2008

A Process Variation Tolerant Self-Compensation Sense Amplifier Design

Aarti Choudhary
University of Massachusetts Amherst

PROCESS VARIATION TOLERANT SELF COMPENSATION SENSE AMPLIFIER DESIGN

A Thesis Presented

by

AARTI CHOUDHARY

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

September 2008

Electrical and Computer Engineering

PROCESS VARIATION TOLERANT SELF COMPENSATION SENSE AMPLIFIER DESIGN

A Thesis Presented

by

AARTI CHOUDHARY

Approved as to style and content by:

Sandip Kundu, Chair

Massimo V. Fischetti, Member

Wayne P. Burleson, Member

C.V. Hollot, Department Head
Electrical and Computer Engineering

To Family and my Teacher.

ACKNOWLEDGMENTS

I am deeply indebted to my advisor, Professor Sandip Kundu for his constant encouragement and support throughout my graduate studies. His vast technical expertise and insight has given me an excellent background in the field of vlsi design. I would like to thank Professor Max Fischetti and Professor Wayne Burleson for their guidance and valuable feedback throughout my work. I have been fortunate to have worked with knowledgeable and friendly people at the VLSI Circuits and Systems laboratory at UMass. I am thankful to my family and my friends Vandita and Mayank for their constant encouragement at the time when it was needed the most

ABSTRACT

PROCESS VARIATION TOLERANT SELF COMPENSATION SENSE AMPLIFIER DESIGN

SEPTEMBER 2008

AARTI CHOUDHARY

M.S.E.C.E., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Sandip Kundu

As we move under the aegis of the Moore's law, we have to deal with its darker side with problems like leakage and short channel effects. Once we go beyond 45nm regime process variations have emerged as a significant design concern as well. Also embedded memories are now very popular for both processor and ASIC designs. Embedded memories uses sense amplifier for fast sensing and typically, sense amplifiers uses pair of matched transistors in a positive feedback environment. A small difference in voltage level of applied input signals to these matched transistors is amplified and the resulting logic signals are latched. Intra die variation causes mismatch between the sense transistors that should ideally be identical structures. Yield loss due to device and process variations has never been so critical to cause failure in circuits. Due to growth in size of embedded SRAMs as well as usage of sense amplifier based signaling techniques, process variations in sense amplifiers leads to significant loss of yield for that we need to come up with process variation tolerant circuit styles and new devices. In this work impact of transistor mismatch due to process variations

on sense amplifier is evaluated and this problem is stated. For the solution of the problem a novel self compensation scheme on sense amplifiers is presented on different technology nodes up to 32nm on conventional bulk MOSFET technology. Our results show that the self compensation technique in the conventional bulk MOSFET latch type sense amplifier not just gives improvement in the yield but also leads to improvement in performance for latch type sense amplifiers. Lithography related CD variations, fluctuations in dopant density, oxide thickness and parametric variations of devices are identified as a major challenge to the classical bulk type MOSFET. With the emerging nanoscale devices, SIA roadmap identifies FinFETs as a candidate for post-planar end-of-roadmap CMOS device. With current technology scaling issues and with conventional bulk type MOSFET on 32nm node our technique can easily be applied to Double Gate devices. In this work, we also develop the model of Double Gate MOSFET through 3D Device Simulator Damocles and TCAD simulator. We propose a FinFET based process variation tolerant sense amplifier design that exploits the back gate of FinFET devices for dynamic compensation against process variations. Results from statistical simulation show that the proposed dynamic compensation is highly effective in restoring yield at a level comparable to that of sense amplifiers without process variations. We created the 32nm double gate models generated from Damocles 3-D device simulations [25] and Taurus Device Simulator available commercially from Synopsys [47] and use them in the nominal latch type sense amplifier design and on the Independent Gate Self Compensation Sense Amplifier Design (IGSSA) to compare the yield and performance benefits of sense amplifier design on FinFET technology over the conventional bulk type CMOS based sense amplifier on 32nm technology node effective in restoring yield at a level comparable to that of sense amplifiers without process variations. We created the 32nm double gate models generated from Damocles 3-D device simulations [25] and Taurus Device Simulator available commercially from Synopsys [47] and use them in the nominal

latch type sense amplifier design and on the Independent Gate Self Compensation Sense Amplifier Design (IGSSA) to compare the yield and performance benefits of sense amplifier design on FinFET technology over the conventional bulk type CMOS based sense amplifier on 32nm technology node.

INTRODUCTION

Moore's law has been the key foundation and the driving force for breakthrough and evolution in the semiconductor industry. It has been serving the industry and academia marvelously since its evolution but as we continue to move in sub nanometer regime we have to deal with the darker side of the Moore's law [27] which comes from its fulfillment as by allowing nearly exponential increase in the device integration density and performance, it faces some major road blocks due to intrinsic physical limitation of the devices. One of the major barriers that the CMOS devices face at nanometer scale is increasing process parameter variations. Due to limitations of the fabrication process (e.g. sub-wavelength lithography and etching) and variation in the number of the dopants in the channel of short channel devices, device parameters such as length (L), width (W), oxide thickness (T_{ox}), threshold voltage (V_T) etc suffer large variations.

A 50 GHz microprocessor with 2 billion logic transistors using 22 nm drawn channel length (7 nm of effective channel length) devices operating at 250mV supply voltage by first half of the next decade - this is the expected roadmap should the scaling trends continue. Can we achieve this - maybe, maybe not!. To be able to even dream about such a processing system, it is important to be able to do predictive design. The old and easy way of designing for worst-case will not be adequate. It is important to accept that process variation is a reality and that one has to design circuits, with variations in mind.

Variations in the device parameters both systematic and random lead to loss in the parametric yield. The circuit yield loss due to process and device parameter variations has never been so critical and is getting some serious attention now. Moore's law

enabled us to integrate large memory blocks with logic circuits on a single chip,but this is also true that performance of the on-chip memory limits the speed and performance of the overall system. The key limiting factor is the increasing bit line capacitance, which results in increased time to develop bit line differential voltage.

For fast and power efficient memory design, both time and signal swing on the bit lines needs to be minimized. A sense amplifier is used to generate full rail output voltage using minimum bitline differential voltage or current making fast read write possible in memories. Designing high performance and robust sense amplifier is extremely important for designing SRAMS but with increasing parameter variations developing a robust, reliable and fast sense amplifier is becoming a task in itself. With each shrinking node it is important to accurately model the impact of device parameter variations at the circuit level and develop process tolerant design for sense amplifiers with improved performance and reduced impact of the leakage to reduce the gap between the CPUs and memories.

Proceeding in this direction we came up with the novel process variation tolerant Self Compensation Sense Amplifier (SSA) design on 32nm technology node on conventional bulk type CMOS technology. It demands only few additional transistors which compensates for the process variations since the single failing sense amplifier implicates the whole memory, there is one sense amplifier across each the column of the SRAM (128X8) so additional few transistors doesn't hurt the area overhead that much.

The trade off on added transistors in each sense amplifier in a column of SRAM to an improvement in the yield is definitely a solution to a problem. The fact that CMOS technology is at the cross-road today. Oxide scaling has halted due to gate leakage problem [40]. Changing oxide material provides an alternative solution for only 1-2 generations of technology [40]. Power density has become so high that power supply voltage must be scaled down which will require scaling of V_T which aggravates

the sub-threshold leakage problem. The ITRS predicts that static power dissipation per device will surpass the dynamic power dissipation by 2008.To reduce leakage while improving performance, ITRS predicts using strained silicon channels,ultra-thin single-gate FETs, and metallic gates. Two types of structure and materials changes must be considered. First, there are structures that allow a shorter channel length to be fabricated. Second, there are materials that enable higher performance for a given channel length. For the past few decades, new materials like strained silicon is used by IBM, Intel [17]. SOI based device structures incorporated by AMD, Honeywell, IBM has resulted in exponential growth of performance and integration density of silicon CMOS technology. One common condition for each innovation is that it was built on the advantages of existing ideas, without compromising the previous. Historically, the switch from metal gates to polysilicon gates, the switch from diffusion to ion implantation,and the incorporation of silicides were major changes that were difficult to implement, even though the transistor structure itself was little changed. Ultra thin body SOI FETs employ very thin silicon body to achieve better control of the channel by the gate and hence also gets advantage of reduced leakage and short channel effects [36].

Use of intrinsic or lightly doped body reduces threshold voltage (V_T) variations due to random dopant fluctuations and enhances the mobility of careers in the channel region and improvement in ON current. Researchers are exploring alternative technologies and this is a known notion that better scalability can be achieved by introduction of multiple gates at the other side of the body of each transistor resulting in a Double Gate (DG) SOI structure. CMOS device with a second gate for each device are referred to as 4-Terminal devices. In such technologies, one can choose to connect the back and front gates together or to control them separately while designing a circuit resulting in new circuit styles [36]. Connected back and front gates provides a simple way of mapping circuits designed in single gate technologies to double gates

technologies. 3 Terminal configurations provide more ON current for transistors as well. On the other hand, independent gate control (4 Terminal configurations) can be used for designing new circuit styles.Back gate bias can be used to dynamically adjust the threshold voltage of the front gate to tune the power and performance requirement of a circuit. It can also be used for merging parallel transistors, the nature of these devices can be exploited to come up with new circuit styles. The three alternatives that are most widely considered for conventional MOSFET replacement are: planar device, vertical (pillar) device and the FinFETs .

The FinFET technology is the most promising among the alternatives to conventional bulk CMOS [15]. FinFETs increase drive current through larger gate area while they reduce sub-threshold leakage through reduced channel doping. FinFETs have been successfully fabricated by multiple laboratories [20].A FinFET is a vertical double gate device that is promising below 45nm technology [20, 34].A double-gate FinFET structure offers additional configuration possibilities such as single channel or double-channel, whereby two gates create their own independent channels or one common channel. Similarly, the gates may be symmetric or asymmetric in terms of gate capacitance and current drive [19, 22].FinFETs increase drive current through larger gate area while they reduce sub-threshold leakage through reduced channel doping.A FinFET is a vertical double gate device that is promising below 45nm technology [50] motivated by these considerations and feasibility for ease of manufacture and implementation, we chose FinFET to solve the yield problem. We propose Independent Gate Self Compensation Sense Amplifier Design (IGSSA), which is process variation tolerant self compensating FinFET based sense amplifier design.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

This document is organized as follows, Chapter 1 presents the background and builds up the motivation on the basis of road-maps and where the trend of semiconductor industry is leading to. It point us to present known problems and possible solutions to those. It builds up the motivation as to why we chose the sense amplifier to solve the problem of process variations. Chapter 2 describes different types of sense amplifiers. Chapter 3 describes the impact of process variations and mismatch effects on the sense amplifier causing problems with yield and performance. Chapter-4 we propose the new self compensation sense amplifier (SSA) technique on one of the current mode latch type sense amplifier gives a comparison of this circuit with the existing current latch type sense amplifier. In Chapter 5 we describe the modeling of the double gate cmos its promises and challenges. Chapter 6 describes the Independent Gate Self Compensation Sense Amplifier design (IGSSA) on FinFET technology and we compare it with the SSA and nominal latch type sense amplifier. We generated our models through TCAD simulation from synopsys [47], and from DAMOCLES [25] 3D device simulation and used those in our sense amplifier depicting significant improvement in yield, performance and pvt variations.

1.1 Benefits of MOSFET Scaling

Significant advances in silicon integrated circuit(IC) technology led by the continued miniaturized of the MOS transistor has enabled us increase in the computing power. The rapid progress in the semiconductor industry has been driven by im-

proved circuit performance and functionality together with reduced manufacturing costs. Since the 1960s MOS transistor have been shrinking 30% every three years, as predicted by Moore's law [27] shown in Figure 1.1.

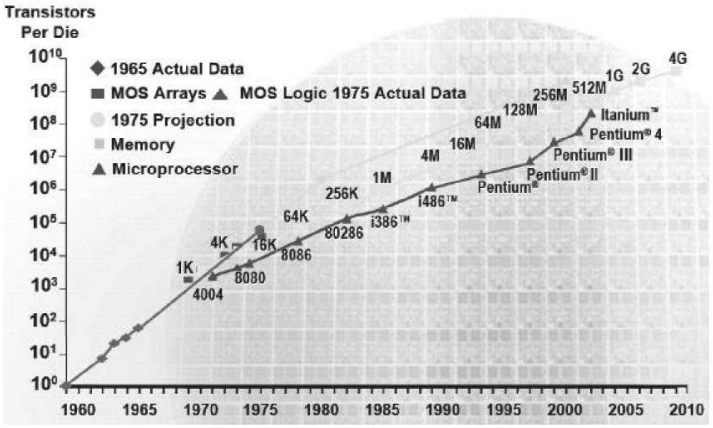


Figure 1.1. Moore's law of scaling the number of transistors on a chip has been increasing exponentially.

While Moore's law only describes the rate of increase in transistor density, reduction of the physical MOS device dimensions has improved both circuit speed and density in the following ways:

a) Circuit operational frequency increases with a reduction in gate length, L_G as $F_max \quad 1/L_G$; allowing faster circuits. b) Chip area decreases ${L_G}^2$ enabling higher density and cheaper ICs. c) Switching power density is constant allowing lower power per function or more circuits at the same power.

Device scaling has been a relatively straightforward affair thus far, but physical limits are fast being approached, and new materials and device structures are needed to continue scaling trends.

1.2 Issues in Planar Bulk-Si MOSFET Scaling

The planar bulk-silicon MOSFET has been the workhorse of the semiconductor industry over the last 40 years. However the scaling of the bulk MOSFETs becomes increasing difficult for the gate lengths below 20nm (sub-45nm half pitch technology node). As the gate length is reduced, the capacitive coupling of the channel potential to source and drain increase relative to the gate, leading to significantly degraded short channel effects (SCE). This manifests itself as a) increased off-state leakage b) threshold voltage (V_T) roll-off,i.e. smaller V_T shorter gate lengths, and c) reduction of V_T with increasing drain bias due to a modulation of the source-channel potential barrier by the drain voltage, also called drain induced barrier lowering (DIBL).

In order to maintain the relatively strong gate control of the channel potential in bulk devices, various technological improvements such as ultra-thin gate dielectrics,ultra shallow source/drain junctions, halo implant and advance channel dopant profile engineering techniques such as super steep retrograde well have been necessary. Each of these technologies is now approaching fundamental limitations which may in turn, limit further scaling of device dimensions. Significant scaling difficulties have been encountered already and are expected to worsen in the next few years as the gate length (L_G) is projected to scale to well below 32nm [40]

With transistor scaling Vdd is also scaled down, However the threshold voltage, V_T cannot be scaled down significantly since the source/drain sub threshold leakage current Isdleak increase sharply with decreased V_T and it is important to keep Isdleak within tolerable limits because I_{dsat} depends on (Vdd-V_T), the scaling of Vdd tends to reduce I_{dsat} and hence make it difficult to improve the transistor. Short-channel effects (SCEs) such as drain induced barrier lowering [DIBL] are becoming very difficult to control as transistors are continuously scaled. This tends to lead to increased leakage current and reduces I_{dsat} and hence reduced transistor speed. Channel doping is becoming very large (both to set the threshold voltage correctly and to control SCEs),

leading to degradation in the mobility and to increased leakage current due to band to band tunneling.

Scaling of the gate dielectric equivalent thickness, T_{ox} is limited by gate leakage current. With scaling, random dopant fluctuations and line edge roughness can cause significant statistical variation in V_T and effective channel length L_{eff} [32].

In MOS devices, the gate dielectric thickness is single most important device dimension to enable device scaling and has also been the most aggressively scaled one. A thin gate dielectric increase capacitive coupling from gate to the channel, thereby reducing the source/drain influence on the channel. A larger gate capacitance also leads to a larger inversion charge density, or increased ON-state drive current. However gate dielectrics are already so thin that quantum mechanical direct tunneling through them results in significant gate leakage current below 20Å, The use of alternative high - K dielectric material can provide a small effective oxide thickness to maintain adequate gate control needed for L_G scaling while proving a large physical barrier to gate-oxide tunneling reducing gate leakage. Reduction of the source/drain extension junction depth directly decreases capacitive coupling of the drain to the channel, thus also reduces drain-induced short channel effect. Shallow source/drain to the channel formation required that low-energy ion implantation together with low thermal budget dopant activation to minimize dopant diffusion. The downside to this is the increase in the parasitic series resistance of the source and drain extension regions.

Raised source and drain technologies can alleviate the extrinsic resistance problem while maintaining shallow junctions. The contact resistance associated with the metallic contact to the source and drain regions is another source for parasitic series resistance and is expected to dominate total parasitic resistance of the device. In order to scale bulk-Si transistors, heavy body doping is also necessary to eliminate leakage path far from the gate dielectric interface and to increase back-gate (substrate)

control of the body. For sub-100 nm gate length devices, a strong halo implant is generally used to suppress sub-surface leakage but this tends to increase the average channel doping in small L_G devices. However high channel doping concentration, reduce carrier mobility due to impurity scattering and increase transverse electric field, increase sub threshold slope, enhances band-to-band tunneling leakage and increase depletion and junction capacitances. These factors combine to significantly degrade devices performance.

In summary from device design point of view, in order to achieve good electrostatic integrity or good control of short channel effects (SCE), the gate dielectric thickness, T_{ox}, the source,drain junction depth, X_j and the channel depletion depth X_{DEP}, needs to be scaled down. The scale length for a bulk device λ_{BULK}, is an indication of how short L_G can be made before the SCE are excessive and is expressed in equation 1.1.

$$\lambda_{BULK} = 0.1(T_{OX}X_J X_{DEP}^2)^{1/3} \tag{1.1}$$

1.3 Process Induced Variations

As the scale gets smaller, consistency of process control also decreases [39]. The semiconductor process cannot be perfectly controlled, which leads to statistical variation of many process variables. Several types of process variations can occur: line to line, batch to batch, wafer-to-wafer, die-to-die and intra die (within chip) [8, 30].

As per Sani Nassif [30] as shown in 1.2, there is 10% is V_T and 40% mismatch in the Leff.

If we extrapolate this graph as shown in 1.3for another 10 yrs,we see mismatch effect get even worse within a next decade.

Control of critical dimensions (CDs) in device and interconnects such as L_G in devices continues to be a difficult challenge, as the physical gate length is considerably smaller than the lithography printed line width [18, 8, 30]. In order to limit

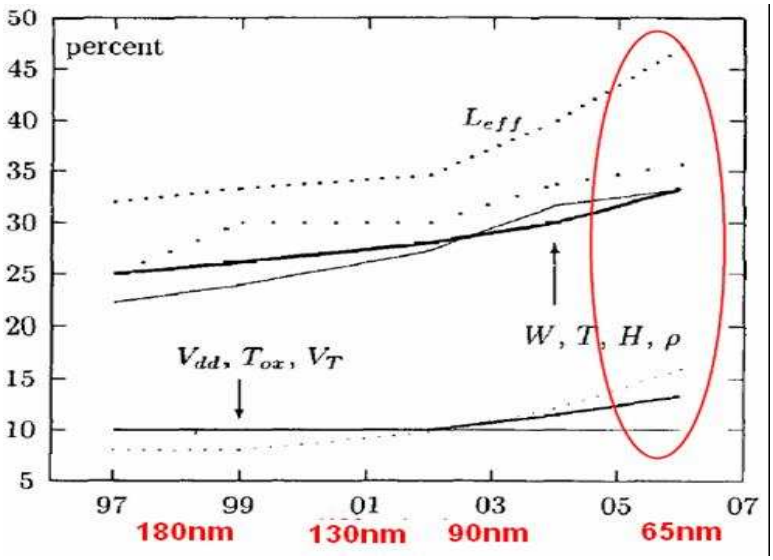


Figure 1.2. Device and interconnect (intra-die) variation trends.

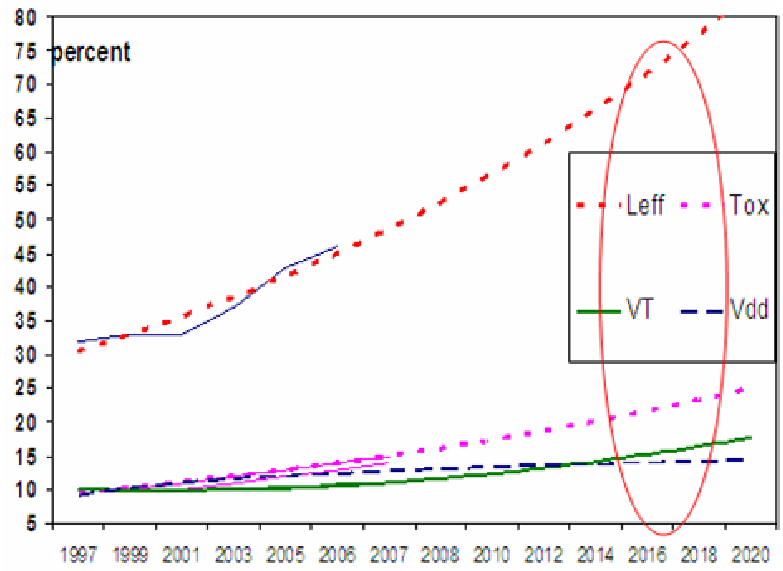
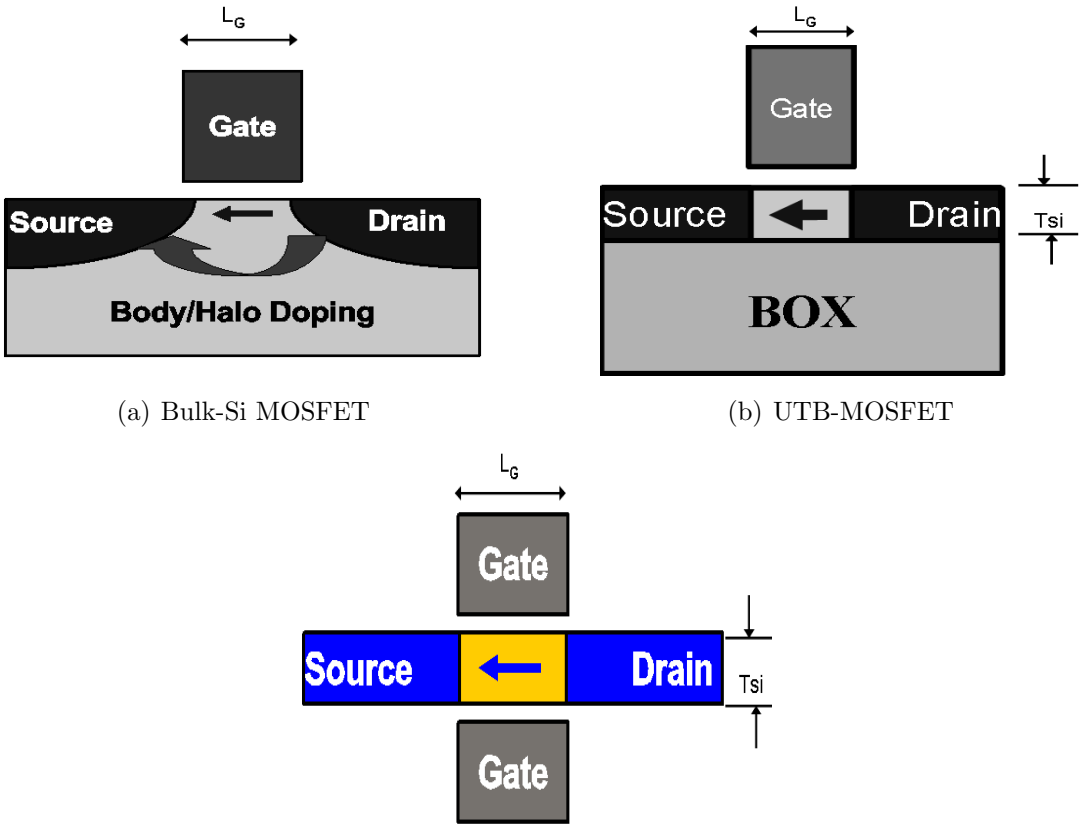


Figure 1.3. Technology Parameter Variations for next 12 yrs.

the impact of variations, the semiconductor industry is actually using slightly larger physical gate lengths than those specified in the ITRS, especially for memory applications .The slowing down on L_G scaling is now unavoidable in the future since the control of process variable does not track the scaling of minimum feature sizes. This is particular important for memory arrays, because if the desired degree of dimensional control were not achievable, design margins would need to be relaxed to achieve large functional memory arrays. While advanced process control can minimize systematic shifts in the CD the role of random variation arising from statistical dopant fluctuations and line edge roughness is expected to increase, so that variation will impact the overall power dissipation and performance [3]. Therefore, statistical treatment of random variation of circuits (statistical design) is becoming increasingly important. New transistor structures should have better immunity to process variations, and devices with tunable V_T are beneficial to counter any systematic shifts in transistor characteristics.

