# DESIGN OF A FLEXIBLE HIGH TEMPERATURE SRAM WITH REDUCED DESIGN TIME

By

SRIKANTH VELLORE AVADHANAM RAMAMURTHY

Bachelor of Engineering in Electronics & Communication

Anna University

Chennai, Tamilnadu, India

2005

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December 2007

DESIGN OF A FLEXIBLE HIGH TEMPERATURE

SRAM WITH REDUCED DESIGN TIME

Thesis Approved:

Dr. Chris Hutchens

Thesis Adviser

Dr. Louis G. Johnson

Dr. James Stine

Dr. A. Gordon Emslie

Dean of the Graduate College

# Acknowledgements

First of all, I would like to thank my advisor Dr. Chris Hutchens for his continuous advice, support and patience throughout the process of accomplishing this work. I would also like to express my gratitude to the committee members Dr. Louis Johnson, and Dr. James Stine for their valuable suggestions.

I am fortunate to have worked as a research assistant at MSVLSI lab and would like to extend my gratitude to all the members of MSVLSI team, and especially Dr.Liu, Singaravelan, Srinivasan, Ranganathan, Vijay, Usha, Henry, Dr.Hooi Miin Soo, Sheshnag, and Vibhore for being an inspiration to me and also helping me with their efforts to make the learning process more interesting. Special thanks to Dr. Liu, Henry & Dr. Hooi Miin Soo for their unselfish efforts in testing the fabricated circuits.

I would like to dedicate my work to my parents Mr. V.A Ramamurthy and Mrs. Gowri, my brothers, Srihari and Sridhar and all my friends for having unwavering belief in me and encouragement.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION ..... 1

1.1 Motivation ..... 2
1.2 Thesis organization ..... 3

II. OVERVIEW OF MEMORY ARCHITECTURES ..... 5

2.1 Conventional Array Architecture ..... 5
2.2 Divided Word-line Architecture ..... 7
2.3 Divided Bit line Architecture ..... 8
2.4 Folded Bit line Architecture ..... 10
2.5 Conclusion ..... 10

III. OVERVIEW OF SENSE AMPLIFIERS & DECODERS

3.1 Sense Amplifiers ..... 12
3.1.1 Conventional Latch-type Sense Amplifier ..... 13
3.1.2 Linear Differential Sense Amplifier ..... 15
3.1.3 Current Sense Amplifiers ..... 17
3.1.4 Voltage Latch-type Sense Amplifier ..... 20
3.2 Address Decoders ..... 24

IV. METHODOLOGY FOR REDUCED TIME SRAM SYNTHESIS ..... 28

4.1 SRAM cell 2-by-2 Block ..... 29
4.2 Sense Amplifier & Column Read/Write Logic ..... 31
4.3 Bank Control circuitry & Row Drivers ..... 34
4.4 SRAM Bank ..... 35
4.5 Global Control Circuitry ..... 36
4.6 Global Address Decoders ..... 38
4.7 Conclusion ..... 39

V. HIGH TEMPERATURE SRAM DESIGN EXAMPLE ..... 41

5.1 General considerations & device geometry selection ..... 45

5.2 Row & Column Cell Libraries ..... 47
5.2.1 Row Cell Library ..... 49
5.2.2 Column Cell Library ..... 50
5.3 SRAM cell 2×2 Module ..... 51
5.3.1 Standard 6-T SRAM cell Design ..... 51
5.3.2 SRAM cell module Testbench ..... 54
5.3.3 SRAM cell Module Layout ..... 57
5.4 Sense Amplifier & Column Read/Write Logic ..... 58
5.4.1 Sense Amplifier and Column read/write circuit Design ..... 58
5.4.2 Column Logic Simulation ..... 61
5.4.3 Column Logic Layout ..... 62
5.5 Bank control circuits & local row drivers ..... 63
5.6 SRAM Bank ..... 64
5.7 Global Control Circuitry ..... 66
5.8 Global Address Decoders ..... 67
5.9 Results & Waveforms ..... 70

VI. CONCLUSION & FUTURE WORK ..... 77

REFERENCES ..... 79

LIST OF TABLES

Table Page

Table 3.1 General delay expressions for various sense amplifiers .....22

Table 3.2 General Comparison of performance between various sense amplifiers.....22

Table 5.1 Design Specifications for OSU-HC11 SRAM.....44

Table 5.2 Typical model parameters of 0.5um Peregrine SOS process at 27C.....45

Table 5.3 Measured Ion/Ioff Ratios for 0.8um PMOS and 1.4um NMOS.....48

Table 5.4 Measured Maximum Leakage currents for 0.8um PMOS and 1.4um NMOS.....48

Table 5.5 Design Summary for OSU-HC11 SRAM.....71

Table 5.6 SRAM cell timing parameters & cell noise margins .....73

Table 5.7 Simulated Sense amplifier & column logic timing characteristics.....74

Table 5.8 Summary of measured characteristics of SPI SRAM.....76

LIST OF FIGURES


| Figure | | Page |
|---|---|---|
| Figure.2.1 | Conventional Array Architecture | 5 |
| Figure.2.2 | Divided Word Line Architecture | 7 |
| Figure.2.3 | Divided Bit Line Architecture | 8 |
| Figure.2.4 | Folded Bit line Architecture | 10 |
| Figure.3.1 | Inverter Latch Sense Amplifier | 14 |
| Figure.3.2 | Linear Differential Sense Amplifier | 16 |
| Figure.3.3 | Clamped Bit Line Sense Amplifier (CBLSA) | 18 |
| Figure.3.4 | Alpha Latch Sense Amplifier | 20 |
| Figure.3.5 | Cross Coupled Sense Amplifier | 23 |
| Figure.3.6 | (a) A simple 4-16 Decoder (b) 4-16 Decoder with two 2-4 predecoders | 25 |
| Figure.3.7 | A 3-to-8 Tree Decoder | 26 |
| Figure.4.1 | SRAM 2-by-2 Cell Module used in Test Bench | 30 |
| Figure.4.2 | Sense-Amp & Column Logic Module | 32 |
| Figure.4.3 | Column Logic Module Testbench | 33 |
| Figure.4.4 | SRAM Bank | 35 |
| Figure.5.1 | (a) On-Chip OSU-HC11 SRAM Write timing (b) On-Chip OSU-HC11 SRAM Read timing | 42 |
| Figure.5.2 | Layouts of row cell library (a) 1X Inverter (b) 2-input NAND gate | 49 |

Figure.5.3 Layouts of Column cell library (a) D-Latch (b) 3-state Buffer (1X) .....50

Figure.5.4 Circuit Schematic of 6-T SRAM cell with PMOS access transistors .....52

Figure.5.5 Circuit Schematic of SRAM cell 2×2 Testbench .....55

Figure.5.6 Cadence ADE window of SRAM cell 2×2 Testbench .....56

Figure.5.7 Layout of SRAM cell 2×2 Module.....57

Figure.5.8 Schematic of Sense amplifier for OSUHC11 SRAM .....59

Figure.5.9 Cadence ADE window of Column logic testbench.....62

Figure.5.10 Schematic of Local Control Signals.....64

Figure 5.11 Layout of SRAM Bank (a) Column logic and control circuitry
(b) small section of SRAM cell array along with its row drivers .....66

Figure 5.12 Schematic of 4-to-16 column decoder using 2-input pre-decoding .....69

Figure 5.13 Layout of 4k SRAM for OSU-HC11.....70

Figure 5.14 Simulated waveforms of SRAM cell read operation.....72

Figure 5.15 Simulated waveforms of SRAM cell write operation .....72

Figure 5.16 Simulated waveforms of Sense amplifier & column logic read cycle ....74

Figure 5.17 Logic Analyzer waveform of one SRAM Bank read operation.....75

# CHAPTER I

# INTRODUCTION

Memory arrays account for large portions of most of the modern day CMOS chips. Large sizes of Random Access Memories (RAM) are essential for microprocessors manufactured these days to manage large amounts of data and instruction bits. Static random access memories (SRAM) are predominantly used on-chip for these microprocessors in the form of high speed caches or simply as scratch memories. Hence the design techniques for the SRAM have evolved continuously to match the increased speeds of microprocessor systems as well as to comply with the process technology scaling. The transistor density for modern sub-micron memories is extremely high and hence results in higher current densities in areas where memories are concentrated thereby raising the issues of parasitics affecting the reliability of operations.

The power dissipation in today's CMOS chips is significantly contributed by the leakage power through the sub-micron devices which are primarily concentrated in the memory modules. It is hence a basic requirement of CMOS memories to achieve minimal power consumption and area overheads while striving for maximum possible speeds and yields. Hence the design complexities and the need to achieve stringent performance metrics, make the design of SRAM difficult, yet challenging.

This work takes up the design of a high temperature SRAM capable of operating over a wide range of temperatures to meet the requirements of downhole drilling application rendered by a high temperature micro-controller. At these elevated temperatures, the leakage power dominates any other mechanisms of power dissipation and if proper care is not exercised, the functionality and performance may be lost in addition to incurring power wastage.

## 1.1 Motivation

Although the applications requiring very high temperature operations of an electronic chip are less common, their need offers lot of opportunities that allow innovation and exploration of limits of technology. The design of such electronic systems intended to work over extreme temperature conditions have limitations in terms of achievable performance over these operating ranges, yet their need overrides the limitations.

This work is driven by the requirements of a high temperature micro-controller (OSU-HC11) which needs a specified size of on-chip and off-chip Static RAMs to be operated upto 275C in a sub-micron SOI process. In addition to being a complex custom design procedure, the design of such high temperature SRAMs are further complicated by the design time involved in characterizing and testing the components involved the process. It would be a lot easier to have re-usable design structures for high temperature SRAMs just as a standard digital cell library, which could be utilized to design variants of an SRAM design with much lesser effort and reduced time.

Commercial solutions are available for generating SRAMs using memory compilers just as the usage of standard cell libraries for digital circuit designs. But they are

expensive and often do not include the extreme temperature requirements of certain applications as in this work. This led to the idea of providing a simpler solution to re-use components needed to design a high temperature SRAM, and to develop a methodology that could reduce the design time of SRAMs by simple partitioning into several critical sub-modules and characterizing them for performance across the design corners. This methodology could provide ready-to-use/simpler-to-redesign components of a high temperature SRAM.

## 1.2 Thesis organization

This thesis is organized as 6 chapters, with the first chapter, being discussed here. Chapter 2 provides an overview of common types of memory architectures utilized to form an SRAM with its final section justifying the chosen architecture. An overview of various types of sense amplifiers, their performance characteristics and comparisons are discussed in chapter 3 while specifying the type of sense amplifier used in this work. This chapter also provides a quick insight into few address decoder design techniques with emphasis on usage of pre-decoders. Chapter 4 proposes the methodology of design time reduction for an SRAM where the memory system design is partitioned into several critical sub-modules. Chapter 5 discusses the implementation details of OSU-HC11 high temperature SRAM systems simultaneously developing on the methodology proposed in chapter 4. At the end of this chapter, various simulation and measurement results of the designed high temperature SRAM are presented. The summary of the design of high temperature SRAM and the future work in this direction is presented as the last chapter 6.

# CHAPTER II

# OVERVIEW OF MEMORY ARCHITECTURES

Judicious selection of architecture for SRAM can prove beneficial in terms of delay and power. A lot of architectures have been proposed in the literature in an effort to optimize the performance metrics of the memories. Most of these architectures utilize array partitioning to reduce word line and data path delays. Partitioning SRAM cell arrays also helps reduce bit line swings and active power. This chapter presents the basic structure to realize an SRAM in the first section, with the following sections discussing few commonly used partitioning strategies, and their advantages over the basic structure.

## 2.1 Conventional Array Architecture

This is an elementary architecture to realize an SRAM. As shown in Figure.2.1, it consists of a single array of memory cells with each memory cell being formed at the intersection of a word line and a bit line. It forms a matrix structure with $2^m$ rows by $2^n$ columns, where m + n forms the total number of address lines.

During a read/write only one out of $2^m$ word lines is enabled and it drives all the cells connected to this word line. But then only one cell is selected for a read/write by the column multiplexer using the $2^n$ lines of the column decoder. If a group of bits are to be

stored/read as a single entity such as a word (let us say an 8-bit word), then instead of selecting a single cell for each operation, the column multiplexer now selects all the cells corresponding to a word (in our case 8 cells). Each 8-cell group functions as a word, and accordingly the data bus is made 8-bit wide.

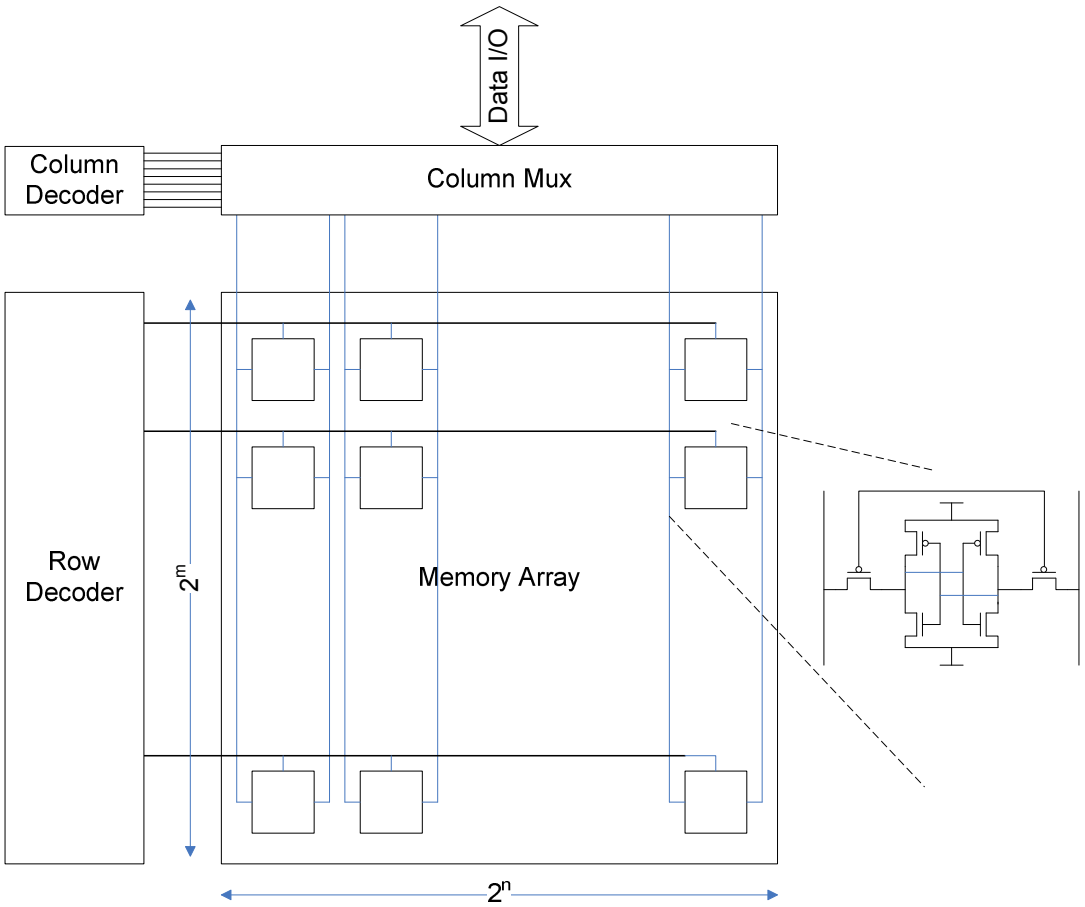



Figure.2.1 Conventional Array Architecture

## 2.2 Divided Word-line Architecture

The basic SRAM architecture shown in previous section suffers from two serious drawbacks. First and foremost the RC delay associated with word lines and bit lines grow proportionately with more number of cells along the Columns and Rows respectively. The word line is loaded by the SRAM cell's access/pass transistors (gate capacitances)

along the row and is proportional to the number of columns or bit lines. Similarly the Bit line is loaded by the diffusion capacitances of the SRAM cell pass transistors, and is proportional to the number of rows or word lines. Secondly the power dissipation on the bit lines and word lines increases linearly with capacitance.

These two drawbacks can be addressed by using an architecture called the divided word line architecture [1]. In this architecture, the single large memory array of the conventional architecture is broken down into several small sub-arrays called macros or banks. Each single macro can now generate its own word lines from the global word lines, by decoding them locally. Hence the RC delay of the global word lines is reduced because each single macro now presents only a small load compared to all the cells in a single row of the conventional architecture. Since the unused bit lines need not be enabled, it reduces the bit line power dissipation as opposed to the conventional architecture where all bit lines are enabled. Likewise the number of word lines may also be reduced and more macros may be added in vertical sense and using column multiplexing to access the word needed. This enables the delay on the bit lines to go down further.

The Divided word line architecture is shown in Figure.2.2. As shown in the figure, the structure has several sub-arrays forming the whole memory. Each sub-array has a certain specified number of Cells along its rows and columns. All the cells along the same row have a common word line. This common word line is derived from the global word line using a macro selector that runs vertically through the whole sub-array. Hence the capacitive load on the global word lines is considerably reduced because each macro/bank loads the global word line with just one gate. Also the word lines in each

macro have a reduced capacitance as the number of cells along the row is lesser. The macro/bank selector lines that run vertically along the columns are actually the output lines of the Column address decoder.

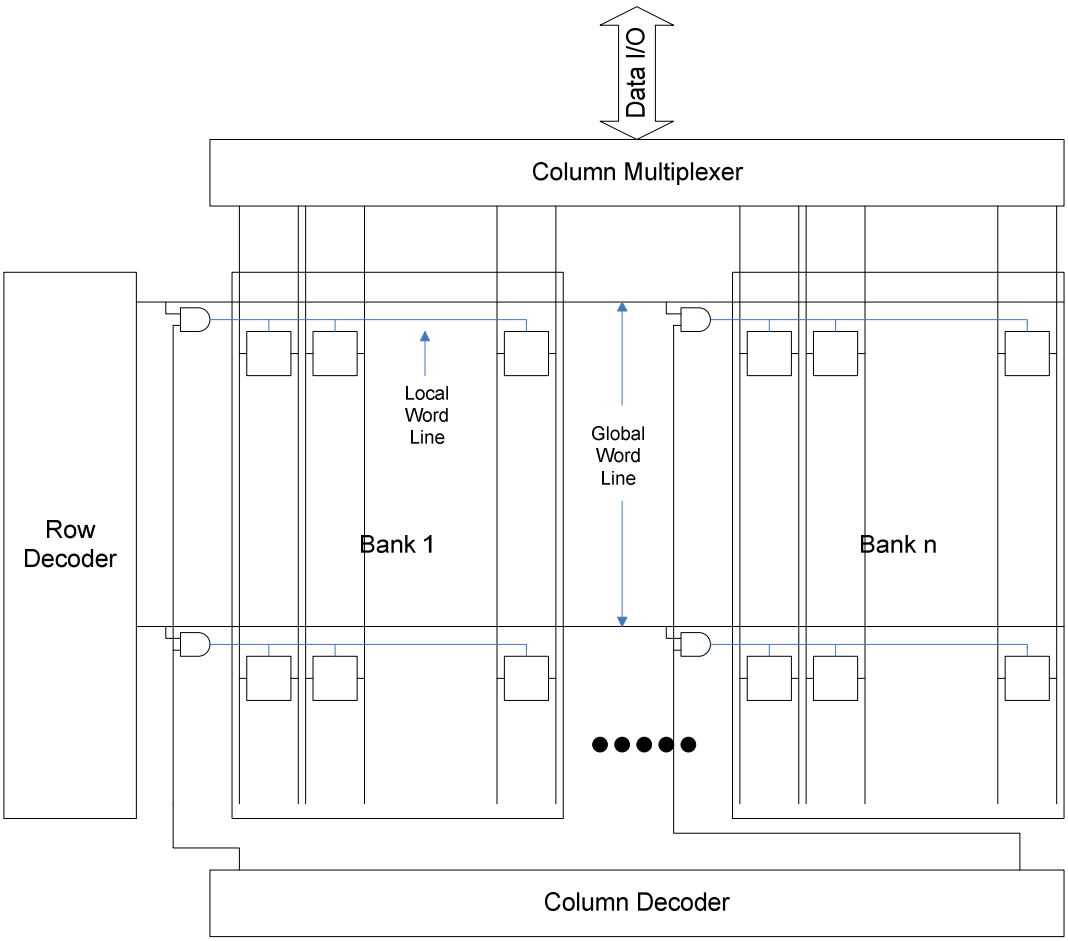



Figure.2.2 Divided Word Line Architecture

This structure helps in the reduction of dynamic power dissipation owing to the fact that the unselected macros don't switch. Bit lines of only the selected macro swing resulting in reduced power dissipation and often improved or reduced read delay.

