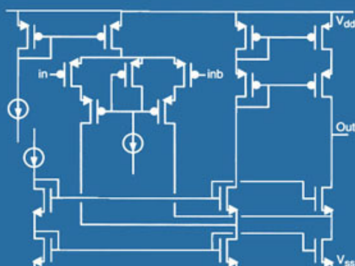
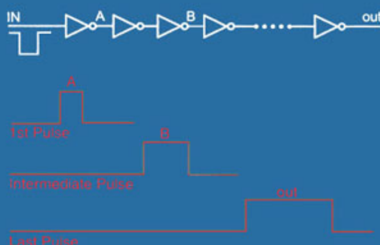
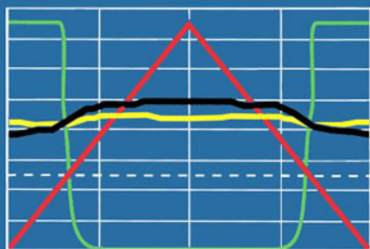
SOI Design:

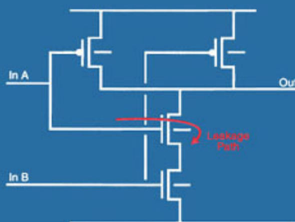
Analog, Memory and Digital Techniques

Andrew Marshall
Sreedhar Natarajan









Kluwer Academic Publishers

SOI DESIGN

ANALOG, MEMORY AND DIGITAL TECHNIQUES

SOI DESIGN

ANALOG, MEMORY AND DIGITAL TECHNIQUES

by

Andrew Marshall
&
Sreedhar Natarajan

Texas Instruments Incorporated

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-306-48161-8
Print ISBN: 0-7923-7640-4

©2003 Kluwer Academic Publishers
New York, Boston, Dordrecht, London, Moscow

Print ©2002 Kluwer Academic Publishers
Dordrecht

All rights reserved

No part of this eBook may be reproduced or transmitted in any form or by any means, electronic, mechanical, recording, or otherwise, without written consent from the Publisher

Created in the United States of America

Visit Kluwer Online at: http://kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

To Judith, Amy and Elise for their enduring patience and encouragement.

-- Andrew Marshall

To my wife Chandra for her patience and support, my Parents, my brother Sriram and his family, my wife's family, and friends for their continued encouragement.

-- Sreedhar Natarajan

This page intentionally left blank

Table of Contents

Preface

Chapter 1: Overview

Silicon on Insulator – a brief Introduction

Circuits and SOI

Technology and SOI

7.10 Dynamic 2-Way OR Circuit 145

7.11 Dynamic

8.8 Decoders

8.9 SRAM Architecture

8.10 Bit Line Related Architecture

8.11 Sense Amplifiers
8.11.1 Differential Amplifier
8.11.2 Clocked Dual Slope Sense Amplifiers
8.11.3 Dynamic Body Charge
Dynamic 8.11.4 Body Charge
ic Body Charge

9.5.2 Quasi- SOI technology

9.6 Influence Of SER on SOI DRAMs

9.7 Cosmic Ray induced Soft Errors in SOI DRAMs

9.8 DRAM Refresh aOI DRAMs

9.21 Destructive read out characteristics of DRAM

Chapter 10: SOI Analog Design

10.1 Introduction
10.1.1 Benefits of SOI for Analog Design
10.1.2 Drawbacks o DRAMChapter 10: SOI Analog Design10.1 Introduction10.1.1
Drawbacks o DRAMChapter 10: SOI Analog Design10.1 Introduction10.1.1
Drawbacks o DRAMChapter 10: SOI Analog Design10.1 Introduction10.1.1
Drawbacks o DRAMChapter 10: SOI Analog Design10.1 Introduction10.1.1
Drawbacks o DRAMChapter 10: SOI Analog Design10.1 Introduction10.1.1

β

I_{DD})

This page intentionally left blank

Preface

In the 1850's my wife's great-great grandfather wrote a medical textbook. In the year and a half that it has taken to complete this work I have grown to appreciate not only how difficult it is to write a book, but also how much more difficult it must have been back then. One hundred and fifty years ago Dr. Julius Jeffreys had to write his entire tome by hand, without the aid of error correcting word processing software or even electric light. More obviously in electronics than in many other fields, the pace of technology development since that time has been astounding. Even in just the past fifty years semiconductor technology has progressed from little more than a laboratory curiosity to an integral part of our daily lives.

Now silicon on insulator technology is becoming the process of choice for high performance consumer electronics, it is appropriate that memory systems and analog circuitry are considered in a single volume with the SOI process and logic implementation. Such an overview helps put SOI into perspective. What we see is a very impressive technology, which holds a great deal of promise. There are applications where SOI does not appear to be a good fit, but generally it appears that SOI has a bright future.

We have compiled this book, based on available literature and SOI circuit development work at Texas Instruments.

Chapter 1 introduces the reader to silicon on insulator, with a very brief review of the history of SOI technology and circuits.

Chapter 2 briefly reviews the materials of SOI, and methods used to generate the most common forms of SOI material used commercially. Manufacturing considerations and materials properties are highlighted in this section.

In chapter 3 we review silicon on insulator devices, with discussion of how they differ from devices available to the designer of bulk silicon integrated circuits. Electrical properties associated with SOI devices are introduced in the discussion of MOS devices, diodes, bipolar transistors and passives.

Chapter 4 explores the challenges of modeling SOI material and devices, with particular emphasis on circuit simulation. The chapter will be of particular interest to the designer of analog circuits, as an introduction to circuit simulation of SOI components.

Chapter 5 introduces some concepts for successful SOI circuit layout. We review various options for the reduction of thermal self-heating effects and

consider the conversion of designs from bulk to SOI from a layout perspective.

Chapters 6 and 7 provide a review of static and dynamic logic circuits by way of introduction into the memory sections. While this has been covered in detail elsewhere (notably Kerry Bernstein and Norman Rohrer' SOI Circuit Design Concepts), the summary here permits a full understanding of the possibilities and challenges of SOI essential to a good cognition of more complex circuitry.

Chapters 8 and 9 are devoted to the dominant memory structures in current use, the SRAM and DRAM. Methods to improve memory performance at the memory cell and sense amplifier level are discussed.

Chapter 10 addresses techniques to permit the integration of analog circuitry into SOI designs. This is an inevitable requirement as SOI use transfers from just microprocessors to DSP and mixed signal applications.

Chapters 11 and 12 examine global design issues, those issues that affect both digital and analog circuit design, and are more easily discussed in a general setting. Low power and low voltage circuit design is of particular importance to the industry, and is thus considered in a chapter by itself.

Chapter 13 takes a look at some of the possibilities for SOI in development, both from the material and circuit viewpoints.

For updates, corrections etc., the web site associated with this publication is given in appendix 1. This site also has contact information where you the reader may communicate suggestions, errata etc. We hope this will prove to be a valuable source of data in this rapidly evolving field.

In summary, we echo the words of Kerry Bernstein and Norman Rohrer, when in the final words of their preface to "SOI Circuit Design Concepts" they state, "Rather than designing to avoid SOI's idiosyncrasies, the true benefit of the technology may be realized when these features are exploited".

Andrew Marshall
Sreedhar Natarajan

October 2001

Acknowledgements

It is inevitable with a work of this length and scope that a huge number of individuals have had a hand in enhancing the content. We thank everyone involved in the work, but would especially like to thank the following:

We extend a special thanks to the reviewers, for their very helpful suggestions: Brock Barton, Charvaka Duvvury, John Fattaruso, David Scott, Howard Tigelaar, Tom Vrotsos.

For discussions and insight: Terry Blake, James Burns, Xiaowei Deng, Olivier Faynot, Jim Gallia, Keith Green, Ted Houston, Keith Joyner, Weideng Liu, Homi Mogul, Betty Prince.

For drafting, simulation, and silicon evaluation: Erich Caulfield, Andres Diaz, Imran Hossain, Rick Little, Gary Mathews, Brian Neal, Michelle Nguyen, Mouli Vaidyanathan.

For enabling the work: Randy McKee, Yoshio Nishi, David Spratt, Tom Thorpe and especially Wah Kit Loh.

This page intentionally left blank

Chapter 1: Overview

1.1 Silicon on Insulator – a brief Introduction

Silicon on insulator (SOI) is a very attractive technology for large volume integrated circuit production and is particularly good for low-voltage, low-power and high-speed digital systems. SOI has also proved to be effective in various niche and growing markets. IC processes based on SOI are known to reduce susceptibility to radiation, and have been used for many years in high radiation environments. SOI is also used for power integrated circuits, micro-electromechanical systems (MEMS), integrated optics and high temperature applications.

Devices built on SOI have lower capacitances to substrate than comparable devices on bulk silicon. MOS transistors on SOI are also, in most cases, capable of higher current density than obtained with equivalent devices on bulk. These improvements are achieved without reductions in process geometries, and thus have the benefit of delaying the introduction of higher precision production equipment, resulting in significant capital equipment cost deferral. The performance advantages associated with SOI have led to consideration of the technology for high performance applications, such as microprocessors, DSPs and wireless applications.

Since being introduced into limited production in the early 1990s the variety of SOI explained by the term "active silicon, separated from a silicon substrate by a silicon oxide insulator" has been developed into a mainstream technology, especially for digital applications [1.1 – 1.4]. Market predictions suggest that SOI will receive a growing proportion of the total integrated circuit market, particularly for high performance requirements.

1.2 Circuits and SOI

SOI offers numerous opportunities and challenges in the design of low-voltage and low-power CMOS circuits for both analog and digital applications. The benefits of this technology for digital applications have been clear for many years. The exploitation of SOI for analog and memory subsystems, meanwhile, has lagged behind digital developments, but is now beginning to attain a level of parity, with circuits that are in some cases improved over their bulk counterparts. SOI is suitable for digital, memory and analog designs, although it is not necessarily straightforward to convert circuits developed for bulk processes into SOI. Device operation is affected by SOI specific factors, including resistive body ties and localized self-heating.

The circuit design examples covered in this book have, for the most part, been developed specifically for use with partially depleted (PD) SOI rather than fully depleted SOI. This is because:

(a) The current lack of availability of good quality fully-depleted (FD) SOI material in abundance, and;

(b) Once PD-SOI analog and digital circuit designs are available the conversion of these circuits to FD-SOI is expected to be straightforward.

Memory and most analog circuits either interface to, or are incorporated within, a digital environment. The design of analog circuits on SOI, in a mixed signal environment, and memory design in an embedded memory application are discussed. Various processes are examined and comparison is made between bulk and SOI circuit design concepts. SOI is the process of choice in various RF applications, particularly when digital circuitry is required. In addition to known and well understood advances in digital circuit operation, the use of SOI can directly lead to improved passive component capability (particularly inductors) and active RF component frequency response.

The potential to integrate high power and high voltage components is a significant benefit of SOI processes. In addition SOI permits simpler higher temperature circuit design (potentially negating some of the detrimental consequences of increased temperature circuit elements due to local heating). As is well known, SOI also gives improved immunity to radiation effects (so-called rad-hard circuit design). SOI CMOS allows fabrication, design and operation at low supply voltages for analog, digital and microwave components with properties significantly superior to those obtained using bulk silicon CMOS.

1.3 Technology and SOI

There are a number of books on SOI technology [1.5 – 1.7]. This work examines some of the basics, but is primarily concerned with circuit related issues. Static and dynamic logic circuit design has previously been studied in some detail [1.5], however, memory design for SOI and analog circuit designs have hitherto been examined only in a piecemeal manner.

SOI material is considered here in terms of implementations that are promising or have been used elsewhere in circuit development, with historical perspective where appropriate. The two basic methods for producing SOI substrates are SIMOX and bonded silicon technologies. In most cases these are interchangeable in terms of circuit design.

References

[1.1] F. Assaderaghi, et. al., “Transient Pass-Transistor Leakage Current in SOI MOSFET's”, IEEE Electron Device Letters, June 1997, pp. 241-3.

[1.2] G. Shahidi, et. al., “Mainstreaming of the SOI Technology”, IEEE International SOI Conference, Oct 1999, pp. 1-4.

[1.3] F. Assaderaghi, et. al., “History-Dependence of Non-Fully Depleted Digital SOI Circuits”, VLSI Tech. Digest, 1996, pp. 122-3.

[1.4] M. Canada, et. al., “A 580MHz RISC Microprocessor in SOI”, ISSCC 1999, pp. 430-431.

[1.5] K. Bernstein & N. J. Rohrer, “SOI Circuit Design Concepts”, Kluwer Academic Publishers, January 2000, ISBN 0-7923-7762-1

[1.6] J-P. Colinge, “Silicon-on-Insulator Technology Materials to VLSI, 2nd Edition” Kluwer Academic Press, September 1997, ISBN 0-7923-8007-X

[1.7] S. Cristoloveanu & S. S. Li, “Electrical Characterization of Silicon-on-Insulator Materials and Devices” Kluwer Academic Press, June 1995, ISBN 0-7923-9548-4

Chapter 2: SOI Materials

2.1 Silicon on Heteroepitaxial Substrate

SOI has been created by both epitaxial growth and by transfer of an active layer onto various types of substrate. Some of the most common heteroepitaxial substrates for silicon include diamond, sapphire, cubic zirconia, calcium fluoride, quartz and glass. Among these processes silicon on sapphire (SOS) and silicon on glass or quartz are the only ones that have seen a level of commercial success.

Historically, SOS material has been available since the 1970s and has led to specialized commercial products. SOS technology requires a silicon epitaxial layer to be grown on sapphire. Defect density of the silicon is relatively high due to lattice mismatch, but decreases with silicon thickness. Electron mobility is significantly lower in SOS than bulk, due to the high defect density and compressive stress in the silicon caused by the lattice mismatch [2.1]. Silicon on sapphire has been used principally for defense and space (rad-hard) applications, where insensitivity to circuit latch-up due to radiation is important. The process is relatively expensive.

Silicon on quartz and silicon on glass are important areas of current development. Most modern processes for heteroepitaxial SOI are similar, and use bonding techniques similar to the majority of the silicon-insulator-silicon processes. This approach produces an active silicon region that is monocrystaline, so is electrically superior to polysilicon films grown directly on quartz or glass. Applications for silicon on glass include liquid crystal displays (LCDs), charge coupled device (CCD) detectors and high definition television (HDTV) projection systems.

One type of silicon-on-glass is termed "silicon on anything" (SOA) [2.2]. If a high resistivity substrate is chosen, SOA is particularly good for integrated inductors. A transparent substrate may be useful for opto-electronic devices. However, these do suffer from inferior thermal characteristics compared to bulk substrates. SOA starts as a fully processed SOI wafer. Additional processing is then required as follows: A glass plate is bonded on top of the fully processed circuits, then the original silicon substrate is etched back to the buried oxide layer. Bondpads are contacted by creating openings in the buried oxide. With this technology it is possible to fabricate low-power transistors. It is also possible to produce integrated inductors that have a quality factor greater than 25 at 1GHz [2.2, 2.3].

2.2 Silicon-Oxide-Silicon SOI substrates

While not exhaustive, this section considers most commercially viable and some experimental techniques for creating SOI. These are separable into two basic methods: (1) separation by implantation, and (2) wafer bonding.

2.2.1 Separation by Implantation of Oxygen

The Separation by Implantation of Oxygen (SIMOX) process is a technique to create SOI through the use of implantation. Most versions of this process use oxygen implantation. Implantation is performed before CMOS processing. High-energy implantation forces oxygen deep below the wafer surface (figure 2.1). The implanted dose to form a buried oxide layer is usually over 200 times higher than required for ion implantation doping [2.1]. After implantation, a 1300°C anneal of around 6 hours is performed, which forms a silicon dioxide layer and recrystalizes the epitaxial layer of any damage sustained from the high-energy implant. The oxygen typically forms an oxide region of thickness from about 50 to 500nm, with a thickness uniformity of +/- 2.5nm. The oxide region normally starts below a 40-70nm silicon region (for fully depleted material), and approximately 100-200nm for partially depleted material. In general terms the oxide thickness produced is categorized as; full dose, where the buried oxide (BOX) layer is about 380nm; medium dose, where the BOX layer is about 150nm; and low dose (approximately 100nmBOX thickness) [2.4].

Implant dose with O_2 ions is usually between 4.0 and 4.5el7 cm^2 at 200 keV. Several companies manufacture implanters aimed at the SOI market, including Hitachi [2.5] and Ibis [2.6], and numerous manufacturers produce SIMOX wafers.

Patterned Insulators

Secective patterning of the buried oxide of SIMOX permits mixed bulk and SOI on the same wafer. The body of a SOI device may be contacted from the

substrate, providing not only true SOI devices, but also structures that are bulk in nature. Selective patterning does leave a stress region within a few microns from the edge of the buried silicon dioxide layer. This must be taken into account in the layout for selectively patterned designs. Patterned insulators permit various combination structures that may otherwise not be practical. These include DRAM chips combined with SOI logic [2.7], and vertical bulk power devices combined with SOI circuitry. High performance logic may be fabricated in SOI, while the DRAM is fabricated on bulk areas. Selective patterning is shown pictorially in Figure 2.2.

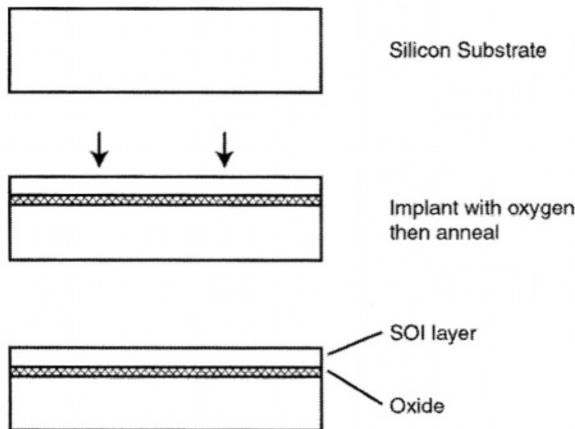


Figure 2.1: Representation of the preparation of SOI using SIMOX process. The starting material is a conventional silicon substrate. This is implanted with oxygen and annealed, to form an SOI layer over oxide, on a silicon support.

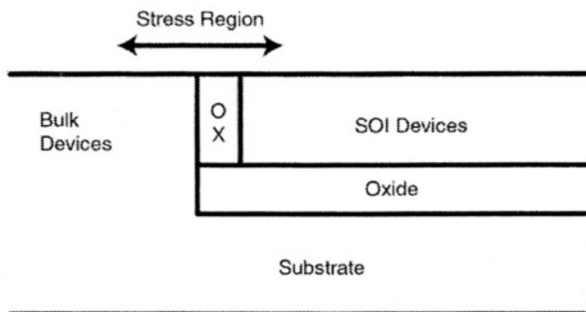


Figure 2.2: Selective SIMOX pattern, showing the possibility of combining bulk and SOI devices on a single substrate. The damage area around the edge of the buried oxide should be avoided for the purposes of building electrically active components.

2.2.2 Wafer-Bonding methods for preparing SOI

Silicon-oxide-silicon wafer technologies, with the exception of SIMOX and its variants, are based on wafer bonding. The basic process sequence for the fabrication of bonded SOI wafer is as follows: The active wafer (also called the donor wafer) is bonded to the support wafer (also called the substrate or handle wafer). In most cases both active and support wafers have been oxidized, although some processes do not require the support to be oxidized. The bonded pair is annealed to increase bond strength. Processing differs depending on the techniques employed, but in all cases the active wafer is reduced to the desired thickness after bonding. Several alternatives are possible to perform thinning, including grinding and chemical etching, plasma assisted chemical etching, bond and selective etch of porous silicon and implant-enhanced wafer splitting.

Physical Thinning and Etchback

The grind and etchback method of creating SOI is the most mature of the practical forms of wafer bonding based SOI techniques, but is least economical in terms of raw material. Two wafers are required, a support wafer and a wafer to contain the circuitry (the active wafer). Both support and active wafers are oxidized. The two wafers are bonded together (oxide to oxide), and the wafer designated the active wafer is ground, annealed and polished.

For the basic grind-back method the active wafer is ground and thinned down to the thickness required for the active layer. SOI quality is usually very good with this process technique [2.8], but thickness uniformity is poor. Uniformity of the active region depends on flatness of the substrate wafer. This is usually inadequate for use except in thick film SOI.

To bypass uniformity problems of grinding, etch-stop methods have been developed. In etch-stop processes a high concentration of an impurity is implanted into the active wafer, prior to bonding. The active wafer is ground and chemically etched using a selective etch. Boron [2.9] and nitrogen-implanted layers [2.10] are common impurities used as etch-stops to improve SOI layer uniformity. Nonetheless, uniformity remains limited since the concentration profile of the etch-stop dopant blurs during annealing. The crystalline quality of this type of material may also be degraded due the high dose implantation.

Figure 2.3 shows the method of preparation of ground and etch-back silicon on insulator. Bonding has the advantage over SIMOX SOI that the substrate can have a wide range of resistivity. Typically the substrate is moderately doped P-type, but there is no fundamental restriction. Higher resistivity

substrates may have beneficial properties in some applications, especially when the circuitry contains inductors, where higher resistivity substrates can lead to higher Q devices. Typically for bonded wafers, the oxide between substrate and active region is 400-500nm.

SmartCut®

SmartCut Technology from SOITEC [2.11] is a variant of the bonded technique. Here, however, the wafer to be used for the active silicon is implanted with hydrogen (protons) at approximately 1µm depth. Annealing creates a stress fracture. The active and support wafers are bonded, and the active wafer is split along the stress fracture to generate a Unibond® wafer. The unused portion of the active wafer can be reused (either as the support wafer or active wafer), resulting in a cost saving over the traditional ground-back method (described above). The bonded combination is the SOI wafer. It is annealed and polished in readiness for use. Figure 2.4 shows the method of preparation for SmartCut wafers.

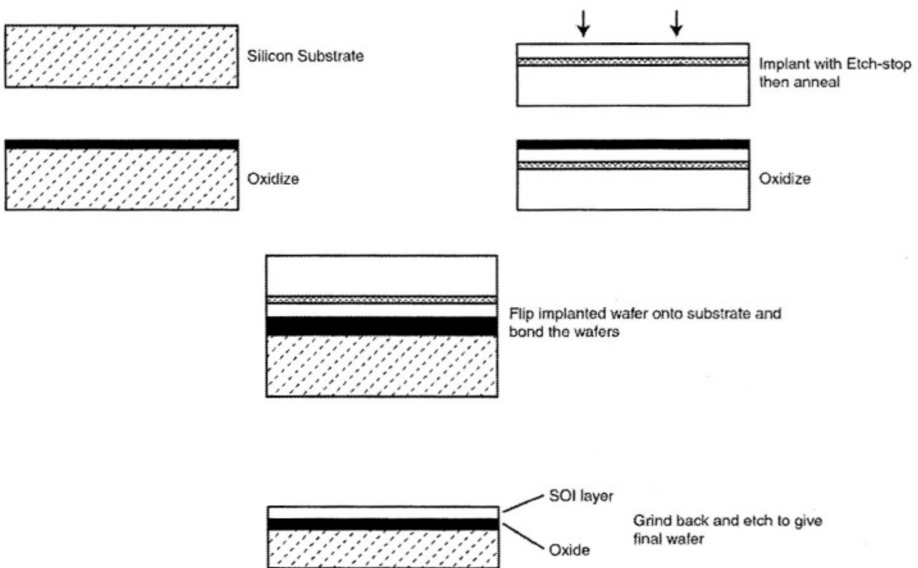


Figure 2.3: Process steps for the grind and etch-back processes for creating SOI material.

Eltran(tm)

Eltran (Epitaxial Layer TRANsfer) is another method of creating SOI wafers [2.12]. This process uses porous silicon to obtain a uniform SOI thickness. A region of about 12µm of porous silicon is formed by anodizing the surface of a boron- doped silicon wafer. An epitaxial silicon layer is grown on the

porous silicon. The layer is bonded to the oxidized substrate wafer. The porous silicon is split from the epitaxial material, using a water jet, and results in a re-usable split wafer. The porous silicon remaining on the epitaxial silicon is selectively etched away [2.13]. The quality of epitaxial silicon grown on porous silicon is poor due to stacking faults. However, stacking faults can be significantly reduced with a pre-bake in hydrogen before the epitaxy process. The material can be produced with an active layer thickness range between 50nm and 5μm, although the standard thickness ranges between 100nm and 200nm. The oxide thickness range available is also of the order of the same range. Surface roughness is comparable to bulk.