1.4 Thin-body MOSFETs

Figure 1.4 shows the transistor structures such as Ultra Thin Body (UTB) and the Double Gate MOSFET. They eliminate sub-surface leakage paths and extend scalability of Si CMOS technology. As the bulk MOSFET is scaled down, the control of short channel effects become increasingly difficult leading to increased sub threshold leakage current.This is because the source/drain influence over the channel potential becomes significant relative to the gate control. From equation 1.1, it is clear that if X_j and X_{DEP} can be reduced aggressively, it is possible to scale the MOSFET down to every small L_G. This is precisely what is done in the case of ultra-thin body (UTB) silicon on insulator (SOI) devices [5, 4]. Where X_j and X_{DEP} are physically limited to the thickness of an ultra-thin silicon film equation 1.1 qualitatively described the scaling behavior of UTB thus the scalability of MOS devices can be improved by using



(a) Bulk-Si MOSFET

(b) UTB-MOSFET

(c) DG MOSFET

Figure 1.4. Advanced transistor structures such as the UTB and the DG-MOSFET eliminate sub-surface leakage paths

an ultra thin silicon body such that all point in the silicon channel are close enough to the gate and well controlled by it, thereby eliminating sub-surface leakage currents [21]. The conventional fully depleted SOI MOSFET (with a thick body) is known to have worse short channel effects than bulk MOSFETs and partially depleted SOI MOSFETs [43]. Also partially depleted SOI (PDSOI) have the floating body effects [26].

Double Gate MOSFET has the best scalability, down to sub-10nm L_G devices [44, 16, 11]. The improved scalability of thin body devices makes them attractive for future generations of CMOS technology and so they have been included in the international technology road map for semiconductors (ITRS) [40] .

UTB devices can be implemented in a straightforward manner as planar single gate fully depleted silicon on insulator (FDSOI) devices. While the planar double-gate device has been demonstrated [53], the fabrication of a planar double gate FET with a bottom gate that is aligned to the top gate and source/drain regions imposes numerous process challenges. Among all DG structures proposed so far, the FinFET is the most manufacturable because it eliminates the need for the bottom gate by rotating the channel by 90 and placing the gate electrodes on the two sidewall of the silicon fin [1, 34, 20, 15] Independent gate FinFETs in which the front and back-gate electrodes can be independently biased have been demonstrated as well [22, 37]. The front gate can be used to switch the device, whereas the back gate can be used to set the correct V_T. The back gate is as strong as front gate, and therefore the devices have degraded sub-threshold slope and transconductance due to capacitive division of the channel potential between the two gates [37].

Figure 1.5 shows the top view of the FinFET devices. In this figure,L_{eff} = physical gate length, L_{gate} = drawn gate length, T_{si} = silcon thickness, T_{oxf} and T_{oxb} are the front gate and back gate oxide thickness

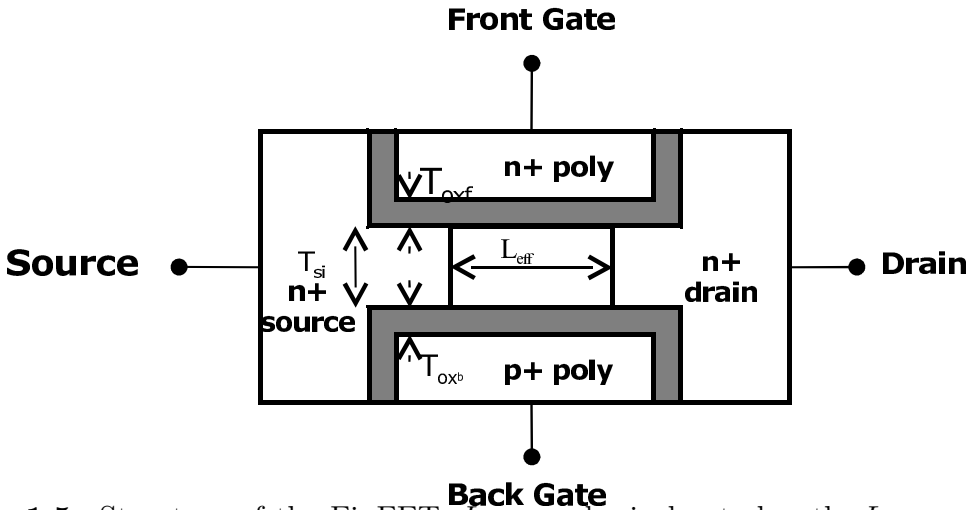


Figure 1.5. Structure of the FinFET. L_{eff} = physical gate length, L_{gate} = drawn gate length, T_{si} = silcon thickness, T_{oxf} and T_{oxb} are the front gate and back gate oxide thickness.

In this research work we have modeled the FinFET devices, it is explained elaborately in chapter-3. The planar FDSOI MOEFET can be extended to include a conducting electrode underneath the buried oxide (BOX) layer to form a second gate to control the channel from below. This ground plane or the back gate act as a second gate to shield the field penetration from the drain into the channel, and improves SCE. In a way it serve the role of the retrograde doping in a bulk MOSFET, by raising the body backside potential and by terminating drain electric fields shown in Figure 1.6

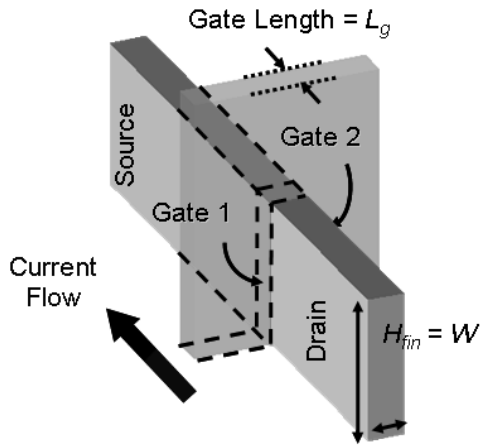


Figure 1.6. Double Gate FinFET.

In addition the, BOX eliminates source/drain-to-substrate depletion capacitance. In order to prevent electric field penetration through the BOX, the BOX layer should be thin. Another benefit of a thin BOX is the "back gate effect' similar to the "body effect' in bulk SI devices wherein the V_T can be tuned by the back-gate voltage . However the sub threshold slope and the transconductance are degraded due to capacitive division of the channel potential between the front and the back-gate potentials. While the early FinFET devices were fabricated on SOI wafer, FinFETs on bulk-Si wafers have been demonstrated as well [1, 34, 20, 15] .

Bulk FinFETs have the advantages of being potentially cheaper and can be easily integrated with conventional bulk-Si CMOS technology. Bulk FinFETs combine the benefits of good leakage and short channel effects controlled together with a cheaper manufacturing process, making them attractive for high density memory applications.

1.5 Motivation and Road Maps

1.5.1 Motivation

As CMOS technology progresses rapidly towards the nanometer regime,the integration level, performance and fabrication cost increase tremendously. Thus only high performance system chips design that integrate CPU (Central Processing Unit), DSP (Digital Signal Processing) processors or multimedia processors,memories,logic circuits analog circuits etc can afford the sub nanometer technology. Embedded memory has become a key component of any system and more practical than ever for at least two reasons. 1. Data processing and storage are the most primitive and basic components of the digital circuits. 2. Memory bandwidth is now one of the most serious bottlenecks to system performance.

The speed gap between the MPUs and memory devices has been increased in the past decade and we have already hit the 'memory wall' [54]. The MPU speed has improved by a factor of 4 to 20 in the past decade. On the other hand, in spite of the exponential progress in storage capacity, minimum access time for each quadrupled storage capacity has improved only by a factor of two. Memory cell is the fundamental component of a memory system. The most important objective in the design of the memory cell is to minimize the size of the cell, which decreases the cost per bit, access time and power dissipation of the memory systems. Caches and other on chip memories requires very fast access time and also at the same time easy implementation. Both of these objectives are satisfied by SRAM Figure 1.7 shows the basic 6T SRAM cell.

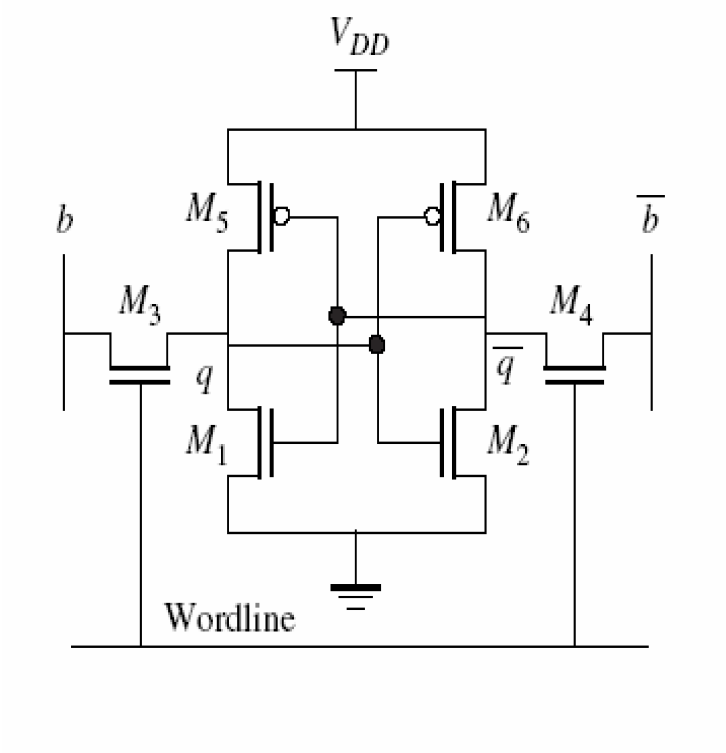


Figure 1.7. Basic 6T SRAM Cell.

Since the size of the memory cell needs to be small most of the transistors are maintained close to minimum feature size of the process technology.One of the most critical circuits in the periphery of a memory is the sense amplifiers (SA). These circuits are strongly related to read access time of a memory, as they are used to retrieve the stored data in the memory array by amplifying small signal variations on the bit-line. The design of fast low-power and robust sense amplifier circuits is a challenge, due to the fact that in modern memory designs bit-lines exhibit significant capacitance,since there is higher number of cells per bit line, which sets limits in the sensing speed and introduces extra signal delays. Smaller the sense voltage, the faster the memory will be. The sense voltage also reduces power dissipation as the bit line can be restored to their default state as soon as a value has been sense and latched. If the sense transistors are perfectly matched, the lower limit on the sense voltage is set by the various noise margins and can be as low as 60-70 mV [38]. However

if the transistor mismatch the sense voltage must be increased to account for such mismatches, otherwise yield will suffer as the sense amplifier may show incorrect values [41]

Predicting and improving the design quality in terms of performance, yield and robustness are a central concern in designing sense amplifiers for use in SRAM cells. In a large embedded SRAM there may be many thousands of sense amplifiers and each one of them will have a very high yield requirement for the product to have a good overall yield in simple words "*All units of one chip have to work in order to make the whole chip work*".

A single failing sense amplifier implicates the whole memory but as CMOS IC technology becomes more and more advanced and Moore's law still governing the semiconductor industry, the control of process variations and manufacturing uncertainty becomes more and more critical. The sense amplifier performance degradation due to process variation and resulting yield loss is more pronounced than before [38]. It is important for the designer to be able to understand mismatch effects, since sensing should be done as fast as possible, subject to sensitivity constraints imposed by the parameter variations inherent in fabrication processes.

For low power application it is always desired to have low bit line differential voltage. The differential sense amplifiers are in general designed to be electrically and topologically symmetrical. Despite careful designing, small variations in parameters like threshold voltage and effective channel length due to process operating conditions [41, 41], lead to an input offset which affects the performance of the sense amplifier. The analysis of these variations provides a good understanding of the impact of device mismatches in differential sense amplifier. [42]. To guarantee reliability of the sense amplifiers either we have to come up with new process variation tolerant circuit styles, techniques or the replacement of conventional bulk type MOSFET and this is specially required for those circuits which rely on symmetry or matched transistors like sense

amplifier, srams, latches [7, 10, 29, 48] There has been lot of work done on similar lines which we will be discussing more in detail in chapter -2.

1.5.2 Road Maps

To understand how Moore's law has run into trouble at 45 nm, we need to understand something about how CMOS transistors got evolved fulfilling Moore's law at every technology node. In an ironic twist of fate, the new CMOS transistor technology actually hearkens back to the earliest transistor implementations. When we make a layout through any VLSI design tools, we use the Mead-Conway design Rule: "Poly over silicon" produces a transistor The Mead-Conway rule was actually shorthand for a layer of poly-Si over an implied silicon dioxide insulator over a doped, over a Si substrate with implied source and drain, leads to a transistor being produced from photolithographic masks generated by VLSI CAD tool. Gorden Moore was trotted out to publicly proclaim: *"The implementation of high-k and metal gate materials marks the biggest changes in transistor technology since the introduction of polysilicon gate transistors in the late 1960s*"

Here Hafnium metal is used in the so-called high-k gate oxide layer, not the metal gate itself. The key issues at each technology node defined are usually the same.Each of these issues comes into play with different level of significance at each SIA node. In fact, it turn out that Moore's law has died many little deaths. The 2007 International Technology Roadmap for Semiconductors (ITRS) makes projections covering the next 15 years, through 2022 [40]. For the transistors these projections include the scaling of key parameters such as gate length, gate dielectric thickness, transistor leakage and drain saturation current, transistor speed performance, etc. In addition the ITRS assesses the main challenges to scaling MOSFETs, as well as the key technology innovations needed to overcome these challenges. The key target of ITRS scaling is to continue the historic 17 percent per year improvement in transistor performance, i.e.

in CV/I. This is important because the transistor performance improvement is critical components of the overall chip speed increase the scaling. The current mainstream transistor is planar bulk (or partially depleted silicon on insulator [PDSOI] which scales similarly to planar bulk).In order to help deal with scaling issues and to enable the targeted 17 percent per year transistor performance improvement while holding the leakage current to reasonable levels, a number of key technological innovations are necessary. The first of these is enhanced mobility due to applied strain, which was implemented in production in 2004, and has been continuously improved since then. This mobility enhancement allows increased I_{dsat} and is critical enabler for the desired transistor performance improvement. To continue to meet the targeted performance improvement, it is essential that the enhanced mobility be maintained as transistors are scaled. High K metal gate has already been incorporated, with AMD moved to Ultra thin body on 65nm technology. SIA road-map puts FinFETs as the prime candidate for post-planar CMOS device. FinFETs typically have un-doped, fully depleted channels because of the structure the electrostatic integrity and hence the ability to control SCEs are generally superior to planar bulk transistors. Furthermore the mobility should be superior because of the lack of doping. Finally because V_T is set by the work function of metal gate electrode random dopant fluctuations do not impact the statistical variability of V_T for these transistor types. Because of all these advantages the multiple-gate transistor will significantly improve scaling. Because SCEs increase notably, clearly the multiple gate transistor scales the best, while the Ultra thin Body Fully Depleted SOI scales in between the planar bulk and the multiple gates, which improve the electrostatic integrity and hence the control of SCEs with scaling. ITRS projects the ultimate MOSFET (for sub-15nm gate length transistors in the Road-map) is the multiple-gate transistors. Multiple parallel paths are envisioned, where for several years two or even all the three transistors types coexist.

CHAPTER 2

SENSE AMPLIFIERS

Sense Amplifier is the most critical circuits in the periphery of CMOS memory [23]. The performance of SA's strongly affects both memory access time, and overall memory power dissipation. As with other ICs today, CMOS memories are required to increase speed, improve capacity and maintain low power dissipation. These objectives are somewhat conflicting when it comes to sense amplifier in memories. With increased memory capacity usually comes increased bit line parasitic capacitance. This increased bit-line capacitance in turn slows down voltage sensing and makes bit-line capacitance swings energy expensive resulting in slower more energy hungry memories. Due to their great importance in memory performance sense amplifiers have become a very large class of circuits. Their main function is to sense or detect stored data from a read-selected memory cell. Figure 2.1 shows a typical position of sense amplifier in a column of memory cells.

Sense amplifiers are used to translate small differential voltage to a full logic signal that can be further used by digital logic. The need for increased memory capacity, higher speed, and lower power consumption has defined a new operating environment for future sense amplifiers. Below are the some of the effects of increased memory capacity and decreased supply voltage.

1) Increase in the number of memory cells per bit-line increases while as increase in the length of the bit-line increases

2) Decreasing memory-cell area to integrate more memory on a single chip reduces the current that is driving the now heavily loaded bit-line. This coupled with increased

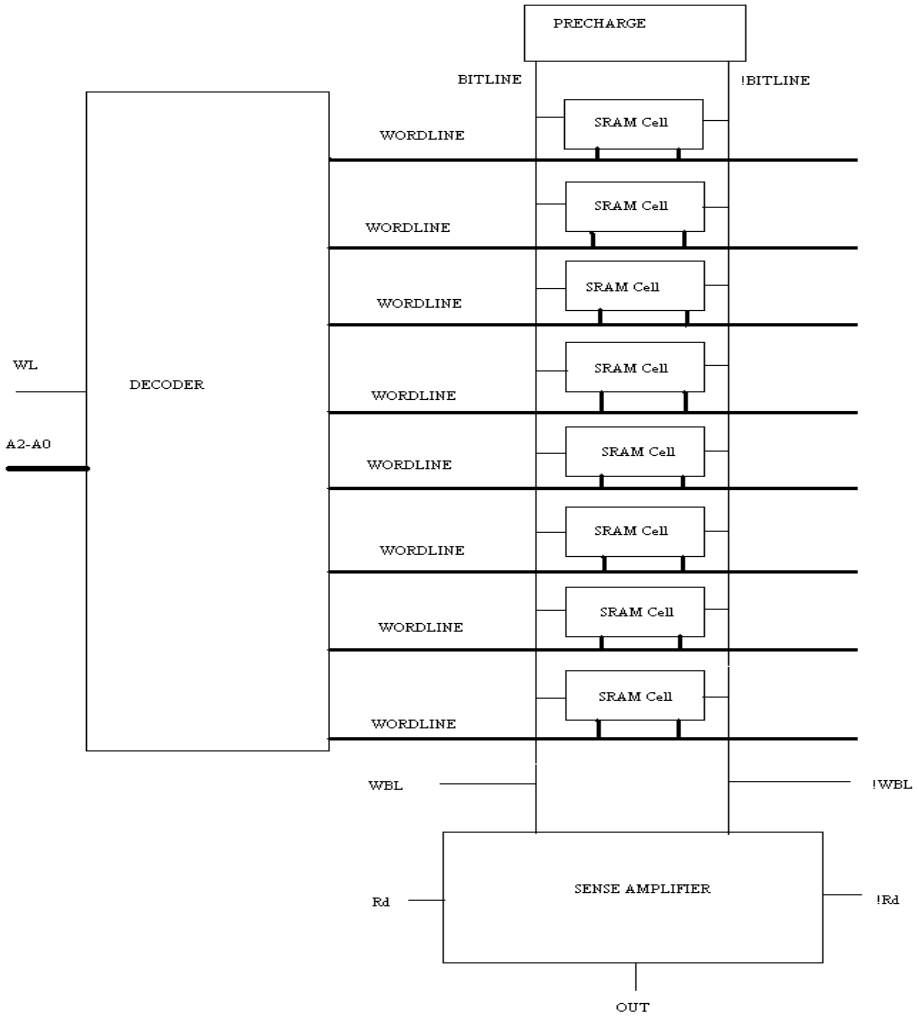


Figure 2.1. Sense Amplifier across a rows of SRAM cells.

causes an even smaller voltage swing on the bit-line.

3) Decreased supply voltage results in smaller noise margins which in turn affect sense amplifier reliability. Shekhar Borkar from Intel suggests

"*Sense amplifier are not tractable in future and will go away but not in memories*"

Projections from the National Technology Roadmap for Semiconductors call for operating voltage of CMOS logic to drop by about a factor of 0.7 to 0.8 per generation to keep power dissipation in check. For the sense amplifier and other support circuit to work properly at reduced voltage requires scaling of the V_T of those devices. For a differential amplifier to sense correctly has a current mirror type of configurations as shown in Figure 2.2.

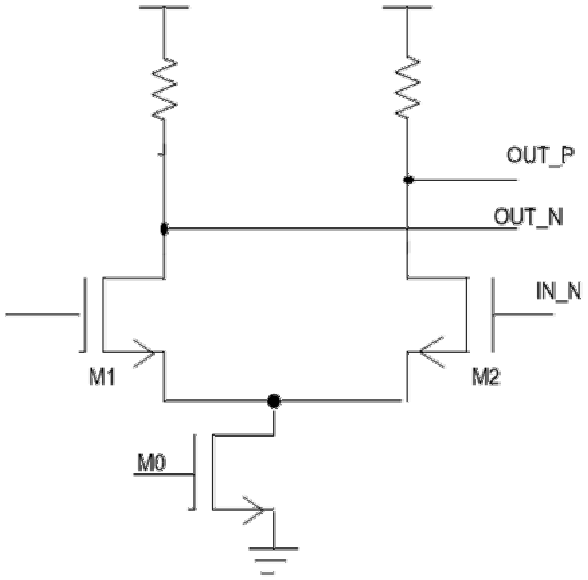


Figure 2.2. Differential Senseamplifier

For Sense amplifier to work in differential mode, M0, M1 and M2 should operate in Saturation Mode. The tail current should be I_{ref}=I_{DS} = $\frac{k}{2}\frac{W}{L}$(V_{GS}-V_T)^2 For M0, M1 and M2 to remain in Saturation Vdd > V_{Tn} (in 0.18 um, a 0.25 V margin is typical)

At VDD < 1V scaling V_T leads to off current problems, such low V_T devices are not feasible in order to meet power requirements and reduce standby power post 45

nm technologies alternatively different sensing circuits may be developed for lower voltage operation.

Since SRAM requires balancing act between delay, area and power consumption. The circuit styles for the decoders and the sense amps,transistor sizing of these circuits,interconnect sizing and partitioning of the SRAM array can all be used as a tradeoff. With technology scaling the transistor mismatch doesn't scale, thus the delay of the output mux doesn't scale. The nonscaling of threshold mismatches with process scaling causes the signal swings in the bit line and data lines also not to scale, leading to an increase in the gate delay of the SRAM across technology generations. This delay increase for most of the SRAM organization can be mitigated by using more hierarchical designs for the bitline and data line path and using offset compensations scheme in sense amplifiers.

