EXPLORING PROCESS-VARIATION TOLERANT DESIGN OF NANOSCALE SENSE AMPLIFIER CIRCUITS

Oghenekarho Okobiah

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

December 2010

APPROVED:

Saraju P. Mohanty, Major Professor
Elias Kougianos, Co-Major Professor
Mahadevan Gomathisankaran, Committee Member
Ian Parberry, Chair of the Department of Computer Science and Engineering
Costas Tsatsoulis, Dean of the College of Engineering
James D. Meernik, Acting Dean of the Robert B. Toulouse School of Graduate Studies

Okobiah, Oghenekarho. Exploring Process-Variation Tolerant Design of Nanoscale Sense Amplifier Circuits. Master of Science (Computer Engineering), December 2010, 71 pp., 8 tables, 37 illustrations, bibliography, 33 titles.

Sense amplifiers are important circuit components of a dynamic random access memory (DRAM), which forms the main memory of digital computers. The ability of the sense amplifier to detect and amplify voltage signals to correctly interpret data in DRAM cells cannot be understated. The sense amplifier plays a significant role in the overall speed of the DRAM. Sense amplifiers require matched transistors for optimal performance. Hence, the effects of mismatch through process variations must be minimized. This thesis presents a research which leads to optimal nanoscale CMOS sense amplifiers by incorporating the effects of process variation early in the design process. The effects of process variation on the performance of a standard voltage sense amplifier, which is used in conventional DRAMs, is studied. Parametric analysis is performed through circuit simulations to investigate which parameters have the most impact on the performance of the sense amplifier. The figures-of-merit (FoMs) used to characterize the circuit are the precharge time, power dissipation, sense delay and sense margin. Statistical analysis is also performed to study the impact of process variations on each FoM. By analyzing the results from the statistical study, a method is presented to select parameter values that minimize the effects of process variation. A design flow algorithm incorporating dual oxide and dual threshold voltage based techniques is used to optimize the FoMs for the sense amplifier. Experimental results prove that the proposed approach improves precharge time by 83.9%, sense delay by 80.2% sense margin by 61.9%, and power dissipation by 13.1%.

ACKNOWLEDGMENTS

I would like to sincerely thank my major professor, Dr. Saraju Mohanty, for the encouragement and guidance he has given me throughout this process. His constant support and technical feedback has made this research possible. I would also like to thank my co-major professor, Dr. Elias Kougianos, who also encouraged and motivated me through the completion of this work. I am grateful for his guidance on EDA tools. I would also like to thank Dr. Mahadevan Gomathisankaran, my committee member, for agreeing to examining this research. Special thanks and appreciation goes to my father, mother, brothers and sisters whose constant moral support, encouragement and financial support has helped me finish this research. I would also like to acknowledge the NanoSystem Design Laboratory (NSDL) members: Garima, Oleg, and Asha who have helped me when I had any problems. I also wish to thank the staff of the Department of Computer Science and Engineering for being very helpful and caring in times of need. Lastly, I would like to thank God, who has been my rock through this all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Dynamic random access memory (DRAMs) are very crucial to the performance of computer systems. They are right in the middle of memory hierarchy consisting of cache (made of static random access memory), main memory, and hard drive (magnetic memory or solid state memory). The improvement of other system components' performance and speed, particularly for microprocessors, has outpaced the performance of DRAMs. Thus, the DRAM has increasingly become a bottleneck for the performance of many computer systems. DRAMs however are continually used in many computing applications because of their ability to be densely packed. The current trend of miniaturization of devices and growing popularity of embedded systems, which employ DRAMs, increases the importance of DRAMs.

The evolution of the DRAM has been an instrumental part to the advancement and performance of modern computer systems. DRAM chips are used in almost all types of computing systems because they are cheap, relatively fast, and possess high density storage [33]. This in essence serves the purpose of the memory hierarchy of computer systems which is making the "maximum possible and fast memory available to the users at a minimal possible cost." Static random access memory (SRAM) is another popular type of memory. It is faster than DRAM but more expensive and also has a lower memory density. The SRAM is often used as a supplement or a buffer for the DRAM which can be used in more capacity. With the development and design of many memory intensive devices such as smartphones, media players, and other similar portable devices, the impact of DRAMs on electronic systems continues to grow. Therefore, there is a large body of research on improving the performance of the DRAM and still maintain its cheap cost and high memory density.

Aggressive scaling of the silicon process technology has led to dramatic improvements in processor performance. The developments in chip technology, which have been steered by Moore's

law has seen the number of transistors that can be placed on a single integrated circuit approximated double every two years. The ability to place more transistors with increased switching speeds has led to increases in performance and speed of processors and increase in memory capacities. The rate of increase in the speed performance of memory systems however do not match up to that of processors. One of the primary causes is the capacitive component of the DRAM memory system which limits speed and performance. It takes longer for capacitive components to charge and discharge [30]. This has resulted in drastic and increasing processor memory performance gap [10] as shown in Figure 1.1. Arbitrarily increasing the switching speed of the transistors on the DRAM chip will exponentially increase the power consumption. This is a major consideration in DRAM design. Another reason for the lagging performance of DRAM systems compared to processor performance is that DRAM devices have been generalized enough to be produced with the minimum components to ensure device operation. As such it is very sensitive to fabrication costs and only improvements resulting in significant performance benefits are incorporated in new generations [30]. Determining features that could significantly enhance performance is quite complex. This is because of there are a large number of independent factors that could affect the peformance enhancement including memory system architecture and configurations [7, 30]. The

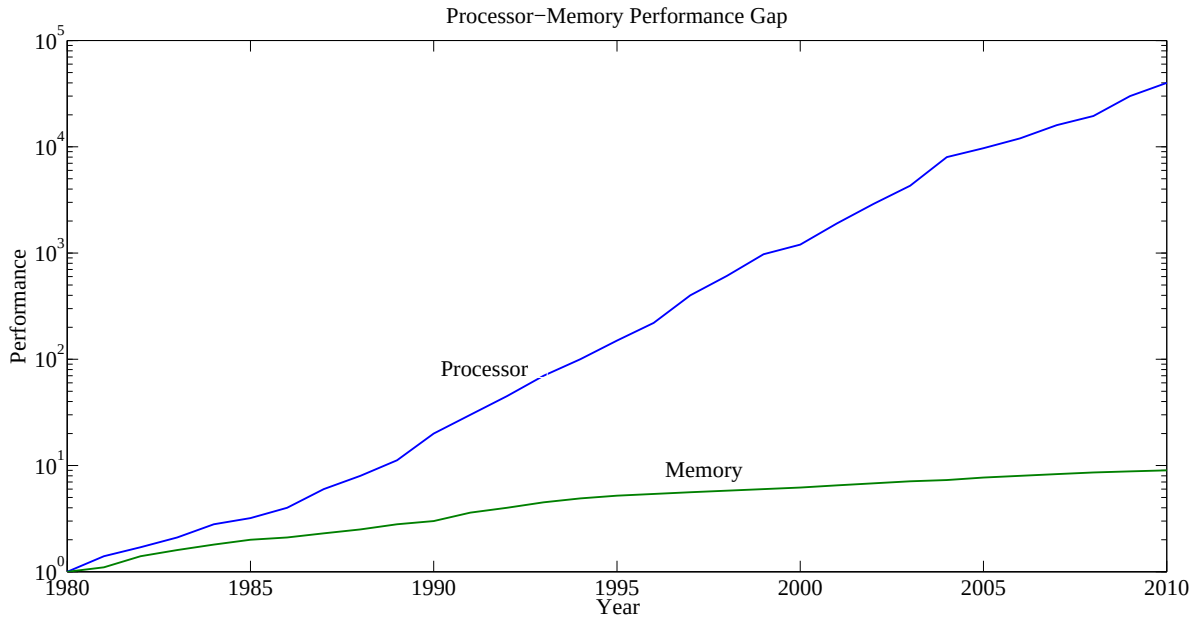


FIGURE 1.1. CPU memory speed gap trend.

lagging performance of memory systems has increasingly become a bottleneck to the overall performance of computing systems as DRAM circuits have become integral to the design of many computing systems. While ongoing research continues to find ways to decrease the cycle times of DRAM memory systems, the speed gap in processor technology and memory circuits continues to increase at an exponential rate. The exponential increase of the processor performance and almost linear increase in the memory performance is clearly demonstrated in Figure 1.1. Part of the lag in speed is attributed to the fact that most DRAM chips are manufactured separately from processor circuits. One of the approaches used to alleviate the burden memory systems place on computer systems is to place both memory systems and processors on the same die. The increasing size and capacity of memory systems driven by the memory demand of recent applications makes them however too large to place on the same circuit with the microprocessors. Hence, this approach has not become popular. The approach adopted instead is to integrate a small size memory on the chip which is used as buffer during execution. SRAM, which is faster than DRAM, is used and it occupies about 70% of the chip area of microprocessors [12].

1.1. Historical Development

A brief historical development of the DRAM is presented in this section. The first MOS circuit was designed in the 1960s. The design of the 1 Kb DRAM was a major milestone in the development of DRAM circuits. Advances in process and circuit design technology have seen the DRAM evolve from the early 1,024 bit (1K) DRAM to the conventional gigabit (Gb) DRAMs. The first 1 Kb DRAM was designed in the early 1970s with a three transistor circuit that used PMOS technology. Current DRAM circuits are made with a one transistor-one capacitor circuit (1T-1C). Computer memory consisted of magnetic cores before the use of CMOS DRAMs. A small magnetic ring, which had 2 wires strung to perform read and write functions made up memory cells [28]. The first DRAM was invented by Robert Dennard in 1968. However, the first fabricated DRAM chip was the Intel 1103 chip in 1973.

1.1.1. 1 Kb DRAM

The 1 Kb DRAM was developed in 1973. It was the first fabricated DRAM chip. It used the three-transistor technology and was built with PMOS transistors. The voltage levels used for the power supply were -12 V and 5 V as V_{DD} and V_{SS}, respectively. This was because the circuits had to interface with logic circuits using transistor-transistor logic (TTL) [15].

1.1.2. 4 Kb - 64 Mb DRAM

With advancement in process and design technology, the 4 Kb - 6 Mb generation of DRAMs was introduced shortly after the 1 Kb DRAM. This generation features major advances in the historical development of the DRAM, a list of which includes the introduction of the following:

- 1T-1C memory cell: Although this was the first cell design, it was finally adopted by the mainstream in this era.
- Multiplexed addressing: It was used to reduce the input/output (I/O) pins on the chip.
- Multiple memory banks or array: It was used to reduce the effect of increasing capacitance of bitlines.
- CMOS technology
- Single 5 V power supply.

The memory size in this era varied from 4 Kb to 64 Mb with different variations of 4K$\times$1, or 16M$\times$4, 8M$\times$8, and 4M$\times$16.

1.1.3. Gb-SDRAM

The Gb or synchronous DRAM (SDRAM) generation featured the next major innovation which was the addition of a synchronization circuit between the DRAM circuit and control from the chip. This was to make all operations of the DRAM execute on the rising edge of a master clock. This enables the DRAM to operate faster. Sizes in this era began to feature Gb DRAMs.

1.2. Basic Configuration and Operation

This section gives a brief overview of the major circuits of the modern DRAM. The basic configuration of a DRAM memory system is shown in Figure 1.2. It demonstrates the basic functional parts of a DRAM system. The major circuit components include the following:

- memory array,
- sense amplifiers,
- address decoders,
- high-speed I/O circuits, and
- the data in/out buffers.

These are discussed in the rest of this section.

1.2.1. Basic Circuits

The major circuits in a modern DRAM include the memory array, the decoder driver and the sense amplifier. They are briefly described below.

1.2.1.1. The Mbit Cell and Memory Array

The "Memory Array" is an array of DRAM bit (mbit) cells. The cell design used in modern DRAM circuits is the 1T-1C cell circuit shown in Figure 1.3. The transistor acts as a switch which controls current flow while the capacitor stores the charge which is interpreted as a logic "1" or a logic "0" value. For a write, read, or sensing operation, the transistor is switched on by the wordline and the capacitor discharges or shares its stored charge with the bitline, which is initially precharged to a predetermined voltage. This charge sharing causes the voltage of the bitline to either increase or decrease. An important component of the DRAM circuit, the sense amplifier, amplifies this voltage change to a full swing of V_{DD} or 0 for a "1" or "0", respectively.

The mbit cells are arranged together in a matrix fashion to form the memory array. There are two major types of array architectures: the open bitline and the folded bitline architectures. The folded bitline is preferred over the open bitline because the wordline couples with both bitlines, (the bitline of the memory cell and the dummy cell); this injects a common noise into the sense amplifier, whereas in open bitline, the wordline couples with only one bitline presenting a noise mismatch to the sense amplifier. Figure 1.4 shows the configuration of a 4×4 DRAM array.

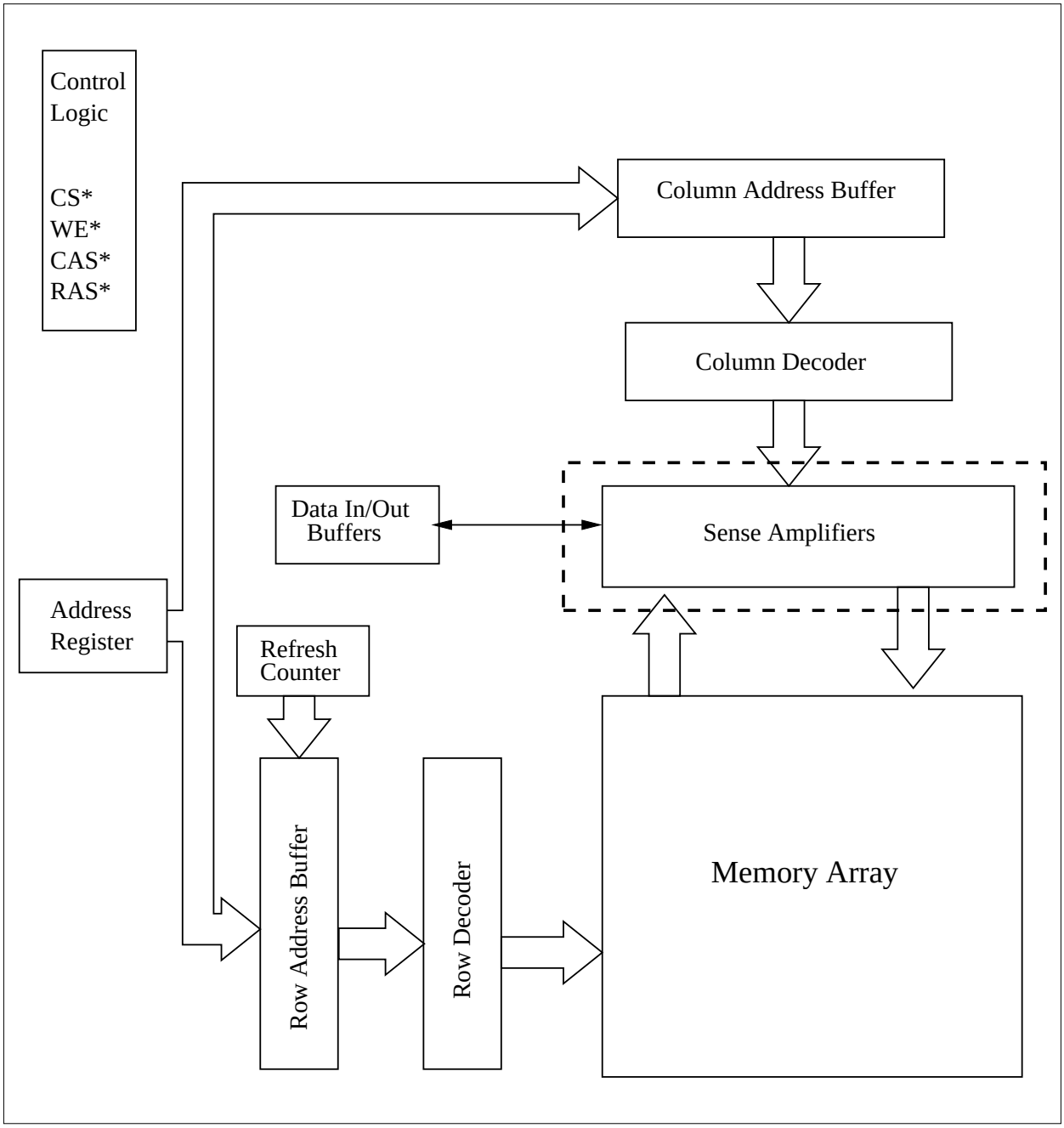


FIGURE 1.2. Block diagram of a DRAM system demonstrating the major components.

1.2.1.2. Decoder

The decoder constitutes the wordline driver and an address decoder tree. The wordline driver enables the wordline to be boosted up to the wordline voltage. There are three basic configurations for the wordline driver which are as follows:

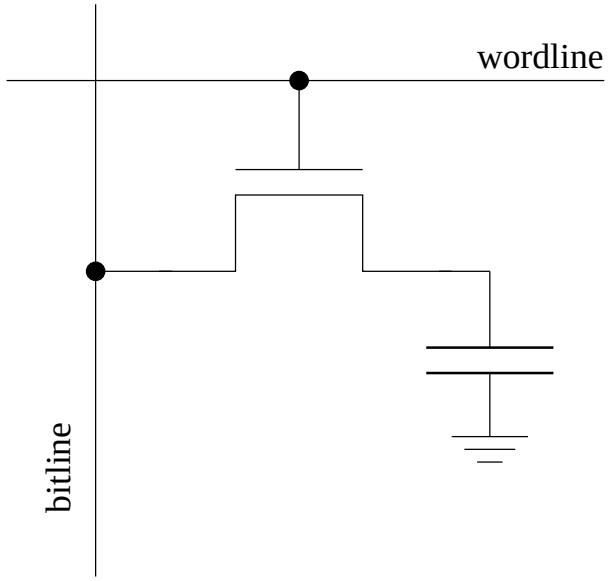


FIGURE 1.3. Circuit diagram of a 1T DRAM.

i. Bootstrap

ii. the CMOS

iii. NOR drivers

The address decoder involves designing to optimize the speed of decoding with minimal area design. Design logic used includes static, dynamic, pass gates or any combination.

1.2.1.3. Sense Amplifier

The sense amplifier is one of most important circuits of the DRAM. They perform the very important function of detecting and amplifying the minimal voltage change in the bitlines of the DRAM. The sense amplifier is connected to a pair of bitlines. One of the bitlines is the output data from a cell, while the second bitline is set at a precharged level and used as a reference. The sense amplifier detects the difference in the voltage level of the bitlines. It amplifies the difference to an extreme to be interpreted either as a logic “1” or a logic “0” value. The sense amplifier is also used to refresh the memory cells after a read cycle. It writes the amplified value back to the DRAM cell to restore it to either a strong logic “1” or a “0” value. The amplification of the voltage change by the sense amplifier is very important to the performance of the DRAM as this operation determines to a large extent the correct operation of the memory system. Sense amplifiers have increasingly

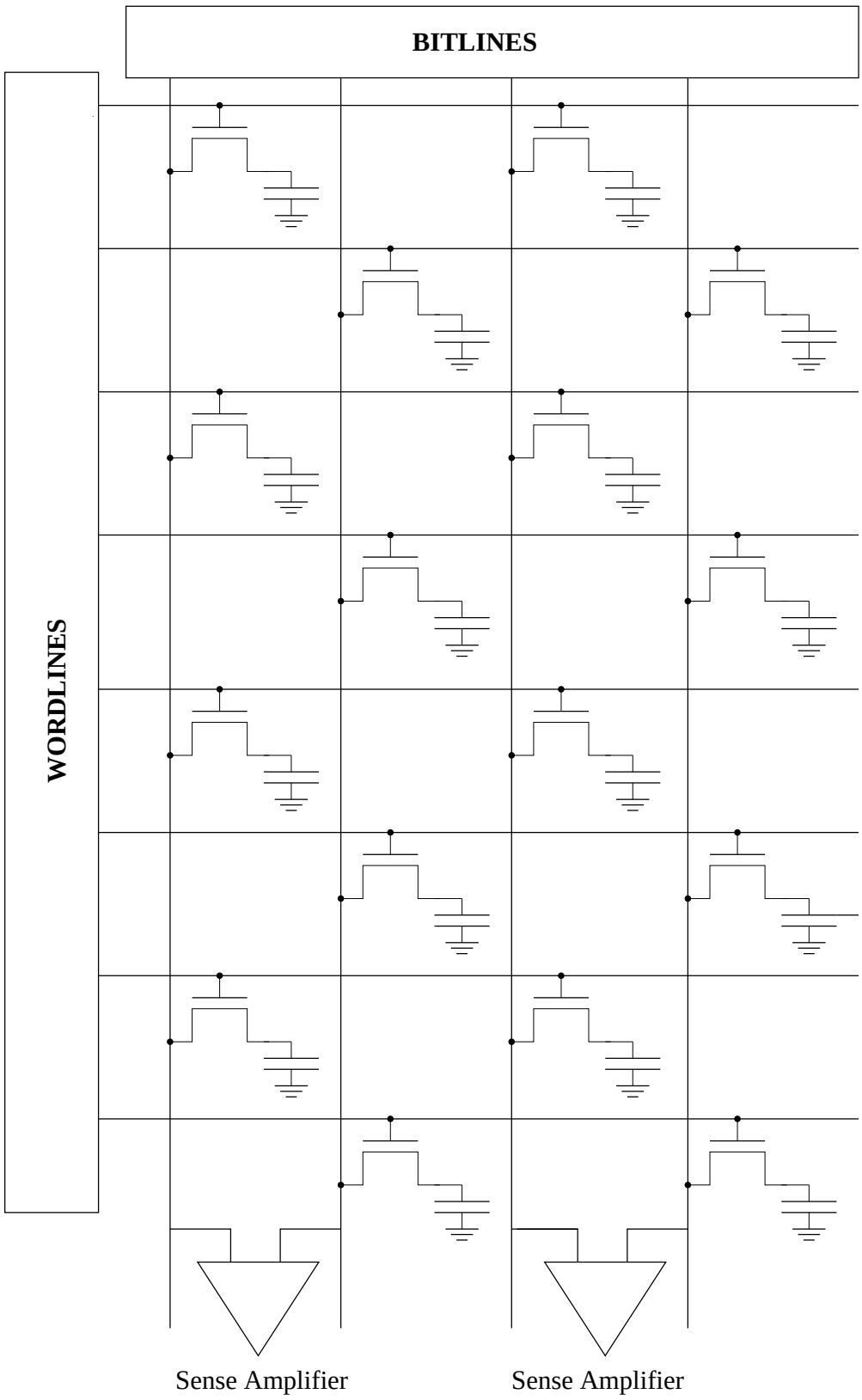


FIGURE 1.4. A 4×4 folded bitline memory array showing DRAM cells and sense amplifiers.

become one of the most critical circuits in the design of the DRAM with the current scaling of device technology below 50 nm. This thesis makes a contribution to the study and analysis of the process variations, which greatly impact the performance of the sense amplifiers.