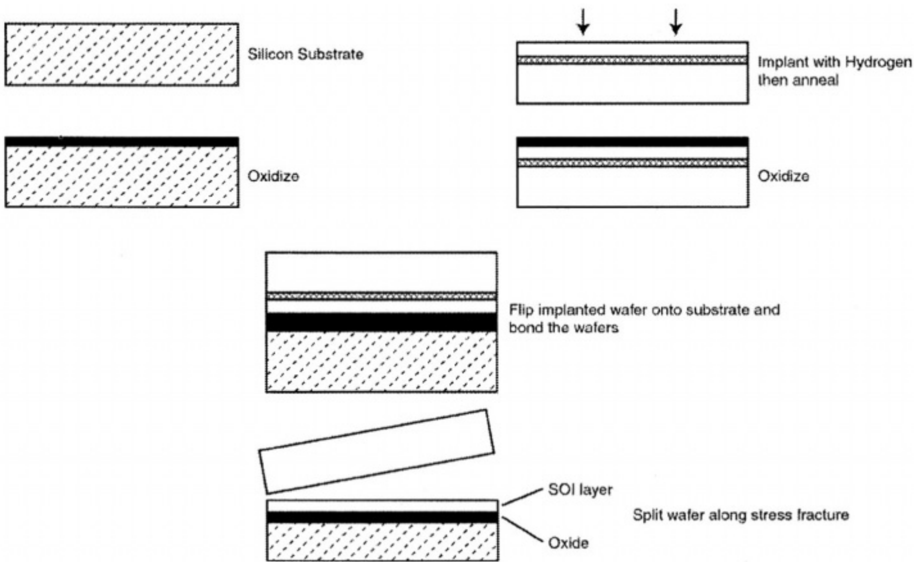


Figure 2.4: SOI wafer produced using the SmartCut® process. This process has the advantage of not wasting an entire wafer, as the split wafer can be re-used.

AcuThin

The AcuThin process uses a local dry etch process, using plasma assisted chemical etching (PACE) to obtain a uniform SOI layer. Measurements are made as etching proceeds, and non-uniformity is reduced through feedback of the local etch rates. With this technology better than 1% SOI thickness uniformity can be obtained [2.14].

The processing proceeds by bonding the substrate and active wafer. During processing the buried oxide thickness is determined by reflectance

measurement. Uniformity of the silicon on insulator layer is determined by the plasma characteristics and number of etching cycles. As the plasma area is reduced, and etching rate slowed, uniformity can be improved, at the expense of throughput.

Nanocleave

Nanocleave is an atomic layer cleaving process for bonded SOI [2.15], with a local surface roughness of less than the lattice constant of silicon (0.54nM). This SOI fabrication process is based on forming a cleave plane on an active wafer, using ion implantation. A bonded wafer pair is created with the active wafer. The bond strength of the donor-handle interface is higher than the cleave plane. The cleaving is done at room temperature using the force of a nitrogen jet. The resultant SOI wafer has an active silicon thickness within ±1% [2.16]. In this process the handle wafer can be silicon, glass or quartz, and the active wafer can have a layer of SiO_2, or be just silicon. The donor wafer is re-usable.

2.2.3 SOI Materials Summary

Various methods of producing SOI material have been presented. Some are more applicable to particular requirements than others. For example bonded techniques have the option of tailored resistivity for the support wafer, while the processing of SIMOX may provide a simple method to obtain mixed bulk, PD-SOI and FDSOI on the same wafer. For all methods there appear to be no fundamental restrictions on producing material up to or beyond 300mm wafers, and 200mm SOI is readily available. With several manufacturers now producing the material, it is becoming more readily available. As material costs reduce designers are considering SOI as a means to improve performance, without upgrading photolithographic equipment.

2.3 Comparison of SOI and Bulk

Designs in SOI benefit from reduced junction capacitance, reduced threshold voltage and improved I-drive. Threshold voltage reduction however is often at the expense of leakage. Process design usually accounts for this by matching worst-case leakage. This is usually considered to be at the strong process corner, high temperature and nominal supply voltage. Using this process corner for matching of the IDDQ leakage leads to I-drive gains throughout the operating range, while maintaining the maximum specification of leakage. It should be noted, however, that this method of device design normally results in a higher average leakage (figure 2.5).

2.3.1 Isolation Techniques

Bulk silicon processes use either junction isolation, or dielectric isolation. While both techniques are effective, junction isolation has the advantage of

low cost. However, it has the disadvantage of requiring a relatively large silicon area to isolate structures, since lateral diffusion of an isolation tank is about 70% of the vertical diffusion (figure 2.6). Sufficient vertical diffusion of a tank for a pMOS structure requires the tank to be diffused well below the deepest possible junction of the source and drain. For speciality MOS devices such as high voltage components diffusions may become substantial. With the reduction of feature sizes adding to the cost per square millimeter of silicon, the area lost to isolation becomes more significant with successive process nodes.

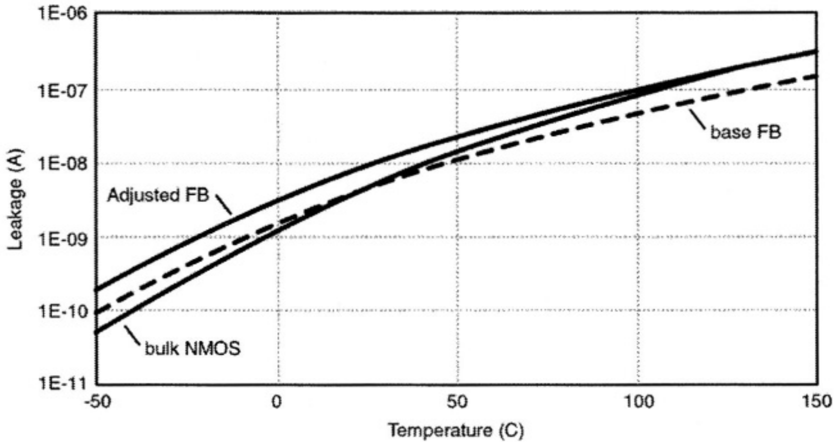


Figure 2.5: Drain-source leakage characteristics of SOI vs. Bulk. The bulk has a steeper leakage slope, so is normally lower leakage at room temperature and higher leakage at high temperature than its SOI counterpart. To maximize the performance enhancement of SOI it is common for the leakage of the SOI process to be adjusted, so it matches that of bulk at the high temperature limit.

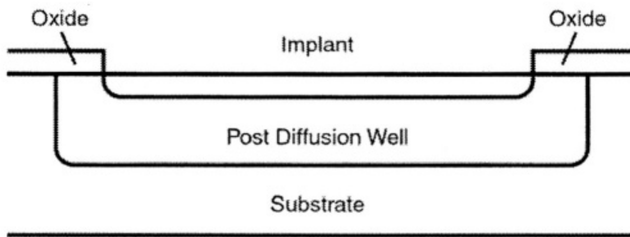


Figure 2.6: Diffusion of the Nwell or Pwell into which the MOS source and drain diffusions are made. For junction isolated schemes, the sideways diffusion of the dopant can add significantly to chip area.

In SOI, individual devices can be dielectrically isolated from each other using a trench-type isolation, resulting in isolation that is not susceptible to electrical latchup, and can be significantly more area efficient than junction isolation.

Figure 2.7 shows conventional bulk nMOS and pMOS structures in a junction isolated process. Figure 2.8 shows these structures in a dielectric isolated process, on a bulk substrate.

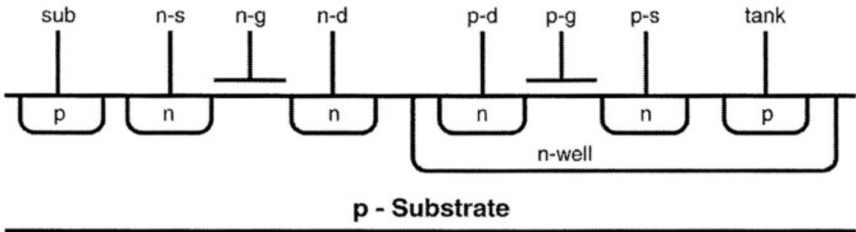


Figure 2.7: Simplified junction-isolated cMOS processed cross-section: The nMOS structure is generally constructed directly into the substrate or epitaxial material, and the pMOS normally is fabricated into an N-well diffused into the p-type silicon. The sources and drains are more heavily doped, shallower diffusions.

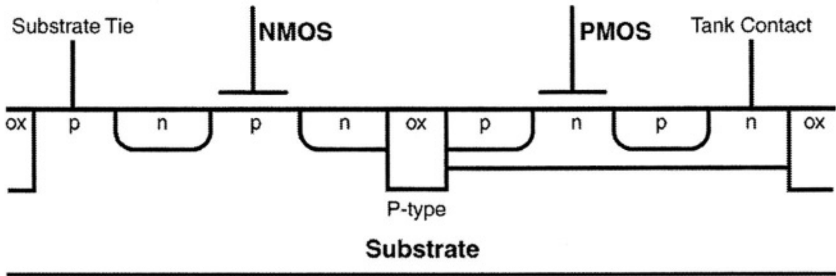


Figure 2.8: Simplified dielectric isolated CMOS process cross-section. Here the nMOS source/drains are diffused into a lightly doped P-type region, while the pMOS source drains are diffused into a lightly doped N-type region. Electrical isolation between structures is achieved with dielectrics, but the devices are electrically connected to the substrate either directly, or via a N-P junction diode.

The use of oxide isolation can result in a reduction of interconnect, source and drain capacitance, but body and local substrate contacts are still required. Figure 2.9 shows the basic SOI structure. This is similar to the oxide isolated bulk process, figure 2.8, except for the buried oxide or BOX

layer. The basic transistor changes from body tied (to ground for the nMOS, and usually to supply or source for the pMOS), to one where the body floats. This allows the parasitic bipolar device, which is in parallel with the MOS device, to be affected by switching conditions, allowing both the MOS and bipolar characteristics to modify device performance.

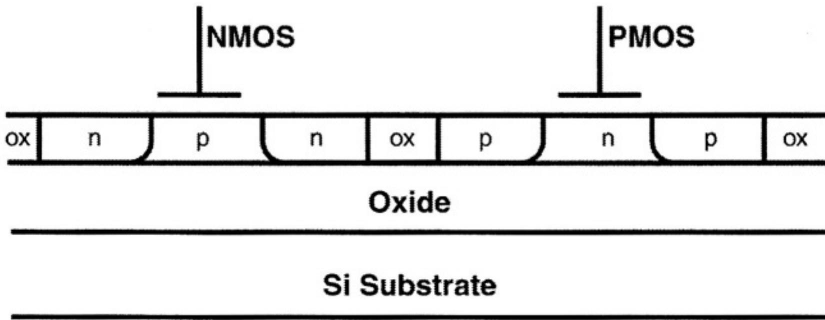


Figure 2.9: Simplified PD-SOI process cross-section: The nMOS source/drains are diffused into a lightly doped P-type region, and the pMOS source/drains are diffused into a lightly doped N-type region. Electrical isolation between structures is achieved with an oxide dielectric.

Source and Drain Doping Profiles.

The sources and drains of SOI devices have shallow junction depths compared to junctions in bulk material. The use of SOI results in a well-defined source/drain, which is relatively easy to maintain with technology scaling. This is not the case with bulk material, where junction scaling becomes more difficult as geometries reduce. Reduction in volume of the SOI source/drain does, however, tend to lead to higher resistance transistors than their bulk counterparts due to their shallower diffusions.

Partial Trench / Shallow Trench

There are several configurations that permit body ties through silicon. These are more area efficient than their metal connected counterparts, and may have advantages in reducing the effective body resistance, but normally require partial trench isolation and extra processing steps. A shallow, or partial, trench is achieved by adding an extra processing step that creates a trench in the active silicon which does not reach down to the buried oxide region. Their drawback is that the silicon under a shallow trench is usually fairly high impedance, thus not recommended for connections where low-resistance contacts are required. Partial trench may also be used to maintain thermal coupling between devices. The shallow or partial trench concept is shown in figure 2.10.

A degradation of the electrical characteristics of SOI devices with shallow trench isolation (STI), which depends on device size, is due to mobility decrease caused by interface roughness. The degradation is the result of plasma damage and can possibly be removed with a thermal oxidation [2.17].

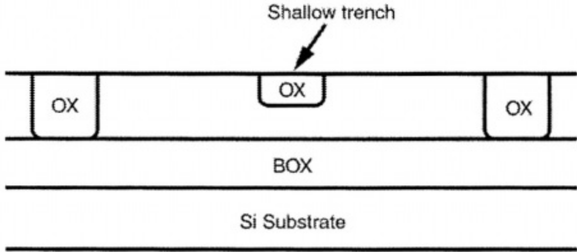


Figure 2.10: A shallow trench is created by etching part way through the active silicon and stopping before the buried oxide region is reached. This provides a resistive electrical contact between areas separated by the shallow trench, and a relatively good thermal contact between the regions.

Substrate Ties

Despite the active silicon being only capacitively coupled to substrate it remains normal practice to ground the substrate. This precaution prevents charge build-up on the substrate, which can affect transistor performance through the creation of a channel region in the body region, just above the buried oxide. A scheme to achieve a substrate tie in PD-SOI is shown in figure 2.11. In addition to connection to ground within the die, it is easy to connect the entire substrate to ground around the scribe line seal. This is area efficient since it is common practice to use a grounded metal ring at the outer periphery of many ICs.

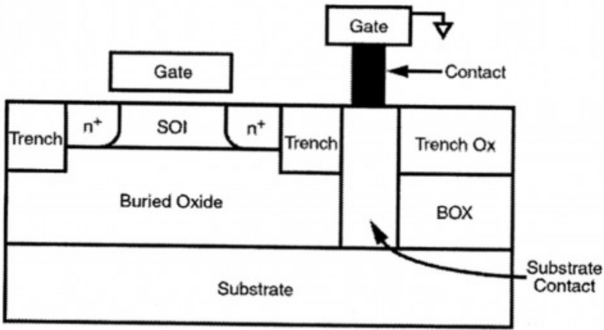


Figure 2.11: Method to tie the substrate to ground. This is achieved by etching through the active region and buried oxide. A substrate contact is included that extends from the substrate to the top of the active region.

2.4 SOI Technology Advantages

2.4.1 Capacitance Reduction

SOI devices have low drain / source to substrate capacitances, due to the buried oxide region, which replaces the junction capacitance of bulk material with oxide capacitance to substrate. Compared to some IC processes which use junction isolation, sidewall capacitance is also reduced. Mostly, reduction in source capacitance does not confer significant device performance improvement, although operating frequency improvements are observed in circuits where the source voltage varies, such as the upper nMOS device of a NAND gate, or when a source follower configuration is used.

Reduced capacitance of drain to substrate (or drain to tank in a pMOS) provides performance enhancement, although it does not provide a benefit in certain source-follower configurations. The reduction in source/drain capacitance usually provides incremental speed improvement, or conversely power reduction. However, reduced capacitance can also be a disadvantage. Supply noise is reduced in bulk material as a result of the large pMOS tank to substrate. This capacitance is not present in SOI material.

2.4.2 Reduced Short Channel Effect

Short channel effect increases as channel lengths are reduced. This compromises the threshold voltage and leads to increased device leakage. The threshold voltage roll-off decreases as substrate bias increases. Since the bodies of PD-SOI devices float, this tends to negate the short channel roll-off effect [2.18]. However, threshold voltage may be degraded in short channel floating body devices due to impact ionization [2.19]. The dominant mechanism depends on process, layout and operating conditions. PD-SOI can utilize many techniques used in bulk CMOS for controlling short channel effect, such as retrograde wells and pocket or halo implants. Controlling short-channel effects is a challenge as device channel lengths are scaled below 100nm. For bulk silicon, the method used to contain short channel effects is through careful design of the dopant profiles [2.20 - 2.22].

2.4.3 Lower Vt Device Threshold

Unless adjustments to processing are made, SOI features lower threshold voltage (Vt) compared to bulk. It is therefore normal practice for threshold adjust ion implants to be modified to achieve a similar threshold voltage for PD-SOI and bulk, to maintain similar drain-source leakage levels. Without threshold voltage correction drain-source leakage may become excessive, especially at high temperatures.

2.4.4 Soft Error Rate (SER) Effects

One benefit recognized early in the development of SOI has been reduction in memory soft-error rate resulting from cosmic rays and background radiation. SER is a greater concern in small area memory cells, where there is less charge to disrupt. With the sizes of all types of memory cells reducing at a rate close to that of process feature size reductions, any advantage that can be gained from reduced SER is significant. In addition, soft error immunity (radiation hardness) of SOI devices is important for high data rate network servers and global data transfer links [2.23]. SER is discussed in more detail in chapters 8 and 9.

2.5 Performance

There continues to be considerable debate surrounding the benefits of SOI, and questions of scalability. This is largely due to difficulty in obtaining direct comparisons between SOI and bulk. The floating body gives different levels of benefit depending on the circuit being tested, operating conditions such as supply voltage and temperature, and history of the signals to the circuitry being compared. In addition, performance also varies depends on the comparison point of device leakage. If circuit leakage is normalized to nominal operating conditions the result is very different to circuit leakage being normalized to the high temperature, high leakage corner. However, SOI appears to provide better performance than bulk material in virtually all cases, so additional processing costs aside, SOI appears to have a future.

When compared to bulk, MOS devices fabricated on SOI have lower capacitances from source to substrate and drain to substrate, but with many devices, interconnect capacitance contributes significantly to total capacitance. Therefore reductions in drain and source capacitance observed in the move to SOI are not as significant as initial appearances might suggest. Circuits exploiting the reduced reverse body effect in SOI obtain approximately 15% performance enhancement compared to bulk, but added design margin required due to history effect may take away much of the advantage, unless care is exercised in design. Differences between SOI and bulk are given in table 2.1. Metal and poly-silicon etches may need re-tuning between bulk and SOI processes due to different etch rates on SOI and bulk. The yield of bulk silicon is the same as that of bonded wafer SOI and SIMOX material.

Gate lengths of planar CMOS transistors on bulk silicon are not expected to reduce much beyond 50nm [2.16]. According to some reports a 20-50% increase in transistor speed can be achieved by switching from bulk material to SOI [2.24, 2.25]. This improvement represents one to two generations of scaling, without the cost of higher resolution lithography [2.16].

Circuits built on SOI are more versatile than those on bulk because the buried oxide can be used to optimize scaling [2.26]. As devices are scaled source/drain capacitances begin to dominate gate capacitance. However, as this occurs interconnect capacitance becomes more significant in defining overall chip speed, moving the focus towards improved drive currents and a combination of reduced interconnect capacitance and resistance as the most important performance drivers. Library cell layout can be between 10 and 30% more compact with SOI because nMOS and pMOS devices may be randomly placed, since there are no dielectric isolation requirements for SOI.

Table 2.1: Principal benefits and disadvantages of SOI vs. bulk material

Principal Benefits of SOI:

- Reduced drain and source capacitance, resulting from oxide isolation rather than junction isolation techniques.
- Reduction in the "MOS reverse body effect" in stacked transistor devices.
- Performance enhancement due to floating body effects, particularly in stacked components such as the upper nMOS devices in NAND gates, where a floating body replaces a body biased at lower than the source voltage.
- Marginally reduced device area may lead to a reduction in metalization capacitance, and slightly reduced die area, and layout of library cells are more compact with SOI because nMOS and pMOS devices can be randomly placed.
- Improved latchup, noise and current immunity through the substrate
- Improved high temperature performance (lower device and parasitic leakage)
- Improvements to passive components – Inductor Q improvement, parasitic capacitance reduction
- Reduced capacitance from substrate to metal interconnect.

The primary disadvantages of SOI are:

- Self heating
- Reduced supply to ground capacitance for noise reduction on the supply rail.
- Undesirable kink effect (chapter 10) in analog applications
- Floating body can lead to increased drain-source leakage (transient and DC)
- Floating body can lead to high or excessive gate to body voltages, with potential reliability issues
- History effect complicates design, and requires greater design margin in many cases.

2.6 Partially and Fully Depleted-SOI

The depletion region of partially Depleted (PD) SOI does not reach through the entire silicon channel/body region. PD material usually has a silicon thickness greater than 0.15μm. Claimed performance benefits of using PD-SOI compared to bulk range from 7-11% to 20-35% over similar bulk

material [2.27]. Actual comparison is difficult due to the non-exact equivalence of bulk to SOI in terms of optimum transistor design.

In the early 1990's SOI was hampered by poor material quality. SOI still has to be measured against the continuing scaling achieved in bulk. Over the last few years SOI material quality has improved and bulk scaling appears to be reaching a point where improvements can only be made at greater costs. This makes the improvements gained from the use of SOI relatively more cost effective. PD-SOI is easier to manufacture than fully Depleted (FD) SOI due to its relatively thicker substrate [2.28].

PD SOI devices display the 'kink effect', a circuit behavior not present in bulk technology, which also reduces further as the active region reduces, to the extent that it can be eliminated in FD-SOI. The kink effect is caused as a result of the floating body. It has the beneficial effect of higher drive current for digital designs (chapter 3). The floating body also gives rise to "history dependence", causing different switching rates on subsequent switching edges. In PD silicon, "pass-gate leakage" can occur, and is a result of the body voltage floating to a point where the parallel, parasitic bipolar transistor associated with most PD MOS devices, can turn on.

Partially Depleted SOI is often considered to be an evolutionary advance in process development, as the topology is essentially the same as bulk, with the exception that bulk does not have the oxide layer. In many analog and transmission-gate logic applications floating body effects are unacceptable. In these cases a body contact can be added to PD-SOI, although body contacts are resistive, and may not be very efficient at high frequencies. It is not realistic to add a body contact to FD-SOI, since there is no guarantee of a body region being available under most operating conditions.

The body potential of a floating body PD-SOI device depends on switching history, circuit topology, component, chip temperature and supply voltage [2.29]. This may vary typically by approximately 8% faster or slower than a nominal state. As a minimum, test routines are required to force the body voltages to the extreme delays. This approach however, can cause over-compensation, resulting in excessive margin. Minimization of design margin is discussed in chapter 6.

Variations in the floating body voltage of PD-SOI MOS transistors causes an uncertainty in the gate-source threshold voltage. This in turn reduces the noise margins for dynamic and memory circuits [2.29]. As noise margin is reduced, additional attention must be paid to dynamic circuit and memory design. Predischarging can be used to force floating nodes into a known

state, to improve noise margin. This discharges body nodes of indeterminate voltage (chapter 10).

The use of fully depleted silicon on insulator material has been proposed as a way to minimize floating body effects. For FD devices, the SOI film is far thinner than the device depletion width (a few nanometers to a few tens of nanometers). Thus there is no body region of the MOS device that can be charged. With no floating body, FD does not exhibit significant kink or history effects.

It may be assumed that fully depleted MOS transistors lend themselves more readily to analog implementations, as they do not have the kink and history effects of their PD-SOI counterparts. However, PD-SOI is more difficult to manufacture, and threshold voltage control, together with matching difficulties may occur due to 2-D field fringing. Devices on FD-SOI display a steep sub-threshold slope, with low parasitic capacitances, making them appropriate for low-voltage, low-power applications. A significant advantage of FD circuits is that for the same bandwidth they typically consume only about half the power of their bulk counterparts, alternatively, bandwidth may be improved a given power dissipation. Table 2.2 shows differences in PD and FD-SOI characteristics.

Table 2.2: Comparison of the component differences between Partially Depleted-SOI and Fully Depleted-SOI

	PD-SOI	**FD-SOI**
Floating body	yes	no *
Active Silicon Thickness	>0.15μm	<0.1μm
Kink effect	yes**	no
History effect	yes	no
Source/Drain resistance	moderate	high
Manufacturing	reasonable	difficult
Power	low	very low

*Floating body is not present in the saturation region of operation
**Kink effect is suppressed in some operational states

Many device and circuit performance related limitations are postponed or eliminated for CMOS fabricated on very thin SOI. In bulk material the control of very shallow MOS source/drain junctions is difficult due to limited thermal budgets [2.30], but with SOI material junctions are determined by the thickness of the active silicon [2.16].

The term 'Ultra-thin' film SOI is usually used to refer to material with an active silicon thickness in the range of 20-50nm, although functional transistors on an active silicon thickness of just 1nm have been demonstrated [2.26]. In such devices, thickness-related effects on carrier mobility and threshold voltage become important. At thicknesses of 1nm, devices are quantum mechanical in nature, and the structures behave as quantum wells (chapter 13).

Here we consider mostly PD-SOI circuits, rather than FD-SOI, due to the clearer applicability of PD-SOI at this time, and also the particular challenges of designing with PD-SOI. PD-SOI will likely dominate through the 50 to 70nm gate length generation, and will therefore be used by mixed signal designers for a number of years. FD-SOI may follow with two different gate electrodes. This is because the mobility of double gate devices is close to 4x that of single gate structures as a result of electric field effects. Although mostly described with an emphasis on PD-SOI, many of the process effects, thermal effects and some of the circuit techniques, are applicable to FD-SOI. FD-SOI gives more improvement in circuit speed and soft-error immunity than PD-SOI [2.16], and also permits a reduction in power requirement.

Thick film SOI (SOI material that has an active region that is thicker than about one micron) is used mostly for power integrated circuit applications, RF bipolar applications, MEMS and opto-electronic devices. Floating body CMOS devices designed for thick film SOI exhibit similar floating body effects to PD-SOI devices, although due to the additional thickness, the parasitic bipolar is less affected by depletion caused by the source/gate/drain of the MOS structure.

2.7 Technology Scaling

Scaling device dimensions has been used as a fallback method to improve device performance. It has proved highly successful as patterning technology has improved ahead of the processing requirements. However, as simple scaling has proved more difficult to achieve, additional process improvements have been required to maintain the trend of matching the Moore's Law curve. This includes combinations of the best improvements from all areas of IC production [2.31].

SOI processes also make use of the process improvements that have occurred during the last few years, including high-performance CMOS logic technology where gate lengths are measured in the tens of nanometers, together with multi-level copper [2.32] and dual-damascene interconnects with low dielectric constant (low-k) intermetallic material [2.33].