The sense amplifiers that are used in SRAM are mainly differential in nature. The differential sense amplifiers can distinguish smaller signals from noise due to their high common mode rejection ratio providing good reliability.

The differential sense amplifier is designed to be electrically balanced symmetric circuits. However due to process variations all the devices in the circuit does not have the same characteristics and this leads to variation in the design parameters if the circuit. The results of such variation in the device parameters are an offset voltage or current. An offset is a voltage or current difference which appears between the two output nodes, when an identical input voltage or bias circuit is applied at the inputs. This offset might affect the sensing delay or even the functionality of the circuit depending on the extent of process variation on each device. Therefore determining the worst case possibility of variation in each circuit is highly significant in the design of a sense amplifier.The offset voltage can be calculated as a function of the threshold voltages and dimensions of the transistors in the sense amplifier as shown in equation 2.1.

$$V_{OS} = f(V_T, I_D, W_i, L_i) \tag{2.1}$$

The following section discusses and analyzes different type of differential sense amplifier circuits. Especially latch type implementation due to their capability of positive feedback they are extensively used in memories [23, 52] This configuration makes possible to restore data in DRAM cells simply with increase in differential gain of the sense amplifier and reduces switching times and delays in the sense circuits thus has the benefit of fast speed and low power consumption, which makes them hot favorite for caches and low power embedded SRAMs [52, 23].

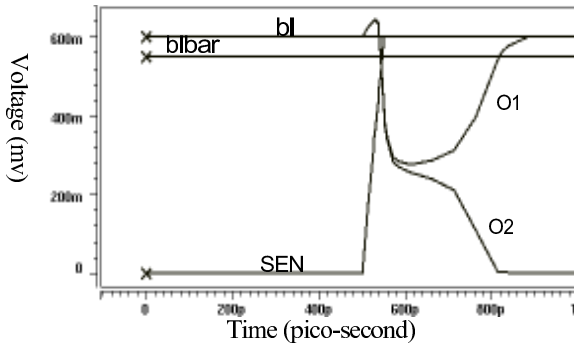


Figure 2.3. Sense Amplifier Operation.

To act as a sense amplifier, the cross coupled latch is initialized in its metastable point by equalizing bit-line and bit-linebar. Once a voltage difference is built over the two.The sense amplifier is enabled by raising 'SEN' [52]. Depending on the value of bit-line and bit-linebar the cross coupled pair traverses to one of its stable operation point as shown in the Figure 2.3.

Offset is particular and inherent to differential sense amplifiers,and it is the voltage or current difference which appears between the two output node potentials or between the two output currents,when an identical voltage or bias current is applied to the two inputs. The offset voltage or current has to be counteracted by the memory-cell generated signal for correct sense operations. Theoretically, differential sense amplifiers are electrically balanced symmetrical circuits. In practical implementa-

tions both the transistors and the passive elements have slight parameters differences inspite of the utmost design efforts to assure their symmetrization. These parameter difference and the resulting sense amplifier offsets are distributed spatially throughout the chips,wafer and lots and the signal generated by a memory cell has to act against and neutralize the appearing maximum offset before the sensing of the data signal could start. Thus, the offset limits the sensitivity i.e. the minimum data signal amplitude that the circuit can detect and it delays the effective start of data sensing. To improve both sensitivity and sensing speed the offsets should be kept small by minimizing the imbalances between the halves of differential sense circuit. Imbalances may results from the effects of process variations and can be mapped to non uniform variations in threshold voltage V_T gain factors, gain factors, load resistances, load capacitance which can be mapped to gate length variations.

2.1 Voltage Mode Sense Amplifier

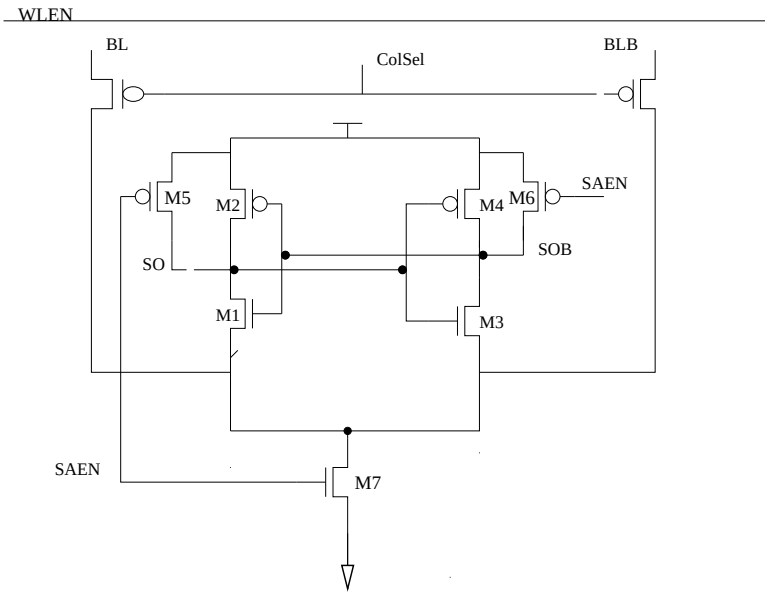


Figure 2.4. Voltage Latch Type Sense Amplifier.

In conventional memories, voltage-mode sense amplifier is used that present a high input impedance to the bit-lines. This allows the sense amplifier to provide a high

voltage gain with the use of simple circuits .The sense amplifier are designed with operating margins that constrain the minimum and maximum input signal amplitudes. The minimum input signal amplitude is necessary to provide reliable operation of the sense amplifier. To meet this constraint, a sufficient differential voltage in the bit-lines is allowed to develop before enabling the sense amplifier. The Voltage mode sense-amplifier also operates in two phases, in the pre-charge phase, the bit-lines and the sense-amp output are pre-charged high. SEN is pulled high for sensing the bit-lines. The voltage mode sense amplifier requires differential discharging of the bit-line capacitance for sensing the voltage difference. It reacts after a certain amount of differential voltage is developed on the bit-line capacitance. Hence the time to develop a certain differential voltage to appear depends on the bit-line capacitance. Hence the time to develop a certain differential voltage will increase with the increase in the capacitance (i.e. number of cells in the column). The time to develop a certain differential voltage (for proper operation of sense-amplifier) will increase because of the increased leakage current and this problem will worsen with the coming technology generation. This section describes two differential voltage sense amplifiers cross-coupled inverter latch Figure 2.4 and 2.5 shows the typical Voltage Latch Type Sense Amplifier. Figure 2.6 shows the timing scheme of VLSA

2.2 Current Mode Sense Amplifier

In advanced memories the capacitances of the bitline is increasing due to technology scaling and the increasing number of cells attached to the column as shown in Figure 2.1. In such memories voltage mode cannot keep up to their performance thereby leading to the need for faster sensing techniques that are not affected by the bitline capacitance. Current mode sense amplifier are applied to reduce the sense circuit delays as they provide low common input/output impedences. The small input impendence presented to the bit-lines result in reduced voltage swings,cross-talk and

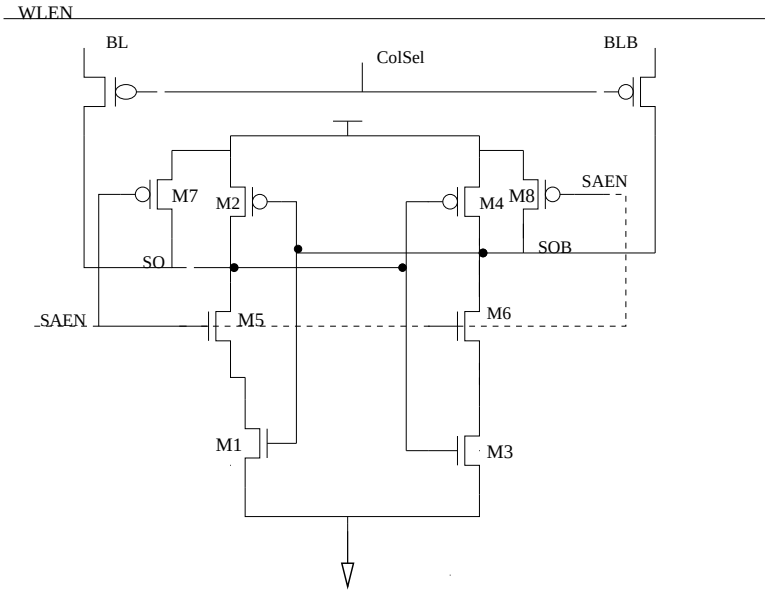


Figure 2.5. Voltage Mode Latch Type Sense Amplifier

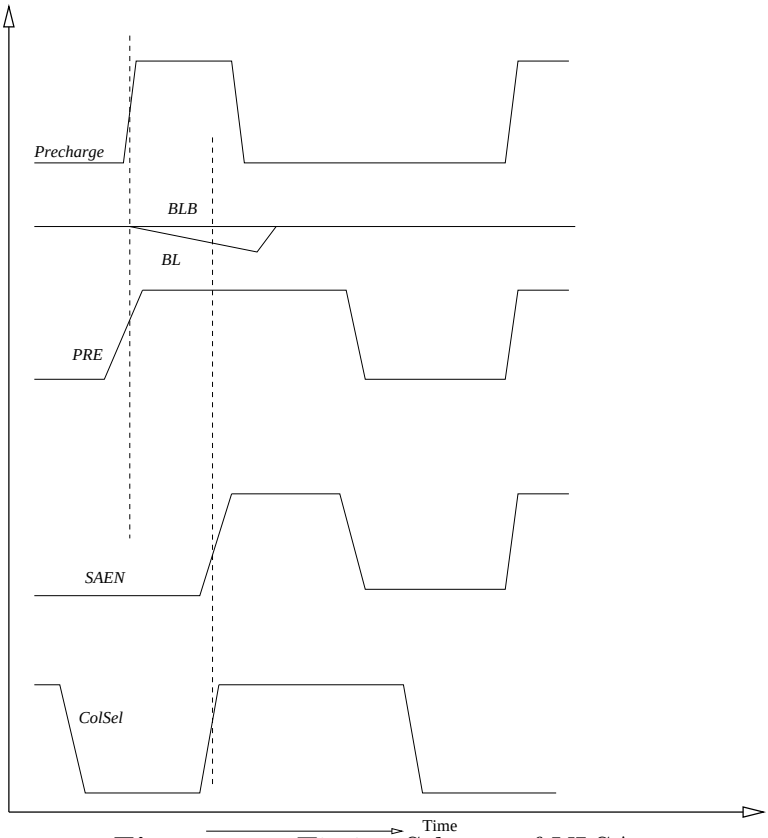


Figure 2.6. Timing Scheme of VLSA.

substrate currents. The current mode sense amplifier converts and amplifies a small current difference into CMOS voltage levels. This section describe different type of current mode sense amplifier and due to its inherent advantage of performance over voltage mode sense amplifier, we will demonstrate the compensation technique on the current mode sense amplifier both on Bulk CMOS technology and FinFET technology.

2.2.1 Current-Mirror Sense Amplifier (Conventional)

The traditional form of the primitive current amplifier is the current mirror sense amplifier. In the current-mirror amplifier if the devices M1 and M2 are identical, then the bitline or input current I1 is the same as the readline or output current Io, because the gate-source voltage V_{GS} is common for both devices M1 and M2. If M1 and M2 differ only in their aspect ratios W/L otherwise they are identical. As shown in the Figure 2.7 [24] In this circuit, the gates of the two nMOS transistor M2 and

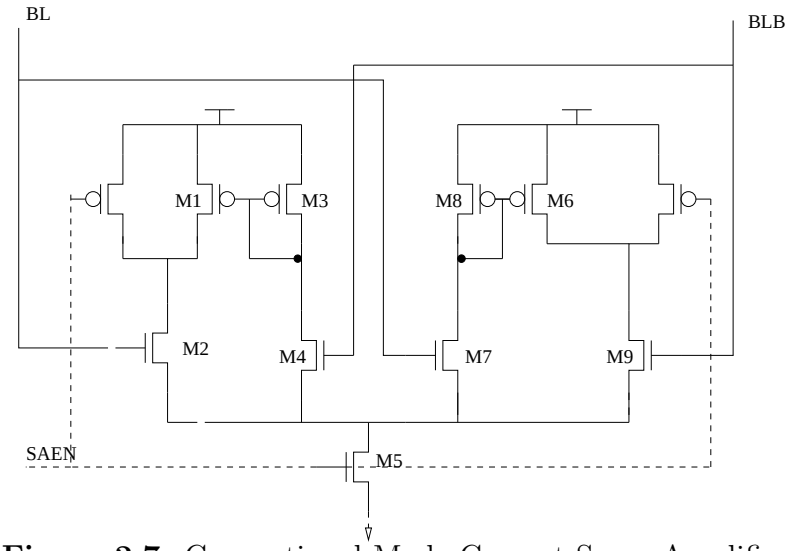


Figure 2.7. Conventional Mode Current Sense Amplifier

M9 are connected to the bit-lines. The substrate terminals are tied to their respective source terminals in order to remove substrate-bias effect. Notice that each bit-line is represented by a large parasitic capacitance. The nMOS transistor M5 is a long

channel device which acts as a current source for both branches, and is controlled by a SAEN signal. Before beginning of a "read" operation, the two bit lines (columns) are pulled up for equalization. The SAEN signal is low during this phase, so that the nMOS transistor M5 remains off. Since both M2 and M9 conduct source node is pulled up, and the output node of the amplifier also goes high. Therefore the output of the inverter is at a logic low level initially. Once a memory cell is selected for the "read" operation, the voltage on one of the complementary bit line will start to drop slightly. At the same time as the row selection signal, the SAEN signal driving M5 is also turned on.if the stored data on the selected SRAM cell forces the bit line to decrease slightly, transistor M5 turns off, and the output voltage of the differential amplifier drop immediately.

2.2.2 Current Controlled Latch Sense Amplifier

Many sense amplifiers are activated simultaneously to achieve wide bandwidth, the current of sense amplifiers increases. Therefore power reduction in sense amplifiers is important to reduce the total power consumed in the memory. A current mirror sense amplifier is shown in Figure 2.7. It is easy to control and the speed of a current-mirror sense amplifier can be easily accelerated by increasing the operating current. Therefore memories frequently use this type of sense amplifier However, the static current flows through the pull down MOSFET connected to ground as long as SAEN is activated. To realize low-power and automatic power-saving scheme Kobayashi et all [23] came up with current controlled latch sense-amplifier shown in Figure 2.8.

In the read cycle , the data of memory cell appear as a small difference on the bit lines (BL and BLBAR) The gate of two NMOS's (M5 and M6) are connected to bit-line and bit-line bar. The current flow through M1 and M3 controls the serially connected latch circuit. A small difference between the current through M5 and M6 converts to a large output voltage, shown in Figure refLatch3.

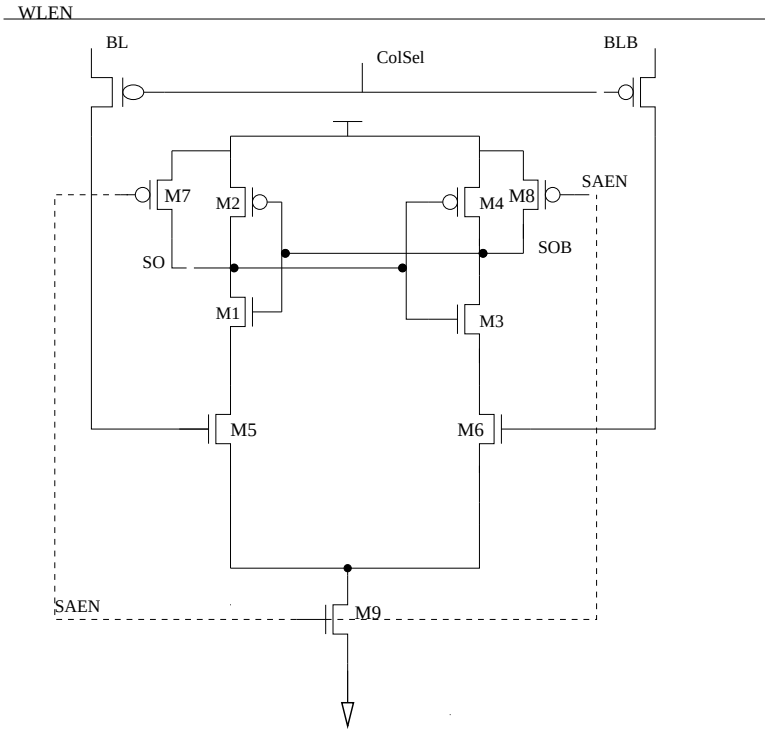


Figure 2.8. Voltage Mode Current Latch Type Sense Amplifier.

Operation Latch Type Sense Amplifier (LSA) starts by turning on M9. After LSA is activated, the operation current (I_{sa}) flows during the transition of output nodes. The current flow only during switching of inverters that compose the latch sense amplifier. The operating of LSA is shown in figures. It is difficult to control the latch timing, because of early latching often causes errors. A latch sense amplifier must start the sensing operation after the bit line are separated enough to sense, can be understood more clearly with the Figure 2.9 [52].

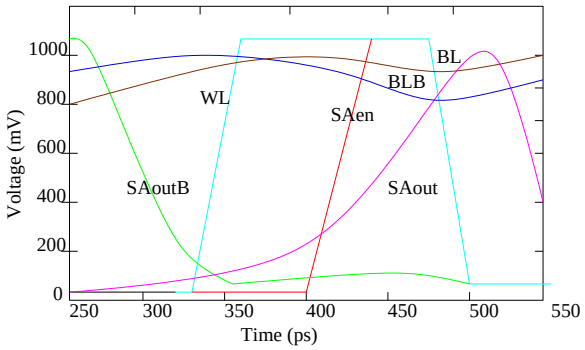


Figure 2.9. Working of Voltage Mode Current Latch Type Sense Amplifier.

2.2.3 Clamped Bitline Sense Amplifier

Another type of current mode sense amplifier is Clamped Bitline Sense Amplifier(CBLSA),It is another current mode type of sense amplifier, which amplifies a small differential current between the bit-lines into a full rail-to-rail voltage. It requires an extra timing signal (equalize signal) for its operation. This presents a two fold problem: one is the routing overhead required for the equalize signal and the other is the generation of timing pulse for sensing every data bit. The CBLSA circuit has no provision for stopping the flow of static current, and hence dissipates power, even when there is no data activity on the interconnect. The CBLSA has an equalize transistor that keeps the voltage at outputs of the cross-coupled inverter at half-V_dd. This presents a problem to the succeeding buffer, as the input voltage at half-V_dd will lead to continuous static power dissipation during the equalize period. Shown in Figure 2.10

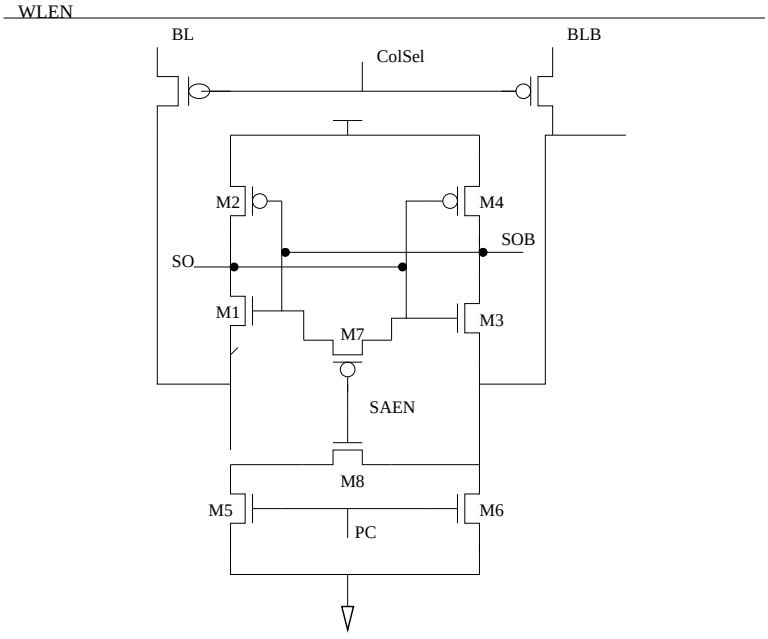


Figure 2.10. Schematic of Clamped Bitline Latch Type Sense Amplifier.

The bit-lines in CBLSA are connected to a low impedance node, which is away from the sense amplifier output. The transistors M1, M2 and M3, M4 form the cross-

coupled inverters. Transistors M5 and M6 provide the low impedance termination. During the precharge phase transistor M8 equalizes the output node of the sense amplifier and the output nodes are precharged to Vdd. Transistors M5 and M6 are off during that time thereby preventing any static currents . During the read operation, transistors M5 and M6 are on and once a sufficient differential current is developed the SAEN signal is pulled low to enable the sense amplifier. The cross coupled inverter pair is enabled which converting and amplifying the difference in the currents in the two legs of the sense amplifier circuits.

The working on CBLSA is shown in Figure 2.11

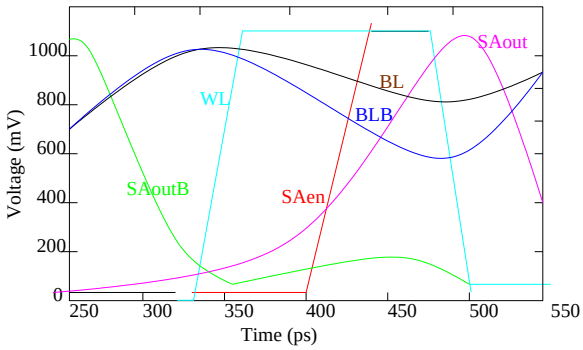


Figure 2.11. Working of Clamped Bitline Latch Type Sense Amplifier.

CHAPTER 3

PROCESS VARIATIONS

Parameter variations have become increasingly important for microprocessor design.It occur for multiple reasons, including non-uniform ion implantation or photo resist exposure, or lack of uniformity in oxidation, diffusion or polishing. Variation between the production lines can occur because of different mask, stepper and the optics at each line. There can be variation between each exposed field on a wafer because of slight changes in the distance between wafers and the optics due to steppers or the wafer not being perfectly flat. Varying illumination and lens aberrations can lead to a large intra-field variation of effective channel length and speed. The process variation can cause the delays of wires and gates within a chip to vary.As a result, some chip may also operate correctly at slower speeds. In order to understand that we need to classify the type of variation and components of variations. The process variations in semiconductor devices can be classified into environmental variations and physical variations.

Environmental variation includes variation in power supply voltage,noise coupling among nets and temperature.Changes in supply voltage accounts for IR drop and Package Noise.Physical Variation are caused by processing and mask imperfections and reliability-related degradation.

As geometries shrink, the impact of process variations on interconnect reliability is expected to increase. Varying etch rates can modify the profile of wire cross-section. At the same time, the thickness of the metallic and dielectric layers can change over the die surface. As the degree of control of processing methods and

techniques are not able to keep up with the degree of scaling in feature size, the manufacturing methods introduce variations. These variations are permanent and result from limitation in the fabrication process. As technology scales, the features size reduces thereby requiring a sophisticated fabrication process. All of these variations effects speed, power (primarily leakage) and yield.

3.1 Trends in Variability

The semiconductor process cannot be perfectly controlled, which leads to statistical variation of many process variables. Several types of process variations can occur: line to line, batch to batch, wafer-to-wafer, die-to-die and intra die. There can be variation between each exposed field on a wafer, because of slight changes in the distance between wafers and the optics, due to steppers or the wafer not being perfectly flat. Varying illumination and lens aberrations can lead to a large intra-field variation of effective channel length and speed.