The other circuit components in the DRAM include the following:

- High speed I/O circuits
- Data bus amplifier
- Voltage regulator
- Reference generator

The voltage regulator is used to derive the boosted wordline voltage and the precharge voltage.

The control block provides important timings for the correct operation of the DRAM. The "row address strobe (RAS)" clocks in the row address and the "column address strobe (CAS)" clocks in the column address. For modern DRAMs, the address inputs are multiplexed to read addresses in each direction. In other words both the row and column address can be read with the same pins. After the row address is latched by the row address buffer, the CAS clocks in the column address on the same pins which, are then latched by the column address buffer. A DRAM cell is accessed by a row and column address. A cell is located at the intersection of a row and column. The *R/W* input indicates whether it is a read or write access while the *CE* input has to be enabled. The Data in/out bus manages the input or output data for the DRAM.

1.2.2. Basic Operation

The three basic operations, read, write, and refresh are now discussed.

1.2.2.1. Read Operation

It is performed by setting the *R/W* pin high and then enabling the *CE* pin. The important times are the "read cycle time" (t_{RC}) and "access time" (t_{AC}). t_{RC} indicates the speed of the memory system. How fast the memory can be read is a crucial design constraint.

1.2.2.2. Write Operation

This operation is performed by setting the *R/W* signal low and enabling the *CE* pin. The "write cycle time" (t_{WC}) indicates how fast we can write data to a cell. The "address to write delay time"

(t_{AW}) indicates the time between the address changes to the write pin enabled, while the "write pulse width" (t_{WP}) indicates how long input data must be present before it is valid and the write pin is disabled.

1.2.2.3. Refresh Operation

This is accomplished by accessing every possible address location of the DRAM. The *CE* pin is enabled, while the *R/W* pin is used to signal a read and a write back of data read to refresh the DRAM cell.

1.2.3. Modes of Operation

In the evolution of DRAMs, design engineers have introduced different techniques of accessing the DRAM in an attempt to increase the access time. There are different modes of operation some of which include the following:

- Fast Page Mode (FPM)
- Extended Data Out (EDO)
- Burst EDO (BEDO)
- Synchronous DRAM (SDRAM)
- Rambus DRAM (RDRAM)

The different modes of operation are briefly described below.

1.2.3.1. Fast Page Mode (FPM) DRAM

In this approach, multiple DRAM cells within the same row or on the same page are accessed successively. In conventional DRAMs, for each cell access, after the address is decoded, the page has to be set up (opened) for access. If a stream of cells to be accessed are on the same page, the performance of the DRAM is enhanced by keeping the page open and then accessing different column addresses. This is called "fast page mode (FPM)", because the row (also known as a page) is kept open for multiple column reads. It improves access time by eliminating the time required to open a page for successive reads on the same row [7, 23]. A variation of the FPM includes the "static column mode", in which the column address required is changed without strobing in a

new access. The "nibble mode" where four serial bits are accessed for a selected bit is no longer a popular method for DRAM design.

1.2.3.2. Extended Data Out (EDO) DRAM

Another technique used to enhance the performance of the DRAM is the Extended Data Out (EDO). It is a modification of the FPM. A latch is added to the output buffer of an FPM DRAM. This modification allows a new read to be started while the output of the previous data is latched at the output buffer. This configuration increases the valid output data time, effectively decreasing the cycle time. The EDO DRAM provides about 10% to 15% decrease in access times of the FPM DRAM [23].

1.2.3.3. Burst EDO (BEDO) DRAM

The Pipelined Burst EDO (BEDO) is an improvement of the EDO DRAM. It improves the EDO by processing four read or write cycles in one "burst." This is needed when accessing a block of addresses. This mode is implemented by adding a two-bit internal address counter to keep track of the next address to be accessed. Only the first address to be accessed is decoded with the addition of the counter. The internal address counter provides the remaining addresses, thus reducing the number of clock cycles required to access the addressed bits. A pipelined stage is also included to replace the output latch in the EDO. This increases the access latency, but improves the bandwidth.

1.2.3.4. Synchronous DRAM (SDRAM)

The next major improvement to the performance of the DRAM was the design of Synchronous DRAMs (SDRAMs). In an Asynchronous DRAM, a read or write cycle begins access whenever the *RAS* and *CAS* signals are available. In SDRAMs, in addition to the *RAS* and *CAS* signals, a clock signal is used to determine when the read or write cycle begins. This allows the DRAM to be synchronized with the CPU clock. This improves the memory latency by making it a multiple of the CPU clock time as opposed to the asynchronous DRAM [7]. This reduces the skew and improves performance [23]. SDRAMs support the BEDO DRAM method. The SDRAM has a few different variations called the Double Data Rate (DDR). This improves the performance by

performing multiple reads or writes in one cycle by using both the rising and falling edges of the clock. Different versions are DDR1, DDR2, DDR3 (currently commercially available), and DDR4 (which is proposed for the future). DDR1 performs two operations in one cycle, while DDR2 and DDR3 perform 4 and 8 read or write operations per cycle.

1.3. Types of DRAM Topologies

The DRAM basic memory cell has seen two major circuit designs. The Three-Transistor design and the One Transistor-One Capacitor circuit topologies are the most common cell designs. Some other DRAM cell designs have been proposed and analyzed in [27, 18, 17, 3]. However, the 1T-1C is the conventional DRAM cell used today. The following section briefly describes the 3T and 1T-1C DRAM cells.

1.3.1. Three Transistor (3T)

The three-transistor DRAM (3T) circuit cell is shown in Figure 1.5. It was the cell design for the 1 Kb 1103 DRAM, the first ever fabricated DRAM chip by Intel. It was also used for some of the 4 Kb and 16 Kb DRAMs [15]. The 3T cell, as the name implies, consists of 3 transistors and two pairs of column and row lines. There is a write column line and read column line, and there is also a write row line and a read row line. The data to be stored is charged to the gate capacitance of transistor M2. Transistors M1 and M3 are used to control access to the data stored in M2. To write to the cell, the write rowline is set to "high", which switches M1 on. The data to be written flows from the write column line and either charges or discharges the input capacitance C_S of M2. After the input capacitance has been sufficiently charged or discharged, the write row line is then turned off. To read data from the cell, the read column line is precharged to a predetermined voltage. M3 is switched on by setting the read row line to high. When M3 is switched on, the read column line is either driven low or unchanged based on the data value stored on the input capacitance of M2. If the charge stored at the input capacitance is low (storing a logic "0"), then M2 is turned off and the read column line is unchanged. Otherwise transistor M2 is on and this will drive the read column line down indicating a stored logic "1."

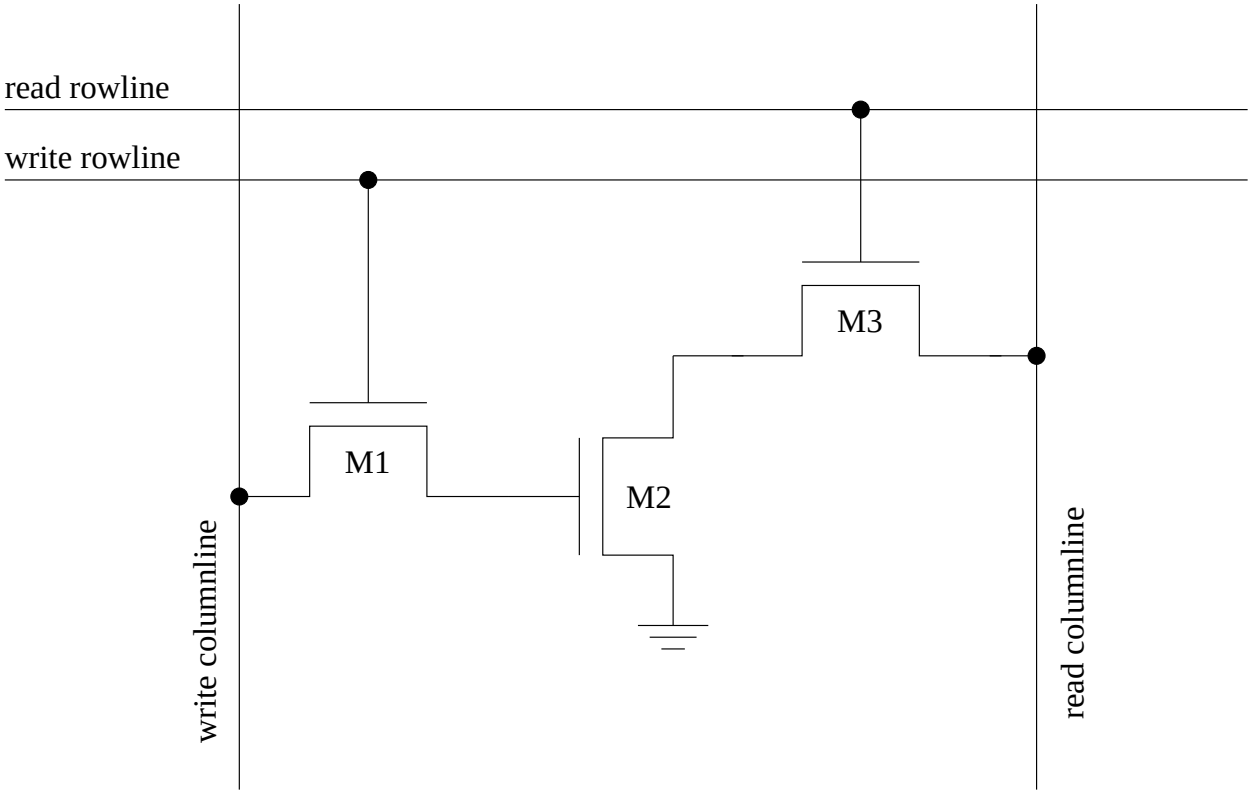


FIGURE 1.5. Circuit diagram of a 3T DRAM cell.

The value of a 3T DRAM cell is stored in the input capacitance of M2. Over time, the stored charge will dissipate as the capacitor discharges. Substrate leakage also causes the value to decay. The value has to be refreshed periodically to retain correct values. The disadvantage of the 3T DRAM is that it uses two pairs of column and row lines and hence requires a larger layout area. The 1T-1C DRAM, however, with only 1 transistor uses smaller layout area. With the trend to optimize layout area, the 1T-1C quickly replaced the 3T cell and became the more conventional cell used in design. The 3T DRAM cell is however still used in some application specific integrated circuits (ASICs) [32] as it can still be packed more densely than the SRAM, but does not require as much processing and is faster than the 1T-1C DRAM.

1.3.2. One-Transistor/One Capacitor (1T-1C)

The 1T-1C DRAM technology grew popular quickly because of its ability to be packed more densely. It requires only 1 transistor and a capacitor and thus uses fewer components than the 3T

cell. A wordline and bitline are used to store and access data from the cell. Figure 1.3 shows the circuit diagram of a 1T-1C DRAM.

To write to the cell, the wordline is set to "high", which switches the transistor on. With the transistor on, the bitline, which is set to the value to be written, either charges or discharges the capacitor to write a logic "1" or a logic "0", respectively. To read data from the cell, the bitlines are precharged to a predetermined voltage (usually $(V_{DD}/2)$), and then the wordline is set to "high" to switch on the transistor. Charge sharing occurs between the bitline and the capacitor. Charge will either flow from the capacitor to the bitline or vice versa depending on the value of the stored data. If a "1" was stored, charge will flow from the capacitor to the bitline and the reverse will occur if a "0" was stored. The charge shared appears as a very small voltage change on the bitline. A sense amplifier is required to detect the voltage change of the bitline [27, 25, 32]. The effect of the charge sharing between the transistor cell and the bitline capacitance makes the read process destructive. When charge is shared between the bitline and the capacitor, the data stored on the capacitor is modified and might be corrupted. If the cell stored a logic "1", it loses charge to the bitline and may not have enough charge to be considered as a logic "1." The reverse occurs if the cell stored a logic "0." Thus, the sense amplifier also helps refresh the value of the accessed cell. After amplifying the signal of the accessed cell, the value read is rewritten back to maintain its validity. The charge stored in the capacitor also dissipates over time. Hence the cell must also be refreshed periodically to retain correct values. This can be done by performing a read operation.

1.4. Effect of Process Variation on Sense Amplifiers in DRAM

Manufacturing process variability induces variations on device parameters. These are due to several sources, including the following [19]:

- Ion implantations
- Random doping
- Chemical vapor deposition (CVD)
- Sub-wavelength lithography
- Thermal processes
- Lens aberrations

- Materials flow

Non uniform conditions during random doping and diffusion of impurities could lead to variations in process parameters, which include impurity concentrations, diffusion depths, and gate-oxide thickness. These variations lead to variation of device parameters, including the following:

- Supply voltage
- Threshold voltages
- Oxide thickness
- Transconductance
- Channel lengths
- Channel widths
- Source and drain doping concentration

Threshold variations are affected by different dopant concentrations and gate-oxide thicknesses. The channel length and channel width variation results mainly from lithographic process and etching dependencies, while transconductance variations are due to gate-oxide thickness. These device parameters cause transistors to have different electrical characteristics including the following:

- Driving current
- Threshold voltage
- Subthreshold leakage
- Sheet resistance
- Source/drain resistance

Thus, two transistors which are fabricated on the same die, which were designed to be identical may end up having different device parameters. The voltage supply of the device is also subject to variability.

Process variations can be grouped under different classifications. They could be inter-die or intra die, random or systematic, correlated or uncorrelated, and spatial or temporal. Inter-die variations are those that occur between devices on the same die while inter-die variations are those that occur on devices fabricated on different dies, wafers, lots or manufacturing lines. Random variations include effects from random dopant fluctuations affecting the threshold level and effective

lengths while systematic variations are due to the physical layout patterns of the device. Parameter variations could be correlated or uncorrelated. For example, variations in effective length are not correlated to variations in threshold voltage, but variations in threshold voltage are directly affected by variations in oxide thickness.

Process variation has become a significant problem in nanoscale CMOS technology based semiconductor chips. As the process technology used in DRAMs and in semiconductor chips as a whole continues to scale into the ultra nanoscale region, the effects of process variation continue to be enhanced and become more significant. Limitations of the fabrication processes and the sheer scale of the physical device (current trend is towards 22 nm) make it impossible to have perfectly matched devices. The major factors for process variation are varying levels of dopant concentration and channel dimensions such as the length, width, and oxide thickness for the devices. Such variations lead to different threshold voltage and affect the driving capability of the transistor [5, 14].

A sense amplifier is used to amplify minimal differential voltages to full scale logic "1" or "0" and relies on matched transistors to operate effectively. Designing a sense amplifier for high performance is crucial to the performance of DRAMs. Sense amplifiers need matched transistors for optimal performance. The supply voltage, length, width, threshold voltage (V_{th}), and capacitance of the transistors must be matched accurately for optimal performance. However, with increasing process variations, this continues to pose a great challenge. With the scaling of technology increasing, the supply voltage is reduced to keep power consumption low. This implies that the voltage change detected in a read operation of the DRAM also reduces making the sense amplifier more sensitive to variations. As the capacity of DRAMs increase, there are more memory cells in the bitline connected to a sense amplifier. This increases the bitline capacitance and thus leads to a reduction of voltage change detected during read operations. This increase in bitline capacitance from the different memory cells invariably increases the capacitance mismatch [8]. CMOS technologies in the nanometer region are very sensitive to channel length variations. As the minimum length scales below 50 nm, the transistors become sensitive to channel length variations affecting the driving capabilities, which is crucial to the sensing operation of the sense amplifiers. Also

increasing variations in threshold voltage are expected as device sizes shrinks. These are the main factors affecting the performance of sense amplifiers. Hence, the accuracy of the performance of the sense amplifier must be increased to accommodate the reduction in voltage change detected resulting from reduced voltage supply and increased bitline capacitance.

The scaling trend of technology is expected to decrease to below 22 nm. As process technology improves, the tolerance of mismatch by the sense amplifier diminishes and thus aggravates its effect [26]. Hence the effects of mismatch of process variations must be minimized by designing sense amplifiers, which are less susceptible to these effects. This thesis contributes to the analysis and study of process variation tolerant designs of sense amplifiers. The aim is to design sense amplifiers which are robust enough to accommodate these parameter variations.

1.5. Novel Contributions of this Thesis

The novel contributions of this thesis are described in this section. This thesis presents a study and analysis of the effects of process variation on the performance of three common sense amplifier designs and provides a method to choose parameter values that minimize the impact of these effects. This is done through the simulation of a 1-bit memory cell, a reference cell and sense amplifiers using analog simulators with a 45 nm process technology. The major contribution of this research is providing a design process to mitigate the effects of process variations on the performance of sense amplifiers for DRAMs. There has been extensive research to improve the performance of sense amplifiers, specifically to improves its tolerance towards process variation. Most of these techniques involve adding components to the basic circuit of conventional sense amplifiers. The approach and contribution of this work is providing methodology of identifying parameters and figures of merit which are severely affected by process variation. The final parameters for the sense amplifier design are chosen to provide the best tolerance for process variation while optimizing the performance of a figure of merit.

1.6. Organization of this Thesis

The rest of this thesis is organized as follows. The next section gives a brief overview of the terms and acronyms used in this work. Chapter 2 discusses some of the related research on performance of sense amplifiers with respect to process variation. The basic design and functionality of

the sense amplifiers studied are presented in Chapter 3. Chapters 4 and 5 discuss the methodology of improving and optimizing the sense amplifier design for power dissipation using dual oxide and dual threshold voltage. Conclusions and future work are presented in Chapter 6.

1.7. List and Definitions of Acronyms and Symbols Used

This section lists all acronyms and symbols used throughout this thesis. Most of the acronyms are typically used in DRAM technology, however their definitions are described here for completeness.

Acronym	Definition
$1T-1C$	The one transistor-one capacitor DRAM cell
$3T$	The three transistor DRAM cell
BL	Bitlines
CAS	Column access strobe signal
CE	Chip enable Signal
C_S	Memory cell capacitance
EDO	Extended data Out
DDR	Double data Rate
FoM	Figure of merit
FPM	Fast page Mode
L_{eff}	Effective gate length of a transistor
L_n	Gate length of an NMOS transistor
L_p	Gate length of a PMOS transistor
PRE	Precharge and equalization signal
RAS	Row access strobe signal
R/W	Read/write enable signal
SE	Sense amplifier enable signal
T_{ox_n}	Oxide thickness of the NMOS transistor
T_{ox_p}	Oxide thickness of the PMOS transistor

t_{AC}	Access time
t_{AW}	Write delay time
t_{RC}	Read cycle time
t_{WC}	Write cycle time
t_{WP}	Write pulse width
V_{DD}	Voltage supply of the circuit
V_{CS}	Stored voltage of the memory cell
V_{th_n}	Threshold voltage of the NMOS transistor
V_{th_p}	Threshold voltage of the PMOS transistor
W_n	Gate width of an NMOS transistor
W_P	Gate width of a PMOS transistor

CHAPTER 2

RELATED PRIOR RESEARCH

The existing literature is rich in research on the impact and performance analysis of sense amplifiers towards the general performance of dynamic access random memory (DRAMs). Analysis of trends and performance of modern DRAMs has been undertaken by various researchers [7, 21, 16]. Some research has also been performed to analyze the operating parameters and key characteristics of the sense amplifier operation. In this chapter, a selected number of basic sense amplifier topologies are examined along with some of the techniques and methods employed to mitigate the effect of process mismatch in sense amplifiers.

2.1. Different Sense Amplifier Topologies

This section gives a review of different sense amplifier topologies. Two common categories for sense amplifiers include the following [5, 22, 31]:

i. Voltage mode: Voltage mode sense amplifiers amplify a small differential voltage.

ii. Current mode: Current mode sense amplifiers amplify a differential current.

There are various sense amplifier designs in use. Based on the application scenario, the sense amplifier could range from a basic gated flip-flop circuit to a full latch cross coupled sense amplifier [27, 25, 13, 22, 14]. Voltage based sense amplifiers are the most common sense amplifiers used in DRAM designs [22, 31]. Current sense amplifiers are more useful in larger memory designs because they have a lower input capacitance. This provides an advantage over voltage based sense amplifiers. The lower voltage swings result in a faster sensing operation [2, 4]. The following sections describe some common voltage and current based sense amplifiers.

2.1.1. Voltage Based Sense Amplifier

Voltage based sense amplifiers are typically used in DRAM circuit designs. The most common form is the full latch cross coupled voltage sense amplifier. Figure 2.1 shows the cross-coupled

sense amplifier circuit. Figures 2.2 and 2.3 show the full latch cross coupled amplifier, and some of its variants topologies [25, 13, 24, 31, 4]. Two inverters cross coupled together to provide a positive feed back form of the core of this sense amplifier. The sense amplifiers is basically a differential couple with two bit lines. The cross-coupled sense amplifier shown in figure 2.1 does not have inverters cross coupled but has a two PMOS-transistor cross coupling arrangement. In these sense amplifiers, during a read or refresh, the bit lines which are precharged are amplified to a full logic "1" or "0" values. The amplified signal requires discharging and charging of the bitlines. This poses a problem as the size of memory design increases. The bitline capacitance of the memory cells will make it difficult for the bitlines to develop a minimum signal for detection for the bitlines to be quickly charged or discharged [4]. The disadvantage of the high input capacitance is mitigated by the current-based sense amplifiers. The cross-coupled sense amplifiers and the full-latch cross-coupled sense amplifiers are briefly described below.

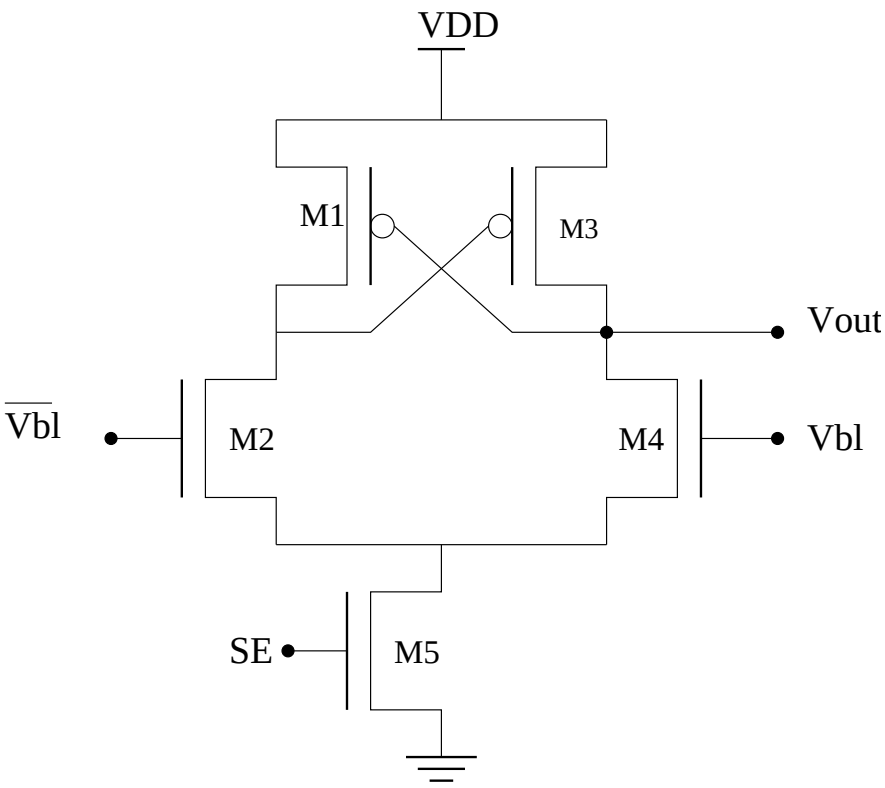


FIGURE 2.1. The cross coupled sense amplifier circuit with a total five transistors.