The word line division procedure can be recursively followed on both the global and macro/block selector lines to achieve what is referred to as the Hierarchical Word Decoding (HWD) technique [2]. The number of hierarchies, in other words, the level of

word-line division, is determined by the total load capacitance of the word decoding path. HWD results in a tree type structure and as such may be optimized the same manner as clock trees or pad drivers, albeit considerably more complex. This method of sub-dividing the SRAM array increases area overhead at the boundaries of the partitions [3]. Also the use of word line drivers for each macro adds to the area overhead. This increase in area has a minor impact on power dissipation, and its effect is acceptable when compared to the benefits achieved in delay and overall reduction in dynamic power.

## 2.3 Divided Bit line Architecture

The division of column access lines of an SRAM array into several levels can also be applied in memory design. In other words the HWD technique can be applied to the column lines sub-dividing them into levels saving power and reducing delay in the SRAM data path. This idea of partitioning bit lines is referred to as Divided Bit Line (DBL) architecture [4], as shown below in Figure.2.3.

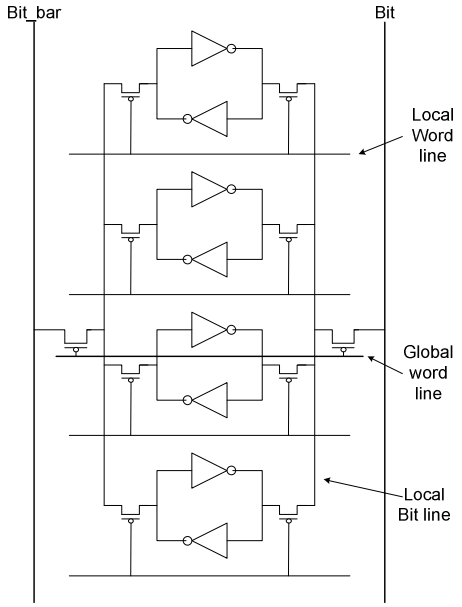



Figure.2.3 Divided Bit Line Architecture [4]

The bit line capacitance is mainly composed of the drain to body capacitance of the pass transistors of the SRAM cell and metal capacitance of bit lines. The bit line capacitance is significantly reduced by reducing the number of pass transistors loading the bit lines. As shown in Figure.2.3, the divided bit line architecture has a global bit line running all the way through the SRAM array columns. It also has several sub bit lines that connect to a small group of SRAM cells vertically. These sub bit lines connect the global bit lines through a pass transistor, thus effectively reducing the number of pass transistors loading the global bit lines. In the case shown above there are 4 SRAM cells combined together and connected via one pass transistor to the global bit line.

In high density memories, the number of sub bit lines will increase and even with divided bit line architecture, the bit line capacitance can be significant. This can be addressed by using Hierarchical Divided Bit line architecture (HDBL) [4]. Again this is a tree type structure and is readily optimized. In this architecture, the bit line is divided into more than two levels. Quite obviously, the number of hierarchies is then determined by the number of rows in the SRAM array.

## 2.4 Folded Bit line Architecture

The folded bit line architecture [5] can be visualized as the structure formed by folding one SRAM array over the other. In this architecture, alternate SRAM cells along the column form a group, as shown in Figure.2.4. It can be seen from the figure that the bit line capacitance is reduced by half effectively. It can be thought of like splitting a bit line into two, and hence reducing the capacitive loading on each by half of the original. There is yet another advantage of this architecture. It is possible to activate two word lines

almost simultaneously since this structure has two separate bit lines. This can increase the data rate almost a factor of two. By using proper word line activation scheme, it is possible to reduce the size of the decoder. For example if we assume each group to have 128 word lines in them, a 7-to-128 decoder would suffice, even though the total effective word lines in the whole array are 256. This can be achieved by introducing an effective delay between activation of word lines in two different groups.

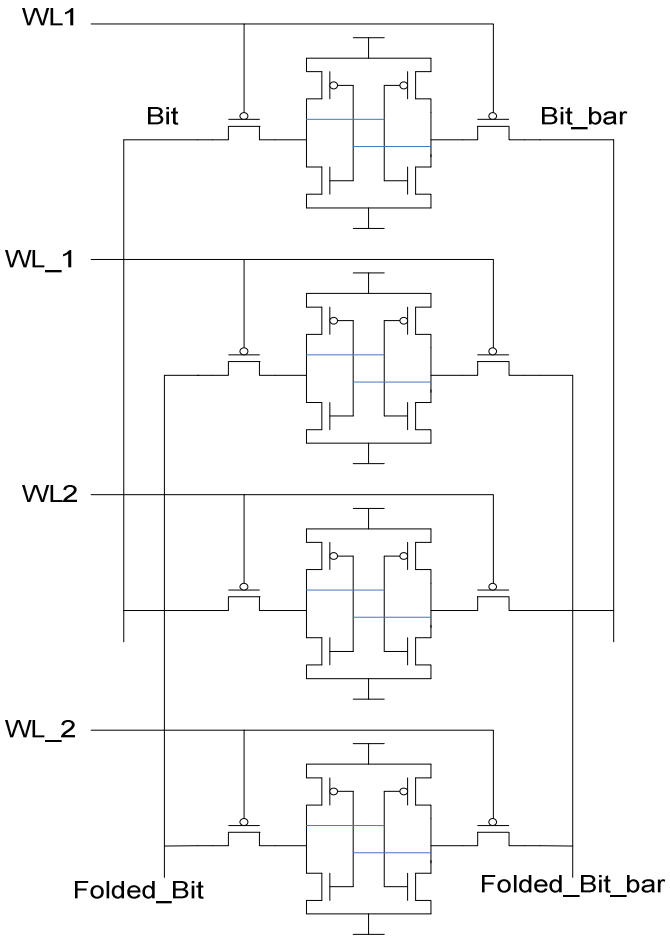



Figure.2.4 Folded Bit line Architecture [5]

## 2.5 Conclusion

The folded bit line architecture and the divided bit line architecture prove useful in conjunction with the divided word line architecture. The folded bit line architecture is

primarily used in DRAM. It requires interleaving of two separate cell arrays and can be complex from the layout perspective. The divided bit line architecture on the other hand has no interleaving, but with more hierarchies, it has more sub bit lines which can again affect the layout issues. Also the use of these methods can add more parasitic capacitance due to additional metal lines in the cell arrays running parallel to one another. They incur a small area penalty in order to compensate for the increase in parasitic capacitances. This increased area penalty, in addition to layout complexity may play a part in choosing these architectures over the standard array. The divided word line architecture on the other hand reduces the word line delay quite efficiently and with the optimal number of word lines in an array, it can generate more acceptable bit line loads, in view of the area considerations and aspect ratio. The trade-off between the performance and area; complexity will decide the right architecture for any design. In this work, the divided word line architecture is chosen considering the area limitations already incurred by high temperature leakage currents.

# CHAPTER III

# OVERVIEW OF SENSE AMPLIFIERS & DECODERS

There is a tradeoff between performance and area in the SRAM cell arrays, in which mostly area (memory cell density) is the crucial factor for designers [6]. Hence the performance is traded for area as far as the cell arrays are concerned. In order to make up for the performance loss caused by cell arrays, the peripheral circuits have to be designed with efficiency and utmost care. The performance of peripheral circuits is primarily dominated by sense amplifier and address decoders. This chapter discusses these two circuits in detail, providing various commonly used implementations. Section 3.1 provides a discussion of commonly used types of Sense Amplifiers, their speed vs. area/power analysis. This is followed by section 3.2 where various address decoding techniques are presented.

## 3.1 Sense Amplifiers

Sense amplifiers are needed in the SRAM data path in order to insure a reliable read process. When reading an SRAM cell, the time required by the cell to charge the bit lines to their final values can be quite large. Also rail to rail signal swings are not desired in most cases in order to reduce read delay times. Hence some form of amplification and

data latching are required to ensure speed-up of the read process in SRAM. This is achieved by using the sense amplifiers. Sense amplifiers detect small voltage variations in the bit lines caused by the SRAM cell and present the data to the data bus.

There are certain key performance objectives to be considered while designing the sense amplifiers, like high sensitivity, high speed of operation, low power, low offset, and finally low area. It is hard to achieve all of these performance objectives simultaneously, and requires extensive design procedure. The choice of sense amplifier for a design depends on how well it meets all the above listed objectives. A number of sense amplifiers have been proposed in the literature. The following sections describe the functioning and performance of some of the most commonly used sense amplifiers, viz. Conventional Latch-type Sense amplifier, Linear Differential Sense Amplifier, Current Sense Amplifiers, and Voltage Latch-type Sense Amplifier. Each of these sense amplifiers are evaluated for the performance objectives listed before and finally the chosen architecture is justified.

### 3.1.1 Conventional Latch-type Sense Amplifier

A simple cross coupled pair of inverters can function as a sense amplifier. An inverter has high gain in its transition region. This high gain can be used in conjunction with a positive feedback as in a latch to sense the bit lines. This idea is used in the latch-type sense amplifier as shown below in figure 3.1, where the cross coupled inverters function as a latch. In this sense amplifier, there is no isolation between the inputs and outputs which are cross-coupled and once the amplifier senses the bit, both the bit lines are eventually driven to full logic levels. The speed with which the bit lines are driven to full

logic levels is much faster than being driven by the memory cell, thereby providing a sense-amplifier action.

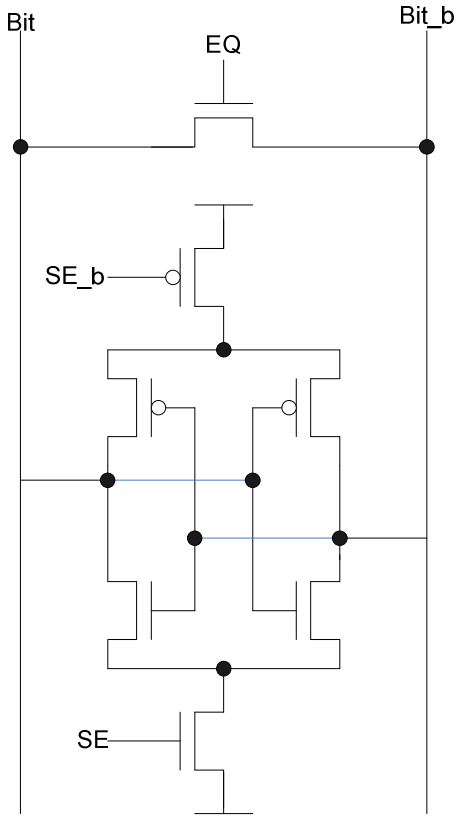



Figure.3.1 Inverter Latch Sense Amplifier

In a typical read cycle, the latch is initialized to its high gain meta-stable state by equalizing the bit lines using signal EQ and pre-charging them in most cases to mid-rail. When the SRAM location to be read is activated, the bit stored in the SRAM cell starts driving appropriate voltage difference across the bit lines. After sufficient voltage difference is developed across the bit lines, the sense enable signal SE is raised to enable the sense amplifier. The latch inside the sense amplifier then settles the bit lines to their final values thus enabling logic detection. The transition to final logic voltages by the bit lines is quick owing to the positive feedback of the latch. The time delay equation of the latch type sense amplifier can be expressed in general as below in equation 3.1. In this

equation, the $C_{BL}$ represents total capacitance at the bit lines and Gm refers to the transconductance of a single inverter in the sense amplifier. $\Delta V_0$ is the initial voltage difference across the bit lines just when the sense amplifier starts up. $\Delta V_{logic}$ is the final voltage difference in logic levels of the bit lines.

$$Td \quad \alpha \quad (C_{BL}/G_m) \times \ln(\Delta V_{logic}/\Delta V_0) \tag{3.1}$$

It can be seen from the equation 3.1 that the delay of the latch type sense amplifier is directly proportional to the total capacitance at the bit lines making it less attractive for higher bit line loads. The delay equation shown above does not include the time taken to charge the bitlines to initial value $\Delta V_0$ by the SRAM cell, and if the sense amplifier is enabled before this time, there is a risk of bit being detected incorrectly. The latch type sense amplifier is not the fastest, but given its small area advantage, it performs well enough to be used in low and medium speed SRAMs. But the offset in the switching threshold of the cross-coupled inverters considerably impacts the sensitivity. Also since the bit lines must swing full logic values during the sensing action, the latch-type sense amplifier suffers from considerable dynamic power dissipation. Hence low power SRAM designs typically use sense amplifiers which limit the bit line swings.

### 3.1.2 Linear Differential Sense Amplifier

A differential amplifier takes in a small signal voltage difference across its inputs and provides an amplified signal at its output. A differential amplifier has high voltage gain, and good common mode noise rejection which is desirable in an SRAM design for sensing the bit lines which are differential in nature. Such a differential sense amplifier is linear in a sense that it simply provides an amplified version of input without driving the

outputs to full logic levels. Since in a linear amplifier, voltage gain and speed are inversely related, increasing the gain for a particular design results in relatively slowing down the read process. To overcome this issue, sense amplifier designs using linear amplifiers generate the full logic outputs by incorporating multiple stages [9]. A simple differential amplifier with an inverter as its output stage is shown in figure 3.2.

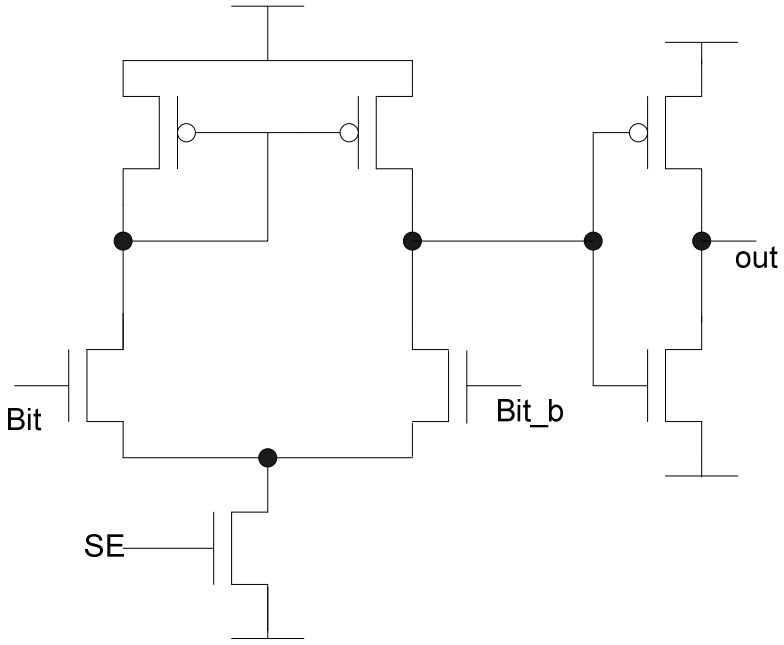



Figure.3.2 Linear Differential Sense Amplifier [3]

The differential amplifier has good common mode rejection ratio (CMRR) as well as power supply rejection ratio (PSRR). CMRR is desirable in an SRAM because of the fact that bit lines of an SRAM inject common mode noise (via the row select lines) which differs across process. It is beneficial to have the bit lines decoupled from the outputs of the sense amplifier and reduce the voltage swings in the bit lines, thereby dissipating less dynamic power. But in most cases the linear differential amplifier requires a DC bias which causes some static power dissipation. If we assume the output impedance of the first stage amplifier as $1/G_o$ , where Go = Gdsn+ Gdsp, and its total output capacitance as

$C_L$ (including the input capacitance of second stage), the delay contributed by the linear amplifier follows the relation 3.2. The total sense amplifier delay is the sum of delays of linear differential amplifier and the succeeding digital logic stages (significantly less) as given by equation 3.3.

$$Td_{amp} \quad \alpha \quad (C_L/G_O) \tag{3.2}$$

$$Td_{SA} \quad = \quad Td_{amp} + Td_{Dig\_stages} \tag{3.3}$$

This type of sense amplifiers has good speed, good sensitivity and usually a low offset which comes at the expense of area. Since they consume biasing power and operate over limited supply voltages they are less preferred for low power and low voltage designs and are primarily used in high performance designs [3].

### 3.1.3 Current Sense Amplifiers

The bit-line capacitances will limit the speed of any sensing scheme that requires a voltage difference to be introduced on the bit lines to initiate sense amplifier operation. In the case of current mode sense amplifiers, the sensing nodes offer low impedance to the bit lines, and respond to current signals rather than voltage differences across the bit lines. There are several current mode sense amplifiers being used in SRAMs, one of the popular ones being the Clamped Bit Line Sense Amplifier (CBLSA) [7]. The clamped bit line sense amplifier is shown in figure 3.3.

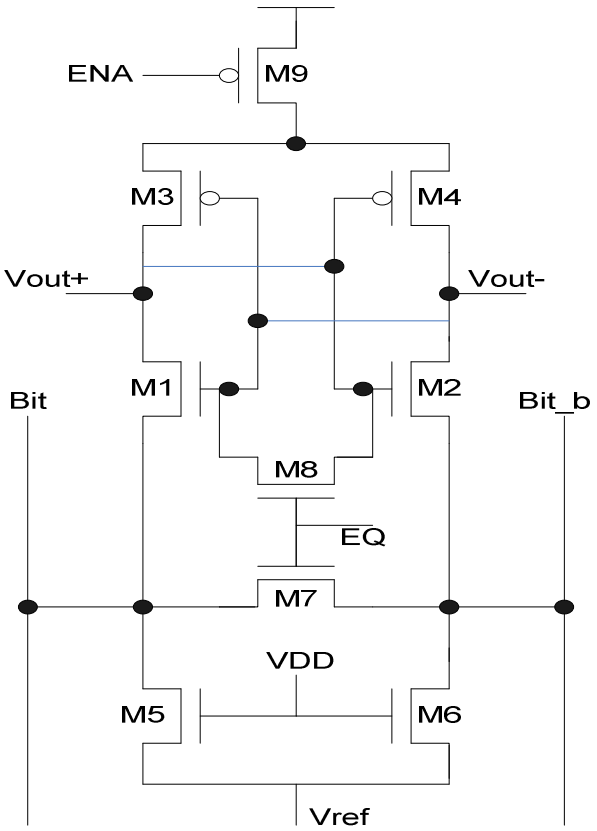



Figure.3.3 Clamped Bit Line Sense Amplifier (CBLSA) [7]

The CBLSA has transistors M1-M4 forming conventional CMOS cross coupled inverter latch. Transistors M5 and M6 provide a low impedance path for the bit lines to the reference potential Vref. M5 and M6 are biased in linear region and the reference potential Vref can be as low as 0V. It is this clamping action of transistors M5 and M6 that gives rise to the name Clamped Bit line Sense Amplifier. The sense amplifier is powered using large transistor M9. During the precharge phase, both the outputs of the sense amplifier are equalized using the transistor M8 and signal EQ. Also the bit lines are equalized using the transistor M7. At the end of precharge phase, transistors M7 and M8 are switched off and SRAM cell is opened to the bit lines. This bit line current difference $\Delta I$ flows through M5 and M6, which is sourced by M1 and M2 respectively. Transistors M1-M4 then latch the data due to high gain positive feedback.