Inter die variation is the difference in the value of a parameter across nominally identical die (whether those die are fabricated on same wafer, on different wafers or in different lots) and is typically accounted for in circuit design as a shift in the mean of some parameter values (e.g. V_T or wire width) equally across all devices or structures on any one chip. For purposes of circuit design it is usually assumed that each contribution or component in the inter die variation is due to different physical and independent sources, and it is usually sufficient to lump these contribution together into single effective die-to-die variation component with a single mean and variance and are modeled using worst case corners, substantially all of existing practical and theoretical work on yield analysis and maximization techniques are focused on intra-die variation. Intra die variation or within die variation occurs spatially within a die, which are random and due to the semiconductor manufacturing process. These variations are classified into three categories which include devices variation, interconnect

variation and dynamic variations. Device variations are fluctuation in MOS parameters and include effective gate length (L_{eff}), Threshold voltage (V_T) Thickness of the gate oxide (T_{OX}) , and the drain/source region parasitic resistance (R_{dsw}), effective gate length variation could arise due to masking difference and threshold variations occur due to dopant variations.

3.2 Source of Local Variation

The source affecting device and interconnect variation on a die are: Random Dopant fluctuations, Sub-Wavelength Lithography, Increased use of Chemical Mechanical Polishing Gate oxide thickness is critical but generally well controlled parameter variation tends to occur from one wafer to another.

As we scale the transistor further,physics based random variation in the threshold voltage due to random dopant fluctuations, and in the channel length due to line edge roughness, add random components of variability to the systematic layout silicon just mentioned. These variations play a significant part in sub threshold leakage and other important device performance metrics. This uncertainty is expected to continue as we scale our silicon devices to the level of atomic scaling, with oxides a few atoms thick and channels with countable numbers of dopant atoms. The rise in the inherent systematic and random non uniformity will have effects that are far-reaching in every aspect of design, manufacturing, test and overall reliability *"The law of large number no longer applies"* . It's obvious that no single solution or advancement can solve this problem of process variation. We have to come up with ways with new devices and process variation tolerant technique both at the circuit level and architecture to solve our problem of process variations V_T is adjusted to desired level by implanting dopants (threshold implant) in the channel.

Every technology generated, transistor area reduces by half and number of dopant atoms in channel decreases exponentially.

Dopant fluctuation results in: two transistors next to each other will have different V_T from one another due to random placement and concentration fluctuation of the dopant atoms.

Sub wavelength photolithography is the primary reason for line edge roughness and causes of variation in length and width. Lithography variation is one reason why poly is oriented in one direction and dummy cells are used in SRAM arrays.

For short channel MOSFET, variation in L_{eff} can cause further variation in V_T due to drain source charge sharing, halo implants are used these days to minimize this effect. As circuit designers we do not want parametric yield to have large impact on the final yield. The dilemma which we face today: Process and device variation data/model is not available until sufficient silicon has been processed most of the circuit blocks are designed before this data is available, so we need to come up with process variation tolerant techniques both at the architecture level and circuit level, there has been lot of efforts in the past [7, 48]. For dealing with these variations, numerous process technology, circuit and architectural solution have been proposed to combat variation. Designer verifies the circuit functionality and performance under the extreme conditions.

3.3 Traditional Approach: Worst Case Design

Traditionally, 5 different bins are defined by two letter acronyms describing the relative performance characteristics (T=typical, F= fast, S= slow) of nmos and pmos devices. There are differential spice models for fast, slow and typical nmos and pmos devices and they are generated in accordance with the maximum and minimum values of the saturation currents and threshold voltage of the transistors found on a sample chip. Using these models, the circuit/chip is simulated on spice.

MOS transistor dimensions and the threshold voltage are among the key parameters that control CMOS transistor's drive current. MOS circuits become extremely

sensitive to length and threshold voltage as we tend to see minimum length transistors and reduced supply voltage. The variation in substrate doping concentration and gate oxide thickness can be accounted for variation in the threshold voltage. Charge centers near the oxide-substrate interface, fixed oxide charge and surface state densities - are likely to be dominant mismatch source for carrier mobility. Random variations in deposition, etching, annealing temperature and chemical mechanical planarization etc contribute to channel length variation. Dependence of transistor current is increasingly non-linear to channel length. A 10 % transistor gate length variation can translate to as much as variation of -15 % to + 25% in gate delay [30, 8]. Therefore parameter variations introduce asymmetry in the sense amplifiers which in turn reduces yield and performance.

3.4 Previous Work to Combat Process Variations

There has been commendable and different techniques proposed in the past to compensate for process variations.

Some of the work is even now implemented on sub nanometer technology node. Variation Tolerant Design can come under two category, first would be reducing SOURCE of variation and second should be reducing EFFECTS of variation at the time of design and reducing effects of variation in post silicon DRAM designers have proposed several offset compensation schemes [2]. Circuits which rely heavily of symmetry of the devices like latches,flipflops, SRAMs, Ring oscillators requires special attention and there has some recent proposed technique "A process variation tolerant technique for sub-70nm latches and flip-flops" [14].

3.4.1 Supply Voltage

Changes in supply voltage cause transistor subthreshold leakage variation across the die, which results in uneven voltage distribution and temperature hotspots. Adap-

tive Vdd has been shown to be useful to reduce impact of parameter variations. Even though it helps increasing yield in high frequency bins, adaptive Vdd does not solve the problem of having voltage droops ΔVdd. One well-known technique is using on-die decoupling capacitors. If an appropriate number of decoupling capacitors are placed on die,ΔVdd can be reduced by 50However, decoupling capacitors have a cost in silicon area. Moreover, gate oxide leakage is a problem in sub-90nm technologies, and the layout of decoupling capacitors tends to increase the gate oxide area.

3.4.2 Temperature

Increasing temperatures affect processor design in many different ways. For instance, the cooling system of a processor is targeted to support a peak temperature, even though the processor spends most of the time running at much lower temperatures. In order to reduce dynamic power dissipation, chip designers have relied on scaling down the supply voltage. To counteract the negative effect of a lower supply voltage on gate delay, the threshold voltage is also scaled down along with the supply voltage. However, lowering threshold voltage has a significant impact on leakage current. Moreover, it also impacts temperature, due to the exponential relationship between leakage current and temperature. The most common technique to control temperature, already implemented in several commercial processors, is throttling. When the temperature goes beyond the maximum limit, the operating frequency is decreased. This is followed by a decrease in Vcc, which helps reducing power consumption and temperature. Once the processor cools down, the process is reversed. However, chip temperature has a very high timing constant and thus, these thermal emergencies have a significant impact on performance.

In case of sense amplifier there has been some work done and proposed in the recent past. These techniques try to compensate the mismatch by either precharging the bit-lines or the two output to different initial values. However in Voltage Mode

Sense Amplifiers for SRAM, bit-line and the outputs are pre charged and equalized to V_{DD} [31, 52].These techniques can not be directly applied to voltage mode sense Amplifier in SRAM,

Kaushik Roy and Hamid Mahmoodi [28], proposed "Leakage current based stabilization scheme for Robust Sense amplifier design for yield enhancement in Nano-Scale SRAM". In this work they developed a method to analyze the probability of access failure in SRAM array (due to random V_T variation in transistors) by jointly considering variations in cell and sense amplifiers. By introducing two PMOS transistors, PL-R and PR-L (collectively called the PMOS stabilizers), between the inputs) i.e. the bit-lines BL and BLB) and the drain of the driver transistors N1 and N2.

Kaushik Roy et all [29], have also proposed robust sense amplifier using independent gate in symmetric and asymmetric double gate devices for sub-50-nm technologies. The double gate devices can be design in different structures, namely

1) symmetric devices with same gate material (e.g near-midgap metals) and oxide thickness for the front and back gate (SymDG) [19] asymmetric device with different front and back oxide thickness (AsymOxDG) [51] and 3) asymmetric devices with material of different work function (e.g, n+ poly and p+ poly) in the front and the back gate (AsymWfDG) [22].

In this work they have design symmetric and asymmetric devices (both AsymOxDG and AsymWfDG) with 50-nm gate length in the 2D device simulator MEDICI [46].

Here they have shown directly translated current latch type sense amplifier and directly translated voltage latch sense amplifier with the independent gate technique they are proposing robust sense amplifier but this does not prove helpful incase the double gate MOSFET has process variations the threshold voltage is insensitive to random Dopant fluctuations incase of FinFETs since the FinFETs have undoped channel but the absolute value of V_T increases significantly with variations body

thickness and this is a consequence of two effects. First as the silicon body gets thinner, the two gates get closer and have better control over the short channel, reducing short channel effect and V_T roll-off. So simply translating sense amplifier from symmetric and asymmetric double gate devices would definitely an advantage in terms of performance, short channel effects and process variation, but still it will require some added circuit styles or logic to compensate further on process variations due to parameter variations. As FinFET do have process variation due to parameteric variations.

Joyce Yeung and Hamid Mahmoodi proposed "Robust Sense Amplifier under Random Dopant Fluctuation in CMOS Technologies" they have considered three design options They are (A) Transistor sizing, (B) Double V_T assignment and (C) Multiple finger layout structure.

As transistor sizing can directly determine the threshold variation due to random dopant effect. It also affects the trip point mismatch and current mismatch because both of them as functions of V_T variation. Upsizing the NMOS of the inverter lower its trip voltage, which makes the cross-coupled inverters flip at a lower output voltage when there is more output differential developed. This improves the probability of correct operation.

Double V_T assignment is a technology that provides optionally low or high threshold voltage for each transistor. It was originally developed for leakage power reduction. The impact of high or low V_T on V_T variation is negligible, however similar to sizing, V_T affects the amount of current that flows through a transistor.

Multiple finger structure is also commonly used technique in layout design to implement large transistors because it provides better aspect ratio and less area for layout of large transistors. Since V_T variation is inversely proportional to the size of a transistor, overall effect of variability of multiple-fingered transistor was found to be canceled out due to the parallel structure in multiple finger layouts (the finger that

shifts to high V_T cancels the effect of the finger that shifts to low V_T and vice-versa) due to the property of randomness in dopant fluctuation.

It is expected that multiple finger structure does not impact the robustness of the sense amplifiers. But, all of the above techniques are not a solution to a problem. We need to come up with ways that improves the robustness i.e. yield of sense amplifiers and for the fixed yield the scheme should result in faster sensing.

3.5 Impact of Mismatch on Sense Amplifiers

MOS transistor dimensions and threshold voltage are among the key parameters that control CMOS transistor's drive current. MOS circuits become extremely sensitive to length and threshold voltage as we tend to use minimum length transistors and reduced supply voltage. The variation in substrate doping concentration and gate oxide thickness can be accounted for variation in threshold voltage. Charge centers near the oxide-substrate, fixed oxide charge and surface state densities - are likely to be the dominant mismatch sources for carrier mobility. Mentioned above parameter variations introduce asymmetry in the sense amplifiers, which in turn reduces yield and performance. In case of process variation V_{OS} can be computed by partial sum of derivatives of mismatches in V_T, Width and Length as in equation 3.1

$$V_{OS} = \sum (\frac{\delta f}{\delta V_{t,i}} \Delta V_{t,i} + \frac{\delta f}{\delta W_i} \delta W_i \Delta W_i + \frac{\delta f}{\delta L_i} \Delta L_i) \tag{3.1}$$

Transistor mismatch effect vary considerably between different regions(weak inversion, linear, saturation etc.) of operation. Various mismatch models are studied in the past and the drain current model is found to be less complex and fairly accurate for mismatch analysis. The current mismatch model for linear and saturation regions of operation are given by equations 3.2 and 3.3 below [6].

Linear Region:

$$\frac{\Delta I_D}{I_D} \cong \frac{\Delta\beta}{\beta} - \frac{\Delta V_T(V_{SB})}{V_{GS} - V_T(V_{SB}) - (1/2)V_{DS}} - \frac{\Delta\theta_{eff}(V_{GS} - V_T(V_{SB}))}{1 + \theta_e ff(V_{GS} - V_T(V_{SB}))} + \frac{\Delta V_{DS}}{V_{DS}} + \frac{\Delta V_{GS}}{V_{GS} - V_T(V_{SB}) - (1/2)V_{DS}} \tag{3.2}$$

Saturation Region :

$$\frac{\Delta I_D}{I_D} \cong \frac{\Delta\beta}{\beta} - \frac{2\Delta V_T(V_{SB})}{V_G S - V_T(V_{SB})} - \frac{\Delta\theta_{eff}(V_{GS} - V_T(V_{SB}))}{1 + \theta_{eff}(V_{GS} - V_T(V_{SB}))} + \frac{2\Delta V_{GS}}{V_{GS} - V_T(V_{SB})} + \frac{\lambda\Delta V_{DS}}{1 + \lambda V_{DS}} \tag{3.3}$$

Where θ_{eff} is empirical effective mobility reduction factor and is function of β and source and drain resistances. Other parameters in equation 3.2 and 3.3 have usual meanings. From the above equations it can be seen that the variation in drain current is strong function of variations of β, V_T, θ_{eff} and drain to source voltages.

In the present work we have considered the effect of variation in V_T and L_{eff} on the performance of the different latch type of sense amplifier due to their extensive use in memory applications. Figure 2.4 and Figure 2.8 shows latch type voltage sense amplifier and latch type current sense amplifier. As mentioned earlier latch type implementation has benefit of positive feedback which results in faster and stable sensing, thus used extensively in memories. There is also a low point of that due to positive feedback and high gain these sense amplifiers quickly produce the output, however that also makes them very sensitive to parameter variations. Figure 2.10 shows the schematic Clamped Bitline Sense Amplifier (CBLSA).Its low input impedance makes it suitable for large memories, we will comparing all the three in terms of performance, yield and transistor mismatch effect on yield and performance.

All simulations were run on 90 nm cadence generic pdf design kit available to us as a part of university program. [9]

All the Sense amplifier delay variation comparison with bit-line differential voltage is shown in Figure 3.1 for the four types of sense amplifiers As bit-line differential voltage plays a crucial role in terms of sensing delay. Ideally all the sense amplifier should start sensing or should have minimum sense delay with bit-line differential voltage. From the Figure 3.1 we see that the sensing delay decreases with increase in

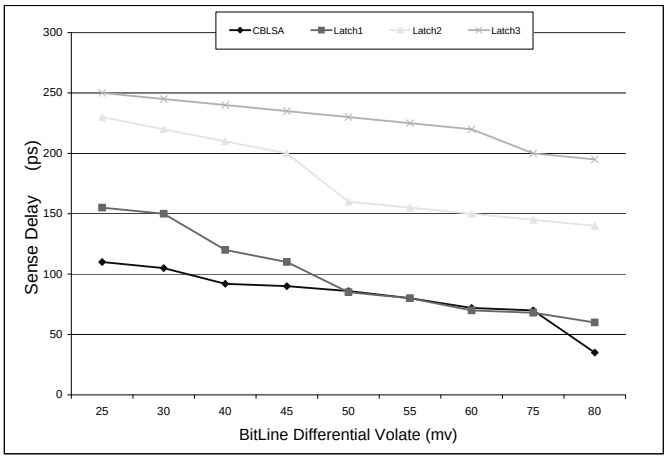


Figure 3.1. Sense Delay versus Differential Voltage for different type of Sense Amplifier

differential voltage. CBLSA is found to be fastest among the fours type of sense amplifier. However rate of decrease of delay and increase in differential voltage is highest in voltage sense amplifier and lowest in current mode latch type sense amplifier.

3.5.1 Experimental Setup

The sense amplifier is said to be perfectly balanced or symmetric when all the transistor parameters on the left-hand side are equal to parameters on the right-hand side. It is said to be vertically matched when the W/L's of NMOS and PMOS have been ratioed so that the sense amplifier has equal pull-up and pull-down capabil-

ity. For mismatch analysis we consider only the variation in the channel length and threshold voltage of the transistors. Our analysis assumes that a particular parameter x(where x is either L,W or V_T of the device) is varied by decreasing the values of that parameters for the NMOS transistors in the left hand side of the sense amplifiers by Δx and increasing it by the same amount for corresponding right hand side transistors.

We assume perfect matching in PMOS transistors and vice-versa. In all sense amplifiers total width of all the transistors is kept nearly same for fair comparison. Simulations are carried out for 90nm CMOS technology from cadence generic pdk available to us [9]and 32nm using CMOS technology, using Predictive Technology Model [49] and considering nominal values of NMOS and PMOS transistor threshold voltages at 0.26B (V_{Tn}) and -0.303V (V_{Tp}) respectively. A nominal output load capacitance of 5fF is assumed at sense amplifier output nodes. The circuits are tested to read both logic low and logic high from consecutive memory locations. 200fF bit line capacitance is assumed for all the simulations. Power supply voltage is assumed to be at 1.1 V. The timing scheme used in all of the sense amplifier is shown in Figure 3.2.

The wordline enable (WLEN) and column select (ColSel) are activated for a short duration the sensing operation. Both the signals are deactivated after sufficient differential voltage/current is developed between the bit-lines. This has the effect of isolating bitline capacitances from the sense amplifier. This kind of timing scheme helps in reducing signal swing on bitlines, thereby reducing power dissipation.

3.5.2 Results

Extensive simulations are carried out to check the effect of mismatch in V_{Tn} and channel length. For all mismatch analysis time interval between WLEN and sense amplifier enable (SAEN) is kept constant at 110ps in all the four circuits under

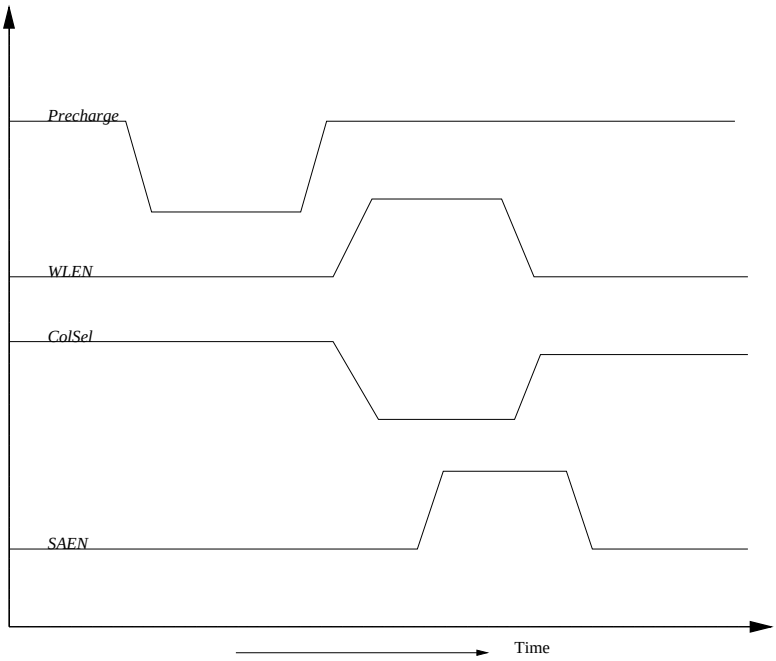


Figure 3.2. Sense Amplifier timing scheme

consideration. The sense amplifier delay was measured as the time taken for the sense amplifier's output to reach 90% of the full-rail voltage after the SAEN signal goes high.

a) Mismatch in NMOS Parameters:

For the circuit shown in Figure 2.4, V_{Tn} of M1 is decreased and that of M3 is increased by same amount. For circuits shown in Figure 2.5, 2.8 and 2.10, the V_{Tn} of M1 and M5 is decreased while that of M3 and M6 is increased. Similar analysis is carried out to study channel length mismatch in NMOS transistors. For all the simulations we assume that the SRAM cell content logic is high, since this gives the worst case results for the above setup.

The results of delay versus NMOS threshold voltage mismatch are shown in Figure 3.3

It is found that, Current Latch-Type Voltage Mode Sense Amplifier is extremely sensitive to V_T mismatch.

From Figure 3.4 we find that sensing delay increase with the mismatch.

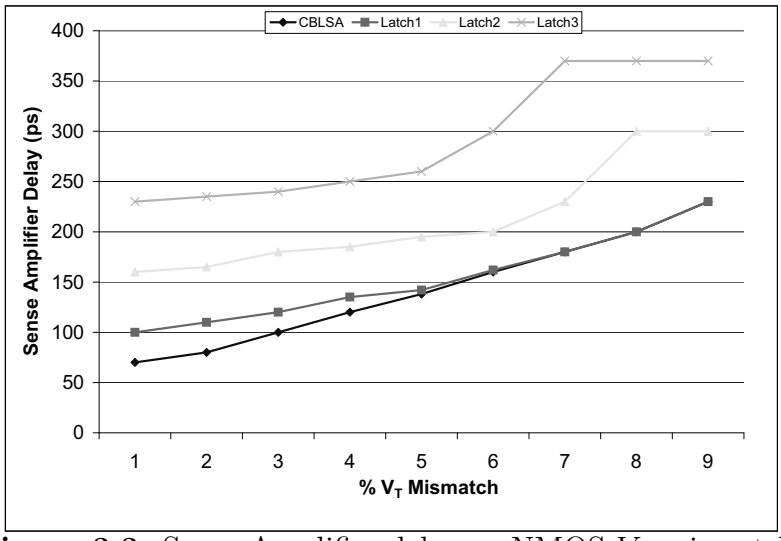


Figure 3.3. Sense Amplifier delay vs NMOS V_T mismatch

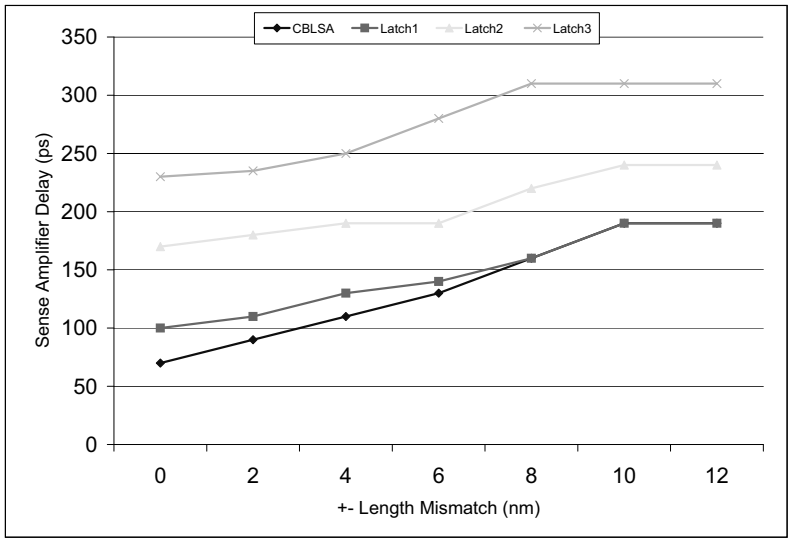


Figure 3.4. Sense Amplifier delay vs NMOS channel length mismatch

b) Mismatch in PMOS Parameters:

Parameter mismatch of PMOS transistors does not drastically affect performance of the sense amplifier. The reason being, when the sense amplifier is activated NMOS transistors immediately start operating in saturation region, however incase of PMOS transistors operate either in cut-off or conduct in deep triode region. Mismatch analysis is carried out by decreasing V_{Tp} of M2 and increasing V_{Tp} of M4 by the same amount. Similar variations in channel length are imposed to understand the effect of length mismatch. simulation results for threshold voltage and length mismatch are shown in Figure 3.5 and 3.6