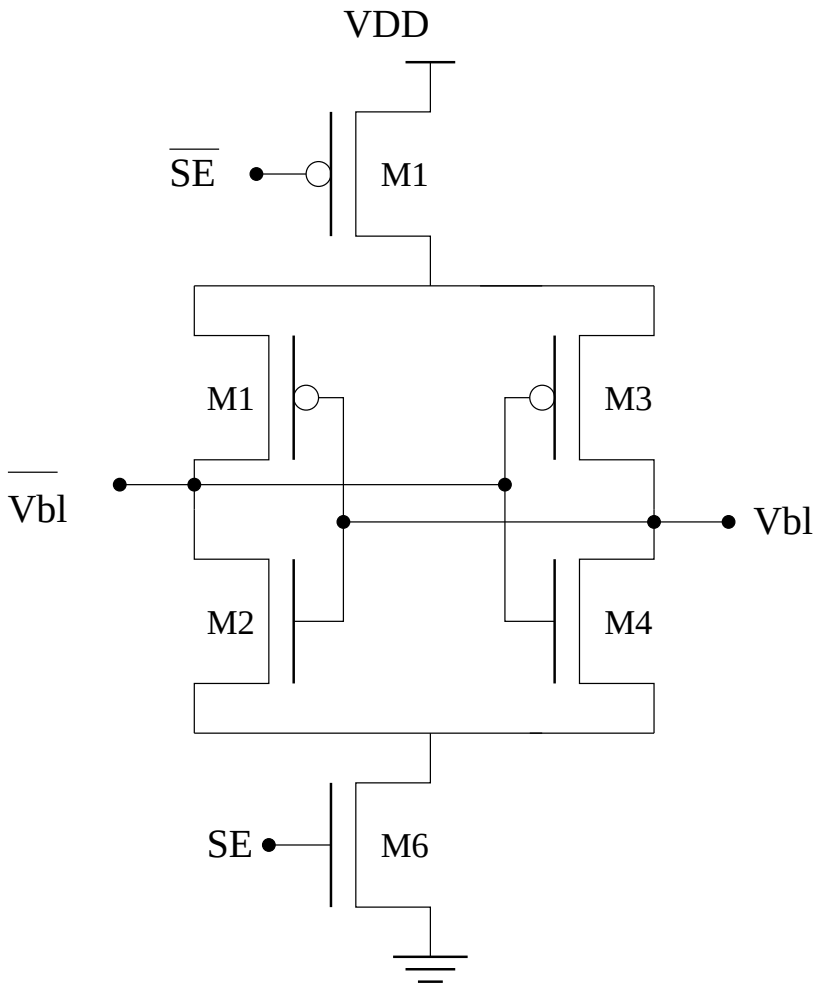


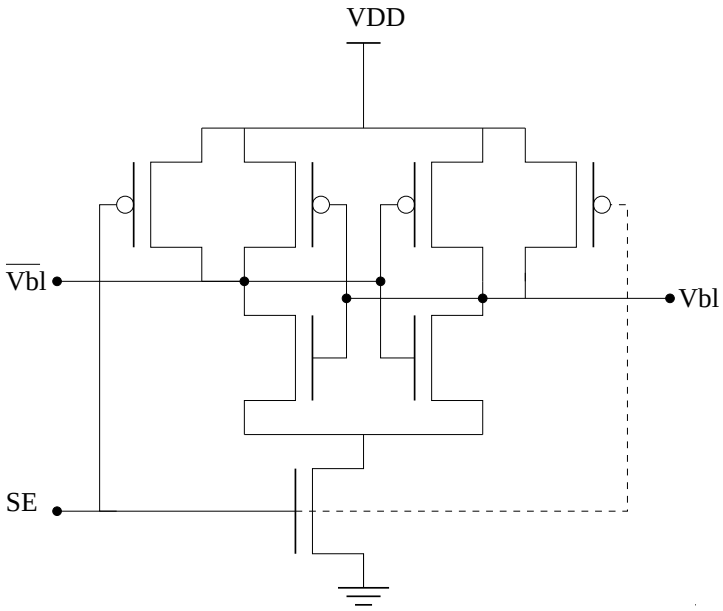
FIGURE 2.2. The full-latch cross-coupled sense amplifier circuit.

2.1.1.1. Cross-Coupled Sense Amplifier Circuit

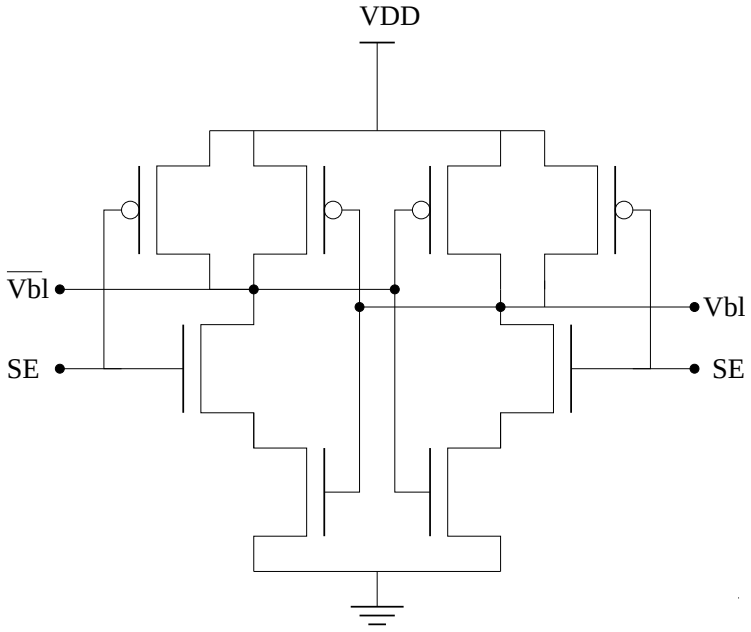
The cross coupled sense amplifier circuit is shown in figure 2.1. This circuit consists of two cross-coupled PMOS transistors (M1 and M3) and three NMOS transistors (M2, M4 and M5) [1]. The sense amplifier is activated by transistor M5, the inputs from BL and $\overline{BL}$ are connected to the gates of M2 and M4, while the PMOS transistors act as a positive feedback from the output.

2.1.1.2. Full-Latch Cross-Coupled Sense Amplifier Circuit

The full-latch cross-coupled sense amplifier (also referred to as simply full latch) consists of two cross coupled inverters. Transistors M1-M4 form a pair of cross coupled latch inverters. The coupled inverters are set to the transient region by a precharge circuit. Transistors M5 and M6 are



(a) Full Latch Type 1



(b) Full Latch Type 2

FIGURE 2.3. Different variations of the full-latch cross-coupled sense amplifier circuits.

used to switch on the sense amplifier. The two cross-coupled inverters provide a positive feedback and are connected to the bitlines BL and $\overline{BL}$. The circuit diagram is shown in figure 2.2. Some variants of the common full-latch sense amplifiers are shown in figure 2.3.

2.1.2. Current Based Sense Amplifier

Current sense-amplifier circuit operation is two fold: to detect and to amplify differential currents between its input. Current sense amplifiers have an advantage over voltage based sense amplifiers because they have a lower input capacitance. This is useful in increasing the sensing speed of the sense amplifier circuits. Three sense amplifiers, the current mirror sense amplifier [1, 13] and clamped bitline sense amplifier [24, 4] are briefly described below.

2.1.2.1. Current-Mirror Sense Amplifier

The circuit of the current-mirror sense amplifier is shown in figure 2.4. The circuit consists of an NMOS differential amplifier with a simple PMOS mirror as its load. The PMOS mirror provides a stable bias current and allows for a large voltage swing. The signals from the BL and $\overline{BL}$ are the inputs through the gates of transistor M2 and M4. Both bitlines are initially precharged to a predetermined voltage. After the word line goes high and there is sufficient voltage change on the bitline, the amplifier is activated by raising the SE signal. It has the disadvantage of requiring a high current supply.

2.1.2.2. Clamped-Bitline Current Sense Amplifiers

Another configuration of the current based sense amplifier is the clamped bitline amplifier [2, 24, 4]. Figure 2.5 shows the clamped bitline current sense amplifiers. This configuration clamps the bitline voltage to a reference voltage. It also uses an internal sense amplifier which receives the bitline current. This eliminates the charging and discharging of the bitline capacitance. This dramatically increases the sense speed of the amplifier and improves its power consumption.

2.2. Related Sense Amplifier Circuit Design Approaches

The operation of a sense amplifier is significantly affected by mismatch of transistors in the circuit. The effect of mismatch of transistors caused by process variation continues to increase as process technology continues to decrease below 50 nm [21]. Manufacturing processes such as

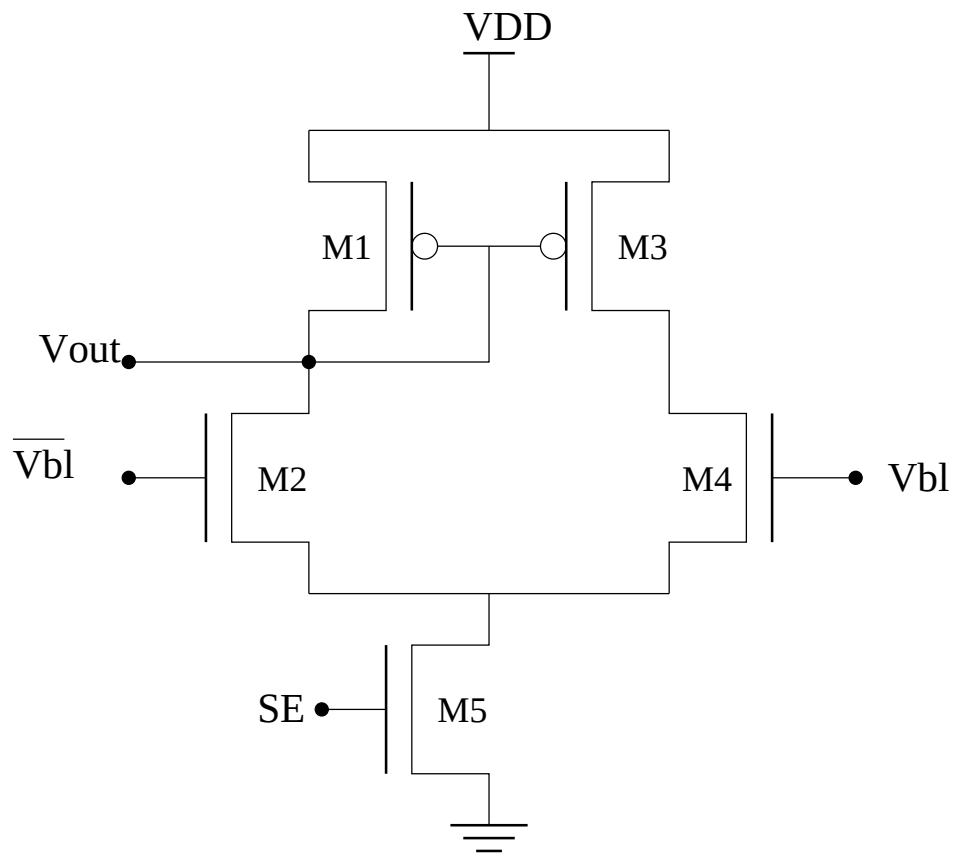


FIGURE 2.4. The current-mirror sense amplifier circuit.

accurate random doping of silicon, and effective sizing are becoming very difficult to achieve as technology continues to scale. The miniature size of transistors aggravates the effect of process tolerance. Mueller *et al.*, detail the trend of technology scaling for the DRAM to 40 nm and operating voltage V_{DD} of sub 1 V. The scaled down operating voltage also reduces the voltage signal change on the bitlines sensed by the sense amplifier. An effort to keep the cell capacitance constant is made. Some of the parameters that affect the correct operation of the sense amplifier the most are the bitline capacitance (C_{BL}), effective length of the transistor L_{eff}, and threshold voltage (V_{th}) [5, 16].

There have been several studies on the analysis and effects of parameter and process variations on the performance of sense amplifiers. In [1], Sherwin presents a study on the performance of several sense amplifiers when the supply voltage is varied. The voltage gain, voltage signal, current supply margin, and noise margin were among some of the performance measures recorded.

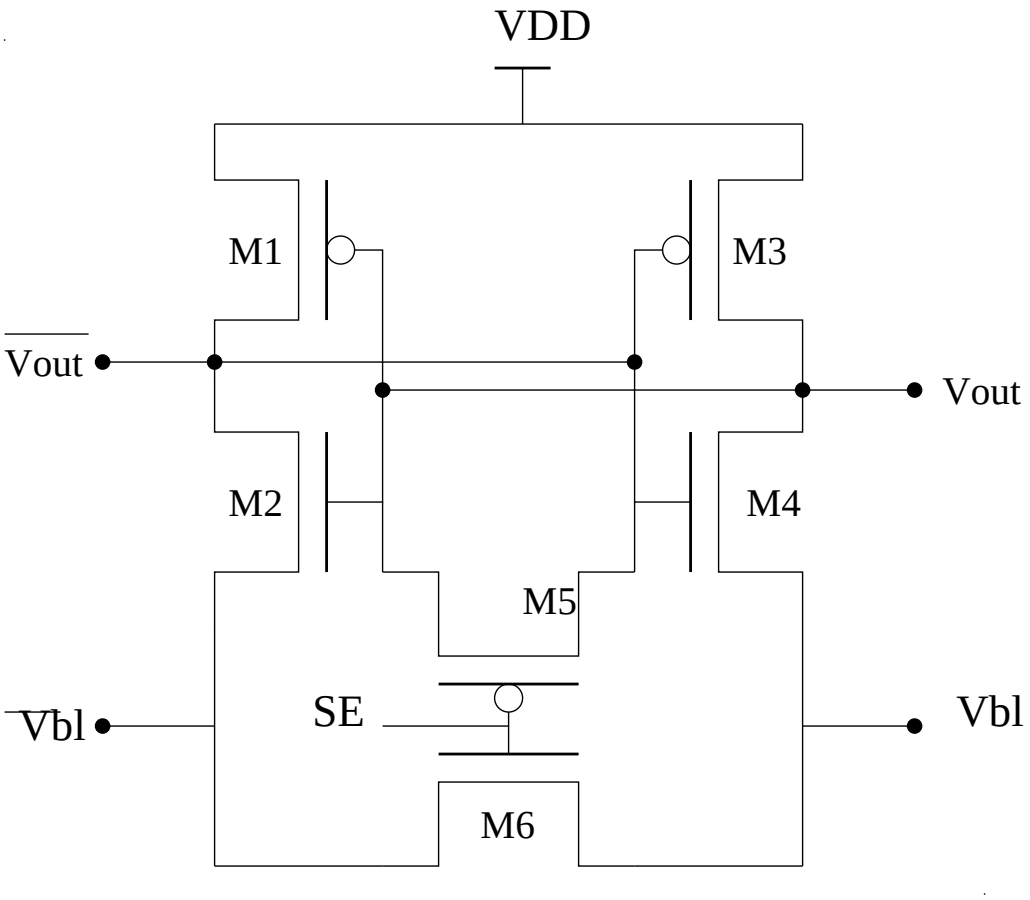


FIGURE 2.5. The clamped-bitline sense amplifier circuit.

Simulation results showed that although each sense amplifier had different advantages and disadvantages for each performance measure, a reduction in voltage supply leads to an increased sensing time. This is one of the design conflicts encountered, especially with scaling technology. One of the goals of CMOS design is to reduce the power consumption. Reducing voltage supply or V_{DD} is an obvious and easy way to accomplish this. However as shown in [1], decreasing V_{DD} increases the sensing time of the sense amplifier, which decreases the speed performance of the DRAM. The research in [29, 16] also studied the effect of sense amplifier performance on minimum signal margins. In determining the signal margin of sense amplifiers [16], Laurent also shows the effect of process variations on the signal margin. The variations of transconductance and V_{th} are shown to have a more significant effect on the signal margin than the bitline capacitance.

Some studies proffer solutions by modifying the sense amplifier circuit to mitigate some of the parameters that significantly affect the operation of the sense amplifier [4, 5, 6]. The sensing time or sense delay is the time required for sufficient voltage sharing to be sensed by the sense amplifier. It constitutes a major part of the sensing time and is a significant factor to the overall speed of the sense amplifier. An approach to increase the speed of a low V_{DD} amplifier was proposed by [6]. A modification is made to the sense amplifier circuit to provide a boost capacitance, which increases the speed and provides a larger voltage difference after transfer.

Choudhary *et al.* in [5] also modify the full latch voltage sense amplifier to provide a process variation tolerant sensing scheme. In their technique, one of the major factors leading to an incorrect sensing operation was identified as the voltage offset, which must be smaller than the change in voltage signal on the bitline. The process parameters that significantly affect the voltage offset are V_{th}, and the length and width of the transistors. Choudhary *et al.* add a training capacitor to mitigate the effect of process variations. Their results show an improved performance sense amplifier with greater process variation.

Another factor that is affected by process mismatch and variation is the input voltage offset. The input voltage offset also affects the minimum signal margin. Some studies have been conducted to model techniques that identify factors that create offset and modify parameters to mitigate this offset [11, 26]. In [11], two major sources of offset were identified in bitline sense amplifiers: the variance in the transistor parameters and specifications such as gate length and threshold voltage. This factor is increased by technology scaling trends which results in low process tolerance. The second source is from cross coupling noise resulting from closely packed memory cell. This will also increase as the trend to densely pack more memory cells increases. Singh *et al.* present an analysis in [26] identifying intrinsic and extrinsic factors of offset in the NMOS and PMOS transistors of the latch amplifier. By modifying the size of the PMOS transistors, the rise time of the amplifier enable signal is increased, effectively reducing or completely eliminating the offset voltage mismatch.

2.3. Summary of Prior Research

Table 2.1 shows a summary of the related research. The impact of process variations of process parameters such as length, width, and V_{th} are the most severe factors affecting the performance of sense amplifiers for DRAMs. With the current trend of technology scaling and specifications in nanoscale, the device tolerance for variation will continue to decrease. Methods presented in [6, 5, 11], which modify the circuit design of the sense amplifier by adding components to boost performance, increase the circuit complexity of the conventional sense amplifier. Although the modification of the circuit design improves the yield and performance, the added element increases the complexity of the circuit and burdens the constraint of dense packing. In [11], the modified sense amplifier is however able to support 44% more memory cells than the conventional sense amplifier. In this thesis, we propose a design process that maintains the circuitry of the sense amplifier and is still tolerant towards process variations in NanoCMOS circuits.

TABLE 2.1. Summary of Research

This section provides a summary of related research

Research	Parameter	Feature	Approach	Result Improvement
Sherwin[1]	V_{DD}	Voltage gain	Physical measurements	-
Chow [6]	C_{BL}	Sense speed	Spice simulations	Sense speed - 40%
Laurent[16]	C_{BL}, V_{th}, β	Signal margin	Spice simulations	-
Vollrath[29]	C_{BL}, V_{th}, β	Signal margin	Physical measurement	-
Choudhary [5]	V_{th}, L, W	Yield	Monte Carlo analysis	Improved yield
Hong [11]	-	Sense scheme	Spice simulations	Improved sense amplifier cell ratio
Singh [26]	-	SE rise time	Spice simulations	Eliminated offset voltage
This research	Dual-V_{th}	Process variability	Optimization	Sense delay - 80.2% and Sense Margin - 61.9%

CHAPTER 3

DESIGN AND SIMULATION OF THE SENSE AMPLIFIER CIRCUITS

The major function of the sense amplifier in DRAM memory is to amplify the minimal voltage change detected and to refresh the data in the bit cell. In the 1T-1C DRAM, when cells are accessed, there is a very small change detected on the bitlines. The capacitor is used as the storage mechanism in the 1T-1C cell and the stored charge gradually dissipates, hence the sense amplifier is periodically used to refresh the data values in the memory bit cell of the DRAM. The sense amplifier also plays a significant role in the overall speed of the DRAM. The speed is affected by how fast the sense amplifier can sense the voltage change on the bitline. As was discussed in chapter 2, the major classifications for sense amplifiers are the voltage based sense amplifiers and current based sense amplifiers. In this chapter, the core design of a conventional sense amplifier is discussed. The full latch voltage sense amplifiers is one of the most common sense amplifiers in use. The basic full latch sense amplifier is made from two cross coupled inverters, as shown in figure 3.1.

The pair of NMOS and PMOS transistors make up the Nsense and Psense amplifiers. The latch amplifier is preferred because of the positive feedback voltage. The output of the sense amplifier is connected to a pair of bitlines, BL and $\overline{BL}$. To speed up the sensing operation and reduce the voltage swing and thus power consumption, the bitlines are precharged to a predetermined voltage, which is usually $(V_{DD}/2)$ [25, 15]. The equilibration and bias circuit, also known as the precharge circuit, functions to ensure this. The precharge circuit consists of one or more NMOS transitors connected between the bitlines. NMOS transistors are used because of their higher driving capability [15]. Figure 3.2 shows the circuit schematic of the precharge circuit. The sources of transistors M1 and M2 are connected to the precharge voltage $(V_{DD}/2)$. The gates of the transistors are connected to the *PRE* signal. The *PRE* signal goes high just before the WL is activated to ensure both bitlines are precharged.

