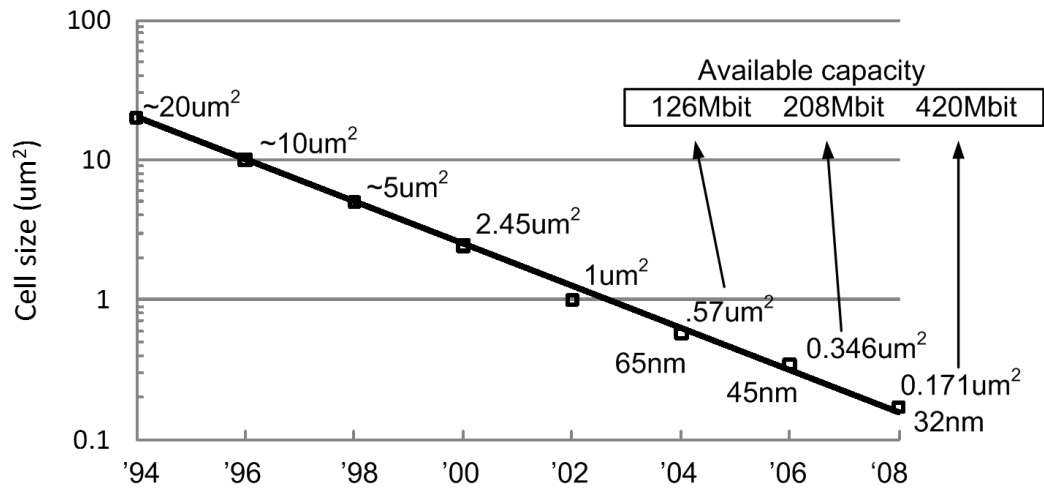


Figure 1.5: Scaling trend of SRAM bit-cell size [7].

### 1.2.1 Process Variations

The process technology is approaching the regime of fundamental randomness in the behavior of silicon structures. At the present technology nodes, we are trying to operate the devices at a scale where quantum physics is needed to explain the device operation and we are trying to define materials at the dimensional scale that is comparable to the atomic structure of silicon. In other words, the key dimensions of MOS transistor approach the scale of the silicon lattice distance, at which point the precise atomic configuration becomes critical to macroscopic device properties [8]. These are giving rise to increased process variations in the transistors' various properties, such as threshold voltage.

The transistors are fabricated on silicon by defining the N-well, diffusion area, the gate polysilicon and the metal connections. Photolithography with ultraviolet light is used to define these areas. Wavelength of ultraviolet light is in the range of 10 nm to

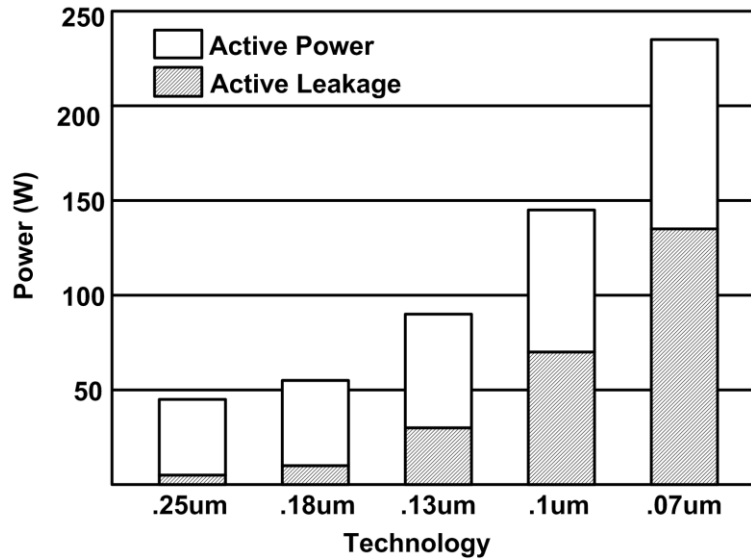


Figure 1.6: Leakage power and total power consumption of microprocessors with technology scaling [9].

400 nm. Since the dimensions of the minimum sized transistors are comparable to the wavelength of ultraviolet light, the photolithography process suffers from increased diffraction. As a result, the dimensions of the minimum sized transistors suffer from increased variation of length and width.

### 1.2.2 Leakage Power Consumption

An inescapable trend of the scaled process technologies is the increasing proportion of the leakage power consumption. Transistors in sub-100nm technologies exhibit higher leakage current because the geometry of the transistor keeps shrinking, which leads to higher leakage current in channel, gate and junction. Subsequently the leakage power consumption of SRAM has become more pronounced because high-performance VLSIs demands more and more on-chip SRAMs. As a result, leakage power consumptions in microprocessors and SOC have become dominant with technology scaling as shown in Figure 1.6. In fact, being the largest block and consisting of the

maximum number of transistors, SRAM leakage power consumption plays the cardinal role in sustaining battery life of portable devices.

### **1.2.3 Single Event Upset (SEU)**

The node capacitance decreases by about 30% in each new process technology due to transistor scaling [10]. As a result, the minimum amount of charge that can flip the logic state of any memory device decreased. Thus, electronic memory devices fabricated in the current process technologies have become very vulnerable against particle-induced SEU.

## **1.3 Motivation and Thesis Outline**

Extensive effort is being put to overcome the various SRAM design challenges. A number of SRAM topologies and techniques have been proposed in recent years to address these challenges [11], [12], [13]. However, most of these topologies usually incur high overhead in terms of silicon area, power consumption, and delay. As a result, the use of these topologies remained limited to specific applications. In this thesis, we propose a seven-transistor (7T) SRAM cell and low-leakage array architecture in order to increase the SRAM yield and minimize the leakage power consumption and SER.

The proposed cell utilizes decoupled read bit line from the write bit lines. Thus, the cell has higher data stability during read operation and yield under varying process, voltage, and temperature (PVT) conditions. The cell utilizes a unique write

mechanism which reduces the write power to less than half of the write power consumed by the traditional 6T SRAM cell. It also exhibits lower SEU or soft error rate (SER). It can be laid out on silicon without any area overhead compared to the 6T SRAM cell. By integrating with a column-based gated-ground or virtual ground technique, the leakage power is significantly reduced. The column virtual grounding technique also supports multiple words per row, enabling efficient bit-interleaving to achieve even lower SER with conventional error correcting codes (ECC). The proposed bit-cell being single-ended, a 7-transistor single-ended sense-amplifier is also proposed in this thesis.

The thesis document is organized as follows. Chapter 2 presents an overview of the SRAM architecture. Chapter 3 discusses the impact of process variations on SRAM data stability and existing solutions to tackle that. Chapter 4 presents the proposed 7T cell and sense-amplifier, and their operation principles. Chapter 5 compares the performance of the proposed 7T SRAM cell with the conventional 6T SRAM cell. Chapter 6 presents a low power array-architecture utilizing the column virtual grounding techniques. Finally, Chapter 7 summarizes the contributions of this work to the field of embedded memory and presents some directions for future work.

# Chapter 2

## SRAM Architecture and Operation

### 2.1 Basic SRAM Architecture

A typical SRAM consists of an array of memory cells along with some peripheral circuits. The peripheral circuits include the row decoder, column decoder, address buffer for row and column decoders, sense amplifier, precharge circuitry, and data buffers. While the construction of the SRAM array can be very complex depending on the memory size, area, and speed requirements, a basic array consists of  $2^L$  rows and  $N \times 2^K$  columns of cells. Here  $L$  is the number of address bits for the row decoder,  $K$  the number of address bits for the column decoder, and  $N$  the number of bits in a word (Figure 2.1). There are  $2^L$  word lines, only one of which is activated by the row decoder based on the row address bits (bits  $A_0$  to  $A_{L-1}$  in Figure 2.1) at a given time instant. On the other hand,  $K$  address bits are decoded to select one of the  $N$ -bit words from a given row. Most of the recent microprocessors operate with 64-bit words and hence are referred to as 64-bit processors. Thus, the SRAM array for such systems will have  $2^K \times 64$  (or  $2^{K+6}$ ) columns of cells in total. Usually  $K$  and  $L$  are selected in such a



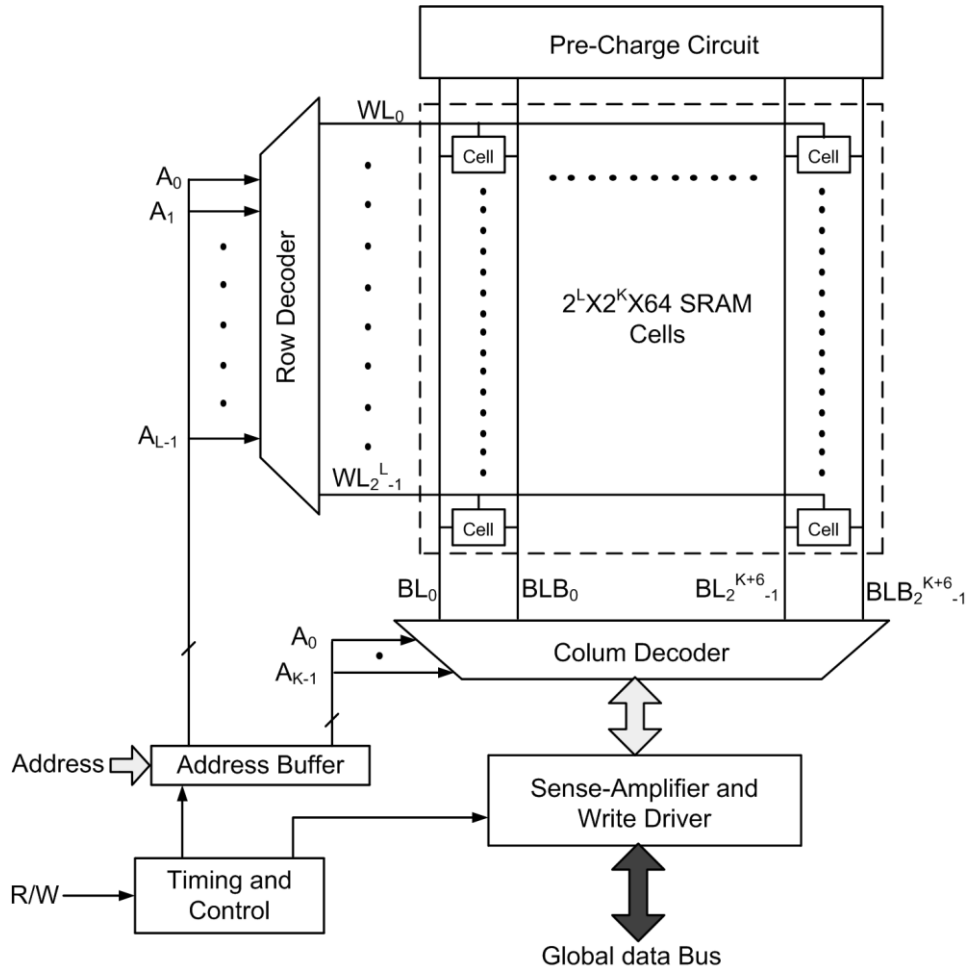


Figure 2.1: A typical SRAM architecture.

way that the overall array assumes a square shape when laid out. Thus,  $2^{K+6} = 2^L$  or  $K+6=L$  can be tentatively used for a layout optimized array for square-shaped cells. The choice of using row select bits as MSB and column select bit as LSB of the entire address bits or vice versa is arbitrary. The timing of the activation of sense amplifier, write driver, decoders and other peripherals are controlled by a timing circuitry. The read/write (R/W) signal determines whether the SRAM is to be read or written.

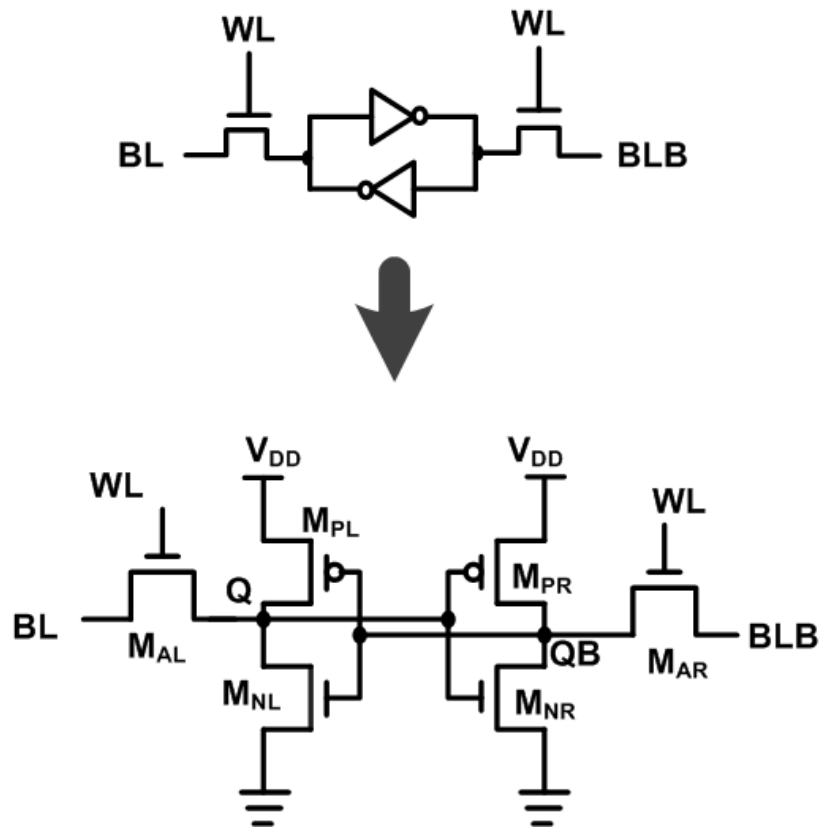


Figure 2.2: Conventional 6T SRAM cell.

## 2.2 6T SRAM Cell

The most widely used SRAM bit-cell is the six transistor (6T) cell shown in Figure 2.2. It consists of a back-to-back inverter latch and two access transistors. The latch holds the data bit while the access transistors are used for read and write operation. Access transistors also isolate the cells from the bit lines (BL and BLB) when they are not accessed. As opposed to DRAM, an SRAM cell has to provide non-destructive read operation and the ability to indefinitely retain data without any refresh operation (given the power is supplied to the cell).

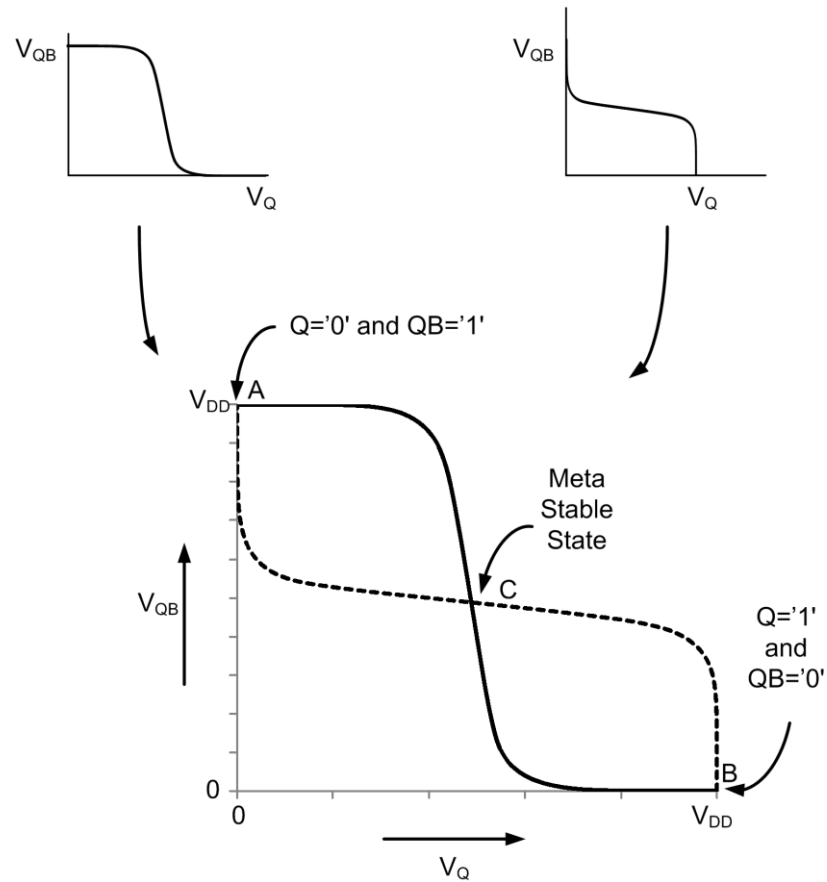


Figure 2.3: The VTCs of two cross-coupled inverters forming the butterfly curve of the SRAM cell.

The 6T SRAM cell has been used by the semiconductor industry in today's SOCs and microprocessors. Accordingly, the 6T SRAM cell will be discussed in detail, paving the foundation of the development of a new bit-cell in this thesis.

The two cross-coupled inverters inside the 6T cell form a bistable circuit with a positive feedback. The voltage transfer characteristics (VTC) of the inverters can be combined to generate the butterfly curve shown in Figure 2.3. When the access transistors are OFF, the cell acts as an isolated latch and the VTCs have three

intersecting or operating points  $A$ ,  $B$ , and  $C$  (see Figure 2.3). Among these three points, the latch can remain in either  $A$  or  $B$ . The third point  $C$  represents an unstable state where the latch cannot practically stay. A small deviation from this state, caused by a small noise, is amplified and regenerated around the feedback loop. As a result, the latch either goes to state  $A$  or  $B$  and remains there.  $A$  and  $B$  states correspond to the storing of two complementary values, namely '0' and '1'. When the latch is in state  $A$ , it can be said that the cell is storing '0' ( $Q=0$ ) and when in state  $B$  the cell is storing '1' ( $Q=1$ ). As long as the power supply is ON, the cell will continue to store that data without any refresh operation. The stability of state  $A$  (and  $B$ ) is quantitatively denoted by static noise margin (SNM). SNM is defined as the maximum sized square that can be inscribed inside the butterfly curve [14].

### 2.2.1 Read Operation

The read operation is initiated in a 6T SRAM cell by asserting WL in order to turn on the access transistors. Another pre-condition for the read operation is that the bit lines be precharged to the supply voltage,  $V_{DD}$ . However, the bit lines have to be kept floating to avoid any contention with the driver NMOS transistor inside the cell. If the driver NMOS transistor discharges a bit line, it has to be ensured that no other circuitry charges the bit line at the same time.

Let us now assume that the cell is in state  $A$  ( $Q=0$  and  $QB=1$ ). When WL signal is asserted,  $M_{AL}$  is turned ON while  $M_{AR}$  remains OFF as its gate-to-source voltage is 0 (see Figure 2.4). Consequently, no current will flow through  $M_{AR}$  and BLB will stay at the precharged voltage ( $V_{DD}$ ). Conversely, the voltage difference

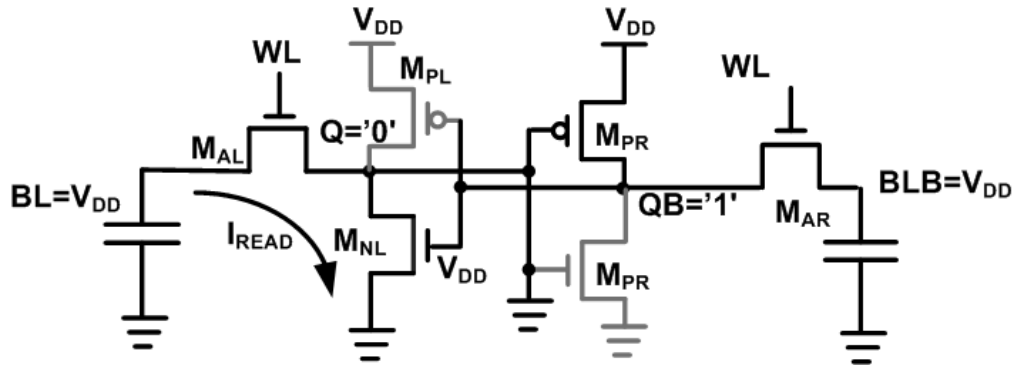


Figure 2.4: 6T SRAM cell during a read operation (The transistors in grayscale are OFF).

across  $M_{AL}$  will cause a current ( $I_{READ}$ ) to flow from BL to ground, discharging BL. Had the cell been read while being in state  $B$  ( $Q='1'$  and  $QB='0'$ ) BLB would have been discharged and BL would have stayed at  $V_{DD}$ .

As shown in Figure 2.4,  $I_{READ}$  forms a voltage divider between the BL and ground with  $M_{AL}$  and  $M_{NL}$ . As a result, the potential at node Q ( $V_Q$ ) is elevated from 0V to a non-zero potential,  $\Delta V$ .  $\Delta V$  can be termed as the logic '0' degradation as it increases the logic '0' voltage and reduces the SNM. The value of  $\Delta V$  should be as low as possible for the data stability. In fact, in order to avoid any unintentional flipping of the stored data,  $\Delta V$  should be less than the switching threshold voltage,  $V_{TRIP}$ , of the cross-coupled inverter pair.

From Figure 2.4 it can be seen that the magnitude of  $\Delta V$  depends on the relative strength of  $M_{AL}$  and  $M_{NL}$ . A quantitative measure of  $\Delta V$  can be easily found out by equating the currents ( $I_{READ}$ ) through  $M_{AL}$  and  $M_{NL}$ . Assuming  $M_{AL}$  in the saturation region and  $M_{NL}$  in the linear region of operation, some mathematical manipulation yields [15]:

$$\Delta V = \frac{V_{DSATn} + CR(V_{DD} - V_{Tn}) - \sqrt{V_{DSATn}^2(1 + CR) + CR^2(V_{DD} - V_{Tn})^2}}{CR}, \quad (2.1)$$

Here,  $V_{Tn}$  is the threshold voltage and  $V_{DSATn}$  is the saturation drain-to-source voltage of the NMOS, and CR is called the cell ratio, which is defined as  $\frac{W_{MNL}/L_{MNL}}{W_{MAL}/L_{MAL}}$ . It should be noted that CR is the same for also  $M_{NR}$  and  $M_{AR}$  since the cell is symmetrical by design. In our study with a commercial 65nm technology, CR=1.5 showed a reasonable read stability under various process and mismatch corners.

During the read operation, since one of the bit lines (BL in the above discussion) is discharged by  $I_{READ}$  while the other bit line remains at the precharged voltage, there will be a voltage difference between the bit lines. Based on the differential voltage at the bit lines, the sense amplifier makes the decision of which value ('0' or '1') was stored and hence is being read from the SRAM cell.

### 2.2.2 Write Operation

The write operation on the cell is also done by asserting the WL. However, before the WL assertion, one of the bit lines is pulled down to 0 V from its precharged state based on the data intended to be written. For an example let us assume that  $Q='0'$  (and  $QB='1'$ ) in a cell and the cell is to be written to  $Q='1'$  ( $QB='0'$ ). To do that, BLB is discharged to 0V and BL is precharged to  $V_{DD}$ . Then, WL is activated.