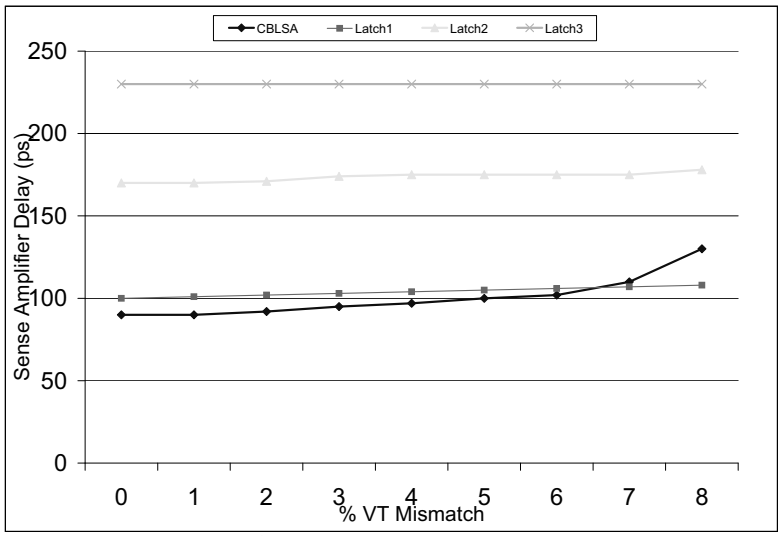


Figure 3.5. Sense Amplifier delay vs PMOS V_T mismatch

Simulation results shows that except CBLSA, all latch sense amplifier are less sensitive to PMOS parameters.

c) Yield Analysis

For successful read operation, the sense amplifier should be activated only after development of sufficiently large bit-line differential voltage or current. However to decrease the overall sensing delay, it is desirable to activate the sense amplifier as early as possible. Sense amplifier may fail to read correctly due to its asymmetry, which

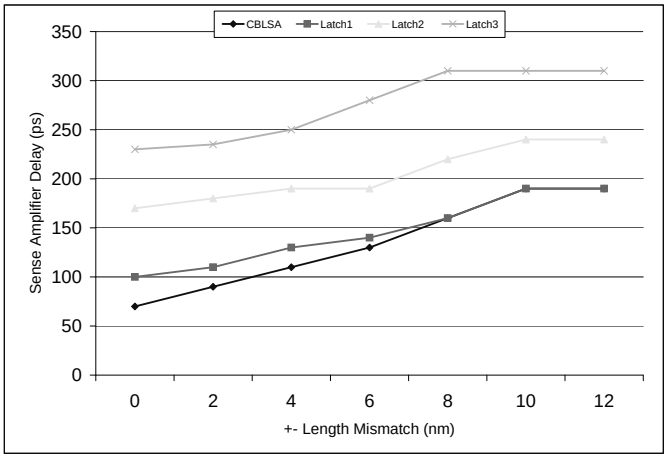


Figure 3.6. Sense Amplifier delay vs PMOS channel length mismatch

is a result of manufacturing process variations. Yield analysis is carried out using Monte-Carlo simulations on HSPICE [45]. Threshold voltage and channel length of all the NMOS transistors (M1 and M3 in Latch Type1, M1, M3,M5 and M6 in the remaining circuits) is randomly varied simultaneously. All Monte-Carlo simulations incorporate 3σ variation where σ is standard deviation of the parameter. The yield analysis results are shown in Figure 3.7

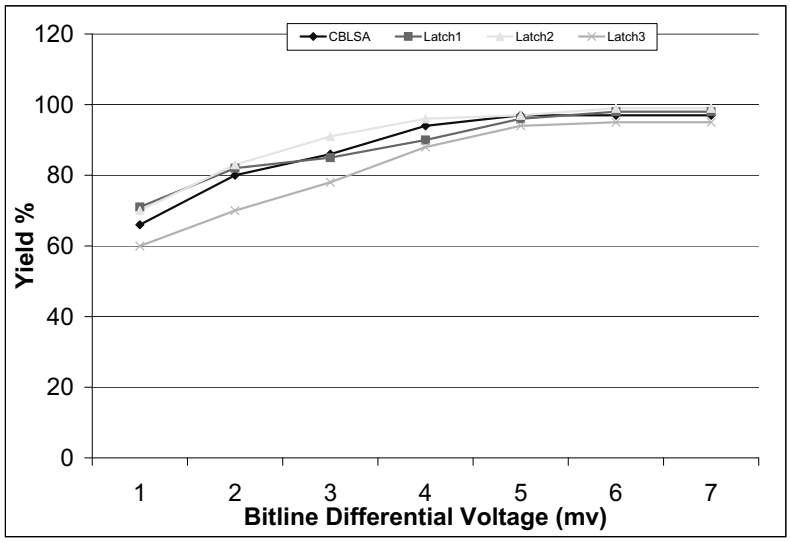


Figure 3.7. Yield vs bit-line differential voltage

Yield measurement for a sample size of 10000.Yield is calculated using the following simple equation 3.4

$$Yield = \frac{number of correct decisions}{number of samples = 10000}.100\% \tag{3.4}$$

Hence, Impact of process variations induced transistor mismatch on different sense amplifiers configurations is presented. It was found that as the mismatch increases sensing delay also increases, eventually leading to failure of the sense amplifier. Performance of the sense amplifier is worst affected by mismatch in NMOS transistor threshold voltage and channel length whereas performance degradation due to PMOS parameter variations is found to be marginal.

3.5.3 Characterization of mismatch in MOS transistor

The characterization of mismatch in MOS transistor is complex and require further understandng.The drain current matching not only depends on the device dimensions but also on the operating point. So in the section the focus is on developing a characterization methodology which determines the mismatch in drain current over operating region. The simulation of the effects of mismatch on an arbitrary analog circuit requires a statistical model which can predict drain current mismatch at a wide range of gate,drain and bulk bias points. It is the goal of this work to determine if the current mismatch of a pair of transistors at any bias can be represented solely by mismatches in four standard MOS parameters : V_{T0}, γ,β and θ.

For a MOSFET the current voltage relation in linear or triode region is given by 3.5

$$I_D = \frac{\beta}{1 + \theta(V_{GS} - V_T)}(V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2 \tag{3.5}$$

where I_D is the drain current, β is the conductance constant, V_T is threshold voltage and V_{DS} is the drain to source voltage. The statistically significant parameters of this model are V_T and β conductance. The mismatch in V_T accounts for the variations in the different charge quantities and on the gate oxide capacitance per unit area are measured as mismatch in β. As both V_T and β are dependent on the gate oxide capacitance per unit area and we need to measure the correlation between the mismatches in V_T and β.

If the MOS transistors are operating in the saturation region in the analog circuits. The measured mismatch in V_T and β to the saturation region, where the drain current is given by **??**

$$I_D = \frac{1}{2}\frac{\beta}{1+\theta(V_{GS}-V_T)}(V_{GS}-V_T)^2 eqn:sat) \tag{3.6}$$

With

$$V_T = V_{T0} + \gamma(\sqrt{V_{SB}+2\phi_F} - \sqrt{2\phi_F}) \tag{3.7}$$

and $2\phi_F$ is the surface potential.

Assuming the mismatch between the transistors in V_T, γ , β and θ are ΔV_{T0},Δ γ,$\Delta\beta$ and $\Delta\theta$ respectively and that gate and drain are mismatched by ΔV_{GS} and ΔV_{DS} respectively.

Expanding 3.5 and 3.9 in taylor series around the nominal bias point and neglecting higher order difference term, the percent drain current mismatch, in the linear region is given by 3.8

$$\frac{\Delta I_D}{I_D} = \frac{-\Delta V_T}{V_{GS}-V_T} + \frac{\Delta V_{GS}}{V_{GS}-V_T} + \frac{\Delta\beta}{\beta} + \frac{\Delta V_{DS}}{V_{DS}} - \frac{(V_{GS}-V_T)\Delta\theta}{1+\theta(V_{GS}-V_T)} \tag{3.8}$$

and in the saturation region

$$\frac{\Delta I_D}{I_D} = \frac{-2\Delta V_T}{V_{GS} - V_T} + \frac{2\Delta V_{GS}}{V_{GS} - V_T} + \frac{\Delta\beta}{\beta} + \frac{\Delta V_{DS}}{V_{DS}} - \frac{(V_{GS} - V_T)\Delta\theta}{1 + \theta(V_{GS} - V_T)} \tag{3.9}$$

If a comparison of drain current mismatch of 3.8 from 3.9. The device in saturation has almost twice the drain mismatches due to V_T mismatch.

If we consider both device n-channel and p-channel both operating in same operation modes then ideally due to difference in fabrication incase of p-channel mosfet, as the threshold adjust implant is very shallow, a considerable portion of the implanted ions in retained in the gate oxide. The presence of these impurity atoms in the oxide may cause a degradation in the capacitance matching of the p-channel devices as compared to n-channel ones. Thus devices which use a compensating threshold adjust implant have a higher mismatch in threshold voltage due to differential doping occurring at the surface. This is the major reason for significantly larger mismatch noticed in the p-channel devices as compared to n-channel transistor operating in same operating mode [35]

Now, incase of sense amplifier the parameter mismatch of PMOS transistor does not drastically affect performance of the sense amplifier. The reason being, when the sense amplifier is activated NMOS transistor immediately start operating in saturation region, however the PMOS transistor either in cut-off or conduct in deep triode region. As shown by equations above he mismatch effect are practically doubled when the device is in saturation mode.

CHAPTER 4

SELF COMPENSATION SENSE AMPLIFIER DESIGN ON BULK CMOS TECHNOLOGY

As concluded in chapter-3 performance of the sense amplifier is worst affected by mismatch in the NMOS transistor threshold voltage and channel length whereas performance degradation due to PMOS parameter variation is found to be marginal.

In this work, our objective is to address the problem of yield loss due to process variation by redesigning the latch type sense amplifier with compensation circuit to compensate for process variations. To illustrate our technique we will use simple current mode latch type sense amplifier shown in Figure 4.1 (simplified for illustration purpose) though this technique is applicable to any of the sense amplifier.

4.1 Operation

Suppose, we apply a small voltage offset at the bit line inputs such that $V_{BLBAR} < V_{BL}$ and then turn on SEN. During the pre-charge phase, the output nodes O1 and O2 are set to V_{DD} by the precharge transistors. The sense enable signal SEN starts the sensing operation by turning on M4, The nodes O1 and O2 are discharged by current I1 and I2 respectively. However, since $V_{BLBAR} < V_{BL}$, I2 > I1 hence O2 discharges faster than O1.

The delay up to this point is relatively small. As the nodes O1 and O2 begin to discharge, O2 will reach (V_{DD}-V_T) sooner than O1. This in turn will turn on PMOS M5 earlier than PMOS M6. Turning on of M5 charges O1 and counters the effect

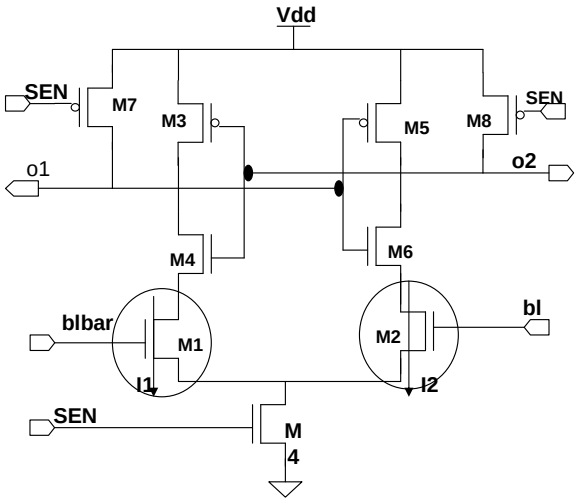


Figure 4.1. Current Latch Voltage Mode Sense Amplifier scheme

of I1. This further increases the voltage difference between O1 and O2. The cross coupled inverter action amplifies the voltage difference. The strong positive feedback enhances the output voltage difference starting from the initial voltage difference V_{BL} - V_{BLBAR}. The total time until the sense amplifier latches the data is termed as sensing delay. Correct operation of sense amplifier is shown in Figure 4.2

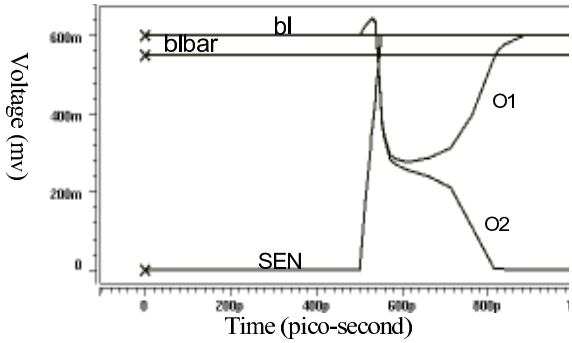


Figure 4.2. Sense Amplifier Correct Operation

4.2 Impact of Process Variations

It can be seen that the offset voltage depends on the mismatch of device parameters, and only for V_{in} larger than this offset the sensing circuit flips in right direction. Applied to certain number of samples this effect is referred to as parametric yield.

As already mentioned initially yield is an important characteristic of sense amplifiers as single failing amplifier implicates whole of memory. We performed statistical investigation by monte carlo simulations on variations of sense transistor parameters like V_T mismatch and L_{eff} variations between the two differential transistors of a nominal LSA. Figure shows the effect on Yield on nominal sense amplifier without mismatch, comparing it with nominal LSA with V_T variation in the transistor. Plot shows three is considerable yield loss with V_T variation on the nominal type sense amplifier. We wish to fix this problem of yield loss due to process variation in present sense amplifier design by our Self Compensation Sense Amplifier design (SSA)

4.3 The Self Compensation Sense Amplifier (SSA) Technique

The proposed sense amplifier design is shown in Figure 4.3. Wrapper script is written for the sense amplifier with minimum transistor size for minimum sense delay and minimum capacitance size and maximum Yield. In the SSA design, M1 and M2 are the primary sense transistors. Like in the previous scenario, let us assume that M1 suffers from a lower V_T than M2. We have added shadow transistors M6 and M5 for M1 and M2. When SEN is turned on, M5 is driven by node voltage at capacitance C2 while M6 is driven voltage at capacitance C1 through transistors M13 and M14 Ordinarily, M5 and M6 are identical transistors. Let us for a moment assume that is the case even under process variation . If the voltage at C2 can be set higher than C1 appropriately, then the effect of V_T mismatch at M1 and M2 can be offset by pumping more current through M6 and less through M5. Include Current I5 and I6 for compensation transistor equation along with I1 and I2.

4.3.1 The Operation of SSA

The proposed sense amplifier has two added phases for training the capacitances. In the first training phase, the capacitors are fully discharged by asserting CAP_DIS.

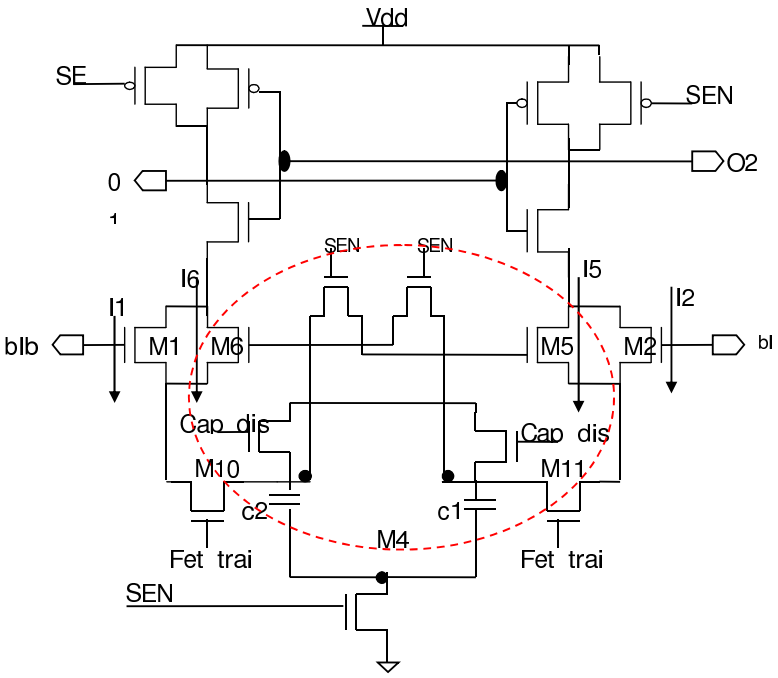


Figure 4.3. Self Compensation Sense Amplifier Technique

Next, CAP_DIS is turned off,and FET_TRAIN is turned on after a pre-charge while $V_{BLBAR} = V_{BL}$. Since M1 conducts more current that charges C2, while M2 has less current that charges C1, C2 will develop a higher voltage than C1. The capacitors C1 and C2 along with FET_TRAIN pulse width is chosen such that voltage across the capacitance ¿ V_T of the shadow transistors M5 and M6. We have a wrapper script in which transistors are sized with minimum sense delay and maximum yield. The capacitances are sized of the value for minimum sense delay and maximum yield. The operation of SSA amplifier is shown in Figure 4.4 with $\triangle$V of 50 mV.

FET training transistor plays a crucial role, in mitigating the mismatch due to process variation between the sense transistors. The training transistor M10 and M11 are responsible for charging the capacitor C2 and C1 according to V_T, W and L variation in M1 and M2. To demonstrate this we ran Self Compensation Amplifier (SSA) with no FET training phase which again causes the incorrect sensing of the sense amplifier due to V_T mismatch in transistor M1 and M2 as shown in Figure 4.5

.

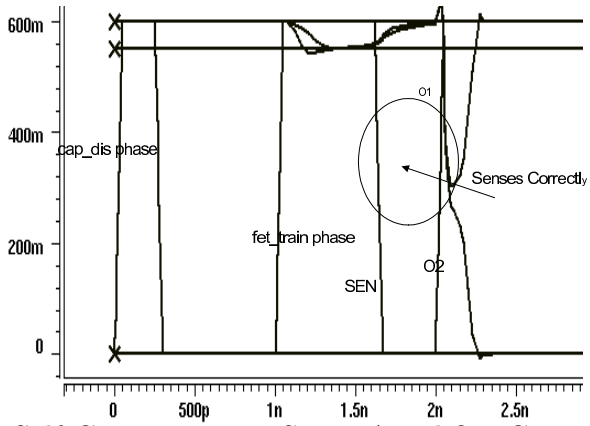


Figure 4.4. Self Compensation Sense Amplifier Correct Operation

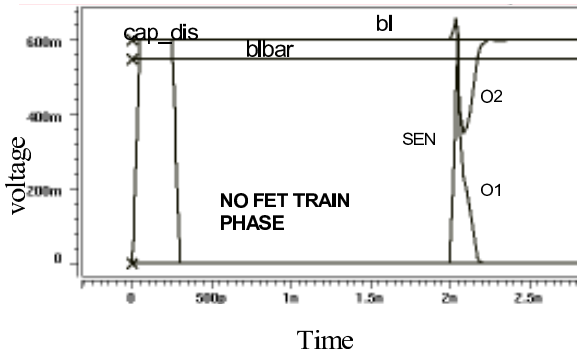


Figure 4.5. Self Compensation Sense Amplifier Without FET_TRAIN Phase

Once we introduce V_T variation between the two differential transistors M1 and M2 and introduce M2 with an high V_T, so as in nominal case due to V_T mismatch variation I2 should be lower than the I1 and result in incorrect sensing but once FET_TRAIN is turned on, C2 enables the compensation due to mismatch through the shadow transistor M5 as shown in Figure above even with V_T variation of 50mV. The sense amplifier performs correct sensing.

It is understood that all the added devices will have normal process variation. Therefore, the results can only be verified by statistical simulation. We ran Monte Carlo simulations and for each of 10000 samples for each voltage difference. The results shows on SSA that number of working sense amplifier without any V_T variation is almost the same with V_T variations on the sense amplifier with the compensation scheme shown for 180nm to 32nn in Figure 4.6 and 4.7

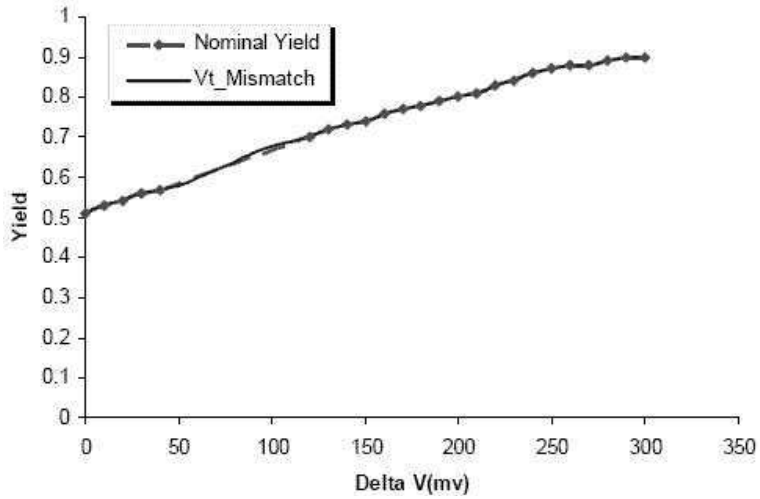


Figure 4.6. Yield vs Bitline Differential Voltage on 180nm

4.3.2 Impact of PVT Variation

As per Sani Nassif [30] in the present technologies it is expected to have 40% variation in the effective channel length From the road map there is prediction of 12% variations in the supply voltage which cause transistor subthreshold leakage variation across the die and results in uneven voltage distribution also causing temperature hotspots. There is a need to understand change in supply voltage and statistical simulation across 10000 samples of sense amplifier. With 12% variation in supply voltage there is considerable yield loss as significant difference in yield in nominal latch type sense amplifier compared to self compensation type sense amplifier shown in Figure 4.8

Increasing temperature affect processor design in many different ways.In order to reduce dynamic power dissipation chip designers have relied on scaling down the supply voltage.To counteract the negative effect of a lower supply voltage on gate delay, the threshold voltage is also scaled down along with the supply voltage. Lower threshold voltage impacts temperature. On normal instances temperature varies 25^oC to 125^oC, though in nominal latch type sense amplifier and self compensation sense amplifier. It doesn't impacts the yield significantly as shown in Figure 4.9

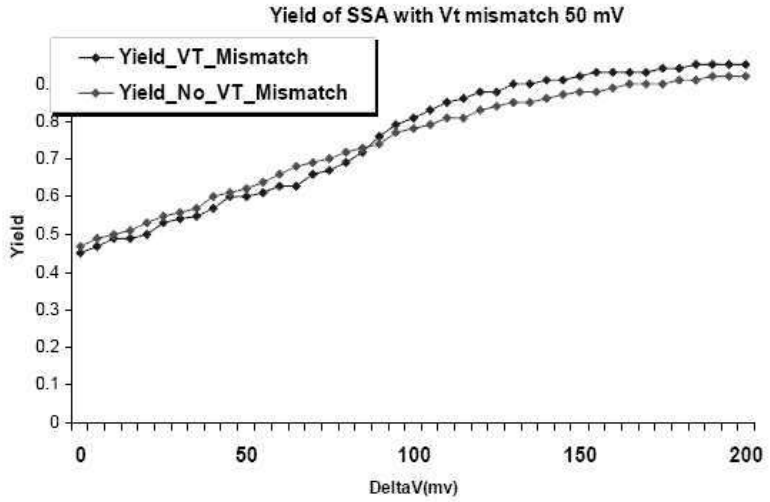


Figure 4.7. Yield vs Bitline Differential Voltage on 32nm

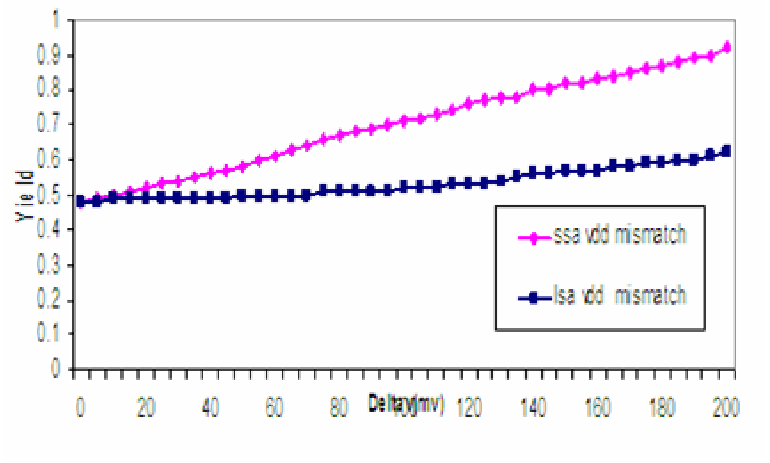


Figure 4.8. Yield vs Bitline Differential Voltage on 32nm with Vdd variation

4.3.3 Quantifying Area increase with Power and Delay

Self compensation technique come with the area overhead of eight more transistor from nominal latch type sense amplifier. The area of nominal self compensation sense amplifier on TSMC18rf at $47um^2$ has an average power as 6.35mW and Sense delay from SEN is 20ps.