When high velocity carriers flow from source to drain, impact ionization induced electron-hole pairs are generated. The threshold that impact ionization occurs depends on many device characteristics, including gate oxide thickness, doping profile and supply voltage. Impact ionization causes hole charging of the body, which reduces with gate length scaling. However, while impact ionization reduces, gate leakage increases due to the thinner gate oxides, which are required for the shorter channel lengths. One solution to reduce gate leakage is to use high dielectric constant (high-K) gate dielectrics, though this has limitations. Parasitic bipolars can be tuned to minimize device gain. This is achieved through a combination of source doping and channel doping to reduce emitter efficiency and gain. There is general agreement that SOI provides an improvement in terms of operating frequency for digital applications over bulk at technologies down to gate lengths of between 50 and 100nm. However, there has been some debate on the level of scalability of PD-SOI with future technology generations [2.34 - 2.36]. This stems from the reduced role of junction capacitance in the obtainable improvements, and the possibility of increased history effect with device scaling. (Increased history effect leads to the possibility of over-designing to guarantee circuit functionality and hence a loss of much of the expected SOI performance improvement).

2.8 SOI Device Properties

2.8.1 Source/Drain-to-Substrate Capacitance

Devices built on SOI have lower source-to-substrate and drain-to-substrate capacitances than bulk silicon [2.37], due to the oxide layer between source/drain and substrate. These junctions are not diode-like in behavior in SOI, which eliminates the possibility of forward biasing the node to substrate or to the tank (in the case of the pMOS device). High-resistivity SOI substrates, used for high frequency applications, further reduces these capacitances. At low to moderate frequencies capacitances to substrate may not need to be included in simulations, but at higher frequencies, above 500MHz to 1GHz, specific substrate capacitance modeling is usually required [2.38].

2.8.2 Gate Leakage

Along with shorter gate lengths there is a trend to reduced gate oxide thickness. This results in increased gate leakage, due to tunneling. When thin gate oxides are combined with SOI the body voltage is affected in some conditions. Leakage may not cause significant effects in simple inverter circuits, since when a large potential is present across the gate and body of a device, the device is usually in the 'on' state. However, there may be some effects on stacked stages, such as the upper MOS of a stacked scheme, as

found in NAND or NOR gates, and also in transmission gates (chapters 6 and 7).

The gate leakage mechanism is described by tunneling [2.39], providing a leakage that is not greatly affected by temperature. Thus there may be little noticeable effect on circuit operation of relatively gate-leakage-tolerant circuits at room temperature or higher. The impact at lower temperatures, however, becomes significant, when gate leakage may dominate other sources of leakage that do scale with temperature. This is of particular concern in such circuits as loadless four-transistor (4T) SRAM cells, where gate leakage contributes to a significant proportion of the overall leakage. Gate leakage is an important consideration in analog designs where charge storage is required (chapter 10). Gate leakage is usually low enough that static logic operation is unaffected for bulk designs, but in many other circuits gate leakage may be problematic, and requires consideration when transferring a bulk design to SOI.

Gate leakage depends on the voltage gradient across the oxide. For nMOS devices gate leakage reduces the body potential if Vgb is below zero, and increases it if the Vgb is above zero. In terms of switching sequence for a nMOS device in an inverter, the starting condition can be with the gate at either zero volts or the supply voltage. If the starting condition is zero volts on the input, then the body of the nMOS will drift to a few tenths of a volt, as a result of the drain-to-body and body-to-source leakage. Under these conditions tunneling between gate and body is very low, and has a minimal effect on the starting condition.

If the starting condition is an input state of Vdd, the effect of gate leakage is to increase body voltage from the expected value of 0V, to a value that can be several tenths of a volt above ground [2.40]. As switching occurs, drain capacitance tends to dominate the switching effects on the body, raising the voltage still further. The overall effect is an increase in history effect, together with a marginal increase in frequency.

2.9 Body Effects

2.9.1 Bipolar Effect

Figure 2.12 shows the bipolar structure of a PD-nMOS. The base of the parasitic NPN is not connected, but capacitive switching and leakage effects can combine to increase the base voltage sufficiently that the structure could turn on. An example of NPN turn-on is readily seen when the nMOS source and drain are high (e.g. in a transmission gate), the body drifts high (neglecting gate leakage), until source, drain and body are at the same

potential. Assuming that the gate remains low during this time the circuit configuration is in the 'off' state, and does not draw current. If the source is then pulled low, the base is pulled down due to the base-emitter (body-source) diode turning on. The body voltage will initially be at about 0.7V higher than the source. The result is that the parasitic NPN will turn on for a transient period, until carriers are swept from the base region. This can take many picoseconds, during which time current is being drawn from the drain (collector) region. It is important to consider this effect in several circuit configurations, and circuit designs have been proposed to address this effect (chapters 6 and 7).

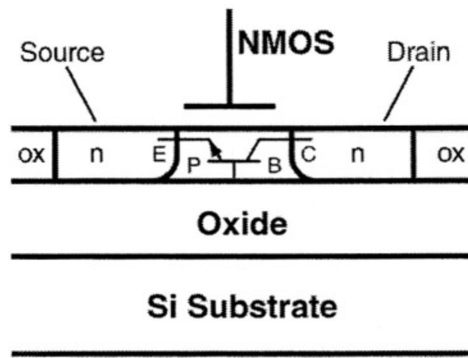


Figure 2.12: The parasitic bipolar device of a PD-nMOS. Unlike in bulk silicon, the body of the bipolar is floating, and can be turned on through capacitive coupling.

When power is supplied, or when there is either input or output state change to a PD-SOI component, the body voltage can be perturbed by capacitive coupling. The time taken for the body to settle back to its steady state value ranges from a few microseconds, to a few milliseconds. During this time Vt and Idrive can change. This can seriously impact digital and analog circuits, and can also significantly increase the initial parasitic leakage current (although once steady state body bias is reached, leakage current will settle to the expected value). This is discussed further in chapters 6 and 10.

2.9.2 Kink Effect

The "Kink Effect" is observed once impact ionization begins. This effect is the result of an abrupt saturation current increase in the strong inversion condition. It is mainly caused by an increase in body potential due to carrier generation caused by impact ionization. The rise of body potential is accentuated in thin depletion layers because the lateral field is stronger and the recombination region is restricted [2.41]. The effects of this phenomenon can be used to advantage in digital designs to improve performance. The

kink in the output current curve (figure 2.13) shows where the drain to source leakage current is boosted, and is observed in PD-SOI, but not in bulk devices. It is detrimental to the design of analog circuits, and most analog designs avoid operation of SOI devices in the region where the kink effect occurs.

Figure 2.13: Graph showing the kink effect. As source-drain voltage is increased, the output current passes through the impact ionization threshold region, and output current is boosted.

As operating frequency increases the kink effect is suppressed [2.42]. It has been suggested that this may be beneficial when designing RF circuits for SOI applications [2.43]. Other work suggests that for optimum PD-SOI device performance in RF applications the use of body-source tied devices is necessary for lowest noise [2.44], and that parasitic bipolar contributions should be minimized.

The large output swings and matching requirements of high speed operational amplifiers, analog-to-digital converters, and a variety of other analog circuits still necessitate the removal of the I-V kink. In some cases body contact structures are used, but body contacts are not ideal due to the high resistance of the body region and resistance variability with body bias. Other circuits make use of circuit biasing to operate outside the kink region. The low-frequency noise overshoot has been correlated to frequency dependence of the kink-related noise in SOI FETs, even when the devices are body-tied [2.45].

The kink effect observed in PD-SOI can be minimized through process optimization (minimizing the bipolar gain and leakage), and/or by layout design (such as adding body ties). Body ties can be source or ground (or supply) connected. Body to source ties have specific layout requirements, to maintain low resistance, for example 5μm maximum between body ties. For

body tied structures it is normal, though not universal, practice to contact both ends of the body region (figure 2.14). This is because the double contact effectively reduces the body resistance by a factor of four compared to a single resistive contact at one end of the body.

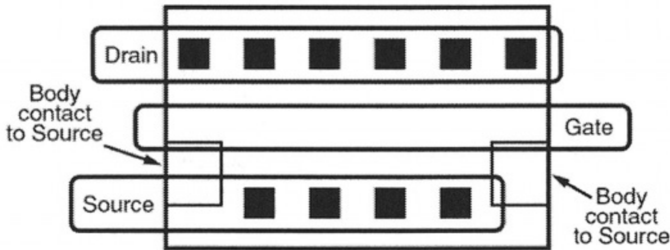


Figure 2.14: Body contacted nMOS device, showing typical body contacting sites. Maximum spacing between body contacts is determined from the body characteristics and MOS transistor leakages.

At drain voltages, a secondary kink can occur, attributable to the lateral parasitic bipolar device turn-on. This occurs at the point when the body is sufficiently forward biased to allow the bipolar device to turn on. At this point the parasitic and MOS devices are conducting in parallel. This becomes less significant as supply voltages reduce below 1.5V.

2.9.3 Capacitive Body Effects

Perhaps the single most significant feature of MOS devices on SOI, compared to bulk, is that the SOI body is connected electrically to the substrate through a capacitance (insulator of the silicon-on-insulator). The junction in bulk material is capacitive in negative bias, and behaves as a diode with positive bias. Studies have compared circuits and devices on bulk to fully and partially depleted SOI [2.46-2.48]. In bulk silicon the nMOS body is typically substrate, and therefore directly connected to ground. The pMOS is normally built in an N-type well, which makes a PN junction to substrate. The construction of MOS devices in bulk silicon can lead to the production of a highly efficient PNP structure to substrate, consisting of the pMOS source or drain, N-type tank and P-substrate. This PNP must be maintained in the 'off' state to prevent excessive supply-to-substrate leakage. Figure 2.15 shows the parasitic transistor of bulk integrated circuits. The floating body in SOI is capacitively coupled to the substrate. The body voltage of a SOI device is determined by rapid transient switching of the source, drain and gate, and their associated leakages. Capacitive coupling between gate and drain to the floating body leads to the so-called 'history' effect, where the response to device switching is dependent upon earlier switching events.

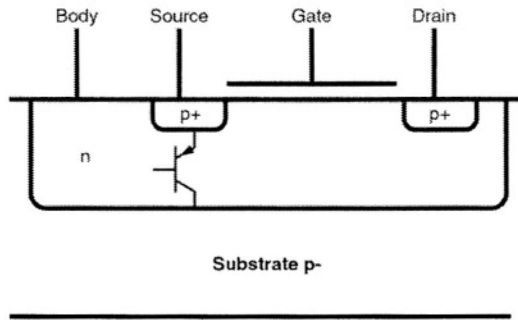


Figure 2.15: A parasitic transistor is created in bulk ICs by a pMOS source or drain, the N type diffusion for the pMOS well acts as the PNP base and the substrate is the collector.

With any type of circuitry converted from one technology to another, even when it is merely from one bulk process node to the next, the most challenging aspect of the transfer is a complete understanding and characterization of the new circuit behavior. This is generally achieved through modeling and silicon analysis. With conventional redesign, process adjustment is largely limited to the various advances made resulting from shorter channel length devices, such as Idrive improvements, changes to parasitics, and supply voltage reduction. The technology transfer from bulk to SOI is vastly more challenging, largely due to the floating body behavior of SOI.

Figure 2.16 shows the circuit elements that contribute to the body voltage of a device built in SOI. The DC component of body voltage is determined by the p-n diode leakages of body to source and body to drain. The body to drain current includes impact ionization current and gate induced drain leakage. In addition, there is the possibility of gate leakage current resulting from tunneling through gate oxide. The AC component of body voltage is determined by capacitive coupling during switching, in particular the gate to body coupling (Cgd) and drain to body coupling (Cdb). The Source to body capacitance (Csb) attenuates some of the body coupling in most circuits. Time constants for the leakage currents can be in the order of milliseconds, whereas response time of capacitive coupling is instantaneous.

Gate thickness continues to be scaled with technology generations. Thin gate dielectric is important in maintaining performance at reduced supply voltages and in controlling short-channel effects. With very thin dielectrics, especially at reduced temperatures, gate tunneling current dominates over other leakage mechanisms. Part of the tunneling current is between body and gate, affecting body voltage [2.40].

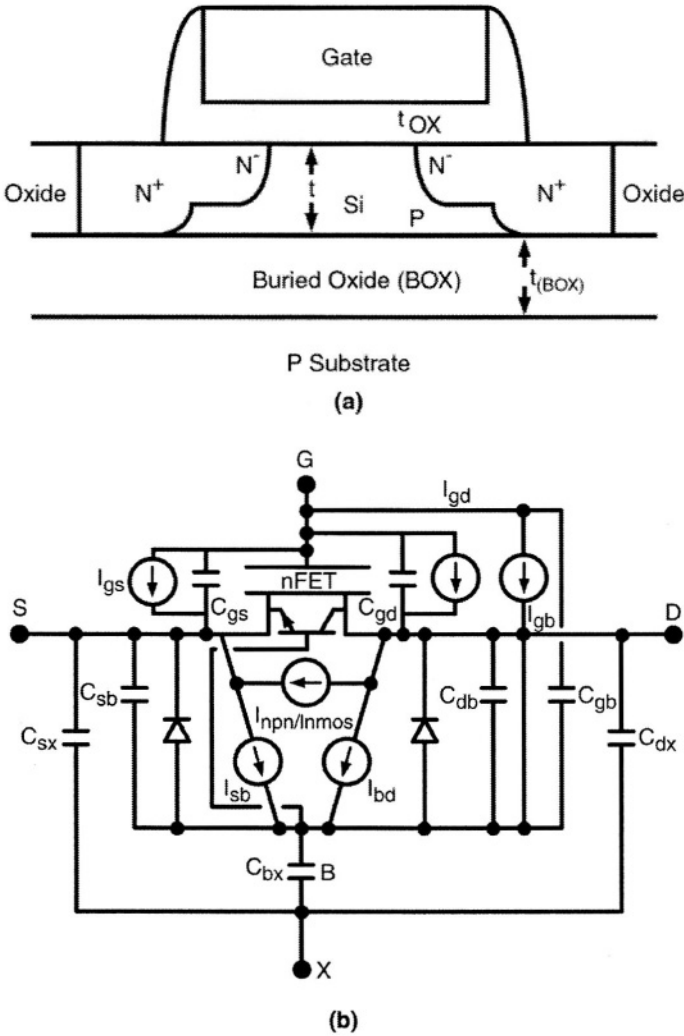


Figure 2.16: Cross section and equivalent parasitic components associated with the nMOS structure in PD-SOI. Note the body voltage is determined from DC leakage to drain, source and gate, and AC coupling to gate, source, drain and substrate.

Floating-body effects depend strongly on supply voltage. The baseline (DC) body voltage depends on leakage and the AC perturbation of the DC value is capacitance driven. Thus as supply voltage is reduced, floating body effects are also reduced. The kink effect gives rise to reduced MOS resistance at high Vds conditions, improving the efficiency of the structure.

A significant 'floating body' effect is the MOS device's dependence on past switching events. History dependence (or history effect) is the effect of switching history on gate delay [2.49]. As the input pulse train changes, delay through the gate changes. In the DC state the body voltage is determined by leakage through the drain, source and gate, and by impact ionization.

2.10 Body Ties

Additional variation of the threshold voltage in PD-SOI (compared to bulk CMOS) is attributable to floating body effects. There are instances where effects caused by the floating body cannot be tolerated, particularly in circuits previously been used for designs on bulk silicon. In other cases, however, such variation is merely a minor inconvenience or may even be beneficial. In most designs, especially those specifically for SOI, the variation of threshold voltage can be used to advantage, permitting faster turn-on for circuits, without increasing leakage current.

Body ties [2.50] can be achieved in a number of ways and applied to a variety of applications [2.51]. They have been designed to minimize or eliminate floating-body effects and effectively suppress DC effects. However, they are not as effective as devices constructed using bulk material in eliminating body transients in AC conditions, since the body charge and discharge time may have a time constant much longer than the transient. In order to verify effective body tie design, dynamic charging due to the body resistance and capacitance must be included in simulations. Depending on the type of circuitry and frequency of operation, this can be as simple as adding a lumped equivalent for body and gate resistance, or can require more detailed simulation of the resistive and capacitive effects.

For some analog applications it is permissible to use more complicated or expensive device engineering mechanisms [2.52]. The body tie has some advantages, but carries a die area penalty [2.53]. It is effective under DC conditions, but body potential is affected by fast switching transients, and is not well regulated at high frequencies, where gate, source and drain capacitances form a distributed resistor-capacitor network to the body contact, allowing the body to be perturbed by the switching MOS device. In many analog designs threshold voltages must be closely matched. This can be achieved through the use of body ties [2.54] which may be:

(1) tied to source (figure 2.17(a)),
(2) tied to ground (figure 2.17(b)), or
(3) coupled to another device that requires matching [2.54] (figure 2.17(c)).

In addition transient techniques can be used such as precharging or predischarging.

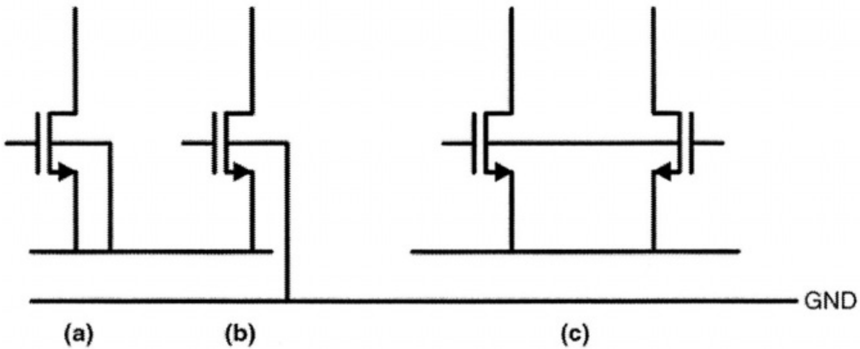


Figure 2.17: (a) Body tied to source configuration. This is the most area efficient form of body tie for most SOI. (b) Body tie to ground. This matches the body of most bulk nMOS structures. (c) Body tied to body - used where two bodies need to be maintained at the same potential but may float.

Body contacts provide a means to maintain the body at a known potential, and allow modulation of the body voltage. Body modulation can be used to improve performance through the application of a voltage to the body that is closer to the parasitic bipolar turn on potential. Positive body modulation reduces device threshold voltage, which can be a useful circuit design technique. In addition to using body biasing for improved performance, the body can be biased to provide lower IDDQ or quiescent currents, biasing the body to a more negative potential than the source (for nMOS), or to a more positive potential than the source (for pMOS).

The resistance along the body is normally tens of kilo-ohms per micron width. Depending on processing, the body resistance may not reduce much with increased gate length, due to the body diffusion profiling. If the body tie is to source, the capacitance between drain and body must be considered in the overall efficiency of the body tie, although the parallel body to source capacitance will help stabilize body voltage. If the body tie is independent of the source, both body to drain and body to source capacitances should be included in the calculation of body effect. There is also body to gate capacitance to be considered when the device is off, and body to channel effect when the device is on.

In general capacitances affecting the body voltage vary with process. It is desirable to include the distributed effects in simulations, to determine the effectiveness of particular body tied configurations. Figure 2.18 shows the

resistance and capacitance coupling of a body tied to source, and figure 2.19 shows the resistance and capacitance coupling of an isolated body tie.

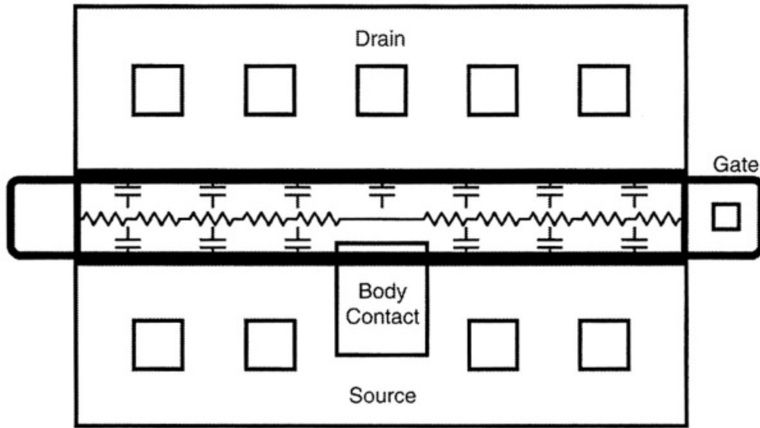


Figure 2.18: Representation of body resistance and capacitance of a body to source tied body contact (BC). The body resistance is many kiloohms per micron gate width, thus at high frequencies, or during fast transients the local body voltage may be perturbed from that of the body contact.

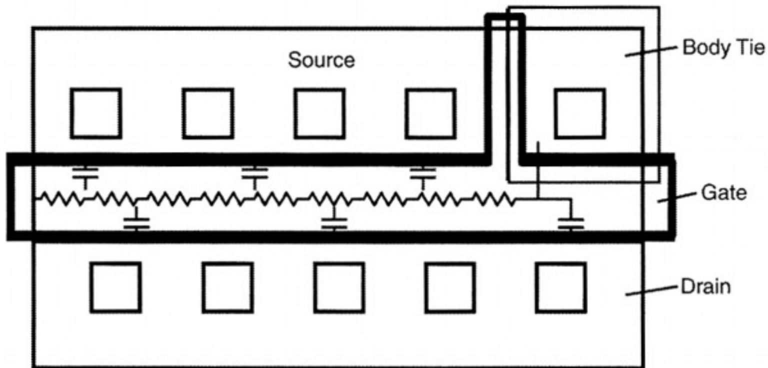


Figure 2.19: Representation of body resistance and capacitance of an independent body tied contact. The distributed body resistance causes local body voltage to vary from the contact potential.

2.10.1 Body Tied to Substrate

An area efficient variant of the body contact is the Low-Barrier Body-Contact (LBBC) Scheme [2.55]. Here a low barrier body contact is created under the source or drain region, to facilitate collection and removal of current carriers generated by impact ionization. Reduced flicker noise (compared to floating body devices) is claimed, making them useful for

analog applications. The scheme is more efficient for substrate current collection compared than conventional SOI contact schemes, but may be somewhat more difficult to implement.

2.10.2 Body-tied to gate Configurations

The body-tied-to-gate transistor has been proposed for a number of designs, especially high-speed digital applications [2.56]. This configuration has the effect of increasing drive current in some cases, but it has disadvantages, the most significant being that the gate voltage is limited by the parasitic bipolar transistor. As gate voltage is increased, the base-emitter (body-source) voltage becomes forward biased, limiting the voltage that can be applied between gate and source (figure 2.20).

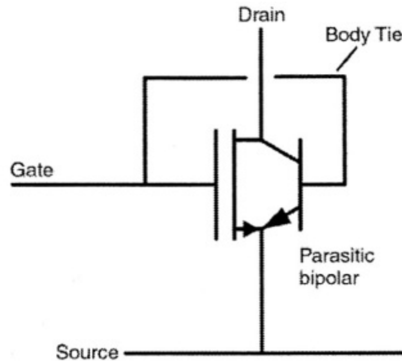


Figure 2.20: Base-emitter forward biasing caused by a direct gate-to-body tie. This can be alleviated with the use of a current controlled connection, or a voltage level shifter.

A further limitation of this configuration is that the circuit driving the gate-body tied input sees a higher capacitance than if it were just driving a gate (figure 2.21). The conventional gate configuration of a nMOS sees capacitance of the gate to source (GS), gate to drain (GD) and gate to body (GB), initially, and after reaching the threshold voltage this will become GS + GD + gate-to-channel (GC) capacitance. The gate to body tied device however, sees GS + GD + body to source (BS) capacitance + body to drain (BD) capacitance, in addition to the smaller body to substrate (BSU) capacitance. As gate voltage is increased beyond the threshold voltage, the input capacitance becomes GS + GD + GC + BS + BD + body to channel (BC) capacitance + BSU. Hence the body tied to gate configuration becomes viable if gate voltage does not change significantly, or if the gain in output current (from the similarly constructed previous stage) is higher than the extra capacitance that needs to be driven.

An additional issue is that the body resistance changes with gate bias: As gate bias increases the depletion region increases. Under these conditions body resistance becomes higher than the condition where the gate is at the source voltage.

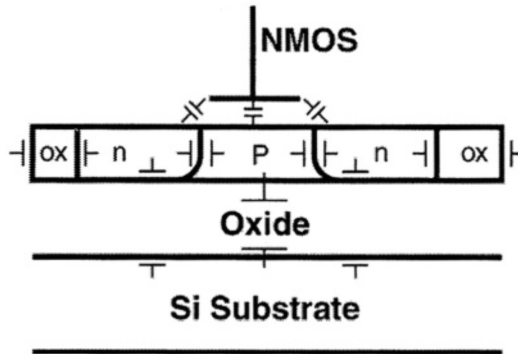


Figure 2.21: Parasitic capacitances associated with the gate and body of an SOI device.