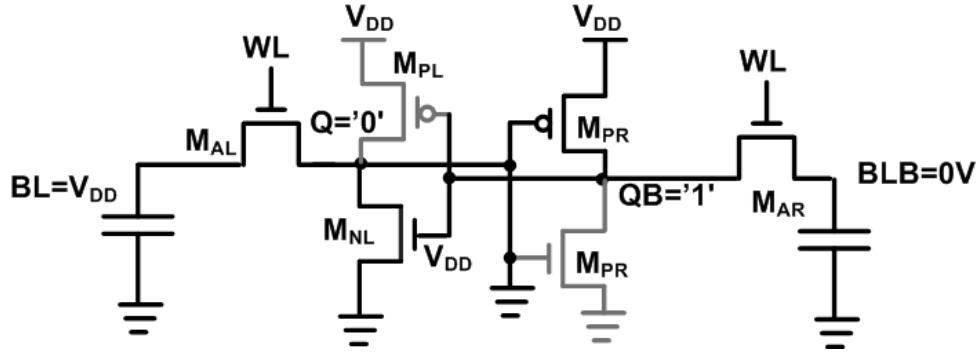


Figure 2.5: 6T SRAM cell during a write operation (The transistors in grayscale are OFF).

Since BL is precharged to  $V_{DD}$ , activating WL puts  $M_{AL}$  in a condition similar to the read operation (see Figure 2.5). Since the node Q stores '0',  $V_Q$  will be elevated to  $\Delta V$ . However, the sizing of  $M_{AL}$  and  $M_{NL}$  (or  $M_{AR}$  and  $M_{NR}$ ) is determined by CR, which is chosen in such a way that  $\Delta V$  stays well below  $V_{TRIP}$ . As a result, the write operation cannot be accomplished from the side that stores '0' (node Q in Figure 2.5).

On the other hand, since  $QB='1'$  and BLB is pulled to 0V,  $V_{QB}$  will be pulled down from '1' ( $V_{DD}$ ) to an intermediate voltage level by  $M_{AR}$ . If  $V_{QB}$  falls below  $V_{TRIP}$  of the inverter  $M_{PL}-M_{NL}$ , then  $M_{PL}$  will be turned ON and  $M_{NL}$  will be turned off, pulling node Q to '1' and flipping the cell. Thus, the write operation is always accomplished from the side that stores '1' before accessing the cell. In order to ensure that  $V_{QB}$  falls below  $V_{TRIP}$  of inverter  $M_{PL}-M_{NL}$ ,  $M_{AR}$  has to be made stronger than  $M_{PR}$ . The quantitative condition to meet this requirement can be derived by equating the current through  $M_{PR}$  and  $M_{AR}$  [15]:

$$V_{QB} = V_{DD} - V_{Tn} - \sqrt{(V_{DD} - V_{Tn})^2 - 2 \frac{\mu_p}{\mu_n} PR \left( (V_{DD} - |V_{Tp}|) V_{DSATp} - \frac{V_{DSATp}^2}{2} \right)}, \quad (2.2)$$

Here,  $V_{Tn}$  and  $V_{Tp}$  are threshold voltages of NMOS and PMOS, respectively,  $V_{DSATp}$  is the saturation drain-to-source voltage of PMOS, and  $\mu_p$  and  $\mu_n$  are the mobilities of PMOS and NMOS transistors, respectively. PR is called the cell pull-up ratio, which is defined as  $PR = \frac{W_{MPR}/L_{MPR}}{W_{MAR}/L_{MAR}}$ . From a design perspective, the stronger  $M_{AR}$  (or  $M_{AL}$ ) is, the lower  $V_{QB}$  is pulled down to. Since an NMOS typically has a higher mobility than a PMOS, the minimum-sized PMOS pull-up and NMOS access transistors and hence a PR of 1 is used. PR is the same for  $M_{PL}$  and  $M_{AL}$  since the cell is symmetric.

From above discussion, it can be seen that the cell access transistors have to be weak enough to ensure stability during a read operation on one hand, and have to be strong enough to ensure writability during a write operation on the other hand. This apparent contradictory design requirement makes the 6T cell design challenging, particularly in scaled CMOS technologies, which suffer from increased process variations. Nonetheless, the 6T cell has been the workhorse for the embedded memories over the past decades because of its excellent noise margin, minimal leakage power consumption, and high speed of operation. In addition, it is fully compatible with the standard logic process that is used to realize the rest of the logic processing circuits on the same silicon die.



## 2.3 Row Decoder

Row decoder is primarily a binary decoder. The inputs of the decoder are the address bits while the outputs are the word line (WL) signals, each of which is used to select a row of the SRAM cell array. For an n-bit address input, the row decoder enables one of  $2^n$  word line signals. Typically, the address bits for the row decoder are a subset of the total address bits. For example, if L=8 and K=3 in Figure 2.1, then the total address will be 11-bit long. Out of those 11 bits, 8 bits will be used as input to the row decoder, which will control 256 WLs.

If  $A_0$ - $A_7$  are the input bits of a row decoder, the logical function of the row decoder can be expressed as:

$$WL_0 = \overline{A_7 A_6 A_5 A_4 A_3 A_2 A_1 A_0} = \overline{A_7 + A_6 + A_5 + A_4 + A_3 + A_2 + A_1 + A_0} \quad (2.3)$$

$$WL_1 = \overline{A_7 A_6 A_5 A_4 A_3 A_2 A_1} A_0 = \overline{A_7 + A_6 + A_5 + A_4 + A_3 + A_2 + A_1 + \overline{A_0}} \quad (2.4)$$

$$WL_2 = \overline{A_7 A_6 A_5 A_4 A_3 A_2} A_1 \overline{A_0} = \overline{A_7 + A_6 + A_5 + A_4 + A_3 + A_2 + \overline{A_1} + A_0} \quad (2.5)$$

An obvious way to implement these function is by using a wide NAND or NOR gate. But that poses a number of design challenges. First, the layout of the wide NAND (or NOR) gate must fit within the word line pitch. Second, the large fan-in of the gate will have negative effect on the performance of the circuit, particularly in terms of delay (delay is usually proportional to the square of the fan-in). Thus, implementing wide NAND (or NOR) is not a practical solution [15].

An efficient way to implement the entire row decoder is by utilizing the large amount of redundancy, which is inherently present at the decoder outputs. For example, the three logical functions shown in (2.3) – (2.5) can be re-arranged to yield the following:

$$WL_0 = (\bar{A}_7\bar{A}_6)(\bar{A}_5\bar{A}_4)(\bar{A}_3\bar{A}_2)(\bar{A}_1\bar{A}_0) \quad (2.6)$$

$$WL_1 = (\bar{A}_7\bar{A}_6)(\bar{A}_5\bar{A}_4)(\bar{A}_3\bar{A}_2)(\bar{A}_1A_0) \quad (2.7)$$

$$WL_2 = (\bar{A}_7\bar{A}_6)(\bar{A}_5\bar{A}_4)(\bar{A}_3\bar{A}_2)(A_1\bar{A}_0) \quad (2.8)$$

We can see that the term  $(\bar{A}_7\bar{A}_6)(\bar{A}_5\bar{A}_4)(\bar{A}_3\bar{A}_2)$  is used in more than one case (4 to be exact). Thus, it is not necessary to generate  $(\bar{A}_7\bar{A}_6)(\bar{A}_5\bar{A}_4)(\bar{A}_3\bar{A}_2)$  in all 4 instances. Instead, it can be generated only once and then used 4 times with  $(\bar{A}_1\bar{A}_0)$ ,  $(\bar{A}_1A_0)$ ,  $(A_1\bar{A}_0)$ , and  $(A_1A_0)$ . This is equivalent to splitting a complex gate into two or more layers of logic. It results into faster and cheaper implementation in terms of power and silicon area. Thus, the address is decoded in segments where the segments other than the final decoding segments are called predecoder (see Figure 2.6).

The final stage of the row decoder has maximum number of transistors. For the 8-to-256 row decoder, there will be 256 word line drivers each consisting of a NAND gate and an inverter, as shown in Figure 2.7. Since the inverter has to drive a highly capacitive word line, its transistors have to be relatively larger. However, larger transistors consume higher leakage current. It should be noted that in the active mode only one of the word line driver is activated. The rest of the circuit still remains inactive. In inactive mode, all  $WL_K$  ( $K = 0, 1, 2, \dots, 255$ ) are LOW and all  $P_K$  nodes

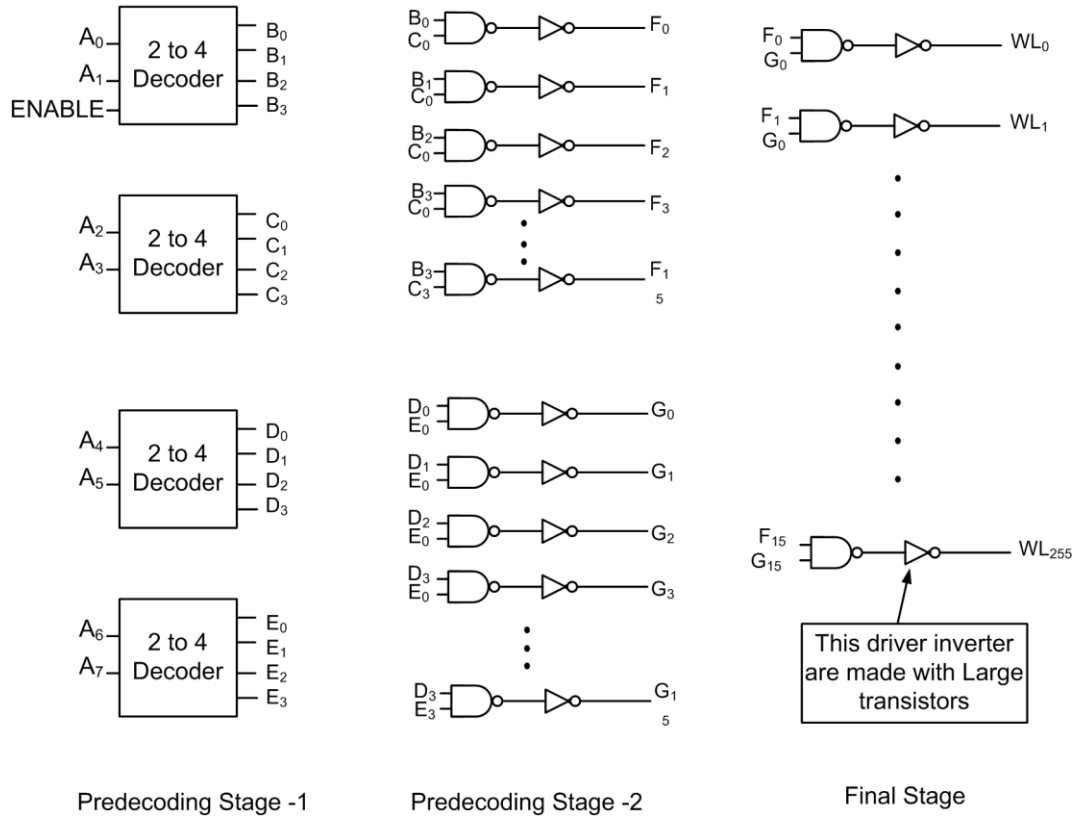


Figure 2.6: Segmented decoding of address bits in a row decoder

are HIGH i.e.,  $V_{DD}$  ( see Figure 2.7). When the input of an inverter is HIGH, the leakage is determined by the PMOS transistor, which is in the sub-threshold region. Therefore, the PMOS transistor connection inside the inverter has to be modified for reducing the leakage power consumption. An efficient way to achieve this goal is to apply the gate-source self-reverse biasing (GSSRB) [17] by using stacked transistor, as shown in Figure 2.7 by  $M_{P1}$  and  $M_{P2}$ . The gate-source voltage of  $M_{P1}$  is 0V. However, the voltage of  $S_K$  is approximately midway between 0V and  $V_{DD}$ . Thus, the gate-source voltage of  $M_{P2}$  is positive and  $M_{P2}$  will have reverse gate-source biasing. As a result, the leakage current will be drastically reduced by  $M_{P2}$ .

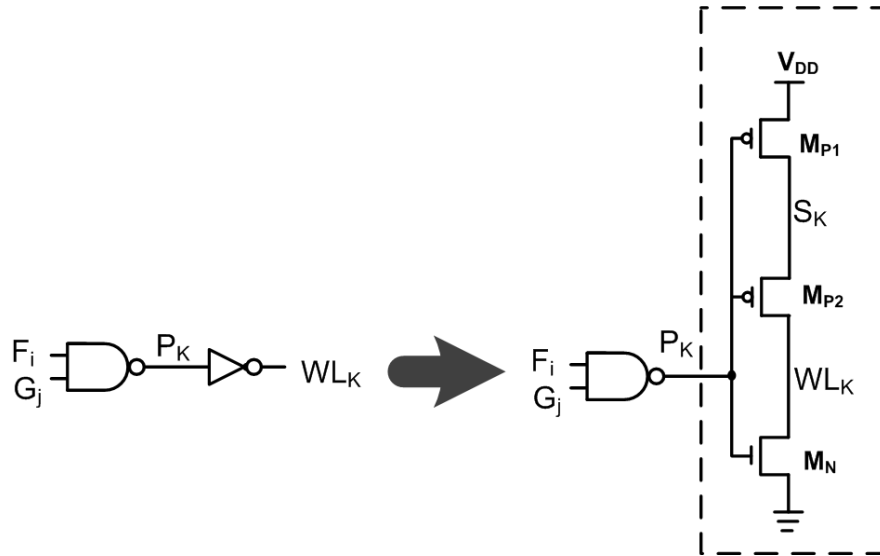


Figure 2.7: A word line driver circuit to reduce PMOS leakage current.

## 2.4 Column Decoder or Multiplexer

The aspect ratio of an SRAM array is typically made close to unity so that the bit line and word line capacitances are in the same order of magnitude. This is achieved by putting multiple words per row. For example, if a word consists of 64 bits and an SRAM array of 1024 words needs to be constructed, then putting one word per row would result in 64 cells per row and 1024 cells per column (see Figure 2.8(a)). Consequently, the bit line would become too long and its capacitance would become significantly larger than the capacitance of a word line. On the other hand, placing four words per row results in 256 cells per row and 256 cells per column. If the cell is assumed square shaped, the latter arrangement is preferable to balance the bit line and word line capacitances. However, in order to accommodate multiple words per row, a

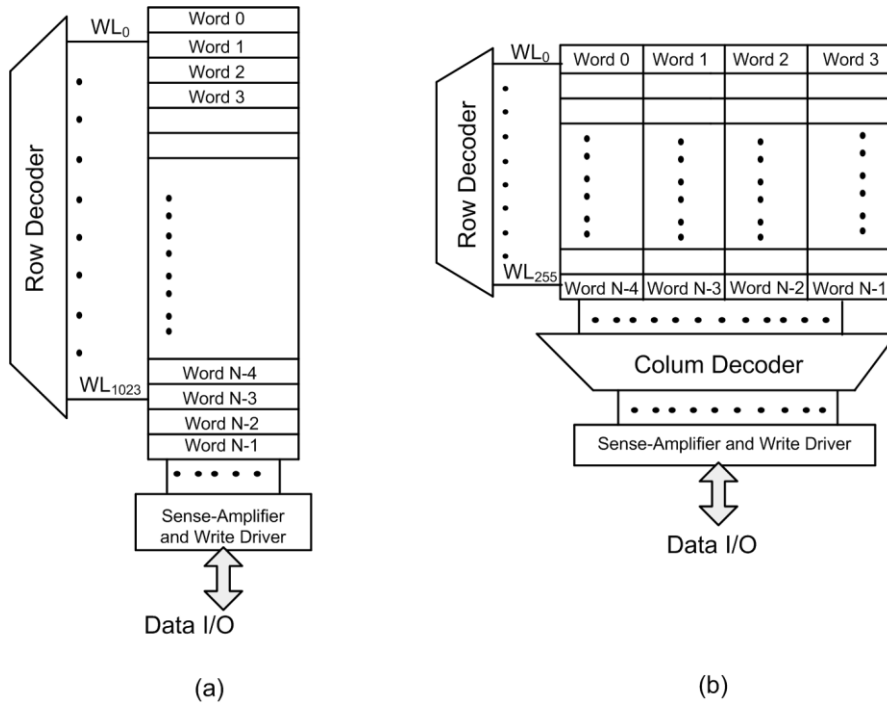


Figure 2.8: An SRAM array with: (a) single word per row and (b) multiple words per row.

column decoder or multiplexer (MUX) is needed to multiplex the words of a row to a set of sense amplifiers, which equal the number of bits in a word.

Two typical implementations of the column decoders are shown in Figure 2.9. Figure 2.9(a) shows a column decoder with PMOS pass-transistors and a 2-to-4 pre-decoder. Based on the inputs  $A_1$  and  $A_0$ , only one of the PMOS is turned on at a time and passes the bit line voltage from one of the four columns to the inputs of a sense amplifier. A more efficient version of the column decoder is shown in Figure 2.9(b). It is called a binary tree decoder formed by PMOS pass transistors. The tree decoder does not require any predecoding stage and utilizes fewer transistors. However, the propagation delay in the tree decoder increases quadratically with the number of

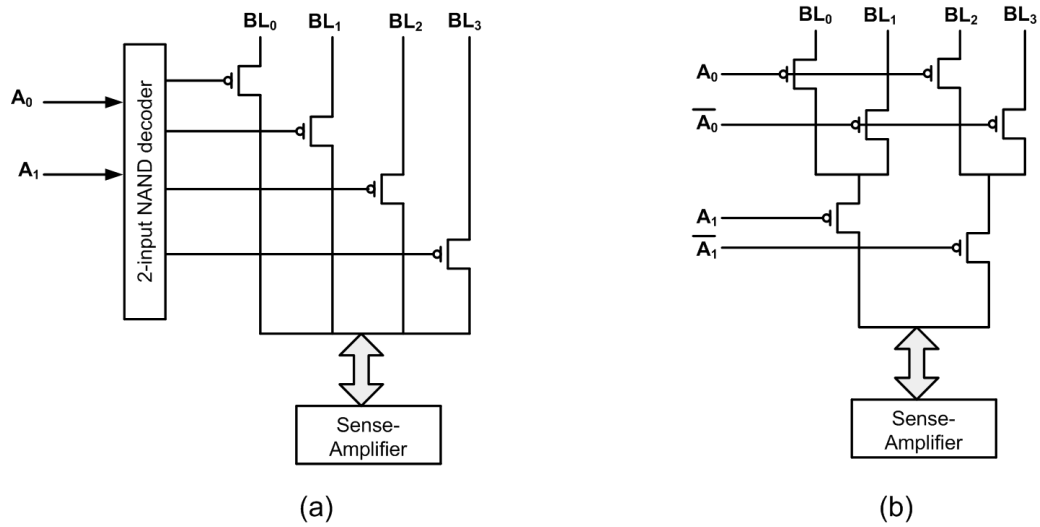


Figure 2.9: 4-to-1 column MUX: a) pre-decoder based and b) tree based.

PMOS transistor sections. A large tree-based column decoder introduces too much delay, which can affect performance, limiting the application of the tree decoder [15].

## 2.5 Sense Amplifier

The sense amplifier is used to facilitate the read operation. The read operation in the conventional 6T SRAM cell is differential. During a read operation the stored data inside the SRAM cell appears on BL and the complement of the stored data appears on BLB. However, the data is not directly read from the bit lines. If the data is directly read from the bit lines, then one of the bit lines has to be discharged to 0V. Since the bit lines are highly capacitive, discharging a bit line to 0V would make the subsequent precharging consume a significant amount of power. In addition, SRAM cells are made as small as possible in order to maximize the memory capacity in a given silicon area. The current driving capability of the SRAM cell's read discharge path is very

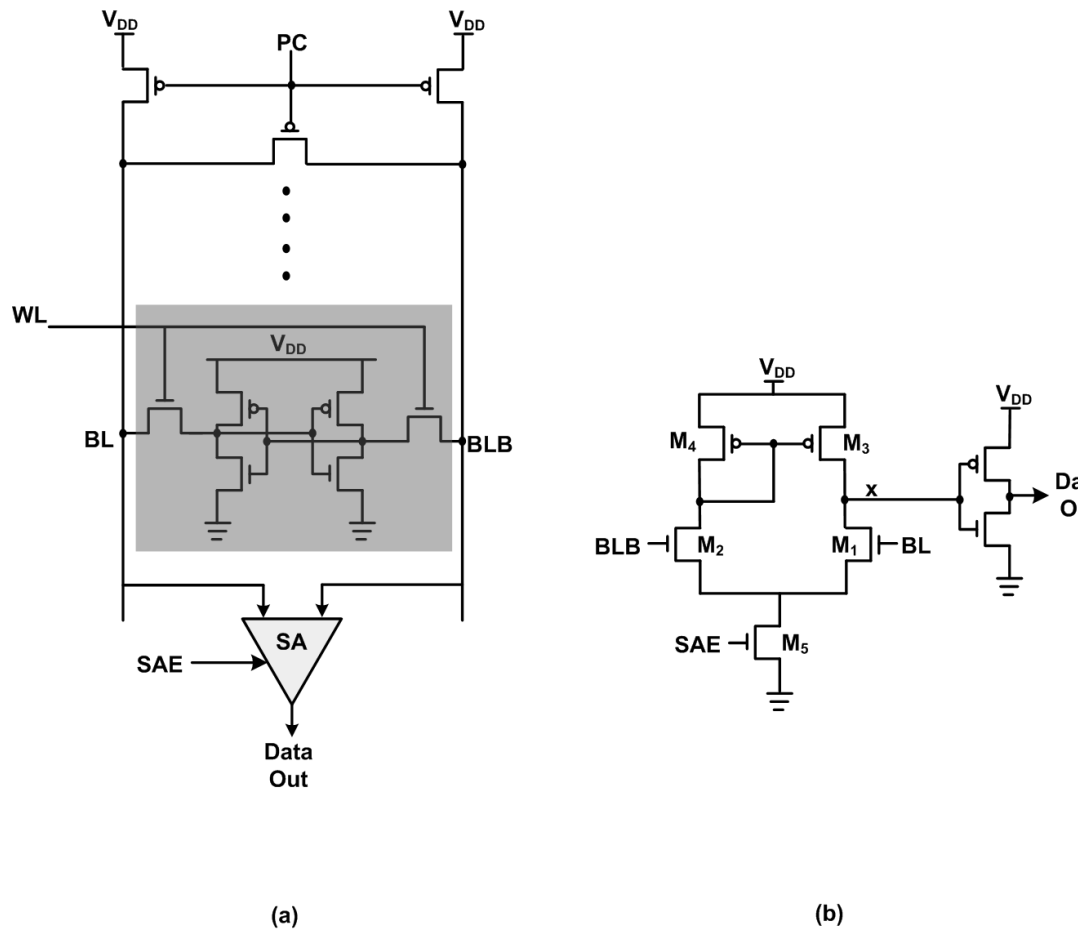


Figure 2.10: (a) SRAM column with the sense amplifier and precharge circuits and (b) Basic differential sense amplifier with current mirror load.

low. If such a low current drive is used to discharge the highly capacitive bit lines, it would take a large amount of time. Sense amplifier is used to avoid these problems. The sense amplifier works as a buffer (see Figure 2.10(a)) between the bit lines and the node from where ultimately the data is read, which is comparatively less capacitive than the bit lines. Instead of being completely discharged, the bit lines are typically discharged by 10%-15% of  $V_{DD}$ . That way both the subsequent precharge power and the discharge delay is reduced.