The delay equation governing the CBLSA can be approximated as shown in equation 3.4. It takes the same form as in the latch-type sense amplifier, except that the bit line capacitance is replaced by $C_L$, the internal capacitance of the cross-coupled inverter pair, which is significantly less. $\Delta V_0$ is initial voltage difference caused by the difference in current flowing through the and can be written as $( \Delta I_{Sense} - I_{OS} ) \times G_m$, where $I_{OS}$ is the offset current of the sense amplifier, and $G_m$ is the transconductance of the inverter pair.

$$(\Delta I_{Sense} - I_{OS}) \times G_m = \Delta V_0$$

$$Td \quad \alpha \quad (C_L/G_m) \times \ln(\Delta V_{logic}/\Delta V_0) \tag{3.4}$$

In this sense amplifier the bit line capacitance is taken to a node within the amplifier such that it has minimal effect on the performance of the circuit. The signal current from the memory cell can now be injected directly into the sense amplifier due to the fact that sensing node presents low impedance to the memory cell while reading and it has to be ensured that this signal current is greater than offset current $I_{OS}$. As a result the sense amplifier has minimal dependency on the bit line capacitance, and is only sensitive to the difference in current flowing through the bit lines. If the memory cells are designed to produce at least the minimum required current difference through the bit lines, the current sense amplifiers operate almost independent of bit line capacitances. Since the sense amplifier's operation does not require the bit line capacitances to swing full logic values, the power consumption is reduced significantly during the read process.

Current-sense amplifiers in general provide increased speed and low power dissipation at moderate area overhead. They also provide fairly constant delays over a range of bit line capacitances, making them easier to reuse for different size memory arrays. But their use is justified primarily in high performance memory systems owing to their complexity.

### 3.1.4 Voltage Latch-type Sense Amplifier

The Voltage Latch-type sense amplifier [12] is similar to the Inverter latch sense amplifier presented in section 3.1.1, except that the bit lines are separated from the outputs of the sense amplifier. The voltage latch-type sense amplifier or the Alpha latch [10] looks as shown below in figure 3.4. It can be seen from the figure that the bit line inputs to the sense amplifier are terminated at the gates of the transistors M5 and M6, thus isolating them from sense amplifier's outputs. The functioning of the amplifier is similar to the Inverter Latch sense amplifier. A sufficient voltage is allowed to develop across the bit line inputs before the Sense Amplifier is enabled using the signal SE. The transistors M5 and M6 then act as transconductors and the differential current flowing through them enables the cross coupled inverters to latch the data. Outputs are obtained from the inverters placed at the cross-coupled nodes of the latch.

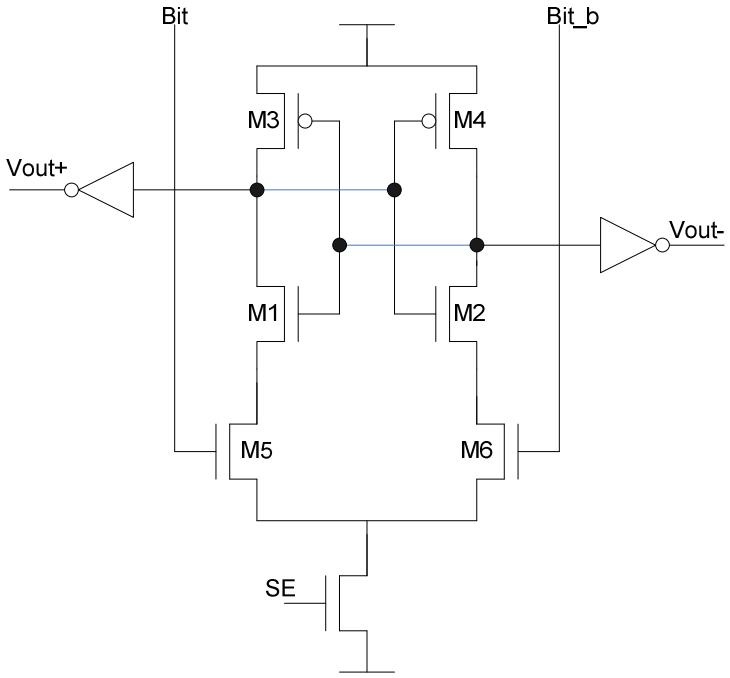



Figure.3.4 Voltage Latch-type Sense Amplifier [12]

The advantage of this circuit over the Inverter Latch sense amplifier is that the isolation of inputs and outputs will reduce the signal swings on the bit lines and also sense amplifier's delay becomes less dependent on the bit line capacitance. However the addition of an input pair in series with the inverter latch slows down the circuit a little bit unless they are made very large relative to the cross coupled devices. The equation governing the delay of the sense amplifier can be written as the sum of two components, as shown in equation 3.5.

$$t_d = t_0 + t_{latch} \tag{3.5}$$

In the above equation, $t_0$ refers to the time taken to start the regeneration by the cross-coupled inverter latch, which can be written as below in equation 3.6, where $C_L$ is the load capacitance of the cross-coupled inverter pair and I is the current through the input transistors when the sense amplifier is enabled. The second term in equation 3.5 [8] is the latching delay which can be written as in equation 3.7 just as in the conventional latch-type amplifier.

$$t_0 = C_L \Delta v_{tp} / I \tag{3.6}$$

$$t_{latch} = (C_L/G_m) \times \ln(\Delta V_{logic}/\Delta V_0) \tag{3.7}$$

The linear amplifier type sense amplifiers (differential amplifiers in the first stage) and current mode sense amplifiers are used typically in high performance systems. When acceptable speeds are needed at lesser complexity and power, latch type sense amplifiers are preferred. A comparison of sense amplifiers discussed so far are shown in table 3.1. Table 3.1 shows the general expressions for delay of each of the above discussed sense amplifier, while table 3.2 gives a general perspective into the performance metrics of the sense amplifiers.

Table 3.1 General delay expressions for various sense amplifiers

| **Sense Amplifier** | **Delay Expression** |
|---|---|
| Inverter Latch SA | $(C_{BL}/G_m) \times \ln(\Delta V_{logic}/\Delta V_0)$ |
| Linear Differential SA | $\alpha\,(C_L/G_O) + Td_{Dig\_stages}$ , where<br>$G_O = G_{dsn}+G_{dsp}$ |
| Current Sense Amplifier | $(C_L/G_m) \times \ln(\Delta V_{logic}/\Delta V_0)$ ,where<br>$\Delta V_0 = (\Delta I_{Sense} - I_{OS}) \times G_m$ |
| Voltage Latch-type SA | $C_L \Delta v_{tp} / I + (C_L/G_m) \times \ln(\Delta V_{logic}/\Delta V_0)$ |

Table 3.2 General Comparison of performance between various sense amplifiers

| **Sense Amplifier** | **Speed** | **Vos/Ios** | **Power** | **Complexity** |
|---|---|---|---|---|
| Inverter Latch SA Big $C_{BL}$ | fastest to slowest | lowest – 2 pairs | Moderate/High | Easier Analysis |
| Linear Differential Sense Amp | Moderate to slowest | Lowest– 2 pairs | Moderate/High | Straight forward Analysis |
| Current Sense Amplifier | fastest | Highest– 3 pairs | Low/Moderate | Somewhat more Analysis |
| Voltage Latch-type SA | Moderate to fastest | lowest/Highest – 2 or 3 pairs | Low/Moderate | Straight forward Analysis |

In this work, the speed requirement for sense amplifier is modest and hence the current mode sense amplifiers or linear amplifiers were not chosen. The alpha latch was preferred over the conventional Inverter latch sense amplifier due to its input/output isolation. But as mentioned just before, the addition of input pair in series with inverter latch makes the circuit little slower than the Inverter Latch sense amplifier. In this work, a variant of Alpha Latch is used where the cross coupled inverter pair is replaced by a cross coupled PMOS pair to compensate for added input transistors.

The circuit schematic for the modified alpha latch sense amplifier also known as the Cross Coupled Sense Amplifier used in this work is shown below in figure 3.5. The PMOS transistors M3 and M4 can be sized appropriately to achieve enough positive feedback gain. Both the output nodes Vout+ and Vout- are charged up to VDD when SE signal is not active. This will reset the latch every time a new read operation is initiated. The outputs of the sense amplifier can be applied to a D-latch to obtain full logic outputs.

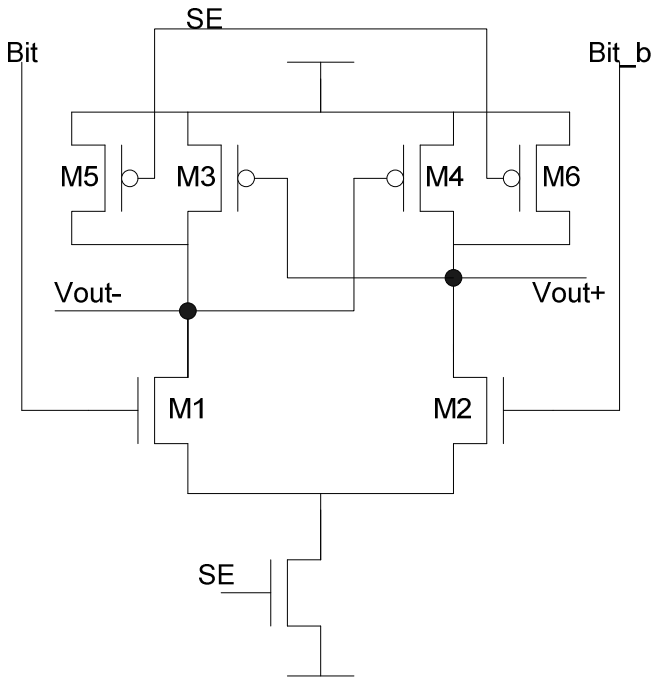



Figure.3.5 Cross Coupled Sense Amplifier

## 3.2 Address Decoders

Address decoders constitute the major portion of word line delay in selecting the appropriate memory word. In any design fast low power decoders are desired, such that the total memory access times are optimized with consideration to decoder energy expenditure. One other primary concern while designing decoders is the area spent. For larger memories, the decoder area becomes significant, next only to SRAM cell array. Hence smaller decoders are preferable keeping in mind that the decoders be pitch-matched to the cell array so that additional wiring area can be minimized. Pitch-matching not only helps reduce the area, but also prevents wiring delay from rising up.

There are two main categories of decoders that are commonly used; CMOS decoders, and Tree decoders. Conceptually in a CMOS decoder, an $n:2^n$ decoder consists of $2^n$ n-input AND gates, one for each row of memory. The AND gates are made out of NAND or NOR gates. But for n larger than 4, the decoder becomes slower and hence there is a need to use either multiple fewer input gates to make the larger input gates or break the decoder into two or more levels. For example, a 4-input AND gate can be generated from three 2-input NAND gates. But this method results in tremendous increase in number of gates used, as the decoder's size becomes larger and also power & area becomes a concern. Often, predecoders [11] are used in the first level followed by $2^n$ final AND gates. The primary advantage of predecoding is that it does not change the path effort, but at the same time takes up lesser number of gates thereby reducing the area and power dissipation. A simple way to illustrate this issue is shown in figure 3.6a and 3.6b.

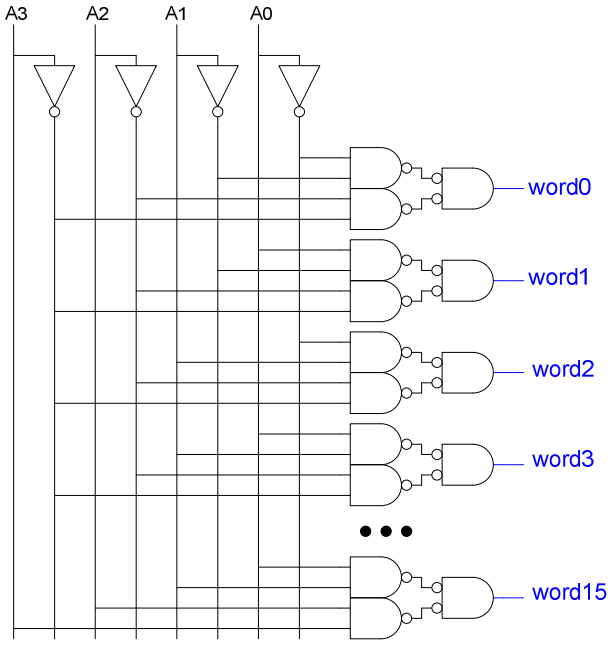



Figure.3.6a A simple 4-16 Decoder [11]

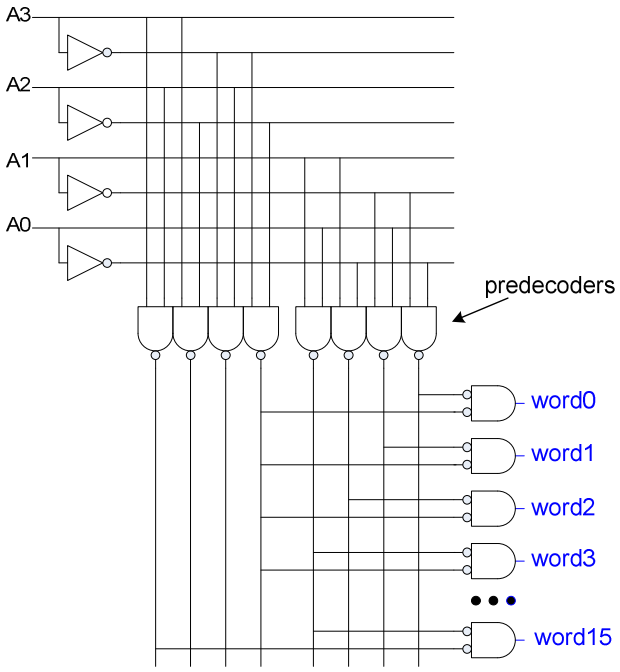



Figure.3.6b 4-16 Decoder with two 2-4 predecoders [11]

It can be analyzed from the figure that the number of 2-input NAND gates used in the ordinary decoder circuit is 32 while the predecoded circuit uses only 8. This considerably reduces area and power. But as the decoder size increases, there needs to be more buffering added internally to reduce the fan-out of gates. In spite of this additional buffering, the use of pre-decoder proves beneficial over normal CMOS decoders.

There is another prominent type of decoder used in memories, known as the Tree decoders. The area occupied by CMOS decoders is considerably high, even after using the predecoders. To reduce the layout area further, a pass-transistor based decoder structure can be used as shown in the figure 3.7.

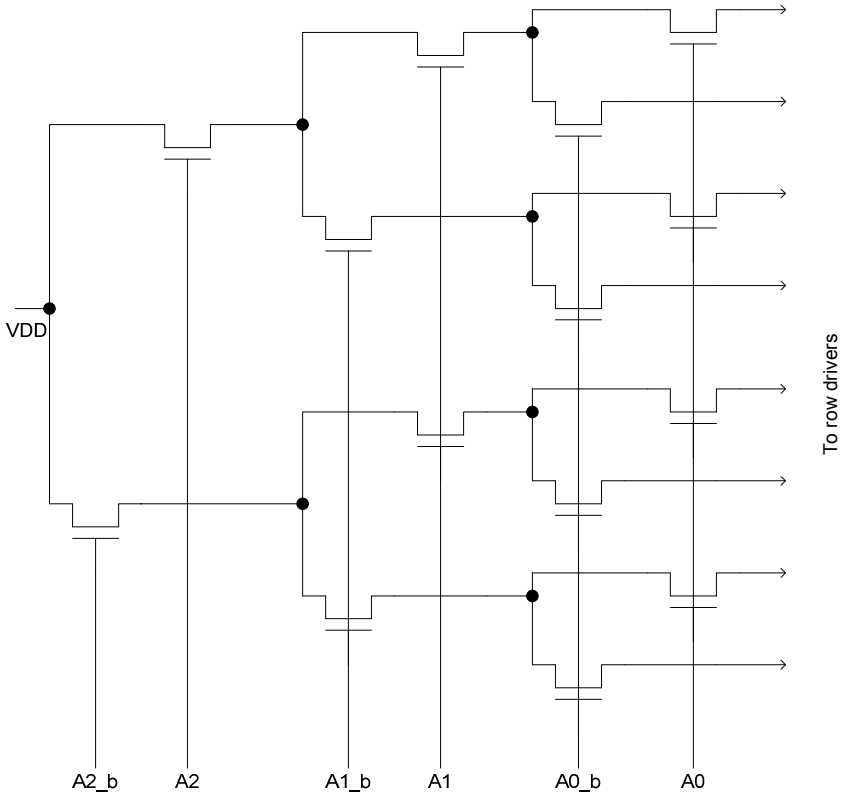



Figure.3.7 A 3-to-8 Tree Decoder [13]

It can be seen from the figure above that, the 3-to-8 tree decoder implementation results in use of minimum number of transistors. But the structure shown above suffers

from two problems. The non-selected outputs float and NMOS pass-transistors do not pass valid high efficiently, the effect of which can be mitigated by using buffers every 3 or 4 gates resulting in an optimal delay tree structure. Both these issues can also be addressed by using a simple circuit that pulls the outputs of the decoder low, when they are not selected, as well as using a skewed inverter to maximize the noise margins. Tree decoder structures are predominantly used in n-to-1 column multiplexers. This reduces the area of column multiplexer considerably.

# CHAPTER IV

# METHODOLOGY FOR REDUCED TIME SRAM SYNTHESIS

Designing an SRAM includes some critical tasks like SRAM cell design & stability analysis; simulation of word line activation delay, data path delay; control signals generation & timing analysis. Most of these tasks are done manually since automatic synthesis, place & route algorithms cannot be directly applied to memory design. When it comes to laying out the memory, the task becomes even more difficult and each sub-module historically has to be laid out by hand. This chapter proposes a methodology adopted in this work to reduce the design time and layout effort by partitioning the memory into several critical modules and setting up a test bench for characterization of each critical module. For each of these modules, repeatability in layout is also analyzed and this factor is tapped to simply array them out as required.