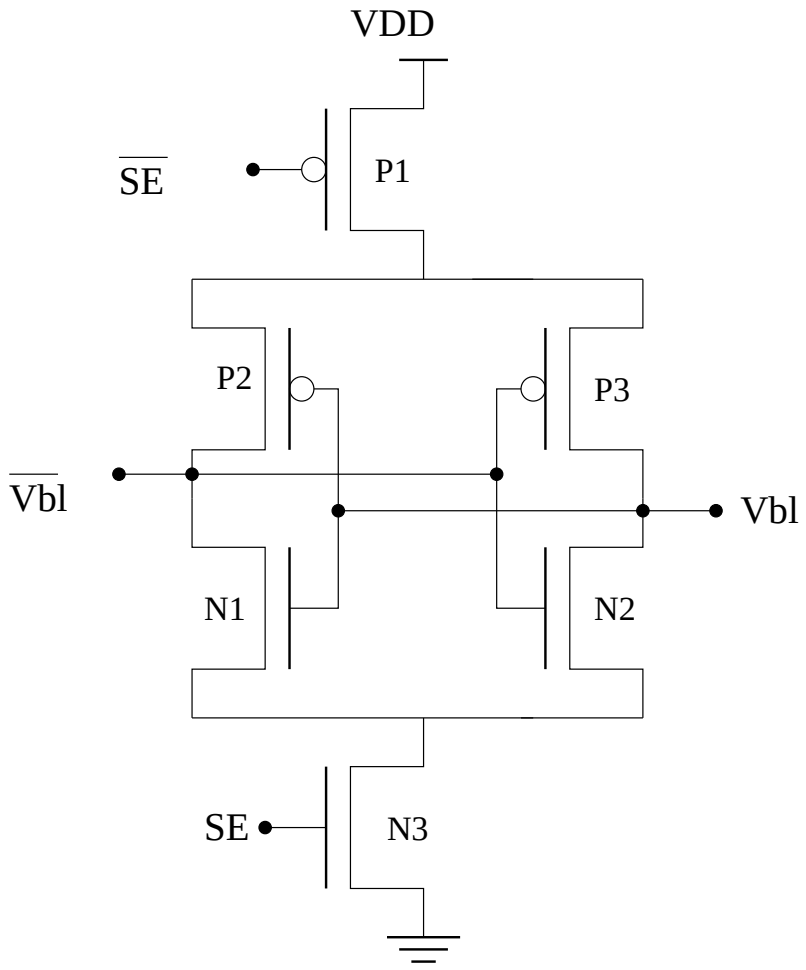


FIGURE 3.1. Conventional Cross Coupled Latched Sense Amplifier: N1-N3 transistors form the Nsense amplifiers while P1-P3 form the Psense amplifiers

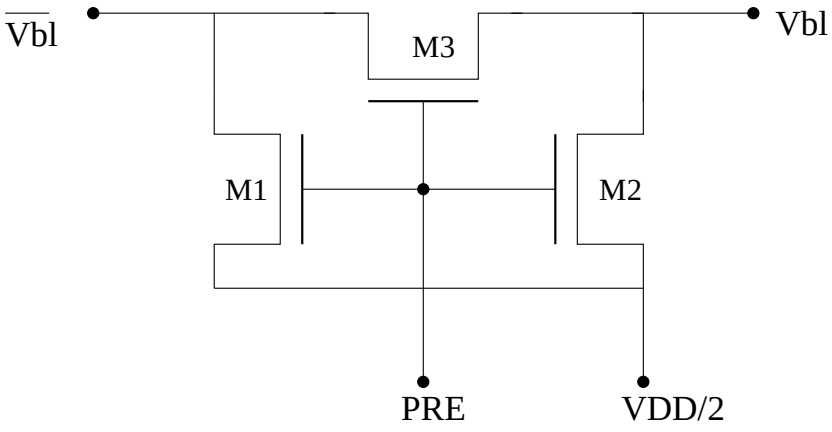


FIGURE 3.2. Precharge circuit schematic

The sensing operation, which is the detection of signal change and amplification to the full logic voltage is described below [15]. It is simulated with the 1T-1C DRAM cell. The 1T-1C DRAM cell has become the widely conventional DRAM design. We assume a read operation on the bit cell with a logic "0" stored with BL connected to the bit cell and $\overline{BL}$ connected to the dummy cell. After the wordline is raised, charge sharing begins to occur between the bit cell and BL. After sufficient voltage margin has developed on BL, the sense amplifier is activated. When the sense amplifier is activated, the Nsense amplifier is pulled to ground through transistor N3. The gate of transistor N2 which is connected to $\overline{BL}$ begins to conduct first in the subthreshold and then in the saturation region. This causes BL to discharge to ground through transistor N3. Because BL is eventually discharged, transistor N1 never turns on. A little bit after the Nsense amplifier turns on, the Psense amplifier is also activated and brought to high through transistion P1. BL is connected to the gate of transistor P2, and at this point is low enough making P1 able to conduct in the subthreshold state. As BL discharges closer to ground, $\overline{BL}$ is charged to V_{DD} through transistor P2. Again because the gate of P3 is connected to $\overline{BL}$ and remains high, P3 is not turned on [13].

For the sensing operation described above to work correctly, there must be sufficient voltage change on both BL and $\overline{BL}$. When reading a value from a bit cell, the direction of charge flow depends on the value in the bit cell. When reading a value of logic "1", the charge flows from the storage capacitor to the bitline capacitor and the reverse occurs when reading a value of logic "0". A positive voltage gain on the bitline signifies a logic "1" value read, while a negative charge gain signifies a logic "0" value .

The resulting voltage shared is expressed as follows:

$$\Delta V = \frac{C_S}{C_S + C_{BL}} \left(V_{CS} - \frac{V_{DD}}{2} \right). \tag{1}$$

where C_S and C_{BL} are the cell and bitline capacitances respectively, V_{CS} is voltage stored in the cell and V_{DD} is the operating voltage.

Usually $C_{BL} \gg C_S$, thus equation 1 can be reduced to

$$\Delta V \cong \frac{C_S}{C_{BL}} \left(V_{CS} - \frac{V_{DD}}{2} \right) \tag{2}$$

When the bit cell value is 1, $V_{CS} = V_{DD} - V_{th}$ and

$$\Delta V(1) \cong \frac{C_S}{C_{BL}} \left(\frac{V_{DD}}{2} - V_{th} \right) \tag{3}$$

When the bit cell value is 0, $V_{CS} = 0$ and

$$\Delta V(0) \cong -\frac{C_S}{C_{BL}} \left(\frac{V_{DD}}{2} \right) \tag{4}$$

Typically, ΔV is usually very small. It is about 30 $\sim$ 150 mV.

The read operation of the 1T-1C DRAM is *"destructive"*, hence a mechanism is required to refresh a bit cell after a read operation. The sense amplifier also functions to refresh the bit cell after a read operation. During the sensing operation described above, the wordline remains high, hence, after the BL has be been charged or discharged to V_{DD} or ground, the fresh value is written back to the memory bit cell. The importance of a correct sense operation is seen here as a wrong amplification could lead to an error in the data refreshed.

3.1. Characterization of Sense Amplifier

This section presents the FoM used in characterizing the performance of the sense amplifier. In characterizing the circuit, simulations were carried to quantify some performance measures for the baseline design. The following figures of merit were chosen: the precharge and voltage equalization time, the power consumption for a read and write cycle, the sense delay and sense margin. The FoMs are briefly described below.

3.1.1. Precharge and Voltage Equalization Time

The precharge and voltage equalization time is the time required for the bitlines BL and $\overline{BL}$ to both be equally charged to $(V_{DD}/2)$. The bitlines are precharged to $(V_{DD}/2)$ to reduce the voltage swing of the sense operation. It is significant in quantifying the overall speed of the sense amplifier, because the bitlines have to be set up for a sensing operation. For every read access, the row or page is opened. To open up a page, the bitlines have to be precharged. Some DRAM operation modes limit the effect of this by keeping a row opened for multiple reads with only a column change address. The main factor which affects the precharge time is the capacitance of the bitlines BL and $\overline{BL}$.

3.1.2. Power Consumption

Power consumption is the measure of the average power consumption of the sense amplifier circuit. With the scaling of technology, the circuit voltage supply decreases which directly reduces the power consumption. On the other hand, the increasing ability to pack more cells on the same chip area increases the power consumption per area. Two components for power consumption in CMOS circuits are the dynamic and static power. The dynamic power is due to the switching activity of the transistors, short circuit current, and recently gate current leakage due to reduced oxide thickness. The static power is due to subthreshold current, gate leakage, and reversed biased diodes. Both dynamic and static power are measured. The formula for dynamic power in a CMOS circuit is given as follows [25]:

$$P_{dyn} = \alpha f C V_{DD}^2, \tag{5}$$

where α is the activity factor and C is the capacitive switching load. The static power dissipation is given as follows:

$$P_s = I_s * V_{DD} \tag{6}$$

With the trend of scaling, static power consumption continues to be an important factor. The decrease in oxide thickness increases the gate leakage current and thus increases power consumption. For this study, the average power consumed measured is from simulation runs only. The power dissipation due to gate current leakage is also taken into consideration. The equation used to capture this for each transistor in the circuit is:

$$P_{avg} = (V_d * I_d) + (V_g * I_g), \tag{7}$$

where V_d and I_d are the drain voltage and current and V_g and I_g are the gate voltage and current.

3.1.3. Sense Delay

Sense delay is the time required for a sufficient amount of voltage sharing to produce the minimal voltage change that can be detected. It is perhaps one of the most important characteristics of the sense amplifier as it constitutes a significant portion of the memory access time [8]. The

voltage signal change that develops on the bitlines is very small and there is some minimal signal that must be developed for the sense amplifier to operate correctly. When the wordline is activated, there is a delay before the sense amplifier is activated to allow for the minimum signal change to appear. The sense delay is affected by the supply voltage. A higher supply voltage will result in a faster sense delay [1], however one of the major goals of the sense amplifier design is to reduce power consumption. So reducing the supply voltage for a lower power consumption presents a design conflict. The power-delay product is often used to optimize the power consumption while maximizing the delay. For this work, the sense delay is measured from circuit simulation. The delay is the time required for the bitline to develop 90% of the sense margin. The sense delay is different based on where the cell data value sensed is a "1" (high) or "0" (low).

3.1.4. Sense Margin

The sense margin is the minimum amount of voltage change that can be correctly detected by the sense amplifier. When voltage is shared between the memory cell and the bitlines, the sense amplifier detects the change on the bitline to be amplified to a logic "1" or a "0" value. It is one of the important measurements for the operation of the sense amplifier. If there is insufficient signal change on the bitline, the sense amplifier may not correctly interpret the signal on the bitlines. The equation for the sense margin is given in 3 and 4. The sense margin is different based on whether the cell data value sensed is a "1" (high) or "0" (low). It is higher when a logic "1" is being read. The sense margin is measured from the simulations because the equations do not take into account the effects of process variations, capacitive asymmetry, and cross coupling noise which produce an offset [11, 9]. The input offsets impacts the minimal voltage required to correctly interpret the data from the memory cell.

3.2. Full Latch Cross Couple Sense Amplifier Design

The cross coupled latch sense amplifier is shown below in figure 3.3. [25]

It is formed by two cross coupled inverters. Transistors M1-M4 form a cross coupled latch inverter. The NMOS transistors M8-M9 are used to precharge and equalize the bitlines to $V_{DD}/2$.

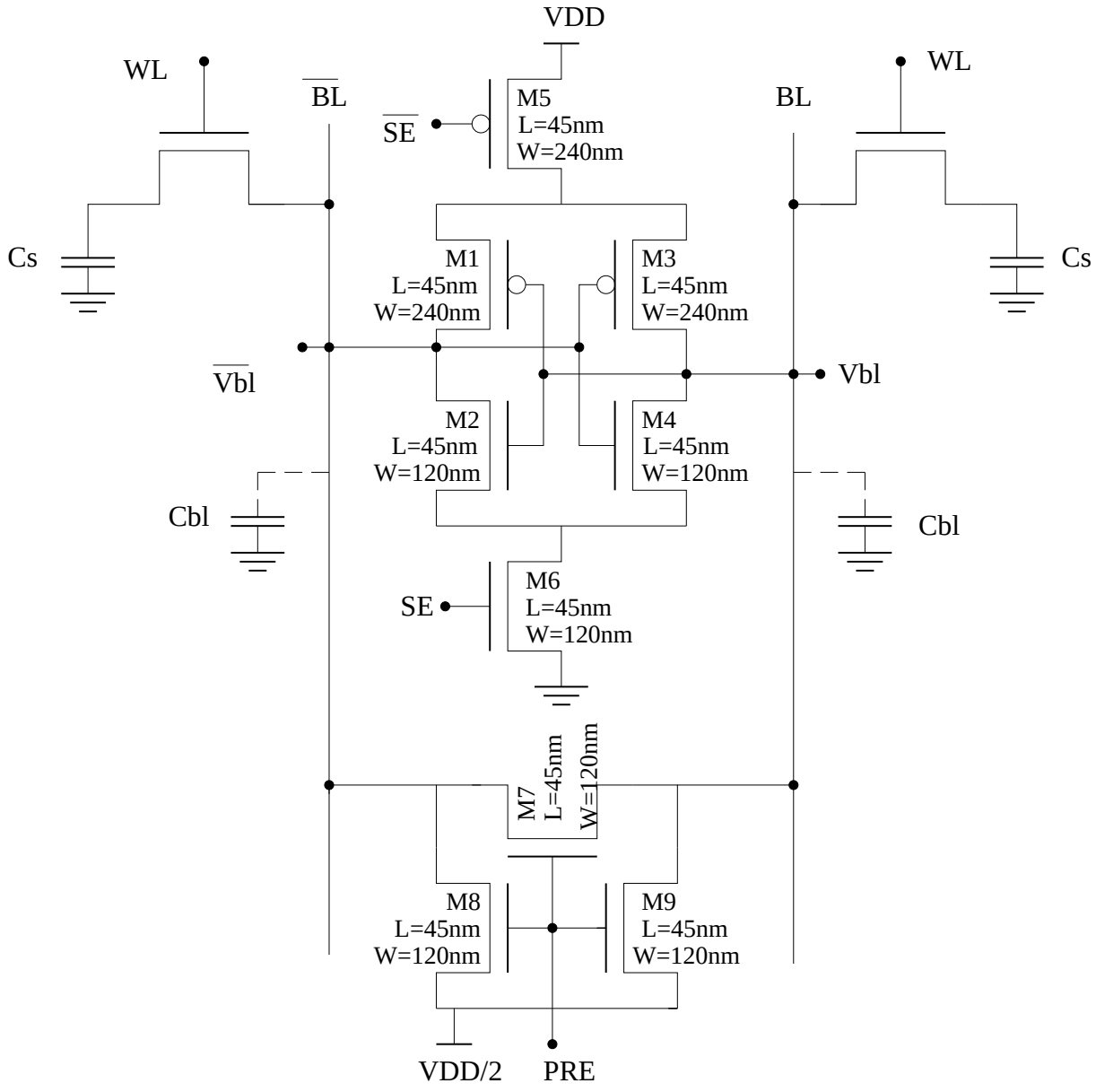


FIGURE 3.3. Circuit diagram of a cross couple latch sense amplifier

Transistor M5 and M6 are used to switch on the sense amplifier. The two cross coupled inverters provide a positive feedback and are connected to the bitlines BL and $\overline{BL}$.

The operation of the sense amplifier is as follows:

i. The precharge circuit is activated and turned on by the PRE signal, which drives both bitlines equal to $(V_{DD}/2)$. The precharge circuit is then turned off and the bitlines are left floating at $(V_{DD}/2)$.

ii. The wordline is raised to high to turn on the access transistor. A voltage difference appears on both bitlines with V_{BL} higher than $\overline{V_{BL}}$ for a logic "1" being read.

iii. The sense amplifier is then turned on by the SA signal, the signal difference is detected and then amplified to a full swing of "1" logic value. Since the wordline is still raised, the value on the bitline is written back to the cell.

3.2.1. Functional Simulation of Full Latch Cross Coupled

Figure 3.4 shows a waveform from a simulation of the operating states of the DRAM. The read of a "0" and "1" are shown. The simulation is run for 600 ns.

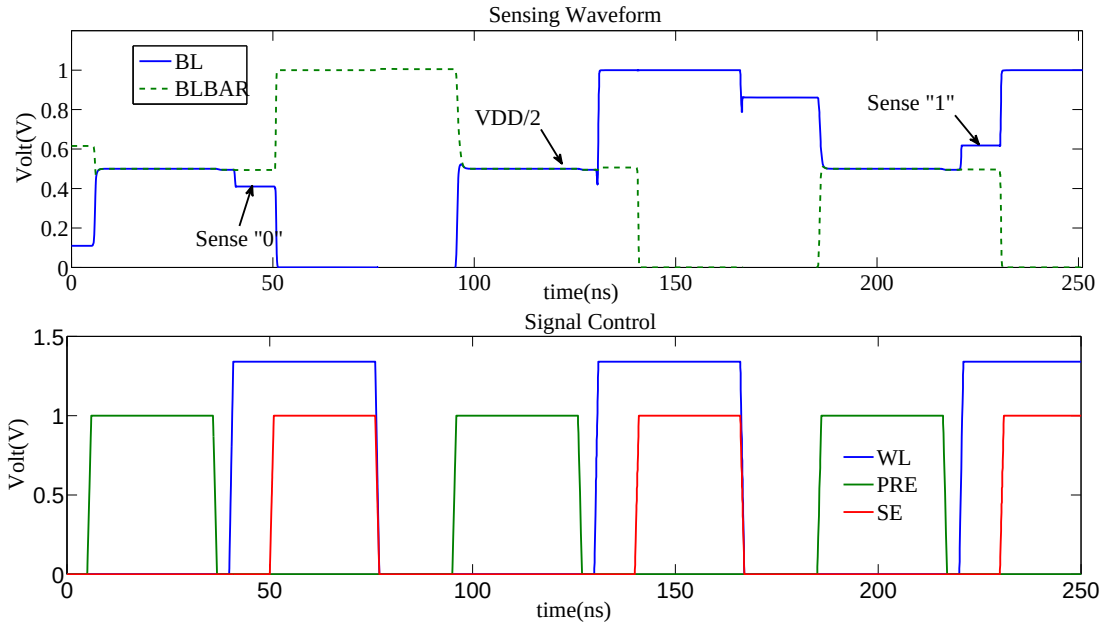


FIGURE 3.4. Waveform of full latch cross coupled cense amplifier

Table 3.1 shows a summary of the values for the FoMs. The value of the precharge time for the baseline design is 7.508 ns. The power consumed is 144.04 nw. The sense delay is 4.39 ns while the sense margin is 43.46 mV.

TABLE 3.1. Figures of Merit Characterization

FOM	Precharge (ns)	Power (nW)	Sense Delay (ns)	Sense Margin (mV)
Value	7.50	153.101	4.39	43.46

3.2.2. Variant of Full Latch Sense Amplifier

A variant of the full latch sense amplifier is shown in figure 3.5. The sense amplifier is enabled by transistors M3 and M7 at the Nsense region and transistors M6 and M2 at the Psense region. It also has the same precharge and equalization circuit. Figure 3.6 shows the waveform of the functional simulation.

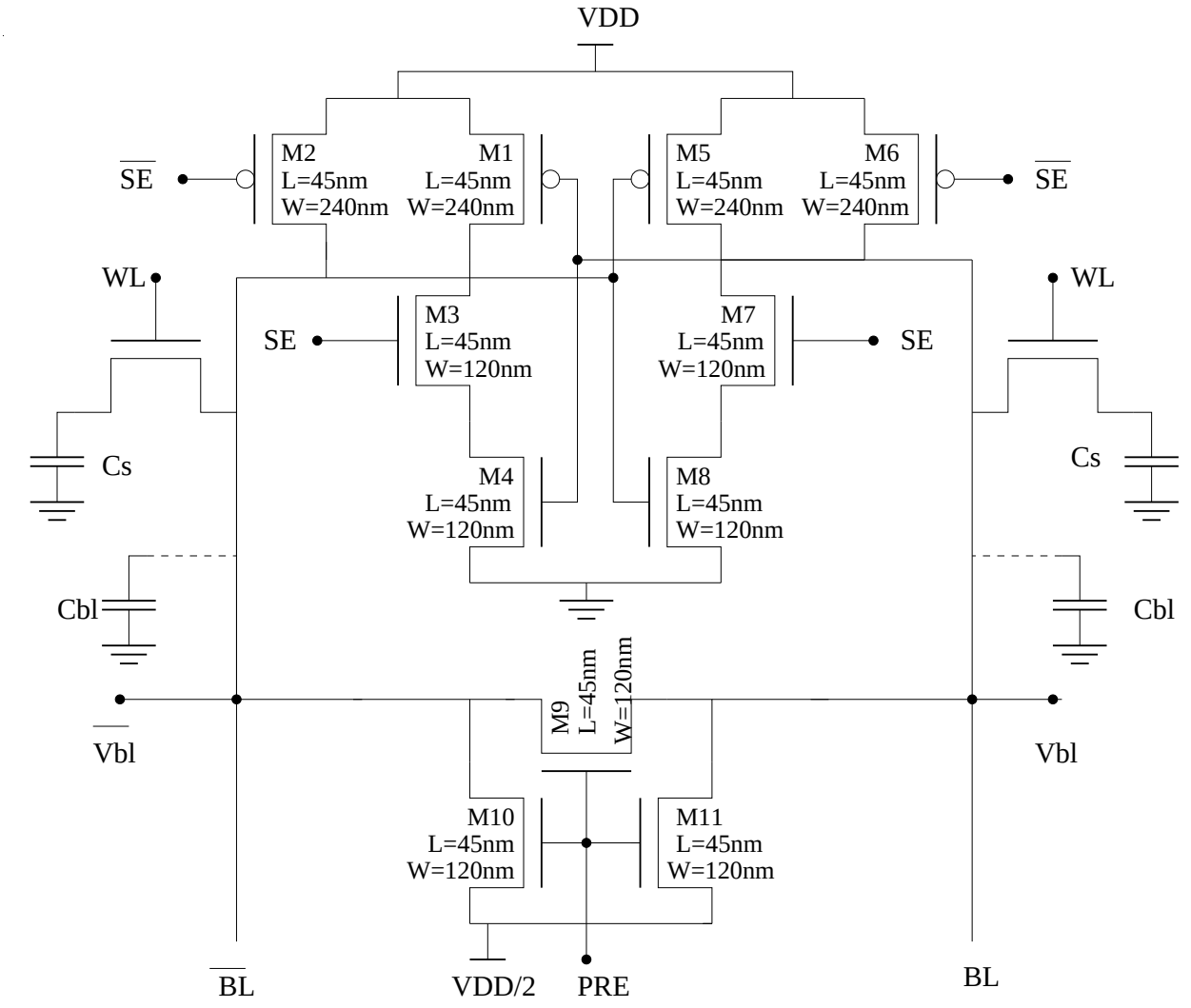


FIGURE 3.5. Circuit diagram of a variation of a full latch sense amplifier

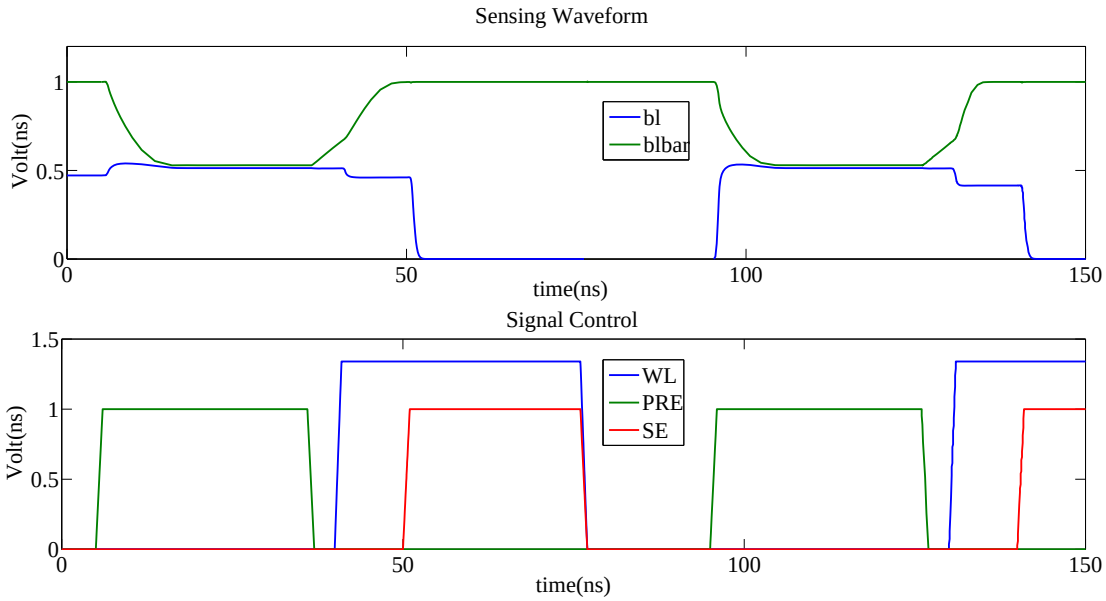


FIGURE 3.6. Waveform of bit latch sense amplifier

CHAPTER 4

DUAL OXIDE TECHNOLOGY BASED DESIGN OF THE LATCH CROSS COUPLED SENSE AMPLFIER

This chapter presents a design optimization method using dual oxide thickness for the sense amplifier design. A parametric analysis is performed on the latch cross coupled sense amplifier to analyze the effects of process and design parameters on the various FoMs defined in chapter 3. The results of the parametric analysis will guide the design optimization process. Using the results from the parametric analysis, parameter values are chosen for optimal performance of each FoM. A statistical analysis is performed using these parameter values to analyze the effect of variance on the performance. Finally, a design optimization algorithm is presented using the dual oxide technology to optimize a specific FoM, in this case the precharge time.