Sense amplifier is an amplifier that has very high gain when activated. The bit lines are used as input to the sense amplifier. During a read operation, one of the bit lines is discharge and a voltage differential between them is generated. At the same time, the sense amplifier is biased in an operating point with high gain. In some sense amplifiers this high gain is achieved by positive feedback. When the bit line voltage differential is applied, it is amplified due to the high gain of the sense amplifier. As a result the output of the sense amplifier will either saturate to 0V or  $V_{DD}$ .

There have been several topologies of sense amplifiers. Each has been developed with a particular type of operation and goal in mind. However, since sense amplifier is an additional component in the read critical path, it should have a number of performance characteristics. In general, a sense amplifier should exhibit small delay, consume low power, and use a small number of transistors to limit the layout area, which has to be pitch-matched with the cell columns.

The basic single-stage differential sense amplifier with current mirror load is shown in Figure 2.10(b). Actually, this sense amplifier does not utilize positive feedback. It derives its high gain from the current mirror load ( $M_3$ ) and transconductance of  $M_1$ . A gain of around 100 can be achieved by this sense amplifier. However, the primary goal of the sense amplifier is to minimize the response time, i.e., to quickly generate the full logic-level output signal. Thus, gain of the sense amplifier is secondary to the response time and a gain of around 10 is typically used [15].



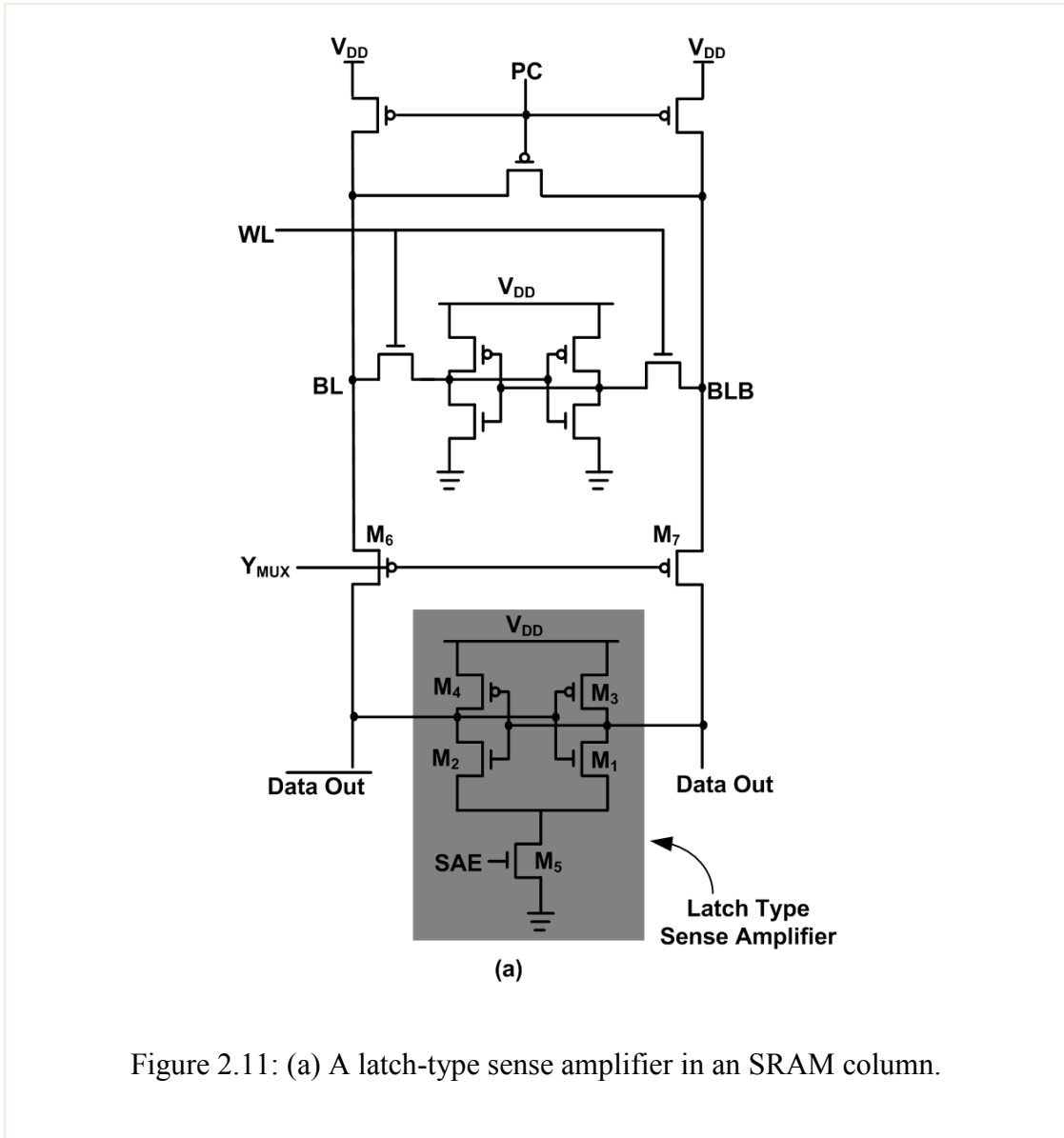


Figure 2.11: (a) A latch-type sense amplifier in an SRAM column.

Another topology of the SRAM sense amplifier is the latch-type sense amplifier shown in Figure 2.11. This sense amplifier utilizes a positive feedback to achieve a high gain. The amplifier consists of a pair of cross-coupled inverters. The sensing is initiated by biasing the sense amplifier in the high-gain region (i.e., at the metastable point of the inverters) by precharging and equalizing its outputs  $\overline{\text{Data Out}}$  and Data Out to  $V_{DD}$ . Thus, the inputs (bit lines) are not isolated from the outputs.

Additional transistors,  $M_6$  and  $M_7$  are used to isolate the latch-type sense amplifier from the bit lines. When word line is asserted and sufficient voltage differential is generated between the bit lines, the transistor  $M_6$  and  $M_7$  are turned off, thus isolating the bit lines from the output of the sense amplifier. Then, the sense amplifier is activated and based on the data stored in the cell, i.e., the differential voltage on the bit lines, either one of  $\overline{\text{Data Out}}$  and Data Out becomes 0V while the other one becomes charged to  $V_{DD}$ , which will produce a full logic level output.

## 2.6 Write Drivers

The write driver is used during the write operation in order to discharge one of the bit lines. In the 6T SRAM array, write drivers typically discharge the bit line to 0V to ensure successful write operation in all process and mismatch corners. When write driver is enabled, the precharge circuit is usually deactivated to avoid any contention. Based on the application, a write driver circuitry can be implemented in different ways. A typical write driver circuit is shown in Figure 2.12(a).

In 6T SRAM cells, same bit lines are used for read and write operations. For other SRAM cells ([12], [13]), which have bit lines dedicated for the write operation only, the write driver can be modified to include the precharge circuit as well. In such cases, write bit line is only discharged during write operation. Thus, the discharge and subsequent precharge of the write bit line can be solely controlled by the write enable signal. The write driver for such an SRAM is shown in Figure 2.12(b).

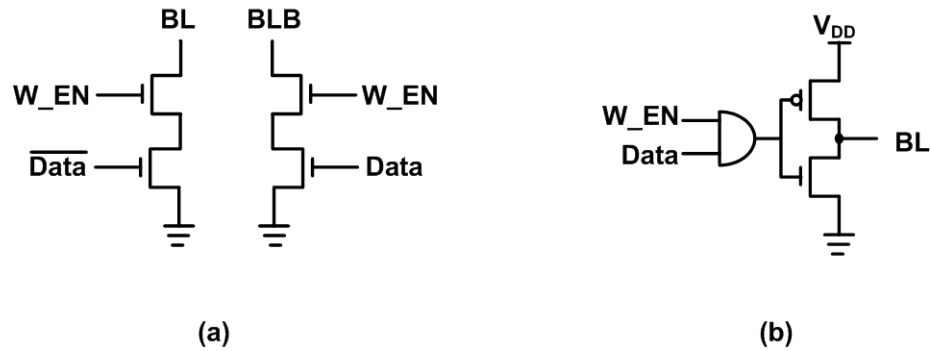


Figure 2.12: (a) A typical write driver used for conventional 6T SRAM cell. (b) A write driver for SRAM cells with distinct write bit lines.

It should be noted that one write driver is needed for one entire column. Thus, the strength of the write driver transistors is not constrained by size. They can be made large to expedite the discharge speed. As a result, the large area required by the large pull-down transistor of a write driver does not pose any challenge in the array layout.

## 2.7 Timing and Control Circuits

The operation of the SRAM consists of a strict sequence of actions such as address latching, word line decoding, bit line precharging and equalization, sense-amplifier enabling, and output driving. For proper operation, this sequence must be maintained under all operating conditions. This necessitates a precise timing and synchronization among the different actions. A timing and control circuitry is used to serve this purpose.

The various timing approaches used for designing the timing and control circuitry can be primarily categorized into clocked approach and self-timed approach. A

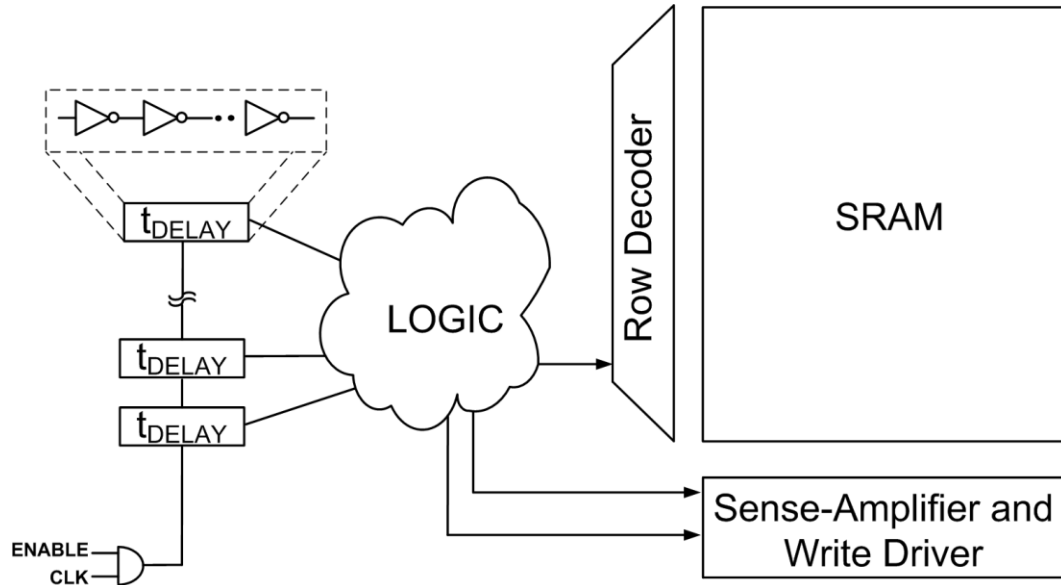


Figure 2.13: Functional diagram of delay-line based clocked timing block.

detailed discussion of these timing approaches would be very long and hence is beyond the scope of this thesis. Figure 2.13 shows a timing control circuit based on the clocked approach. The circuit takes the clock as the reference signal and generates a series of control signals using inverter chain-based delay elements. The control signals are then fed to different sub-block of the SRAM. Such a timing control circuit has been employed for the simulation test bench used in this thesis.

# Chapter 3

## Impact of Process Variation on SRAMs

### 3.1 Process Variation

The most prominent challenge in semiconductor process technology is the increased process variations. These variations deviate the transistor operations from their expected behavior. When the deviation is too large, the electronic circuit ceases to function as it was designed to do which result in yield loss. To address this problem design level and process level measures are taken. Process level measures are beyond the scope of this thesis. In this thesis, only design level measure is discussed. During design stage of any electronic circuit sufficient margin is kept so that even after the deviation in behavior, the resulting IC still performs as it was intended to do. However, keeping too much margin in the design level means increased cost in terms of power consumption and silicon area. Thus, it requires careful analysis of the circuit operation and various process variations which are the most critical to electronic circuit operations, especially memory circuit operations. The performance, power

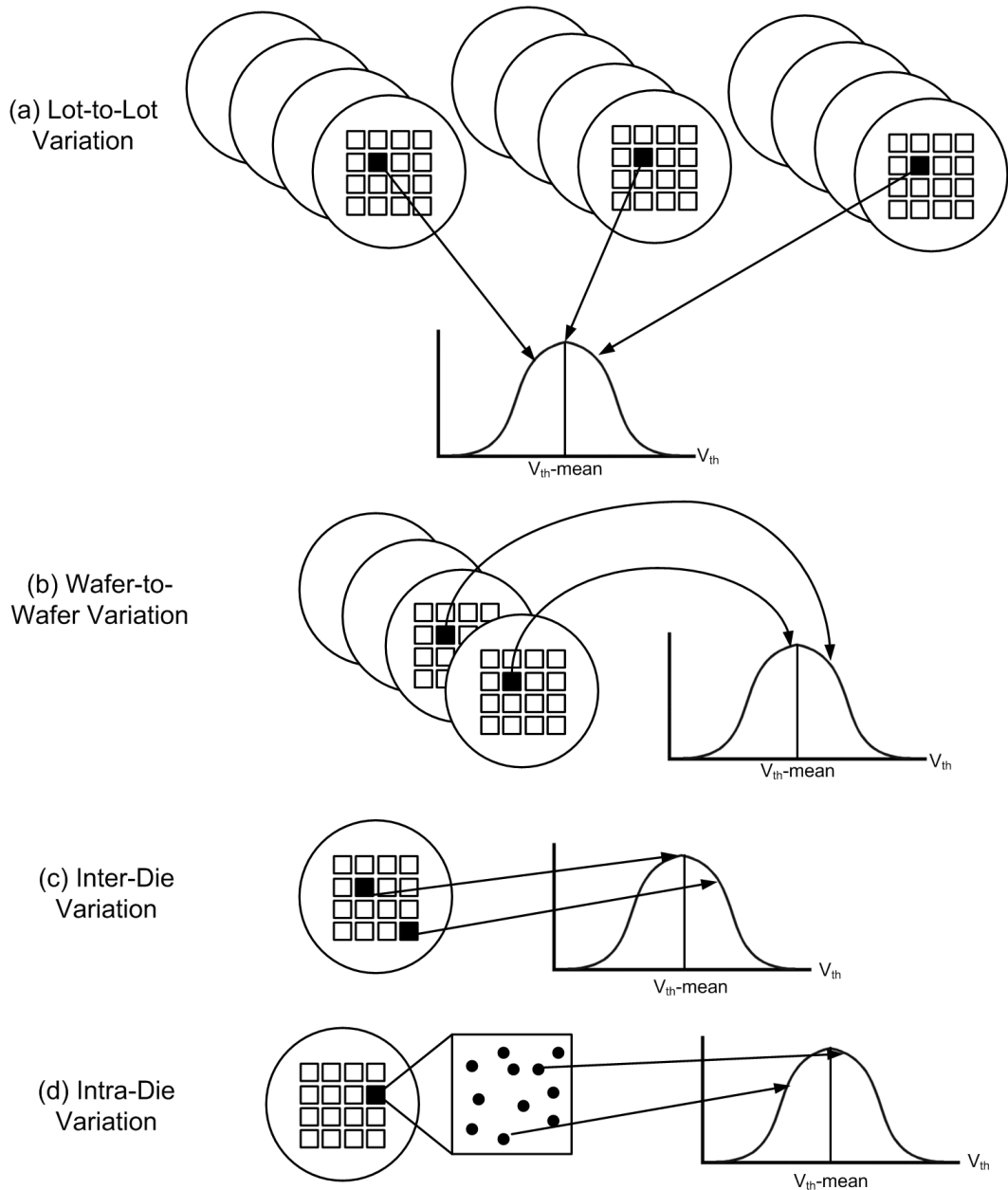


Figure 3.1: Types of process variation. Due to the variation, threshold voltage (or any other property) of any two (or three) transistors selected from different (or same) dies will be different.

consumption, and the yield of any integrated circuits are impacted by four types of variation (Figure 3.1). If three dies are randomly selected from three different lots and the threshold voltage of any transistor from each die is measured, the values will be

found to be different (Figure 3.1(a)) and will be termed lot-to-lot variation. Similarly, if two dies are randomly selected from two wafers and the threshold voltage of any transistor from each die is measured, the values will be found to be different (Figure 3.1(b)) and will be termed wafer-to-wafer variation. Similarly, if two dies are randomly selected from a wafer and the threshold voltage of any transistor from each die is measured, the values will be found to be different (Figure 3.1(c)) and will be termed inter-die variation. If two transistors are selected randomly within a die and their threshold voltage is measured it will be found out to be different (Figure 3.1(d)) and will be termed intra-die variation.

Lot-to-lot and wafer-to-wafer variation is due to the use of different fabrication facility to produce the same chip. Different fabrication facility may use different version of equipment. These variations can also be due to the use of same fabrication facility over a long span of time. Any piece of equipment in a fabrication facility may slowly shift out of calibration over time. These two types of variations can be addressed in the process level.

Inter-die variation is the variation due to the different location of each die within the same wafer. Inter-die variation can be modeled as a shift in the mean of any parameter value (e.g., threshold voltage or channel length or width) in the transistors fabricated on any silicon chip. Typically, this type of variations is the simplest to analyze [18].

Among these four types of variations, intra-die variation is the most dominant factor that affects the performance of memory circuit. It is the deviation occurring

spatially within one die (e.g., variations between transistors located side by side). Examples of such intra-die variations are threshold voltage ( $V_{th}$ ) mismatch due to random dopant fluctuations and channel length and width variations due to line edge roughness (LER). They are unavoidable and cannot be predicted. Their effects are discussed in detail in the next section.

### **3.1.1 Impact of Intra-die Process Variation on Memory Cells**

Current nanoscaled semiconductor technologies push the physical limits of scaling, making precise control of process parameters exceedingly difficult. Particularly the intra-die variations significantly increase in these technologies. Intra-die variations cannot be taken care of in the process level. These types of variations can affect two adjacent transistors in the opposite direction. For example,  $V_{th}$  variations can make the NMOS of an inverter weaker (by making the  $V_{th}$  higher) and the PMOS stronger (by making the  $V_{th}$  lower). That will strongly affect the switching threshold voltage ( $V_{TRIP}$ ) of the inverter. Since an SRAM cell is basically built from cross-coupled inverters, such variation can strongly affect the stability of the SRAM. In order to address this type of variation, design level measure has to be taken. For example, sufficient margin during design level has to be maintained.

Any asymmetry in the SRAM cell structure, due to cell transistor's mismatch, will make the affected cell less stable. If the mismatch is too intense, such cells may unintentionally flip during a read operation or even in retention, corrupting the stored data. Since, modern microprocessors are utilizing more and more embedded memory,



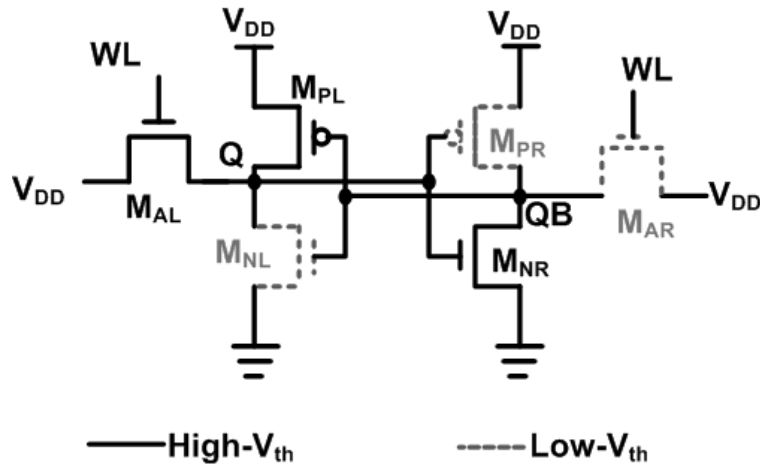


Figure 3.2: An example of process-induced threshold voltage variation affecting read stability

which is primarily implemented by SRAM cells, the probability of data corruption due to mismatch is also increasing [16].

### 3.1.2 Impact of Process Variation on Read Stability

The transistors in 6T cell may have different deviations in  $V_{th}$ . As a result, some transistors will have their  $V_{th}$  higher than the mean while some will have  $V_{th}$  lower than the mean. In order to better understand the effect of  $V_{th}$  variation on the 6T SRAM cell, Figure 3.2 shows the schematic of a 6T SRAM cell subjected to worst case intra-die  $V_{th}$  variations which can potentially compromise the cell stability during a read operation. Let us assume, the inverter  $M_{PL}$ - $M_{NL}$  has a high- $V_{th}$  PMOS and a low- $V_{th}$  NMOS, implying a reduced switching threshold. On the other hand, the inverter  $M_{PR}$ - $M_{NR}$  has a low- $V_{th}$  PMOS and a high- $V_{th}$  NMOS, causing an increased switching threshold. Also  $M_{AR}$  is a low- $V_{th}$  NMOS and  $M_{AL}$  is a high- $V_{th}$  NMOS. Assuming  $Q=1$  (and  $QB=0$ ), at the onset of the read operation, there is a slight increase in voltage level at QB due to the voltage division on the read discharge path.

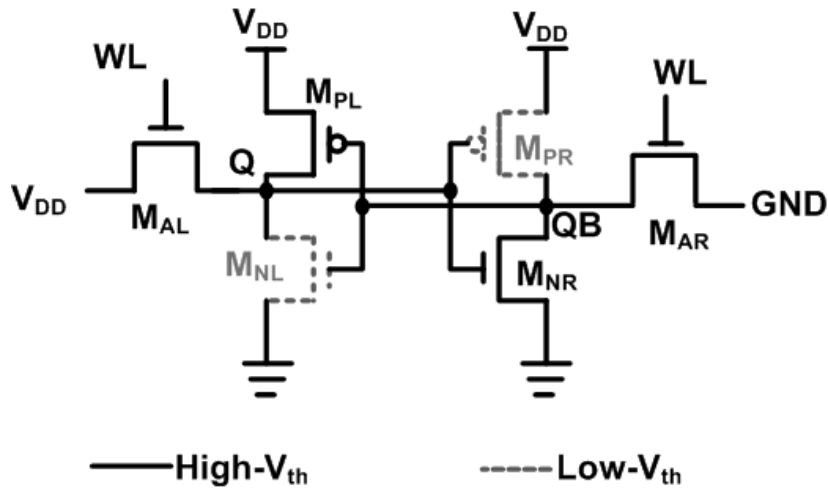


Figure 3.3: An example of process-induced threshold voltage variation affecting the writability to the cell.

The increase in QB voltage can toggle the state of the inverter  $M_{PL}$ - $M_{NL}$ , due to its reduced switching threshold. Consequently, the stored data value can be lost. This is one of the major challenges in SRAM design and yield under the unavoidable process variations at nanoscale CMOS technologies.