There are two perspectives to be considered for finding out the critical modules: simulation (for functionality and timing characterization), and layout. The individual modules should facilitate timing simulations, and also be repeatable in layout. Once the critical modules are found out, it would be a lot easier for the designer if all the test benches required to characterize these modules are available at hand and that the designer only needs to replace modules for each design to meet the requirements. This idea is

utilized here in this work, and also the layouts for most modules are done by hand, and depending on the size of the memory, they are simply arrayed out as required. The critical modules identified in this work are listed below:

- SRAM cell 2×2 Block
- Sense Amplifier & Column Read/Write Logic
- Bank Control circuitry & Row Drivers
- SRAM Bank
- Global Control circuitry
- Global Address Decoders

The following sections discuss these modules in detail, with a concluding section summarizing the methodology

## 4.1 SRAM cell 2×2 Block

The very basic module involved in a Static RAM is the SRAM cell which forms the most part of the layout. It has to be designed such that a read process does not flip the bit stored inside the cell, while the write operation reliably writes a value into the cell. Also the SRAM core consists of identical SRAM cells arrayed out in a regular fashion according to the memory size. Hence in this work due to the layout considerations, the SRAM cell module was chosen to be of size 2×2, i.e. a total of 4 cells, 2 cells each along the row as well as the column.

A test bench was developed to characterize the SRAM cell module for its functionality as well as the read/write delays. To emulate the real bit line loads, a generic capacitance was used on the bit lines, which requires the user input of number of words(W) and bit

line capacitance per unit cell($C_{unit}$). The testbench then computes the total capacitance on the bit lines and uses it in the simulations. The testbench uses initial conditions on the internal nodes of the SRAM cells so that when the write operation begins, it has to overwrite the SRAM cell, ensuring the functionality of write process. Also during the read process, the bit lines are precharged to a user specified value Vpre. The user also needs to input the clock frequency F_CLK. Figure 4.1 shows the SRAM cell module used in the testbench for characterizing the performance and functionality of 6-T SRAM cell.

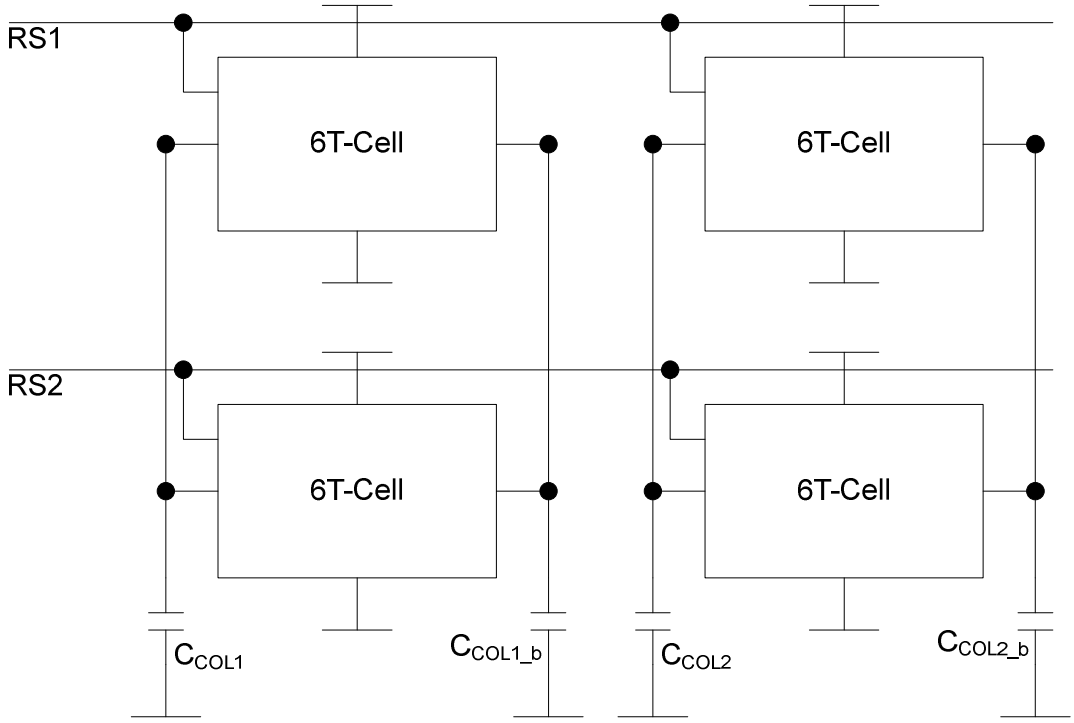



Figure.4.1 SRAM 2×2 Cell Module used in Test Bench

It can be seen from the figure that the bit lines have been loaded with capacitors. As mentioned before the value of this capacitance is calculated during the simulation from the inputs given by the user. The testbench has output expressions to calculate the write time, read time, voltage fluctuations in the SRAM internal nodes named Q & Qb. These expressions present the user with various timing parameter values without having to

manually measure the delays on the waveform window. When it comes to the layout, the SRAM cell is flipped along vertical as well as horizontal directions and abutted to form a compact and efficient module, which can then be arrayed easily.

## 4.2 Sense Amplifier & Column Read/Write Logic

The sense amplifier along with its associated read logic is the most critical part of the SRAM peripheral circuitry. The tradeoff in performance for area of the SRAM cell in the core has to be regained by the efficient design of read circuitry. Sense amplifier stands at the top of the read circuitry in terms of contributing to the speed of read process. Hence it is necessary to design and characterize the sense amplifier to get the required performance across all worst case design corners.

Even though the write process is less complex than the read process, the write logic has to be validated to ensure reliability across corners along with timing characterization. Hence in this work the sense amplifier and read/write logic have been grouped as a one module forming the column logic. The block diagram of the column logic module along with its sense amplifier is shown below in Figure. 4.2. As shown in the figure the column logic consists of read and write data paths. The read data path includes a cascade of sense amplifier, D-Latch and a 3-state Buffer. The inputs to the read path, Bit & Bit_b are received from the SRAM cell array column.

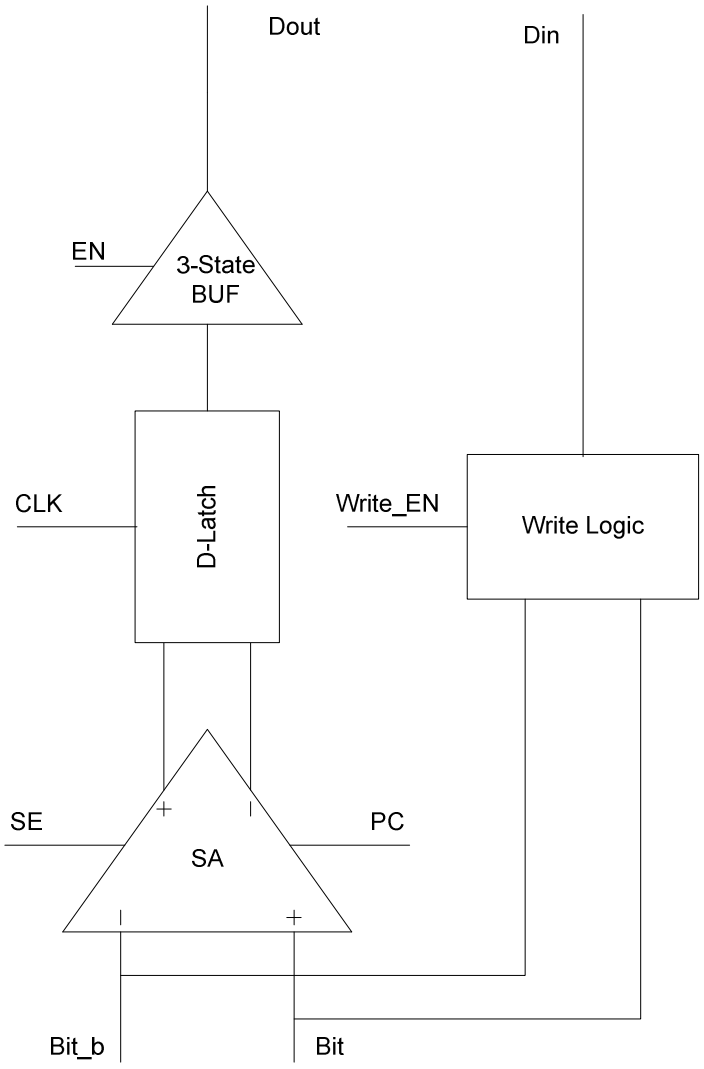



Figure.4.2 Sense-Amp & Column Logic Module

The write logic gets the data input and when enabled, it writes the appropriate values onto the Bit and Bit_b lines which in turn get written into the selected SRAM cell. This column logic structure is repeated across each column of the SRAM cell array and hence gains significance in layout.

Now that the column logic is defined, it can be easily characterized by a test bench for read and write access times along with variation across process corners and different loading conditions of the data bus and bit line capacitances. A comprehensive test bench

is developed which includes two column logic modules and parameterized capacitors being used at each block's outputs to account for variation across different loads capacitances. The testbench taps the outputs of each block in the column logic module thus enabling the characterization individually. This is primarily helpful in testing the sense amplifier.

The testbench for column logic gets the inputs COL and COL_b from two sources, viz. the SRAM cell and external voltage sources. The purpose of providing SRAM cell is to trigger the Column logic's read path involving sense amplifier so as to mimic the real behavior of SRAM read process. But when it is essential to characterize across different sensitizing voltage inputs, an external voltage source is used thereby making testbench more generic. For the better understanding of the reader the testbench structure is shown in figure 4.3. As shown in the figure, the testbench has two identical column logic modules, and when row select RS is selected, the SRAM cell drives the bit lines for a read process, otherwise, external voltages are used for read time characterization.

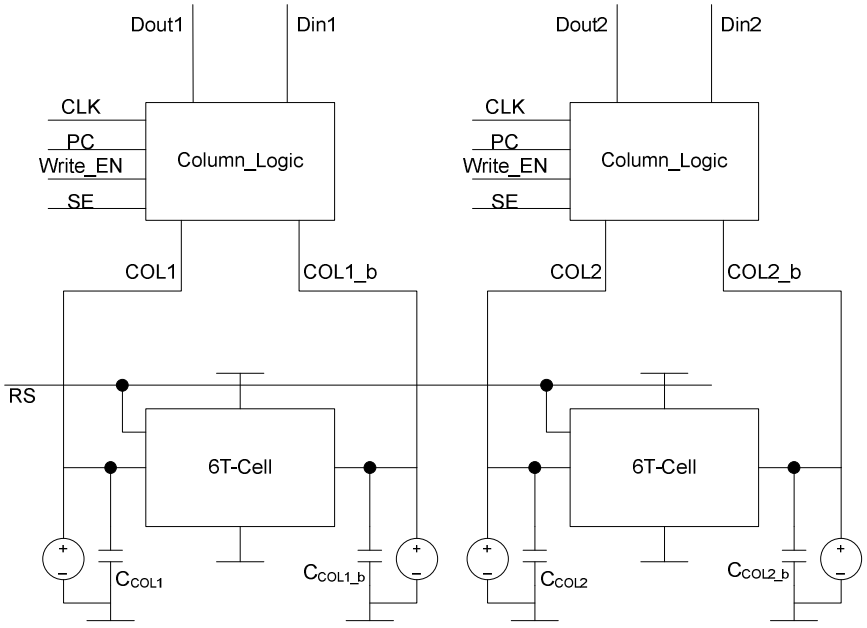



Figure.4.3 Column Logic Module Testbench

## 4.3 Bank Control circuitry & Row Drivers

SRAM arrays are commonly divided into several sub-arrays called banks as discussed in chapter 2. Each of these banks uses several common control signals for handling read/write process. But at a given read/write cycle only one bank is enabled while the rest are idle, thereby reducing the switching power dissipation. This is usually achieved with the help of unique bank select signals for each bank generated from column decoder. Additionally, the row drivers needed for each SRAM cell array bank is also gated using these bank select signals.

The bank control circuitry is a regular repeatable structure used for each bank and its characterization is essential to validate the timing of the control signals against the data path. Since the bank control circuits are entirely built from standard digital gates, this process can be automated completely using a standard cell library and with the help of synthesis & place/route tools.

The row drivers for each SRAM bank, being a part of bank control circuits can be synthesized for appropriate capacitive loading provided by each bank's SRAM cell array. Since the row drivers are identical for all the row select signals, the layout of these drivers can be done by simply arraying them out by hand. But it must be remembered that the row drivers need to be pitch matched in layout area to the SRAM cell in order to achieve efficient area utilization.

For simplicity, it is easier to use a pitch-matched standard cell library for both the row drivers and other bank control circuits. Basic 2 input and 3 input logic gates along with tapered inverters or buffers would suffice for the standard cell library. Once all of the required control signals for the SRAM bank are synthesized, the module is then validated

across various temperature and process corners, in conjunction with consideration given to variations in load capacitances. This is achieved using a testbench developed in Cadence Analog Design Environment for more realistic timing analysis.

## 4.4 SRAM Bank

An SRAM bank can be considered as a miniature version of the whole memory in a sense that it has all basic circuits needed to read and store data, including the control signals. It is common for an SRAM bank to have an array of SRAM cells, associated row select signals, column logic circuits for each bit column, and required control signals for the column logic. The structure of an SRAM bank used in this work is shown in figure.4.4. Once the functionality of this regular, repeatable structure is ensured, it then just becomes a job of arraying the banks in a regular fashion to achieve the required size of memory. Thus it is essential to characterize an SRAM bank module precisely to predict overall performance of the SRAM.

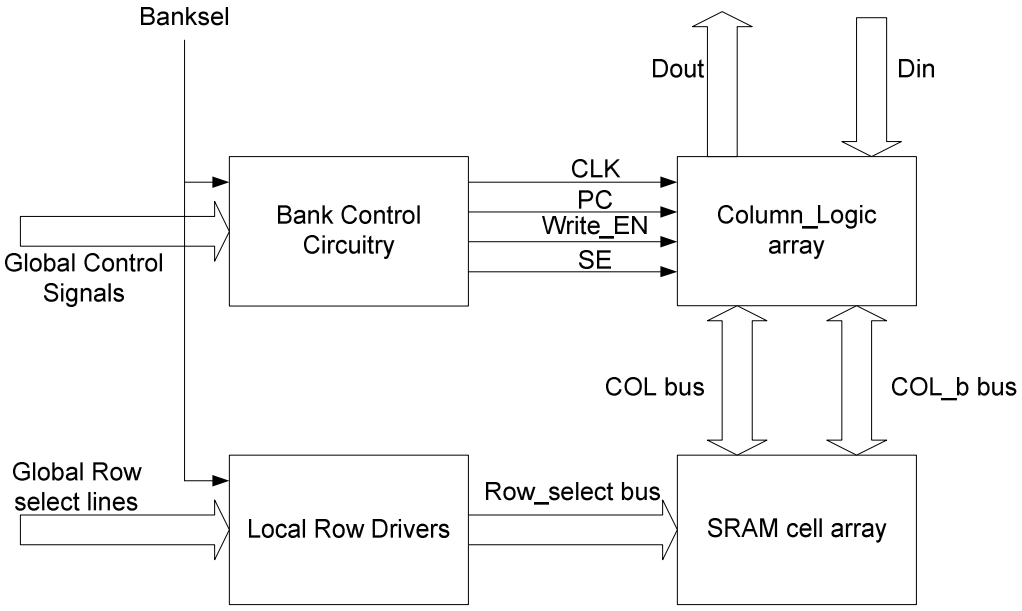



Figure.4.4 SRAM Bank

As shown in figure.4.4, the SRAM bank is essentially formed from other critical modules discussed in sections 4.1 to 4.3. The column logic array shown in the figure corresponds to a group of identical column logic blocks, the total number being set by the number of bits in the cell array. Since all of these modules will have to be characterized prior to forming a bank type structure, a firsthand performance of the SRAM bank can be predicted easily from their timing information. But the timing relations of the controlling signals are very crucial to maintain the functionality of the read/write cycles. Hence a comprehensive testing procedure was developed to ensure the reliability of operation and to measure the read and write access times across various corners with due consideration for variation of capacitive loads at data bus and internal bit lines.

Let us now consider the task of laying out an SRAM bank. Since the SRAM array is just a regular structure, it can be replicated in layout with ease, either manually or using automation tools. Similarly automatic layout tools can be applied for bank control circuitry, while row drivers can simply be arrayed out by hand since they are pitch-matched with the cell array. But the layout of Column logic decides the additional area overhead and in most cases it has to be done manually to get efficiency in area. Since the column logic includes sense amplifier which is a quasi-analog circuit, it has to be laid out by hand. But once the column logic is laid out completely with corresponding routing lines, it then becomes a single block which can be parameterized and treated as a standard cell for place and route. All of these sub-blocks constituting the bank can now be grouped together in layout to form an SRAM bank.

## 4.5 Global Control Circuitry

We have discussed earlier about the control signals utilized in a single bank. But these bank control signals are actually the gated & buffered version of global control signals running across all the banks. The global control signals are generated from input clocks and read/write signal provided to the memory. Global control circuitry is responsible for controlling the read/write or idle state of the whole memory block. It is a complete digital circuitry and hence it can be synthesized automatically and placed & routed as a block.

The design of global control circuitry is straightforward and can be accomplished either manually or using synthesis tools. One important issue to be taken care of is the buffering required to drive all the banks of the memory. Since the final place & route of the whole memory is going to be done with the help of individual critical modules, buffering the control signals automatically is not straightforward. It is hence easier to buffer the control signals manually after an estimation of the number of banks that are to be driven, with due consideration given to parasitic capacitances arising out of long routing metal lines.

The characterization of global control circuitry is best done using a testbench that automatically measures the various timing parameters of control signals. Hence a testbench was developed that allows the designer to measure pulse widths, delays, rise/fall times of all output signals with ease and to characterize their variations across different loads and corners.

## 4.6 Global Address Decoders

Address decoders control the accessing time of memory locations and set a limit on minimum overall read/write access times. Global address decoders are prioritized in importance along with sense amplifiers because of their role in providing reliable access to all the memory locations. Hence it is essential to efficiently design and accurately characterize the performance of address decoders. Address decoders can be classified considering to the dimensions of SRAM cell array into Row Address decoders and Column Address Decoders. Row address decoders send the decoded row select lines across the whole SRAM cell array, which are then decoded locally at each bank using a column decoder to enable the cells located in a particular bank.

The design of address decoders can be done using HDL, and then synthesized using a standard cell library, thereby saving considerable design time. Also, there are many architectures available for the designer to optimize area, speed and power, requiring meticulous design process which is simplified via the HDL synthesis process. But the task becomes even more difficult in laying out the decoders because, the output stage must be pitch-matched to SRAM cell to minimize routing problems and maximize efficient area usage. Hence a standard cell library with pitch-matched layouts is preferable for the synthesis and place & route using automatic tools. Use of standard cells for the design also allows coarse estimation of the delays of critical paths of the decoder. Nevertheless, precise decoder delays are required to estimate the maximum speed of the memory, thereby mandating the need for a testbench which can characterize the decoder across variations in loads and process corners.

Functionality of the decoder can be verified and critical paths can be identified easily using HDL simulation tools. This way the simulation and optimization times of device level simulations can be reduced drastically, especially when the decoder input size becomes larger. Once the critical paths are identified, their delays can be more realistically characterized using a custom-built testbench in a device level simulator such as spectre. Hence in this work a testbench was developed in cadence to characterize the delays of identified critical paths across a wide range of temperatures and process corners. The delays are also measured across variations in load capacitances of each output.