4.1. Parametric Analysis

This section presents the results of the parametric analysis of the sense amplifier. The variation of several parameters is studied. As discussed in chapter 2, process variations in semiconductor technology affect the device parameters and result in different operating conditions for the devices. Because of the inherent function of differential amplification, sense amplifiers are very sensitive to process mismatches and the transistors must by closely matched to ensure proper operation. The following device parameters were considered for the parametric analysis: the lengths and widths (L_n, W_n) and(L_p, W_p) of the NMOS and PMOS transistors in the sense amplifier and precharge and equalization circuit are considered as design parameters. The voltage supply (V_{DD}) and oxide thickness (t_{ox_n} and t_{ox_p}) are also varied along with the memory cell capacitance (C_S) and the bitline capacitance (C_{BL}). Each parameter was varied, while the others were kept constant to analyze the sensitivity of the FoMs on each parameter. The simulations were done using the SPECTRE analog simulator with a 45 nm CMOS technology. The circuit schematic used for the simulation is shown in figure 3.3. The following subsections show the effect of varying the parameters on the FoMs.

4.1.1. Sensitivity on L_n and L_p

The effects of varying both L_n and L_p are shown in figures 4.1 and 4.2. L_n and L_p are both varied from 45 nm to 120 nm. As L_n increases, the precharge time decreases from about 7.5 ns to approximately 3.5 ns. The average power consumed increases from about 141 nW to 145 nW as the sense delay and sense margins are only slightly affected by variations of L_n. The sense delay decreases from 4.4 ns to 4.3 ns, while the sense margin increases from 43.4 mV to 43.7 mV.

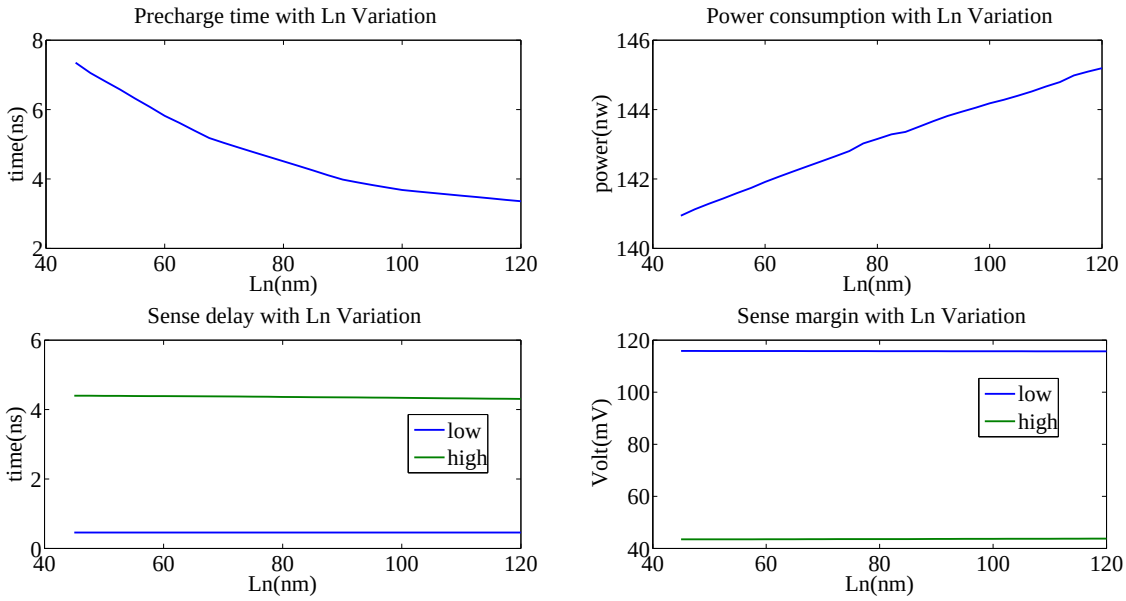


FIGURE 4.1. Effect of L_n Variation on FOMs: Precharge time decreases from about 7.5 ns to 4.5 ns and average power consumed increases from about 141 nW to 145 nW as L_n increases from 45 nm to 120 nm. Sense delay and margin are mildly affected

Varying L_p has a reverse effect on the precharge time as it increases with an increase in L_p. It increases from about 7.3 ns to about 7.8 ns. The power consumption decreases from 141 nW to 136 nW. The sense delay and sense margins are mildly affected by L_p variations. The sense delay decreases from 4.39 ns to 4.37 ns. The sense margin decreases from 43.4 mV to 43.3 mV. Power consumption decreases with an increase in L_p.

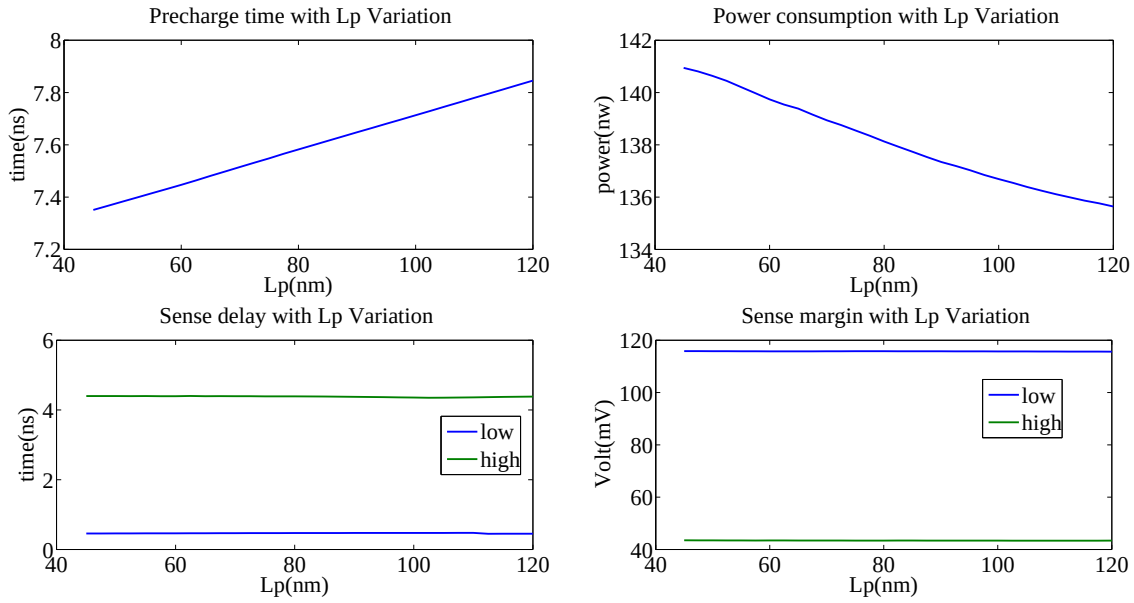


FIGURE 4.2. Effect of L_p Variation on FOMs: Precharge time increases from about 7.3 ns to 7.8 ns and average power consumed decreases from about 141 nw to 136 nw as L_p increases from 45 nm to 120 nm. Sense delay and margin are mildly affected.

4.1.2. Sensitivity on W_n and W_p

The results for the variation of W_n and W_p are shown in figures 4.3 and 4.4. W_n is varied from 120 nm to 360 nm, while W_p is varied from 240 nm to 720 nm. Increasing W_n decreases the precharge time, the sense delay and sense margin, while the power consumption increases. The precharge time decreases from 7.5 ns to 4.6 ns, while the power consumption increases from 141 nw to 151 nw. The sense delay decreases from 4.4 ns to 3.8 ns. The sense margin is mildly affected and decreases from 43.4 mV to 43.3 mV. The precharge time and power consumption increase from 7.5 ns to 8.6ns and from 141 nW to 237 nW respectively with an increase in W_p. The sense delay and sense margin both decrease slightly 4.4 ns to 4.1 ns and from 43.4 mV to 41 mV respectively as W_p increases

4.1.3. Sensitivity on V_{DD}

The voltage supply V_{DD} is varied from 0.8 V to 1.2 V. Figure 4.5 shows the results. Increasing V_{DD} increases the average power consumption, and the sense margin. Power consumed increases

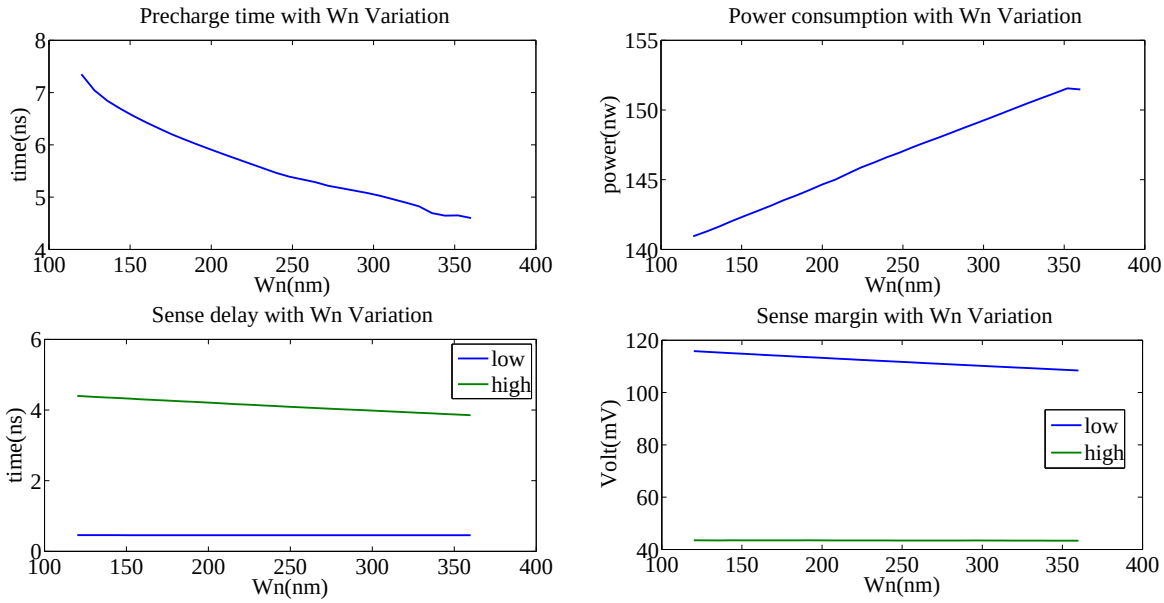


FIGURE 4.3. Effect of W_n Variation on FoMs: W_n is varied from 120 nm to 360 nm. The precharge time decreases from 7.5 ns to 4.6 ns, while the power consumption increases from 141 nw to 151 nw. The sense delay also decreases from 4.4 ns to 3.8 ns.

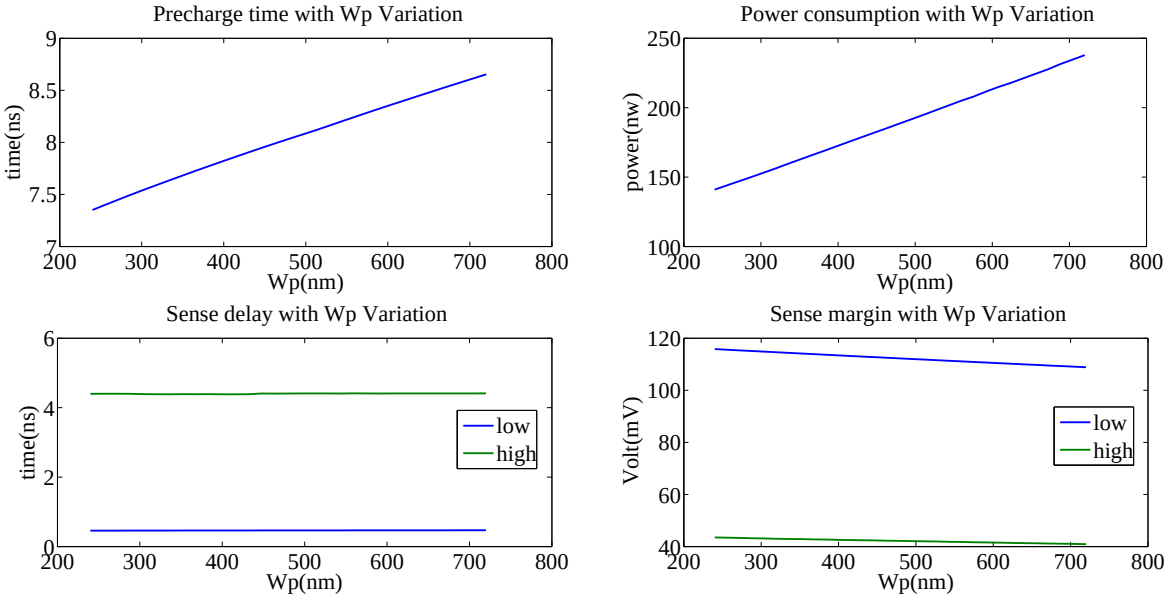


FIGURE 4.4. Effect of W_p Variation on FoMs: W_p is varied from 240 nm to 720 nm. The precharge time increases from 7.5 ns to 8.6 ns, and the power consumption increases from 141 nw to 237 nw. Sense delay and margin are mildly affected.

from 68 nW to 218 nW, while the sense margin increases from 3 mV to 64 mV. An increase in V_{DD} however, decreases the sense delay which is similar to the conclusion by Sherwin[1]. Sense delay decreases from 8 ns to 1.4 ns. The precharge and equalization time decreases from 20 ns to 2.1 ns as V_{DD} increases.

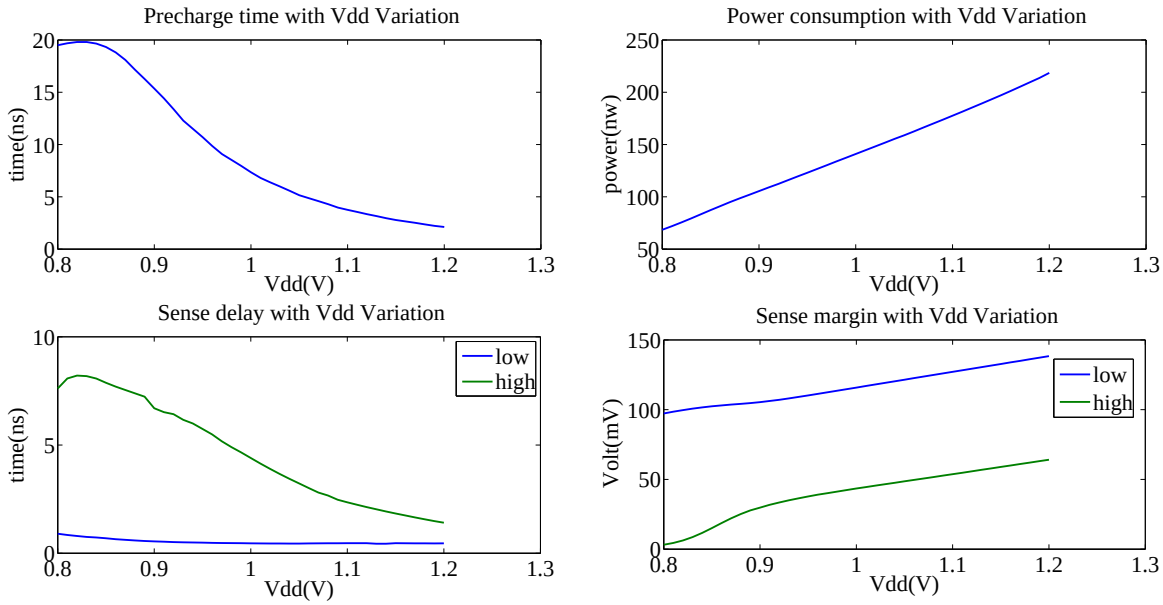


FIGURE 4.5. Effect of V_{DD} Variation on FoMs: V_{DD} is varied from 0.8 V to 1.2 V. Precharge decreases from 20 ns to 2.1 ns, power consumed increases form 68 nW to 218 nW. Sense delay decreases from 8 ns to 1.4ns while the sense margin increases from 3 mV to 64 mV.

4.1.4. Sensitivity on Capacitance

The cell storage capacitance C_S is varied from 2 fF to 30 fF and the bitline capacitance is varied from 10 fF to 90 fF. As C_S is increased, the precharge time and average power consumed increases as expected. Precharge time increases 7.35 ns to 7.36 ns and power consumed increases from 139 nW to 143 nW. The sense margin also increases from 38 mV to 56 mV. The sense delay slightly decreases from 6.7 ns to 6.3 ns as C_S increases.

As C_{BL} increases, there is an increase in the precharge time, average power consumed and a slight increase in sense delay. Increasing C_{BL} however decreases the sense margin. Precharge time increases from 5.26 ns to 19.6 ns and then decreases when C_{BL} reaches 40 fF to 17.3 ns. Power

consumed increases from 112 nW to 895 nW. The sense delay also increases form 4.42 ns to 7.6 ns as C_{BL} increases. The sense margin decreases from 56.9 mV to 3 mV. Figures 4.6 and 4.7 show the effects of C_S and C_{BL} variations on the FOMs, respectively.

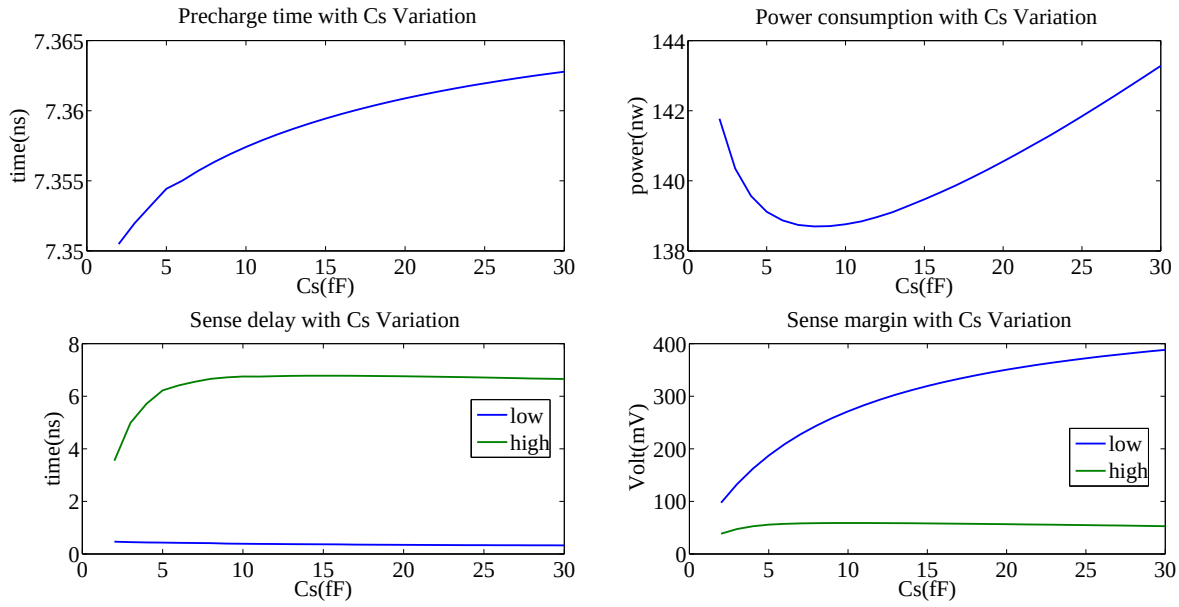


FIGURE 4.6. Effect of C_S Variation on FoMs: C_S is varied from 2 fF to 30 fF. Precharge time increases 7.35 ns to 7.36 ns and power consumed increases from 139 nW to 143 nW. Sense margin increases from 38 mV to 56 mV. sense delay slightly decreases from 6.7 ns to 6.3 ns.

4.1.5. Oxide

The oxide thickness for both NMOS and PMOS transistors was varied from 2 nm to 6 nm. The effects are shown in Figures 4.8 and 4.9. Increasing t_{ox_n}, precharge decreases from 7.43 ns to 1.2 ns. Power consumption increases from 140.9 nW to 142.5 nW. The sense delay time decreases form 4.42 ns to 0.90 ns. Sense margin increases from 43.3 mV to 87 mV.

Increasing t_{ox_p} reduces the precharge time, power consumption and mildly affects the sense delay and sense margin. Precharge time reduces from 7.35 ns to 7.26 ns, while power consumption decreases from 140.9 nW to 129.5 nW. The sense margin increases from 43.49 mV to 43.5 mV. The sense delay barely decreases from 4.399 ns to 4.394 ns.