Self Compensation Sense Amplifier, has an area increase of $78.4um^2$, average power with almost double area overhead, power is 16.9mW and sense delay from cap_dis signal is 112ps.

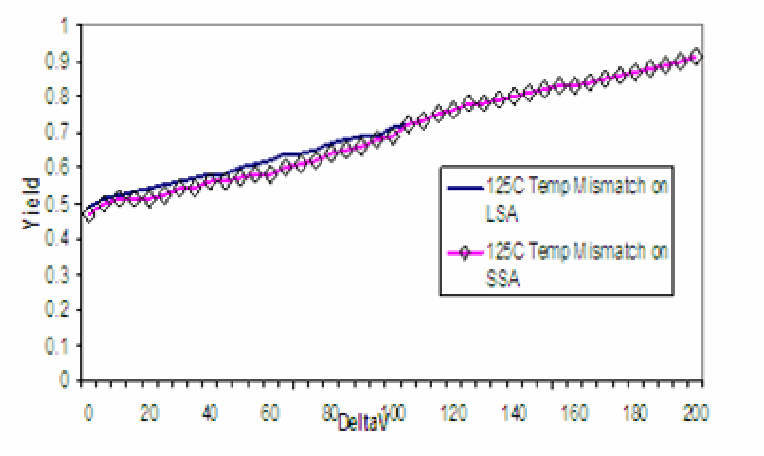


Figure 4.9. Yield vs Bitline Differential Voltage on 32nm with Temp variation of 125^oC

SRAM of the size 128X8 with Area $51345um^2$ would have one sense amplifier for each sram column, so ideally area increase shouldn't be that much. SRAM with nominal LSA has area of $51721um^2$ and SRAM with SSA has area of $51972um^2$. % increase in area is only 0.48%. So to deal with process variation which ensures robustness and reliability in case of sram and sense amplifiers increase in area with this much percentage should be tolerable.

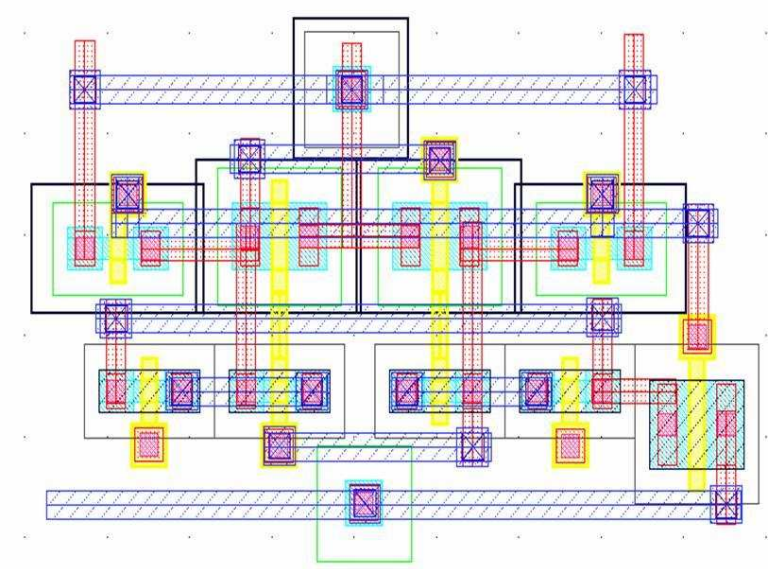


Figure 4.10. Layout of nominal latch type sense amplifier

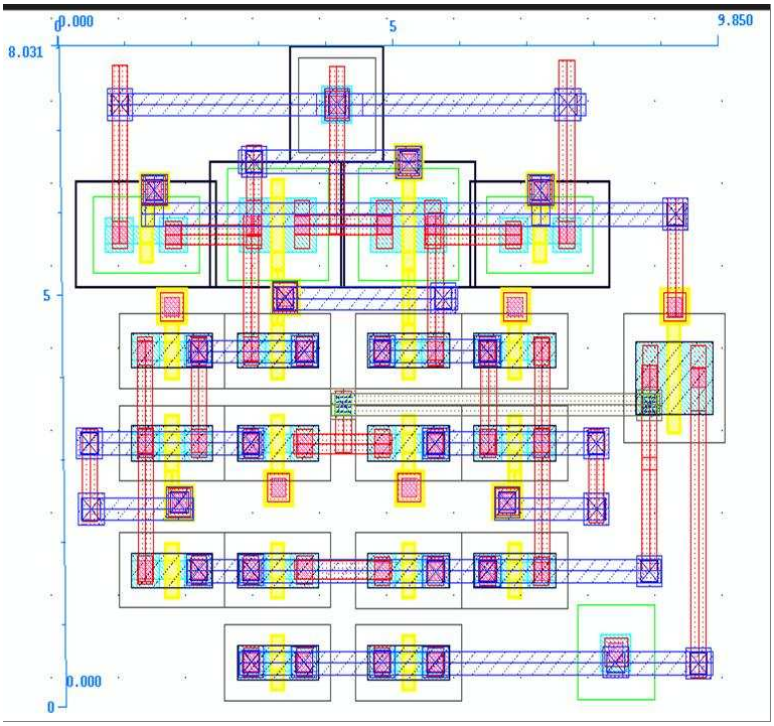


Figure 4.11. Layout of self compensation latch type sense amplifier

CHAPTER 5

MODELING OF DOUBLE GATE CMOS-DGFET

5.1 Device Scaling

Significant work has been done in deriving optimal transistor scaling methodologies. In order to maintain predictable circuit behavior, much more must be considered beyond the actual resizing of the devices. The following list summarizes one approach to the scaling of MOSFETs and the corresponding circuit parameters. Each parameter change presented here is related to the scaling of channel length.

1. The channel width should scale proportional to length in order to allow for overall circuit design scaling to occur without drastic changes in circuit performance. Much of circuit design focuses on the width-to-length ratio and if certain ratios are made unavailable due to scaling, then the approach to circuit design itself may need to be reconsidered.

2. The electric field from drain to source should remain constant (not scale) in order to maintain a channel behavior consistent with existing device models. Increasing the drain-to-source field will likely result in velocity saturation and eventually ballistic carrier transport across the channel. Increasing fields may also lead to drain induced barrier lowering (DIBL)

3. To maintain the constant electric field just mentioned, the drain-to-source voltage (V_{DS}) must also scale down with length.

4. The gate oxide thickness (T_{OX}) should scale down with length to ensure sufficient control over the energy barrier within the channel. Losing control of the energy barrier would most likely result in high subthreshold currents which lead to excess power

dissipation.

5. The doping density of the channel must increase inversely with length to maintain a reasonable number of dopants in the decreasing channel area. Each of the items just presented can be controlled in the design and fabrication steps of circuit production. Unfortunately, there is one critical component that is somewhat beyond design control and which limits our ability to scale many of the other parameters just discussed which is "threshold voltage" (V_T). The value of V_T may be expressed as in equation 5.1.

$$V_T \approx V_{fb} + 2\psi_B + \frac{\sqrt{2\epsilon_{si} q N_a (2\psi_B + V_{bs})}}{C_{OX}} \tag{5.1}$$

where V_{fb} is the flat-band voltage, ψ_B is the difference between the fermi level and the intrinsic level within the channel, ϵ_{si} is the silicon permittivity, q is the electron charge, N_a is the acceptor density, V_{BS} is the substrate reverse-bias voltage and C_{OX} is the oxide capacitance per area . The expression is approximate because the threshold value differs somewhat in the various regions of device operation. While equation shows that some manipulation of the threshold voltage is possible, when the device performance trade-offs are considered (i.e. transconductance, etc.), some control of the threshold is forfeited voluntarily. It is possible to scale V_{DS} and T_{OX} with considerable consistency, V_T tends to deviate from the desired value as length scales down beyond 100 nm.

5.2 Scaling Issues

5.2.1 Short-Channel Effects

It has been shown that optimal circuit performance (i.e. speed, noise immunity, etc.) is achieved when the ratio $V_T \ / V_{DS} < 0.3$. As a result, the supply voltage, V_{DS}, may not be scaled exactly with length. The inability to scale V_{DS} as needed results

in an increased electric field from drain to source leading to short-channel effects like velocity saturation and DIBL. DIBL occurs in short-channel MOSFETs when the depletion region around the drain expands due to an increased drain bias voltage. As the size of the depletion region becomes comparable in size to the channel length, the field lines extending from the drain effectively begin to compete with the gate field lines for channel control resulting in additional lowering of the energy barrier. This barrier lowering is typically modeled as a down-shift in V_T.

While a lower V_T is what was wanted in the first place, now it is a side-effect (not controlled), accompanied by increased sub threshold current and undesired power dissipation. Much work has been done to reverse the short-channel effects. Most of this work revolves around varying the doping profile of the channel.

One doping profile known to suppress DIBL is the"superhalo", shown in Figure 5.1 . In this configuration, the field lines from the drain terminate on an increased number of dopants local to the drain area. This has the effect of decreasing the drain's depletion region extension into the channel, restoring a more symmetric electric potential throughout the channel area.

5.2.2 Mobility Degradation

A second problem associated with device scaling is the effect that decreasing T_{OX} has on carrier mobility within the channel.

T_{OX} is decreased in order to increase the electric field from the gate to the channel, which is a critical part of MOSFET operation as this field controls the energy barrier allowing finite currents to flow through the device. As the gate-to-channel field strength is increased, the channel carriers are attracted closer to the channel surface. Imperfections at the Si - SiO_2 interface result in a new scattering mechanism (surface scattering), which lowers the effective mobility of carriers in the channel. In addition to mobility degradation, thinner oxides result in electron tunneling from

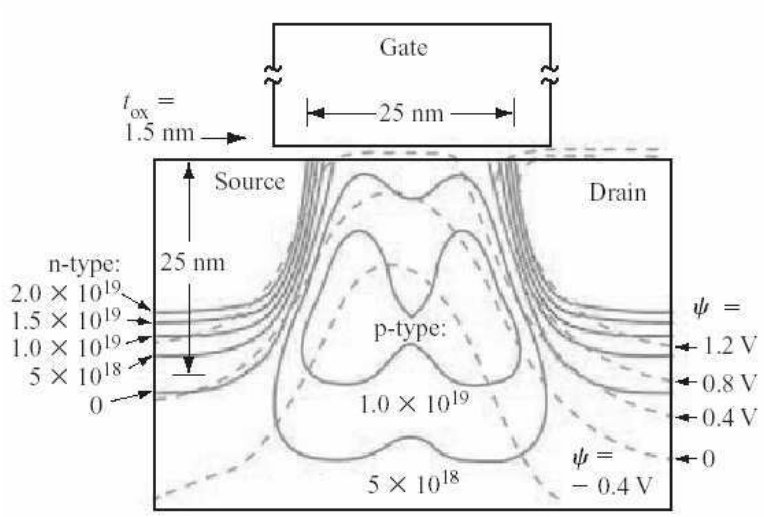


Figure 5.1. The Superhalo doping profile has been shown to suppress DIBL through restoring some of the symmetry in electric potential that is lost as increased drain voltage tends to increase the drain depletion region and field lines. (Y. Taur - March/May 2002)

the gate to the channel, manifested as increased gate leakage current at lower gate voltages. An alternative to scaling the thickness of the oxide layer is to generate the gate-to-channel insulation from a material with a dielectric constant greater than that of SiO_2. While this is possible, it has been difficult to find another insulating material that interfaces as well with the silicon channel. In addition to increased surface scattering, interface imperfections can result in interface charge trapping, which may permanently alter the characteristics of the device. This is why SiO_2 has been an integral part of MOSFET design for so long.

While increasing the dielectric constant increases the electric field from the gate to the channel without the increased gate leakage of a thinner oxide, it does not remedy the degradation of mobility in the channel.

One solution to the decreasing carrier mobility is the"strained-silicon" approach shown in Figure 5.2. in which germanium atoms are scattered throughout the silicon lattice in the channel area. As the silicon atoms attach to the larger germanium atoms, the distance between atoms, and consequently between potential collisions,

are increased. By increasing the spacing between collisions, the average velocity, and hence the mobility, of the carriers is also increased.

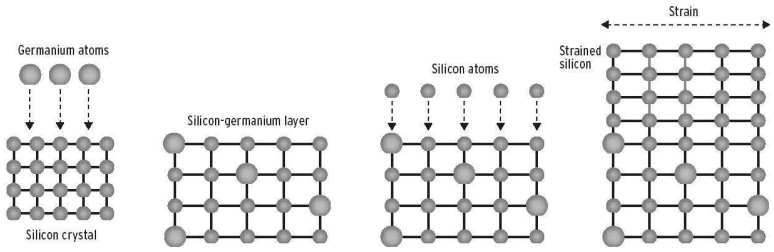


Figure 5.2. Increasing mobility through straining the silicon lattice. As the silicon attaches to the larger germanium atoms, the spacing between atoms and therefore between potential collisions is increased.L. Geppert - Oct 2002

5.2.3 Dopant-Density Fluctuations

A third problem, resulting from the necessary increase in channel doping is increased fluctuation in the actual doping density of the channel. Even with densities greater than 1020/cm3, submicron channel dimensions only allow for tens of dopants in a given channel volume. As dopant numbers are so few in smaller devices, any difference in the existing number of dopants may result in significantly different device behavior. Additionally, while dopants have traditionally been an important part of the MOSFET process, they tend to add to any mobility degradation by creating imperfections within the lattice structure. Channels void of dopants provide a smoother path for charge carriers.

5.3 Proposed Structures

While much has been done to counteract the major side-effects of device scaling, there is still a limit to how far the conventional MOSFET structure can go. As a result, much new architecture has been proposed [53, 12, 15], with the most widely accepted solution being a double-gated structure. shown in Figure 5.3

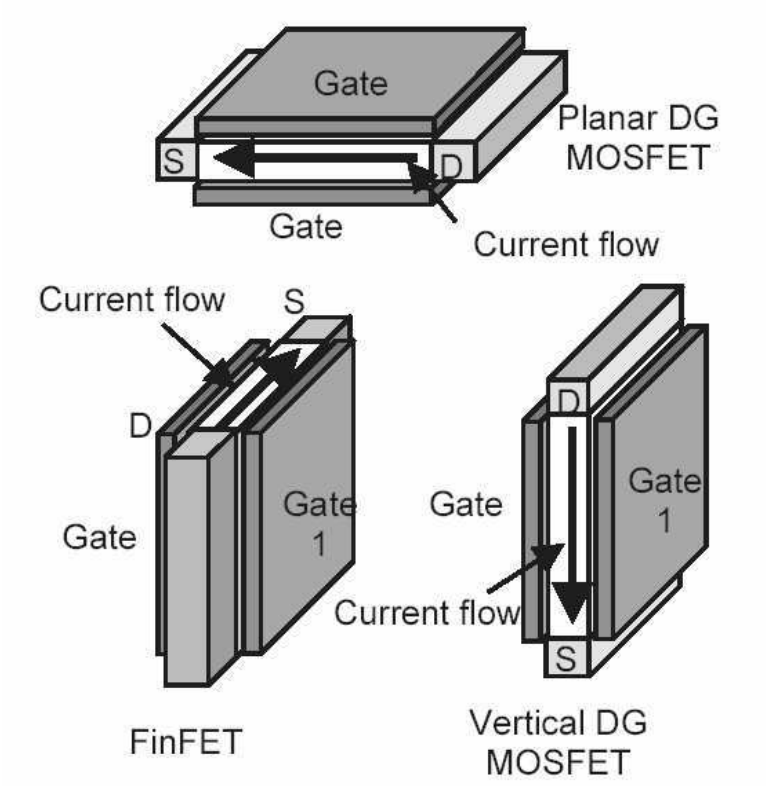


Figure 5.3. Planar, Vertical and FinFET Architectures (M.Chan et all 2003).

The double-gate approach provides increased control over the channel energy barrier, even in the case of thicker gate oxides. Double gates also allow for improved threshold control through engineering the gate material work functions. For example, manipulating the ratio of Si to Ge in a SiGe gate allows for continuous control of the gate work function and hence V_T .Finally, the fact that most double-gate architectures are built upon an SOI (silicon on insulator) foundation removes the body (substrate) terminal and as a result, the body effect.

These three characteristics alone remedy,to some degree, the increased subthreshold current and gate leakage currents of the short-channeled conventional device, the difficulties in scaling V_T and V_{DS} and the body effect, which aggravates timing issues in multi-input digital gate circuits. The use of double-gates to control a silicon channel is not new. Junction Field Effect Transistors (JFETs) have used double gates for decades to control the conduction of a source-drain channel by depleting the channel area of carriers. The difference here is that now the Double gates must be integrate-able into standard VLSI designs.

5.3.1 Doublegate Alternatives

The three double-gate alternatives most widely considered for conventional MOSFET replacement are the planar device, the vertical (pillar) device and the FinFET as shown in Figure 5.3

While the FinFET is the most promising of the three, it is helpful to have an overview of each, in order to better understand the tradeoffs and the reason why FinFETs have been incorporated in my research work for designing Process Variation tolerant Sense Amplifier on FinFET technology.

For our technique we are considering a double-gate FinFET structure with independent gate controls. A double-gate FinFET structure offers additional configuration possibilities such as single channel or double-channel, whereby two gates create

their own independent channels or one common channel. Similarly, the gates may be symmetric or asymmetric in terms of gate capacitance and current drive. The proposed sense amplifier design works for all of the above configurations, while we prefer Double-channel FinFETs due to reduced capacitive coupling between gates.

5.3.2 The Planar Device

The planar device is perhaps the closest alternative to the traditional MOSFET. The critical difference is that there is a second gate located below the channel. As in typical FETs, electrons flow through a channel parallel to the surface of the wafer from source to drain, with additional control provided by the second gate, in this case. While this device structure is only a short step from the conventional MOSFET, its fabrication is significantly more complicated. Alignment of the two gates requires that the gates be developed simultaneously, which is nearly impossible using current fabrication techniques. Creating the gates separately is a simpler process, but often results in gate misalignment, leading to increased gate-to-source (drain) overlap capacitance and mismatch in channel length. To resolve this, some have suggested creating the gates together with some sort of sacrificial layer in between. This layer can then be removed, leaving a tunnel through which the silicon channel can be grown. While this has been shown to work, it requires many additional steps, as well as some complicated processes. Another issue with the planar device is the potential for asymmetry in the two gates. This asymmetry may result from differences in the doping of the poly-Si gates or the oxide thicknesses, offset in the gate voltages or differences in the metal work-functions, where metal gates are used. Early on, it was believed that asymmetry might actually lead to improvements in the suppression of the short-channel effects. By purposely creating asymmetry, likely with the use of n+ and p+ gates, it was believed that the carriers could be forced to one side of the channel allowing for a high level of control with the nearest gate. It was shown,

however, that this line of reasoning was incorrect. Rather, the asymmetry resulted in increased short-channel effects and larger off-state leakage currents. While more recently proposed fabrication techniques for the planar device have been shown to improve the alignment of the gates, the fact remains, that they are still two separate pieces and are therefore very difficult to match precisely.

5.3.3 The Vertical (Pillar) Device

The vertical device is arranged with the two gates running vertically on either side of a vertical channel with source and drain regions at either end. In this configuration, current travels perpendicular to the surface of the wafer. One of the advantages of this configuration over the planar device is that the two gates are more easily created during the same fabrication step as neither is masked by some other layer. But again, asymmetry is an issue as the gates are separate pieces of the same material and may be slightly mismatched.

5.3.4 The FinFET

The FinFET is by far the option being investigated most widely. It resolves many of the concerns mentioned previously. In fact, it improves some of the scaling problems so well, that practically industry has started looking at implementing it even while there is still considerable lifetime in the conventional MOSFET.

5.4 FinFET Fabrication

While a variety of FinFETs have been developed, generally they resemble the device as it was originally proposed.

The reported fabrication process for this structure is as follows : 1. The process begins with a SOI wafer with a 400 nm oxide layer serving as the insulator below 50 nm of silicon.

2. After covering the silicon with a Si_3N_4 and SiO_2 stack layer, a silicon fin is patterned by electron beam lithography.

3.The SOI layer is then etched away exposing the sides of the fin.

4. The source and drain materials (phosphorus-doped-amorphous silicon) and subsequently SiO_2 are deposited, covering the entire fin.

5. A narrow gap between the source and drain are formed through electron beam lithography and etching.

6. With the protective layer above the fin, the amorphous silicon along the sides of the fin is removed.

7. SiO_2 is deposited between the source and drain as spacers. The distance between the SiO_2 spacers is the gate length.

8. The silicon between the spacers is exposed, cleaned and oxidized to create a gate oxide around 2.5 nm.

9. The gate, a boron-doped $Si_0.4Ge_0.6$ material is deposited as the gate. It is also interesting to note that the thick protective layers of SiO_2 above the source and drain regions allow for lower temperature processes through the later part of fabrication making the FinFET compatible with high k dielectric oxides as well as metal gates.

5.5 FinFET Modeling

The FinFET Modeling is done though 3D device simulator and we are using DAMOCLES [25], to run our the device simulation. Transport occurs under the action of collisions and of the external field, which is usually obtained self-consistently with the evolution of the charge density of the particles. Its main (but far from unique) goal is to predict the response of a single device to external voltage and/or current waveforms applied at its contacts. The term device simulation also hints to the existence of a more-or-less complicated and realistic geometrical description of the device, such as the inclusion of ohmic and Schottky contacts, of spatially-

varying doping profiles, of a nonhomogeneous atomic composition due to the presence of various materials (semiconductors, insulators, metals) or various semiconductors (such as in heterostructures).