The final task in the Global Address Decoders module is that of laying them out pitch-matched to the SRAM cell array. It is necessary for the decoder output buffer stage to be pitch-matched to the cell array, but the internal decoder logic leading to the output lines may be laid without stringent area rules. Even though the layout of decoders can be done automatically using place & route tools thereby saving time in complex routing, to achieve the best of area efficiency, the layout has to be meticulously done manually.

## 4.7 Conclusion

This section summarizes the methodology employed in this work to reduce the design time of an SRAM. The idea is to efficiently partition the SRAM into several critical modules identified based on functionality and repeatability in layout; then develop generic comprehensive testbenches; use the test benches to design the critical modules until desired performance specifications are met; then lay them out in a way easier to array them out if required. Adopting this methodology, six most critical modules have

been identified in this work, viz. SRAM cell, Sense Amp & Column Logic, Bank Control Circuitry, SRAM bank, Global Control Circuitry, and Global Address Decoders. After careful consideration to all the necessary specifications, the testbenches needed to characterize these critical modules have been developed. It is with the help these testbenches and pitched-matched cell libraries that the design time of the critical modules will be reduced. This is because whenever a change in the existing design is desired or a new design is developed, it is beneficial to use the generic testbenches with minimal or no changes in these testbenches. Once the design is validated across all possible PVT corners, the layout of the whole memory can be done with relatively lesser effort by using the individual critical module layouts to array them out wherever required. Once the critical module designs and their base testbenches are formed, design of larger size memories or their variants can be accomplished with lesser effort, and reduced time.

# CHAPTER V

# HIGH TEMPERATURE SRAM DESIGN EXAMPLE

It is expected of an SRAM to have high reliability over its operating range and desired to have lesser power consumption with minimum possible area overhead. SRAM is primarily used for on-chip memory access needed for a microprocessor or a microcontroller. Memory systems account for about 50% of die area in modern day microprocessors. Hence it becomes necessary to limit the area occupied by memory on a die, simultaneously providing high reliability across its specified operating range. In addition to these common factors, the style of SRAM design is governed by the application requirements and process limitations. All these factors make the design of an SRAM a true custom design specific to a process and application.

Previous chapters discussed the various architectures of SRAM and the circuit techniques for its constituting modules. The methodology of design time reduction has been introduced in chapter IV. This chapter builds on the concepts of previous chapters and presents the design of a high temperature standard 6-T cell based SRAM. The design of SRAM in this work is driven by the requirements of a high temperature microcontroller (OSU-HC11) which needs a 32k-bit on-chip RAM and 32k-bit off-chip

RAM being accessed with the help of an SPI controller, each of 8-bit wide data bus. The read and write timing diagrams of On-chip OSU-HC11 SRAM is shown in figure 5.1.

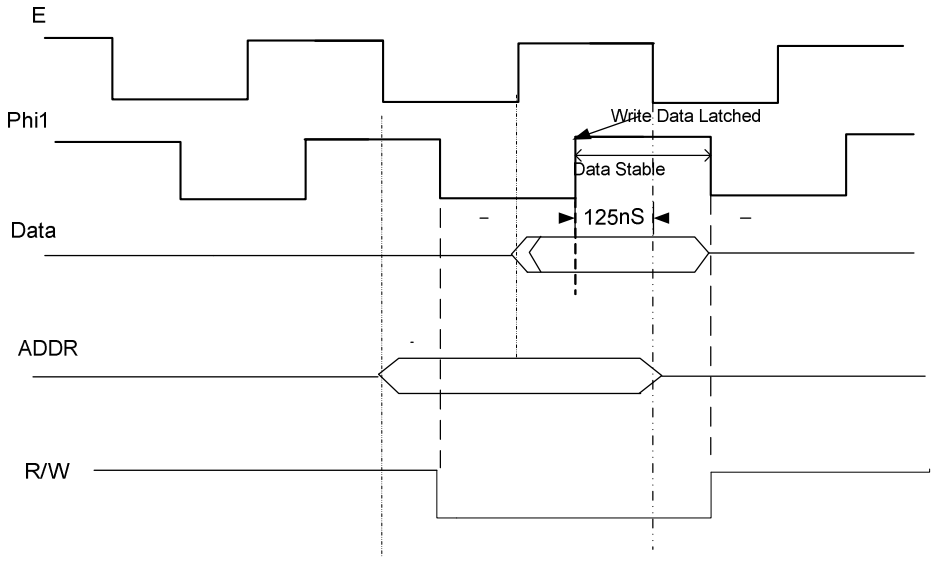



Figure.5.1 (a) On-Chip OSU-HC11 SRAM Write timing

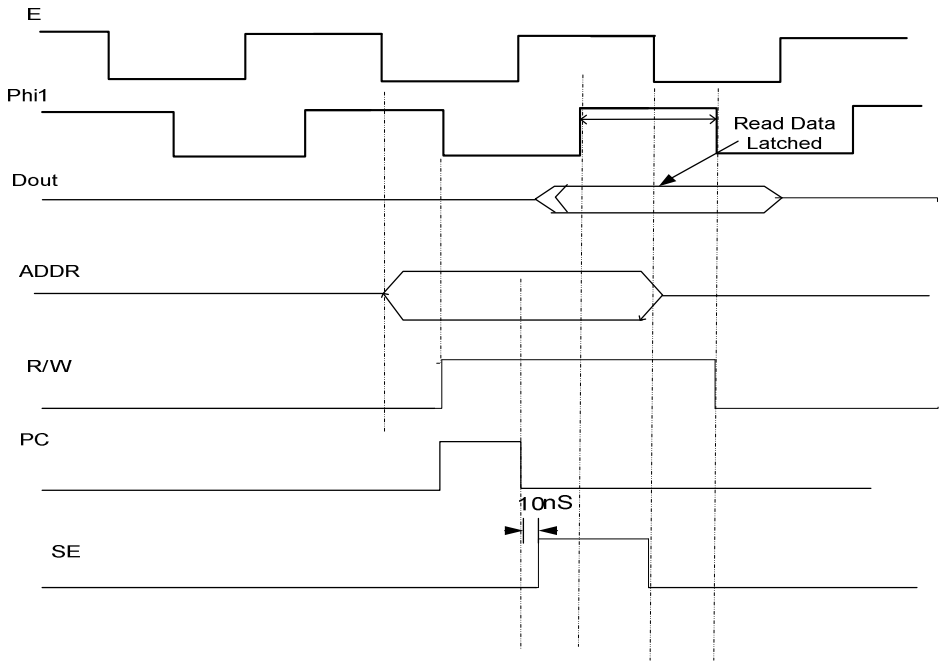



Figure.5.1 (b) On-Chip OSU-HC11 SRAM Read timing

The SRAM write and read timings are shown in figure 5.1(a) and 5.1(b) respectively, where the signals E and Phi1 stand for the clock signals of frequency 2MHz. Phi1 is the delayed version of the system clock E and R/W stands for read/write signal which specifying a read or write memory cycle. The addressing scheme used in OSU-HC11 microcontroller is memory-mapped addressing in which each module has unique address by which it gets selected and new address is provided every negative edge of E clock. The R/W signal is synchronized with respect to Phi1 clock and during a write cycle, the input data from the databus is latched into the memory during the rising edge of Phi1 which allows 125ns for the data value to be written into the SRAM cells. During a read cycle, the bitlines are first precharged for 125ns from falling edge of Phi1 until the rising edge of E clock, and then the read data is latched onto the databus during the falling of E. The additional signals shown in the read timing of figure 5.1(b) are generated internally in the SRAM to effectively control the read process.

OSU-HC11 is an 8-bit microcontroller with 16-bit addressing capability. As discussed earlier, two versions of SRAMs are accessed by the micro-controller, on-chip and SPI SRAMs. The timing diagrams of figure 5.1 represent the on-chip SRAM and are almost the same for SPI SRAM except for Phi1 clock being by the internal SPI clock. The basic design specifications for the two versions of SRAMs are same and are as shown below in table 5.1.

Table 5.1 Design Specifications for OSU-HC11 SRAM

| **Specification** | **Value/Range** |
|---|---|
| Memory Size | 4k bytes (32k-bit) |
| Temp. Range | 0C-275C |
| Frequency of Operation | 2 MHz |
| Power Supply Range | 2.2V-3.3V |
| Process | 3.3V, 0.5um Peregrine SOS (Silicon on Sapphire) |

As seen from the table 5.1, the temperature range specified is quite wide. Even though the speed of operation is significantly lower for an SRAM, the design is still complicated by the fact that the temperature range should be met reliably. Hence in this chapter before proceeding to the actual design of SRAM, general considerations for high temperature design and device geometry selection for these elevated temperatures are discussed as the first section, followed by the discussion on the digital standard cell libraries custom built for applicable circuitry in the SRAM. The digital standard cell libraries will be primarily used for the design, synthesis and layout of row and column decoders, and for circuits used in the generation of control signals to coordinate the read/write cycles.

Following these general requirements for any SRAM design, the actual implementation of various critical modules identified previously in chapter 4 will be discussed in detail in succeeding sections. Following are the critical modules identified in this work: SRAM cell 2×2 block; Sense amplifier & Column read/write logic; Bank control circuitry & row drivers; SRAM bank; Global control circuitry; and Global address decoders. Each of these critical modules emphasizes the design, creation of the

testbenches wherever applicable, simulation, and finally the layout, thereby providing all the necessary steps involved in generating these modules. The last section presents the simulation results and measurement results to validate the design.

## 5.1 General considerations & device geometry selection

The semiconductor process used in this work is an SOI process (Silicon-on-Sapphire, or SOS) from peregrine semiconductor corporation of minimum channel length 0.5um and power supply of 3.3V. The key technology parameters of this process are summarized in table 5.2.

Table 5.2 Typical model parameters of 0.5um Peregrine SOS process at 27C

| **Parameter** | **PMOS** | **NMOS** | **Unit** |
|---|---|---|---|
| $V_{th0}$ | -0.654 | 0.755 | V |
| $T_{ox}$ | 95 | 95 | Å |
| $C_{gso}$ | 3.44 | 2.9 | $10^{-10}$ F/m |
| $C_{gdo}$ | 3.44 | 2.9 | $10^{-10}$ F/m |
| $L_{min}$ | 0.5 | 0.5 | $10^{-6}$ m |
| $W_{min}$ | 1.2 | 1.2 | $10^{-6}$ m |

The minimum length and width numbers as seen from the table are good starting points for consideration into the digital cell library. From the designated bus load it is possible to choose a device length and width satisfying the drive strength required. But this kind of minimum length design is not suitable for a wide range of temperature, and lot of leakage and reliability issues arise at elevated temperature ranges. The threshold

voltage reduction and mobility degradation have to be measured for these increased temperature ranges to meet the design specifications.

To a first order, the operation of a transistor at the high temperature is affected by drop in threshold voltage of the device, degradation in mobility, and significant sub-threshold leakage current mechanism, which is enhanced by the reduction in threshold voltage. It is necessary to consider two factors for a high temperature design – the operating current to leakage current ratio and maximum leakage current in the device. The first factor is the Ion/Ioff ratio which is a good figure of merit for the performance of a transistor in digital circuits. Ion/Ioff ratio which is defined as the ratio of accepted minimum operating current when the transistor is 'ON' to the maximum leakage current when the transistor is 'OFF', indicates the reliability of transistor operation. The second factor which is the maximum leakage current through the transistor will be crucial to estimate the power wasted through leakage. Since modern day chips constitute millions of transistors, it becomes almost mandatory to monitor the leakage current mechanism in a process to enable reduced leakage power design.

All of the above mentioned factors in conjunction with the drive strengths required at a given frequency of operation have to be considered in choosing the device geometries for any digital design to reliably satisfy the wide temperature range. Hence in this work, the device length selection process involved obtaining Ion/Ioff ratio and $I_{leak}$ measurement results for channel lengths varying from Lmin to 3×Lmin for temperatures ranging from 27C to 275C and then choosing a value based on above considerations. Once the lengths for PMOS and NMOS were chosen, their widths for a 1X load value were easily calculated, and used to build the standard cell library.

## 5.2 Row & Column Cell Libraries

The use of a standard cell library for synthesizing and laying out the digital circuitry needed in the SRAM simplifies the design process and helps reduce the time involved in functionality & timing validation, and generating layouts. It is easily seen that the standard cells could be used in critical modules such as global address decoders, global control circuitry, bank control circuitry and row drivers. But the type of standard cells and their cell layout heights may be dependent on whether they needed to be pitch-matched with the SRAM cell's row height or column width. In addition to different cell heights, the standard cells may be oriented horizontally or vertically in their layouts to make efficient use of area. This is especially true in cases of row decoder, and column logic & its control, where the cells may be horizontally and vertically oriented respectively, to simplify routing. Hence to make the design process more flexible, two standard cell libraries were developed in this work: the row cell library and the column cell library, which primarily differ in cell heights, layout orientation style – horizontal or vertical, and buffer drive strengths.

The first task in developing the cell libraries was to select the device geometries for PMOS and NMOS transistors required to create a 1X inverter. Since the frequency of operation of the SRAM is 2MHz, which is quite low, the channel length selection is influenced primarily by the area limitation, the Ion/Ioff ratio at high temperatures and the maximum leakage currents through the devices at these temperatures. Hence the transistor Ion/Ioff ratio and Ileak measurement results for temperatures ranging from 27C to 275C were utilized to select the channel lengths of PMOS and NMOS transistors. A length of 0.8um for a PMOS device and 1.4um for an NMOS device was chosen. The

corresponding Ion/Ioff ratios for these devices at different temperatures are shown in table 5.3, while the maximum leakage currents are shown in table 5.4.

Table 5.3 Measured Ion/Ioff Ratios per 1µm width of 0.8um PMOS and 1.4um NMOS

| **Temperature** | **27C** | **200C** | **275C** |
|---|---|---|---|
| PMOS | $2.36\times10^{6}$ | 7827 | 1613 |
| NMOS | 84000 | 600 | 155 |

Table 5.4 Measured Maximum Leakage currents per 1µm width of 0.8um PMOS and 1.4um NMOS

| **Temperature** | **27C** | **200C** | **275C** |
|---|---|---|---|
| PMOS | $7.014\times10^{-12}$ A | $1.14\times10^{-9}$ A | $7.94\times10^{-9}$ A |
| NMOS | $458\times10^{-12}$ A | $10.6\times10^{-9}$ A | $46.1\times10^{-9}$ A |

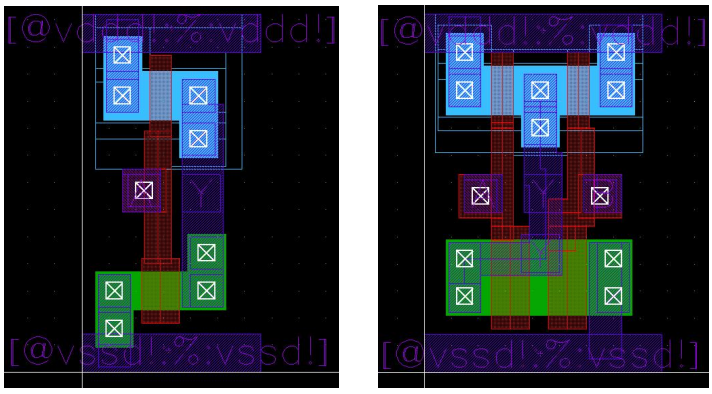
Now that the channel lengths are fixed, the given 1X load of 12fF was used to determine the widths of a beta-matched 1X inverter. Once the geometries of beta-matched 1X inverter are chosen, the task then simplifies to matching the drive strengths for all other necessary logic gates and developing the required buffer drivers.

The SRAM cell was designed, laid-out and its layout dimensions were made available beforehand, so as to enable the layouts of the standard cells. The list of cells included in each library differs slightly in addition to the cell layout heights. The following two sections 5.2.1 and 5.2.2 discuss the row cell library and column cell library respectively from a closer perspective.

### 5.2.1 Row Cell Library

The row cell library was developed primarily to facilitate the synthesis and layout of global row address decoder which needs to be pitch-matched in layout to the SRAM cell's height. In addition to the global row address decoder, the global row buffers, the local decoder and its associated buffers also need to use the pitch-matched cells from the row cell library.

The standard cells needed to meet the requirements of all these circuits primarily include basic 2 and 3 input NAND, AND, NOR, OR and inverter buffers. Hence a total of 15 logic gates were built including inverters of various drive strengths. The transistor level schematics of all these gates were characterized using Cadence Signal-Storm tool across different capacitive loads and input signal slew rates. Once the characterization of these gates was accomplished, all the cells were laid out satisfying the required cell height criterion imposed by the SRAM row pitch, and the laid-out cells were abstracted to help the place & route tool in using these standard cells for auto-layout of any circuit. The layout orientation style is horizontal for all the cells in this library and as an example, the layouts of 1X inverter and 2-input NAND gate are shown below in figure 5.2.

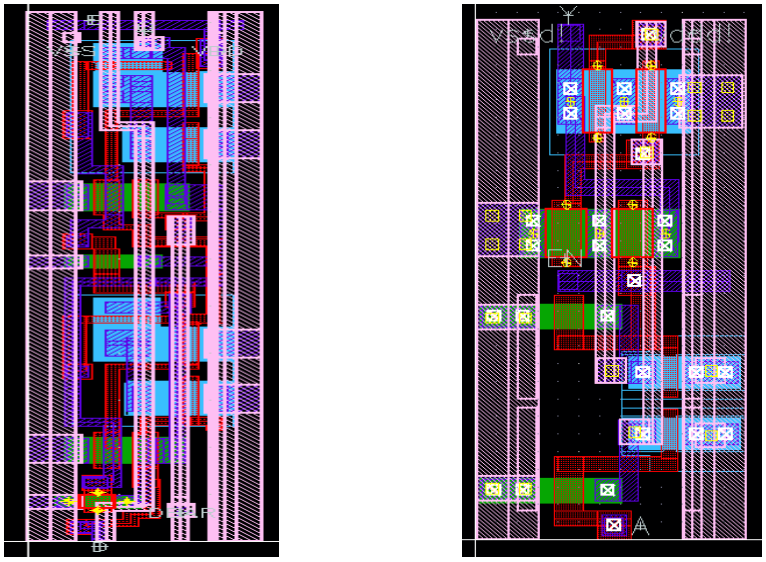


(a) (b)

Figure.5.2 Layouts of row cell library (a) 1X Inverter (b) 2-input NAND gate

### 5.2.2 Column Cell Library

The column cell library is used by bank column control circuits, column read/write logic and global control circuits. The device lengths for PMOS and NMOS devices and geometries for basic gates in the column cell library are the same as used for the row cell library, with the differences being addition of D-latch and tri-state buffers of various drive strengths needed for the column read/write logic and the layout style being vertical for many standard cells. In the column cell library there are certain cells including the D-latch and tri-state buffers, which need to be pitch-matched to the SRAM cell's width. In addition to being pitch-matched in layout, these cells have to be laid out in a vertical fashion so as to efficiently utilize the column logic area. This is exemplified by the layouts of D-latch and 3X tri-state buffer as shown in figure 5.3.