### 3.1.3 Impact of Process Variation on Write Margin

Similarly process variation has detrimental effects on the write margin of the 6T SRAM cell. Figure 3.3 shows a 6T SRAM cell subjected to  $V_{th}$  variations. The inverter  $M_{PL}$ - $M_{NL}$  has a high- $V_{th}$  PMOS and a low- $V_{th}$  NMOS, resulting in a low switching threshold of the inverter. On the other hand, the inverter  $M_{PR}$ - $M_{NR}$  has a low- $V_{th}$  PMOS and a high- $V_{th}$  NMOS with high- $V_{th}$  access transistors. Assuming  $Q='0'$  and  $QB='1'$ , if we want to write '0' to QB, BLB needs to be discharged to '0' during the write cycle. Once BLB is at '0', there will be a voltage division between  $M_{PR}$  and  $M_{AR}$ . Since  $M_{PR}$  is stronger than  $M_{AR}$ , the voltage level of QB cannot fall

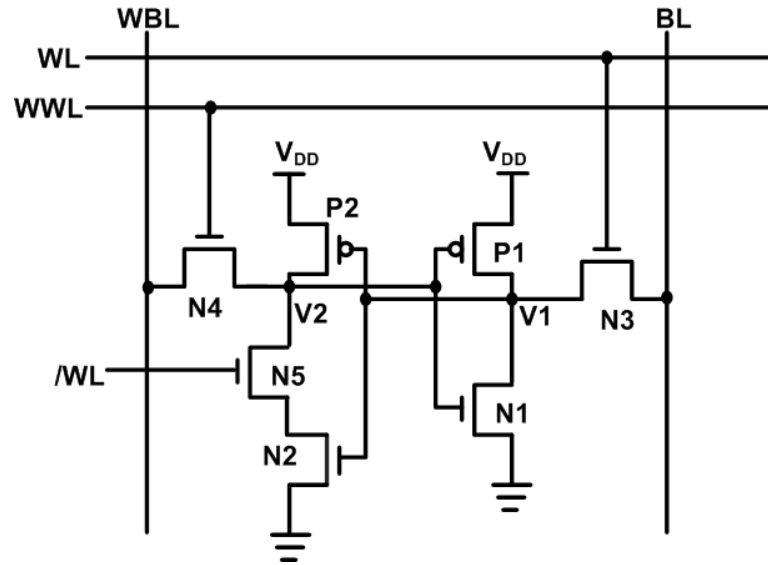


Figure 3.4: 7T cell proposed in [11].

below the ‘low’ switching threshold of the inverter  $M_{PL}-M_{NL}$ . Thus, QB cannot be flipped during the write cycle and the cell cannot be written.

## 3.2 Existing SRAM Designs for Limiting the Impact of Process Variations

There has been considerable effort over the past years to devise SRAM cells that provide high read stability and write ability in the presence of process variations. Three of such cells are discussed in the following sections.

### 3.2.1 7T SRAM Cell

A 7T SRAM (Figure 3.4) cell has been proposed by K. Takeda *et. al.* in [11]. In this cell, the transistor N5 for loop-cutting is added to the 6T cell. During data

retention mode,  $\overline{WL}$  is kept HIGH. Thus, the cell behaves as the conventional 6T cell. During write operation both WL and WWL are asserted HIGH,  $\overline{WL}$  is asserted LOW and WBL/BL are precharged or discharge according to the data intended to be written. The write operation is similar to the 6T cell except for the loop-cutting transistor N5. Since, N5 is turned off during write operation, the positive feedback is momentarily disabled and as a result, it is easier to write data into the cell. During read operation, WL is asserted HIGH and  $\overline{WL}$  is asserted LOW while WWL remains LOW. Based on the data stored in the cell, BL either discharges or not which is subsequently latched by appropriate sense-amplifier. During read operations, the threshold voltage of the inverter driving node V2 increases because the loop-cutting transistor is turned off. Thus, even if V1='0' and the voltage level of V1 is momentarily increased, the possibility of data flipping is greatly reduced. Thus, the 7T cell provides improved read stability. However, compared to the 6T cell, the 7T cell incurs approximately 13% higher area overhead. The cell has three word lines which can pose some area constraint when the array is constructed. Also, driving three word lines in a write operation will entail increased dynamic power.

### **3.2.2 8T SRAM Cell**

L. Chang, *et. al.* proposed an 8T SRAM bit cell, which is shown in Figure 3.5 [12]. The cell eliminates the disturbance to the logic '0' node inside the cell by separating the read bit line (RBL) from the write bit lines (WBL, WBLB). Prior to the read operation the read bit line RBL is precharged to  $V_{DD}$ . The read operation is started by asserting the RWL. RBL either remains at  $V_{DD}$  (if internal node 'QB' contains a '0') or is discharged (if internal node 'QB' contains a '1'). In both cases, the internal nodes

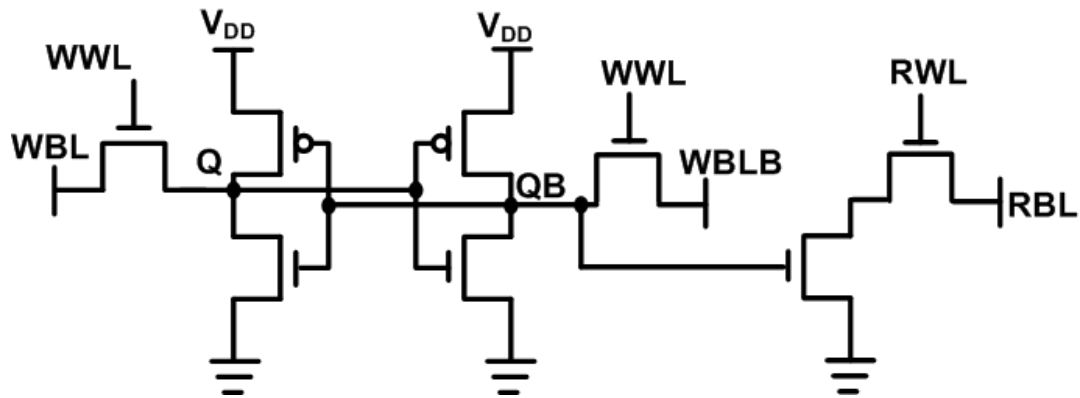


Figure 3.5: 8T SRAM cell proposed in [12].

remain undisturbed. Prior to the write operation, the bit lines are precharged/discharged to the pre-determined values. The write operation is initiated by asserting the write word line (WWL) and the nodes attain the corresponding values from the bit lines. The write operation in this 8T SRAM cell is similar to the 6T SRAM cell. The 8T cell offers improved read stability but incurs an area penalty of 30% over the traditional 6T SRAM cell and it cannot support multiple words in a row.

### 3.2.3 9T SRAM Cell

Similar to the 8T SRAM cell a 9T SRAM cell with enhanced data stability was proposed in [13]. The schematic of the 9T SRAM cell is shown in Figure 3.6. The upper part of the new memory cell is essentially a 6T SRAM cell with minimum sized transistors. The two write access transistors are controlled by a write signal (WR). The data is stored in the back-to-back inverter pair. The lower sub-circuit of the new cell is composed of the bit-line access transistors ( $R_{AX1}$  and  $R_{AX2}$ ) and the read access transistor ( $R_{AX}$ ). The operations of  $R_{AX1}$  and  $R_{AX2}$  are controlled by the value of data stored in the cell.  $R_{AX}$  is controlled by a separate read signal (RD). The write operation

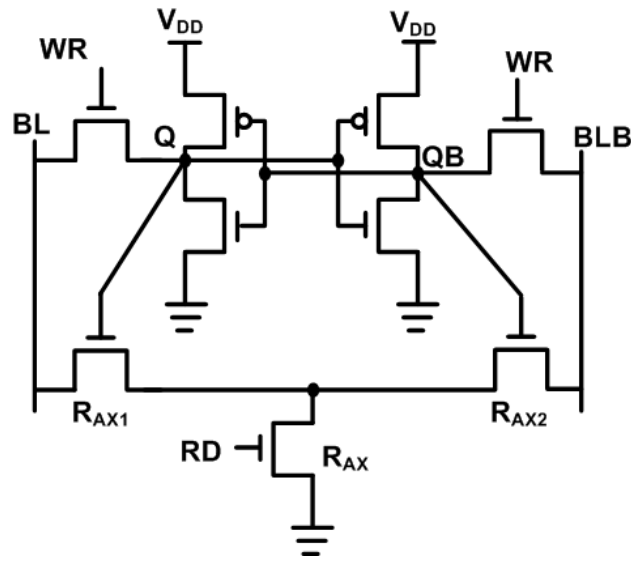


Figure 3.6: 9T SRAM cell proposed in [13].

is exactly as it is in the 6T SRAM cell. During write operation WR signal is HIGH (while RD is LOW) and BL/BLB are precharged/discharged according to the data intended to be written. During read operation, WR is low and RD is high. If Q='1' (and QB='0'), BL discharges and BLB does not. On the other hand, if Q='0' (and QB='1') then BLB discharges and BL does not. Unlike the 6T SRAM cell and like the 8T SRAM cell, the voltage of the node which stores '0' is maintained at the zero voltage level during a read operation in the proposed SRAM cell. So there is no read disturbance in this cell. Also this design provides differential sensing during read operation. But the cell incurs 37% area penalty compared to the traditional 6T SRAM cell and like the 8T SRAM cell cannot support multiple words in a row.

### 3.2.4 Performance Comparison of the Existing SRAM Design

Since more and more amount of memory is being used in various SOC and microprocessors, leakage power consumption and silicon area/cell are two key

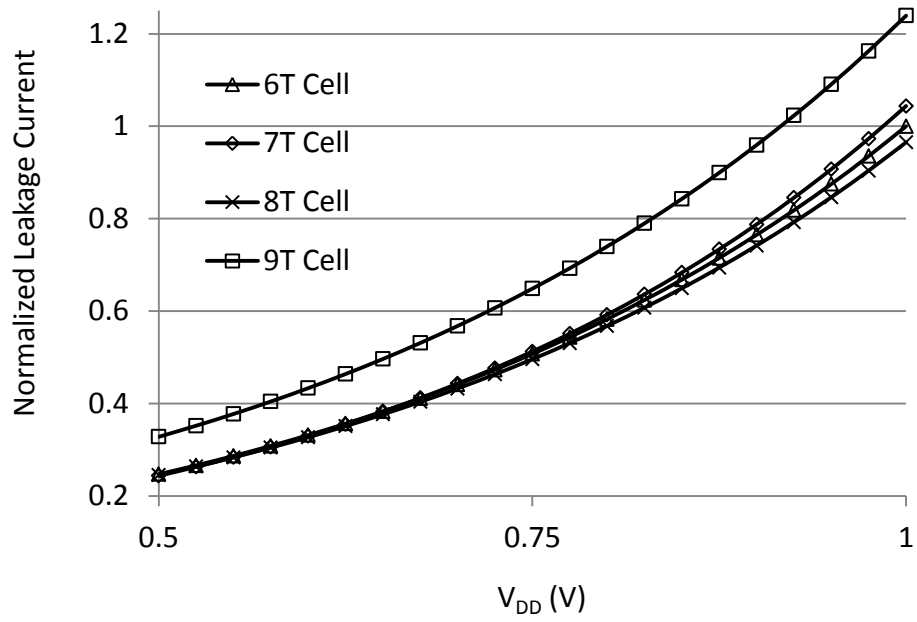


Figure 3.7: Comparison of leakage consumption of various SRAM designs.

performance metrics of any SRAM cell design. A comparison of leakage and silicon area of the above SRAM designs with the conventional 6T SRAM design is shown in Figure 3.7 and Figure 3.8 respectively.

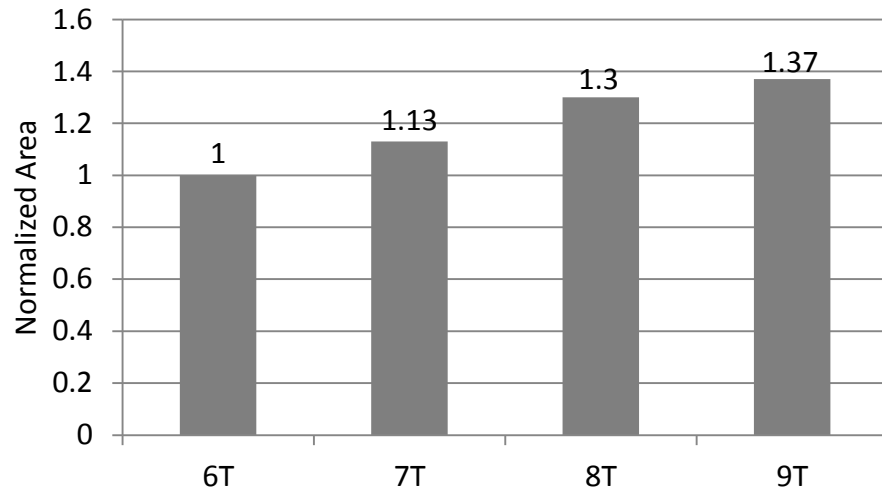


Figure 3.8: Comparison of area of various SRAM designs.



# Chapter 4

## Proposed 7T SRAM Cell and Sense-Amplifier

### 4.1 Cell Design

In order to achieve a high read data stability and writability while minimizing the area overhead, we propose a seven transistor (7T) SRAM bit-cell. The cell is shown in Figure 4.1. The proposed cell utilizes a single access transistor similar to the portless five transistor SRAM cell proposed in [19]. However, using transistors  $R_{AX1}$  and  $R_{AX2}$ , the read bit line has been decoupled from the write bit lines. Transistor  $R_{AX1}$  is controlled by a read word line (RWL). QB is connected to the gate of  $R_{AX2}$ . Thus, during read operation the node QB does not suffer any perturbation, unlike 6T SRAM cell.  $W_{AX}$  is controlled by a write word line (WWL) during write operations. A single transistor similar to  $W_{AX}$  was used in [19] for both read and write operations. As a result, the sizing of that transistor in [19] was very critical. It had to be strong enough to ensure a

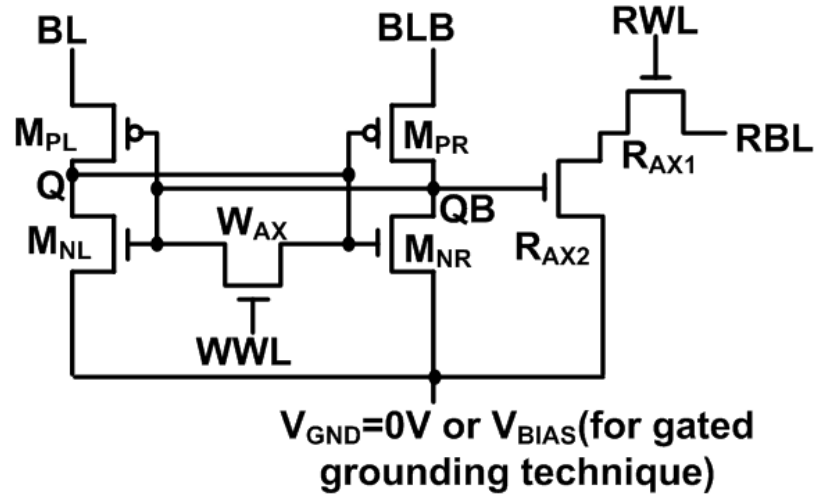


Figure 4.1: The proposed 7T SRAM cell.

successful write in all corners while it had to be weak enough for data retention during the read operation. And due to  $W_{AX}$  being weak, the write operation would have required the bit lines to be discharged by a significant amount. This would have resulted in significant amount of power consumption due to the subsequent pre-charge of the bit lines. In our proposed 7T cell, the write access transistor ( $W_{AX}$ ) is only used for write operation and hence can be optimized as required for write operation. In fact, by making  $W_{AX}$  strong, we have limited the bit line discharge during the write operation, thus making the write power consumption two times less than the write power consumed by the 6T cell. Also, as will be explained later in detail, the bit lines in the 5T cell of [19] has a dependency on the stored data. This variable bit line capacitance would pose severe constraint on reliable sensing during read operation in all process and mismatch corner.

On the other hand, the read operation, being decoupled in the proposed 7T SRAM cell, removes the read stability problem of 6T SRAM cell as well as the variable bit line capacitance problem inherent in the 5T SRAM cell. The worst-case static noise

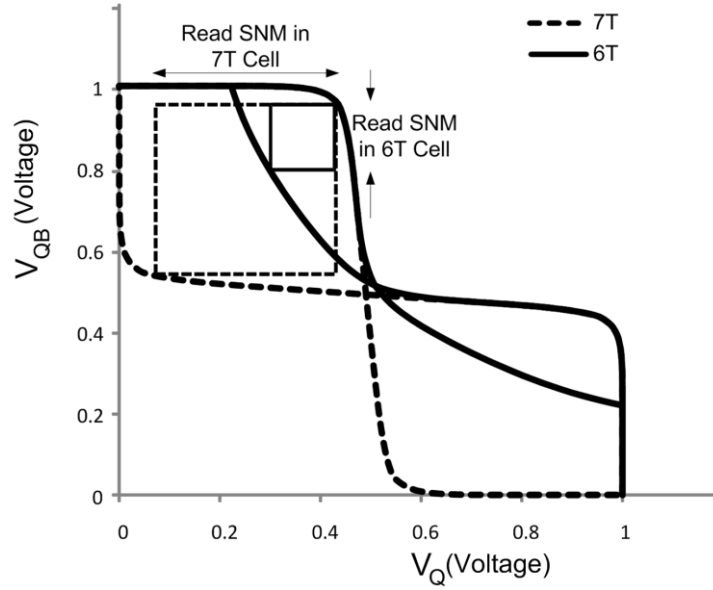


Figure 4.2: Worst-case static noise margin for 7T-SRAM and 6T-SRAM.

margin (SNM), as defined in [14], for the proposed cell is simply that for two cross-coupled inverters (Figure 4.2) as the logic ‘0’ node does not suffer any perturbation during read operation. This improved cell stability does not compromise the writability. As a result, the cell can be designed for higher speed and lower power operation while maintaining high yield. In addition, as the cell does not use multiple  $V_{th}$ , which is often employed to improve cell stability or reduce cell leakage [20], the cell is suitable to realize in the standard CMOS process without any additional process steps like implant masks, gate oxides, etc.

Since the 7T cell reduces the write power by using a method of writing where the cell is intentionally made weak during writing time window, the 7T cell by itself cannot support multiple words in a row because that would expose some cells to “half-selected state” in which due to the cell’s extreme vulnerability the data may be

destroyed. As a result, modifications are required in the array organization. Such array-level changes are necessary to achieve the full stability benefit of the 7T SRAM implementation.

## **4.2 Principle of Operation of the Proposed 7T Cell**

### **4.2.1 Cell Operation**

The write operation is done by asserting WWL (Figure 4.1) signal and discharging BL (for '0' write) or BLB (for '1' write). Assuming,  $Q='0'$  and we want to make  $Q='1'$ , we will assert the WWL. This will pull up the voltage level of Q from 0V and pull down the voltage level of QB from  $V_{DD}$ . But the pulled down level of QB will still be above the pulled up level of Q. Then BLB will begin to be discharged and as a result pulled down level of QB will decrease even more. When the level of QB falls below the pulled up level of Q, WWL will be turned off. Subsequently Q will latch to  $V_{DD}$  while QB latches to 0V and a successful write operation will be accomplished. The stronger the write-access transistor is the weaker the cell becomes when WWL is asserted and easier it is to write data in the cell. 'Easier' means less discharge (of BL or BLB) will be required for successful write operation. This fact is utilized in our cell to make it low-power relative to other cells.

During read operation RWL is asserted. If  $QB='1'$  ( $Q=0$ ), the RBL discharges indicating '0' read. If  $Q='1'$ , RBL does not discharge, indicating '1' read. The read discharge path is similar to the read discharge path of a 6T cell since both constitute of two minimum sized NMOS. Thus, the 7T cell has similar performance in terms of

discharge speed. Unlike 6T cell, the read mechanism is single-ended and thus incurs some noise sensitivity. That can be solved by using a slightly larger NMOS for  $R_{AX1}$  and  $R_{AX2}$  (Figure 4.1), ensuring larger discharge than is usually done for differential sensing.

### 4.2.2 Array Operation

The array implementation of the proposed 7T SRAM cell requires a second set of WL drivers. But this does not add to the area since these word lines run horizontally. And to accommodate these two word lines the height of the cell did not need to be increased.

The cell by itself cannot support multiple words in one row. Because the write access transistor  $W_{AX}$  is purposely made stronger to facilitate write operation. As a result, if multiple words are implemented in a row and one word in a row is to be written, the bit-cells belonging to the other words in the same row will be in a half-selected state (half-selected state is when WWL of a cell is asserted during a write operation and BL/BLB are held at  $V_{DD}$ ). And when WWL of a cell is asserted, due to the cell's extreme vulnerability, the data is prone to flipping even if both BL and BLB are held at  $V_{DD}$ . Thus, conventional array implementation with the proposed 7T SRAM cell cannot support multiple words per row. However, it will be shown in Chapter 6 that by utilizing Column Virtual Grounding techniques, the proposed 7T SRAM cell can support multiple words per row. Implementation of multiple words per row enables protection from multi-bit soft error events. Since the bits of different word

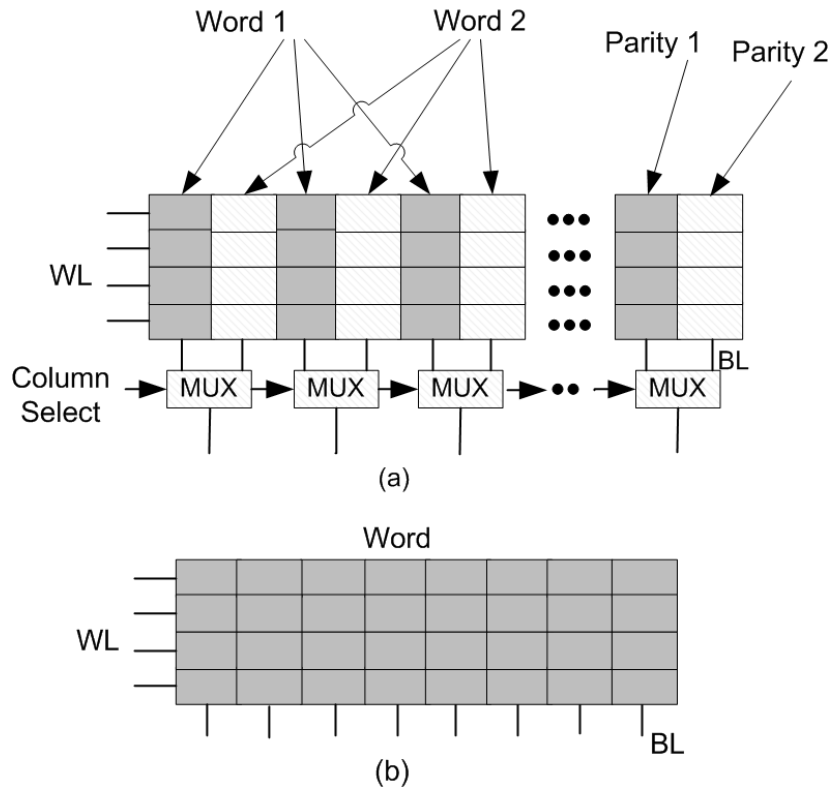


Figure 4.3: (a) Floor plan, where multiple words per row is implemented. (b) Floor plan, where one word per row is implemented. Sophisticated ECC codes are required for multiple bit corruption..

in one row are physically interleaved (Figure 4.3), multi-bit errors resulting from a soft-error even can at most affect only one bit from one word because such multi-bit errors tend to be spatially adjacent. Such one bit error per word can be easily detected or corrected with simple parity checking or error correcting codes (ECC). A single error correcting double error detecting (SECDED) error correction code incurs an overhead of 8 bits per 64 bits of data (i.e., 13%). On the other hand, radiation-hardened cells can have an area overhead of 30-100% [21].