2.10.3 Resistive Body Tie

A variant of the standard body ties the body resistively to a bias potential (often source or ground). This gives a DC state that returns relatively quickly to the source or ground voltage, to avoid high off-state leakage, but permits AC switching to modulate the body voltage for improved transient switching performance. A layout option for this is shown in figure 2.22. As distance from the body contact, along the width of the transistor increases, coupling reduces, allowing capacitive switching to more easily perturb the body voltage. In principle this may offer some advantages of improved operating Idrive, while minimizing off-state leakage. However, if the body resistance is particularly high, impact ionization can pull the body voltage up, even with just a DC bias. This increases leakage, and causes a low threshold voltage region of the device.

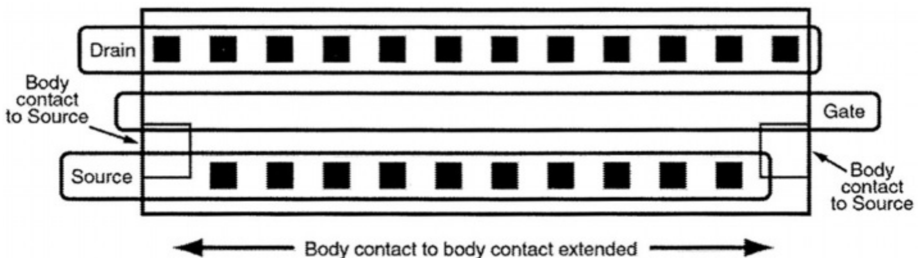


Figure 2.22: Widened body contacts allow the center of the PD-MOS device body voltage to move with capacitive coupling.

Body Pre-Charging and Biasing

There are several options for biasing the body to a known potential. The simplest implementation, in terms of circuit design, is that of body tied to source, or body tied to the appropriate supply or ground. These options do not utilize the full capabilities of the SOI devices, though they do tend to minimize leakage.

Pre-charging of the body has been proposed, and is discussed in chapter 10. This technique uses the fact that charging the body with a known current will raise the body voltage to a forward biased diode level. Thus the body voltages of two or more matched devices can be equalized prior to a measurement being taken.

Applications of body tied to unlike body

Another body tied approach, which is applicable to low voltage applications, is to tie the body of a nMOS to the body of an associated pMOS, such as the nMOS and pMOS of an inverter. This may have certain advantages at sub 1V operation, but is unlikely to be used above 1V. The reason is that in this configuration, two forward biased diodes are formed.

The temperature coefficient of the on-state forward voltage drop (Vf) of a junction diode is about -1.8mV per °C. Since Vf is typically quoted as being around 0.7V at room temperature, the Vf falls to 0.5V at about 125°C and 0.45V at 150°C. Thus, if such a circuit were raised in temperature above normal operating conditions there is danger of turning on the diode stack from supply to ground. The main advantage of this scheme is its higher current capability, without additional drive capacitance. Also, for a given supply voltage, the additional boost to current at high temperatures may give increased performance under these condition, at the expense of higher leakage.

2.11 Device Noise

There are two basic noise sources, although these can be split further into sub-set categories. These are circuit related noise, such as capacitive or substrate coupling (chapter 11) and noise generated within the active or passive devices.

Most significant among device noise considerations is MOS transistor noise caused through impact ionization, and oxide traps. Impact ionization effects are caused when the electrons from the source region (for an nMOS) are accelerated to the point where ionization can occur. This is shown in figure 2.23.

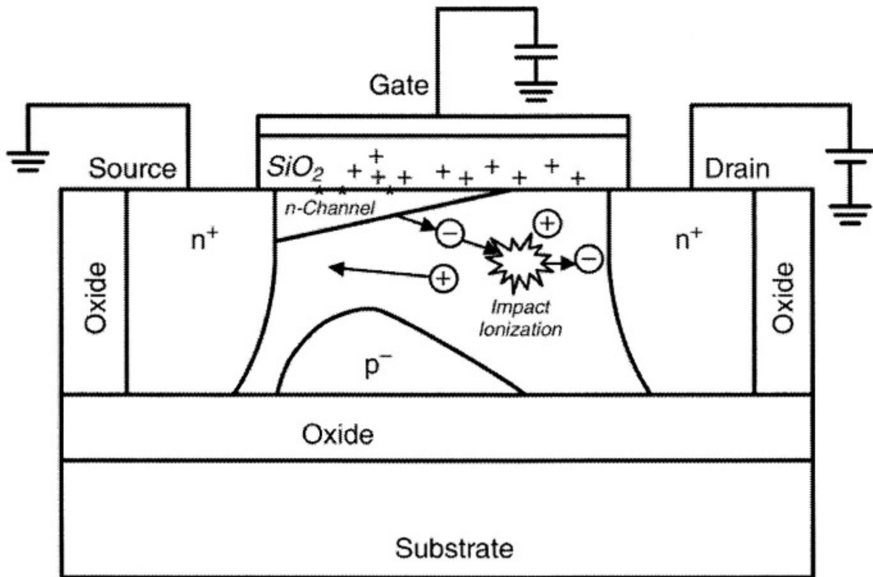


Figure 2.23: Electrons from the source and channel region can gain enough energy as they accelerate across to the drain such that an impact causes an electron hole pair to be generated. The hole tends to migrate to the gate oxide, where it can be trapped, and become a source of noise.

Once a charge unit (electron or hole) has been trapped in the oxide it can become a noise source, which will be more or less significant depending upon the trap depth within the oxide, its activation energy, and the temperature of operation of the circuit.

Figure 2.24 illustrates the trapping mechanism within the oxide. Only traps close to the Fermi level are likely to be active and thus contribute significantly to the noise level. As a result noise varies with temperature. Tunneling of carriers within the oxide is most efficient when occurring between trap centers that are at the Fermi level energy.

While it is possible for tunneling to occur at other trap levels, any traps far away from the Fermi level are statistically likely to be filled or empty, dependent upon their energy position. Thus tunneling far from the Fermi level is not significant enough to affect the overall noise characteristics.

Static noise in PD-SOI is higher than in bulk due to additional noise being induced through the floating-body effects and leakage current [2.57]. In PD-SOI there is a parasitic bipolar which can result in noise [2.58]. Body potential variability, and thus threshold voltage of devices, also affect noise.

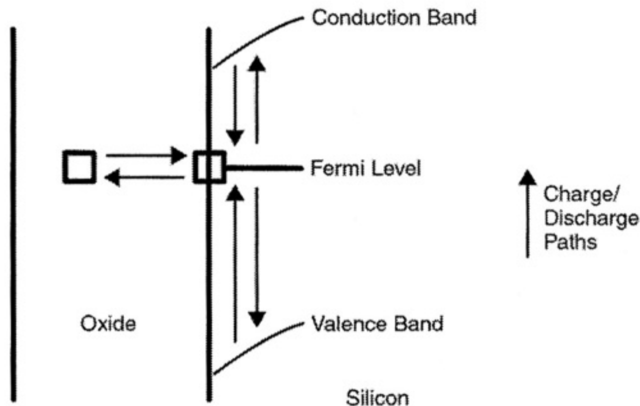


Figure 2.24: This represents a trap deep in the oxide. Traps up to about 15 Angstroms deep contribute to noise. Once filled, the trap continues to affect the channel until it depopulates the trap center.

Without successful prediction of noise, many circuits are at risk of being over designed. Static timing and noise analysis tools have demonstrated their importance in verifying digital designs [2.31, 2.59]. Transistor-level static noise analysis has been shown to be extendable to PD-SOI technology through the addition of floating-body-induced threshold voltage variation and incorporation of parasitic bipolar leakage [2.57].

To establish an initial condition for static noise analysis, a choice must be made between minimum and maximum body voltages for each transistor. Typically for memory and digital cells this might result in a worst case condition of strong noise inducing devices and weak holding devices. Although noise is normally a problem, one analog device has made beneficial use of noise. A single-stage operational transconductance amplifier in SOI has been used as a generator to produce a high level of noise at its output created by the SOI transistors [2.60].

2.12 Self Heating

Self-heating is a serious disadvantage for SOI, particularly when applied to analog circuits, although self-heating can also reduce the performance of logic circuitry and cause an increase in leakage. Self-heating has two manifestations. It can emanate from within the sub-circuit itself or from elsewhere on the same IC. Self and coupled heating can have an impact on analog circuit blocks in SOI.

Temperature rise can be significant even at relatively low power dissipations. As with bulk silicon, these effects are layout dependent.

Particularly in circuits requiring accurate matching, accounting for this type of thermal effect relies on good layout techniques and in some instances additional thermal simulation.

The buried oxide reduces thermal conductivity to the substrate (compared to bulk) resulting in locally increased temperature. This can impact circuit performance, and possibly reliability [2.27]. In addition to buried oxide, devices are relatively isolated from adjacent devices because of the use of trench isolation.

Self-heating reduces mobility and drain conductance. Analysis of drain conductance from DC to 300MHz has been performed to show that thermal dissipation can be modeled by just a single time constant [2.61]. Thus dynamic self-heating models usually provide sufficient accuracy using a single thermal resistance and capacitance. This is the type of thermal model incorporated in the BSIM-PD model (chapter 4).

Typically SIMOX has a thermal resistance of about 300°C/W (for the entire chip) and a thermal time constant of approximately 3μs. However, individual SOI devices are susceptible to local thermal heating as a result of the reduced thermal conductivity caused by the buried oxide and isolation trench. From [2.62] we can assume that the thermal resistance for a gate width of 1μm is approximately 25000°C/W.

In most logic circuits, devices conduct only during switching. Temperature rise in most devices is just a few degrees, which can normally be tolerated. This minimal effect of temperature rise on digital circuits has been confirmed experimentally: Two otherwise identical 11 and 51 stage ring oscillators on SOI were tested. Assuming leakage is low compared to switching dissipation, to a first order approximation the power dissipation per stage of the 11 stage structure is 51/11 times that of the 51 stage structure. Delay per stage of the 11 stage ring was not significantly different to that of the 51 stage ring, indicating any temperature rise difference between the two rings is not significant to circuit operation.

Analog circuits with only internal connections often have higher dissipation, and may experience a larger increase in temperature. This is of most importance in structures that are required to match, as discussed in chapter 10. Output drivers typically see temperature increases of the order of 20°C, but under switching or fault conditions this can be much higher.

Figure 2.25 shows the effect of self-heating on the MOS I-V curve. Most devices in switching circuits follow the solid curves since transient switching

does not result in significant power dissipation, hence temperature increase is minimal. Devices affected by self-heating follow a curve closer to that of the dashed lines, which takes into account the power consumed in the component. Drain-source resistance and drain voltage dictate the maximum power handling, and hence the device's contribution to self heating. This must be considered in both analog and high power design. It is also necessary to correct for self-heating effects when making DC SOI current measurements. For this type of measurement a typical device shows self-heating degradation in Ids values of between 3 and 10%.

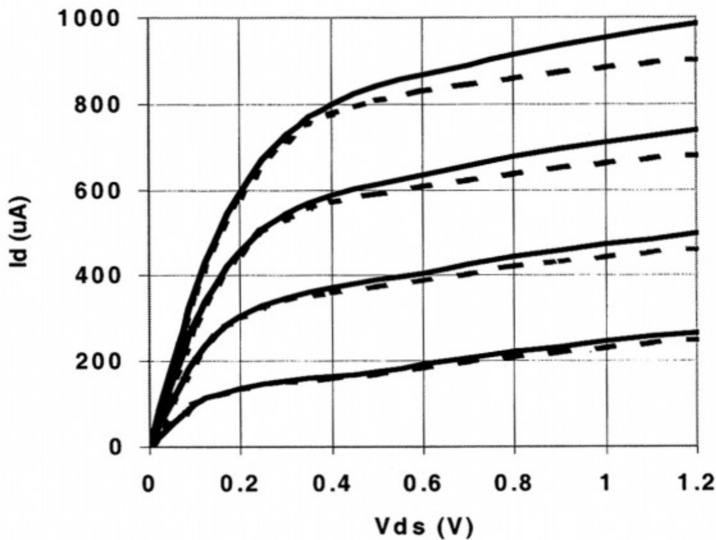


Figure 2.25: Representation of the effect of heating on a MOS device (dashed lines) compared to device operation without heating effects (solid lines). Trend lines shown for various Vgs values.

Dynamic self-heating in SOI devices is known to be an issue with analog designs [2.63]. Most SOI SPICE models, e.g. BSIM-PD and STAG v2.0 (chapter 4) include self-heating components, and it is considered essential to include self-heating in the modeling of analog circuits. This requires accurate thermal resistances and time constants [2.64, 2.65].

2.12.1 Self Heating from elsewhere on the same chip

<u>Digital Circuits</u>
Bulk digital CMOS design rarely needs to consider heating effects. With higher current densities of modern processes and the poor thermal

conductivity of the buried oxide layer, it is not necessarily the case that thermal effects can still be ignored. The buried oxide has a thermal conductivity approximately a factor of 100 worse than bulk silicon, making self-heating effects more pronounced. Assuming a 0.5μm thick oxide is equivalent to about 50μm of silicon, and a typical chip is about 250μm thick, then the additional equivalent thermal thickness of SOI is significant.

Analog Circuits

Within chip heating effects, from high thermal dissipation regions to other areas on the same chip are well understood, and layout techniques have been described in standard texts for bulk silicon [2.66, 2.67]. Many of the same basic configurations such as common centroid matched devices can be applied to SOI. (Cross coupling of matched devices serves more than immunity from thermal gradients, it also helps to minimize the effects of other processing related gradients that can occur). Conventional current mirrors in bulk silicon often use cross coupled devices to eliminate process variability across a chip, and thermal gradients due to external sources. Due to the high thermal resistivity of the oxide isolation between devices, cross coupling used in bulk devices may be less effective in SOI. Improvements to this scheme require all elements of the cross coupled structure to be within the same tank.

2.12.2 Self Heating from within a sub-circuit

Power is dissipated in the MOS channel as a result of the current through the channel and voltage across it. Due to poor thermal conductivity of SOI compared to bulk, the temperature of the layer above the buried oxide may increase significantly in the power-dissipating device, affecting its characteristics [2.65, 2.68]. The most typical effects of self-heating within a device are reduced threshold voltage, drive current, and increased leakage. Most models have the capability to determine dynamic local device temperature, through an externally accessible voltage node, which represents device temperature.

Self-heating effects are most important in devices where continuous current flows through a device with the source and drain biased at significant voltages. This occurs most frequently in analog designs. One of the simplest analog circuits in which this is important is the current mirror. Thermal effects are usually simulated with a RC network, representing thermal capacity of the silicon and θjc. Temperature rise within the device can be calculated from power and thermal resistance as follows:

θjs (junction-substrate thermal res.) = ΔT (Temperature rise)/Power (W)

Therefore: $\Delta T = \theta js * W$

W = power output = power input = Id * Vds

Hence $\Delta T = \theta_{js} * (Id * Vds)$ (1)

Note Id has temperature dependence and usually reduces with increasing temperature.

References

[2.1] C. Y. Chang and S. M. Sze, Editors, "ULSI Devices", John Wiley & Sons, Inc., ISBN 0-471-24067-2

[2.2] A. Wagpmans, et. al., "3.5mW 2.5GHz Diversity Receiver and a 1.2mW 3.6GHz VCO in Silicon-On-Anything" ISSCC98, paper No. FP 16.3:A

[2.3] R. Dekker, et. al., "An Ultra Low-Power RF bipolar technology on Glass" IEDM, 1997. pp. 921 - 923.

[2.4] M. L. Alles et.al., "Investigating optical metrology issues specific to SIMOX-SOI wafers", Micro, June 2000

[2.5] Appendix 1 (2.5)

[2.6] Appendix 1 (2.6)

[2.7] R. Hannon, et.al., "0.25 μm Merged Bulk DRAM and SOI Logic using Patterned-SOI", 2000 Symposium on VLSI Technology Digest of Technical Papers, pp. 66 - 67.

[2.8] J. Haisma, et. al., Jpn.J.Appl.Phys., vol.25, 1989, pp. 1426.

[2.9] C. E. Hunt & C. A. Desmond, Proceedings of the First Symposium on Semiconductor Wafer Bonding, 1992, pp. 165.

[2.10] A. Soderbarg, Proceedings of the First Symposium on Semiconductor Wafer Bonding, 1992, pp.190.

[2.11] Appendix 1 (2.11)

[2.12] K. Sakaguchi & T. Yonehara, "SOI Wafers Based on Epitaxial Technology", Solid State Technology, Jun 2000.

[2.13] K. Sakaguchi, et. al., Jpn. J.Appl. Phys., vol.34, 1995, pp. 842 - 847.

[2.14] P. B. Mumola & G. J. Gardopee, Conference on Solid State Devices and Material, 1995, pp. 256.

[2.15] "New Technologies for silicon-on-insulator", European Semiconductor, Feb 2000, pp. 25.

[2.16] M. I. Current et. al., "Ultrashallow Junctions or Ultrathin SOI", Solid State Technology, Sept. 2000, pp. 66 - 77.

[2.17] H. Lee, et. al., “An Anomalous Device Degradation of SOI Devices with STI”, 2000 IEEE International SOI Conference, Oct. 2000, pp. 132 - 133.

[2.18] K. Bernstein & N. J. Rohrer, “SOI Circuit Design Concepts”, Kluwer Academic Publishers, January 2000, ISBN 0-7923-7762-1

[2.19] T. Houston, et. al. “Novel MESA isolation using CMP for planarization of 0.35/0.25uM SOI”, IEEE International SOI conference, Oct 1995, pp. 110 - 1.

[2.20] G. Shahidi, et. al., Proceedings of the 1992 VLSI Symposium, p. 93 - 94.

[2.21] G.G. Shahidi, et. al., “A high performance low temperature 0.3 µm CMOS on SIMOX” 1992 Symposium on VLSI Technology, 1992, pp. 106 – 107.

[2.22] L. Su, et. al., Proceedings of the 1996 VLSI Symposium.

[2.23] S. T. Liu, et. al., "Radiation Response of SOI Materials”, Electrochemical Soc. Proc. 1999, pp. 225 - 230.

[2.24] G. Shahidi, et. al., 'Mainstreaming of the SOI Technology,' Proc. IEEE Inter. SOI Conf, 1999, pp. 1-4.

[2.25] E. Leobandung, et. al., "High Performance 0.18um SOI CMOS Technology,' IEDM-99, 1999, pp. 679 - 682.

[2.26] S. Cristoloveanu, "Architecture of SOI Transistors: What's Next?", 2000 IEEE International SOI Conference, Oct. 2000, pp. 1 - 3.

[2.27] G. G. Shahidi, et. al., “Device and Circuit Design Issues in SOI Technology”, Proceedings of the IEEE, 1999, pp. 339 – 346.

[2.28] G. Shahidi et. al., "A Room Temperature 0.1µm CMOS on SOI", IEEE Transactions on Electron Devices, Vol 41, Dec. 1994, pp. 2405 - 2412.

[2.29] G. Anthony, et. al., "A 0.2-mm, 1.8-V, SOT, 550-MHz, 64-b PowerPC Microprocessor with Copper Interconnects", IEEE Journal of Solid State Circuits, Vol. 34, No. 11, Nov. 1999, pp. 1430 - 4.

[2.30] Proc. 5th Inter. Workshop on Measurement, Characterization and Modeling of ultrashallow Doping Profiles in Semiconductors, J. Vac. Sci. Technol., 2000, pp. 337 -604.

[2.31] K. Sukegawa, et. al., "High-performance 80-nm Gate Length SOI-cMOS Technology with Copper and Very-low-k Interconnects", 2000 IEEE Symposium on VLSI Technology Digest of Technical Papers

[2.32] D. Edelstein, et. al., "Full copper wiring in a sub-0.25 μm CMOS ULSI technology" IEDM Tech.Dig.,1997, pp.773 - 6.

[2.33] M. Ikeda, et. al., "Integration of organic low-k material with Cu-damascene employing novel process", Interconnect Technology Conference, 1998. pp. 131-3.

[2.34] K. Mistry, et. al., "Scalability Revisited: 100 nm PD-SOI Transistors and Implications for 50 nm Devices", IEEE VLSI Conference 2000, 0-7803-6308-6, pp. 204-5.

[2.35] R. Chau, et. al., IEDM Tech. Digest, 1997, pp. 591 - 594.

[2.36] E. Leobandung, et. al., "Scalability of SOI technology into 0.13 μm 1.2 V cMOS generation", IEDM Tech. Digest, 1998, pp. 403 - 406.

[2.37] D. Flandre, et. al., "Fully-Depleted-SOI cMOS technology for low-voltage low-power digital/analog/microwave circuits", Analog Integrated Circuits and Signal Processing. New York: Kluwer, 1998.

[2.38] J-P. Eggermont, et. al., "Potential and Modeling of 1μm SOI cMOS Operational Transconductance Amplifiers for Applications up to 1 Ghz" IEEE Journal of Solid State Circuits, Vol. 33, No. 4, April 1998, pp 640 - 643.

[2.39] S. M. Sze, "Physics of Semiconductor Devices", John Wiley and Sons, New York.

[2.40] S. K. H. Fung, et. al., "Impact of the Gate-to-Body Tunneling Current on SOI History Effect", 2000 IEEE International SOI Conference, Oct. 2000, pp. 122 - 3.

[2.41] J. Y. Choi & J. G. Fossum, "Analysis and control of floating body bipolar effect in fully Depleted submicrometer SOI MOSFET's", IEEE Trans. Electron Devices, vol.38, 1991, pp.1384 - 91.

[2.42] W. Redman-White, et. al., "Analogue design issues for SOI CMOS" IEEE International SOI Conference, 1996, pp. 6 - 8.

[2.43] Y-C. Tseng, et. al., "Local Floating Body Effect in Body-grounded-SOI nMOSFETs", Proceedings 1997 IEEE International SOI Conference, Oct 1997, pp. 26-7.

[2.44] O. Rozeau, et. al., "Impact of Floating Body and BS-Tied Architectures on SOI MOSFET's Radio-Frequency Performances", 2000 IEEE International SOI Conference, Oct. 2000, pp. 124 - 5.

[2.45] Y. C. Tseng, et. al., "Correlation between Low Frequency Noise Overshoot in SOI MOSFETs and Frequency Dependence of Floating Body Effect", 1997 Symposium on VLSI Technology Digest of Technical Papers, 1997, pp. 99 - 100.

[2.46] J. Gautier, et. al., IEDM Tech. Dig., 1995, pp. 61 - 3.

[2.47] D. Suh & J. G. Fossum, "Dynamic floating-body instabilities in partially depleted SOI CMOS circuits", IEDM Technical Digest, 1994, pp. 661 - 664.

[2.48] A. Wei, et. al., "Effect of floating-body charge on SOI MOSFET design", IEEE Trans. Electron Devices, vol. 45, Feb. 1998. pp. 430 – 438.

[2.49] F. Assaderaghi, et. al., "History-Dependence of Non-Fully Depleted (NFD) Digital SOI Circuits", VLSI Tech. Dig., 1996, pp. 122 - 3.

[2.50] J. P. Colinge, "Silicon-On-Insulator Technology: Materials to VLSI", Kluwer Academic Publishers, 1991.

[2.51] J. Eckhardt et. al., "A SOI specific PLL for 1 GHz microprocessors in 0.25 μm 1.8 V", ISSCC 1999. pp. 436 - 437.

[2.52] A. Nishiiyama, et. al., "Mechanism of the suppression of the floating-body effect for SOI MOSFET's with SiGe source structure", Proc. 1996 IEEE Int. SOI Conf., Oct. 1996, pp. 68 - 69.

[2.53] K. Suma, et. al., "An SOI-DRAM with wide operating voltage range by CMOS/SIMOX technology", IEEE J. Solid State Circuits, vol. 29, Nov. 1994, pp. 1323 - 1328.

[2.54] G. Shahidi et. al., "Partially-Depleted-SOI Technology for Digital Logic", ISSCC 1999, pp. 426 - 427.

[2.55] Silicon-On-Insulator Integrated Circuit Technology, U.S. Patent No. 5,489,792

[2.56] J.G. Fossum & G.O. Workman, "A comparative analysis of the dynamic behavior of BTG/SOI MOSFETs and circuits with distributed body resistance", IEEE Transactions on Electron Devices, Oct. 1998, pp. 2138 - 2145.

[2.57] K.L. Shepard & D-J. Kim, "Static Noise Analysis for Digital Integrated Circuits in Partially-Depleted Silicon-On-Insulator Technology", Design Automation Conference, 2000. pp. 239 – 242.

[2.58] P-F. Lu, et. al., "Floating-body effects in partially Depleted-SOI CMOS circuits", IEEE Journal of Solid-State Circuits, August 1997, pp. 1241 - 1253.

[2.59] K. L. Shepard, et. al., "Harmony: Static noise analysis for deep-submicron digital integrated circuits", IEEE Transactions on Computer Aided Design, Aug 1999, pp. 1132 - 1150.

[2.60] A. Nève, et. al., "Smart Card Circuits in SOI Technology", 2000 IEEE International SOI Conference, Oct. 2000, pp. 48-49.

[2.61] B. M. Tenbroek et. al., Proceedings ESSDERC 1993, Grenoble, France, pp. 189-192

[2.62] H. Nakayama, et. al., "SOI MOSFET Thermal Conductance and Its Geometry Dependence" 2000 IEEE International SOI Conference, Oct. 2000, pp. 128 - 9.