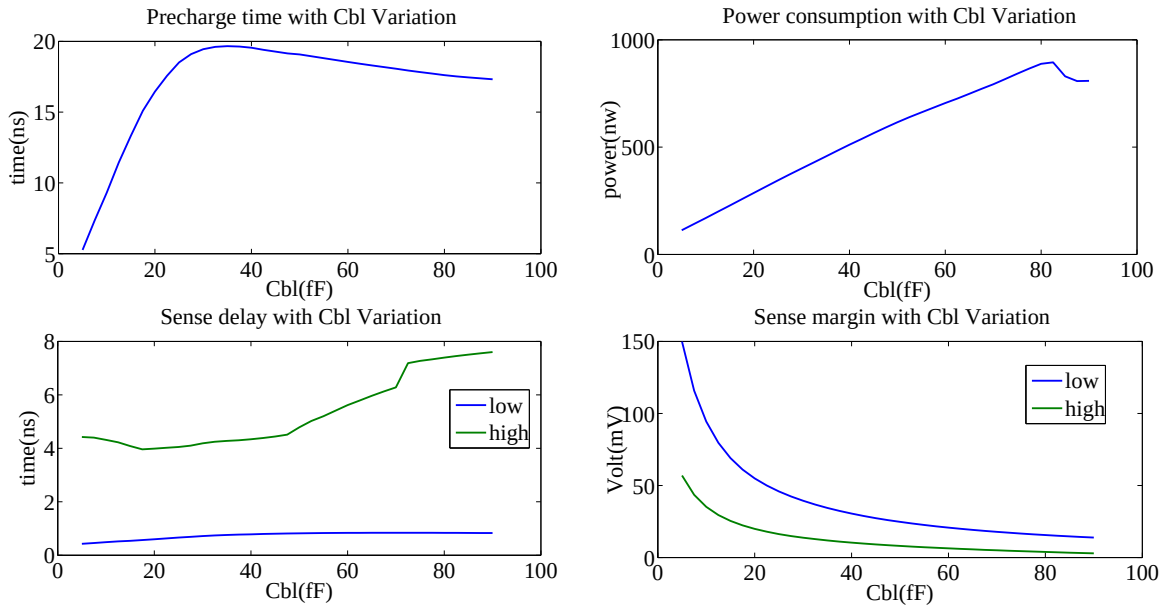


FIGURE 4.7. Effect of C_{BL} Variation on FoMs: C_{BL} is varied from 10 fF to 90 fF. Precharge time increases from 5.26 ns to 17.3 ns. Power consumed increases from 112 nW to 895 nW. Sense delay increases from 4.42 ns to 7.6 ns while sense margin decreases from 56.9 mV to 3 mV.

Table 4.1 shows a summary of the parametric analysis results for the parameters selected. The table shows the effect on the performance for each FoM when the parameter is increased. The performance of each FoM either increases or decreases as each parameter is increased. These results will serve as guidelines to investigate different optimization alternatives.

For the precharge and equalization time, an increase in L_n, W_n, V_{DD}, t_{ox_n}, t_{ox_p} decreases the precharge time and an increase in L_p, W_p, C_S, C_{BL} increases precharge time.

The power dissipation is affected by all the parameters. Increasing L_n, W_n, W_p, V_{DD}, C_S, C_{BL}, and t_{ox_n}, the power dissipation increases, while increasing L_p, V_{th} and t_{ox_p} the power dissipation decreases.

An increase in W_n, V_{DD}, C_S and t_{ox_n} reduces the sense delay. Increasing L_n, L_p, W_p and t_{ox_n} mildly decreases the sense delay and increasing C_{BL} has a mild increasing effect on the sense delay.

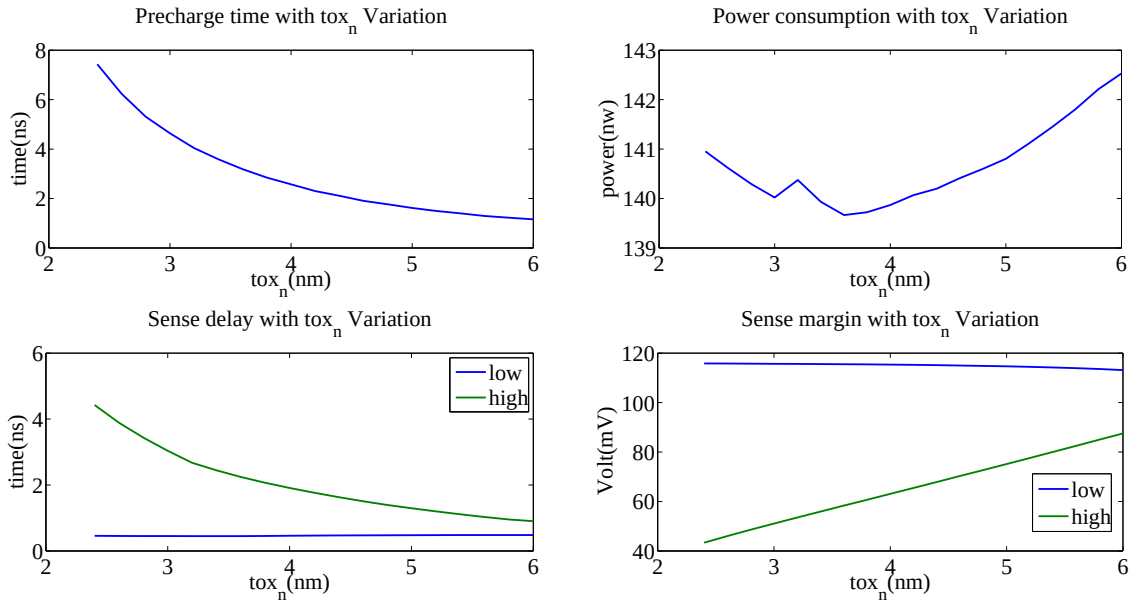


FIGURE 4.8. Effect of t_{ox_n} Variation on FoMs: t_{ox_n} is varied from 2 nm to 6 nm. Precharge decreases from 7.43 ns to 1.2 ns, while power consumption increases from 140.9 nW to 142.5 nW. The sense delay time decreases form 4.42 ns to 0.90 ns. Sense margin increases from 43.3 mV to 87 mV.

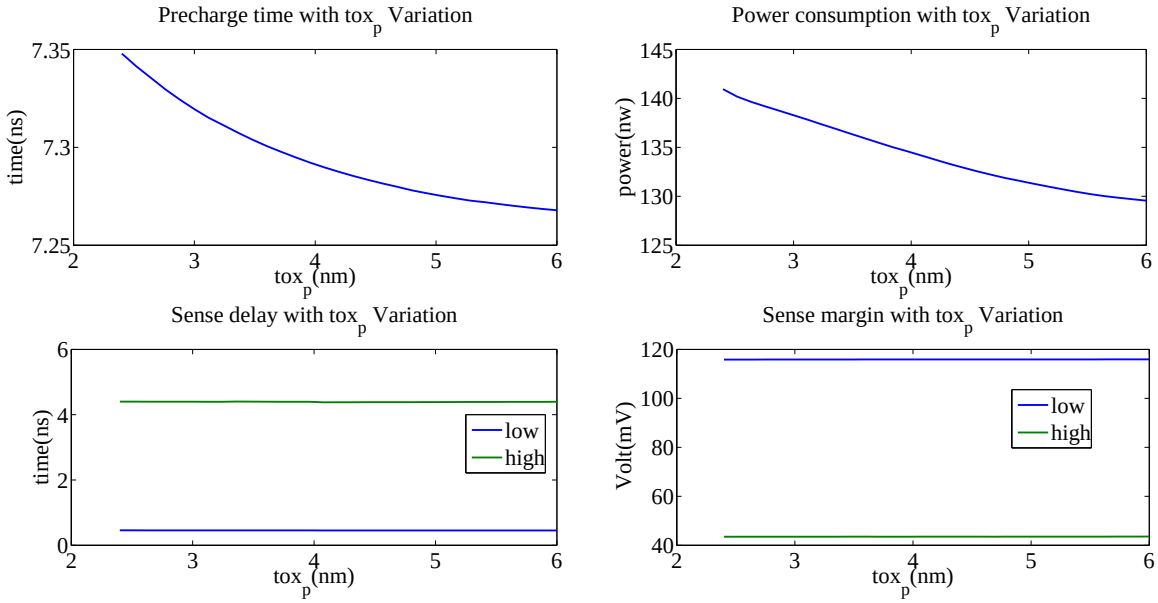


FIGURE 4.9. Effect of t_{ox_p} Variation on FoMs: t_{ox_p} is varied from 2nm to 6 nm. Precharge time decreases from 7.35 ns to 7.26 ns, and power consumption decreases from 140.9 nW to 129.5 nW.

An increase in V_{DD}, C_S and t_{ox_n} will result in an increased sense margin. Parameters L_n, L_p, W_n, W_p, V_{DD} and t_{ox_p} have a mild effect on the sense margin. Increasing L_n and t_{ox_p} mildly increases the sense margin, while increasing L_p, W_n, and C_{BL} mildly decreases the sense margin.

TABLE 4.1. Effects of Different Parameters on Sense Amplifier FoMs

Params	Precharge	Power	Sense Delay	Sense Margin
L_n	decrease	increase	mild decrease	mild increase
L_p	increase	decrease	mild decrease	mild decrease
W_n	decrease	increase	decrease	decrease
W_p	increase	increase	mild decrease	decrease
V_{DD}	decrease	increase	decrease	increase
C_S	increase	increase	decrease	increase
C_{BL}	increase	increase	increase	decrease
t_{ox_n}	decrease	increase	decrease	increase
t_{ox_p}	decrease	decrease	mild increase	mild increase

4.2. Statistical Analysis of Process Variation Effects on Sense Amplifier Circuit

In this section, a statistical analysis for each FOM is presented. The statistical analysis for each FoM is performed for different process parameter variations. A Monte Carlo analysis of a 1000 runs each with a 5% standard deviation from the mean of selected parameters is performed for each FoM. For each Monte Carlo analysis, the parameter values where chosen to provide the best performance for the given FoM. The parametric analysis results, shown in Table 4.1, summarize the effect of each parameter on the FOM. From this table the parameter values were selected for the analysis. The results of the Monte Carlo analysis are shown in figures 5.3, 5.4, 5.5 and 5.6 and are discussed in the following sections.

4.2.1. Precharge and Voltage Equalization Time

Figure 4.10 shows the results for the MonteCarlo analysis for the precharge time. The parametric analysis indicates that all parameters considered (L_n, L_p, W_n, W_p, V_{DD}, C_S, C_{BL}, t_{ox_n} and

t_{ox_p}) affect the precharge time significantly. A higher value of 120 nm and 360 nm is chosen for L_n and W_n, while a lower value of 50 nm and 240 nm is chosen for L_p and W_p respectively. V_{DD} is 1.2 V while 2.5 fF and 7.5 fF are chosen for C_S and C_{BL}, respectively. t_{ox_n} and t_{ox_p} are kept at 3 nm. All parameters were varied with a normal (Gaussian) distribution and a standard deviation of 5%. The mean value for the precharge time is 768.01 ps with a standard deviation of 128.67 ps.

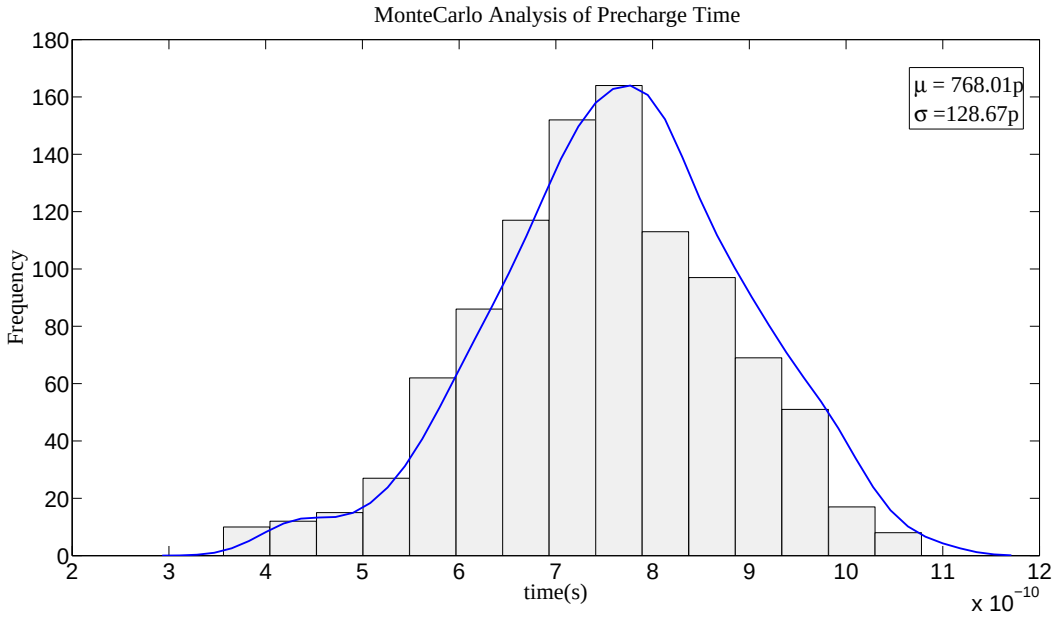


FIGURE 4.10. Probability Density Function (PDF) of precharge time: The distribution is Gaussian in nature with μ = 768.01 ps and σ = 128.67 ps

4.2.2. Power Consumption

Figure 4.11 shows the results for the Monte Carlo analysis for the average power consumption. From the parametric analysis, it is seen that all parameters considered significantly affect the power consumption. A higher value of 120 nm is selected for L_p, while 50 nm, 130 nm, and 240 nm are used for L_n, W_n, and W_p respectively. V_{DD} is 0.9 V while 5 fF and 15 fF are chosen for C_S and C_{BL}. t_{ox_n} and t_{ox_p} are kept at 2.4 nm. All parameters were varied with a normal distribution and a standard deviation of 5%. The mean value for the power consumed is 190.88 nW with a standard deviation of 28.88 nW

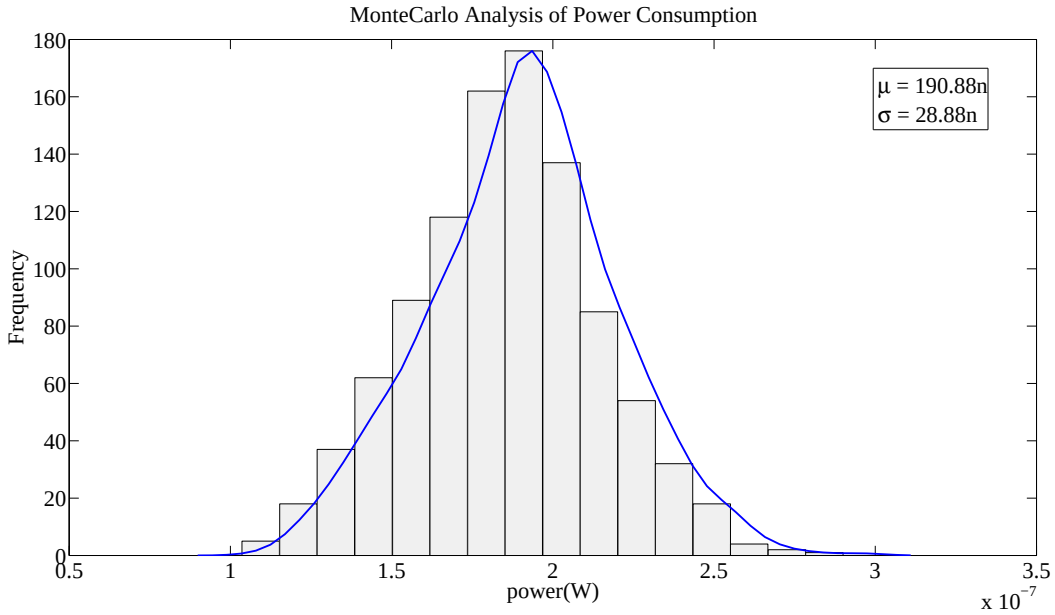


FIGURE 4.11. Monte Carlo Analysis of Average Power Consumption: The distribution is Gaussian in nature with μ = 198.88 nw and σ = 28.88 nw

4.2.3. Sense Delay

Figure 4.12 shows the results for the Monte Carlo analysis for the sense delay time. The parametric analysis shows that only parameters W_n, V_{DD}, C_S, C_{BL}, V_{th} and tox_n significantly affect the sense delay. A higher value of 120 nm, 360 nm and 720 nm is chosen for the L_n, W_n and W_p respectively, while a lower value of 50 nm is chosen for L_p. V_{DD} is at 1.2 V while 2.5 fF and 45 fF are chosen for C_S and C_{BL} respective. t_{ox_n} and tox_p are kept at 3 nm and 2.4 nm, respectively. Only parameters significantly affecting the sense delay are varied with a normal distribution and a 5% deviation. The mean value for the sense delay is 1.32 ns with a standard deviation of 0.38 ns.

4.2.4. Sense Margin

The results for the Monte Carlo analysis of the sense margin are shown in figure 4.13. The parameters significantly affecting the sense margin are W_n, W_p, V_{DD}, C_S, C_{BL} and t_{ox_n}. A higher value of 120 nm, 360 nm and 720 nm is chosen for the L_p, W_n and W_p respectively, while a lower value of 50 nm is chosen for L_n. V_{DD} is set at 0.9 V, while 2.5 fF and 45 fF are chosen for C_S and C_{BL}, and t_{ox_n} and t_{ox_p} are kept at 2.4 nm. The mean value for the sense delay is 10.11 mV with a standard deviation of 2.31 mV.

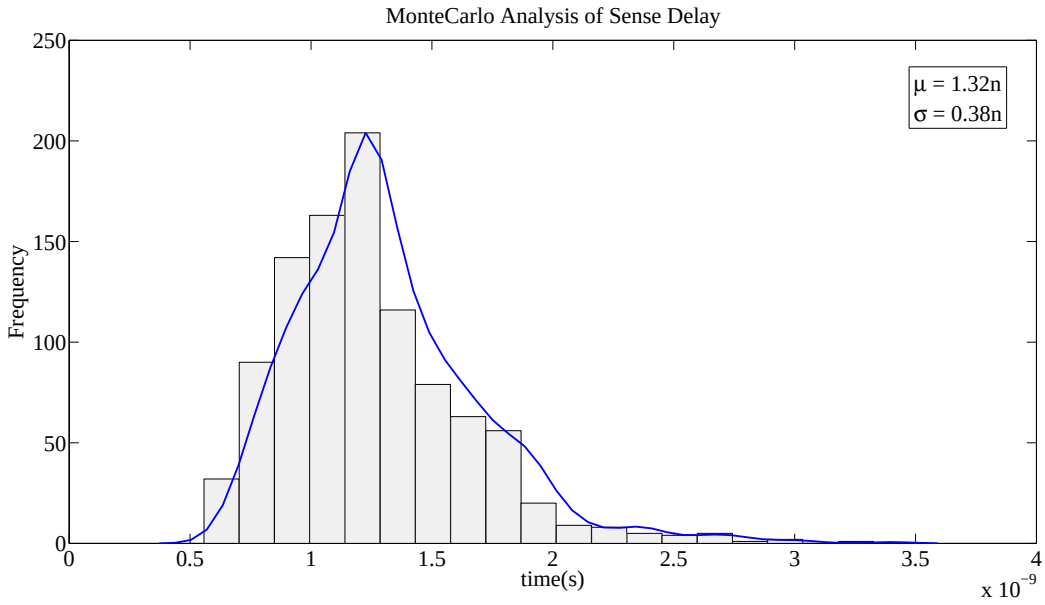


FIGURE 4.12. PDF of sense delay: The distribution is Gaussian in nature with μ = 1.32 ns and σ = 0.38 ns

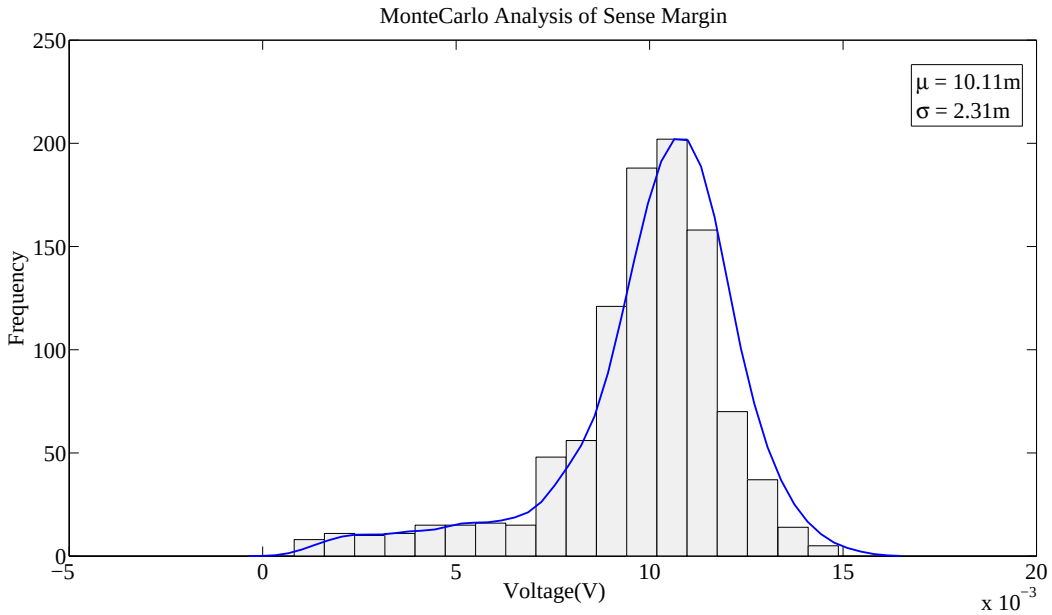


FIGURE 4.13. PDF of sense margin: The distribution is Gaussian in nature with μ = 10.97 mV and σ = 1.87 mV

4.3. Effect on Mean and Standard Deviations

To analyze the effect each parameter has on the mean and standard deviation values for the Monte Carlo analysis, another set of Monte Carlo analysis, 1000 runs each is performed.

For this analysis, only one parameter is varied with a normal distribution and 5% deviation, while the other parameters are kept constant. A total of 36 different tests, one for each parameter and FoM is performed. The results for this analysis are shown in Table 4.2. The table shows the effect each parameter has on the mean and standard deviation of the FoM values. In this work, the criteria used for selecting parameters were based on the smallest effect on the deviation, which effectively improves yield.

TABLE 4.2. Probability Density Function of Different FoMs of the Sense Amplifier

Params	Precharge PDF		Power PDF		Sense Delay PDF		Sense Margin PDF	
	μ (ps)	σ (ps)	μ (nw)	σ (nw)	μ (μs)	σ (ps)	μ (mV)	σ (μV)
L_n	773.65	13.81	191.67	0.56	1.26	16.79	10.67	32.33
L_p	773.44	2.40	191.67	0.06	1.26	0.15	10.68	2.22
W_n	773.18	5.04	191.71	0.50	1.25	15.82	10.67	102.41
W_p	774.08	6.34	190.69	2.07	1.26	0.22	10.68	13.49
V_{dd}	772.72	123.68	190.85	27.25	1.30	0.36	10.21	2074.63
C_s	772.61	0.20	191.68	0.25	1.26	49.81	10.66	367.51
C_{bl}	773.48	14.76	191.73	6.38	1.26	3.76	10.69	591.49
t_{ox_n}	774.15	11.68	191.32	1.47	1.26	25.48	10.62	311.87
t_{ox_p}	773.74	2.05	192.06	1.93	1.26	0.10	10.68	65.80

4.4. Optimal Design Flow

This section discusses the optimal design flow process. The design flow for the optimal design of sense amplifiers is presented in figure 4.14. This flows aims to present a method for an optimal design for a sense amplifier with a maximum variance tolerance without increasing the complexity of the circuitry. This includes starting with the baseline design of the final parameters chosen for the design.

The first step involves choosing a baseline design to start with. After the design schematic is completed, a simulation is run to verify its operation. The design is then characterized to identify key FoMs. More simulations are then performed to identify parameters that affect these FoMs.