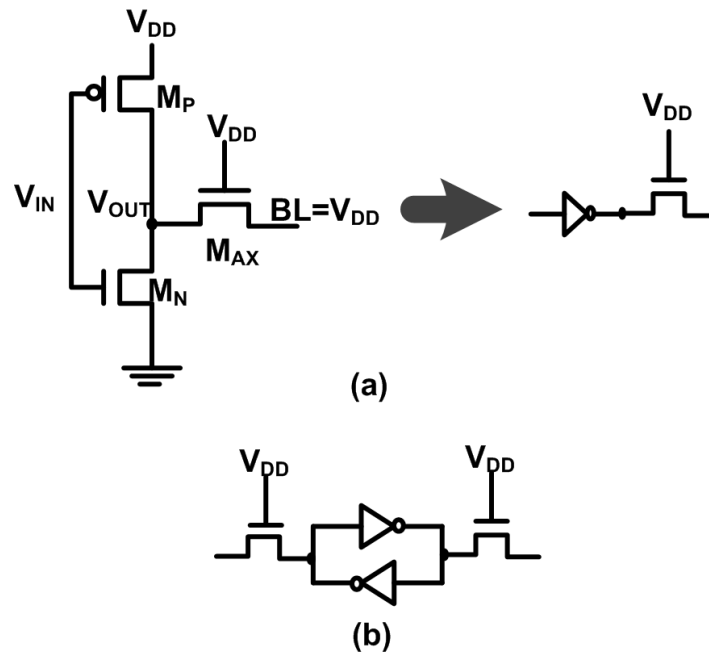


Figure 4.4: (a) Inverter with an access transistor. (b) 6T SRAM cell.

### 4.3 Theoretical Analysis of the Proposed Cell

$M_{PL}$ - $M_{NL}$  and  $M_{PR}$ - $M_{NR}$  constitute the cross-coupled inverters to store data (Figure 4.1).  $W_{AX}$  is used for write operation when WWL is HIGH.  $R_{AX1}$  and  $R_{AX2}$  are the transistors used to decouple the read operation. Unlike 6T SRAM, during read operation the cell will not suffer any stability problem. In Figure 4.4(a) we have an inverter with an access transistor. By cross-coupling such an inverter, the 6T SRAM is constructed, shown in Figure 4.4(b). Figure 4.5 shows the Forward Voltage Transfer Characteristics (VTC) and the Inverse VTC of both inverters with access transistor turned ON. In fact, Figure 4.5 is the butterfly curve of the 6T SRAM, during read operation as well as write operation (when the access transistors are turned ON).

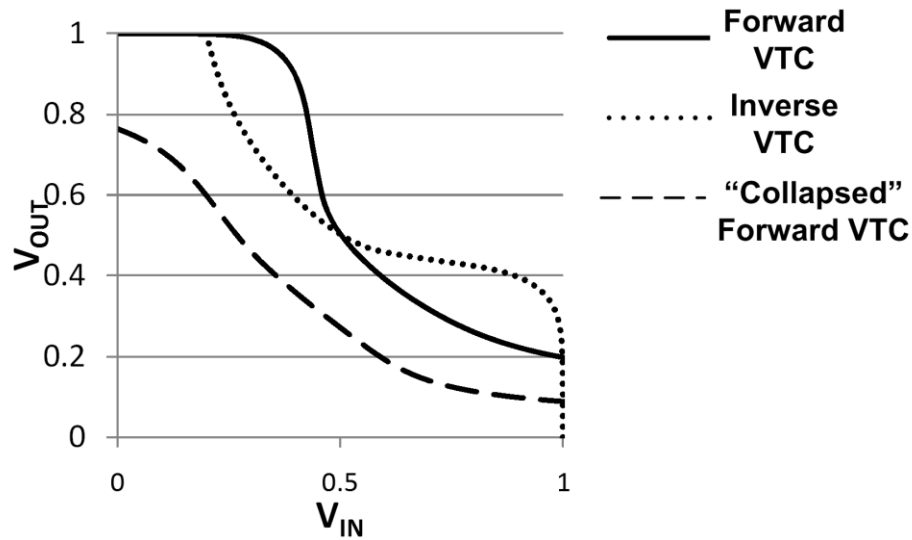


Figure 4.5: The Forward-VTC and the Inverse-VTC form the “butterfly” curve of two cross-coupled inverters.

During write operation, one of the bit line (shown by BL in Figure 4.4(a)) is discharged. As a result, that VTC will “collapse” (the dashed line in Figure 4.5) and there will be only one intersecting point between Forward VTC and Inverse VTC. Subsequently, the SRAM settles into that point, ensuring a successful write operation.

Similarly, as shown in Figure 4.6(a),  $M_P$ - $M_N$  is a basic inverter and  $M_{AX}$  is used to connect the input and output point. If  $M_{AX}$  is kept OFF, the circuit will function like a normal inverter. If  $M_{AX}$  is kept ON (as shown in Figure 4.6(a)) its behavior will be different. For ease of description in this work, the circuit is termed “modified” inverter. When  $V_{IN} = 0V$ ,  $V_{OUT} = V_{DD}$  in a normal inverter. But in the “modified” inverter,  $M_{AX}$ , being ON, pulls down  $V_{OUT}$  midway between  $V_{DD}$  and  $0V$ . Similarly, when  $V_{IN} = V_{DD}$ ,  $V_{OUT} = 0V$  in a normal inverter. But in “modified” inverter,  $M_{AX}$  pulls up  $V_{OUT}$  to a non-zero voltage level. The VTC of the “modified” inverter is given by the solid line in Figure 4.7.



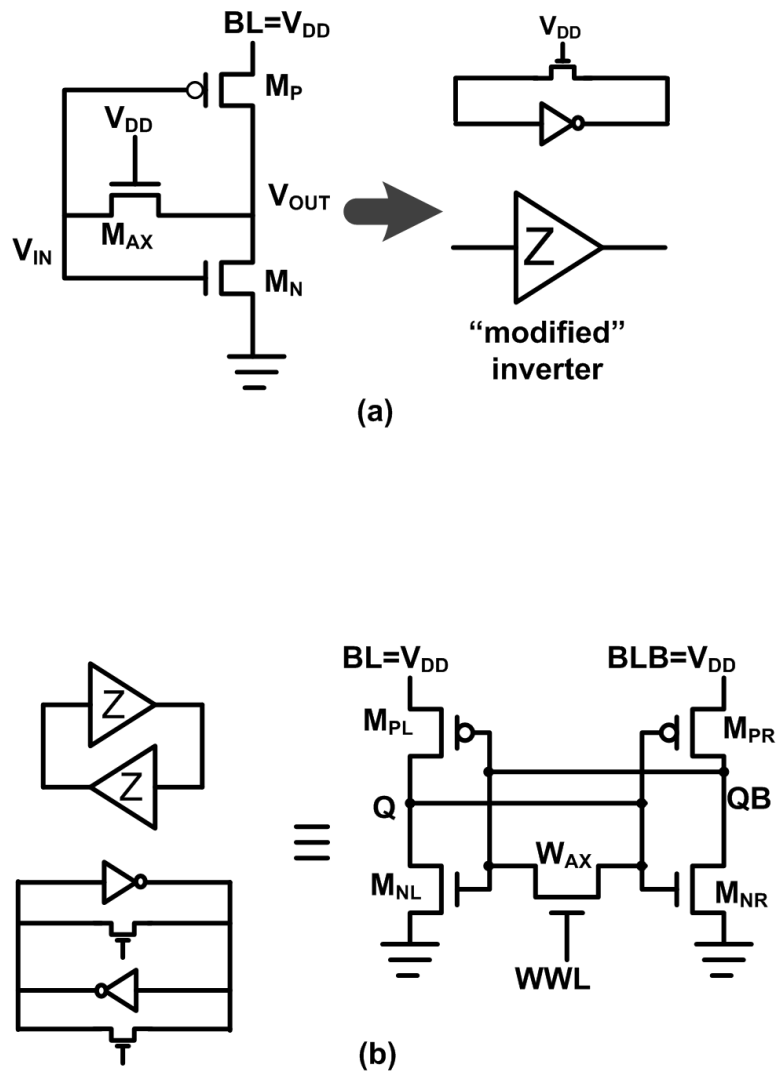


Figure 4.6: (a) A schematic of the “modified” inverter. (b) Two cross-coupled “modified” inverters constituting a memory cell named Portless SRAM Cell.

In Figure 4.6(b), two “modified” inverters are connected in cross-coupled configuration. The  $M_{AX}$  of the two “modified” inverters will be in parallel and is replaced by the equivalent transistor named  $W_{AX}$ . This is the cell proposed in [19]. In Figure 4.7 the Forward VTC (solid line) and the Inverse VTC (dotted line) constitute the butterfly curve of two back-to-back “modified” inverters. There are three intersecting points between Forward VTC and Inverse VTC. As in the 6T cell, to write

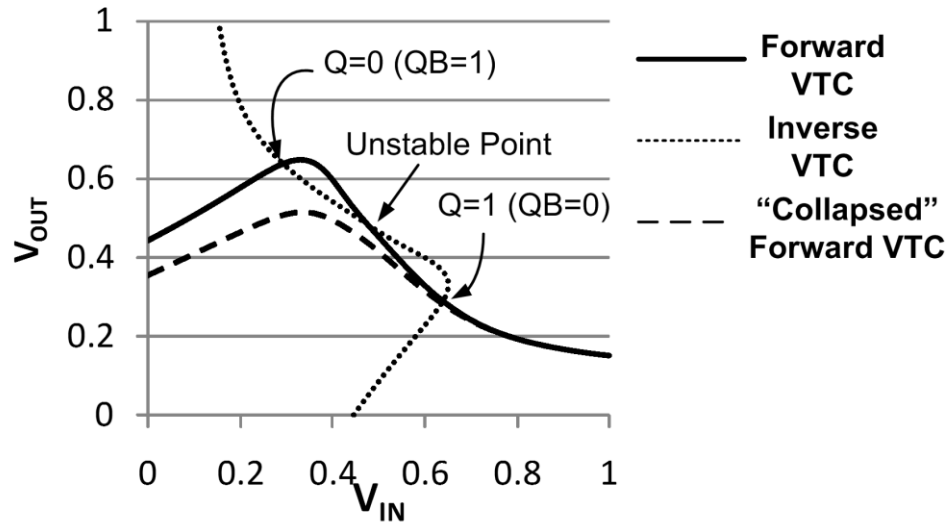


Figure 4.7: The Butterfly curve of the cross-coupled “modified” inverter.

data in a cell we have to collapse one of the VTCs so that there is only one intersecting point between the two curves and the cell will settle into that point. And the “collapse” of the VTC is accomplished by decreasing the voltage level of BL (or BLB) from  $V_{DD}$ .

#### 4.4 The Proposed Single Ended Sense-Amplifier

The read operation of the proposed 7T SRAM cell is single-ended. Thus, the sense amplifier for this bit-cell has to be single ended. Conventional 6T SRAM cell gives differential output. Thus, most of the available sense amplifier topology is differential. A single-ended sense amplifier is proposed in this section, which can be used with the proposed 7T SRAM cell.

An inherent problem of the sense amplifier is the “memory” from the previous evaluation. Let us assume, in the previous evaluation period the sense amplifier made

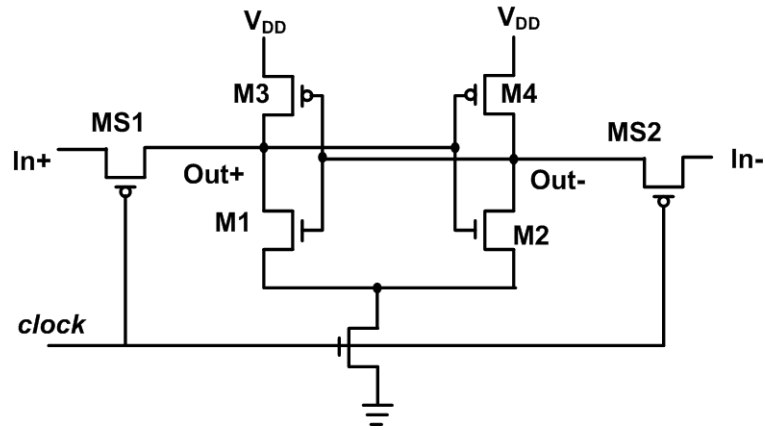


Figure 4.8: A basic clocked sense amplifier.

an evaluation of  $OUT+= '1'$  ( $OUT-= '0'$ ) as shown in Figure 4.8 and in the next evaluation period the sense amplifier should make an evaluation of  $OUT+= '0'$  ( $OUT-= '1'$ ). That means the latching mechanism inside the sense amplifier has to be flipped. But due to mismatch between the transistors, the latching mechanism can be biased towards  $OUT+= '1'$  or the generated voltage differential between the bit lines can be too small for a “successful” evaluation. To remove the sense amplifier’s memory, all nodes in the sense amplifier are driven to a known voltage. None of the nodes are kept floating or dynamically charged, because keeping a node floating can result it into being charged or discharged from the previous evaluation. In another words, the two nodes  $OUT+$  and  $OUT-$  of the sense amplifier are precharged to  $V_{DD}$  before the initiation of the evaluation period and during evaluation period one of those two nodes is driven to zero potential based on the discharging of one of the bit lines. If none of the bit line discharges then a race condition occurs and the latching mechanism of the sense amplifier can latch into any direction.

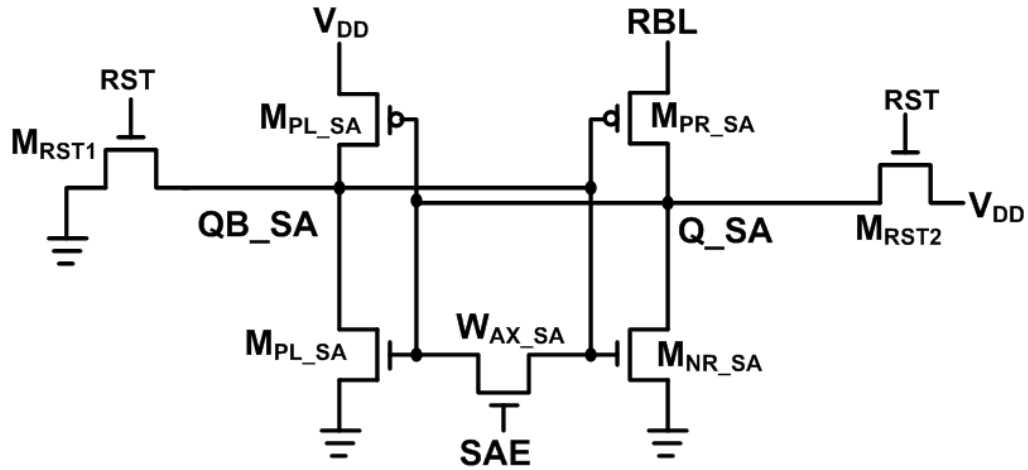


Figure 4.9: The proposed single-ended Sense-Amplifier.

This gives rise to the sensing problem ensued in single-ended sensing. Because in single ended sensing, there is only one bit line and it either discharges or it does not. If it discharges then there is no problem in the evaluation phase. But if the bit line does not discharge then a race condition arises. And a chance arises of making a wrong evaluation. Thus, differential sense amplifier cannot be used for single-ended sensing.

The proposed sense amplifier is shown in Figure 4.9. It is actually based on the proposed 7T SRAM cell. The proposed sense amplifier utilizes the “memory of a previous evaluation” to circumvent the problem of race condition. Instead of precharging both Q\_SA and QB\_SA to  $V_{DD}$ , read operation is initiated by making Q\_SA='1' (and QB\_SA='0') by a reset operation. If the read bit line discharges then the sense amplifier flips to Q\_SA='0' (and QB\_SA='1'). And if the read bit line does not discharge the sense amplifier continues storing Q\_SA='1'. Thus, there is no race condition in the sensing mechanism.

Another advantage of this sense amplifier, for the proposed 7T SRAM cell array, is its similarity to the cell itself. Thus, the sense amplifier can be laid out with same pitch as the SRAM cell column, which is very important for the overall area efficiency of the SRAM array. In 6T SRAM arrays multiple columns are shared by a single sense amplifier. Thus, the space allowed for a sense amplifier is large. But as was explained earlier, multiple words cannot be implemented in the proposed 7T SRAM cell array. Thus multiple columns cannot be shared by a single sense amplifier. The sense amplifier must have equal or smaller width than the column. Since the latching component of the sense amplifier is similar to the cell, that pitch equality can be maintained even under different design rules.

#### **4.4.1 The Principle of Operation of the Proposed Single Ended Sense-Amplifier**

Before the initiation of the read operation, RST is asserted. That will ensure that  $Q\_SA='1'$  (and  $QB\_SA='0'$ ). Since  $M_{RST1}$  has its one end physically connected to GND and  $M_{RST2}$  has its one end physically connected to  $V_{DD}$ , a very short pulse is enough to make  $Q\_SA='1'$ . Then SAE (Figure 4.9) is asserted. As a result, the  $V_{Q\_SA}$  will be pulled down and  $V_{QB\_SA}$  will be pulled up to an intermediate level. If the RBL (read bit line) discharges, the pulled down level of  $V_{Q\_SA}$  will drop below the elevated level of  $V_{QB\_SA}$  and the sense amplifier will flip, indicating that the cell being read is storing  $Q='0'$ . If the RBL does not discharge, the pulled down level of  $V_{Q\_SA}$  will not drop below the elevated level of  $V_{QB\_SA}$  and the sense amplifier will not flip, indicating that the cell being read is storing  $Q='1'$ .

# Chapter 5

## Validation and Comparison of the Proposed SRAM Cell

This section describes the simulation framework used in this thesis. The proposed 7T SRAM cell will require a single-ended sense-amplifier for read operation. Also the cell has two word lines. For an array with 256 cells/column 512 word lines will be required (instead of 256 word lines). Thus, a 9-to-512 decoder was used for simulation purpose, where 8 bits were used as address bits and one bit was used to specify read or write operation.

### 5.1 Simulation Setup

The 7T SRAM cell with its transistor sizing is shown in Figure 5.1. The proposed single-ended sense-amplifier with its transistor sizing is shown in Figure 5.2. The test bench used for analyzing the 7T SRAM cell column is shown in Figure 5.3. This was

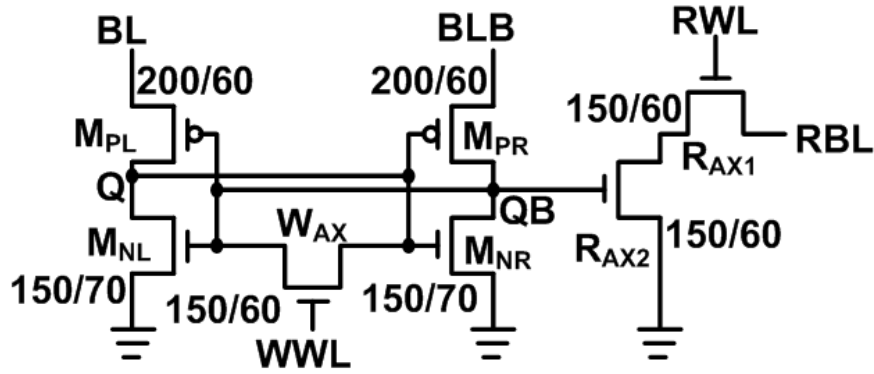


Figure 5.1: The proposed 7T SRAM cell with transistor sizing.

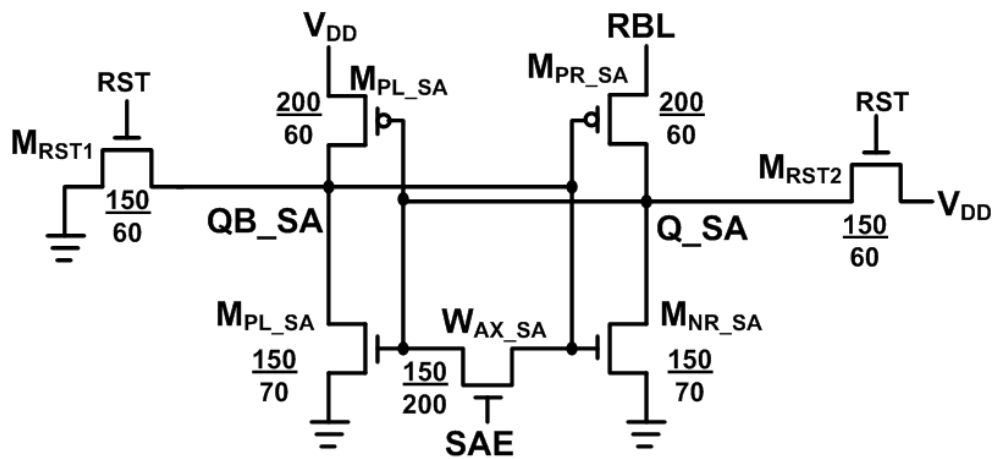


Figure 5.2: The proposed single-ended Sense-Amplifier with transistor sizing.

used to find the equivalent bit line capacitance and the required precharge energy of a column with 256 cells. Since the write bit lines and read bit line are different, their precharge mechanism is slightly different from the ones used for 6T SRAM array. The write bit line is only discharged when a write operation is performed. In all other time it remains precharged to  $V_{DD}$ . As long as  $W\_EN$  is LOW, the write bit lines remain precharged to  $V_{DD}$ . And when  $W\_EN$  is HIGH, based on Data (and  $\overline{\text{Data}}$ ) one of the write bit lines is discharged and a write operation is performed.

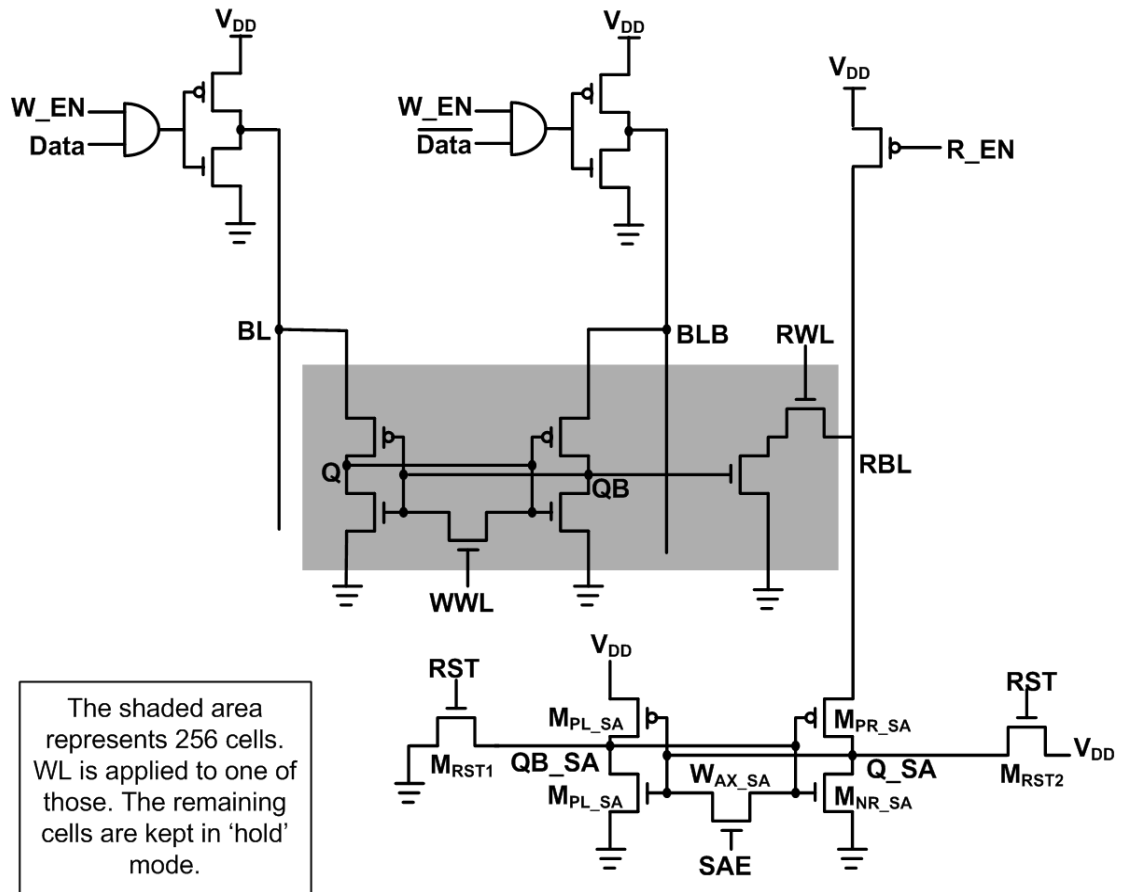


Figure 5.3: Schematic of a column of the 7T SRAM cell along with write driver and sense-amplifier circuitry used to perform read and write operations.