It is very common to invoke the beneficial feedback between device fabrication and device physics as a justification for further studies. Given this situation, any attempt to distinguish between 'theory' and 'simulation' becomes meaningless. This is particularly true for Monte Carlo simulations which aim at the investigation of often sophisticated physical models included in the collision operator. The basic set of approximations which device simulation (or device 'physics') has embraced since its birth can be implicitly defined by stating the equations usually employed

While precise FinFET modeling equations were not found during this work, it is still very useful to study the more general equations derived for the generic double-gated device This derivation begins with the relative Fermi levels of the gates and channel as shown in figure . From this diagram, one can generate the following equation 5.2

$$E_{FG} + q\phi_G + qV_{OX} = E_C + (E_S - E_C) + q\chi \tag{5.2}$$

where E_{FG} is the Fermi level of the gates, ϕG is the work function of the gate material, E_C is the energy level of the conduction band in the center of the channel, E_S is the energy level of the conduction band at the Si - SiO_2 interface and χ is the electron affinity of the silicon channel Dividing each term by the electron charge (q) and defining $V_G = E_FG/\mathrm{q}$, $\phi_s = -E_S/\mathrm{q}$ and $\mathrm{C} = -E_C/\mathrm{q}$ results in 5.3:

$$V_G - \phi_G - V_{OX} = \phi_C + (\phi_S - \phi_C) + \chi \tag{5.3}$$

After rearranging the last equation thus:

$$V_G - (\phi_G - \chi) - \phi_C = V_{OX} + (\phi_S - \phi_C) \tag{5.4}$$

Each of the terms on the right hand side can be expressed in terms of a ratio of charge density per unit area to capacitance.

$$V_{OX} = \frac{q(N_e + N_A - N_D)}{2C_{OX}} \tag{5.5}$$

Where N_e,N_A and N_D are the electron, acceptor and donor densities respectively and $2C_{OX}$ is the combined capacitance of the two gate oxides. The difference between the electric potential at the center and edges of the channel (the bandbending) can be approximated by a similar term:

$$\phi_S - \phi_C \approx \frac{q(N_e + N_A - N_D)}{2C_d} \tag{5.6}$$

Where C_d is the depletion capacitance. To work towards an expression for the channel current, the carrier density in the channel can be solved by substituting equations 5.5 and 5.6 into 5.4:

$$qN_e = 2\frac{C_{OX}C_d}{C_{OX} + C_d}[V_G - (\phi_M - \chi) - \phi_C - q(N_A - N_D)] \tag{5.7}$$

The complete gate capacitance (C_G) may now be defined as the oxide capacitance (C_{OX}) in series with the depletion capacitance (C_d).

$$C_G = \frac{C_{OX}C_D}{C_{OX} + C_D} \tag{5.8}$$

As this is a general double-gate derivation, it must account for channel doping. In the case of the FinFET, the channel is likely undoped and therefore the N_A and N_D terms can be dropped. Similarly, if the gates are made of polycrystalline-siliicon with Fermi levels of E_{FG}, then the ϕ_G and χ terms could be dropped as well. At this point, a more detailed derivation would account for the existence of sub-bands within the channel by splitting the E_C term into the summation of various band energies $E_{n,k}$.

Under this assumption, the channel charge per unit area (Q_C) may be expressed as

$$Q_C = -qN_e = -qN_C e^{\frac{q(\phi_C - \phi_{Fchan})}{K_B T}} \tag{5.9}$$

where N_C is the effective density of states, ϕ_{Fchan} is the electron quasi-Fermi potential, k_B is Boltzmans constant and T is the temperature in Kelvin. Two final terms are needed before the current-voltage equations can be completed: The semiconductor capacitance and the threshold voltage.

$$C_s = -\frac{dQ_C}{d\phi_C} = \frac{q^2 N_C}{K_B T} e^{\frac{q(\phi_C - \phi_{Fchan})}{K_B T}} \tag{5.10}$$

As the double-gated structure is typically undoped, the lack of inversion charge makes the conventional MOSFET threshold definition, relating silicon charge to bulk depletion charge, irrelavent. It is proposed that the threshold voltage of the double-gate structure be defined as the gate-source voltage (V_{GS}) for which $C_S = 2C_G$. This is a reasonable suggestion in that below the threshold C_S will be the smaller, and hence, dominating factor, whereas above the threshold, C_G will be smaller resulting in device behavior dominated by the gate capacitance. Thus, setting $C_S = 2C_G$ and solving for (ϕ_C - ϕ_{Fchan}) leads to

$$V_T = \phi_G - \chi + \frac{k_B T}{q} log(\frac{2C_G k_B T}{q^2 N_C}) \tag{5.11}$$

Assuming an undoped channel and some contribution to the gate voltage due to the difference in work functions (ϕ_G - χ).

5.5.1 Subthreshold Current

The subthreshold current is simply the product of the channel conductivity and the applied voltage. This can be further broken down as the product of channel

charge density, the cross-sectional area of the channel and the electric field along the channel. In the subthreshold regime, it is possible to replace the ϕ_c - ϕ_{Fchan} terms of the channel charge with V'_G - V for the following reasons. If the channel charge qN_e << $2C_G(K_B\mathrm{T/q})$, then ϕ_C may be approximated by V'_G , where V'_G includes the ϕ_G - χ terms. By defining $V'_{GS} = V'_G$ - V_S and assuming ϕ_{Fchan} equivalent to the source terminal voltage (V_S) plus some variable V , then ϕ_c - $\phi_F chan = V'_G$ - (V_S + V) = V'_{GS} - V. With this substitution the subthreshold current can be written as

$$I_D = Wq\mu_0 N_C e^{\frac{q(V'_{GS}-V}{k_B T}} \frac{dV}{dx} \tag{5.12}$$

where μ_0 is long channel mobility of the channel. After a separation of variables and integration across the channel length and drain-to-source voltage (V_{DS}) the final expression emerges

$$I_D = 2\frac{W}{L}\mu_0 C_G(\frac{k_B T}{q})^2 e^{\frac{q(V_{GS}-V_T)}{k_B T}}[1 - e^{\frac{-qV_{DS}}{k_B T}}] \tag{5.13}$$

Comparison of this expression with the equivalent long channel expression

$$I_D = \frac{W}{L}\mu_0 C_G(\frac{k_B T}{q})^2 e^{\frac{q(V_{GS}-V_T)}{nk_B T}}[1 - e^{\frac{-qV_{DS}}{k_B T}}] \tag{5.14}$$

Where n is the subthreshold slope-factor, reveals that the FinFET provides approximately twice the current of the long-channel device.

5.5.2 Above-Threshold Current

The current above threshold is found in a similar way As before, $\phi_{Fchan} = V_S$ + V , so V = ϕ_{Fchan} - V_S, which can be further simplified by considering that ϕ_C will be pinned at the threshold value when operating in strong inversion.

$$I_D = 2\frac{W}{L}\frac{\mu_0 C_G}{1 + \alpha V_{DS}}[(V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2] \tag{5.15}$$

and again, comparing the FinFET current to the long channel current, shows the FinFET producing approximately twice the current of the long-channel device.

$$I_D = \frac{W}{L}\frac{\mu_0 C_G}{1+\alpha V_{DS}}[(V_{GS}-V_T)V_{DS} - \frac{1}{2}V_{DS}^2] \tag{5.16}$$

Where $\alpha = \frac{\mu_0}{v_{sat}L}$

5.6 Conclusion

Careful study of the model equations as well as consideration of the improvements inherent in the novel structure of the device, validates the use of the FinFET as a good alternative to indefinite scaling the conventional MOSFET. Comparing equations and reveals that the threshold voltage of the FinFET is not a function of the body voltage (V_{bs}). This is advantages in digital design where transistors are often stacked in series between the supply voltages. With conventional MOSFETs, the bottom device has a body-to-source voltage (V_{bs}) = 0, while this is not the case for the devices higher up in the chain. As a result, the bottom device has a lower V_T and therefore turns on faster than the remaining devices. With the body effect removed, as is the case with the FinFET, each device in the series chain can theoretically transition at the same time. Equation also clearly shows the correlation between V_T and the work functions of the gate and channel. As mentioned previously, by creating the gate from a combination of silicon and germanium, the difference between gate and channel Fermi levels can be manipulated thereby shifting the threshold voltage as desired. 3.The current equations and show that the current of the FinFET is approximately twice that of the conventional MOSFET, all other parameters being equal. This larger current translates to a larger transconductance, an important characteristic of FET amplifiers. Higher transconductance will result in the speed of shorter channel lengths without less sacrifice in device gain.

4.The similarity of the FinFET "I - V" equations with those of the conventional MOSFET means that it should be easily integrated into existing circuit design theory.

5.Increased control provided by the second gate decreases the need to scale the gate oxides, thereby lowering the potential gate leakage current. 6.The fact that the channel is left undoped improves the mobility of the charge carriers within the channel.

6.While velocity saturation cannot be avoided, DIBL can by suppressed by starting with a fin width "just smaller" than the gate length. The thin fin width allows the gates to effectively control the energy barrier within the channel with little competition from the drain field.

7.The self-aligning nature of the fabrication process means that assymetry, with its associated problems, will be minimized.

8.To increase the effective width of a gate, multiple FinFETs must be placed in parallel. While this does limit the selection of channel width to discrete values, from a layout perspective, this helps to minimize device mismatches throughout a given circuit. In general, large channel widths are traditionally laid out as a set of thinner transistors wired in parallel to avoid thermal (and other) gradients across the wafer.

These points argue the value of the FinFET in future circuit design. Not only does the proposed architecture help to suppress short-channel effects and undesired power dissipation, it also improves the transconductance of the given device even at fin-widths of tens of nanometers. This implies that circuits with higher transistors densities might be created to function at higher speeds for less power. While the economics of overhauling the conventional fab will most likely push the integration of the FinFET out as long as possible, even at current device sizes (near 45 nm) some organizations may decide that the superior performance of the FinFET merits its more immediate use.

CHAPTER 6

SENSE AMPLIFIER DESIGN ON FINFET TECHNOLOGY

Several double gate structures have been proposed and modeled explained in chapter 5 above. For our technique we are considering double-gate FinFET structure with independent gate controls and can be used with tied gates as well as shown in Figure 6.1 and for its advantages mentioned in chapter-5 and above.

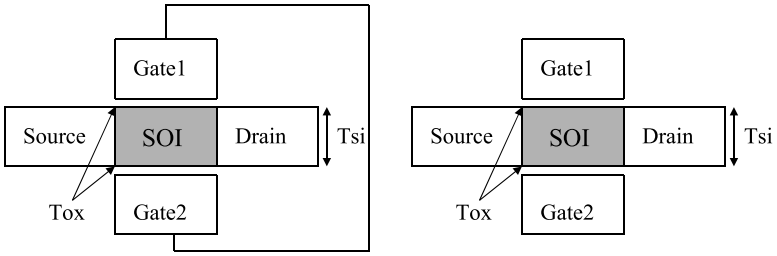


Figure 6.1. FinFET with Tied or Independent Gates.

A double-gate FinFET structure offers additional configuration possibilities such as single channel or Double-channel, whereby two gates create their own independent channels or one common channel. Similarly, the gates may be symmetric or asymmetric in terms of gate capacitance and current drive. The proposed sense amplifier design works for all of the above configurations, while we prefer double-channel FinFETs due to reduced capacitive coupling between gates.In a Double-gate FinFET, the second gate may be biased in a way that changes the threshold voltage for the first gate, similar to the effect of body in bulk CMOS [33, 19]. We take advantage of this property for compensating against process variations.

Previous work has demonstrated performance, power and reliability advantages of FinFETs circuits over the bulk CMOS [33, 1]. Both logic and SRAM FinFET

technologies using independent gate control have been previously demonstrated [13]. In Bulk type conventional MOSFET random dopant fluctuations and oxide thickness variations accounts for variation in threshold voltage and this introduces asymmetry in circuits like sense amplifiers in SRAMs, which in turn reduces SRAM yield and performance. FinFET variations are due to parameter variations. In this study,as of now Predictive Technology Models (PTM) for 32nm FinFET devices is used to demonstrate the effectiveness of the proposed approach [49].

6.1 FinFET based Latch Sense Amplifier

Figure 6.2 shows a basic implementation of latch type voltage sense amplifier design (LSA) on FinFET technology. The technique proposed here can be applied to other types of sense amplifier designs as well. In Figure 6.2 both the gates of the FinFET transistor are tied together so that the threshold voltage of the device is naturally controlled by the input signal. Figure 6.3 are used as the symbols for

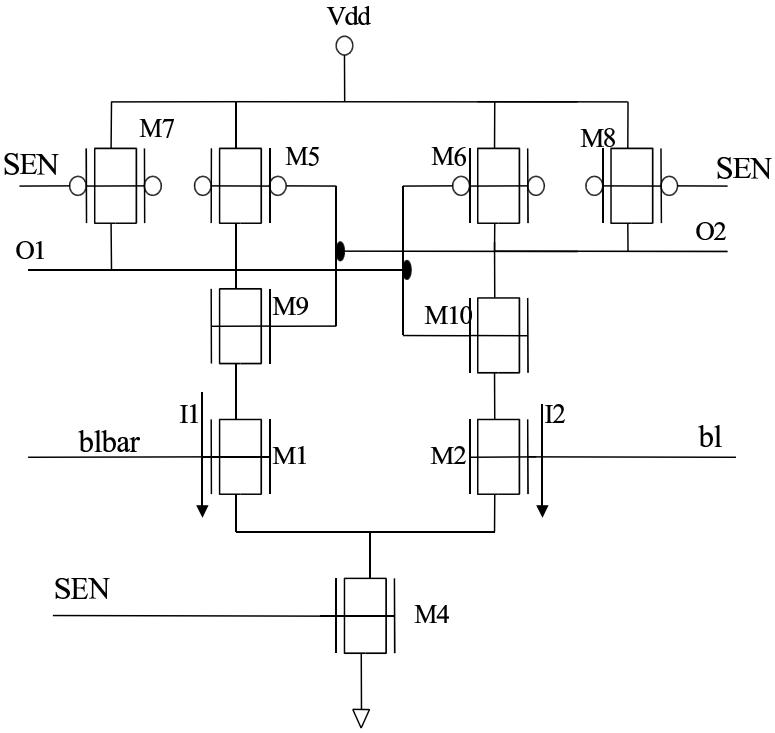


Figure 6.2. Nominal Latch Type Sense Amplifier on FinFET Technology.

n-channel and p-channel FinFET devices in the sense amplifier circuits. Technology specifications are from PTM [49] model for our HSPICE simulation [45].

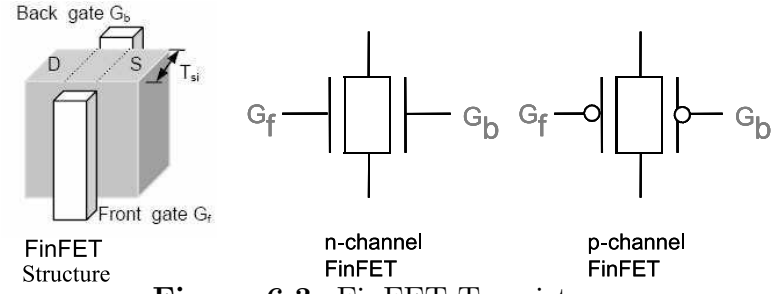


Figure 6.3. FinFET Transistor.

6.2 Working of FinFET based Latch Type Sense Amplifier

The LSA operates in two phases: pre-charge phase and the evaluation phase. The SEN signal decides the phase of the sense amplifier. When SEN is low, sense amplifier is in precharge phase and the outputs are precharged high in this phase. In the evaluation phase, SEN is turned ON for sensing the bit lines. The voltage mode or current mode sense amplifier requires differential discharging of the bit line capacitances and for sensing the input voltage difference. However due to process variations discussed previously and due to feature size scaling the devices mismatch, a wrong output signal may develop for a small input voltage difference which causes access failure in SRAMs. To demonstrate the operation of LSA of Figure 6.2, small offset voltage at the bit line inputs (V_{BLBAR} ¡ V_{BL}) is applied. During the precharge phase, the output node O1 and O2 are initialized to VDD. Sense Enable signal SEN starts sensing the operation by turning on M4. The nodes O1 and O2 are discharged by current I1 and I2 respectively through transistor M1 and M2. As the current directly depends on the voltages applied at the gates of the M1 and M2, I2 ¿ I1 (since V_{BL}¿ V_{BLBAR}). Nodes O1 and O2 begins to discharge at unequal rates. O2 will reach V_{DD} - V_{Tp} sooner than O1. This in turn will switch on FinFET M5 earlier than FinFET M6. Turning on M5 charges O1 and counters the effect of I1, this further

increases the voltage difference between the O1 and O2. The regenerative feedback increases the voltage difference between O1 and O2.

6.3 Impact of Process Variations in LSA

The latch based sense amplifier is designed to be symmetric but with variations due to parameter variations it becomes asymmetric. The effect of the above parameters can be translated into the effective variation in threshold voltage of the sense transistors. Thus we are modeling these variations as V_T variation. This is strictly for an intuitive explanation of the proposed design. To demonstrate the effect of mismatch between the devices let us assume under threshold voltage variations, transistor M2 developed a higher V_T than M1 and that the current I2 becomes lower than I1. Although V_{BLBAR} ¡ V_{BL} this is possible because, I1 is a function of V_{BLBAR} -V_{TM1} while I2 is a function of V_{BL} - V_{TM2}. Since V_{TM2} ¿ V_{TM1}, during bit line discharge it is possible for V_{BLBAR} - V_{TM1} to be greater than V_{BL} - V_{TM2}. Under this condition the stronger current (I1) develops in M1 rather than M2. This results in O1 discharging at a faster rate than O2 causing the sensing to be incorrect. A mismatch of any of the parameters W, L or V_T between the inputs M1 and M2 can result in incorrect operation. Here incorrect sensing happens when $V_{TM1} = V_{TM2} + 50$mV. All parametric failures of memory access can be mapped into offset voltage V_OS which is the minimum bit differential voltage required by the sense amplifier for correct operation (DeltaV). This bitline differential voltage has to be applied to force the inverters in metastable state i.e. V (O1) = V (O2). Only when voltage difference is larger than this offset (V_{OS}), sense amplifier senses the data correctly. It can be seen that the offset voltage of the sense amplifier plays an important role in the access failure probability. The offset voltage can be calculated as a function of the threshold voltages and dimensions of the transistors in the sense amplifier.

In case of process variation V_{OS} can be computed by partial sum of derivatives. The offset voltage depends on the mismatch of device parameters; larger the mismatch greater is the V_{OS}, Seen another way, since Vin must be greater than V_{OS} before sensing it impacts the sensing directly.

As described earlier, the robustness of the sense amplifier is determined by the sense amplifier sensing correctly the inputs at the bit lines even in presence of process variations. Parametric yield describes an important characteristic of sense amplifiers as process variation may affect the ability of the sense amplifier to operate correctly. A single failing sense amplifier may impact the integrity of the entire memory. We performed Monte Carlo simulations for yield analysis varying (W,L, V_T) with 3σ variations to determine the effect on yield with variations in FinFET parameters like V_T, Channel length and width for all the FinFETs versus bit-line differential voltage (DeltaV) in LSA.

A million sense amplifiers were chosen as sample size to illustrate the point. Figure 8 shows the predicted yield of LSA vs bitline differential voltage here DeltaV = V_{BLBAR} -V_{BL} with a 50mV V_T mismatch between the FinFETs M1 and M2. Plot in Figure 8 shows considerable yield loss with V_T mismatch between critical FinFETs of LSA. Please note that all FinFETs are not equally critical. For example, the sense amplifier performance is more sensitive to n-type devices than p-type devices already shown with simulation results in chapter-3.

The proposed sense amplifier is still a latch type sense amplifier with FinFET devices but it has added compensation circuitry as shown in the Figure 6.4.

6.4 Independent Gate Self Compensation Sense Amplifier (IGSSA)

In the modified circuit two critical transistors M1 and M2 (circled) which are most crucial in terms of process variations are now independently biased Rest of the

FinFET transistors has both of their gates tied together. In that case, threshold voltage is controlled by the input signal. The front gate of the M1 and M2 transistor still acts as inputs to BL and BLBAR. The back gate of the critical transistor has the compensation circuit, which includes capacitance C1 and C2 for storing the reference voltage for compensation. C1 and C2 are trained by discharging them first via the CAP_DIS signal. Then M3 and M4 transistors are turned on while V_{BL} and V_{BLBAR} are both at VDD. FET_TRAIN is a controlled pulse signal that initializes C1 and C2. When SAEN is turned ON, back gate of M1 is driven by node voltage at C1 and the back gate of M2 is driven by node voltage of C2 through transistors M10 and M11 respectively.

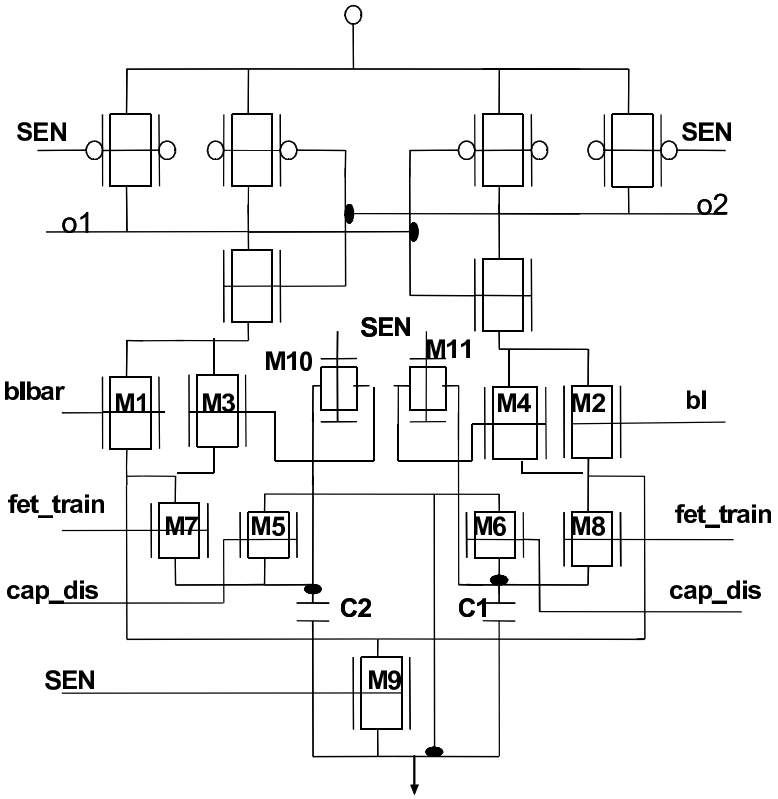


Figure 6.4. Independent Gate Self Compensation Sense Amplifier Design (IGSSA).

Let us assume under process variations in the FinFET V_T increases significantly with decreasing body thickness. As silicon body get thinner, the two gets closer

and have better control over the channel, reducing the short channel effect and V_T roll off. A significant part of the V_T changes with body thickness due to quantum confinement. This is another advantage of using FinFET devices over bulk CMOS. If the voltage at the C2 can be set higher than C1 appropriately, then the effect of V_T mismatch at M1 and M2 can be offset by pumping more current through back gate of the FinFET M2. We understand that the same technique can be applied to conventional bulk type CMOS but in that case we are constrained by short channel effects and optimal performance for the sense amplifier is not achieved. Figure 10 shows that despite V_T mismatch between the two critical FinFETs M1 and M2, the compensation circuitry of IGSSA is tolerant to process variations and senses correctly with the reduced sensing delay. In current bulk CMOS technology, typically a single sense amplifier is deployed for 64 SRAM cells, thus even if IGSSA has some area overhead, it amortizes over the size of the SRAM and final area impact is negligible.