(a) (b)

Figure.5.3 Layouts of Column cell library (a) D-Latch (b) 3-state Buffer (1X)

As it can be seen from the figure, the metal used for power rails is different than that of the cell in row cell library. This does not come as a surprise since the horizontal and

vertical routing of metals are usually accomplished using different metal layers. In addition to these pitch-matched and vertically laid-out cells, other standard cells using less stringent cell heights are also included in the column cell library to synthesize the control circuitry. All the cells in the library are yet again passed through the process of characterization and abstraction to facilitate synthesis and automatic place & route of digital circuits using this cell library.

## 5.3 SRAM cell 2×2 Module

The design of SRAM was partitioned into several critical modules as discussed in the previous chapter. In this section, the first and foremost component, the SRAM cell module is discussed in detail in terms of its design, testbench creation & simulation, and the layout. The terminology 2×2 refers to the layout structure in which 4 SRAM cells, 2 in a row and 2 in a column, are used in a flipped fashion along its two dimensions to minimize the layout area. This same structure of 4 SRAM cells is used in the schematic module so that the entity as a whole can be interpreted as a standard cell, along with its characterization testbench for the benefit of the designer. In the following sub-sections, the design procedure of the SRAM cell used in this work, with the 3.3V 0.5um SOS peregrine process is discussed followed by its testbench creation & characterization finally culminating at the layout of the SRAM cell module.

### 5.3.1 Standard 6-T SRAM cell Design

The design of an SRAM cell is the most important aspect of the entire SRAM module in terms of functionality, performance and reliability. Majority of the die area is occupied

by the SRAM cell array and its building block is the single SRAM cell storage element. It is thus mandatory to design the SRAM cell to provide not just the functionality, but also reliability and performance characteristics such as low operating & leakage power, minimal area across various process, voltage and temperature corners.

In this work, a standard 6-Transistor SRAM cell was used as shown in figure 5.3, with PMOS pass transistors. The reason for choosing the PMOS transistors as the access transistors is that the P-devices have less leakage currents for significantly smaller channel lengths in the peregrine 0.5um SOS process than the N-devices, especially at high temperatures. The reduced leakage currents through the access transistors M5 and M6 of figure 5.4, ensures reliable SRAM cell data retention at high temperatures and also helps reduce significant power dissipation through the entire cell array.

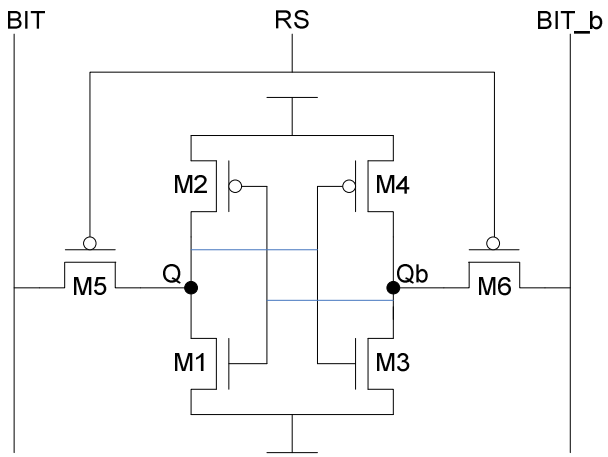



Figure.5.4 Circuit Schematic of 6-T SRAM cell with PMOS access transistors

The actual design of transistor geometries started off with the selection of channel lengths for the NMOS and PMOS devices to be used in the identical cross-coupled inverter. Using the help of measured results for transistor threshold voltage variations and leakage performance at high temperatures, the length of NMOS transistors used in the cross-coupled inverters was chosen to be 1.4um while that of the PMOS is set at 0.8um.

There are two aspects involved with the sizing of transistors in the SRAM cell – prevention of read upset errors when reading a cell and the ability to overwrite the stored bit during write process. With the help of transistor currents flowing into and out of the SRAM cell, the sizing ratios can be determined for all the transistors in the cell. For example, assuming that the bit lines are precharged to VBL, the design equations in general governing the sizing of the transistors in the SRAM cell of figure.5.3 to ensure reliable read operation when the pass transistors are enabled are shown by equations 5.1 and 5.2, where $\Delta V$ stands for the maximum change in internal node voltages of the cell.

$$(VDD - \Delta V - V_{Tpass})(VDD - VBL - \Delta V) - \frac{(VDD - VBL - \Delta V)^2}{2} = \frac{Wpu/Lpu}{Wpass/Lpass}(VDD - \Delta V - V_{Tpu})\Delta V - \frac{\Delta V^2}{2} \quad (5.1)$$

$$(VBL - V_{Tpass})(VBL - \Delta V) - \frac{(VBL - \Delta V)^2}{2} = \mu_n/\mu_p \frac{Wn/Ln}{Wpass/Lpass}\left\{(VDD - \Delta V - V_{Tn})\Delta V - \frac{\Delta V^2}{2}\right\} \quad (5.2)$$

In the above equations, $Wpu$ and L$pu$ refer to the geometries of pull-up PMOS, $Wpu$ and L$pu$ refer to NMOS pull-down of the inverter and $Wpass$ and L$pass$ refer to that of the pass transistor. It has to be ensured that the internal node voltages of the SRAM cell do not exceed the threshold voltages of the cross-coupled inverter pair by sizing the transistors with the help of equations as shown above, thereby preventing read upset errors. Similarly design equations have to be used to size the transistors for writing a value into the cell, in which case the SRAM cell has to be designed such that it can be overwritten reliably during the write process. Usually reading from the cell is more

difficult to design for than writing into the cell and care must be taken to design the cell robust to variations in process, temperature and supplies.

By following the above mentioned design procedure the 6-T SRAM cell was designed to work with a bitline precharge voltage of VDD/2 and simulated to verify the performance and functionality across corners using a comprehensive simulation testbench which is discussed in the next sub-section. The final device geometries achieved to work over all the worst case corners are shown below –

$$W_{pu}/L_{pu} = 1.4\mu m/0.8\mu m$$

$$W_{n}/L_{n} = 1.6\mu m/1.4\mu m$$

$$W_{pass}/L_{pass} = 1.2\mu m/1.2\mu m$$

### 5.3.2 SRAM cell module Testbench

After firsthand device geometries are calculated using the design equations, it is essential to verify the design procedure's validity by simulations and also to gauge its performance across all possible corners along with load changes. This necessitates a series of simulations to verify its functionality, measurement of noise margins, and read/write delays. The simulation results can be interpreted to fine-tune the geometries and achieve the required performance. Hence a testbench was developed as a part of SRAM 2×2 module which could characterize the SRAM cell for all the necessary performance metrics in Cadence Analog Design Environment as shown in figure 5.5.

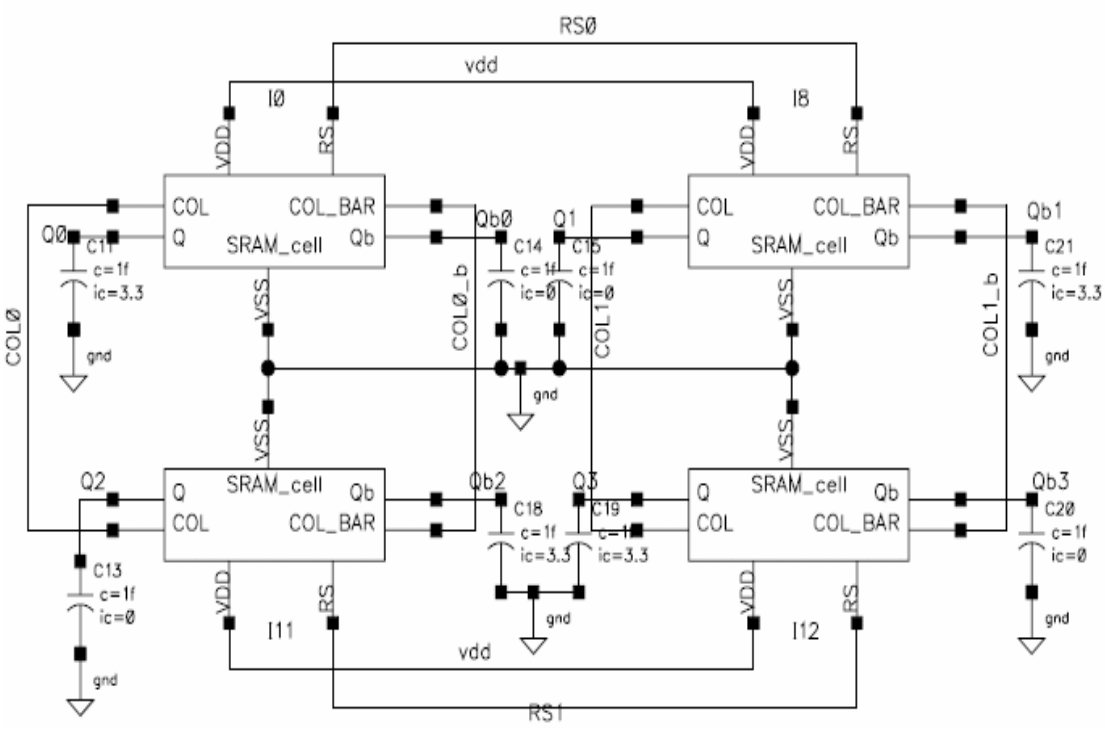



Figure.5.5 Circuit Schematic of SRAM cell 2×2 Testbench

As shown in the figure, there are 4 SRAM cells equally organized along the two rows and columns, representing a miniature cell array. The internal nodes Q and Qb of each cell are pulled out as pins to monitor their voltage fluctuations during a read process and also to measure the write delays. The COL and COL_BAR pins are connected to bitlines named as COL and COL_b respectively, while RS represents the row select. The COL and COL_b lines have been loaded with capacitors whose values are determined by the total number of words which will be given as an input to the testbench. The internal nodes Q and Qb are also loaded with negligible capacitances (1fF in this case) to act as termination as well as help in setting initial bit values to the SRAM cells, thereby allowing the SRAM write test in transient simulation.

The Cadence Analog Design Environment window of the SRAM cell module testbench is shown in figure 5.6. As shown in the testbench, there are two analyses

performed on the testbench circuit – DC analysis and Transient analysis. DC analysis is used to find the cross-coupled inverter pair's switching thresholds. The transient analysis is utilized to find the maximum internal node voltage variations inside the SRAM cell. From these measurements the noise margins of the SRAM cell are calculated and provided to the designer. In addition to the noise margins, the write delay and read delay are measured for a specific voltage difference developed across the bitlines. Using the help of simulation states and design variables, the testbench is made generic to allow user's input of number of word lines and bitlines and also the process variations in conjunction with the temperature, thus helping the designer in measuring performance metrics of any SRAM cell with little effort.

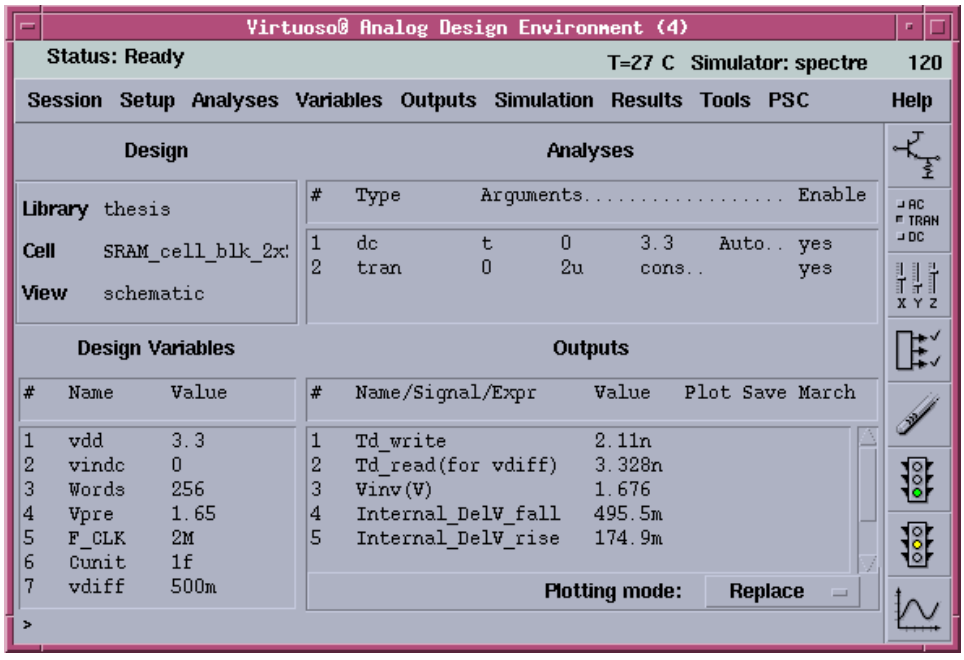



Figure.5.6 Cadence ADE window of SRAM cell 2×2 Testbench

### 5.3.3 SRAM cell Module Layout

The layout of SRAM cell module is organized as a 2×2 array, where 4 SRAM cells are laid-out along the two rows and two columns flipped in layout both horizontally and vertically, to save power rail routing space. The layout of the SRAM cell 2×2 module is shown in figure 5.7. As shown below, the dimensions of the layout structure are 23.4µm×19.6µm. for 4 SRAM cells, making each SRAM cell area as 11.7µm×9.8µm. The SRAM cell module is now complete with all the required components for an SRAM array design, from its initial design and timing characterization to the layout of the array.

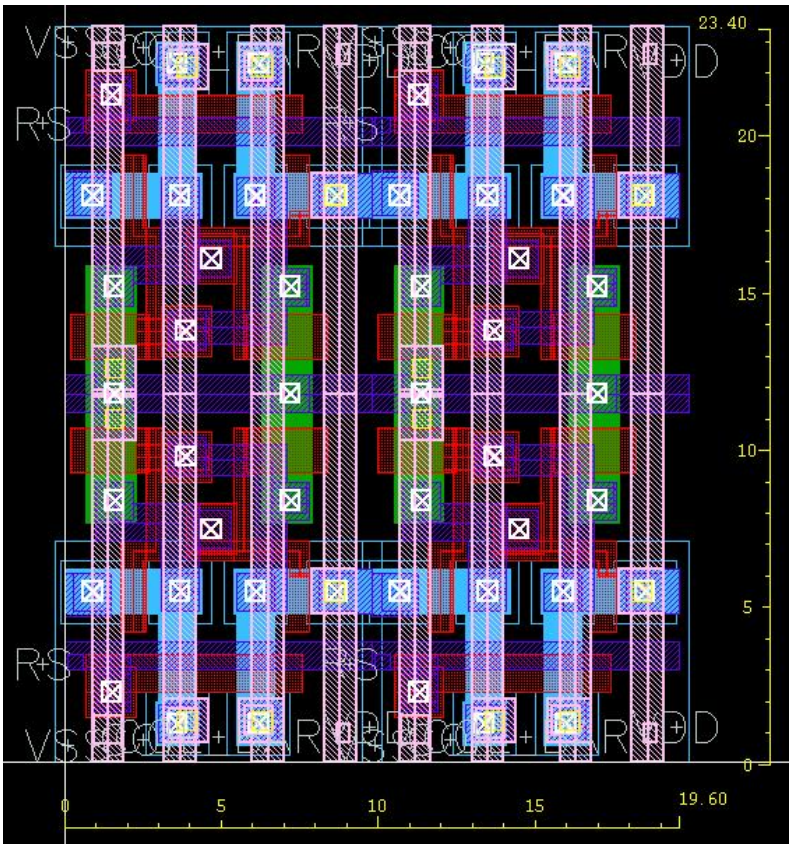



Figure.5.7 Layout of SRAM cell 2×2 Module

## 5.4 Sense Amplifier & Column Read/Write Logic

As discussed earlier, the sense amplifier is grouped with column read/write logic to form a single entity in which except for the sense amplifier, other components of the column logic are digital blocks which can be formed from the column standard cell library developed before. This single entity can be handled easily in the layout as well as characterized for read and write delays. This section deals with the design, simulation and layout of the blocks involved in forming the column logic.

### 5.4.1 Sense Amplifier and Column read/write circuit Design

The sense amplifier used in this work is a variation of alpha latch sense amplifier discussed in chapter 3. It is a clocked regenerative sense amplifier with differential input stage and cross-coupled PMOS loads as shown in figure 5.8. In the figure, the transistor network M1-M9 forms the actual amplifier with transistors M10-M12 being used for precharging and equalizing the bitlines. Transistors M1&M3 and M2&M4 act as single stacked transistor pair used to prevent the kink effect. The sense amplifier is clocked using the sense enable signal SE which controls the NMOS transistor M7.

A typical read cycle starts by precharging the bitlines to a voltage, which is VDD/2 in this case and then enabling the SRAM cell to develop a voltage difference across the bitlines COL and COL_BAR which are the inputs to the sense amplifier. When sufficient voltage is developed across the sense amplifier, the sense enable signal SE is raised, thereby starting the sensing action. The input differential pair produces a corresponding current difference through its legs which enables the cross-coupled PMOS pair to quickly settle the output values to final logic values. The outputs of the sense amplifier are fed to

a D-latch which then latches the data onto the databus through a tri-state buffer. When the sense amplifier is not used, the outputs are pulled high using transistors M8-M9.

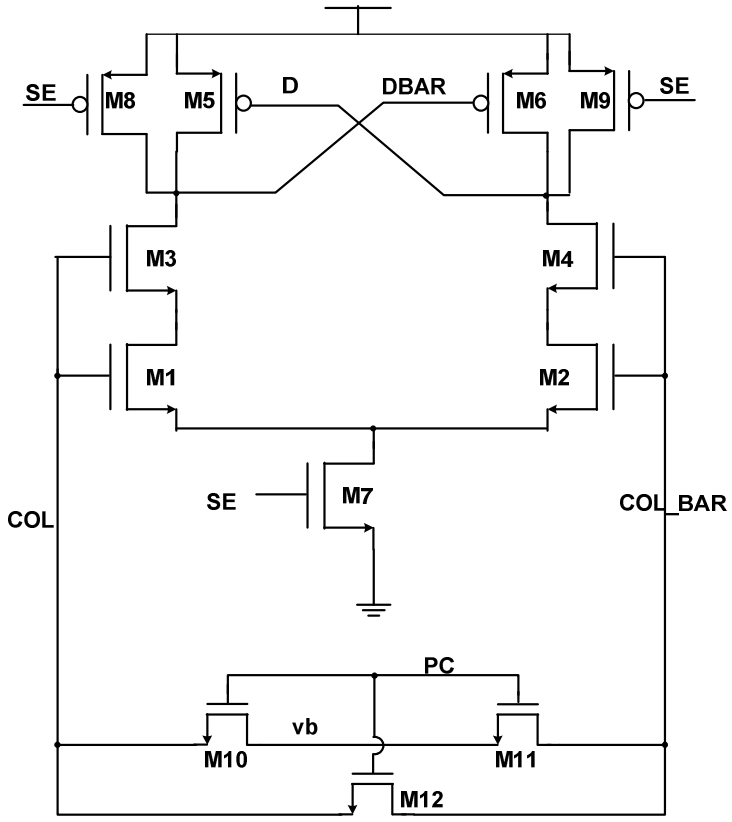



Figure.5.8 Schematic of Sense amplifier for OSUHC11 SRAM

Since the frequency of operation of the SRAM is low (2MHz), the geometries of the sense amplifier are not too critical, and are primarily decided by the input offset voltage and the sensitivity required. The input differential pair was designed to provide less than 5mV of offset, and the pellgrom numbers of the process are used to estimate the device area assuming a channel of 1.4um for the input NMOS pair which satisfies the offset criterion. The input pair M1 and M2 has W/L = 2×10µm/1.4µm. The PMOS loads are beta-matched to the NMOS pair and has W/L = 2×13µm/0.8µm. The offset estimations of these geometries from the pellgrom coefficients are as shown below:

$$nmos : \sigma(\Delta V_t)_n = \frac{11.67mV}{\sqrt{2 \times 10 \times 1.4}} = 2.21mV$$

$$pmos : \sigma(\Delta V_t)_p = \frac{12.51mV}{\sqrt{2 \times 13 \times 0.8}} = 2.74mV$$

$$\sigma(\Delta V_t) = \sqrt{(\sigma(\Delta V_t)_n)^2 + (\sigma(\Delta V_t)_p)^2} = 3.52mV$$

The sensitivity required is assumed to be 6σ which comes to approximately 20mV. Hence this is the minimum voltage required by the sense amplifier to reliably sense the bit. The sense enable signal SE should be enabled only after at least this voltage difference is developed at the input. After worst case estimation of bit line capacitance, the time taken by the SRAM cell to develop this voltage difference across the bitline capacitances at the worst case corner is simulated and the SE signal is triggered sufficiently after this time delay. The designed sense amplifier is simulated for functionality and performance across corners and the worst case sensing delay was found to be about 7ns.