The read circuitry consists of a single-ended sense amplifier as shown in Figure 5.3. The bit value stored in the SRAM cell is obtained on the RBL. The read operation is initiated by making R\_EN HIGH. That will make the RBL floating. Then the RWL of the required row is asserted and based on the stored data inside the cell, RBL either discharges or not. During this period, as explained earlier RST is asserted to make Q\_SA='1' in the sense amplifier. Then SAE is asserted HIGH to make the evaluation. After allowing sufficient time for the sense amplifier to make a valid evaluation the SAE is made LOW and the stored data inside the read cell will be latched into the sense amplifier.



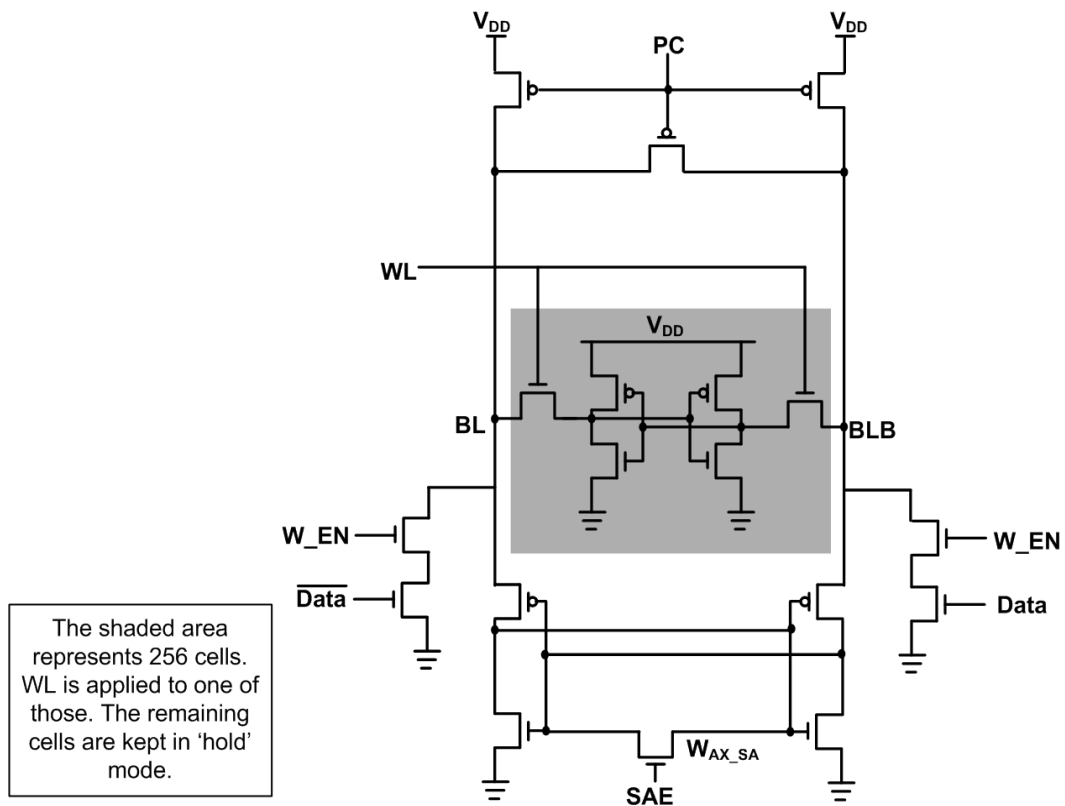


Figure 5.4: Schematic of a column of the 6T SRAM cell along with write driver and sense-amplifier circuitry used to perform read and write operations.

The layout of the 7T SRAM cell was made in 65nm TSMC process and the extracted layout was used to simulate the behavior of the cell under various process corners. 64 cells/row were used to simulate the word line capacitance along a row and the required decoder energy for write or read operation.

Similarly, for comparison purpose, the layout of the 6T SRAM cell was also made in 65nm TSMC process and the extracted layout was used to simulate the behavior of the cell during read and write operation. 256 cells/column was (see Figure 5.4) used to simulate the bit line capacitance and the relevant precharge energy after a successful write and read operation.

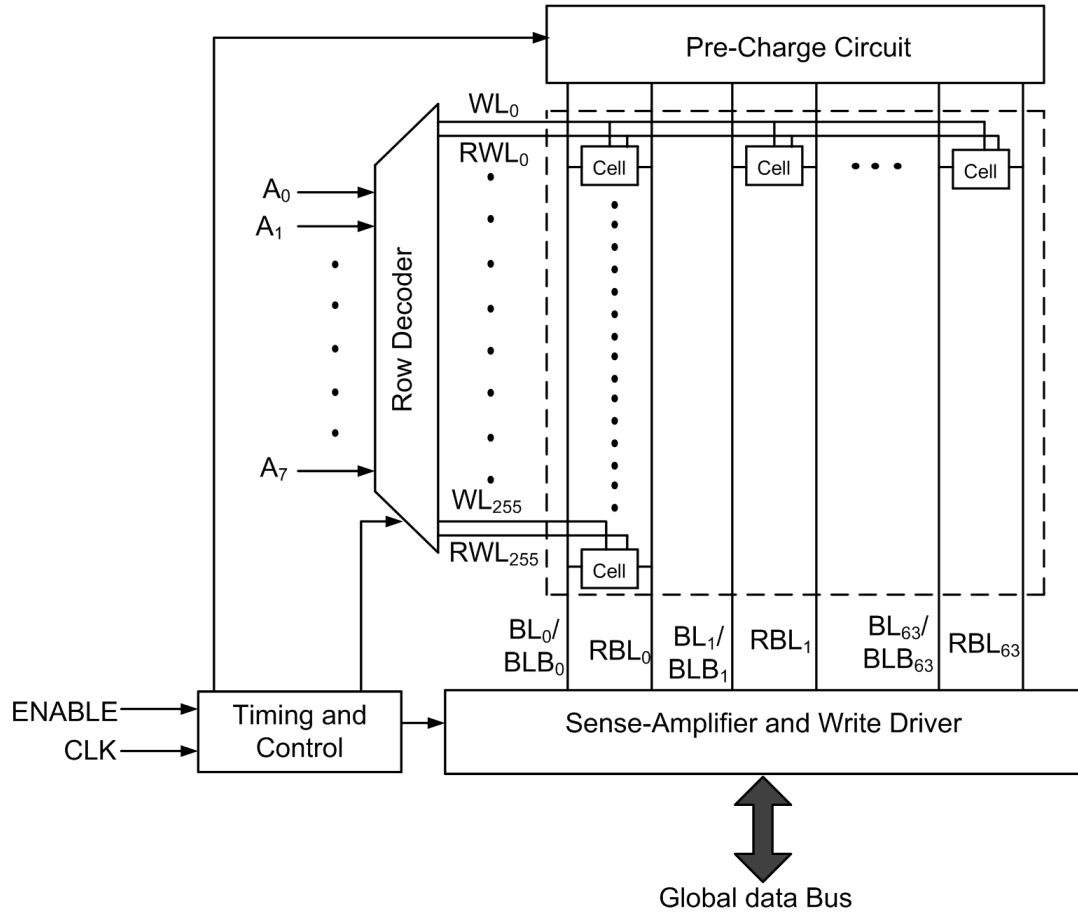


Figure 5.5: Simulating array behavior with peripherals.

To simulate the overall array behavior of the 7T SRAM cell, an array with peripheral circuitry was simulated as shown in Figure 5.5. The First column contains 256 cells. Each of the remaining 63 columns contains one cell with lumped capacitance to mimic the bit line capacitance of a 256 cell-column. From row perspective the first row contains 64 cells. Each of the remaining 255 rows contains one cell with equivalent word line (WWL and RWL) capacitance. The row decoder used was a 9-to-512 decoder. 8 bits were used as address bits and one bit was used as

Read/Write signal. The timing circuit was used to generate all the control signals like sense-amp enable, sense-amp reset, bit line precharge signal, etc.

## 5.2 Write Performance

In the proposed 7T cell when the WWL is asserted, the  $W_{AX}$  transistor turns ON and weakens the cell from inside. As a result small amount of noise (discharge at either of the bit line BL/BLB), in terms of power consumption, ensures flipping of the cell in the desired direction. For 6T cell the bit lines need to be discharged by a large amount (from  $V_{DD}$  to 0V) and as a result, subsequent precharge takes large amount of energy. In 7T cell, bit lines need small amount of discharge for write operation and as a result, subsequent precharge power is significantly smaller. A comparison of total energy consumption in a column after a write operation under different  $V_{DD}$  is given in Figure 5.6. The energy includes the bit line precharge energy and the write driver energy.

It is important to note that the different method of writing (utilized in the proposed design) introduces a dependency of bit line capacitance on cell data, an effect not seen in other SRAM architectures. This relationship results from the direct connection of the cell PMOSs to the bit lines. The PMOS connected to the HIGH data node operates in the triode region while the LOW data node PMOS is effectively off. The parasitic capacitance of the HIGH data node will be included in the HIGH side bit line. The HIGH side bit line will therefore experience a higher effective capacitance in comparison to the LOW side. In the extreme cases, where all the cells in a column store same data, the bit line connected to the high side will have larger (about 3 times

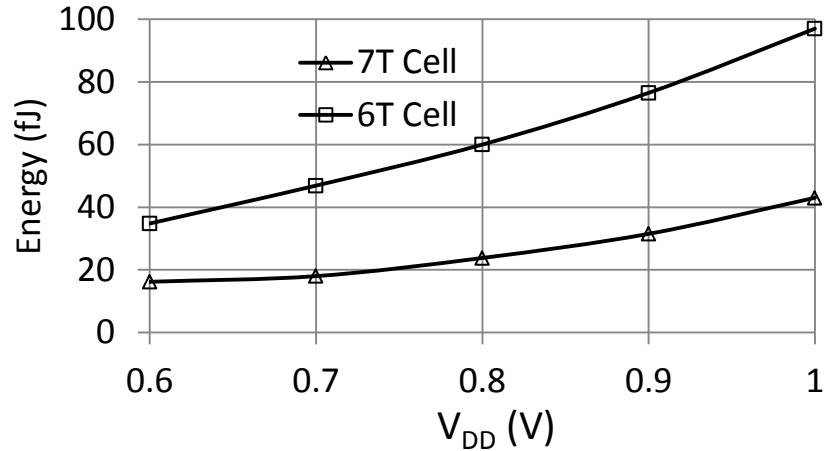


Figure 5.6: Energy consumption per column in a write operation.

Table 1: BL/BLB capacitance dependence to the stored data in the column

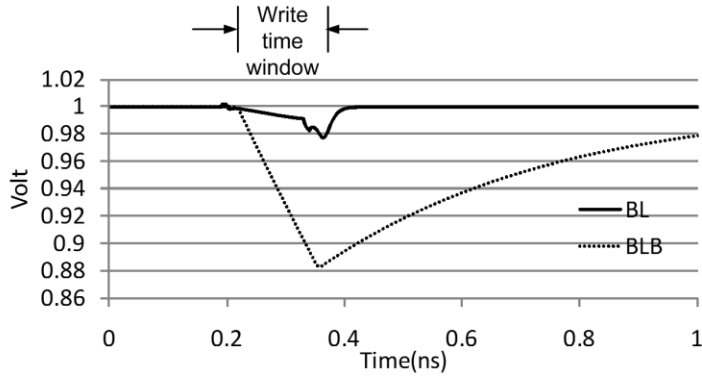
| Data stored in the column | BL Capacitance | BLB Capacitance |
|---------------------------|----------------|-----------------|
| 90% Q='1'<br>10% Q='0'    | 387fF          | 147fF           |
| 50% Q='1'<br>50% Q='0'    | 267fF          | 290fF           |
| 10% Q='1'<br>90% Q='0'    | 140fF          | 432fF           |

of the bit line connected to the LOW side) effective capacitance. As a result, write driver should be strong enough to discharge the maximum effective capacitance bit line (connected to the HIGH side) sufficiently so as to ensure successful write operation. However, if the stored data in all the cells are reversed then the maximum effective capacitance bit line will become minimum effective capacitance bit line and the “strong” write driver will discharge the bit line by a larger amount. The BL/BLB capacitance under various proportions of ‘0’ and ‘1’ is shown in Table 1.

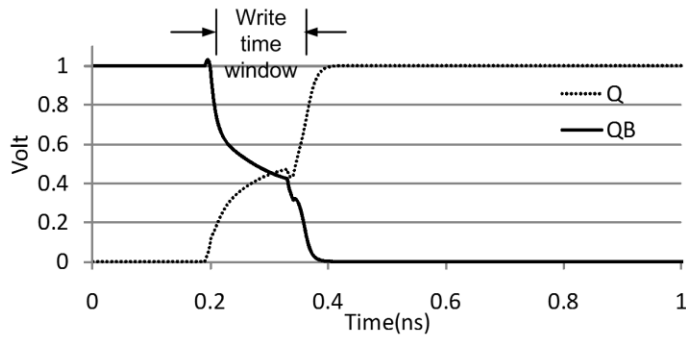
The sizing of  $W_{AX}$  was made to be  $W=150\text{nm}$  and  $L=90\text{nm}$ . A first order analysis would indicate that optimized write operation will require the  $W_{AX}$  to be as strong as

possible. Because stronger  $W_{AX}$  will bring the voltage level of Q and QB closer to each other thus making it easier to flip by discharging BL/BLB. But, due to process variation the  $V_{TRIP}$  of both inverters is not always same. Assuming  $Q='0'$  (and  $QB='1'$ ) and we want to make  $Q='1'$ , it is not enough that the pulled down voltage level of QB is made to fall just below the elevated level of Q by discharging BLB. For successful write operation in all variation corner  $V_{QB}$  should fall below  $V_Q$  by a certain amount to ensure that  $V_{QB}$  itself indeed becomes less than extreme cases of  $V_{TRIP}$ . Though stronger  $W_{AX}$  brings  $V_Q$  and  $V_{QB}$  closer, it also prevents subsequent fall of  $V_{QB}$  (or  $V_Q$ ) by the discharge of BLB (or BL). Thus, there is an optimum sizing for  $W_{AX}$  that will result in the minimum discharge in BL (or BLB) for successful write operation in all variation corners. Extensive Monte-Carlo simulation was done with different sizing of  $W_{AX}$  and it was found out that the sizing of  $W=150\text{nm}$  and  $L=90\text{nm}$  results in the minimum BL/BLB discharge of 100mV for successful write operation in all corners.

Ensuring 100mV of discharge for the case of maximum effective bit line capacitance will translate into a discharge of 290mV for the case of minimum effective bit line capacitance. And a discharge of 290mV does not have any destructive effect on the other cells in the same column. It has been seen that as long as the “discharged state” has a duration of less than 500ps (the bit line gets precharged for the next write operation within that period), discharge of up to 700mV (i.e. the bit line voltage drops to 300mV for a  $V_{DD}$  of 1V) does not have any destructive effect on the other cells. That will give a safety margin of about 410mV. Also, assuming the probability of a cell storing a ‘1’ or ‘0’ to be equal, the probability of such extreme



(a)



(b)

Figure 5.7: Transient waveform during write operation. (a) The write bit lines (BL and BLB). (b) The storage nodes of the cell.

case, where all the cells in a column store same data, is very small ( $\approx 2^{-256}$  or  $10^{-77}$ ). Thus, the write driver was designed according to the maximum effective capacitance when 90% of the cells in a column store same bit-value.

A transient waveform of the storage nodes and the write bit lines during a write operation is shown in Figure 5.7. In this waveform, previously Q was '0' (QB='1') and it is intended to make Q='1' (QB='0'). As a result, write bit line BLB was discharged during write operation. A transient waveform of the storage nodes of one of the other cells in the same column, which are not being accessed, is shown in Figure 5.8. In this waveform Q='0', QB='1' and BLB is being discharged. As a result, voltage of QB is following the discharge of BLB.

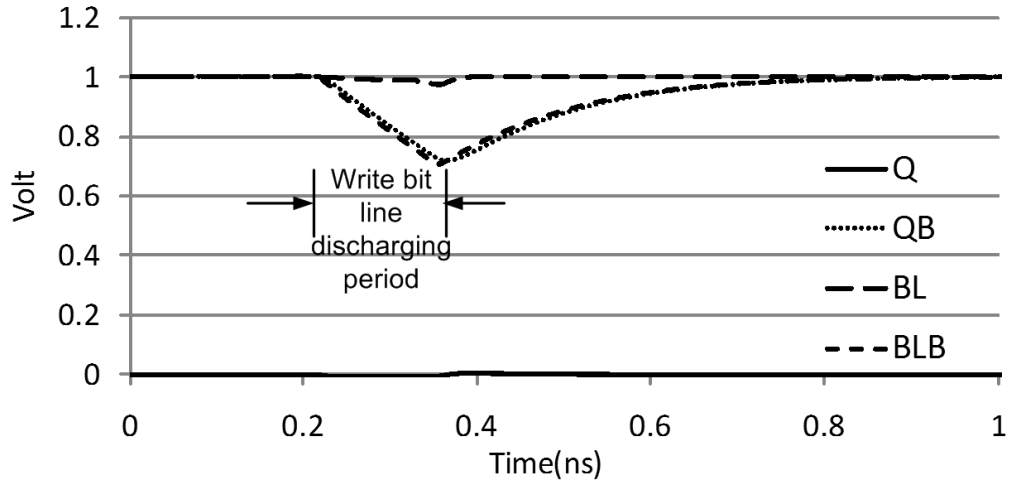


Figure 5.8: Transient waveform of a cell where the write access transistor is OFF but one of the write bit line is discharged maximally.

Table 2: Energy consumption per column in a read operation.

| Cell | Energy consumption in a column for Read operation(fJ) |          |
|------|---|----------|
|      | '0' read  | '1' read |
| 7T   | 20.73   | 12.08    |
| 6T   | 20  |          |

### 5.3 Read Performance

Read operation was performed satisfactorily with a pulse-width of 150ps at RWL for  $V_{DD}=1V$ . For a pulse width of 150ps the RBL discharges by 130mV, which is sufficient to ensure proper sensing by the sense amplifier as was verified by Monte-Carlo simulation under various mismatch corners. The energy consumed in a column during a read operation is given in Table 2. Since the cell is single-ended, the energy consumption for '0' and '1' read is not equal. The energy includes the read bit line precharge energy and the dynamic energy of the sense amplifier.

Table 3: Decoder energy consumption for asserting a word line during a read or write operation.

| Cell | Word line capacitance with 64 cell/row (fF) | Decoder energy consumption (fJ) | Decoder leakage consumption (uA) |
|------|---|---------------------------------|----------------------------------|
| 7T   | 39*   | 140                             | 15                               |
| 6T   | 38  | 125                             | 8                                |

\*The 7T SRAM cell has two word line (read and write word lines). Both have the same word line capacitance.

Table 4: Total read delay.

| 1        | 2                                  | 3                                     | 4   |
|----------|------------------------------------|---------------------------------------|---|
| $V_{DD}$ | decoder delay+WL driver delay (ps) | BL differential generation delay (ps) | Total read delay from the array shown in Figure 5.5. (In addition to the sum of column 2 &3 this column includes some margin). (ps) |
| 1V       | 190                                | 150                                   | 397   |
| .9V      | 234                                | 180                                   | 478   |
| .8V      | 302                                | 250                                   | 590   |
| .7V      | 427                                | 380                                   | 850   |
| .6V      | 701                                | 700                                   | 1460  |

64 cells/row was used to simulate the word line capacitance and the total decoder energy to drive the word line is given in Table 3. The total decoder energy includes the word line driver energy and the dynamic energy consumed in the internal nodes of the decoder. Some of the internal nodes of the decoder circuitry have large capacitance value due to long metal wire used for connection to nodes far apart. The decoder delay and the required discharge delay under different supply voltage are given in Table 4.



Table 5: Cell Leakage Current for  $V_{DD}=1V$ .

| Cell | Leakage Current (nA) |             | Average (nA) |
|------|----------------------|-------------|--------------|
|      | Storing '0'          | Storing '1' |              |
| 7T   | 6.4                  | 4.38        | 5.39         |
| 6T   | 5.63                 |             | 5.63         |

## 5.4 Leakage Power

The proposed 7T SRAM cell is asymmetric. Thus, the leakage current depends on the stored data. When the stored value is '0' ( $Q='0'$ ), one of the NMOS in the read current path is ON and one is OFF while when the stored value is '1' ( $Q='1'$ ) both NMOS in read path are OFF. Thus, leakage current for  $Q='0'$  is higher (rest of the cell remains same for both situation). The leakage current of the 7T SRAM cell is taken to be the average of the two values, Cell leakage current for  $V_{DD}=1V$  is shown in Table 5. A comparison of leakage currents of 6T cell and the proposed 7T cell as a function of  $V_{DD}$  is shown in Figure 5.9. As can be seen, the leakage is similar to the 6T cell.