[2.63] B. M. Tenbroek, et. al., "Characterization of layout dependent thermal coupling in SOI CMOS current mirrors", IEEE Transactions on Electron Devices, Dec. 1996, pp. 2227 – 2232.

2.12 Self Heating

[2.64] B. M. Tenbroek, et. al., "Measurement and Simulation of Self Heating in SOI CMOS Analogue Circuits", Proceedings 1997 IEEE International SOI Conference, Oct. 1997, pp. 156 - 7.

[2.65] B. M Tenbroek, et. al., “Self-heating effects in SOI MOSFETs and their measurement by small signal conductance techniques” , IEEE Transactions on Electron Devices, Vol. 43, No 12, 1996, pp. 2240 - 2248.

[2.66] A. Hastings, “The Art of Analog Layout”, Prentice-Hall, ISBN: 0-13-087061-7, 2000

[2.67] P. R. Gray & R. G. Meyer, “Analysis and Design of Analog Integrated Circuits, third edition”, Wiley, 1993, ISBN 0-471-57495-3.

[2.68] M. S. L. Lee, et. al., "Modelling of thin film SOI devices for circuit simulation including per-instance dynamic self-heating effects", Proc. IEEE Int. SOI Conference, Palm Springs, CA, Oct. 1993, pp. 150 - 151.

Chapter 3: Components

3.1 MOS devices

Figure 3.1 shows the floating body nMOS and pMOS schematic structures. These are the basic SOI components for all PD-SOI except for very specialized options. Figure 3.2 shows cross sections of SOI MOS transistors.

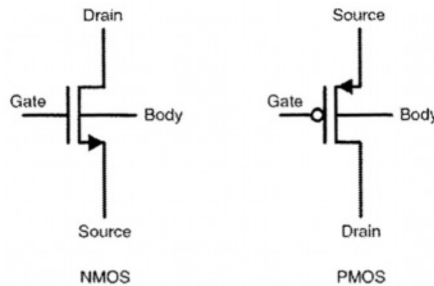


Figure 3.1: Floating gate nMOS and pMOS structures in SOI.

Other MOS devices available are the body tied to source and body contacted structures. For body tied to source, the body is built into the source diffusion. For the body contacted device, the body is made accessible and can be contacted to any appropriate electrical node. Both structures can be readily constructed in SOI, although compared to floating body devices there is some reduction in area efficiency.

Drain Current Temperature Dependence.

A plot of drain current vs. temperature for a body tied (bulk equivalent) and non-body tied or NBT (floating body) device is shown in figure 3.3. The floating body scheme outperforms the body tied scheme, although

performance is very similar at high temperatures. The drain current changes with supply voltage, temperature and device characteristics, making it hard to compare bulk and SOI material performance.

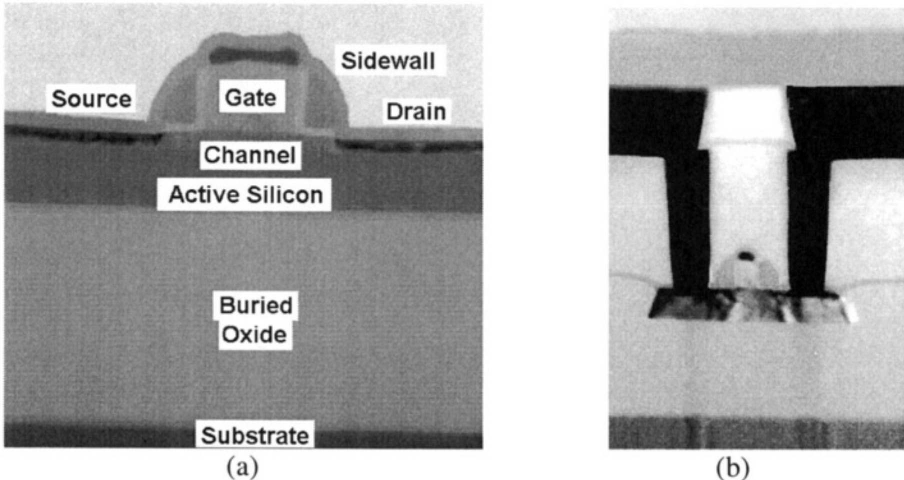


Figure 3.2: Cross sections of a PD-SOI NMOS transistor. (a) shows the basic MOS structure (b) shows the source/drain contacts, and silicon mesa.

Body tied devices are often used in analog applications, as this minimizes the kink effect. At higher frequencies the RC caused by capacitive coupling of the drain/gate to body and the distributed resistance of the body may limit the effectiveness of body contacts, reducing the benefit of body tying. However, the time constant of body tied contacts is still much reduced from that of floating body devices.

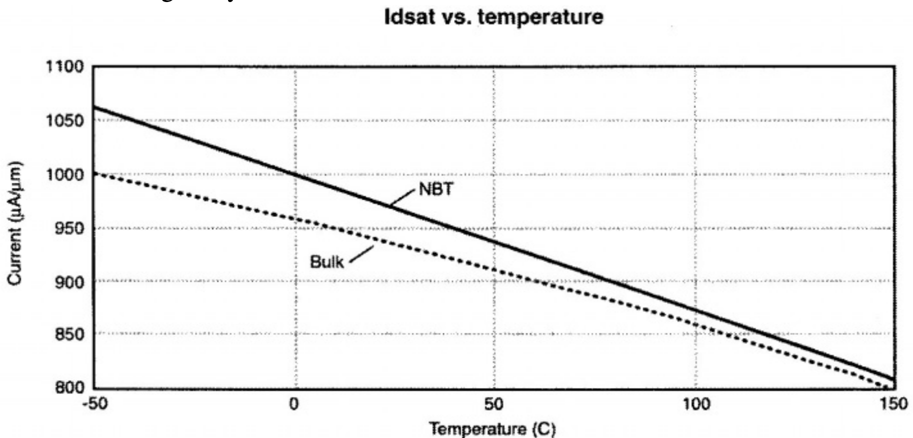


Figure 3.3: Variation of Idsat with temperature for a bulk equivalent nMOS and a floating body SOI device.

Converting designs from Bulk to SOI

It is inevitable, especially in the initial conversion from bulk substrates to SOI, that many SOI designs will be based on prior designs for bulk material. The best approach to converting circuit designs from bulk to SOI is to maintain as much as possible of the bulk design, and only redesign sections that require specific SOI consideration [3.1].

3.2 Diodes

Junction diodes can be created using the same basic diffusions as body contacted MOS devices, or they can be created with the addition of a silicide block option. This is described in the chapter 5. Unlike diodes in bulk material, diodes in SOI do not have parasitic bipolar structures to substrate, and are therefore easier to handle in terms of circuit design, as one type of leakage and latchup mechanism is eliminated.

3.3 Bipolar Transistors

In addition to MOS devices, the other active devices that can be easily constructed are NPN and PNP bipolars. These may be created by accessing the body contact of a MOS device, and connecting its gate to source. Thus the nMOS and pMOS can be used as isolated NPN and PNP devices. Most of these are lateral structures, such as used for the PNP in many bipolar and BiCMOS processes. More detail on the layout of bipolar devices is available in chapters 5.

A disadvantage of the simplest bipolar devices in PD-SOI, which use the parasitic bipolar device of the standard MOS devices is that these structures have a higher base resistance. This can be minimized by process additions or modifications. Figure 3.4 shows a lateral bipolar structure intended for RF applications [3.2]. This uses a cobalt silicided base to achieve an f_{max} of 67GHz. The emitter, intrinsic base, and collector form a lateral structure. The P+ polysilicon base reduces base resistance and is important for achieving high frequency operation. The 'link base' region is formed to contact the silicon base region to the polysilicon electrode. The link base is created by an angled ion implant.

High Current / Voltage Structures

Monolithic high voltage and high power applications are generally designed using power integrated circuit processes. Modern power ICs are most frequently based on CMOS processes, with high voltage lateral DMOS (Double diffused MOS) structures included as a process flow adder to a basic, and often mature process [3.3]. The lateral DMOS or LDMOS structures are a particular variant of a conventional nMOS or pMOS structure, but they are most easily built in the form of a nMOS option. This

makes design more difficult for high-side operation, usually requiring a charge pump to raise the nMOS gate above supply. If the DMOS is required to function as a high side driver it is often a requirement to have additional components with higher voltages than the logic voltage devices. This may include long channel devices or drain extended (DE) devices [3.3].

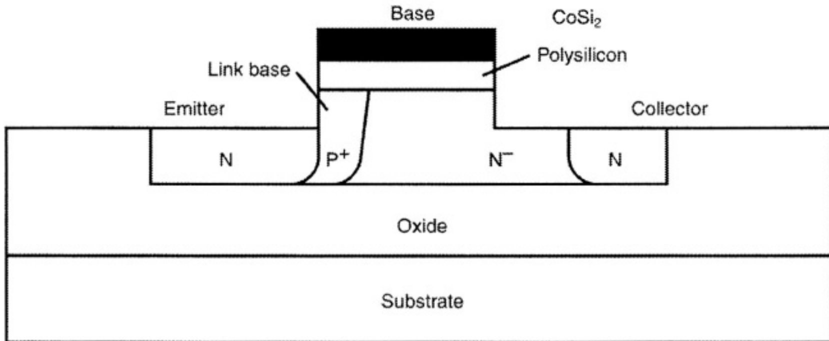


Figure 3.4: Cross-sectional view of the layout of a high frequency bipolar transistor in SOI. The device utilizes a contact to the base region along the length of a silicided gate, to reduce base resistance and improve frequency performance.

One of the most important advantages of SOI is the lack of parasitic transistor action through the substrate. This parasitic has proved to be significant in power integrated circuits, where both high currents and fast switching power transients are, by necessity, adjacent to low power and precision analog and logic circuitry. It has long been understood that parasitic thyristors can be triggered in conventional logic when a differential occurs between logic and power grounds of a power IC [3.4]. While design of conventional DE and LDMOS devices is not part of this work, a brief perspective of the devices is in order, as comparison with the equivalent structures that can be built on SOI. The use of SOI for power integrated circuits is advantageous for die area reduction [3.5] through reduced isolation space.

Power transistors built on SOI use thicker active regions than conventional PD-SOI, to allow high voltage components to be fabricated. Also, most SOI based high-voltage power transistors have thick buried oxides, to minimize the risk of electrical breakdown between active silicon and substrate. However, thicker oxide has higher thermal resistance [3.6]. Replacing conventional SOI with silicon-on-diamond may be used to advantage for eliminating both these problems [3.7]. Despite the fact that diamond is compatible with conventional silicon processing [3.8, 3.9], it remains

unlikely that silicon-on-diamond will become a significant commercial process due to the additional material costs.

3.4 Lateral DMOS

Throughout the 1990s the LDMOS was the workhorse structure of bulk-silicon power integrated circuits [3.10, 3.11]. It appears likely to remain prominent for the foreseeable future since the base processes into which the LDMOS is normally integrated are typical CMOS. An LDMOS cross section is shown in figure 3.5, and construction of the device is as follows: An N-type well is diffused into a P-type silicon substrate. A self aligned boron and arsenic combination is diffused into the N-well. The arsenic produces a n-type diffusion, and acts as the source of the MOS device, and the boron forms a p-type region around the source. In most cases the boron back-gate or body is connected to the source for area efficiency (compared to bringing out separate body and source contacts), and to minimize body resistance. The surface region above the P-type diffusion is covered with a poly-silicon gate. The well creates the drain, which is contacted by a high concentration N-type diffusion. A variation of this basic device includes the addition of a buried N+ layer, to help isolate the structure and minimize the parasitic PNP from body to substrate. Additional processing, often combined with the buried N+ layer, is the addition of deep N drain contacts, to improve efficiency of the drain (through reduced resistance) and further reduce the PNP parasitic to substrate. Figure 3.6 shows these process adders. There are many examples of how this device is used in power IC processes [3.11], which show the capabilities and the problems of combining high power devices on the same silicon as high precision analog and digital circuitry. The LDMOS is usually built as a nMOS structure for isolation, efficiency and self alignment reasons, and most power devices use LDMOS structures for both high and low-side driving, and H-bridge design [3.12, 3.13].

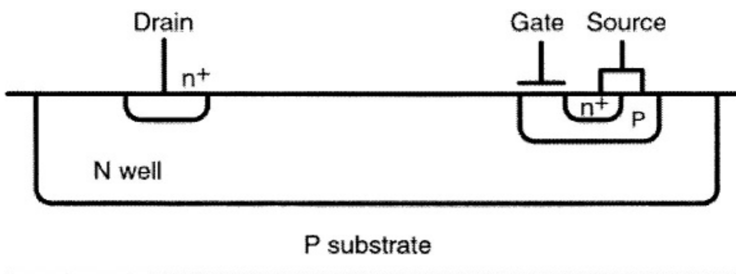


Figure 3.5: Cross section of the LDMOS structure in bulk silicon. The Source contacts the N+ source and P body, shorting the parasitic NPN. The structure is often enhanced with deep N+ drain for reduced resistance and a buried N+ layer under the N- Well, to minimize the parasitic PNP gain.

Even with the PNP parasitic gain reduced significantly with the addition of the buried N+ region and deep N diffusion, there remains the risk of latch-up being caused by the parasitic. An example of this is when the LDMOS is used as a high-side driver. During recirculation of inductive loads, the source and body voltages rise above the drain voltage. With a 10A recirculation current and a gain of just 0.1, a substrate current of 100mA is observed from the vertical parasitic PNP structure. Such a current can easily disrupt the operation of logic, and can cause latch-up to occur.

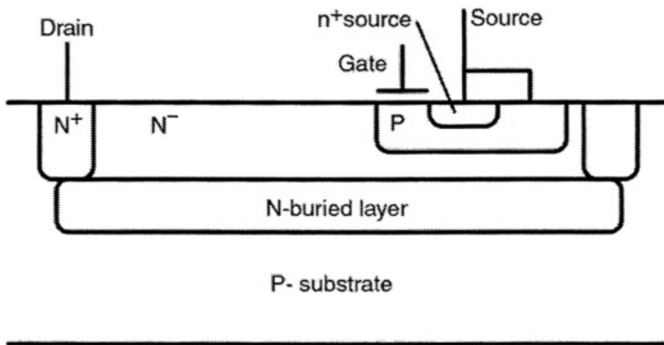


Figure 3.6: Variation on the basic LDMOS device of figure 3.5, showing the inclusion of a buried N+ layer and the addition of deep N drain contacts. The deep-N drain contact usually forms a ring around the N-well, to give maximum isolation and reduce the PNP gain to substrate.

<u>RESURF Devices</u>

"Reduced Surface Field" or RESURF devices are based on creating a very uniform drift region field. The structure is usually an LDMOS. Modification of the drift region is achieved with an appropriate doping profile and thickness. This leads to a depletion of the entire drift region of charge carriers, creating a uniform electric field in the drift region that maximizes device breakdown. A fully depleted drift region is in the RESURF state. Figure 3.7 shows a RESURF SOI device [3.14]. 600V to 850V SOI RESURF LDMOS transistors have been demonstrated [3.15].

<u>Low resistivity SOI MOS structures</u>

With the isolation available from SOI, it is possible to achieve devices with improved efficiency over bulk, and effective elimination of parasitics through the substrate. Structures optimized for the 50-70V range of operation show better than 0.4mΩ-cm^2 typical values, compared to equivalent devices using conventional state-of-the-art power IC processing of 0.5mΩ-cm^2. However, this is not the only benefit, as SOI based power

ICs also benefit from substantial reduction in electrical isolation spacing and guard-rings are eliminated.

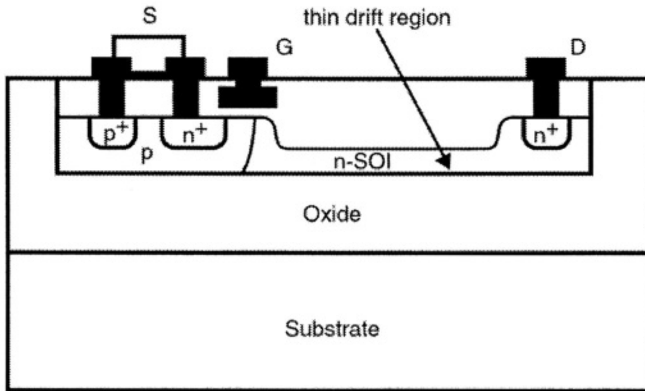


Figure 3.7: Example of how a high voltage LDMOS structure can be produced with RESURF properties. This device makes use of a thin drift region to achieve greater than 600v breakdown.

Although R_{dson} is reduced, SOI power devices have the disadvantage of higher θjc, through both reduced area back-side and a buried oxide layer. This may be partially bypassed with the use of top-side heat spreading or flip-chip mounting [3.16] for improved thermal dissipation. In addition, there are three other options, which should be considered:

(1) Increasing power device size. Larger devices, driving a particular load have reduced on-chip dissipation due to reduced power device resistance, while simultaneously increasing the power dissipation area.

(2) Distributed transistors minimize power dissipation issues. This is practical in SOI due to the minimal electrical isolation requirements of SOI compared to bulk [3.17].

(3) Build the power device into a bulk region of the silicon, through the use of selective SIMOX (chapter 2), or partial silicon-on-insulator [3.18]. In the cases of selective SIMOX or partial SOI, the low power consumption circuitry remains in SOI, thus retaining electrical isolation.

3.5 Drain Extended Devices

Drain Extended (DE) devices are extensions of conventional nMOS and pMOS structures. DE devices are less efficient than LDMOS structures, in terms of on resistance per unit area. They are usually used as pre-drivers for the LDMOS, and therefore need not be very efficient, but often need to be slightly higher in stand-off voltage. The DE-pMOS can also be used in some applications where a charge pump is not available as a driver for a high-side

structure at the expense of some die area trade-off. In bulk material, when only low current is required, they can be made much smaller than DMOS devices due to the added isolation required for the DMOS devices. As with the DMOS devices discussed above, DE devices can be readily produced in SOI. Here we discuss the DE-pMOS, although the nMOS is similar, but is not normally isolated from the substrate. It is normal for the DE devices to be designed for, and rated up to 20 volts higher than the LDMOS structure of a given process. This is done to permit driving of high side LDMOS devices, where the gate must often be driven above the supply voltage for high-side operation.

The DE-pMOS structure is shown in figure 3.8, and is constructed as follows: Nwell is diffused into the P-type substrate (this may or may not be the same diffusion as used for the DMOS N-well). The high resistivity, (p-type) drain extension is diffused and the drain contact and source (both P+, low resistivity) are also diffused. Contact is made also to the Nwell body region. This is very similar to the basic MOS structure, except for the additional high resistivity drain extension. This type of structure can be readily designed to have a blocking voltage up to about a 100V, without process modifications.

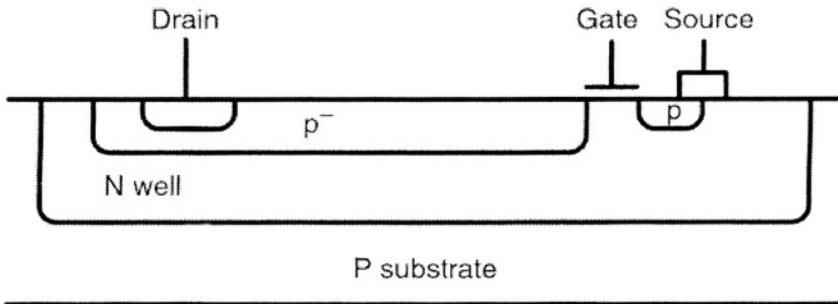


Figure 3.8: Simplified DE-pMOS cross section for bulk silicon applications. The drain extended MOS structure relies on an extension of the drain region using a lightly doped well. Here the Nwell tank is shorted to the p-source through a metal level contact.

Very often DE-MOS devices are built in a circular or square pattern, with the drain as the center and the drain-extension around it. The gate then surrounds the drain extension, and the source surrounds the entire drain and gate. This is advantageous in preventing leakage through surface inversion, which can occur in junction isolated structures exposed to high voltages, but is of reduced importance in SOI trench isolated systems. An example of the center drain layout is shown in figure 3.9.

3.5.1 Design of an SOI high voltage DEMOS device.

As with drain extended devices in bulk silicon, SOI high voltage MOS components are generally extensions of the conventional nMOS and pMOS structures. Here we discuss the high voltage (HV)-SOI-nMOS, although the pMOS is similar. As with bulk devices, it may be that the controller devices have to be rated for higher voltage than the output devices.

HV-nMOS structure is shown in figure 3.10, and is constructed as follows: A high resistivity, n-type drain extension is diffused, followed by drain and source diffusions. This is similar to the basic MOS structure except for the additional high resistivity drain extension. As with high voltage structures in bulk material, care is required to avoid inversion of the active silicon adjacent to the substrate, which could significantly impact leakage.

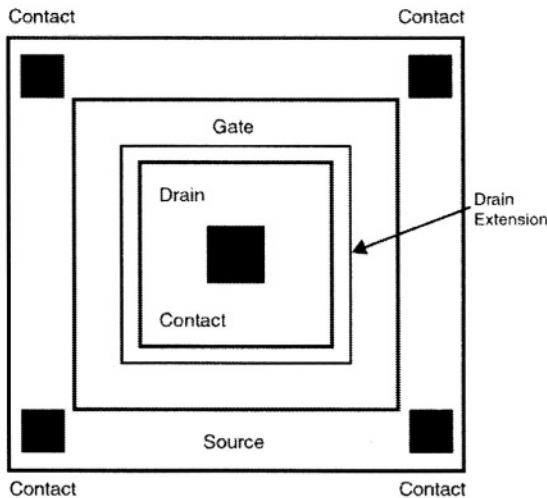


Figure 3.9: Example of a center contact MOS device. This type of device is commonly used in junction isolated circuits, where there is a requirement to isolate the high voltage on the drain from the rest of the circuitry to minimize leakage.

The device shown in figure 3.10 is single sided, and it can withstand high voltage in one direction only. Variants on this device exist that allow stand-off in both directions. These typically involve extending the source as well as the drain, to give a drain-extended-source-extended MOS (DESEMOS) (figure 3.11). It should be noted, however, that the ability to withstand voltage bi-directionally comes at a price. This device has a very high source and very high drain resistance, and is therefore inefficient, and only used when no reasonable alternative is available.

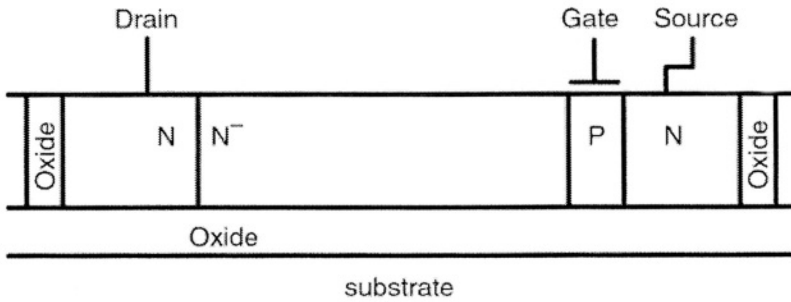


Figure 3.10: Simplified SOI HV nMOS structure cross-section. This structure has a number of similarities to the drain-extended structure used for many years in power IC processes. The major advantages compared to bulk are reduced parasitics, higher voltage capabilities and higher efficiency.

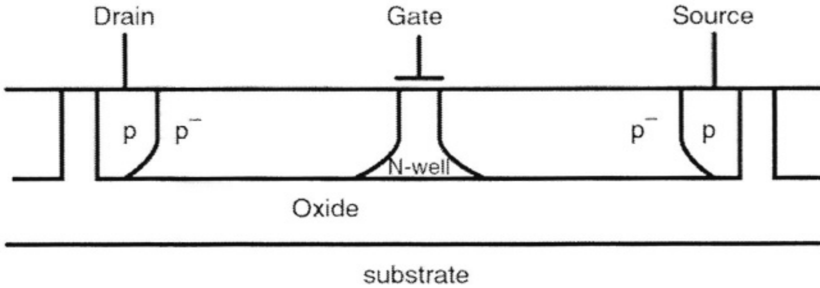


Figure 3.11: DESEpMOS cross section. Here the source is extended, as well as the drain, to give high voltage standoff on both source and drain sides of the device. The Nwell tank is shorted to the p-source through a metal level contact, or may float.

3.6 Compound High Voltage SOI Structures

It is possible to produce a variety of compound device structures in SOI. These usually combine multiple devices, for example nMOS, pMOS, bipolar transistors etc. in a single tank. A 600V compound lateral power device described for SOI [3.19] uses standard SOI CMOS and a single mask level adder, to generate a p-drift layer. The device thus generated is a lateral p-channel power device (figure 3.12).

The equivalent circuit for the structure in figure 3.12 is shown in figure 3.13. The pMOS device is turned on with a negative voltage pulse, which then, activates the nMOS structure. The device performs the function of a high voltage pMOS, but is constructed this way to benefit from the improved performance of the nMOS devices.