In addition to the sense amplifier, the column logic module has a D-latch and tri-state buffer driving the databus as a part of read circuitry. Likewise the write drivers and its logic form the write circuitry and are being included in the column logic. As discussed earlier, the D-latch and tri-state buffer are pulled off from the column cell library and integrated with the sense amplifier to complete the SRAM column read path. The write logic and corresponding drivers are designed such that the bitlines are pulled high to VDD whenever the SRAM cell is not being accessed and drive data values on the bitlines during the write cycle. The write drivers have been sized such that they provide sufficient currents required for a write process while minimizing the leakage when not used.

### 5.4.2 Column Logic Simulation

The testbench developed for the characterization of column logic involves simulating sense amplifier as a standalone entity in addition to the simulation of the entire data path, thereby allowing a single testbench circuit to characterize both the analog and digital portions of the column data path. The sense amplifier input voltage difference is fed to the testbench by the user, using which the delay contributed by the sense amplifier alone is measured and displayed to the user. This is essential to characterize the delay of the sense amplifier as a function of its input voltage difference. Similarly the whole column logic is connected to an SRAM cell to measure the data write times of the column logic into the cell. This way of using an SRAM cell to provide the data values across the bitlines helps in measuring realistic delays of the whole read path including the D-latch and tri-state buffer. The sense amplifier enable signal SE should be triggered appropriately as to allocate enough time for the SRAM cell to develop a voltage difference across the bitlines. This is achieved by specifying a time delay value to the design variable in the testbench to care of the sense enable timing.

The simulation artist window is shown in figure 5.9 to illustrate the significance of using design variables and cadence functions to automatically measure the timing parameters of the output waveforms and display them for user's benefit. As seen from the artist window, various propagation delays are computed by the testbench which provide the overall time taken for the read process upto the databus as well upto each stage in the data read path.

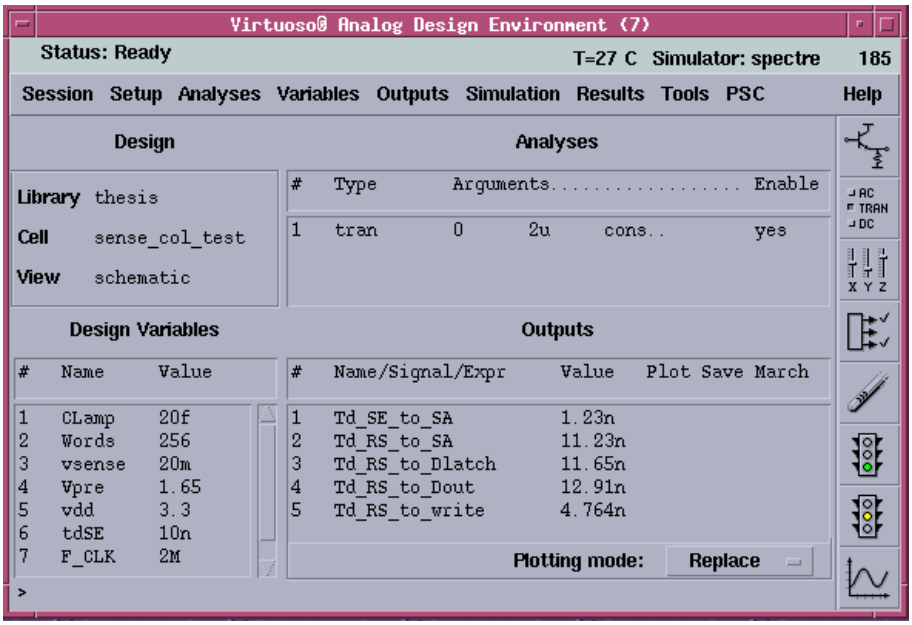



Figure.5.9 Cadence ADE window of Column logic testbench

### 5.4.3 Column Logic Layout

There are four components in the column logic module, the sense amplifier, D-latch, tri-state buffer and write logic. In forming the layout of column logic module, the D-latch and tri-state buffer layouts are pulled off the column cell library while the sense amplifier and write logic are laid-out manually to accommodate additional routing lines in the form of bitlines which cease at the inputs of the sense amplifier. All the control signals are routed horizontally across the column logic. The module has two identical column logic layouts flipped vertically and abutted to make efficient usage of the area. The final dimensions of layout for the column logic are 330µm × 9.8µm, where 9.8µm corresponds to the horizontal pitch of the SRAM cell.

## 5.5 Bank control circuits & local row drivers

The column logic and SRAM cell module discussed in previous sections need control signals to organize their timing for read and write operations. Since the SRAM cell arrays will be organized as distinct banks in this work, only one of them can be active at any given read or write cycle which can be achieved by controlling the column logic and cell array using gated control signals. These bank control circuits are generated from the local control circuitry which is just a repetitive circuitry used by each bank, having common input signals and gated by unique bank select (BS) signal arriving from the column address decoder.

The design of this local control circuit is simple and may be accomplished by using HDL and synthesized using a standard cell library or manually by using schematic capture tools. The column cell library discussed earlier is used to design the local control circuits with sufficient buffers needed by each control signal to drive the load presented by the column logic and SRAM cell array. The bank select signal which gates these signals also need heavy buffering as it virtually drives all the gates in the local control circuitry. Figure 5.10 shows a glimpse of local control signals in the cadence schematic capture tool. Though its design is quite simple, it is essential to verify the timing relationship within these control signals across all corners to ensure the functionality of the read/write process. Hence the control circuitry was simulated in cadence spectre and its timing parameters were verified, in addition to the functional verification done with the help of HDL tools.

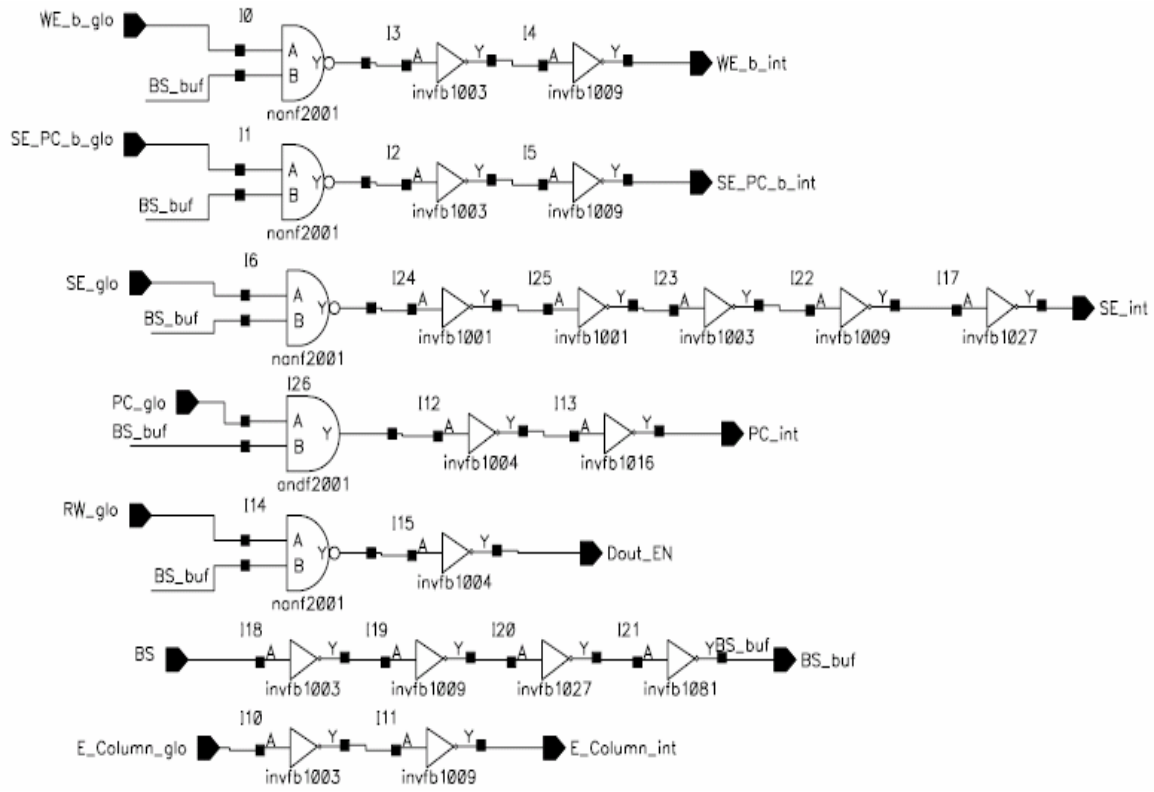



Figure.5.10 Schematic of Local Control Signals

## 5.6 SRAM Bank

The divided word line architecture utilized in this SRAM required the number of banks as 16 for generating a memory of 4k bytes, where each bank is of 256 words and 8-bit wide. The SRAM bank is constructed from the modules discussed so far. It has an array of SRAM cells, in this case 256×8; 8 column logic circuits, one for each bit; bank or local control circuitry and finally the 256 row drivers corresponding to 256 words. The number of words per bank is decided to be 256 based on the layout dimension of each bank, which in turn depends on the layout dimension of each SRAM cell, and total capacitance on the bitlines which affects the read delays. The design of SRAM cell and column logic is carried out by estimating the bitline capacitance offered by 256 words of

the SRAM cell array. The bank control signals are buffered appropriately to drive the 8 column logic blocks needed to access 8-bit data from the array. Similarly the 256 local row drivers are identical circuits which are buffered to drive the access transistors of all the 8 SRAM cells organized in a single row. Since all the necessary components needed to form the SRAM bank are built beforehand, it just becomes the job of grouping them together, both in schematic and layout.

Since the SRAM bank can be considered as a miniature version of the entire memory except for the address decoder delay, the timing characterization of an SRAM bank reveals close to actual read/write access times of the memory. Additionally the switching power dissipation contributed by a single bank will be roughly the same for entire memory with the exception of decoder and global control circuitry power dissipation because of the fact that only one SRAM bank is "on" at a given read/write cycle.

Once the SRAM bank schematic was built, the simulation was carried out in two ways, one with complete SRAM bank schematic and other one with the cell array being reduced in size down to a few bytes to decrease the simulation run time. The bitline capacitance was modeled in the testbench schematic using an ideal capacitance whose value is set by the user during the simulation and the timing parameters like signal rise/fall times, write/ read access times and average switching power dissipation are calculated by the testbench automatically using built-in functions.

When it comes to laying out the SRAM bank, the cell array needs to be laid-out first from the SRAM cell 2×2 module by simply instantiating multiple copies of the 2×2 cell layout block in rows and columns as required. Once the cell array is formed it is connected to the pitch-matched column logic module, pitch-matched row drivers and

local control circuitry to from the SRAM bank as shown in figure 5.8. The figure is split in two owing to its skinny dimensions. Figure 5.11(a) shows the local control circuitry connected to the column logic module and figure 5.11(b) shows a small section of SRAM array in conjunction with its pitch-matched row drivers.

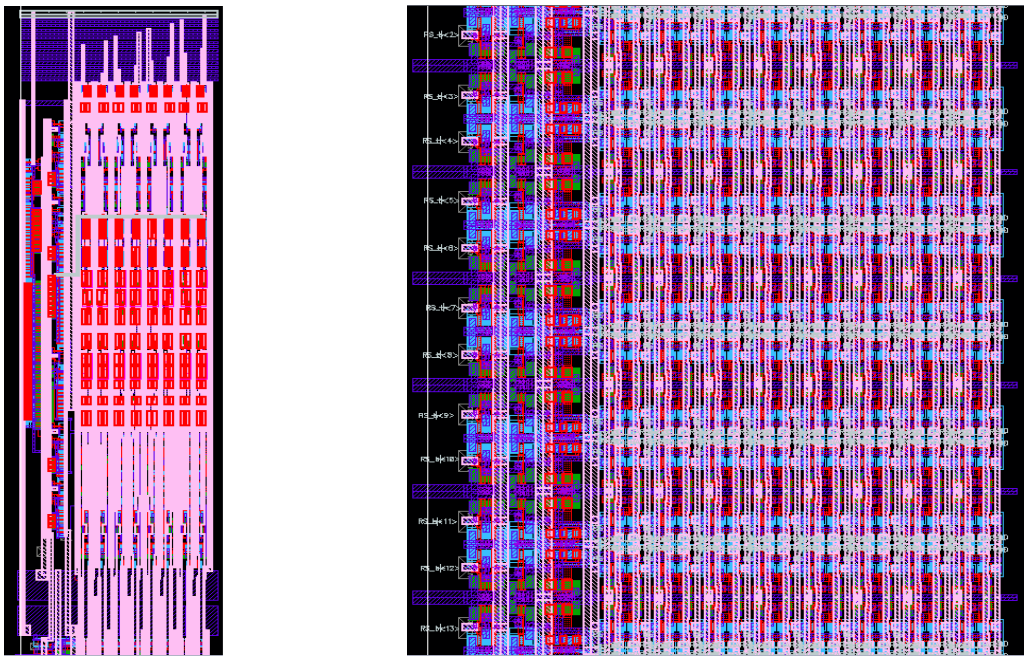


Figure 5.11 Layout of SRAM Bank (a) Column logic and control circuitry (b) small section of SRAM cell array along with its row drivers

## 5.7 Global Control Circuitry

The global control circuitry is the overall controller of memory operations and is responsible for generating signals which control timing of read and write cycles of all the SRAM banks and address decoders. These global control signals run all over the memory and drive the local control circuitry which distributes the buffered control signals to the selected bank. The inputs to the global control circuitry are the memory input pins including, but not limited to, the clock signal (CLK), Read/Write (RW) signal and chip

select (CS). In this work, since two slightly different SRAMs were developed, one for on-chip HC11 RAM and the other off-chip SPI RAM, the global control circuitry was slightly different for these two memories. Except for a few control logic signals, everything else is exactly the same for these two versions of SRAM.

The design of the global control logic was accomplished using HDL and the standard cell library developed earlier. The critical part involved in its design is the buffering of all the control signals needed to drive appropriate bank control logic, as well as buffering the clocks to synchronize the address decoders. The capacitive loading on each of these global control signals are estimated first and then tapered buffers are used. The timing relations of the generated control signals are analyzed using a dedicated simulation testbench and verified across all temperature and process corners. The layout of the final control circuitry was done by utilizing the standard cell layouts of column cell library.

## 5.8 Global Address Decoders

There are two global address decoders used in this SRAM: the row address decoder and the column address decoder. Since the total number of words in each bank is 256, an 8-to-256 row decoder is necessary to select one word out of 256 words of an SRAM bank. Likewise, a 4-to-16 column decoder is necessary to provide 16 Bank Select signals for the 16 SRAM banks where only one is activated for any given address, thus making the total number of words to be 256×16 or 4k, with a word being 8-bit wide. Hence a total of 12 address bits are needed to access all the 4096 words. But the address bus width of OSU HC11 micro-controller is 16, and hence the extra 4 bits are used to enable or disable the memory from accessing the common databus of the micro-controller.

The design of 4-to-16 column decoder is straightforward and was accomplished using pre-decoding the 4 address inputs as two 2-to-4 decoders. The two pre-decoders generate 4 output lines each, which when ANDed two lines at a time generate a total of 16 output lines. This way of pre-decoding reduces number of logic gates required to design the decoder, as well as reduces area and power. But internal lines may need buffers which are usually very small in number when compared to the design without pre-decoding. The actual schematic of the designed 4-to-16 column decoder is shown in figure 5.12. As it can be seen from the figure, the 4 inputs are pre-decoded to generate 8 lines which are then utilized to generate 16 output lines. All the decoder output lines are buffered appropriately to account for the loading. The design was implemented using structural verilog code and was laid-out using the column standard cell library developed earlier.

The 8-to-256 row decoder was designed using two 4-to-16 pre-decoders and its design became much simpler once the 4-to-16 decoder was developed for the column decoder, as the same design can be re-used with very few changes to internal buffering and the addition of final stage comprising of 256 buffered AND gates. The row decoder however has to be laid-out using the row cell library whose standard cells are pitch-matched in layout to the SRAM cell.

Both the decoders were functionally verified using verilog before being subjected to realistic simulations using spectre. The 8-to-256 row decoder was simulated across the worst case corners and its delay was measured to be 7ns at the high temperature corner, which is well faster than the allocated 125ns decoding time. The layout of the decoder was carried out by hand, even though it can be accomplished with the help of automatic place & route tools, to save a lot of layout area. After meticulous hand-layout process, the

final dimensions for the pitch-matched row decoder were 3.3mm × 0.4mm, indicating a skinny decoder.

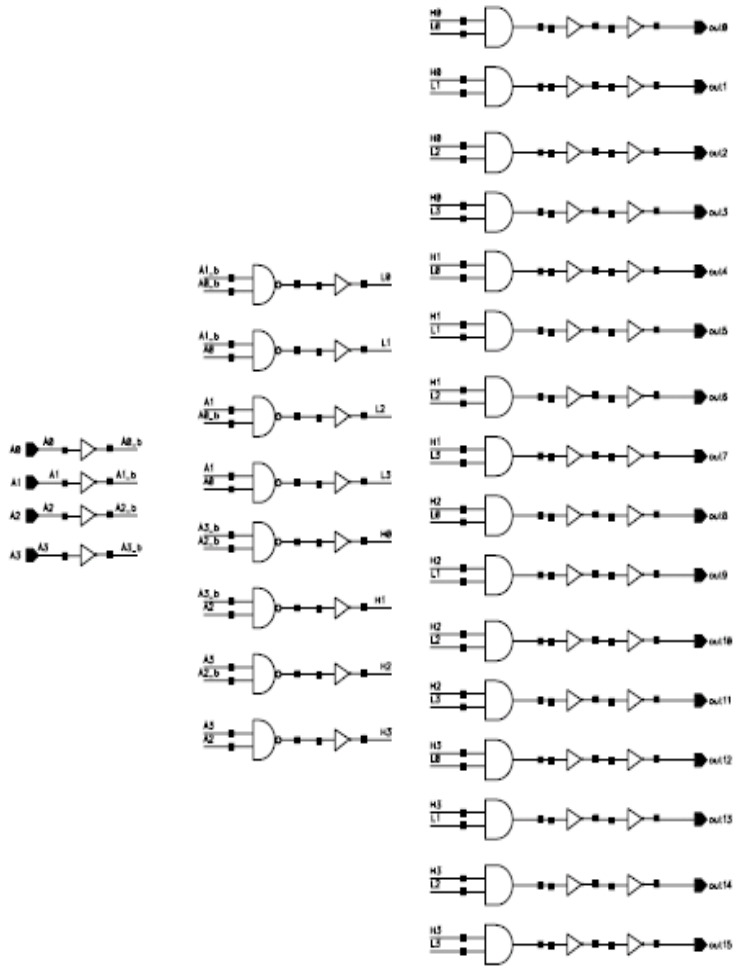


Figure 5.12 Schematic of 4-to-16 column decoder using 2-input pre-decoding

Once all the modules have been characterized individually, they can be grouped together as a library and can be used to construct a complete SRAM with flexibility of changes to existing modules and verifying them using the corresponding testbenches. This procedure was followed to build the complete 4k bytes of SRAM by combining all these block individual blocks and the final layout looks as shown in figure 5.13. The dimensions of the SRAM are 2.2 mm by 3.7 mm or 8.14 $mm^2$.