## 5.5 Soft Error Tolerance

Radiation-induced single event transient (SET) has emerged as a critical reliability concern for integrated circuits in sub-100 nanometer CMOS technologies [22]. When a sensitive node of a memory circuit is affected by alpha-particle or high energy neutrons, a voltage transient is induced at that node. The transient is referred to as an SET, which can flip the stored data ('0' to '1' or vice versa) if the amplitude and

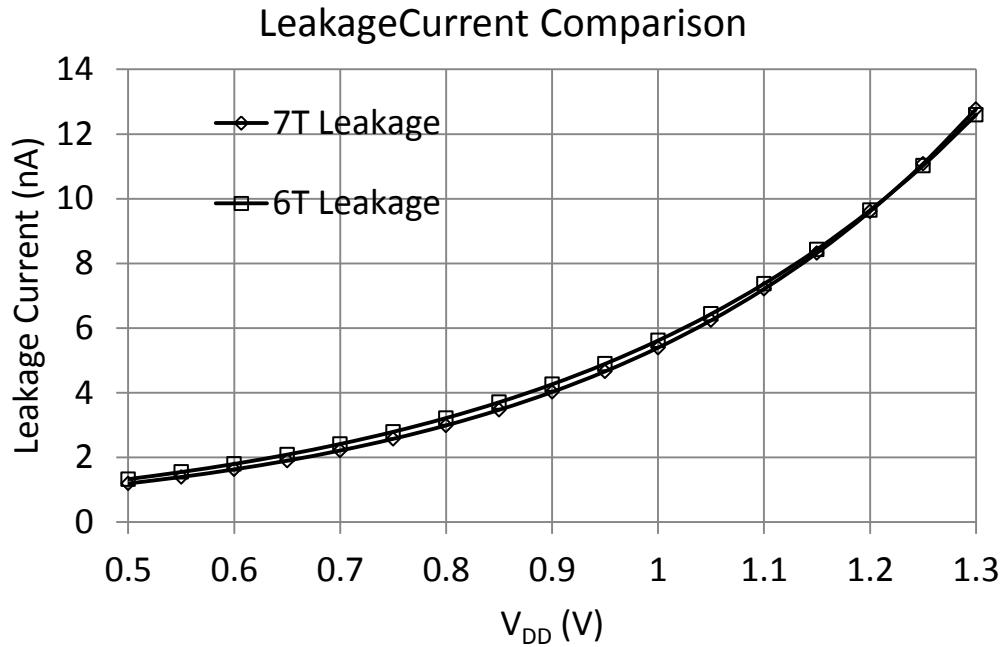


Figure 5.9: A comparison of leakage currents of 6T cell and the proposed 7T cell as a function supply voltage.

duration of the SET is large. Such data flipping is referred to as a single event upset (SEU) or ‘soft error’ as it does not permanently damage the memory circuit. However, SEUs cause computational errors, which can lead to system failure. Accordingly, state-of-the-art microprocessors require SEU protection [23]. Since a microprocessor or an SOC consist of a large number of SRAM cells, making the SRAM cells SEU robust is vital to ensure the overall reliability of the system.

Typically, an SRAM cell experiences a SEU by having an SET at a sensitive node of the back-to-back inverter inside cell. The vulnerability of SRAM to soft error is assessed by its critical charge ( $Q_{crit}$ ) [24].  $Q_{crit}$  is the minimum amount of charge that can flip the data bit stored in an SRAM cell. It exhibits an exponential relationship with the soft error rate (SER) [25]. It should be as high as possible in order to limit the

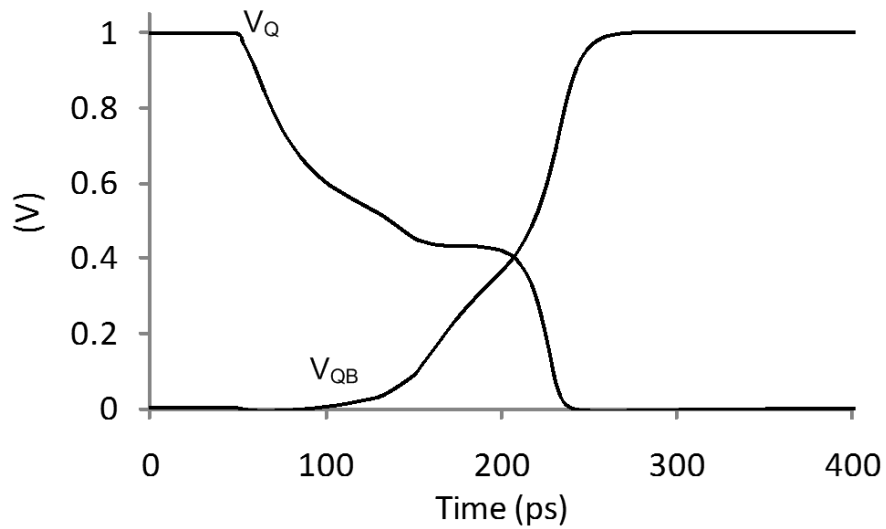


Figure 5.10: Time domain plots of cell node voltages (from Figure 2.2) for a state-flipping case.

SER. The various critical charge models which have been reported to date agree in the qualitative definition. However, they differ in quantitative description. For example, in [24] and [26],  $Q_{crit}$  has been modeled by the following equation,

$$Q_{crit} = C_N V_{DD} + I_{DP} T_F \quad (5.1)$$

Where,  $C_N$  is the equivalent capacitance of the struck node,  $V_{DD}$  is the supply voltage,  $I_{DP}$  is the maximum current of the ON PMOS transistor and  $T_F$  is the cell flipping time. If an amount of charge equal to or greater than  $Q_{crit}$  is drained from (or injected in to) the '1' (or '0') node, the connecting PMOS (or NMOS) will not be able to supply (or drain) that charge and subsequently the data flips as shown in Figure 5.10. In a conventional 6T cell the driver NMOS has a width of 1.5 to 1.7 times more than that of the PMOS for sufficient write margin. The mobility of n-channel is usually 2 to 3 times of that of a p-channel and as a result, the strength of the driver

NMOS is several times higher than that of the PMOS. In a back-to-back inverter data is retained by two nodes having complementary value, namely '0' and '1'. '0' is retained by the connecting NMOS and '1' is retained by the connecting PMOS. If a SET hits the '0' node and tries to change the voltage level, the connecting NMOS is more successful in retaining it than the PMOS when a SET hits the '1' node because the strength of NMOS is higher than the PMOS. Since, vulnerability is to be assessed by the worst case of the two types of possible flipping scenario,  $Q_{crit}$  of an SRAM cell is measured from the '1 to 0' flipping scenario. As a result, the recovering current used in (5.1) is PMOS current.

A dilemma in 6T SRAM cell is that PMOS cannot be upsized, since that would require strengthening the access transistor (for maintaining writability) and subsequently the driver NMOS (for ensuring read stability). But in the 7T cell there is no such restriction. In fact, to maintain equal critical charge for both '0 to 1 flip' and '1 to 0 flip' the aspect ratio of the PMOS should be at least twice of the driver NMOS, which is not possible in 6T-cell. Even in 8T cell [11], where read bit line is decoupled and thus there is no need for the driver NMOS to be stronger than the access transistor, the PMOS cannot be made too strong. Because that would make the write margin too small and thus the writability may totally disappear in worst case variation scenario. But in 7T cell, such design can be accommodated. A comparison of critical charge for 6T and the proposed 7T SRAM cell is given in Figure 5.11. And more importantly if leakage power consumption is not the main issue then the width of the inverter pull-up transistor can be increased for higher critical charge without sacrificing the write margin.

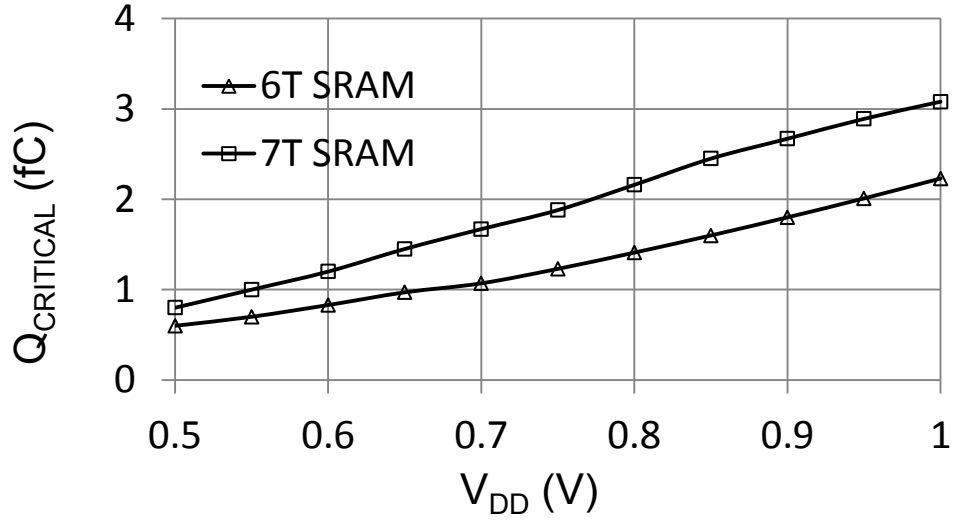


Figure 5.11: Comparison of critical charge between 6T and the proposed 7T SRAM cells.

The SER per bit in an SRAM has been described and experimentally verified by the following empirical model by Hazucha and Svensson [25].

$$SER_{bit} \propto FA \exp\left(-\frac{Q_{crit}}{Q_s}\right) \quad \text{or,}$$

$$SER_{bit} = KFA \exp\left(-\frac{Q_{crit}}{Q_s}\right) \quad (6.2)$$

Here, F is the neutron flux with energy greater than 1 MeV, in particles/cm<sup>2</sup>-s; A is the sensitive area of the circuit, in cm<sup>2</sup>; and Q<sub>s</sub> is the charge collection efficiency of the cell in fC. Typically, Q<sub>s</sub> is dependent on the magnitude of the particle-induced charge, substrate doping, carrier mobility, and the voltage of the collecting node and neighboring nodes. Since different cells have different charge collection volume they may have different charge collection efficiency from a single particle strike. However in the first-order if we assume that the charge collection efficiency of the sensitive

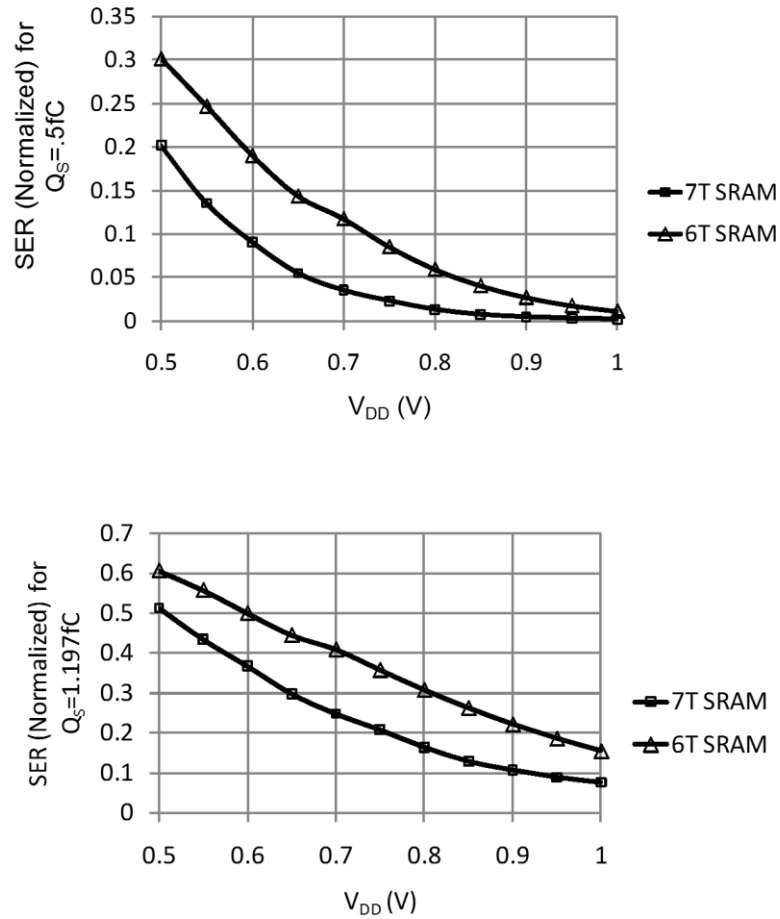


Figure 5.12: Comparison of SER between 6T and the proposed 7T SRAM cell.

node is same in each case, we can estimate the normalized SER of the cells by assuming  $KFA=1$ . From [27] an experimental value of  $Q_s$  is taken to be  $1.187fC$ . Based on that, SER for two test case of  $Q_s = 0.5fC$  and  $1.187fC$  is shown in Figure 5.12.

## 5.6 Cell Area

Silicon die area is a very expensive resource and since memory accounts for as much as 80% of the total area of an SOC, cell area is a very important factor in memory

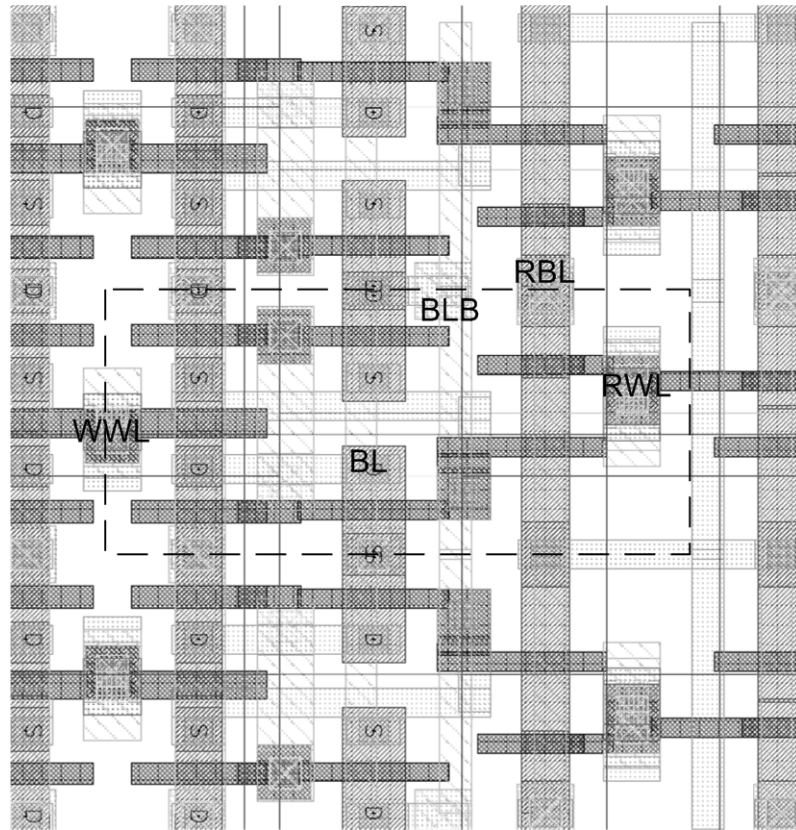


Figure 5.13: 7T cell Layout (The area inside the dotted boundary belongs to one cell).

design. Though 7T cell has one more transistor than 6T cell, the area does not increase because that seventh transistor, which is an NMOS, is accommodated between the two driver NMOS of the inverters. The area of a 7T SRAM cell is same as a 6T SRAM cell.

In the layout, 3 metal layers was used which is the minimum even in conventional 6T SRAM designs. Metal1 is used for interconnections inside the cell, Metal2 is used for bit lines and  $V_{SS}$ , and Metal3 is used for the word lines. The layout is shown in Figure 5.13.

### 5.7 Performance of the Sense Amplifier

The performance of the proposed sense amplifier has been simulated with the proposed 7T SRAM cell. From the operation of the sense amplifier it has been seen that, after resetting when SAE signal is asserted for evaluation, the sense amplifier itself will discharge the read bit line, even if the cell does not. However, in such case the sense amplifier still remains in the reset condition, indicating a '1' read. However, if there is also a discharge by the SRAM cell, then the state of the sense amplifier flips, indicating a '0' read. The waveform of the read bit line voltage during read operation is shown in Figure 5.14. The wave form of '0' read and '1' read are shown in Figure 5.15. During read operation the read bit line discharges by 65mV for '1' read and 160mV for '0' read.

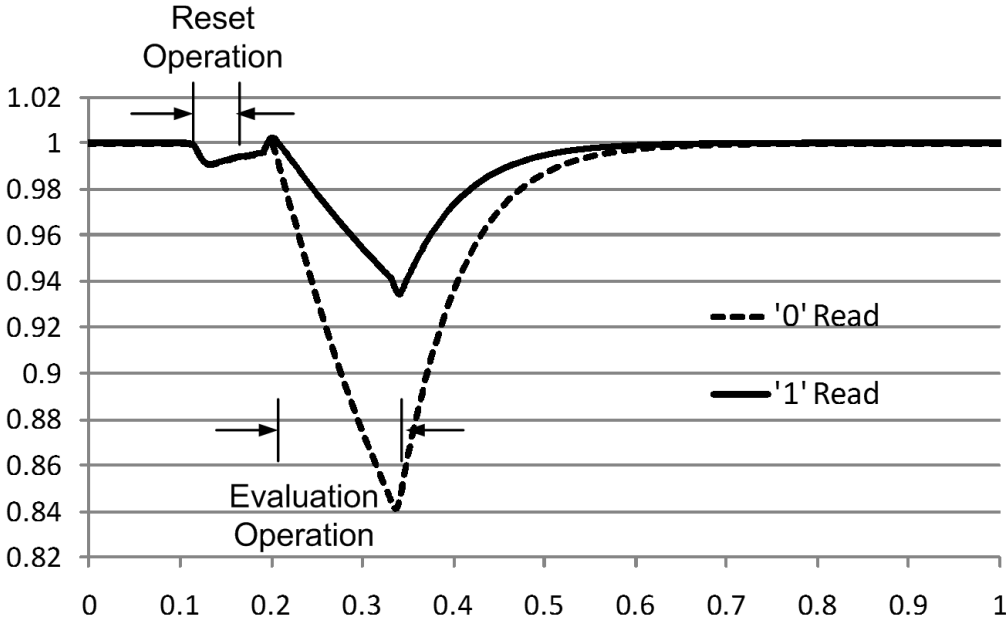


Figure 5.14: Waveform of the read bit line during read operation.



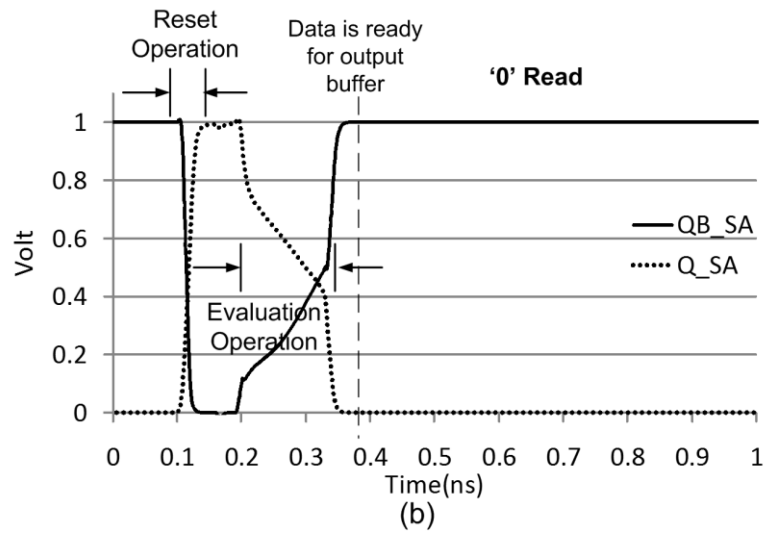
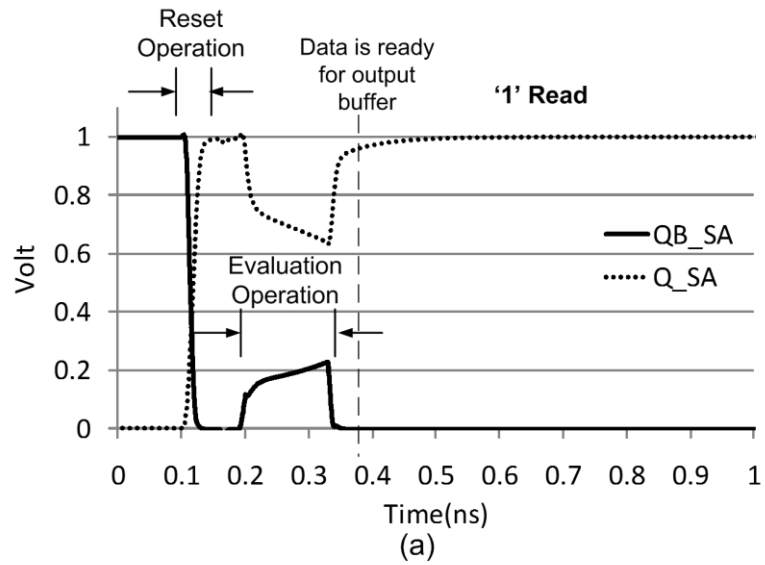


Figure 5.15: Waveform of the two nodes of the latch inside the sense amplifier during read operation. (a) During '1' is being read. (b) During '0' is being read.

# Chapter 6

## A Low-Leakage Array Architecture with Column Virtual Grounding

As was mentioned earlier, the proposed 7T SRAM cell itself cannot support multiple words in one row because during write operation of one word the other words in the same row will be subjected to a vulnerable “half-selected” state. But multiple words per row can be used to improve array efficiency by multiplexing adjacent columns into shared sense amplifiers. It allows the banks to be larger which lessens the required number of banks. And that lessens the required decoding circuitry. It also enables protection from multi-bit soft error events.

The 7T SRAM cell can support multiple words in a row if the array is implemented by column virtual grounding (CVG) techniques, as was proposed in [28]. The half-selected vulnerability can be removed by applying the CVG techniques. The principle of the CVG technique is that all the cells in a column share a common  $V_{GND}$ ,

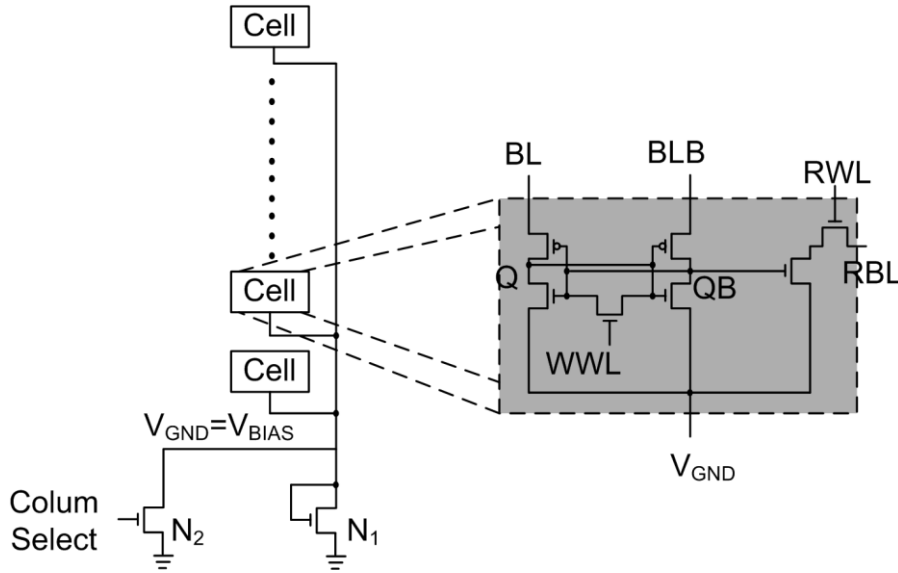


Figure 6.1: Memory array using Column Virtual Grounding (CVG).

which is connected to the source terminals of the three driver NMOS transistors per cell (Figure 6.1).