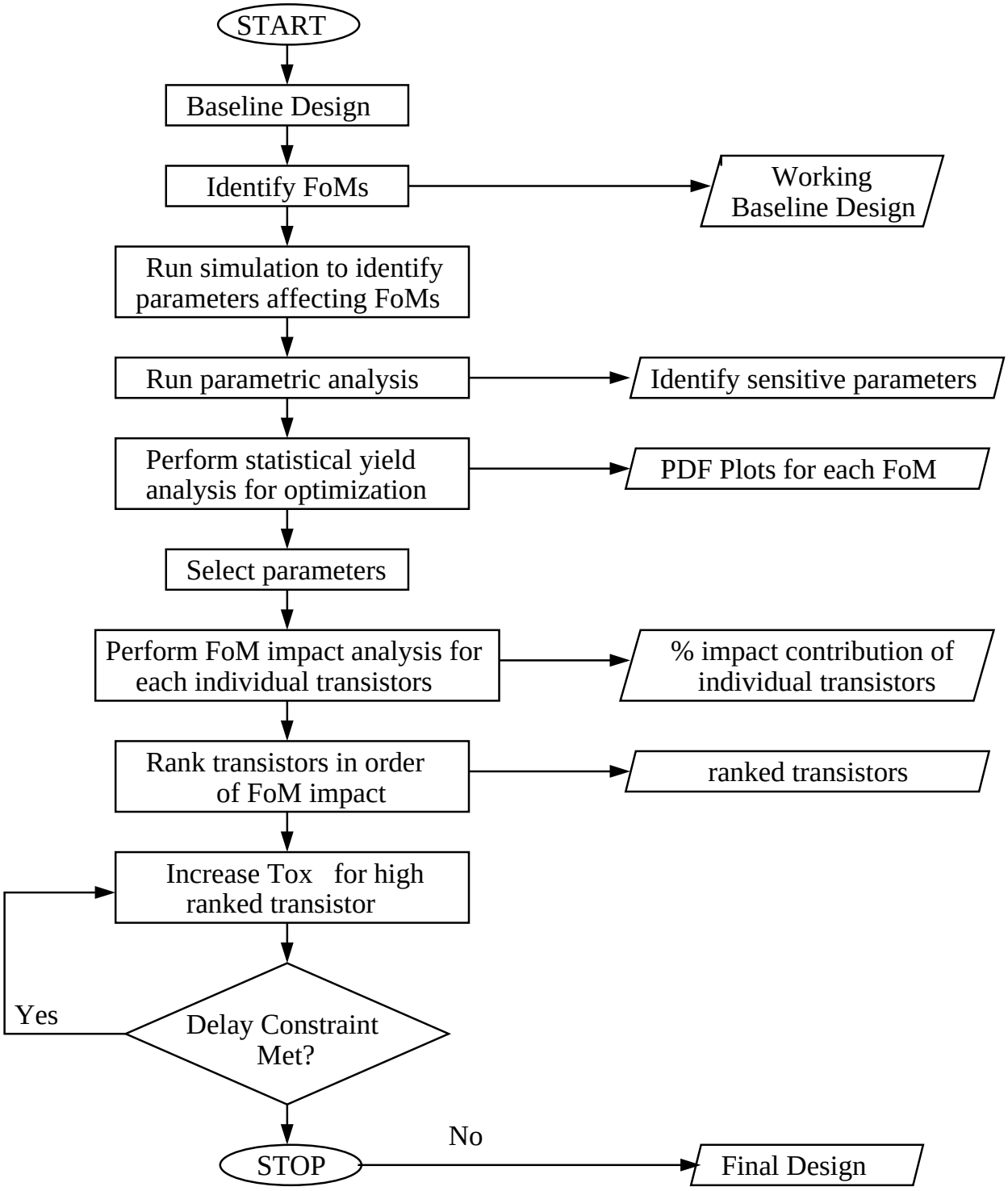


FIGURE 4.14. Flow diagram for optimal design of sense amplifiers applying dual oxide thickness based technology

The next step involves performing a parametric analysis to identify parameters that significantly affect the FoMs. A Monte Carlo analysis is then performed, based on the sensitive parameters and values for parameters that give optimal performance. From the results of the parametric and Monte Carlo analysis, values for the design parameters are then chosen. The values for design parameters where chosen to decrease the precharge time of the circuit. To further optimize the circuit for precharge, an analysis was done on the individual transistors to identify the impact each transistor has on the FoM to be optimized (precharge time in this case).

Algorithm 1 is used to optimize the sense amplifier circuit and deals with individual transistors in the amplifier. The algorithm is heuristic and which relies on simple ranking of the transistors. The transistors are ranked accordingly in order of impact on the FoM (precharge), and t_{ox} is increased from $t_{ox_{low}}$ to $t_{ox_{high}}$ and checked against a given delay constraint. The delay constraint is the minimal delay time that the sense amplifier must have for correct operation. A value of 770 ps is selected based on the statistical analysis of table 4.2. If the constraint is met, the t_{ox} of the next ranked transistor is also increased and the delay constraint checked again. This is continued until all the t_{ox} for all transistors has been increased or until the delay constraint is violated.

The final FoM summary for the design is shown in Table 4.3. The final precharge time and power consumption are lower than their starting values. Further exploration will be performed with respect to varying V_{th} to optimize the process tolerance of the sense amplifier circuit. Chapter 5 explores an optimization algorithm using dual threshold technology.

TABLE 4.3. Final Figures of Merit Characterization

Circuits	Precharge (ns)	Power (μW)	Sense Delay (ns)	Sense Margin (mV)
Value	7.5 ns	153.1 μW	4.4 ns	43.4 mV

Algorithm 1 Heuristic algorithm for sense amplifier optimization using Dual Oxide Based Technology

1: Create the netlist of the sense amplifier circuit.
2: Number each of the transistors in the netlist from 1 to N.
3: **for** Each of the transistors i = 1 to N **do**
4: Rank the transistors for a figure of merit. For example, contributions to precharge.
5: Identify the significant transistors from the ranks, e.g. 30% or 50%
6: **end for**
7: Start with the highest ranked transistor as i = 0
8: **while** Design constraint of the sense amplifier is met and the transistor M_i is a significant transistor that contributes to the overall FoM value of the sense amplifier **do**
9: Increase thickness oxide of the the transistor M_i
10: Move to the next ranked transistor
11: **end while**

CHAPTER 5

DUAL THRESHOLD VOLTAGE TECHNOLOGY BASED DESIGN OF THE LATCH CROSS COUPLED SENSE AMPLIFIER

This chapter presents a design optimization method using dual threshold based technology for the sense amplifier design. A parametric analysis is performed on the latch cross coupled sense amplifier to analyze the effects of process and design parameters on the various FoMs defined in chapter 3. The parametric analysis was done in 4.1. The new parameters introduced in this chapter are the threshold voltages (V_{th_n} and V_{th_p}) of the NMOS and PMOS respectively. The new results and the results from chapter 4 will guide the design optimization process. Again using the results from the parametric analysis, parameter values are chosen for optimal performance of each FoM. A statistical analysis is performed using these parameter values to analyzed the effect of variance on the performance. Finally, a design optimization algorithm is presented using the dual threshold voltage technology to optimize a specific FoM, again in this case, we choose to optimize the precharge time. Dual threshold voltage technology is a method whereby the different transistors in a circuit design are assigned different threshold voltages to optimize the performance of the transistor [20]. The transistor could be assigned a $V_{th_{high}}$ high or low $V_{th_{low}}$.

5.1. Parametric Analysis

This section presents the results of the parametric analysis of the sense amplifier. A discussion on the impact of process variations is given in section 1.4. The parametric analysis of the device parameters is performed to study how they affect the FoMs described in 3. In this section, the parametric analysis of the threshold voltages (V_{th_n} and V_{th_p}) is performed. Each parameter is varied while the others are kept constant. Again, the simulations were done using the SPECTRE analog simulator with a 45 nm CMOS technology. The circuit schematic used for the simulation is shown in figure 3.3. The effect of varying V_{th_n} and V_{th_p} on the FoMs is shown in the next subsection.

5.1.1. Sensitivity Analysis of V_{th_n} and V_{th_p}

The parametric analysis for variation of threshold voltage is shown in figures 5.1 and 5.2. V_{th_n} is varied from 0.18 V to 0.36 V. As V_{th_n} increases, the precharge time and sense delay increase while the sense margin decreases. The precharge time increases from about 2 ns to almost 20 ns. Power consumption varies; it decreases and increases as V_{th_n} is varied. This might be due to the nature of the predictive SPICE models used in this work. The sense delay increases from 1.2 ns to 7.64 ns while the sense margin decreases from 72 mV to 29 mV.

V_{th_p} is varied from -0.18 V to -0.36 V. As V_{th_p} increases, the precharge time decreases from 7.69 ns to 7.42 ns while power consumption decreases from about 74 nW to 52 W. The sense delay and sense margin are mildly affected by V_{th_p}.

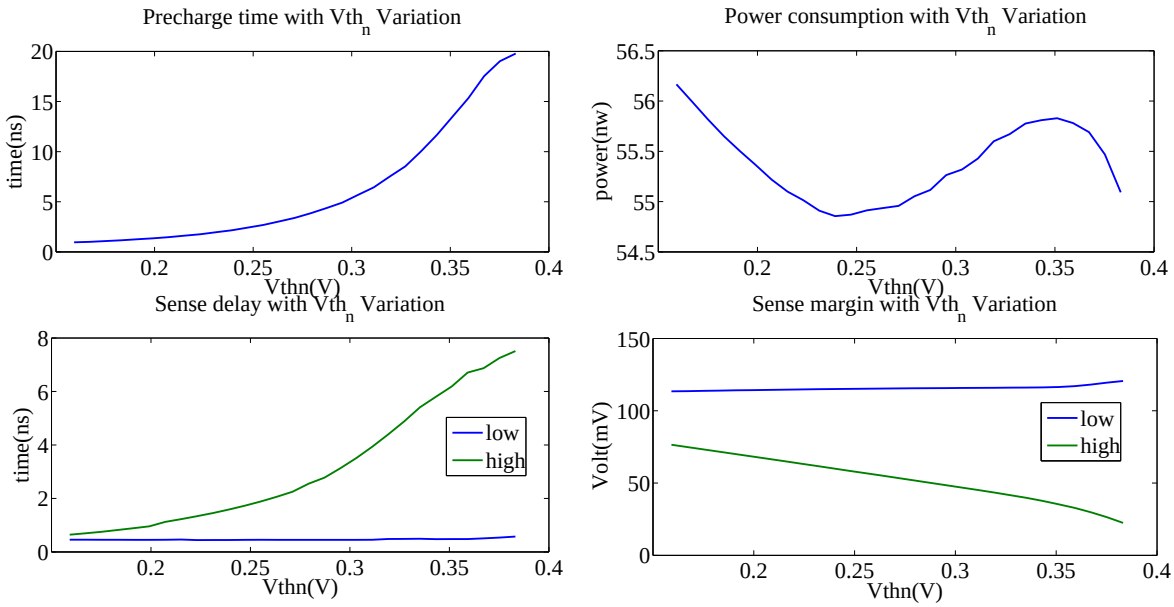


FIGURE 5.1. Effect of V_{th_n} Variation on FoMs: V_{th_n} is varied from 0.18 V to 0.36 V. The precharge time increases from about 2 ns to almost 20 ns. Sense delay increases from 1.2 ns to 7.64 ns while the sense margin decreases from 72 mV to 29 mV.

Table 5.1 shows a summary of the parametric analysis results for the parameters varied. The parameters varied in chapter 4 are also included in this table along with V_{th_n} and V_{th_p}. The table

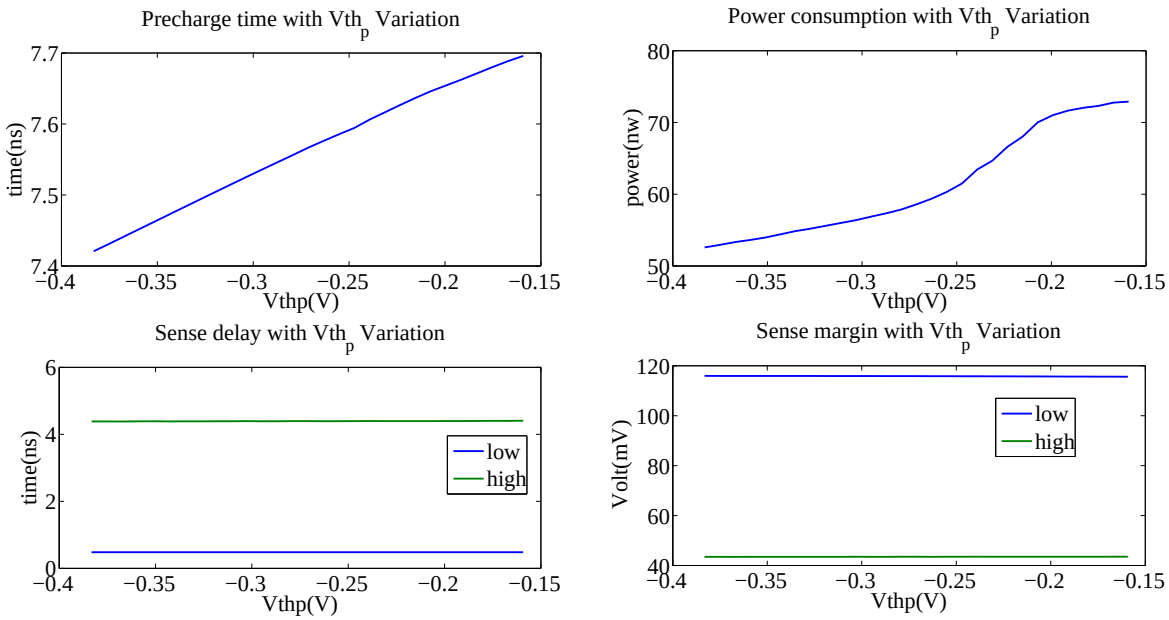


FIGURE 5.2. Effect of V_{th_p} Variation on FoMs: V_{th_n} is varied from -0.18 V to -0.36 V. Precharge time decreases from 7.69 ns to 7.42 ns while power consumption decreases from about 74 nW to 52 W.

shows the effect on the performance for each FoM when the parameter is increased. The performance of each FoM either increases or decreases as each parameter is increased. These results will serve as guidelines to investigate different alternatives.

For the precharge and equalization time, an increase in L_n, W_n, V_{DD}, t_{ox_n}, t_{ox_p} decreases the precharge time and an increase in L_p, W_p, C_S, C_{BL} increases precharge time.

The power dissipation is affected by all the parameters. Increasing L_n, W_n, W_p, V_{DD}, C_S, C_{BL}, and t_{ox_n}, the power dissipation increases, while increasing L_p, V_{th} and t_{ox_p} the power dissipation decreases.

An increase in C_{BL} and a decrease in W_n, V_{DD}, C_S and t_{ox_n} reduces the sense delay. Increasing L_n, L_p, and W_p mildly decreases the sense delay and increasing t_{ox_p} has a mild increasing effect on the sense delay.

An increase in V_{DD}, C_S and t_{ox_n} will result in an increased sense margin while an increase W_n, W_p and C_{BL} results in a lower sense margin. Parameters L_n, L_p, W_p, V_{DD} and t_{ox_p} have a

mild effect on the sense margin. Increasing L_n and t_{ox_p} mildly increases the sense margin, while increasing L_p mildly decreases the sense margin.

TABLE 5.1. Effects of Different Parameters on Sense Amplifier FoMs

Params	Precharge	Power	Sense Delay	Sense Margin
L_n	decrease	increase	mild decrease	mild increase
L_p	increase	decrease	mild decrease	mild decrease
W_n	decrease	increase	decrease	decrease
W_p	increase	increase	mild decrease	decrease
V_{DD}	decrease	increase	decrease	increase
C_S	increase	increase	decrease	increase
C_{BL}	increase	increase	increase	decrease
V_{th_n}	increase	decrease	increase	mild decrease
V_{th_p}	increase	decrease	increase	mild decrease
t_{ox_n}	decrease	increase	decrease	increase
t_{ox_p}	decrease	decrease	mild increase	mild increase

5.2. Statistical Analysis of Process Variation Effects on Sense Amplifier Circuit

In this section, a statistical analysis for each FoM is presented. The statistical analysis for each FoM is performed for different process parameter variations. A Monte Carlo analysis of 1000 runs each with a 5% deviation from the mean of selected parameters is performed for each FoM. For each Monte Carlo analysis, the parameter values where chosen to provide the best performance for the given FoM. The parametric analysis results, shown in Table 5.1, summarize the effect of each parameter on the FoM. From this table the parameter values were selected for the analysis. The results of the Monte Carlo analysis are shown in figures 5.3, 5.4, 5.5 and 5.6 and are discussed in the following sections.

5.2.1. Precharge and Voltage Equalization Time

Figure 5.3 shows the results for the Monte Carlo analysis for the precharge time. The parametric analysis indicates that all parameters(L_n, L_p, W_n, W_p, V_{DD}, C_S, C_{BL}, V_{th_n}, V_{th_p}, t_{ox_n} and t_{ox_p}) considered affect the precharge time significantly. A higher value of 120 nm and 360 nm is chosen for L_n and W_n, while a lower value of 50 nm and 240 nm is chosen for L_p and W_p respectively. V_{DD}, V_{th_n}, and V_{th_p}, is set at 1.2 V, 0.18 V and -0.36 V. while 2.5 fF and 7.5 fF are chosen for C_S and C_{BL}. t_{ox_n} and t_{ox_p} are kept at 3 nm. All parameters were varied with a normal distribution and a deviation of 5%. The mean value for the precharge time is 603.72 ps with a standard deviation of 20.98 ps.

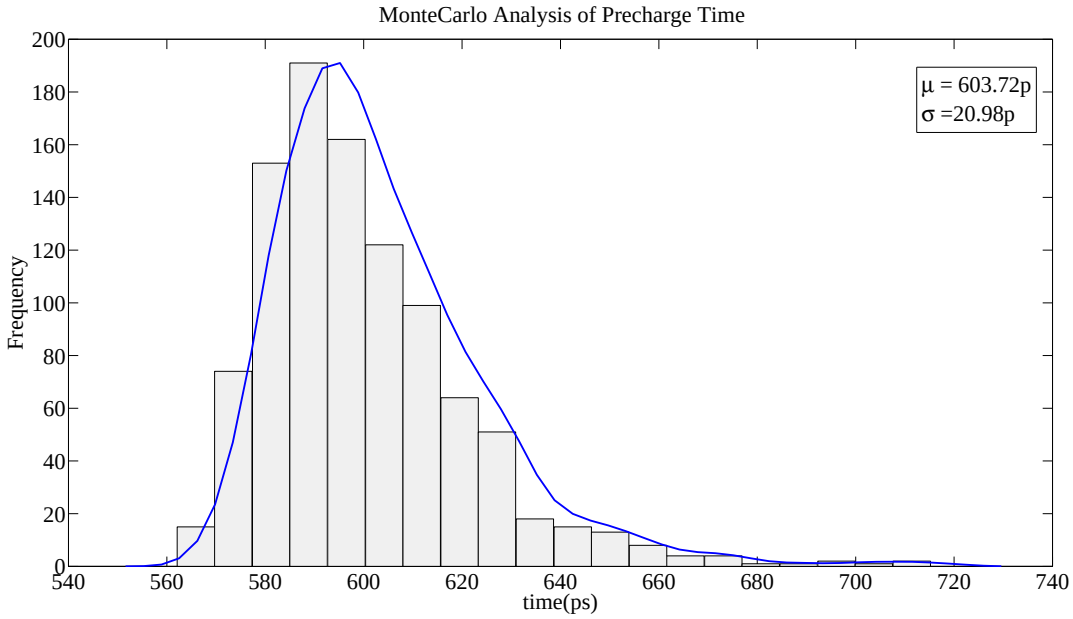


FIGURE 5.3. PDF of precharge time: The distribution is Gaussian in nature with μ = 603.72p and σ = 20.98p

5.2.2. Power Consumption

Figure 5.4 shows the results for the Monte Carlo analysis for the average power consumption. From the parametric analysis, it is seen that all the parameters considered significantly affect the power consumption. A higher value of 120 nm is selected for L_p, while 50 nm, 130 nm, and 240 nm are used for L_n, W_n, and W_p respectively. V_{DD} is 0.9 V while 5 fF and 15 fF are chosen for C_S

and C_{BL}. t_{ox_n} and t_{ox_p} are kept at 2.4 nm. All parameters were varied with a normal distribution and a deviation of 5%. The mean value for the power consumed is 94.88 nW with a standard deviation of 3.11 nW.

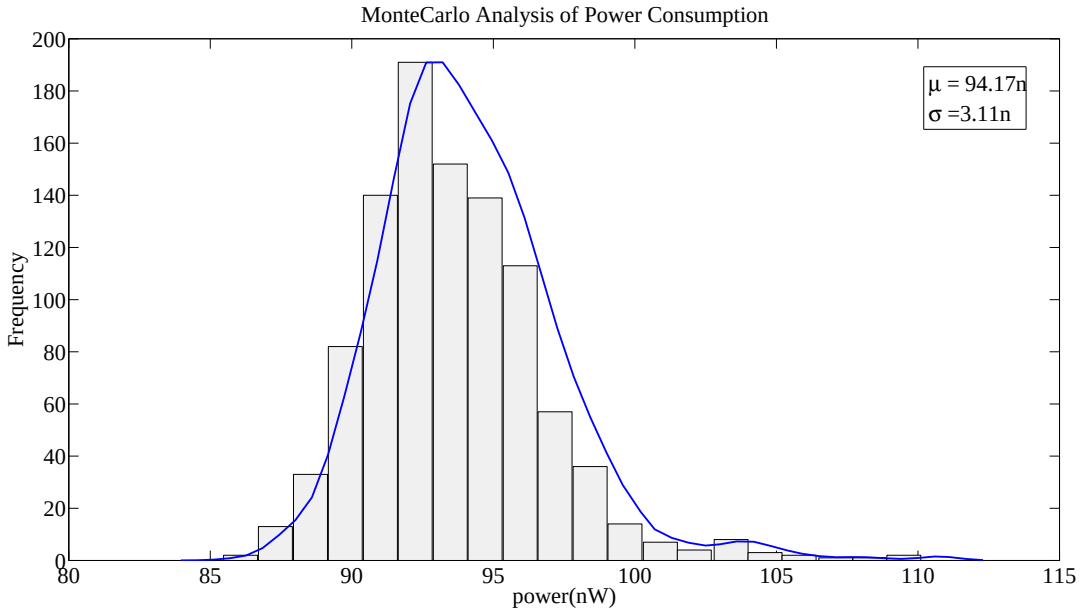


FIGURE 5.4. Monte Carlo Analysis of Average Power Consumption: The distribution is Gaussian in nature with μ = 94.17 nw and σ = 3.11 nw

5.2.3. Sense Delay

Figure 5.5 shows the results for the Monte Carlo analysis for the sense delay time. The parametric analysis shows that only parameters W_n, V_{DD}, C_S, C_{BL}, V_{th} and tox_n significantly affect the sense delay. A higher value of 120 nm, 360 nm and 720 nm is chosen for L_n, W_n and W_p respectively, while a lower value of 50 nm is chosen for L_p. V_{DD} is at 1.2 V while 2.5 fF and 45 fF are chosen for C_S and C_{BL} respective. t_{ox_n} and tox_p are kept at 3 nm and 2.4 nm, respectively. Only parameters significantly affecting the sense delay are varied with a normal distribution and a 5% deviation. The mean value for the sense delay is 812.97 ps with a standard deviation of 244.07 ps.