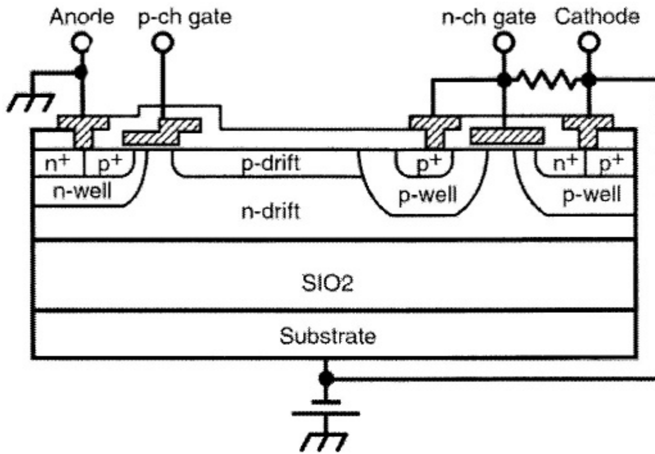


Figure 3.12: Hybrid device built in SOI. The hybrid includes a pMOS actuating device, and nMOS power structure. The combination is capable of operation to several hundred volts.

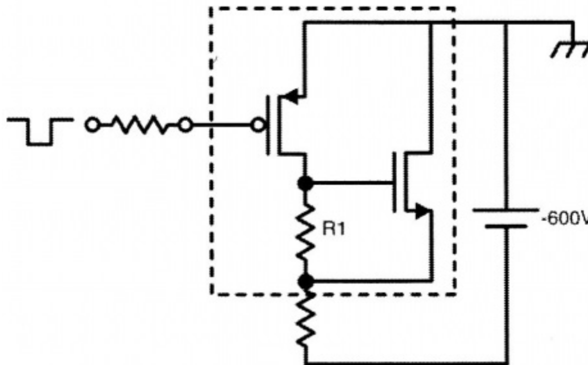


Figure 3.13: Simplified schematic representation of the hybrid structure shown in figure 3.12. The structure contains nMOS, pMOS and resistive elements. The components outside the dashed line are external to the SOI structure.

3.7 Passive Components

3.7.1 Resistors

Polysilicon resistors can be constructed on SOI in just the same way as on bulk processes. They have reduced capacitance to substrate resulting from the BOX region and thus polysilicon resistors are largely unaffected by voltage difference between resistor and substrate.

SOI Well or tank resistors are significantly different to their bulk counterparts. In bulk silicon a diffusion, usually phosphorous (n-type) doping into a p-type substrate, creates a region that can be contacted to produce a resistor. Figure 3.14(a) shows a well resistor, and figure 3.14(b) shows the equivalent SOI structure. The SOI mesa resistor has numerous advantages over its bulk counterpart. Since it is not part of a contiguous silicon region, there is no possibility of placing the resistor into a region where its parasitic diode to ground can turn 'on', or become part of a thyristor structure. Another advantage of SOI mesa resistors is a reduction of the oxide isolation capacitance to ground. Capacitance variation with respect to the resistor voltage variation is also greatly reduced. The most significant advantage of the SOI resistor is its relative immunity to bias conditions. As the voltage from the resistor terminals to ground increases the resistance of the structure increases significantly in bulk tank resistors, as the capacitance depletion extends further into the well as reverse bias across the junction increases. This restricts the applications for well resistors, especially when resistor matching is essential. For devices built on SOI, depletion, and therefore resistance increases, can still occur, especially at high bias levels, but it is normally insignificant for many design requirements.

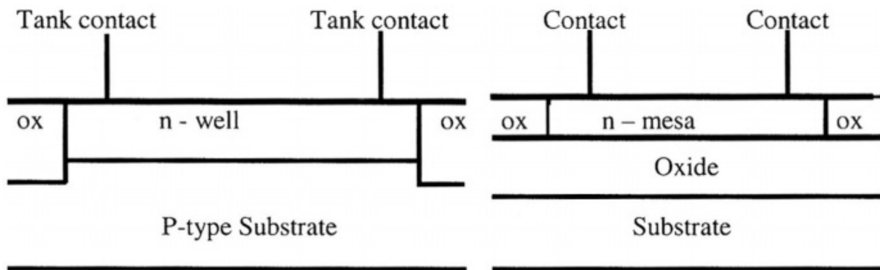


Figure 3.14: Comparison of the cross sections of well-type resistors in (a) bulk silicon, and (b) SOI

3.7.2 Capacitors

Well capacitors are not available in SOI. This is not a serious problem, as they are not used much in modern devices. If required one of the mixed SOI and bulk technologies should be considered. However, it is worth noting that the large supply to ground capacitance present in bulk material is provided by a well capacitor which is absent in SOI.

Polysilicon (gate) capacitors usually have much higher series resistance when built in SOI rather than bulk, because of the thinner silicon region. This can be minimized through the use of a multiple-finger layout. While

this structure performance is worse in terms of resistance, it has less parasitic capacitance to ground.

The metal capacitor is the most common capacitor found in most designs. Often this is formed with multiple metal layers, and the plates are interdigitated for maximum capacitance. Parasitic capacitance to substrate is reduced in SOI compared to bulk.

3.7.3 Inductors

Inductors are normally constructed in copper or aluminum metal layers (often at the top level of metal) as one of the last process steps. As a result, it may at first seem irrelevant to include inductors in a book on SOI. There are, however, two reasons that it is interesting to consider inductor behavior. The first is that SOI is being used as a means to develop RF circuitry in CMOS technology, so any influence that SOI has on integrated inductors is important. The second reason is that SOI does provide certain opportunities for improving the Q-factor of inductances. Inductors and transformers are constructed as a spiral using one or several levels of metal. There has been a good deal of work on optimum design for inductors, but a fairly typical design uses a square spiral (figure 3.15).

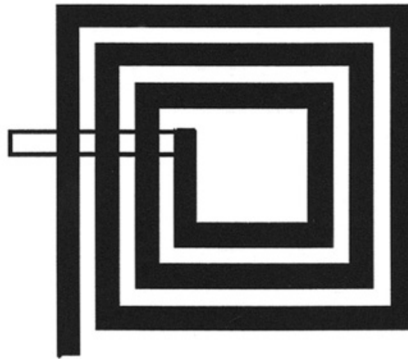


Figure 3.15: Example of an inductor layout on a monolithic IC. The solid trace represents one or more metal levels, and the center tap tunnels under the inductor proper on a lower level of metal.

Typically planar inductors over silicon are simulated using electromagnetic simulators such as Sonnet® (appendix 2), which can be utilized for design and analysis of high-frequency circuits, filters, waveguides and antennas. The S-parameter output from Sonnet may be converted to inductance and Q values, with tools such as Matlab®(appendix 2). Inputs required for a 3D planar EM simulation are conductor width, spacing, thickness, number of turns, diameter of inductor, substrate resistivity, dielectric layer properties

and metal properties. A number of RF circuit designs use inductors for tuning. Typically inductors constructed in the metal layers over silicon exhibit poor Q, often in the single digits.

Topology, metal resistance and the proximity to the substrate all contribute to the final Q value. For an inductor of 2nH, and a diameter of 225x225μm, peak Q increases from about 10 (at 2GHz) to over 20 (at 7GHz) as substrate resistivity increases from 10Ω-cm to 10KΩ-cm. Conventional bulk CMOS uses moderate resistance substrates. For SOI material made using bonding techniques, the restrictions that push conventional substrates to low resistance values, such as minimizing the parasitic transistor gain across a wafer, do not apply. It is easier therefore to use higher resistivity substrates, which lead to improved Q.

The ultimate SOI process for inductors is one of the silicon-on-quartz type technologies. The 'silicon on anything' technology, where substitution of quartz or glass has been made for a silicon substrate, has a quality factor of 29 at 4GHz, with a self-resonance frequency of 10.5GHz [3.20 - 3.22].

References

[3.1] C. R. Tretz, "Circuit Design in SOI: Concept of Floating Beta Ratio", 2000 IEEE International SOI Conference, Oct. 2000, pp. 4 - 5.

[3.2] H. Nii, et. al., "A Novel Lateral Bipolar Transistor with 67GHz f-max on Thin Film SOI for RF Analog Applications", IEEE Trans. on Electron Devices, Vol. 47, No. 7, July 2000, pp. 1536 - 41

[3.3] A. Marshall & K.G. Buss, "Automotive Semiconductor Switch Technologies", IEEE Workshop on Electronics Applications in Transportation, Oct 18-19 1990. IEEE Cat No. 90TH0310-3, pp. 68 - 72.

[3.4] A. Marshall & B. Grose, "Custom Automotive Requirements for Power Integrated Circuits", Texas Instruments Technical Journal, March-April 1994, pp. 93 - 98.

[3.5] T. Ohyanagi & A. Watanabe, "1.3μm cMOS technology merged with 90 V HG-DMOS SOI substrate" Proceedings 1997 IEEE international SOI Conference, Oct 1997, pp. 72-3.

[3.6] B. Edholml, et. al., "Silicon-On-Diamond MOS-Transistors with Thermally Grown Gate Oxide", Proceedings 1997 IEEE International SOI Conference, Oct 1997, pp. 30 - 31.

[3.7] B. Edholm, Ph. D. Thesis, ISBN 91-554-3673-0, Uppsala University, Jan 1996.

[3.8] B. Edholm, et. al., J. Electrochem. Soc., 143, 1996, pp. 1326 -1334.

[3.9] N. K. Annamalai, et. al., "Radiation response of silicon on diamond (SOD) devices", IEEE Trans. Nuclear Sci., 40, 1993, pp. 1780 - 6.

[3.10] A. Marshall, et.al. "Power Device Reliability", Encyclopedia of Electronics, Wiley 1999, Volume 16, pp. 602 - 608.

[3.11] T. Summerlin, et. al. "Technology Driven Design Challenges of Mixed Signal Power ASICs", Power Conversion and Intelligent Motion Conference (PCIM), Poster Session, Nov 10-11, 1999: High Frequency Power Conversion Proceedings (HFPC), Nov 1999, pp. 267 - 273.

[3.12] A. Marshall, “Intelligent Power Integrated Circuits - History and Overview”, Texas Instruments Technical Journal, March - April 1994, pp. 2 -9

[3.13] A. Marshall & J. Devore, “Power IC with EEPROM Programmable Switch Mode Regulators”, IEEE Journal of Solid State Circuits, Vol. 31, No. 9, September 1996, pp. 1351 - 6.

[3.14] C. Eisenhut & J. W. Klein, "SIMOX Voltage References for Applications up to 275°C using the Threshold Difference Principle", Proceedings 1997 IEEE International SOI Conference, Oct 1997, pp. 110 - 1.

[3.15] R. P. Zingg, et. al., “850V DMOS-Switch in Silicon-on-Insulator with Specific Ron of 13 Ohm-mm^2”, 2000 IEEE International SOI Conference, Oct. 2000, pp. 62 - 3.

[3.16] A. Marshall & F. Carvajal, “Power+ Logic Methodology Applied to a Six Output Power Driver”, IEEE 1993 Bipolar/BicMOS Circuits and Technology Meeting, Minneapolis, Minnesota, Oct 4-5, 1993. IEEE Cat No. 93CH3315-9, pp. 72 - 75.

[3.17] A. Marshall, “Power Integrated Circuits”, Encyclopedia of Electronics, Wiley 1999, Volume 16, pp. 699 - 706.

[3.18] Appendix 1: (3.18)

[3.19] K. Watabe, et. al., "An 0.8μm High-Voltage IC Using a Newly Designed 600-V Lateral P-Channel Dual-Action Device on SOI" IEEE Journal of Solid State Circuits, Vol 33, No. 9, Sept. 1998, pp. 1423 - 7.

[3.20] A. Wagpmans, et. al., "3.5mW 2.5GHz Diversity Receiver and a 1.2mW 3.6GHz VCO in Silicon-On-Anything" ISSCC98, paper No. FP 16.3:A

[3.21] R. Dekker, et. al., “An Ultra Low-Power RF bipolar technology on Glass", IEDM, 1997, pp. 921 - 923.

[3.22] P. G. M. Baltus & R. Dekker, “Optimizing RF Front Ends for Low Power”, IEEE Proceedings, Oct 2000, pp. 1546 - 1559.

Chapter 4: SOI Modeling

4.1 Modeling Introduction

Even the simplest IC designs begin with circuit simulation. This requires accurate models and simulators. We discuss the details of modeling and simulation for SOI circuits. Emphasis is placed on SOI modeling for circuit design, but the following list summarizes some of the CAD tools available for development, design and layout (appendix 2):

Process simulation (using such tools as Suprem and PROMIS 1.6, a two-dimensional process simulator),

Device simulation (such as Pisces, which converts the output of the process simulation into I-V characteristics and MINIMOS 6.1, a two- and three-dimensional MOS simulator.),

Model extraction (for example UTMOST III, which extracts SPICE parameters from I-V characteristics),

Circuit simulation (SPICE, ELDO),

Layout verification and Parasitic extraction (for example SAP 1.4, a 2 & 3 dimensional interconnect simulator, RAPHAEL, SIMPLEX fire & ice)

Circuit simulation requires both SOI specific transistor models and SOI enabled simulators. SOI specific transistor models include the LetiSOI model, and the industry standard BSIM-PD models. Most of these are similar in intent, but biased towards particular aspects of SOI simulations. As a result they have different modes of execution, especially in terms of the physical effects taken into account and the number of parameters.

SOI simulation, being relatively new, is still being optimized to make it as robust as bulk models. SOI has resulted in the generation of an entirely different set of modeling problems when compared to bulk. There are several models that can be used and various simulator engines. The major ones include Eldo^tm and a variety of versions of SPICE (such as Berkeley SPICE, HSPICE^tm, PSPICE^tm and SmartSpice^tm). Among the most prominent models are: BSIM-PD, Leti SOI, UFSOI and STAG.

Most early versions of the major simulators are not very compatible with SOI simulations. However, since about the year 2000 the situation has been much improved. More recent releases appear to work well with SOI, but remain three to four times slower than simulations with bulk models. Some problems are worth considering in greater detail. These may apply to certain versions of simulators. The default GMIN is, in many cases, le-12. (For definitions of these BSIM parameters, see the end of this section). If the body is floating, and GMIN is applied to all nodes, then GMIN must be modified, typically to about le-20 to not affect body voltage. Applying the default 1e-12 increases body voltage in the DC state by several tenths of a volt. Increased body voltage leads to apparently higher performance circuit simulation results than observed in silicon, by 1-2%. It also increases drain-source leakage, and IDDQ of the entire chip. Figure 4.1 is a schematic of a simple body bias test, applied to determine the effects of GMIN on the body node. A GMINDC value of le-12 initializes the body voltage, and a GMIN value of le-20 is used for the remainder of the transient run.

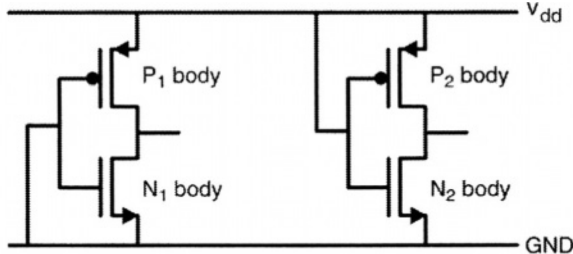


Figure 4.1: Schematic of the test configuration used to determine the effects of GMIN on body bias.

Figure 4.2 shows the effect in a partially depleted BSIM-PD model, using the Hspice 99.2 release. As can be seen, at a GMIN of 1e-12 the pMOS body is biased to about 0.22V below supply. With GMIN= 1e-20, the pMOS body bias is reduced to 0.15V. In some cases a higher value of GMINDC is required to achieve DC convergence than is required to run the transient, which is not a concern in bulk material. However, as seen from figure 4.2, difficulties can arise in PD-SOI, where the GMIN value can interfere with

the circuit, causing the body voltage to stabilize to a wrong value. Figure 4.3 shows how the body voltage transient perturbation is translated into leakage variation. This GMIN induced variation can, in most cases, be resolved by estimating the body voltage, and applying a ".force" statement. Note, however, that even with a GMIN of 1e-20, small devices may be biased incorrectly. Hspice 99.4 does not apply GMIN to an unconnected floating body node relieving this problem somewhat in HSpice.

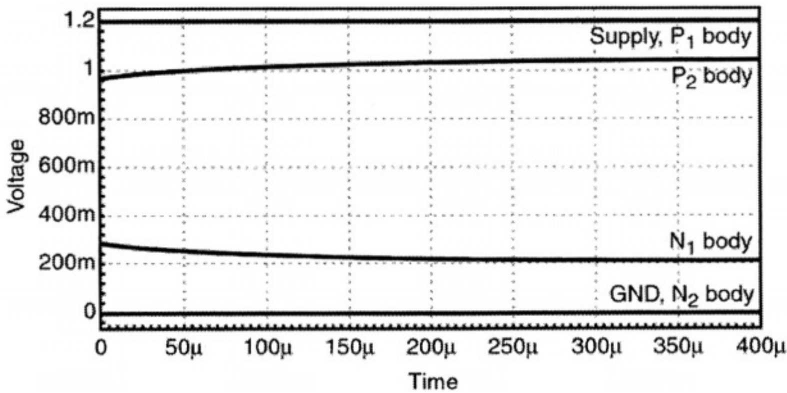


Figure 4.2: Graph showing the body voltage of off-state devices under GMINDC and GMIN conditions of 1e-12 to 1e-20 respectively, showing significant variation over hundreds of microseconds (for referenced components see figure 4.1).

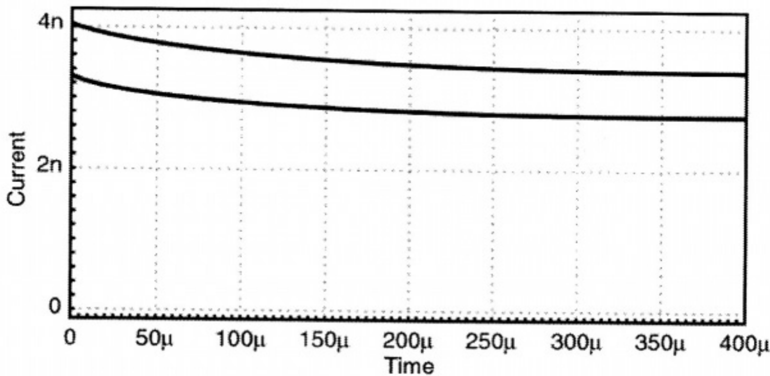


Figure 4.3: Graph showing the drain-source leakage of off-state devices under GMINDC and GMIN conditions of 1e-12 to 1e-20 respectively, showing significant variation over hundreds of microseconds. (Upper trace is nMOS leakage, and lower trace is pMOS leakage. Both devices have leakage that is 18-20% higher at GMINDC of 1e-12, than the GMIN value of 1e-20).

Careful choice of options cards provide improved results, particularly with respect to the body voltage nodes for longer simulations, compared to the values for their bulk counterparts. The values that have been tested experimentally, and found to give reasonable results are:

ABSTOL: This is the absolute node voltage error for DC and transient analysis, set to between 1e-18 and 1e-24 V
ITL1: (maximum DC iteration limit) 1
ITL2: Iteration limit for the DC transfer curve set to 500
PIVTOL: sets minimum value for which a matrix entry is set as a pivot: 1.0e-24
GMINDC: Conductance placed across active device nodes to help DC convergence - make this value too low and the body voltage can be affected: range 1e-15 to 1e-20
GMIN: Conductance placed across active device nodes to help transient convergence: 1e-20
CHGTOL: Capacitance charge error tolerance: 1e-18 to 1e-20 Coulombs
RELTOL: Relative error tolerance for voltages: 1e-4 to 1e-9 V
VNTOL: Sets the absolute minimum DC voltage for DC and transient analysis: 1e-6 to 1e-16
METHOD (Numerical integration scheme): GEAR

4.2 Example SOI SPICE deck

Figure 4.4 shows an example of a SPICE deck for a simple ring of nine oscillator. The main differences between this and the bulk equivalent are the MOS device pin out, where the fourth node is substrate, instead of rail, and the fifth node is body. The body node does not need to be connected, but if it is, some older SPICE simulators (for example HSpice before the 99.4 release) require that the body node be hooked to a dummy load, such as a voltage controlled voltage source. A sixth terminal to the MOS devices is the temperature node.

```
Simulation Inverter Ring Oscillator
****************************

*Included modelfile
.INCLUDE 'modelfilesoi'

*supplies
vdd vdd 0 dc 1.5
vss vss 0 dc 0
Vin in1 out9 pwl 0 1.5 10p 1.5 10.01p 0
```

```
*options and parameters
.options method=gear lvltim=2 abstol=1e-18 ITL1=1 ITL2=500
+PIVTOL=1e-24 GMINDC=1e-15 GMIN=1e-20 CHGTOL=1e-18
+RELTOL=1e-4 VNTOL=1e-6 LIST NODE POST
.temp 25
.param  minsd=0.5u nw=2u pw=3u

inverter subcircuit
.subckt inv out in vdd vss nbody pbody
mn out in vss vss nbody NCH w=nw l=0.18u
+     ad='minsd*nw' as='minsd*nw'
+     pd='2*(minsd+nw)' ps='2*(minsd+nw)'
mp out in vdd vss pbody PCH w=pw l=0.18u
+     ad='minsd*pw' as='minsd*pw'
+     pd='2*(minsd+pw)'  ps='2*(minsd+pw)'

*voltage controlled voltage source as dummy load for body nodes
*this is required in some early SOI simulators
eout1 nbod 0 nbody 0 1
eout2 pbod 0 pbody 0 1
ra1 nbod 0 1
ra2 pbod 0 1
.ends

*ring oscillator
xinv1   out1  in1   vdd vss vnbody1 vpbody1 inv
xinv2   out2  out1  vdd vss vnbody2 vpbody2 inv
xinv3   out3  out2  vdd vss vnbody3 vpbody3 inv
xinv4   out4  out3  vdd vss vnbody4 vpbody4 inv
xinv5   out5  out4  vdd vss vnbody5 vpbody5 inv
xinv6   out6  out5  vdd vss vnbody6 vpbody6 inv
xinv7   out7  out6  vdd vss vnbody7 vpbody7 inv
xinv8   out8  out7  vdd vss vnbody8 vpbody8 inv
xinv9   out9  out8  vdd vss vnbody9 vpbody9 inv

*transient and print
.trans 1p  400u
.print tran v(out4) v(out5) v(vnbody5) v(vpbody5)

.end
```

Figure 4.4: Example H-SPICE ring oscillator program, showing circuit and device variations resulting from the use of SOI.

4.3 Models

There are a number of model options. Most widely recognized is the industry standard BSIM-PD, but others include LetiSOI, UFSOI and STAG. Brief descriptions follow, but the reader is urged to visit the web pages of each model of interest for the latest updates on the models. The web page listing is located in Appendix 1.

4.3.1 BSIM

The Berkeley BSIM models [4.1] have become the industry standard for bulk silicon material. The BSIM4-unified (BSIM4-U) is destined to become an industry standard model for both SOI and bulk applications. BSIM-PD (Berkeley Short-Channel IGFET Model - Partial Depletion) is another Berkeley model, applicable to PD-SOI. BSIM-PD has been chosen as the industry standard SOI model by the Electronics Industry Alliance Compact Model Council. These models have several settings that are required for PD SOI models (compared to the bulk BSIM models), specifically the charge and discharge characteristics of the floating body, and parasitic effects of the diodes and bipolars, that are intrinsic to the basic MOS structure. DC charging mechanisms are derived from the various leakage paths, while AC mechanisms depend on capacitive coupling. BSIM-PD is based heavily on the BSIM3 model, and BSIM4-U combines requirements of both SOI and bulk models, and includes relatively new features such as gate leakage. Table 4.1 shows modifications to the bulk BSIM3 model to create the PD capability. The BSIM-PD model uses up to about 140 parameters in its model definition. BSIM models work with most SPICE simulators.

Table 4.1: List of features added to BSIM3 in the creation of the BSIM-PD model.

- Floating body simulation capability
- Improvements to the parasitic bipolar current model
- Enhanced impact-ionization model
- Improved threshold voltage modeling
- Enhanced bulk charge formulation
- Parameters Pdbcp, Psbcp, Agbcp, Aebcp, Nbc model parasitics for devices with body-contact and isolation structures.
- External body node for modeling distributed body-resistance
- Temperature node for thermal coupling
- Improved low frequency noise model for SOI specific noise modeling
- Body effect width terms

Gate Induced Drain Leakage (GIDL) is controlled in the BSIM-PD model from the Agidl, Bgidl and Ngidl values.

Leakage mechanisms

There are numerous parameters associated with leakage, between any of the terminals. Conventionally these parameters are extracted from leakage current measurements of various length devices. Since many forms of leakage are highly temperature sensitive, it is usual practice for most leakage parameters to be fitted over the full temperature range. An exception to this is for gate leakage, where often a single temperature fit is normally deemed acceptable.

Source-to-body and drain-to-body diode currents are modeled in basically the same way as the same parameters for bulk. There are approximately sixteen tunneling and diode current parameters in the BSIM-PD model. These parameters are extracted by fitting forward and reverse bias currents over temperature.

The parasitic bipolar transistor has an impact on circuits primarily through transient leakage. This is included in SOI models as a simplified bipolar structure, with parameters adjusted to account for diode current and transient pass-transistor leakage current. In practice obtaining the parasitic bipolar of a MOS device can be difficult unless a body tied device is available. A technique that has been used to determine parasitic bipolar characteristics involves connecting the nMOS gate to ground, while initially applying supply voltage to source and drain. Leakage observed as the source is pulsed to ground is mostly due to current through the parasitic bipolar. (The pMOS parasitic bipolar characteristics can be found by similar means).