## 6.1 Array Implementation with CVG

An array implementation of the proposed bit-cell utilizing CVG technique is shown in Figure 6.2. During hold mode the  $V_{GND}$  of all the columns are kept at a non-zero value, namely  $V_{BIAS}$ . During a write operation (as well as read operation)  $V_{GND}$  of only the columns containing the targeted words are pulled down to 0V from  $V_{BIAS}$ . And the respective BL and BLB are discharged according to the data intended to be written. However, the activated WWL signal also turns on  $W_{AX}$  of the other cells in the same row. But their  $V_{GND}$  remain at  $V_{BIAS}$ . Even though reverse body bias is applied to  $M_{NL}$ ,  $M_{NR}$ , and  $W_{AX}$ ,  $W_{AX}$  becomes comparatively weaker than  $M_{NL}$  and  $M_{NR}$ . As a result,

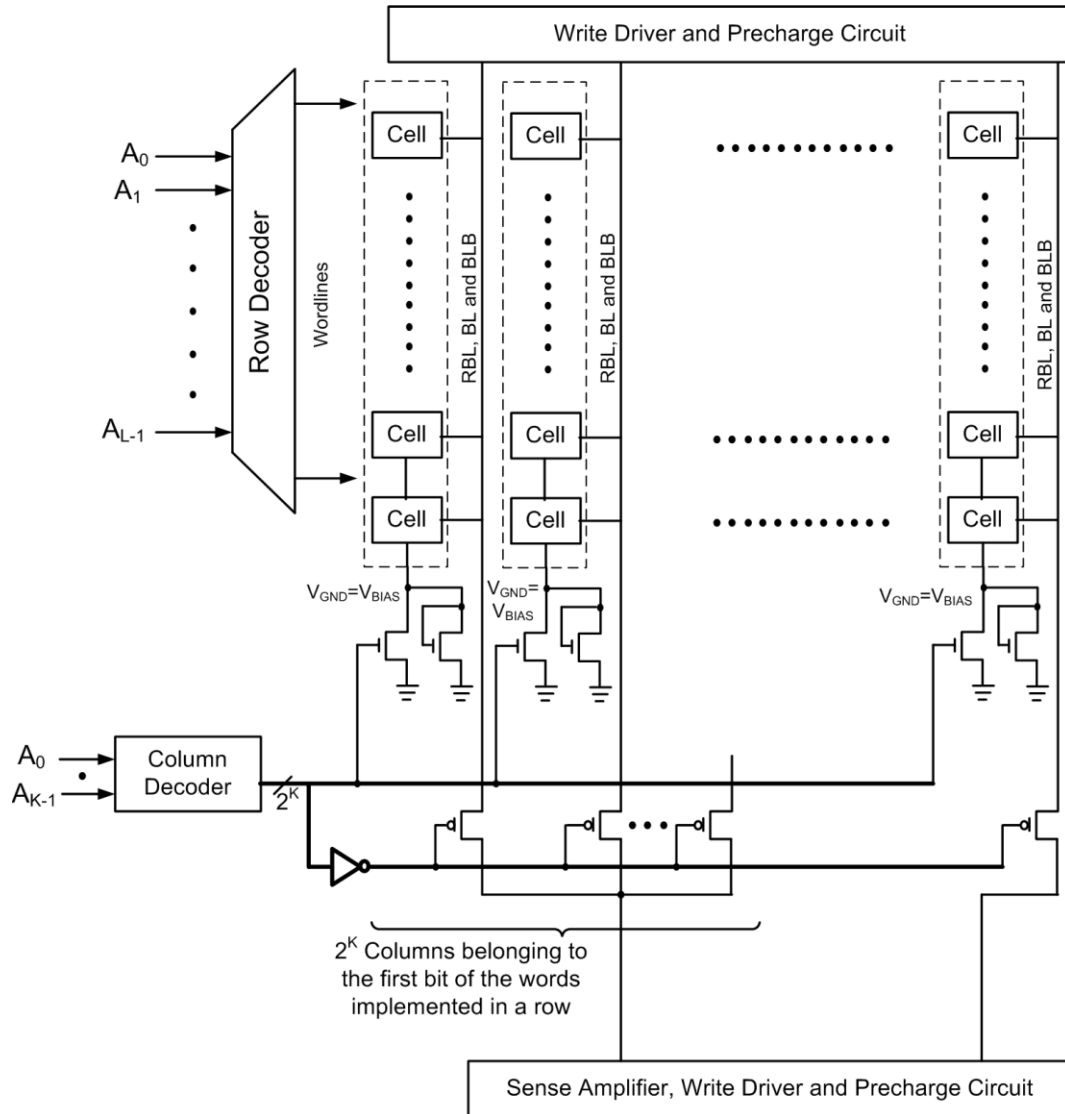


Figure 6.2: Array implementation of the proposed 7T SRAM cell with Column Virtual Grounding.

the cells belonging to the other words in the same row do not flip. The situation of the half-selected cells (belonging to the columns whose  $V_{GND}$  remain at  $V_{BIAS}$ ) becomes tantamount to using a  $W_{AX}$  with longer channel length. Thus, the proposed bit-cell can also provide efficient bit-interleaving structure to achieve soft-error tolerance with ECC.

For read operation, similarly only the columns containing the targeted word are pulled down to 0V. And the respective read bit line discharges (or not) according to the stored data. The cells belonging to the unselected columns do not have sufficient overdrive in  $R_{AX1}$  and  $R_{AX2}$  since their  $V_{GND}$  is kept at  $V_{BIAS}$ . Thus, their respective read bit line discharges by small amount, which saves the subsequent precharge energy.

## 6.2 Performance Results

Monte-Carlo simulation of 1000 run was done with a  $V_{GND}=300\text{mV}$  and  $400\text{mV}$  with  $V_{DD}=1\text{V}$  and no instances of flipping was observed when the cell WWL was asserted and BL/BLB was kept at  $V_{DD}$  (which is the “half-selected state” defined in sub-section 4.2.2). However, when same simulation was performed for  $V_{GND}=0\text{V}$ , which is equivalent to no virtual grounding, more than 200 instance of flipping was observed. A transient waveform of the two storage nodes during half selected state for  $V_{GND}=0\text{V}$  and  $300\text{mV}$  is shown in Figure 6.3. It can be seen that the data does not flip in half-selected state for both cases. But, these simulations correspond to an ideal scenario with no variation. It should be noted from Figure 6.3(a) and (b) that even though data does not flip in both case, the difference between the two voltage levels during half selected state is larger for  $V_{GND}=300\text{mV}$ . Thus, if process variation were included in the simulations, there would have been fewer flipping instances for  $V_{GND}=300\text{mV}$  than for  $V_{GND}=0\text{V}$ .

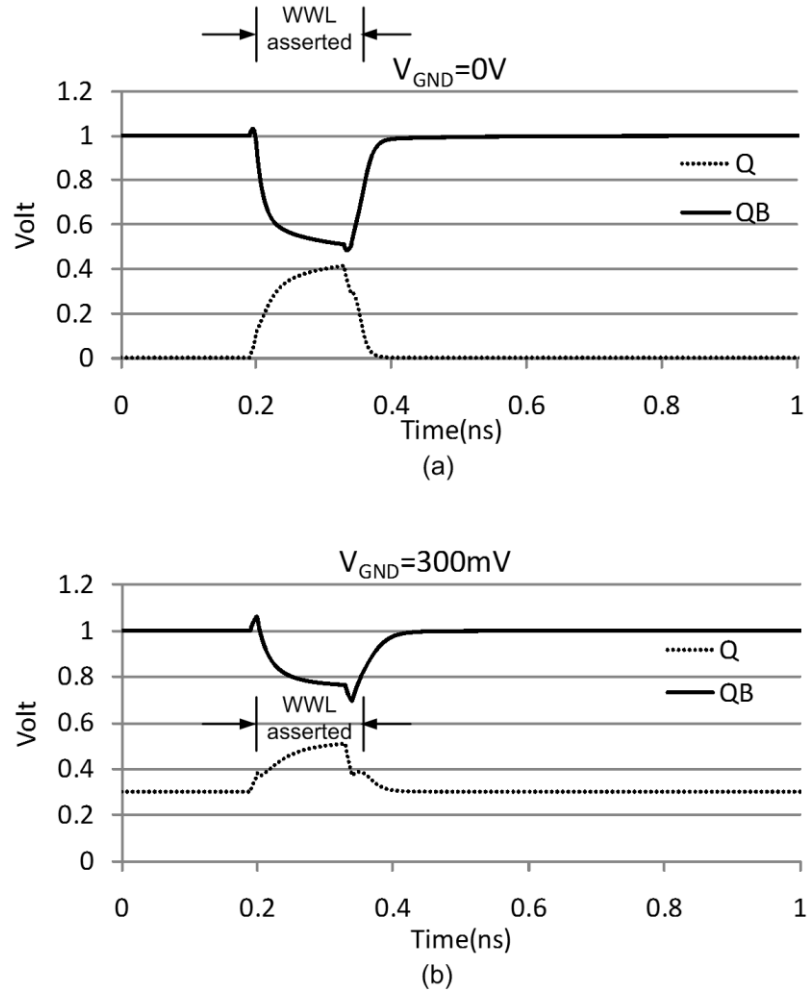


Figure 6.3: Transient waveform of half-selected state. (a) When  $V_{GND}=0V$ . (b) When  $V_{GND}=300mV$ .

A leakage comparison between with and without virtual grounding of the proposed 7T cell is shown in Table 6. A comparison of leakage currents of 6T cell and the proposed 7T cell as a function of rail-to-rail voltage is shown in Figure 6.4 (rail-to-rail voltage for 6T is  $V_{DD}-0V$  while for 7T cell is  $V_{DD}-V_{GND}$ ).

The power savings from any type of virtual grounding techniques (or virtual  $V_{DD}$ ) depend on the switching activity factor (minimum average time between two consecutive accesses). Because, whenever a data is accessed, the  $V_{GND}$  (or  $V_{DD}$ )

Table 6: Leakage comparison between with and without virtual grounding ( $V_{DD}=1V$ ).

| $V_{GND}$ | Leakage current per cell (nA) |             | Average (nA) |
|-----------|-------------------------------|-------------|--------------|
|           | Storing '0'                   | Storing '1' |              |
| 0V        | 6.4                           | 4.38        | 5.39         |
| 300mV     | 1.76                          | 1.58        | 1.67         |
| 400mV     | 1.27                          | 1.18        | 1.23         |

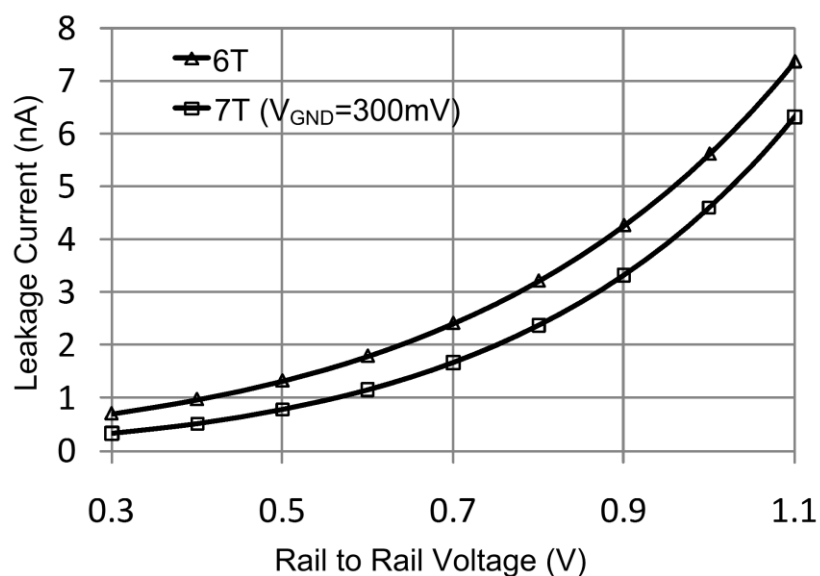


Figure 6.4: A comparison of leakage currents of 6T cell and the proposed 7T cell as a function of rail-to-rail voltage.

lines have to be activated and that consumes some dynamic power. If the switching activity factor is high, the dynamic power consumption for activation may offset the leakage power savings. Also, the power efficiency of the column virtual grounding techniques depends on the number of words implemented in a row. In fact, the CVG technique is more power-efficient when the number of words implemented in a row is large. Based on the first order analysis an estimate of the average time between two

Table 7: The minimum average time between two consecutive access with CVG so that leakage power offsets the dynamic power needed for each access.

| Number of word implemented in a row | Minimum Average time between two consecutive access (ns) |
|-------------------------------------|--|
| 4                                   | 41   |
| 8                                   | 16   |
| 16                                  | 4  |

consecutive accesses for different number of words implemented in a row, so that the leakage power savings offset the dynamic power consumption, is given in Table 7.



# Chapter 7

## Conclusion

### 7.1 Contribution to the Field

Due to scaling, current CMOS process technologies are suffering from increased process variations. As a result, SRAM, which uses the smallest possible transistors but occupies the majority of the die area, is becoming the circuit block most susceptible to process variations. In this thesis, an SRAM architecture, consisting of a bit-cell topology, a sense amplifier, and an array implementation, has been proposed to solve these problems.

#### 7.1.1 The Proposed 7T SRAM Cell

The 7T cell proposed in this work is highly suitable for on-chip L1-cache (e.g., first-level cache in a microprocessor as shown in Figure 1.2). The small bit count and lower array efficiency of such arrays minimizes the impact of the 7T-SRAM's lack of column selectivity when implemented without CVG. The proposed cell incurs reduced write power (half compared to 6T SRAM cell) which will result in reduced heat dissipation.

Such performance is highly desired in those arrays closest to the microprocessor core. The decoupled read operation of the cell removes the “read stability problem” of the 6T SRAM cell. The area of the cell layout is same as that of the 6T SRAM cell while the other proposed cells do incur large area overhead (13% in [11], 30% in [12], 37% in [13]). The leakage power consumption is also same as the 6T SRAM cell. However, the proposed SRAM cell suffers from a limitation. The cell cannot support multiple bits in one row unless used with the virtual ground scheme. Of course, the virtual ground scheme provides a significant reduction in the leakage power, which is a critical concern for SRAM arrays.

### **7.1.2 The Proposed Single-Ended Sense Amplifier**

The proposed sense amplifier is particularly suitable to be used with the proposed 7T SRAM cell. The sense amplifier, being of similar structure as the bit-cell, can be laid out with similar dimensions as the bit-cell itself. Thus, it can be pitch matched with the cell array, even if one word per row is implemented.

### **7.1.3 A Low-Leakage Array with Multiple Words in a Row**

By utilizing CVG, the proposed cell can support multiple words per row. Thus the proposed bit-cell can also be used where larger banks are required (e.g. L2 cache). Multiple words per row also allow simple error correcting codes to be effectively used for soft error protection. Moreover, the CVG has the inherent advantage of reducing leakage power consumption, which is highly desirable where the size of the memory bank is large.

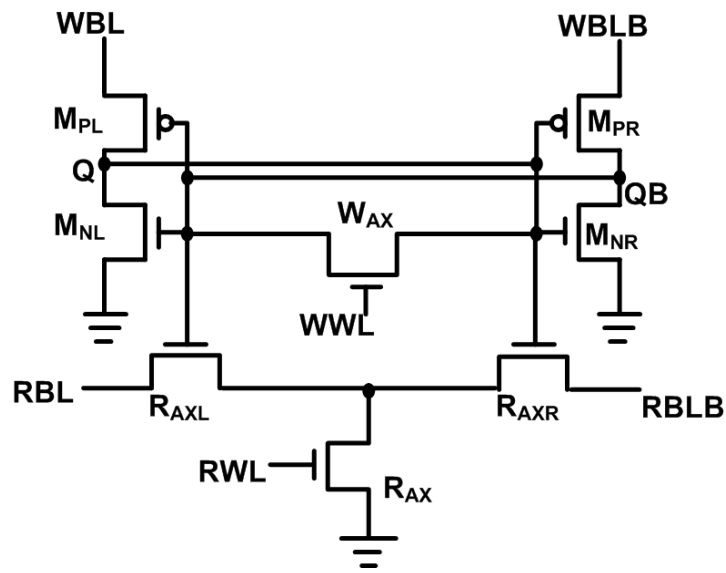


Figure 7.1: An enhanced version of the proposed cell.

## 7.2 Future Works

The most salient feature of the proposed 7T SRAM cell is its low write power (half of the power required in the 6T SRAM cell). Retaining this, enhancement can be made according to specific applications. Some suggestions for future work are:

- 1) The proposed 7T SRAM cell can be further enhanced by applying one more transistor to make the read operation differentially sensed (Figure 7.1). That way conventional sense amplifier can be utilized.
- 2) The increasing use of battery operated portable devices like cell phones, GPS devices, music players, etc. have increased research in decreasing the power consumption of these devices. These devices typically use low power SOCs. Since the caches constitute most of the transistors on SOCs, it is imperative that the cache design incorporates techniques to reduce the power consumption. The

proposed SRAM can be investigated for near threshold or sub-threshold operation.

## References

- [1] International Technology Roadmap for Semiconductors. Link: <http://www.evaluationengineering.com/index.php/solutions/ate/manufacturability-with-embedded-infrastructure-ips.html>.
- [2] S. Rusu, J. Stinson, S. Tam, J Leung, H. Muljono, and B.Cherkauer, "A 1.5-GHz 130-nm Itanium® 2 Processor with 6-MB on-die L3 cache". *IEEE Journal of Solid-State Circuits*, vol. 38, no. 11, pp. 1887–1895, Nov. 2003.
- [3] John L. Hennessy and David A. Patterson, *Computer Architecture – A Quantitative Approach, Fourth edition*, San Francisco, USA, Morgan Kaufmann Publishers, 2007, Chapter 5, pages. 288.
- [4] Kevin Zhang (Ed), *Embedded Memories for Nano-Scale VLSIs*. Springer, LLC, 233, Springer Street, New York, 2009, Chapter 2, page. 7.
- [5] J. D. Schmidt, "Integrated MOS random-access memory." *Solid-State Design*, pp. 21–25, 1965.
- [6] Moore's Law Made real by Intel Innovations. Available: <http://www.intel.com/technology/mooreslaw/>.
- [7] Kevin Zhang, *Embedded Memories for Nano-Scale VLSIs*. Springer, LLC, 233, Springer Street, New York, 2009, Chapter 3, page. 45.

- [8] M. Orshansky, S. Nassif, and D. Boning, *Design for Manufacturability and Statistical Design*, Springer Publications, Springer US, 233 Spring Street, New York 10013, 2007. Chapter 2, page 12.
- [9] R. K. Krishnarnurthy, A. Alvandpour, V. De, and S. Borkar, "High-performance and low-power challenges for sub-70 nm microprocessor circuits," *Proc. IEEE Custom Integrated Circuit Conf.*, pp. 125–128, 2002.
- [10] Shekhar Borkar , "Design Challenges of Technology Scaling" . Available: <http://www.cs.utexas.edu/~hestness/papers/borkar-techscaling.pdf>.
- [11] K. Takeda, Y. Hagihara, M. Nomura, Y. Nakazawa, T. Ishii, and H. Kobatake, "A Read Static Noise Margin Free SRAM cell for Low Vdd and High Speed Applications", *IEEE Journal Solid-State Circuits*, vol. 41, no. 1, pp.113-121, Jan. 2006.
- [12] L. Chang, R. K. Montoye, Y. Nakamura, and K. A. Batson, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches." *IEEE Journal of Solid-State Circuits*, vol. 43, No. 4, pp. 956-963, April 2008.
- [13] Z. Liu and V. Kursun, "Characterization of a Novel Nine-Transistor SRAM Cell." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 16, no. 4, pp.488-492, April 2008.
- [14] E. Seevinck *et. al.*, "Static-noise margin analysis of MOS SRAM cells," *IEEE Journal of Solid-State Circuits*, Vol.22, No. 5, pp. 748-754, October 1987.

- [15] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits - A Design Perspective*. Upper Saddle River, New Jersey: Prentice Hall, 2002.
- [16] A. Pavlov, and Manoj Sachdev, *CMOS SRAM circuit design and parametric test in nano- scaled technologies: Process aware SRAM design and Test*, Springer, Page-1, December 2010.
- [17] K. Itoh (Ed.), Masashi Horiguchi (Ed.), and Hitoshi Tanaka (Ed.), *Ultra-Low Voltage Nano-Scale Memories*. 2007. Springer, LLC, 233, Springer Street, New York, chapter 4, page. 159.
- [18] A. Chandrakasan, W. Bowhill, and F. Fox, *Design of High-Performance Microprocessor Circuits*, Wiley-IEEE Press, 2001, Chapter 6, page. 101.  
Available: <http://0-ieeeexplore.ieee.org/mercury.concordia.ca/xpl/bkabstractplus.jsp?bkn=5266000&tag=1> (cited on 27th July, 2011).
- [19] M. Wieckowski, S. Patil, and M. Margala, "Portless SRAM—A High-Performance Alternative to the 6T Methodology." *IEEE Journal of Solid-State Circuits*, vol. 42, no.11, pp.2600-2610, November 2007.
- [20] G. Torrens, B. Alorda, S. Barceló, J. L. Rosselló, S. A. Bota, and J. Segura, "Design Hardening of Nanometer SRAMs through Transistor Width Modulation and Multi-Vt Combination," *IEEE Transaction on Circuits and Systems—II: Express Briefs*, Vol. 57, No. 4, pp. 280-284, April 2010.

- [21] S. S. Mukherjee, J. Emer, and S. Reinhardt, "The soft error problem: an architectural perspective," in *Proc. Int. Symp. on High-Performance Computer Architecture (HPCA)*, pp. 243–247, Feb. 2005.
- [22] R. C. Baumann, "Soft errors in advanced computer systems." *Design & Test of Computers, IEEE*. volume: 22 issue: 3, pp.258-266, May-June 2005.
- [23] D. Krueger, E. Francom, and J. Langsdorf, "Circuit design for voltage scaling and SER immunity on a quad-core Itanium® processor." *Solid-State Circuits Conference, ISSCC 2008. Digest of Technical Papers. IEEE International*. pp. 94-95.
- [24] P. Roche, J. M. Palau, C. Tavernier, G. Bruguier, R. Ecoffet, and J. Gasiot, "Determination of key parameters for SEU occurrence using 3-D full cell SRAM simulations," *IEEE Transaction on Nuclear Science*, vol. 46, no.6, pp. 1354–1362, Dec. 1999.
- [25] P. Hazucha and C. Svensson, "Impact of CMOS Technology Scaling on the Atmospheric Neutron Soft Error Rate." *IEEE Transaction on Nuclear Science*, vol. 47, no. 6, pp. 2586-2594, December 2000.
- [26] J. M. Palau, G. Hubert, K. Coulie, B Sagnes, M. C. Calvet, and S. Fourtine, "Device simulation study of the SEU sensitivity of SRAMs to internal ion tracks generated by nuclear reactions." *IEEE Transaction on Nuclear Science*., vol. 48, no. 2, pp. 225–231, Apr. 2001.



- [27] S. M. Jahinuzzaman, J. S. Shah, D. J. Rennie, and M. Sachdev, "Design and Analysis of A 5.3-pJ 64-kb Gated Ground SRAM With Multiword ECC." *IEEE Journal of Solid-State Circuits*, vol. 44, no. 9, pp. 2543-2553, September 2009.
- [28] N. Shibata, "A switched virtual-GND level technique for fast and low power SRAM's," *IEICE Trans. Electron.*, vol. E80-C, pp. 1598-1607, 1997.

# Glossary

BL, BLB     Bit line, Bit line Bar (Complementary Bit line)

CPU            Central Processing Unit

DRAM         Dynamic Random Access Memory

ECC            Error Correcting Code

RWL            Read Word line

SET            Single Event Transient

SEU            Single Event Upset

SNM            Static Noise margin

SOC            System on Chip

SRAM         Static Random Access Memory

WL            Word line

WWL          Write Word line