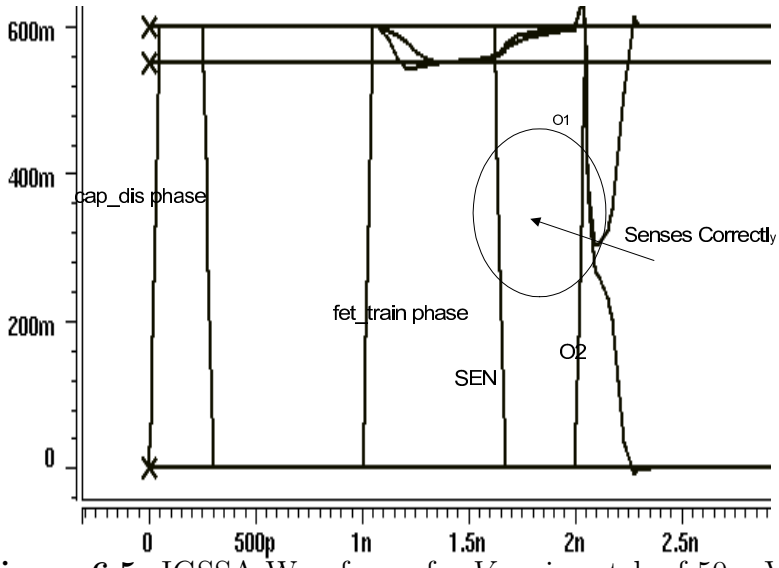


Figure 6.5. IGSSA Waveforms for V_T mismatch of 50 mV.

6.5 IGSSA Operation

The proposed sense amplifier has two added phases of operation, the (i) Capacitance discharge phase and the (ii) Training phase. First phase is for discharging the

capacitances, where the capacitors are fully discharged by asserting CAP_DIS. In the next training phase CAP_DIS is turned off, and FET_TRAIN is turned on, SAEN is off while $V_{BLBAR} = V_{BL}$. Since M1 conducts greater current that charges C2, while M2 has smaller current that charges C1, C2 will develop a higher voltage than C1. The value of capacitances C1 and C2 along with FET_TRAIN pulse width is chosen such that the voltage across the capacitances is greater to overcome the effect of V_T mismatch through the back gate of the FinFET. Observe that the sense delay has also reduced considerably when compared to nominal LSA on FinFET, the technique we are proposing also reduces the delay and this improves the performance. The robustness of the sense amplifier is determined by its yield i.e. by number of failing sense amplifier on account of process variations which we are modeling as V_T mismatch in IGSSA as we did on nominal LSA and in the nominal LSA with no compensation circuitry we experienced incorrect sensing and considerable loss in yield.

6.6 Impact of Process Variations in IGSSA

It is understood that all devices will have normal process variations. Therefore the result can only be verified by statistical simulation. In Monte Carlo simulations, all transistors are subjected to process variations, not just M1 and M2. However, since, M1 and M2 are most sensitive transistors for sense operation, compensation for M1 and M2 provides the greatest benefit. This is also borne out by the results of Monte Carlo simulation in Figure 6.6. To demonstrate the effect of FET_TRAIN phase, IGSSA was run with no FET training phase which caused incorrect sensing of the sense amplifier due to V_T mismatch between transistors M1 and M2 shown in 0. Once the FET_TRAIN phase is turned on, C2 and C1 enables the compensation due to mismatch through the back gate biasing of the FinFET. Even when the V_T mismatch between M1 and M2 is 50 mV, the sense amplifier performs correct sensing. To study the impact of V_T mismatch between M1 and M2 we ran Monte Carlo simulations

with HSPICE [45] with different samples starting with thousand samples to million samples of sense amplifiers with 3s variation of (W,L,V_T). The results in Figure 6.6 show that with the proposed IGSSA, the number of working sense amplifiers (Yield) is nearly identical to that of the design without V_T mismatch or increases with increase in the offset voltage.

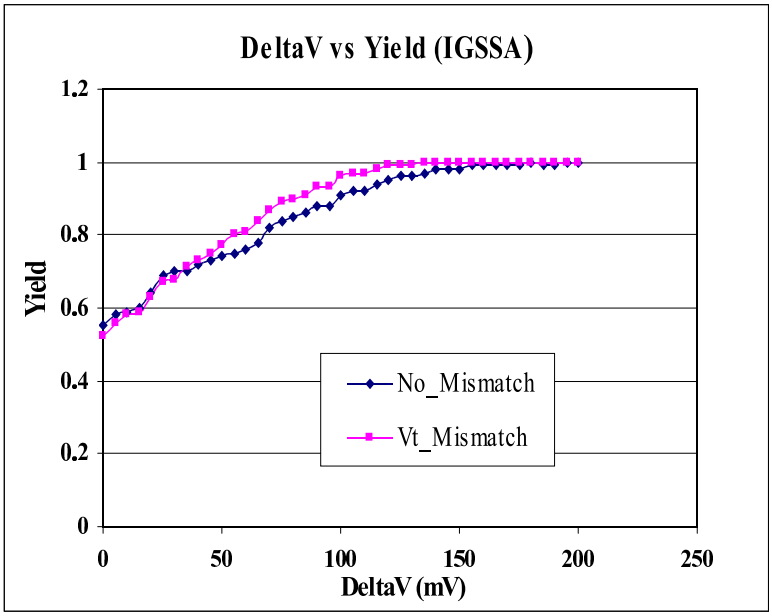


Figure 6.6. Yield versus Bitline Differential Voltage IGSSA technique

The comparison with the yield of IGSSA technique with the normal LSA is shown in Figure 6.7. To study the impact of sample size on yield of the proposed sense amplifier we varied the sample count from thousand to million samples of sense amplifier design. With these sample counts, Monte Carlo simulations were run on the proposed IGSSA. The results are shown in Figure **??**. The yield reaches asymptote within this sample count and further simulations do not impact the result on yield. The effect of increase in sample count of IGSSA on the yield is minimal which can be observed in Figure 6.8 The compensation Circuit introduced is ensuring the robustness and maintains the yield of sense amplifier without affecting its sense delay. IGSSA incurs area overhead due to added compensation circuit. This leads to the question whether the added area could be used in some other way to make the sense amplifier robust.

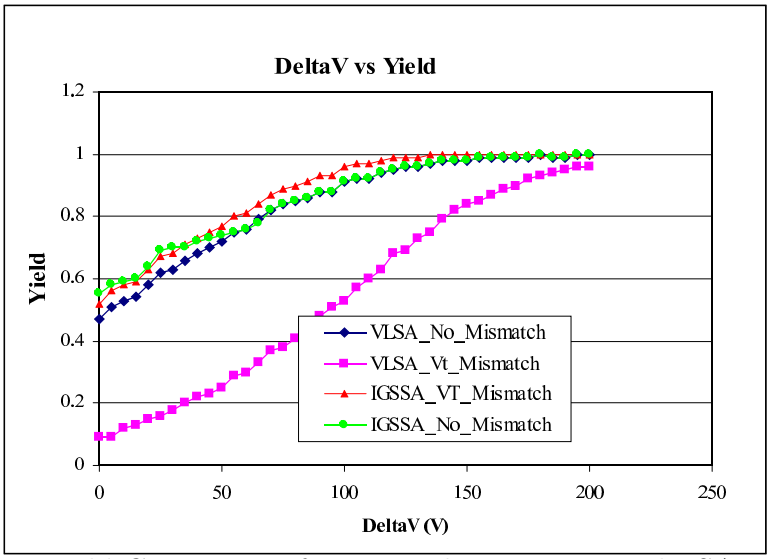


Figure 6.7. Yield Comparison for two techniques nominal LSA and IGSSA

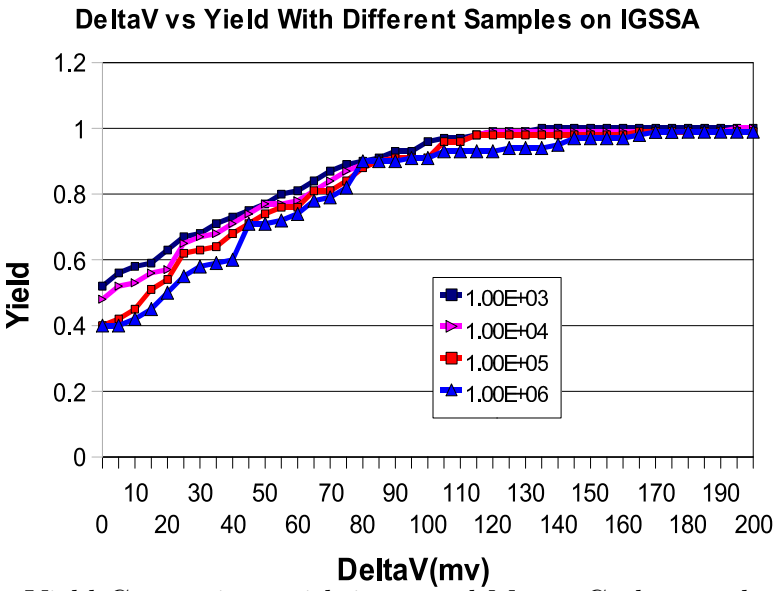


Figure 6.8. Yield Comparison with increased Monte Carlo samples on IGSSA

For example, it is well known that if M1 and M2, the primary sense transistors are increased in size, parameter mismatch between them will decrease. This however is not a viable solution as the bitlines will see a greater load, requiring all cells to be upsized as well, resulting in a bigger area and power penalty. IGSSA incurs a small power overhead. However, it reduces the overall power consumption in SRAM because it senses smaller difference between bitline and bitlinebar that allows the read to be terminated before a larger voltage difference develops between bitline and bitline bar. The actual savings depends on the number of SRAM cells per column.

6.7 Conclusions

Process variation in conventional bulk CMOS is a significant yield detractor for SRAM designs. We presented a process variation tolerant self-compensating sense amplifier design on conventional bulk type CMOS sense amplifier design. The same technique we extended to next generation double gate CMOS devices FinFET catering to the road map of semiconductor devices.We generated our own model from damocles [25] and taurus device simulator [47] for generating our simulation results. Experimental results show that that the proposed design withstands large mismatches in the primary sense transistors and produce yield comparable to designs without any process variation. Statistical simulations show robustness of the design for variations in all transistors and not just the primary transistors. The simulation results further show that for a given performance, our scheme improves yield while for a fixed yield our design scheme results in faster sensing.It does increase the area of a sense amplifier almost twice, but since there is only one sense amplifier required across the column of the SRAM. The layout of sense amplifier with 128X8 SRAM on nominal type and SSA technique to determine area and powers tradeoff was drawn. The area increase for the amount of compensation towards yield and performance is not alarming and is only .48% increase from SRAM with nominal type of sense amplifier.

BIBLIOGRAPHY

[1] Ahmed, Bin Yu Leland Chang, Bell, S. Haihong Wang, Tabery, S. Chih-Yuh Yang, Xiang, C. Chau Ho Qi, Bokor, Tsu-Jae King, Hu, J. Chenming, and D., Ming-Ren Lin Kyser. "FinFET scaling to 10 nm gate length". In *Electron Devices Meeting, IEDM* (Digest. International, Sunnyvale, CA, USA, 2002), pp. 251–254.

[2] Alam, M. A., Kang, K., Paul, B. C., and Roy, K. "Reliability-and Process-Variation Aware Design of VLSI Circuits". pp. 17–25. also cited 1214.

[3] Asenov, A., Slavcheva, G., Brown, A.R., Davies, J.H., and Saini, S. "Quantum mechanical enhancement of the random dopant inducedthreshold voltage fluctuations and lowering in sub 0.1 micron MOSFETs". In *Electron Devices Meeting, 1999. IEDM Technical Digest. International* (Mar. 1999), pp. 535–538.

[4] Barin, N., Braccioli, M., Fiegna, C., and Sangiorgi, E. "Scaling the high-performance Double-gate SOI MOSFET down to 32 nm Technology Node".

[5] Barin, N., Braccioli, M., Fiegna, C., and Sangiorgi, E. "Analysis of Scaling Strategies for sub-30 nm Double-gate SOI N-MOSFETs". *Nantechnology, IEEE Transactions on 6*, 4 (July 2007), 421–430.

[6] Bastos, J. *Characterization of MOS Transistor Mismatch for Analog Design.* PhD thesis, Katholieke Universteit Leuven, Dec. 1998.

[7] Bhunia, Swarup, Mukhopadhyay, Saibal, and Roy, Kaushik. "Process variations and Process-Tolerant Design". In *VLSID '07: Proceedings of the 20th International Conference on VLSI Design held jointly with 6th International Conference* (Washington, DC, USA, 2007), IEEE Computer Society, pp. 699–704.

[8] Boning, D., and Nassif, S. "Models of Process Variations in Device and Interconnect". In *In Design of High-Performance Microprocessor Circuits* (2000).

[9] Cadence. "Cadence 90nm generic product design kit.". www.pdk.cadence.com, 2007.

[10] Cakici, T., Bansal, A., and Roy, K. "a low power four transistor schmitt trigger for asymmetric double gate fully depleted SOI devices". In *SOI Conference, 2003. IEEE International* (Oct 2003), pp. 21– 22.

[11] Chang, Leland, Choi, Yang-Kyu, Ha, Daewon, Ranade, Pushkar, Xiong, Shiying, Bokor, Jeffrey, Hu, Chenming, and King, Tsu-Jae. "extremely scaled silicon nano-cmos devices". In *IEDM* (2003), IEEE, pp. 139–144.

[12] Frank, D. J., Laux, S. E., and Fischetti, M. V. "monte carlo simulation of a 30 nm double-gate mosfet: How short can si go?". In *IEEE Electron Devices Meeting, 2001. IEDM Technical Digest. International, Washington, DC, USA* (2001), pp. 19.5.1–19.5.4.

[13] Guo, Zheng, Balasubramanian, S., Zlatanovici, R., King, Tsu-Jae, and Nikolic, B. "FinFET-based SRAM design". In *Low Power Electronics and Design, 2005. ISLPED* (California Univ., Berkeley, CA, USA, 2005), pp. 2–7.

[14] Hansson, M., Alvandpour, A., Hsu, S., and Krishnamurthy, R. "A process variation tolerant technique for sub-70nm latches and flip-flops". In *NORCHIP Conference, 2005. 23rd* (Nov 2005), pp. 149– 152.

[15] Hisamoto, and Lee, Wen-Chin. "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm". In *Proceedings of the IEEE Electron Devices* (Dec. 1982), pp. 82–85.

[16] Hisamoto, Digh, Lee, Wen-Chin, Kedzierski, Jakub, Takeuchi, Hideki, Asano, Kazuya, Kuo, Charles, Anderson, Erik, King, Tsu-Jae, Bokor, Jeffrey, and Hu, Chenming. "FinFETA Self-Aligned double-gate MOSFET Scalable to 20 nm". In *IEDM* (2000), IEEE, pp. 2320–2324.

[17] IBM, and to add strained silicon, AMD. http://www-03.ibm.com/press/us/en/pressrelease/7511.wss, Dec. 2004.

[18] Iwai, H. "future semiconductor manufacturing: challenges and opportunities". In *Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International* (2004), pp. 11–16.

[19] Kedzierski, J., Fried, D.M., Nowak, E.J., Kanarsky, T., Rankin, J.H., Hanafi, H., Natzle, W., Boyd, D., Zhang, Ying, Roy, R.A., Newbury, J., Yu, Chienfan, Yang, Qingyun, Saunders, P., Willets, C.P., Johnson, A., Cole, S.P., Young, H.E.and Carpenter, N., Rakowski, D., Rainey, B.A., Cottrell, P.E., Ieong, M., and Wong, H.-S.P. "High-performance symmetric-gate and CMOS-compatible vtasymmetric-gate FinFET devices". In *IEEE Electron Devices Meeting, 2001. IEDM Technical Digest. International, Washington, DC, USA* (2001), pp. 19.5.1–19.5.4.

[20] Kedzierski, J. Meikei Ieong Kanarsky, T Ying Zhang Wong H.-S.P. "Fabrication of metal gated FinFETs through complete gate silicidation with ni". In *Proceedings of the IEEE* (IBM T. J. Watson Res. Center, Yorktown Heights, NY, USA, 2004), pp. 2115–2120.

[21] Kim, K., Das, K.K., Joshi, R.V., and Chuang, C.-T. "Leakage Power Analysis of 25-nm Double-Gate CMOS Devices and Circuits". *Electron Devices, IEEE Transactions on 52* (2005), 980–986.

[22] Kim, Keunwoo, and Fossum, J.G. "Double-gate CMOS: symmetrical- versus asymmetrical-gate devices". In *Proceedings of the IEEE Electron Devices* (Feb. 2001), pp. 294–299.

[23] Kobayashi, T., Nogami, K., Shirotori, T., and Fujimoto, Y. "A current-controlled latch sense amplifier and a staticpower-saving input buffer for low-power architecture". In *Solid-State Circuits, IEEE Journal of* (April 1993), pp. 523–527.

[24] Kobayashi, Tsuguo, Nogami, KAzutaka, Shirotori, Tsukasa, and Fujimoto, Yukihiro. "A Current-Controlled Latch Sense Amplifier and a Static Power-Saving Input Buffer for Low-Power Architecture". In *IEEE JOURNAL of Solid State Circuits* (April 1993), IEEE, p. 1413 1422.

[25] Laux, S. E., and Fischett, M. V. "DAMOCLES: Monte Carlo simulation of semiconductor devices". http://researchweb.watson.ibm.com/DAMOCLES/html_files/soi.html, 1999.

[26] Lu, P., Ji, J., Chuang, C., Wagner, L., Hsieh, C., Kuang, J., Hsu, L., Pelella, M., Chu, S., and Anderson, C. "Floating body effects in partially-depleted SOI cmos circuits". In *ISLPED* (1996), Mark Horowitz, Jan M. Rabaey, Brock Barton, and Massoud Pedram, Eds., IEEE, pp. 139–144.

[27] Moore, G.E. "Cramming more components onto integrated circuits". In *Proceedings of the IEEE* (Jan 1998), pp. 82–85.

[28] Mukhopadhyay, Saibal, ArijitRayChowdhury, Mahmoodi, Hamid, and Roy, Kaushik. "Leakage Current Based Stabilization Scheme for Robust Sense-Amplifier Design for Yield Enhancement in Nano-scale SRAM". In *IEEE 14th Asian Test Symopsium 2005* (Feb 2005), pp. 183–192.

[29] Mukhopadhyay, Saibal, Mahmoodi, Hamid, and Roy, Kaushik. "a novel high-performance and robust sense amplifier using independent gate control in sub-50-nm double-gate MOSFET". In *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* (Feb 2006), pp. 183–192.

[30] Nassif, Sani R. Design for Variability in DSM Technologies. In *ISQED* (2000), pp. 451–454.

[31] Natarajan, A., Shankar, V., Maheshwari, A., and Burleson, W. "Sensing design issues in deep submicron CMOS SRAMs". In *VLSI, 2005. Proceedings. IEEE Computer Society Annual Symposium on* (May 2005), pp. 42–45.

[32] Nowak, Edward J. "Maintaining the benefits of CMOS scaling when scaling bogs down". *IBM Journal of Research and Development 46*, 2-3 (2002), 169–186.

[33] Nowak, E.J, Aller, I., Ludwig, T., Kim, K., Joshi, R.V., Chuang, Ching-Te, Bernstein, K., and Puri, R. "Turning silicon on its edge [double gate CMOS/FinFET technology". *IEEE Parallel and Distributed Technology 25*, 10 (2004), 251–254.

[34] Nowak, E.J., Ludwig, T., Aller, I., Kedzierski, J., Leong, M., Rainey, B., Breitwisch, M., Gemhoefer, V., Keinert, J., and Fried, D.M. "Scaling beyond the 65 nm node with FinFET-DGCMOS". In *Custom Integrated Circuits Conference, Proceedings of the IEEE* (Sept. 2003).

[35] Rahul Sarpeshkar, John L. Wyatt, Jr. Nicky C. Lu, and Gerber, Porter D. "Mismatch Sensitivity of a Simultaneously Latched CMOS Sense Amplifier". In *IEEE JOURNAL OF SOLID-STATE CIRCUITS* (10 October 1991), pp. 217–224.

[36] Roy, Kaushik, Mahmoodi-Meimand, Hamid, Mukhopadhyay, Saibal, Ananthan, Hari, Bansal, Aditya, and Cakici, Tamer. "Double-gate SOI devices for low-power and high-performance applications". In *ICCAD* (2005), pp. 217–224.

[37] Sairam, Tarun, Zhao, Wei, and Cao, Yu. "Optimizing finfet technology for high-speed and low-power design". In *GLSVLSI '07: Proceedings of the 17th great lakes symposium on Great lakes symposium on VLSI* (New York, NY, USA, 2007), ACM, pp. 73–77.

[38] Sarpeshkar, Rahul, Wyatt, John L., Lu, Nicky C., and Gerber, Porter D. "Mismatch sensitivity of a simultaneously latched CMOS sense amplifier". In *IEEE JSSC* (Oct 1991), IEEE, p. 1413 1422.

[39] Sery, G. "Impact of process variation phenomena on performance and quality assessment". In *Integrated Circuit Design and Technology, 2004. ICICDT '04. International Conference on* (Intel Corp, Mar. 2004), pp. 25–.

[40] SIA. International Technology Roadmap for Semiconductors. http://www.itrs.net/reports.html, 2007.

[41] Sinha, Manoj, Hsu, Steven, Alvandpour, Atila, Burleson, Wayne, Krishnamurthy, Ram, and Borhr, Shekhar. "High-Performance and Low-Voltage Sense-Amplifier Techniques for sub-90nm SRAM". In *IEEE Int. SOC Conf* (Sept 2003), pp. 113–116.

[42] Siva Narendra, Intel Corporation, Vivek De, Intel Corporation, and Ron Wilson, ISD magazine. "Process Variation: Is it too Much to Handle?". In *ISQED '02: Proceedings of the 3rd International Symposium on Quality Electronic Design* (Washington, DC, USA, 2002), IEEE Computer Society, p. 213.

[43] Su, Lisa T., Jacobs, Jarvis B., Chung, James E., and Antoniadis, Dimitri A. "Short-Channel Effects in deep-submicrometer SOI MOSFETs". In *IEDM* (1993), IEEE, pp. 139–144.

[44] Suzuki, Kunihiro, Tanaka, Tetsu, Tosaka, Yoshiharu, and Horie, Hiroshi. "Scaling Theory for Double-gate SOI MOSFETs". In *IEDM* (1993), IEEE, pp. 139–144.

[45] Synopsys. Hspice. www.synopsys.com/products/hspice/hspice.html, 2007.

[46] Synopsys. "MEDICI:2-D Device Simulation Program.". www.synopsys.com/products/medici/medici.html, 2007.

[47] Synopsys. "TCAD:3-D Device Simulation Program.". www.synopsys.com/products/tcad/tcad.html, 2007.

[48] Tschanz, Jim, Bowman, Keith, and De, Vivek. "Variation-Tolerant Circuits:Circuit Solutions and Techniques". DAC, June 2005.

[49] University, Arizona State. "predictive technology model". http://www.eas.asu.edu/ ptm/, 2007.

[50] Wang, Feng, Xie, Yuan, Bernstein, K., and Luo, Yan. "Dependability analysis of nano-scale FinFET circuits". In *Emerging VLSI Technologies and Architectures, 2006.* (Mar. 2006), pp. 6–.

[51] Wei., L., and all. "vertically intergrated soi circuits for low power and high performance applications". In *IEEE Electron Devices Meeting, 2002. IEDM Technical Digest. International, Washington, DC, USA* (June 2002), pp. 351–362.

[52] Wicht, B., Nirschl, T., and Schmitt-Landsiedel, D. "Yield and speed optimization of a latch-type voltage sense amplifier". In *Solid-State Circuits, IEEE Journal of* (July 2004), pp. 1148– 1158.

[53] Wong, H.-S.P, Chan, K.K., and Taur, Y. "Self-aligned (top and bottom) double-gate MOSFET with a 25 nm thicksilicon channel". In *IEDM* (IBM Thomas J. Watson Res. Center, Yorktown Heights, NY, 1997), IEEE, pp. 427–430.

[54] Wulf, Win. A., and McKee, Sally A. "Hitting the Memory Wall: Implications of the Obvious". In *IEDM* (1994), IEEE.