The body resistance parameter is often modeled as a single lumped resistance value. This, however, is inadequate for high frequency analysis, where body resistance and body capacitance are distributed. Particularly at high frequencies this can result in the body voltage varying along the transistor. This is difficult to simulate without distributed parasitics.

Parasitic Resistance and Capacitance

Individual parasitic capacitances and resistances can be determined experimentally from silicon. Parasitic capacitances to body have a significant impact on transient and AC performance of the model, and hence precision is required when determining these values. Parasitic resistances and capacitances are also highly dependent upon device layout, and should be determined from devices as close in design as possible to the production layout. Sensitivity is often improved through the use of this type of test structure, though care must be taken when using arrayed test structures, that additional interconnect resistances and capacitances are not inadvertently added to the model.

Thermal Node
Most SOI models, such as the BSIM-PD models, incorporate self-heating with an auxiliary circuit. The temperature node (T-node) is initialized in SPICE if the self-heating selector 'SHMOD', in the case of BSIM-PD, is activated, and the thermal resistance is non-zero. The T node is treated as a voltage node.

The SHMOD flag has been included to permit self-heating effects to be incorporated and monitored. Self-heating is simulated by a parallel thermal resistance and thermal capacitance, which can be set to zero if the self heating effect is not used. Figure 4.5 shows the equivalent circuit of the simple thermal self-heating model, incorporated into BSIM-PD. The thermal resistance, R_{th}, represents the dissipation rate to adjacent thermal reservoirs. The thermal capacitance, C_{th}, represents the heat capacity of the device, and I_dV_d is the time varying power supplied to the system.

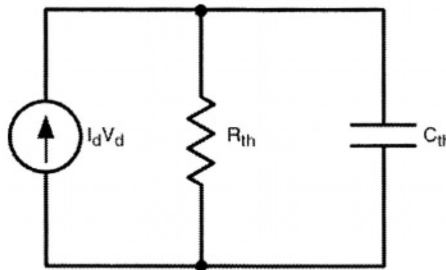


Figure 4.5: Equivalent circuit of thermal self-heating in the BSIM-PD model. Thermal resistance 'R_{th}' represents dissipation to adjacent thermal reservoirs. 'C_{th}' represents the device's heat capacity. I_dV_d is the input power supplied to the system.

4.4 Alternative Model Options

While the BSIM-PD model is the most readily available for many commercial applications, there are a number of other model options, with particular characteristics that may make them preferable in some types of simulation. In most cases information is available on the internet (appendix 1). Although many of the currently available model options are specifically for either partially depleted or fully depleted SOI, there is nothing that forces this to be the case, and compact models covering both FD and PD material have been developed [4.2].

The Leti SOI model
This is in many ways similar in terms of application and operation to the BSIM-PD model, though it is somewhat less physical, and it has some

differences in the basic model development. The Leti SOI model is an analytical and physically based I-V model for PD SOI MOSFET devices [4.3] and contains around 40 parameters. Like BSIM-PD, the drain current is defined by a single, continuous equation valid throughout the operating range, and the impact ionization related floating body effect, inherent to PD SOI devices, is incorporated into the model. The model is specifically designed for low voltage, low-power applications and analog circuit design. It will operate in the ELDO simulator and SmartSpice.

The University of Florida SOI (UFSOI) models

These models run in SOISPICE (a University of Florida enhanced version of Spice2), SmartSpice and Utmost from Silvaco, and Star-Hspice from Avant!. The UFSOI (enhancement-mode) MOSFET models are charge-based with five terminals, and have a floating-body option. Models are available for FD and non-fully depleted (partially depleted) versions. The partially depleted model has an extension which also makes it compatible with bulk silicon. A double gate model is available which permits optimal design of the specific requirements of double gate technologies.

The Southampton Thermal Analog (STAG) model

The STAG model [4.4, 4.5] was designed principally for use as a tool for developing analog applications. It has the capability to determine dynamic local device temperature. An externally accessible voltage node represents device temperature, allowing simulation of heat flow between devices. A thermal contribution ZT can be added to the small-signal equivalent circuit of the MOS device (figure 4.6), which can be included in designs where thermal effects require modeling [4.6].

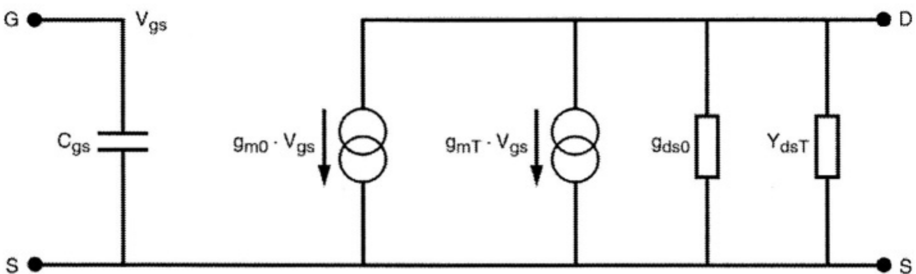


Figure 4.6: Small signal equivalent circuit of the SOI model, including temperature effects.

The STAG model is, as are all of the models considered here, charge conserving for partially depleted SOI MOS devices, and has been developed to operate with SPICE. It is available as a download from the University of

Southampton FTP server. STAG is a physically based, compact model of partially depleted SOI MOSFETs, which maintains a continuous surface potential over all regions of operation.

Southampton University has also developed a compact LDMOS model for Analog SOI applications. In common with STAG, this is a physically based charge conserving model which includes a rigorous physical treatment of self-heating for DC, transient and AC simulations and a thermal node to allow simulation of thermal coupling between the devices. Channel region doping gradient is included.

References

[4.1] Appendix 1: (4.1)

[4.2] J. W. Sleight & R. Rios, “A Continuous Compact MOSFET Model for SOI with Automatic Transitions Between Fully and Partially Depleted Device Behavior”, International Electron Devices Meeting, 1996, pp. 143 - 146.

[4.3] P. Flatresse, et. al., "A Physically Based I-V Model of Partially Depleted SOI MOSFET's dedicated to Deep Sub-micron VLSI/cMOS Technologies", PATMOS '97, Belgium, Sept 8-10, 1997.

[4.4] Appendix 1: (4.3)

[4.5] M. S. L. Lee, et. al., “A Physically Based Compact Model of Partially Depleted SOI MOSFETs for Analog Circuit Simulation”, IEEE Journal of Solid-State Circuits, Vol. 36, No. 1, January 2001, pp. 110 - 121.

[4.6] B. M. Tenbroek, et. al., “Impact of Self-Heating and Thermal Coupling on Analog Circuits in SOI CMOS”, IEEE Journal of Solid State Circuits, Vol. 33, No. 7, July 1998, pp. 1037 - 1046.

This page intentionally left blank

Chapter 5: Layout for SOI

5.1 Introduction to Layout for SOI components

Conventional bulk silicon layout has been discussed in detail [5.1, 5.2]. However, for PD-SOI, components require some specific layout rules that account for floating body use, yet permit body ties. Most PD-SOI design rules remain common to bulk design rules, the most significant difference being floating body effects, and in some cases a change from junction to dielectric isolation. Figure 5.1 shows an example of a floating body nMOS transistor layout. The pMOS is very similar, except for the diffusions being of opposite type.

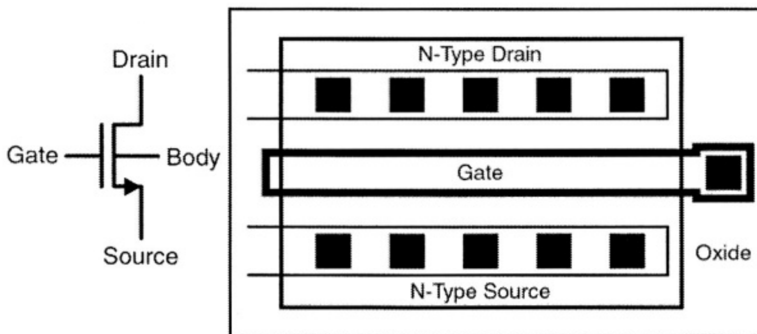


Figure 5.1: Typical layout configuration for a floating gate nMOS structure in SOI. The pMOS device is similar, but has the opposite type diffusions.

An inverter cell is shown in figure 5.2. The drains of the nMOS and pMOS are connected, as are the gates. The configuration using SOI saves some area

compared to junction isolated bulk material, due to reduced isolation requirement of the trench isolated SOI material.

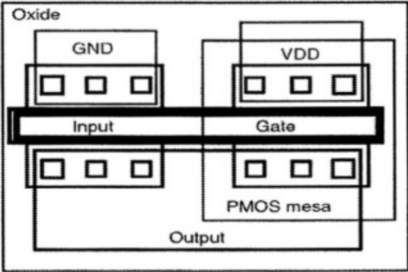


Figure 5.2: Layout configuration for a floating body inverter built in SOI. The gate can be built as a single polysilicon strip. The nMOS is shown as the left hand side MOS, the PMOS is the right hand side component.

Layout of a device with body-tied to source is shown in figure 5.3. Here the body tie is built into the source diffusion. For wide devices, two or more body-ties are required, spaced apart by a few microns, due to the high resistivity of the body region.

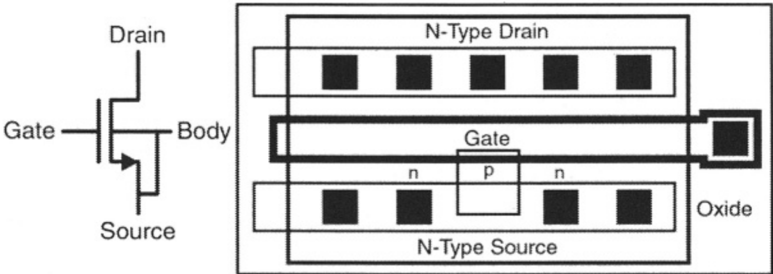


Figure 5.3: One layout example of how a body-tied to source device can be constructed in SOI. The body contact contacts to the source through the low resistivity silicided source diffusion. The body tie reduces the efficiency of the structure.

Body tied devices are often used in analog applications, as this minimizes the kink effect. At higher frequencies the RC caused by the capacitive coupling of the drain/gate to body and the distributed resistance of the body does not permit a solid connection to the entire body region. Thus some of the benefit of tying body to source is lost. At low frequencies body ties can

provide improved matching for input pairs of comparators or operational amplifiers.

Another layout requirement is for when the body tie is not connected to source. The configuration in figure 5.4 is used when the body must be tied to a potential other than the source, such as ground, supply or another body. This is advantageous in transmission gate devices, which are easier to implement if tied to a fixed supply rail.

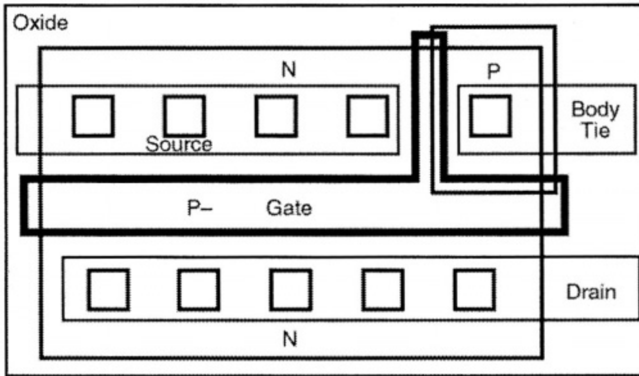


Figure 5.4: Method for producing a body tie that is isolated from the source region. Here the gate is used to mask the source region from the body tie, and the body tie diffusion is of opposite type to the source and drain. This is less efficient than the body tied to source configuration (figure 5.3).

5.2 Converting designs from Bulk to SOI

The typical, though not necessarily the best, approach to converting circuit designs from bulk to SOI, especially when modifying an existing design from bulk to SOI, is to keep as much of the bulk design as is reasonable, but redesign the sections that require it [5.3].

Circuit designs are considered in subsequent sections which use both bulk influenced designs, where many bodies are tied, and designs completely reengineered for PD-SOI, where all bodies float, or are connected in such a manner as to enhance the SOI circuit compared with the bulk counterpart.

5.2.1 Diodes

Diodes can be created using the same basic configuration as used for isolated body ties, although layout may be adapted. This is termed a gated diode, as the structure includes a MOS gate that needs to be biased appropriately to give diode operation. Figure 5.5 shows the gated diode configuration.

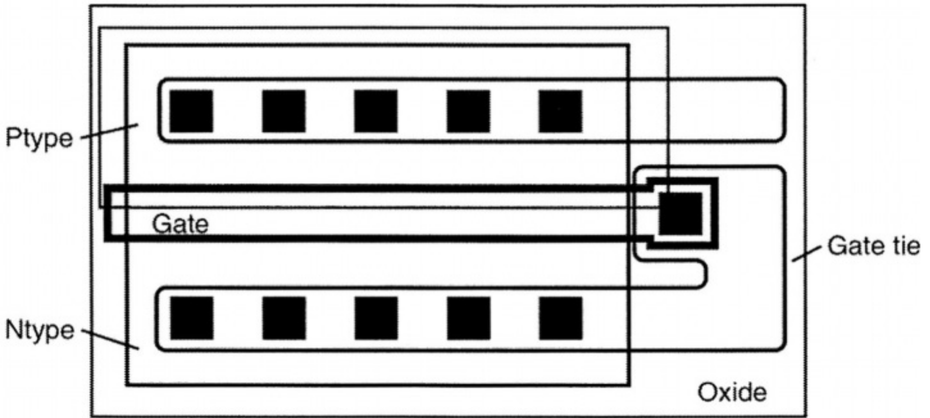


Figure 5.5: Example of a diode constructed in SOI, using the gate as the isolation mechanism between the diode anode and cathode. The gate tie is shorted to the cathode of the diode. The P-type region is the anode.

Silicide is generally included in the processing of modern ICs, to reduce contact resistance and improve MOS device performance. Silicide has the effect of shorting adjacent diffusions, unless separated by polysilicon, or the silicide is blocked.

Most processes include a silicide block option. This can be used to create a silicide block diode, which is more efficient than the gated diode structures. Figure 5.6 shows the silicide-block diode configuration.

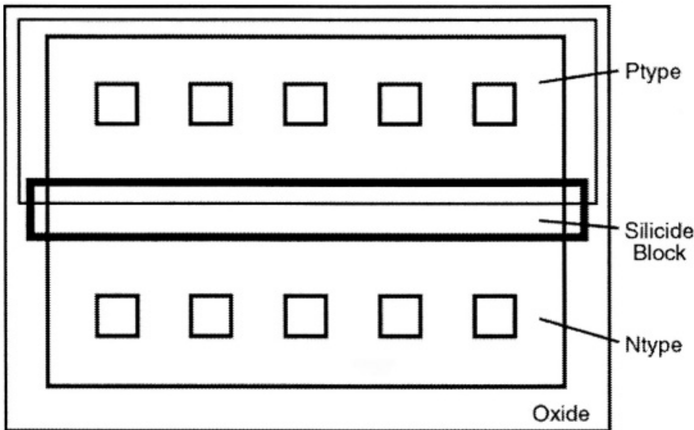


Figure 5.6: Example of a diode constructed in SOI, using silicide block as the isolation mechanism between the diode anode and cathode.

As with gated diode structures, the silicide-blocked diode structure anode is constructed of p-type silicon, and the n-type is the cathode.

5.2.2 Bipolar Transistors

In addition to basic isolated pMOS and nMOS devices, PD-SOI lends itself to high voltage and high frequency applications, and extends the options foractive and passive devices, including isolated bipolar structures. Bipolars are simply created by accessing the body contact, and connecting gate to source. Thus the nMOS and pMOS can be used as isolated NPN and PNP devices. Most of these are lateral structures, such as used for the PNP in many bipolar and BiCMOS processes. The difference being that here the gate length can be much shorter and the isolation greatly improved over the structures used in what are now mostly obsolete processes. Figure 5.7 shows the basic construction of the lateral PNP structure of a bipolar process, and figure 5.8 shows how this may be modified for use in PD-SOI. Although the most efficient variants of the bipolar structures are rather larger than their MOS counterparts, the use of bipolar devices is limited to a small proportion of the circuitry, where for example, higher precision voltage thresholds or higher frequency of operation are required than obtainable from MOS components. Due to base (body) resistance, bipolar structures are generally of higher performance if the active silicon region is relatively thick. With additional processing steps tailored to optimized bipolar structures, improved NPN and PNP devices can be created.

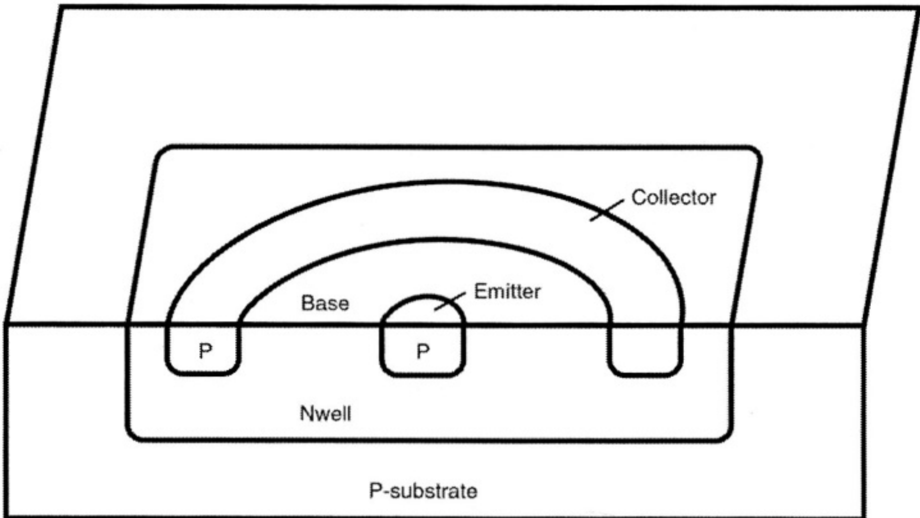


Figure 5.7: Lateral PNP design used in bulk processes.

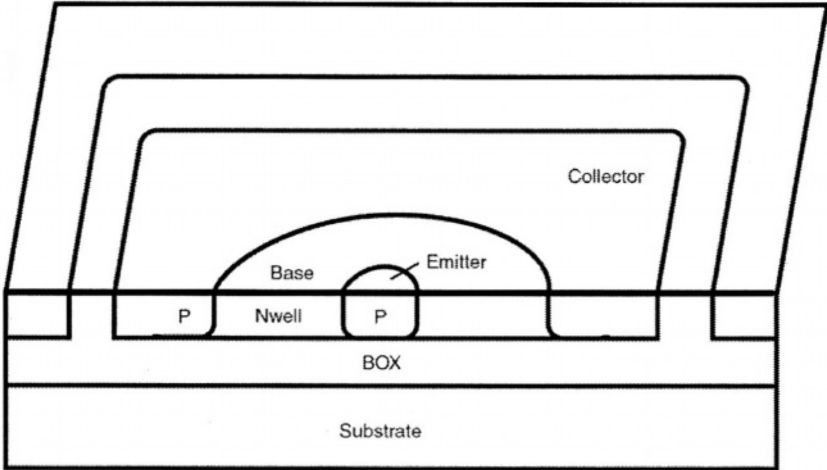


Figure 5.8: Lateral PNP design for PD-SOI.

5.3 Layout for Minimization of Thermal Self Heating Effects

Layout and circuit techniques can be employed to minimize the effect of mismatch due to two devices that have different electrical properties due to the devices being at different temperatures. Often it is possible to combine devices that require matching in the same tank. This helps equalize temperature locally within the silicon mesa. Figure 5.9 shows this for the case of a common-source device, such as a current mirror, or op-amp input pair. Here the source electrically isolates the two transistors, but since the devices are both in the same thermally conducting tank, their temperatures track.

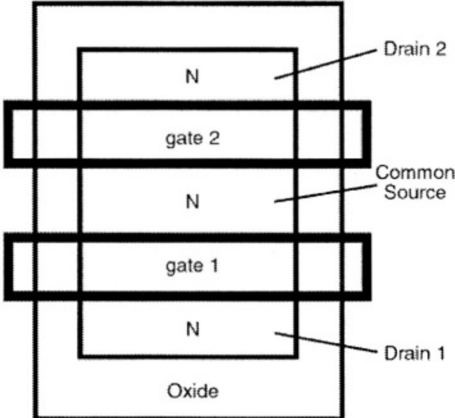


Figure 5.9: Shared tank implementation for a pair of nMOS devices. Here the source electrically isolates the transistors, but since the devices are both in the same thermally conducting tank, their temperatures track.

An alternative implementation for a common mesa also permits shared body voltage, but has specific requirements for operational conditions. Figure 5.10 shows an implementation of two side-by-side MOS devices. Here, the sources are again shared, as are the gates, so this implementation may be used in a current mirror configuration. The drawback here is that there are restrictions on the drain and gate voltages. The drains have to be very close in voltage, or in a state where a channel cannot form between the drains, preventing leakage between nodes. Although this may appear to be restrictive, it may be applicable to such as the lower stage of a cascoded mirror structure.

5.3.1 Cross coupling with thermal coupling

An enhancement to these structures that permit reasonable thermal coupling, while allowing cross-coupling in a common centroid configuration, which can be constructed as shown in figure 5.11. Here the sources are all centrally positioned within the layout and diagonally opposite drains are connected as part of the common device.

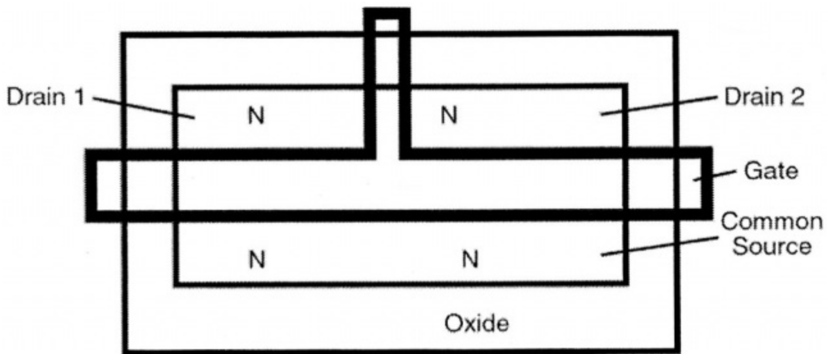


Figure 5.10: Implementation of a nMOS pair, in a side-by-side configuration. Here, the sources and gates are shared, as in a current mirror configuration.

A version of the common centroid structure can be created using partial trench to isolate the adjacent nMOS devices (figure 5.12). Here the thermal coupling is improved, but an additional process step is required for the partial trench.

Thermal coupling can also be achieved in various other components that require matching such as diode arrays and paralleled bipolar devices. There is also a benefit to placing matched polysilicon resistors over the same silicon mesa, rather than over oxide, to obtain the same thermal matching benefits.

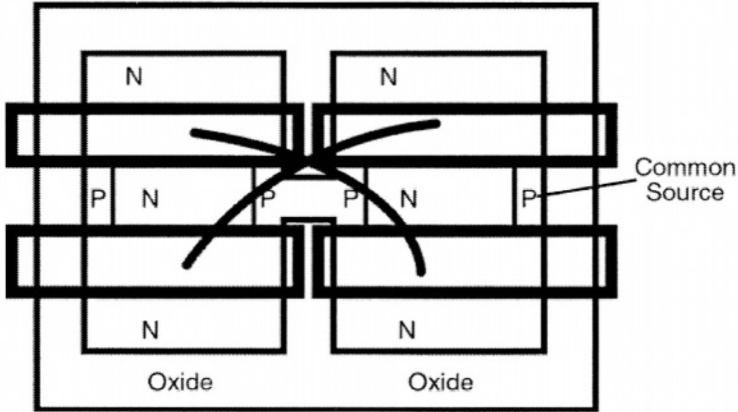


Figure 5.11: Representation of a cross-coupled input pair with common source and isolated drains. This version also ties all the body regions together with a P-type implant. For floating body implementations eliminate the P-type implant.

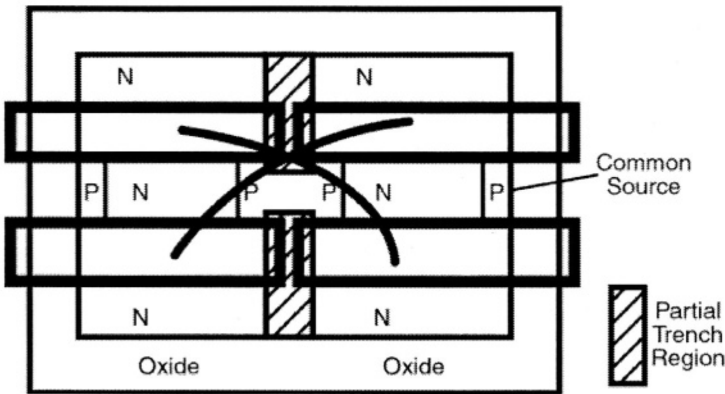


Figure 5.12: Representation of a cross-coupled input pair with common source and isolated drains, utilizing partial trench isolation between adjacent devices.