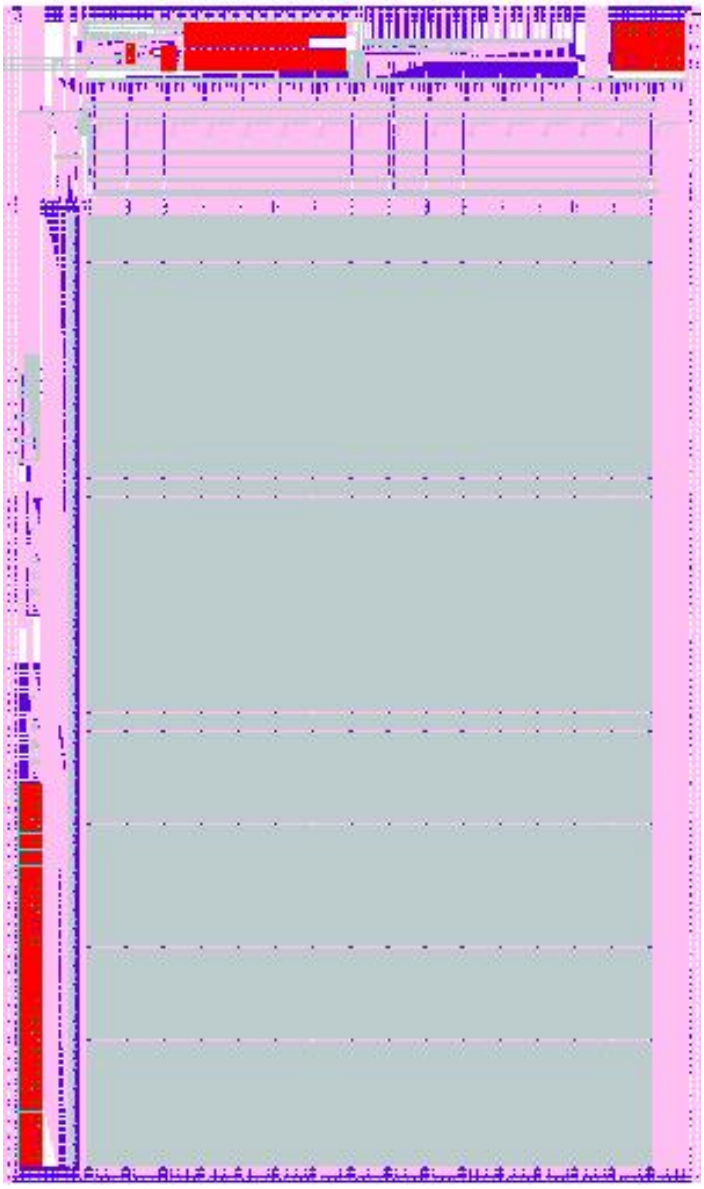


Figure 5.13 Layout of 4k SRAM for OSU-HC11

## 5.9 Results & Waveforms

This section presents the results for various performance parameters of the designed SRAM. The section starts off with a summary of design characteristics and then presents the simulated results for timing parameters and waveforms for SRAM cell and sense amplifier as well as the results for read and write access times. Finally the measured

results for SPI SRAM are presented. A summary of characteristics for the designed SRAM according to simulations are shown below in table 5.5.

Table 5.5 Design Summary for OSU-HC11 SRAM

| **Specification** | **Value/Range** |
|---|---|
| Memory Size | 4k bytes (32k-bit)<br>16 Banks (1 Bank=256×8bits) |
| Temp. Range | 0C-275C |
| Frequency of Operation | 2 MHz |
| Power Supply Range | 2.2V-3.3V |
| Simulated Read time<br>(Decoder + Read data path) | 44ns |
| Simulated Write time<br>(Decoder + Write data path) | 31ns |
| Layout Dimensions | 2.2mm × 3.7mm (8.14 $mm^2$) |

The simulated waveforms of read and write operations for the SRAM cell are shown below in figures 5.14 and 5.15, where Q0 and Qb0 represent the internal nodes of the cell, COL0 and COL0_b represent the bitlines and RS0 stands for the row select signal. The worst case fluctuations of the internal nodes of the SRAM cell are kept under 0.5V so that they are lesser than the threshold voltages of the transistors used in the cross-coupled inverters of the cell.

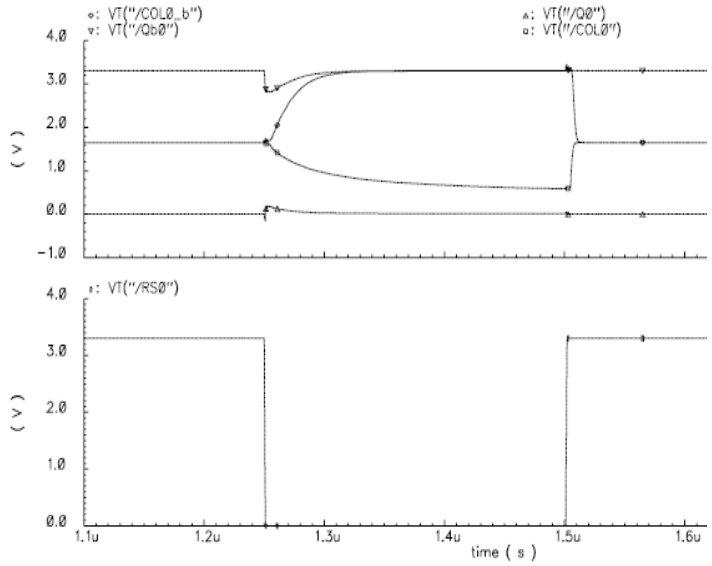



Figure 5.14 Simulated waveforms of SRAM cell read operation

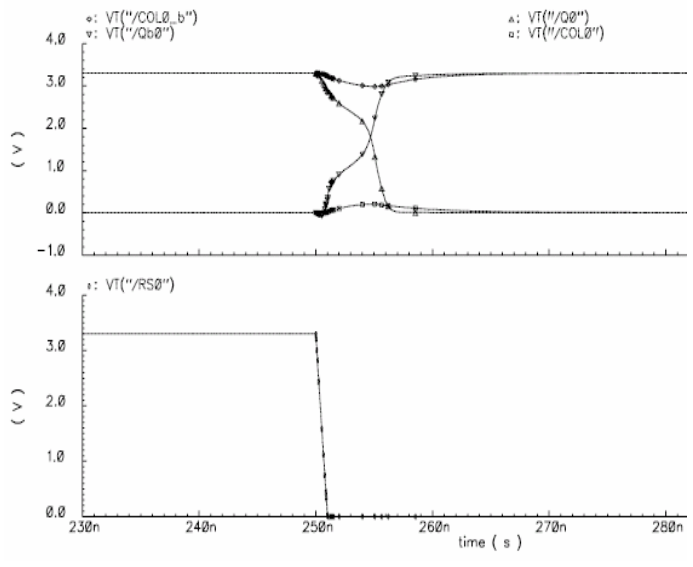



Figure 5.15 Simulated waveforms of SRAM cell write operation

The following table 5.6 summarizes the SRAM cell timing parameters and noise margins for worst case corner of slow process, and 275C.

Table 5.6 SRAM cell timing parameters & cell noise margins

| **Parameter** | **Value/Range** |
|---|---|
| Cell voltage Read Fluctuations | 0.49V (fall)<br>0.18V(rise) |
| Read Noise margin | 1.01V |
| Write Noise margin | 0.9V |
| Read delay (100mV) | 4.56n |
| Write Delay | 5.4n |

The sense amplifier and column logic read timing waveform is shown in figure 5.16, where SA_D0 and SA_Dbar0 are the outputs of the sense amplifier and Qout0 is the output of the D-latch in the column logic, where the total read access time can be measured. The signals COL0 and COL0_b refer to the bitlines and SE is the active high sense enable signal. The design characteristics of the sense amplifier are summarized in table 5.7 for worst case corner of slow process and 275C temperature.

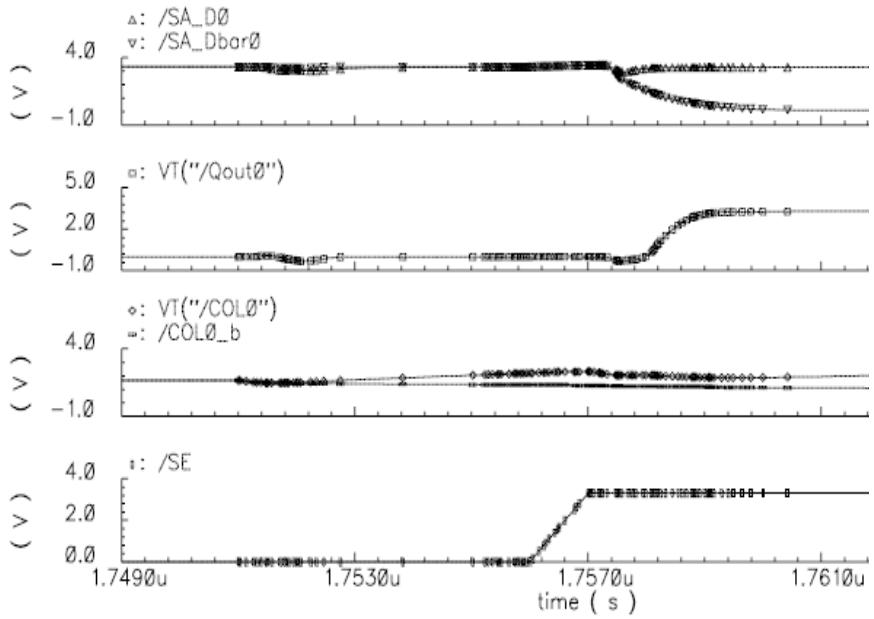



Figure 5.16 Simulated waveforms of Sense amplifier & column logic read cycle

Table 5.7 Simulated Sense amplifier & column logic timing characteristics

| **Parameter** | **Value/Range** |
|---|---|
| Sense amplifier Read delay | 7ns |
| Word line to Dout read delay | 24ns |
| Word line to Write delay | 11ns |
| Average Sense amplifier power/cycle | 55µW |

One bank of SRAM was padded out separately to verify its functionality and measure the performance across the range of temperatures designed for. Various test vectors were applied to verify the functionality, which include, writing all '00' and 'FF' on a background of 'FF' and '00' respectively to all the SRAM cell locations to verify the write test. The written values are then confirmed by reading back the corresponding locations. Similarly the sequences of 'AA' and '55' were also tested to verify the impact of opposite bit patterns on alternate locations. Figure 5.17 shows a snapshot of the SRAM read waveforms for one bank as seen in the Logic Analyzer for reading the sequence 0-255 stored in 256 locations sequentially ,where Do stands for the Data being read out, E is the clock signal, PC stands for the precharge signal and RW is the read/write signal.


Figure 5.17 Logic Analyzer waveform of one SRAM Bank read operation

Similarly tests were conducted on the fabricated SPI SRAM of 4k bytes and the design was verified for functionality across higher temperature ranges. Table 5.8 summarizes the testing results of SPI SRAM.

Table 5.8 Summary of measured characteristics of SPI SRAM

| **Parameter** | **Value/Range** |
|---|---|
| Maximum operating frequency | 8MHz |
| Temperature Range | 27C-275C |
| Maximum Power Dissipation (at 275C) | 26mW |
| Power Supply Range | 2.0V – 3.6V |

## 5.9 Conclusion

There were two different SRAM memories designed for the OSU-HC11 system where one was on-chip and other one was off-chip communicating to the controller through an SPI. Once the on-chip SRAM was designed, methodology discussed in chapter 4 was implemented for designing SPI SRAM where the critical modules that had been characterized earlier were used. The two designs differed in some control signals and clocks used, in addition to providing a dedicated pin for standby operation in the on-chip SRAM. The fabricated SPI SRAM was tested and verified for functionality and the performance metrics were measured for the desired temperature ranges.

# CHAPTER VI

## Conclusion & Future Work

This work was concentrated on implementing a high temperature SRAM to be used in a high temperature micro-controller system for the application of Downhole oil drilling. Importance of reliability concerns caused due to the leakage currents at these higher temperatures formed the basis for designing the standard 6-T SRAM cell geometries and the considerations for choosing the geometries were discussed in the last chapter. Divided word-line architecture was utilized in this work to reduce the delays caused by switching large capacitances of the array, thereby increasing word line activation speeds and lowering the dynamic power dissipation across these large capacitances.

A methodology was proposed and adopted to reduce the design time of high temperature SRAMs by judicious partitioning of SRAM into modules that can be organized and characterized easily using cadence design suite. These individual critical modules can be considered as analogous to the standard cells of digital cell library in a sense that they are characterized for timing parameters and functionality as well as has individual layouts that can be used to form the layout of a complete SRAM. Additionally these critical modules offer flexibility to the memory designer by providing access to generic testbenches that are used to characterize these modules in cadence.

The methodology of design time reduction was adopted for the SPI SRAM after designing the on-chip SRAM. The designed SRAM was fabricated as an SPI SRAM and was tested & verified for its functionality and performance across the specified temperature ranges of 27C-275C.

As a future work, the partitioned SRAM modules and their testbenches can be used as starting points for designing a high temperature SRAM compiler in which the testbenches and simulation process can be done automatically after developing meticulous skill code routines. Adding automatic design capability for SRAM cell and sense amplifier by optimizing the device geometries for a given circuit topology will be hard to achieve, yet provides flexibility to the design process.

# REFERENCES


[1] M. Yoshimoto, et. al., "A 64kb CMOS RAM with divided word line structure", *IEEE International Solid State Circuits Conference, Digest of Technical Papers,* pp.58-59*,1983*.

[2] T. Hirose, et. al., "A 20nS 4Mb CMOS SRAM with Hierarchical Word Decoding Architecture", *IEEE International Solid State Circuits Conference, Digest of Technical Papers,* pp.132-133. 1990.

[3] Bharadwaj S. Amrutur, "Design and analysis of fast low power SRAMs"*, Department of Electrical Engineering, Stanford University, PhD Dissertation, August 1999.*

[4] A. Karandikar and K.K. Parhi, "Low-Power SRAM Design Using Hierarchical Divided Bit-Line Approach", *Proc. of IEEE Int. Conf. on Computer Design* , Austin, Oct. 1998.

[5] Sejun Kim, Ilkwon Chang, Seungyoung Seo, Kaedal Kwack, "A Folded Bit-Line Architecture for High Speed CMOS SRAM", *6th International Conference on VLSI and CAD, 1999.*

[6] Jan M.Rabaey, Borivoje Nikolic and Anantha P.Chandrakasan, "*Digital Integrated Circuits: A design Perspective*", 2ed, Prentice-Hall 2002.

[7] T.N. Blalock et. al., "A high-speed clamped bit-line current mode sense amplifier", *IEEE JSSC*, pp. 542-548, 1991.

[8] B. Wicht, "*Current Sense Amplifiers for Embedded SRAM in High-Performance System-on-a-Chip Designs*", Heidelberg, Germany: Springer Verlag, 2003.

[9] K. Sasaki, et. al., “A 15-ns 1-Mbit CMOS SRAM”, *IEEE Journal of Solid State Circuits,* vol. 23, no. 5, pp. 1067-1071, October 1988

[10] Aiyappan Natarajan et al., “A Study of Sense Amplifiers for Advanced Microprocessor Caches in 70nm Technology,” Intel Circuit Research Labs (CRL), Hillsboro, OR.

[11] Neil Weste and David Harris, “CMOS VLSI Design: A Circuits and Systems Perspective”, Addison Wesley, 3rd Edition, 2004.

[12] Tsugro kobayashi, et al.., “A Current-mode Latch Sense Amplifier and a static Power Saving Input Buffer for Low-Power Architecture”, *IEEE Journal of Solid-State Circuits*, vol28, pp. 523-527, Apr.1993

[13] Jacob Baker, “*CMOS: Circuit Design, Layout, and Simulation*”,2ed, Wiley-IEEE, 2004

VITA

Srikanth Vellore Avadhanam Ramamurthy

Candidate for the Degree of

Master of Science

Thesis: DESIGN OF A FLEXIBLE HIGH TEMPERATURE SRAM WITH REDUCED DESIGN TIME

Major Field: Electrical and Computer Engineering

Biographical:

Personal Data: Born in Macherial, India on February 22nd, 1984, the son of V.A.Ramamurthy and V.A.Rama Gowri.

Education: Graduated from higher secondary school in Chennai, India; received a bachelor's degree from Anna University, India in May 2005 in the field of Electronics & Communication Engineering; completed the requirements for the Master of Science degree with a major in Electrical & Computer Engineering at Oklahoma State University in December 2007.

Experience: Worked as a Graduate Research Assistant in Mixed Signal VLSI Design Lab in the Department of Electrical & Computer Engineering at Oklahoma State University from November 2005 to Dec 2007.

Worked as a Student Intern at Orora Design Technologies Inc., Redmond, WA from July 2007 to Sept 2007.

Professional Memberships: Member of Phi Kappa Phi, 2006-2007

Name: Srikanth Vellore Avadhanam Ramamurthy Date of Degree: December 2007

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: DESIGN OF A FLEXIBLE HIGH TEMPERATURE SRAM WITH REDUCED DESIGN TIME

Pages in Study: 80 Candidate for the Degree of Master of Science

Major Field: Electrical and Computer Engineering

Scope and Method of Study:

The objective of this research is to design an SRAM that works for the temperature range 0C to 275C, with an emphasis on reduction in design time by segregating critical modules and characterizing them. This way of characterizing the critical modules has proven to be helpful in duplicating them multiple times to achieve the size of the memory required, and also to meet the performance requirements.

Findings and Conclusions:

A High temperature SRAM system consisting of two 32k-bit SRAMs (an on-chip SRAM and an off-chip SPI SRAM) was designed and fabricated in 0.5um SOS Peregrine process. The methodology of design time reduction was used for SPI SRAM which is a variant of on-chip SRAM. The fabricated SPI SRAM was tested across the operating temperature range and upto 275C, and the performance parameters were measured. The measured performance parameters were found to be in accordance with/exceed the required performance.

ADVISER'S APPROVAL: Dr. Chriswell Hutchens