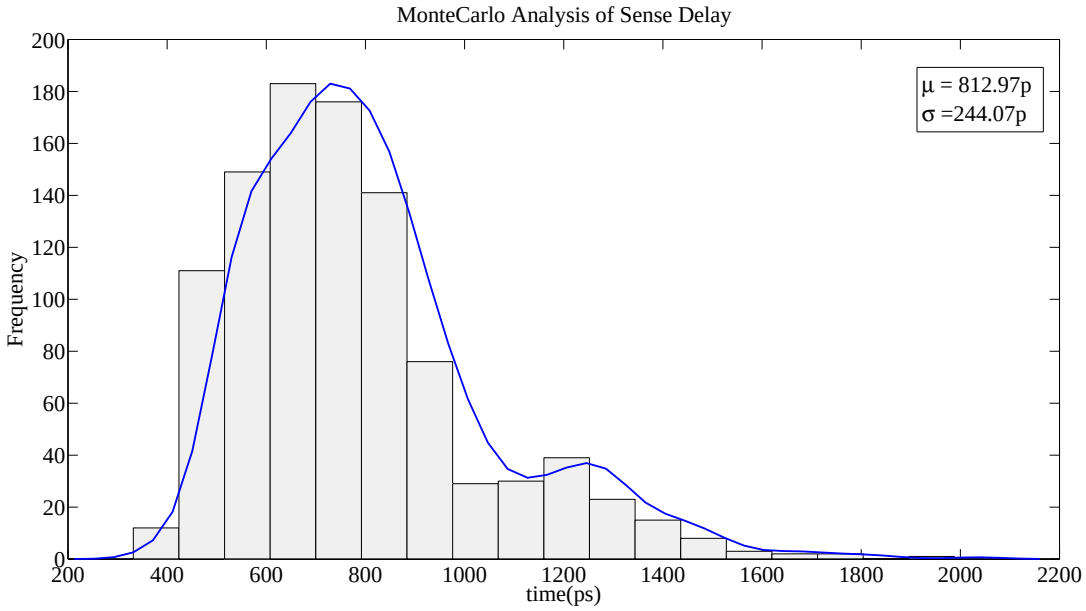


FIGURE 5.5. PDF of Sense Delay: The distribution is gaussian in nature with μ = 812.97 ps and σ = 244.07 ps

5.2.4. Sense Margin

The results for the Monte Carlo analysis of the sense margin are shown in figure 5.6. The parameters significantly affecting the sense margin are $W_n, W_p, V_{DD}, C_S, C_{BL}$ and t_{ox_n}. A higher value of 120 nm, 360 nm and 720 nm is chosen for L_p, W_n and W_p respectively, while a lower value of 50 nm is chosen for L_n. V_{DD} is set at 0.9 V, while 2.5 fF and 45 fF are chosen for C_S and C_{BL}, and t_{ox_n} and t_{ox_p} are kept at 2.4 nm. The mean value for the sense delay is 10.97 mV with a standard deviation of 1.87 mV.

5.3. Effect on Mean and Standard Deviations

To analyze the effect each parameter has on the mean and standard deviation values for the Monte Carlo analysis, another set of Monte Carlo analysis, 1000 runs each is performed.

For this analysis, only one parameter is varied with a normal distribution and 5% deviation, while the other parameters are kept constant. A total of 40 different tests, one for each parameter and FoM is performed. The results from this analysis are shown in Table 5.2. The table shows the effect each parameter has on the mean and standard deviation of the FoM values. In this work, the criteria used for selecting parameters were based on the smallest effect on the deviation.

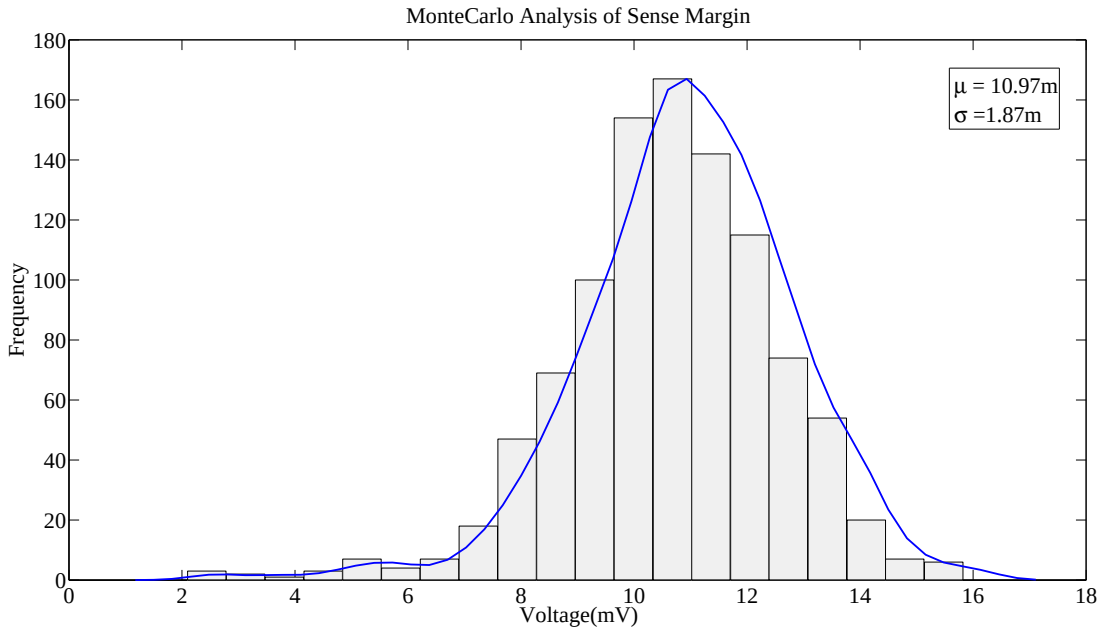


FIGURE 5.6. PDF of Sense Margin: The distribution is Gaussian in nature with μ = 10.97 mV and σ = 1.87 mV

TABLE 5.2. Probability Density Function of Different FoMs of the Sense Amplifier

Parameters	Precharge PDF		Power PDF		Sense Delay PDF		Sense Margin PDF	
	μ (ps)	σ (ps)	μ (nw)	σ (pw)	μ (ps)	σ (ps)	μ (μV)	σ (μV)
L_n	598.01	4.22	94.23	3.13	770.1	80.85	7.89	907.49
L_p	597.92	4.12	94.20	3.08	770.12	80.85	7.88	914.88
W_n	598.2	5.47	94.16	3.03	770.2	81.01	7.87	920.6
W_p	597.93	4.16	94.15	3.25	770.12	80.86	7.88	913.22
V_{dd}	602.37	17.83	94.59	16.88	803.06	211.95	6.93	3154.05
C_s	597.93	4.15	94.18	3.02	769.13	89.73	7.87	943.99
C_{bl}	597.82	6.16	94.11	4.11	770.41	81.4	7.91	999.44
V_{th}	598.69	5.93	94.42	3.12	774.26	98.08	7.89	1007.7
Tox_n	598.19	5.09	94.36	3.53	773.61	90.90	7.82	1220.27
Tox_p	598.11	4.20	94.41	3.09	771.42	85.22	7.96	1003.42

5.4. Optimal Design Flow for Sense Amplifier Optimization

This section discusses the optimal design flow process. The design flow for the optimal design of sense amplifiers is presented in figure 5.7. As in chapter 4, the flow aims to present a method for an optimal design for a sense amplifier with a maximum variance tolerance without increasing the complexity of the circuitry. This includes starting with the baseline design to the final parameters chosen for the design.

This includes starting with the baseline design to the final parameters chosen for the design. The first step involves choosing a baseline design to start with. After the design schematic is completed, a simulation is run to verify its operation. The design is then characterized to identify key FoMs. More simulations are then performed to identify parameters that affect these FoMs.

The next step involves performing a parametric analysis to identify parameters that significantly affect the FoMs. A Monte Carlo analysis is then performed, based on the sensitive parameters and values for parameters the give optimal performance. From the results of the parametric and Monte Carlo analysis, values for the design parameters are then chosen. The values for design parameters where chosen to decrease the precharge time of the circuit. To further optimize the circuit for precharge, an analysis was done on each individual transistor to identify the impact it has on the FoM to be optimized (precharge time in this case).

Algorithm 2 is used to optimize the sense amplifier circuit and deals with individual transistors in the amplifier. The algorithm is heuristic and relies on simple ranking of the transistors. The transistors are ranked accordingly in order of impact on the FoM (precharge), and V_{th} is increased from $V_{th_{low}}$ to $V_{th_{high}}$ and checked against a given delay constraint. The delay constraint is the minimal delay time that the sense amplifier must have for correct operation. A value of 590 ps is selected based on the statistical analysis from table 5.2. If the constraint is met, the V_{th} of the next ranked transistor is also increased and the delay constraint checked again. This is continued until all the V_{th} for all transistors have been increased or until the delay constraint is violated.

The final FoMs for the design are shown in Table 5.3. The final precharge time and power consumption are lower than their starting values. An overall reduction of 83.9% has been achieved on the targeted FoM (precharge). The sense margin is reduced by 17 mV.

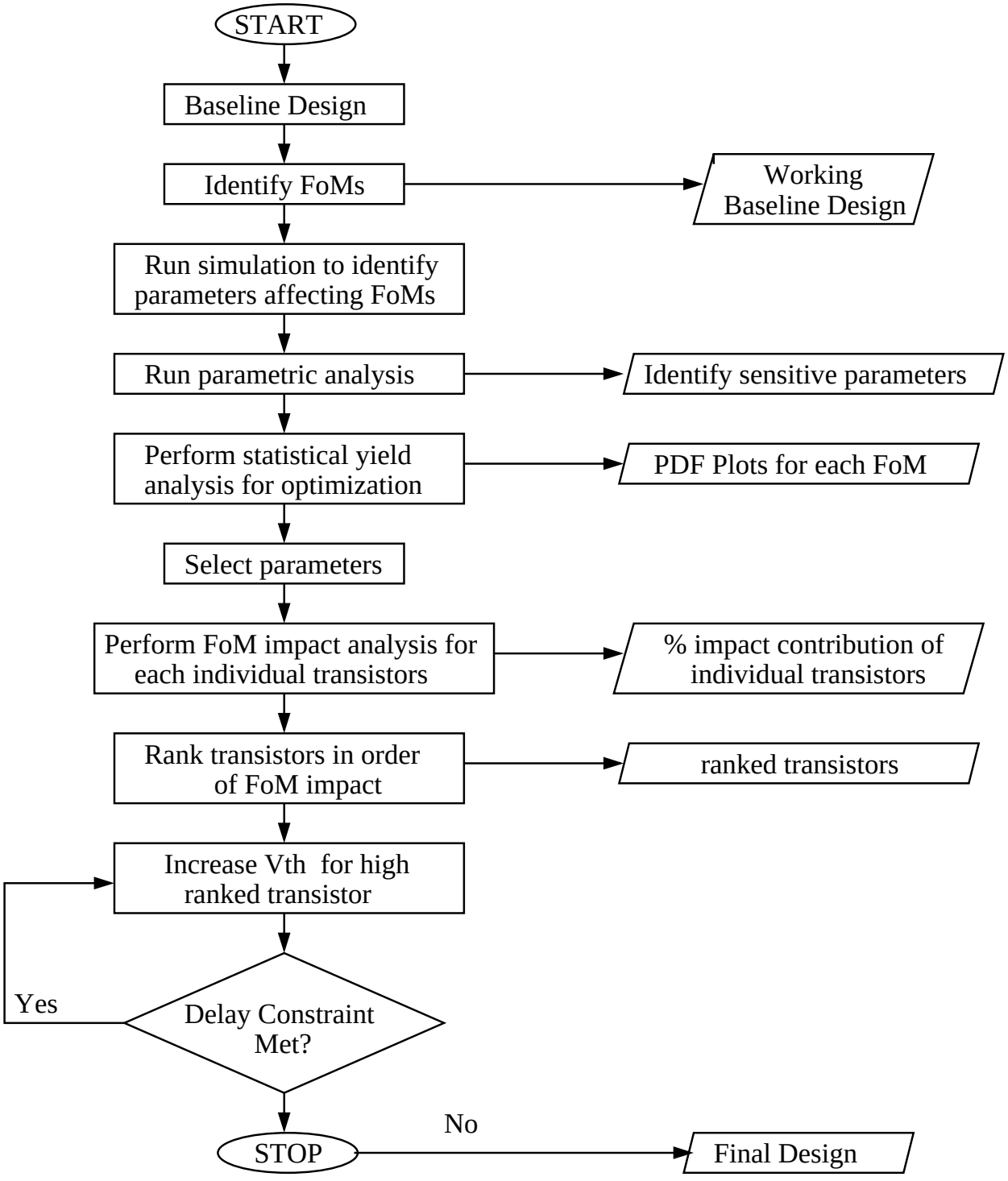


FIGURE 5.7. Optimal Design Flow Diagram

Algorithm 2 Heuristic algorithm for sense amplifier circuit optimization using Dual Threshold Based Technology

1: Create the netlist of the sense amplifier circuit.
2: Number each of the transistors in the netlist from 1 to N.
3: **for** Each of the transistors i = 1 to N **do**
4: Rank the transistors for a figure of merit. For example, contributions to precharge.
5: Identify the significant transistors from the ranks, e.g. 30% or 50%, etc
6: **end for**
7: Start with the highest ranked transistor as i = 0
8: **while** Design constraint of the sense amplifier is met and the transistor M_i is a significant transistor that contributes to the overall FoM value of the sense amplifier **do**
9: Increase threshold voltage of the transistor M_i.
10: Move to the next ranked transistor.
11: **end while**

TABLE 5.3. Figures of Merit Characterization

FoMs	Precharge	Power (μW)	Sense Delay (ps)	Sense Margin (mV)
Baseline	7.5 ns	153.1μW	4.4 ns	43.4 mV
Dual-V_{th}-Optimal	1.2 ns	133.1μW	0.87 ns	70.4 mV
Improvement	83.9%	13.1%	80.2%	61.9%

CHAPTER 6

CONCLUSION AND FUTURE RESEARCH

This thesis studied and analyzed the effects of process variation on the performance of a full latch voltage sense amplifier. A method was presented to choose parameter values that minimize these effects, producing a more tolerant design to process variations. By performing a parametric analysis and a Monte Carlo analysis, a method was introduced to identify optimal parameter values based on FoM's considered. In this research, parameters were chosen to achieve optimal precharge time. The dual oxide and dual V_{th} CMOS technology based techniques were employed in achieving optimal designs. A design flow process was also presented to optimize the sense amplifier design without increasing the complexity of the sense amplifier circuit. The optimization for the precharge time reduced the FoM by 83.9%. Sense delay was also improved by 80.2%, sense margin by 61.9% and power dissipation by 13.1%.

In future research, this type of analysis will be extended to different sense amplifier topologies. The physical layout design of the sense amplifier will also be simulated to analyzed the parasitic effects on the optimal design presented in this research.

BIBLIOGRAPHY

[1] Sherwin P. R. Almazan, Jestoni V. Zarsuela, Anastacia P. Ballesil, and Louis P. Alarcon, *A Study on the Effect of Varying Voltage Supply on the Performance of Voltage Sense Amplifiers for 1-Transistor DRAM memories*, ISCE 2008 Proc. 2008, 2008.

[2] Igor Arsovski, *High-Speed Low-Power Sense Amplifier Design*, Tech. report, University of Toronto, 2001.

[3] Anita Suman Balwant Raj and Gurmohan Singh, *Analysis of Power Dissipation in DRAM Cells Design for NanoScale Memories*, International Journal of Information Technology and Knowledge Management 2 (2009), 371–374.

[4] Aarti Choudhary, *Process Variation Tolerant Self Compensation Sense Amplifier Design*, Thesis, University of Massachusetts Amherst, September 2008.

[5] Aarti Choudhary and Sandip Kundu, *A Process Variation Tolerant Self-Compensating Sense Amplifier Design*, IEEE Computer Society Annual Symposium on VLSI, 13-15 2009, pp. 263–267.

[6] Hwang-Cherng Chow and Chaung-Lin Hsieh, *A 0.5V High Speed DRAM Charge Transfer Sense Amplifier*, Circuits and Systems, 2007. MWSCAS 2007. 50th Midwest Symposium on, 5-8 2007, pp. 1293–1296.

[7] Vinodh. Cuppu, Bruce. Jacob, Brian. Davis, and Trevor. Mudge, *A Performance Comparison of Contemporary DRAM Architectures*, Proceedings of the 26th International Symposium on Computer Architecture., 1999, pp. 222–233.

[8] H. Geib, W. Raab, and D. Schmitt-Landsiedel, *Block-Decoded Sense-Amplifier Driver for High-Speed Sensing in DRAM's*, Solid-State Circuits, IEEE Journal of 27 (1992), no. 9, 1286 –1288.

[9] Heribert Geib, Werner Weber, Erdi Wohlrab, and Lothar Risch, *Experimental Investigation of the Minimum Signal for Reliable Operation of DRAM Sense Amplifiers*, Solid-State Circuits, IEEE Jounal of 27 (1992), no. 7, 1028–1035.

[10] John L. Hennessy and David A. Patterson, *Computer Architecture: A Quantitative Approach*, Elsevier, 2007.

[11] Sanghoon Hong, Sejun Kim, Jae-Kyung Wee, and Seongsoo Lee, *Low-Voltage DRAM Sensing Scheme With Offset-Cancellation Sense Amplifier*, Solid-State Circuits, IEEE Journal of 37 (2002), no. 10, 1356–1360.

[12] Intel.

[13] Anantha Chandrakasan Jan M. Rabaey and Borivoje Nikolic, *Digital Integrated Circuits: A Design Perspective*, Pearson Education, Inc, 2003.

[14] Tao Jiang and P.Y Chiang, *Sense Amplifier Power and Delay Characterization for Operation Under Low-Vdd and Low-voltage Clock Swing*, Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on (2009), 181–184.

[15] Brent Keeth, R. Jacob Baker, Brian Johnson, and Feng Lin., *DRAM Circuit Design: Fundamental and High-Speed Topics*, John Wiley & Sons, 2008.

[16] Duane G. Laurent, *Sense Amplifier Signal Margins and Process Sensitivities*, IEEE Transactions on Circuits and Systems -1: Fundamental Theory and Applications 49 (2002), no. 3, 269–275.

[17] W.K. Luk, Jin Cai, R.H. Dennard, M.J. Immediato, and S.V. Kosonocky, *A 3-transistor dram cell with gated diode for enhanced speed and retention time*, VLSI Circuits, 2006. Digest of Technical Papers. 2006 Symposium on (2006), 184–185.

[18] W.K. Luk and R.H Dennard, *A Novel Dynamic Memory Cell With Internal Voltage Gain*, Solid-State Circuits, IEEE Journal of 40 (2005), no. 4, 884–894.

[19] S. P. Mohanty, *"unified challenges in nano-cmos high-level synthesis"*, Proceedings of the 22nd IEEE International Conference on VLSI Design (VLSID), 2009, pp. 531–531.

[20] Saraju P. Mohanty and Bijaya K. panigrahi, *ILP Based Leakage Optimization During Nano-CMOS RTL Synthesis: A DOXCMOS Versus DTCMOS Perspective,*, Nature Biologically Inspired Computing, 2009. NaBIC World Congress on (2009), 1367–1372.

[21] W. Mueller, G. Aichmayr, W. Bergner, E. Erben, T. Hecht, C. Kapteyn, A. Kersch, S. Kudelka, F. Lau, J. Luetzen, A. Orth, J. Nuetzel, T. Schloesser, A. Scholz, U. Schroeder, A. Sieck, A. Spitzer, M. Strasser, P.-F. Wang, S. Wege, and R. Weis, *Challenges for the DRAM Cell Scaling to 40nm*, Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International, 5-5 2005.

[22] S. Mukhopadhyay, A. Raychowdhury, H. Mahmoodi, and K Roy, *Leakage Current Based Stabilization Scheme for Robust Sense-Amplifier Design for Yield Enhancement in Nano-scale SRAM*, Test Symposium, 2005. Proceedings. 14th Asian, 18-21 2005, pp. 176–181.

[23] Dimitrios S. Nikolopoulos, *DRAM Technology*, University of Crete and FORTH-ICS, December 2009.

[24] S. Rodrigues and M.S Bhat, *Impact of Process Variation Induced Transistor Mismatch on Sense Amplifier Performance*, Advanced Computing and Communications, 2006. ADCOM 2006. International Conference on (2006), 497–502.

[25] Adel S. Sedra and Kenneth C. Smith, *Microelectronic Circuits*, Oxford University Press, 2004.

[26] R. Singh and N. Bhat, *An Offset Compensation Technique for Latch Type Sense Amplifiers in High-Speed Low-Power SRAMs*, Very Large Scale Integration (VLSI) Systems, IEEE Transactions on 12 (2004), no. 6, 652–657.

[27] K.U. Stein, A. Sihling, and E Doering, *Storage Array and Sense/refresh Circuit for Single-Transistor Memory Cells*, Solid-State Circuits, IEEE Journal of 7 (1972), no. 5, 336–340.

[28] Jr. Tasch, A.F. and L.H Parker, *Memory Cell and Technology Issues for 64- and 256-Mbit One-transistor Cell MOS DRAMs*, Proceedings of the IEEE 77 (1989), no. 3, 374–388.

[29] J. Vollrath, *Signal Margin Analysis for DRAM Sense Amplifiers*, Electronic Design, Test and Applications, 2002. Proceedings. The First IEEE International Workshop on, 2002., pp. 123–127.

[30] David Tawei Wang, *Modern DRAM Memory Systems: Performance Analysis and a High Performance, Power-Constrained DRAM Scheduling Algorithm*, Ph.D. thesis, University of Maryland, College Park, 2005.

[31] B. Wicht, T. Nirschl, and D. Schmitt-Landsiedel, *Yield and Speed Optimization of a Latch-type Voltage Sense Amplifier*, IEEE Journal of Solid-State Circuits, 39 (2004), no. 7, 1148–1158.

[32] Wayne Wolf, *Modern VLSI Design: IP-based Design*, Prentice Hall, 2009.

[33] Wangyuang Zhang and Tao Li, *Characterizing and Mitigating the Impact of Process Variations on Phase Change Based Memory Systems*, Microarchitecture, 2009. MICRO-42. 42nd Annuall IEEE/ACM International Symposium on, 12-16 2009, pp. 2–13.