An alternative scheme, which maintains thermal matching within a mesa is shown in figure 5.13. This method is easier to use in terms of electrical circuit effects, but is not as efficient in terms of electrical matching. In this configuration transistors in the same tank are isolated electrically using gates biased to ground for nMOS structures, or to supply for a pMOS. The isolating transistors do not interfere with thermal characteristics. It is a worthwhile precaution to include the isolating transistors in the circuit

simulation, to ensure that transient switching of a shared source or drain does not affect other devices on the same mesa.

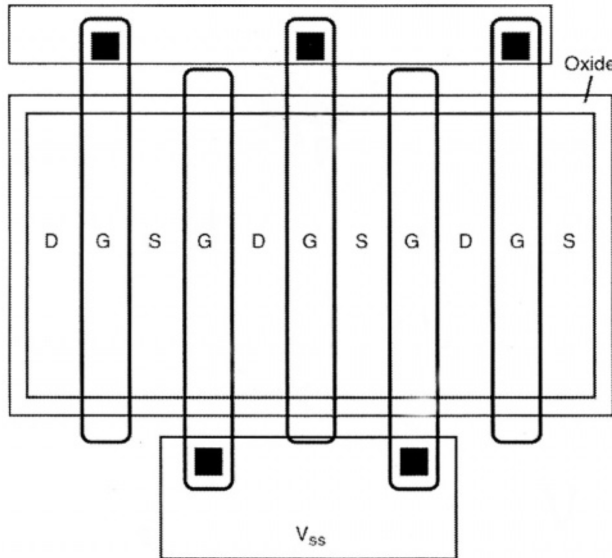


Figure 5.13: Scheme to maintain thermal matching, while permitting widely varying electrical conditions. Here the G' gates are maintained at a potential to ensure that no current can flow between adjacent source and drain nodes.

5.4 Output Stages

Output stages with high power dissipation, such as those common in class "A" amplifiers have been reported to be susceptible to local thermal effects, causing the characteristics of the driver devices to change significantly as output voltage changes. This is because of changes in the small-signal gain, including thermal effects. Self-heating causes a small reduction in gain and introduces an additional thermal zero-pole doublet in the small-signal frequency response.

Layout techniques to minimize these effects include distributed output stages, using a merged SOI with bulk technology (such as selective SIMOX), and thermal coupling through the metal levels. Some of the techniques described previously in this chapter for device thermal matching may also be applicable.

References

[5.1] A. Hastings, “The Art of Analog Layout”, Prentice-Hall, ISBN: 0-13-087061-7, 2000

[5.2] R. J. Baker, et. al., “CMOS Circuit Design, Layout and Simulation”, IEEE Press, ISBN: 0-7803-3416-7, 1998

[5.3] C. R. Tretz, "Circuit Design in SOI: Concept of Floating Beta Ratio", 2000 IEEE International SOI Conference, Oct. 2000, pp. 4-5.

Chapter 6: Static SOI Digital Design

6.1 Introduction

SOI benefits static logic circuits due to the reduced parasitic capacitances, better sub-threshold and trans-conductance characteristics. Device density of a SOI CMOS circuitry can be enhanced substantially over bulk CMOS circuit. NMOS and PMOS SOI devices can be placed adjacent to each other to improve packing density on the chip.

Reduced source/drain junction capacitances are benefits of both partially depleted SOI and fully depleted SOI devices. PD-SOI devices achieve improvement in speed, and fully depleted SOI devices offer improvement in trans-conductance.

Figure 6.1(a) demonstrates typical speed improvement in a fanout of three NAND gate ring oscillators versus temperature for SOI. Figure 6.1(b) compares the propagation delay in various ring oscillators as a function of gate length and gate oxide thickness [L_G*T_{OX}] [6.1]. The SOI ring oscillators show a consistent improvement in speed over those of bulk. Larger circuits, such as SRAMs and microprocessors, exhibit 13%-28% performance improvement, even if the design is directly transferred from bulk to SOI without optimization [6.2, 6.3]. In order to achieve substantial performance improvement in SOI technology, the circuits have to be designed specifically for SOI.

The reduced area source and drain junctions of MOS devices built on SOI helps achieve lower leakage current, and the reduced source/drain

capacitance helps achieve lower active current, compared to bulk. Approximately 47% reduction in current consumption and 26% access time reduction (constant current assumption) is observed in SOI 16K SRAMs compared to bulk [6.4, 6.5].

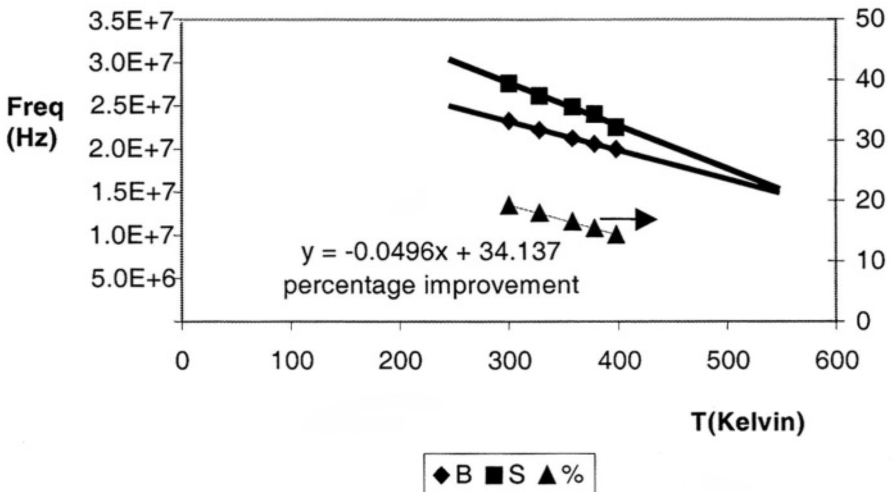


Figure 6.1(a): Frequency versus Temperature for SOI (S) and Bulk (B) Ring oscillators. Percentage improvement is also displayed.

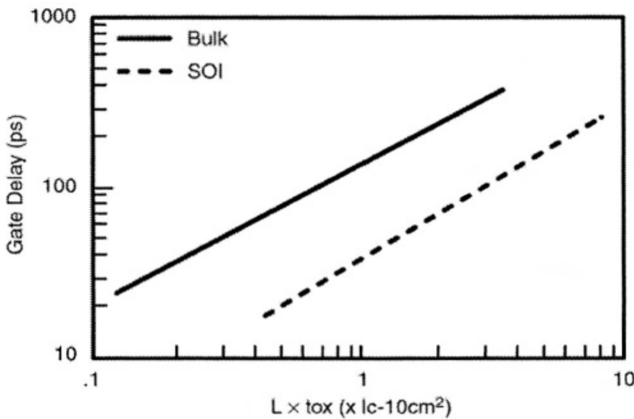


Figure 6.1(b): Propagation delay in SOI Ring Oscillators as a function of gate parameters.

Some of the critical design issues with SOI circuits are fan-out, body effect, threshold effect and the capacitance of circuits.

6.1.1 Lower Fan Out Capacitance

Fanout capacitance of logic circuits is reduced due to the lower MOS source/drain capacitances and metal to substrate capacitance. The MOS device junction capacitance is reduced with source and drain diffusions extending down to the bottom of SOI. Since the BOX is much thicker than the depletion region of the bulk drain to substrate junction, the junction capacitance is much lower. The source/drain capacitances arise from the lateral junctions facing the body. Lateral capacitances in PD-SOI can be higher than corresponding bulk capacitance. The source and drain are more heavily doped to increase the leakage of the junction, resulting in a reduced depletion region and increased lateral capacitance, depending on body voltage. The gate to body and inversion channel capacitance is nearly identical to bulk gate capacitance. The magnitude of charge contained in the body is dependent on the following factors [6.4].

* Previous state of the transistor
* Schematic position of transistor (possible source, drain voltage ranges)
* Slew rate of input, and load capacitance.
* Channel length and processing corner
* Operating supply voltage
* Junction temperature
* Operating frequency

6.2 Decreased Body Effect.

Static logic gates such as NAND gate can suffer from reduced drive capability due to body effect in SOI. Body effect is the variation of threshold voltage, V_T, with respect to the body bias in a stacked MOS device. This exists when a stacked transistor body is connected to supply rail, resulting in large reverse bias with respect to source. This increases the MOS threshold voltage and reduces drive currents. Body effect is reduced in SOI compared to bulk silicon. The result is that PD-SOI components in stacked structures have lower V_T and higher I_{DRIVE} than the same components in bulk silicon. Body effect is also discussed in chapters 3 and 10.

Figure 6.2 shows a bulk nMOS pull-down circuit. Any voltage drop across device MN1 increases the source potential of the transistor MN2 above ground. Transistor MN2 sees negative body voltage with respect to its source, making the transistor MN2's channel region hard to invert. Higher source-substrate bias (V_{SB2}) increases the threshold voltage V_{T2} and degrades performance. SOI devices do not have the body of the device connected to a common substrate. The body of transistor MN2 will drift above ground and create an inverse source follower effect (body effect) reducing voltage V_{SB2}.

With partially depleted SOI, the body will float to a potential that is between the source and drain potentials resulting in a positive V_{SB2}, reducing the threshold voltage and improving the stacked device performance.

The body node for a bulk nMOS is connected to VSS and pMOS is connected to V_{DD}. Capacitive diodes are formed from the body to the source and the drain. Capacitive coupling is not a factor in non-switching MOSFETs, since the forward or reverse bias leakage of the diodes and impact ionization controls the body voltage.

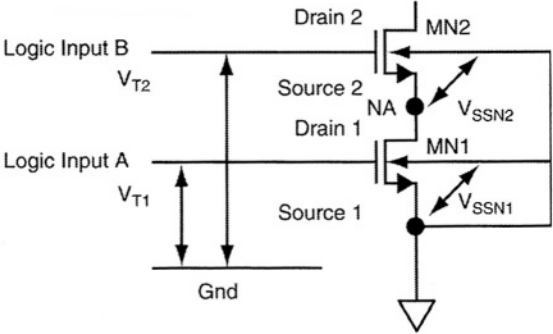


Figure 6.2: Stacked nMOS devices showing body effect in SOI devices.

6.3 Gate Leakage

As gate lengths diminish there is also a trend to reduced gate oxide thickness. This results in increased gate leakage due to tunneling. When thin gate oxides are combined with SOI, the body voltage behavior is questionable. Leakage may not cause significant effects in simple inverter circuits, since a large potential across the gate and body of a device maintains the device in the 'ON' state. However, there may be some effects on stacked stages, such as the upper nMOS of a stacked scheme, as found in NAND gates (figure 6.3). When MN4 and MN3 are 'off', both pMOS devices are 'ON', resulting in the possibility of a high potential between MN3 body and gate. This can lead to gate leakage, lowering the body potential of MN3 and subsequently slowing switching transition.

Logic circuitry can be affected by leakage through the thin gate oxide in transmission gate applications. The gate (nMOS) may be grounded, holding the device 'OFF', while source and drain are both high. Gate leakage can lower the potential of the body, and reduce any leakage between source and drain, while simultaneously increasing the threshold voltage of the device. Reduction in leakage is usually an advantage, but increase in threshold

voltage in most cases is undesirable due to the resultant reduction in switching speed. If the gate voltage has been pulled low and then the input side of the transmission gate is switched low, the parasitic bipolar transistor may turn on and partially drain the stored charge at the output node. If gate leakage has already partially depleted the stored body charge, the transient on time of the NPN is reduced.

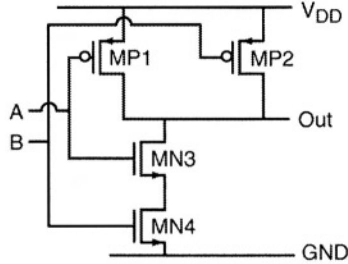


Figure 6.3: Stacked stages are found in all but the simplest static logic gates. Here MN3 can find its body voltage affected by gate leakage.

Gate leakage is caused by carrier tunneling [6.5]. This provides a virtually thermally independent leakage. Thus there may be little effect on circuit operation of gate-leakage-tolerant circuits at room temperature or higher, but significant impact at lower temperatures in circuits such as loadless four-transistor (4T) SRAM cells. Gate leakage is an important consideration in analog designs where charge storage is required. Gate leakage is usually low enough that static logic operation is unaffected for bulk designs, but may cause problems in conversion of bulk designs to SOI.

6.4 Static Inverter Characteristics

Delay of an inverter depends on the threshold voltages of the nMOS and pMOS devices. As the body voltage of an nMOS is increased (reduced for pMOS), threshold voltage decreases resulting in significant delay variations. Delay of a gate depends on the recent history of terminal potentials. This is known as history effect and impacts circuit performance.

Considering the nMOS of an inverter (figure 6.4), when the gate switches from low to high, initial capacitive coupling between gate and body raises the body voltage. Figure 6.5 shows the input low to high transition of the inverter. The nMOS body voltage is initially at 0.197V and once the input low-high transition begins capacitive coupling raises the body voltage to 0.21V. After the channel is formed, close to the saturation threshold voltage (V_{TSAT}) condition, the body is shielded from the increasing gate voltage by the MOS channel. As the gate voltage further continues to increase, the drain

voltage begins to fall (due to the nMOS turning on, and the pMOS turning off), and capacitive coupling between drain and body causes the nMOS body voltage to fall to –0.228V. Depending on the initial conditions and nMOS construction, the body voltage can switch lower than the source voltage, or remain above the voltage of the source node. If the 'high' input condition is maintained, leakage mechanisms return the nMOS body towards a DC equilibrium value of approximately zero volts.

If the gate then switches from high to low, the drain voltage raises influencing capacitive coupling which effectively raises the body voltage. As the gate voltage further reduces, the point is reached where the drain-source channel collapses. Under these conditions the drain has reached the supply voltage, and no longer causes a rise in the body voltage. Instead, as gate voltage further reduces, the gate to body capacitance reduces the body voltage. The nMOS body voltage rises to 0.4V once the gate input is completely low as shown in figure 6.5.

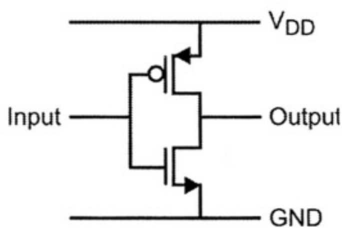


Figure 6.4: Simple static inverter configuration (bn, bp are the body voltages of the nMOS and pMOS).

Both first and second switches are faster than bulk delays since the body of the nMOS is above ground. The second switch is slower than the first switch. This is due to the fact that the body voltage is above ground during the first switch and is negative before the second switch.

The transient leakage can be minimized by modulating the drain-body and gate-body capacitance at a given temperature and voltage. Exact body voltage characteristics vary over temperature and supply voltage, since diode leakage is strongly temperature and voltage dependent.

Variation in delay across a logic circuit can be measured with respect to a varying input period or duty cycle. The SOI-attributable proportion of delay variation is not significantly different to other sources of uncertainty in circuit delays, such as process tolerances and effects of temperature and supply voltage variation. Delay variations due to SOI can be handled in a similar manner to other delay sources [6.6].

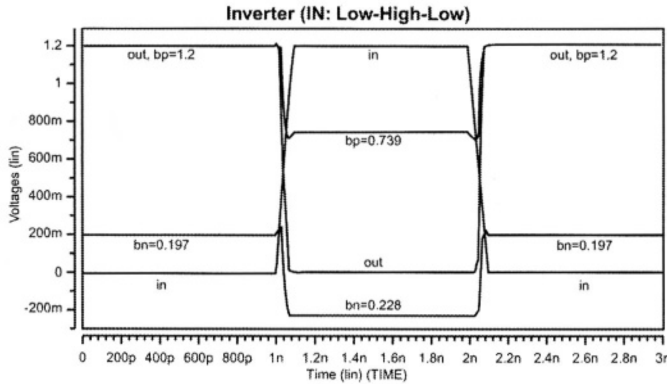


Figure 6.5: Inverter switching characteristics. First Switch 0->1 from an input DC condition, 2nd Switch 1->0 is immediately following first switch.

The average body voltage depends on many external and device-intrinsic characteristics, including duty cycle, supply voltage, physical characteristics of the device and temperature. Variability can be handled by assuming uncertain history effect. An initial maximum or minimum voltage condition is first established on the MOSFET based on its circuit configuration. Several approaches have been proposed:

(1) Full uncertainty simulation is where it is assumed that nothing is known regarding the previous condition of the circuit.

(2) Modified accessibility is used when certain switching conditions are known, such as an internal clock running at a faster rate than the DC relaxation time. This involves assessing the active nets, and re-evaluating the body voltage options accordingly.

(3) Stochastic techniques define the probability of the body state [6.7]. Although considered most accurate, due to complexity of deriving the probability state, stochastic techniques are generally mainly of academic interest rather than applicable to large scale real design situations. The reader is encouraged the read the book by Bernstein and Rohrer on SOI design concepts for additional information [6.4].

Assume the case when the inverter switches from high to low. The inverter characteristics are shown in figure 6.6. Before the inverter begins to fall low, the drain voltage of nMOS is low. Once the inverter starts to switch, the body is lightly coupled to gate, which causes the voltage dip below ground

initially (not shown) and then rises by the capacitive coupling between body and drain. If no more switching were to take place, the body voltage of nMOS would decay to the equilibrium point defined by the body-source, drain-body diode leakages and various other leakage components.

An artifact of the asymmetric delays is that the rising pulse shrinks in width as it propagates down a delay chain. This is known as pulse stretching. Therefore, there is a minimum pulse width for a given inverter chain length. The longer the chain, the wider the pulse-width needs to be before the pulse reaches the end of the chain.

PMOS devices behave in an analogous manner to nMOS devices. The PMOS body (NWELL) is usually connected to supply in bulk material, but floats in PD-SOI. As the body floats below V_{DD}, the magnitude of the pMOS threshold voltage decreases resulting in improved performance. As in the case of the nMOS, it is possible for the pMOS body to rise above the supply voltage due to coupling from the drain.

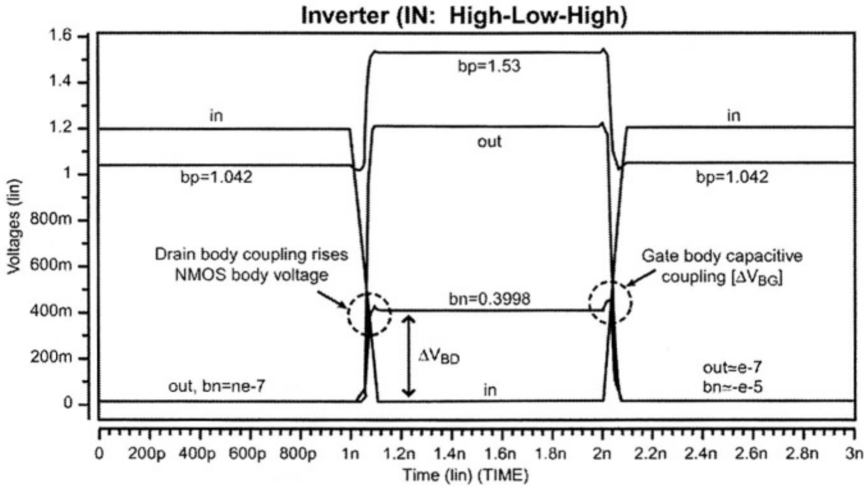


Figure 6.6: Switching characteristics of an inverter. Input transitioning high to low. The nMOS body voltage rises as the output voltage rises. The body to source junction is a forward biased diode and will dissipate charge over a period of time. The final nMOS body voltages settles at 0.4V and the pMOS body voltage settles at 1.53V above the supply voltage.

6.5 Body Voltages in SOI Inverters

Figure 6.7 shows the body voltage convergence of a nMOS and pMOS in an inverter for both the high-low and low-high conditions.

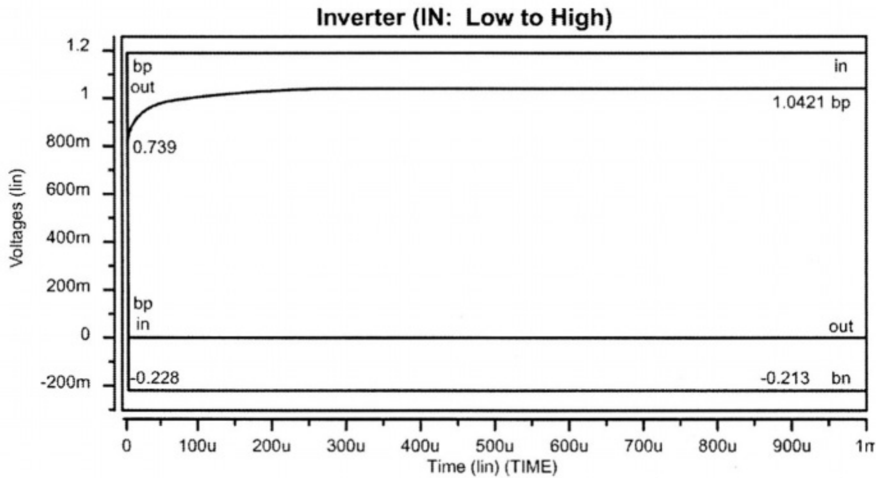


Figure 6.7(a): Body Voltages of a nMOS in an inverter

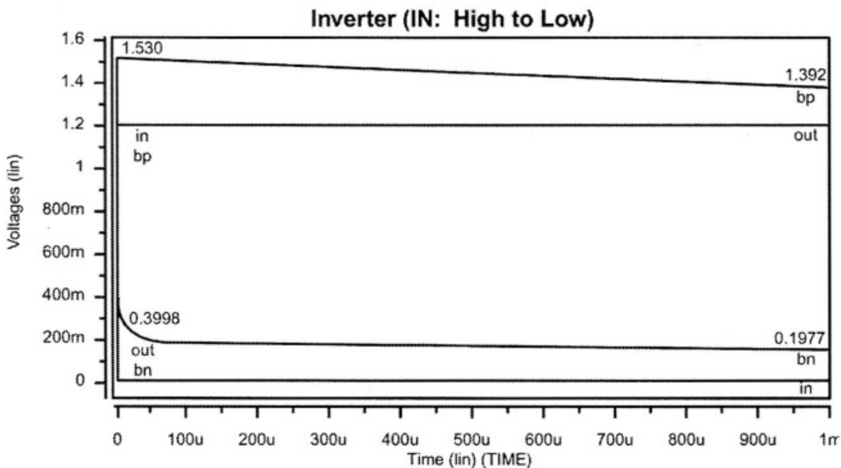


Figure 6.7(b): Body voltages of a pMOS in an inverter
The nMOS body voltage settles at 0.1977V and the pMOS body voltage due the low to high switch settles higher at 1.0421V from the initial value of 0.739V.

When transitioning from a high to low state, the body voltage begins high and within a short period of time reaches a lower DC state, whereas, the low to high state has a larger time constant. The first switch, high to low cannot retain the body voltage due to the forward biased body to source diode (when V_{OUT}=high, and V_{IN}=low).

6.6 Body Voltage Convergence

The speed of any given SOI circuit depends on the body voltage of the transistors, which in turn depends on its recent switching history. During any given circuit operation, most devices switch more than once within several nanoseconds and the circuit always does not return to the DC equilibrium condition. Figure 6.8 shows the body voltage of nMOS and pMOS devices starting from the two DC states of an output low and high in an inverter.

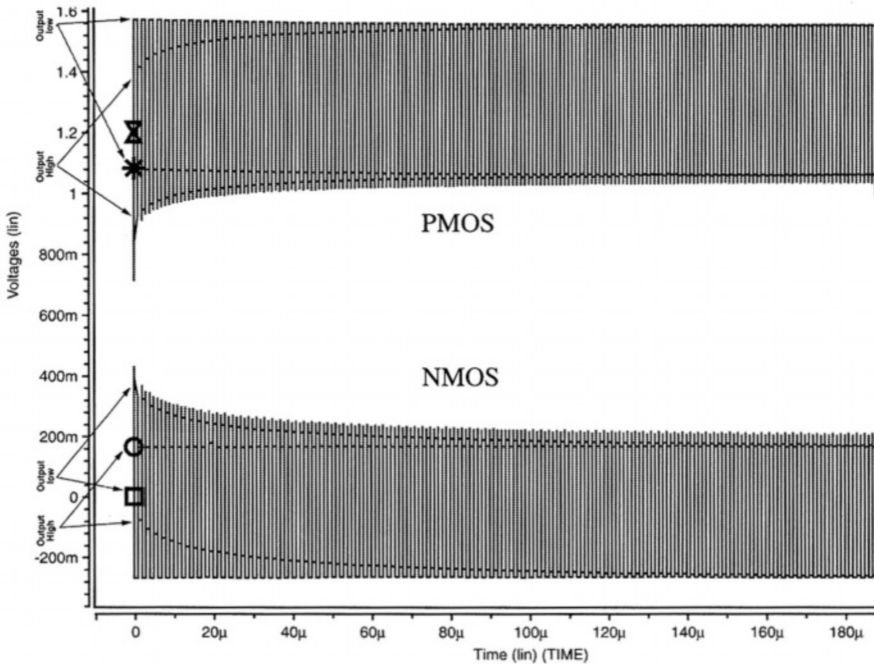


Figure 6.8: nMOS and pMOS body voltage convergence over time in an inverter.

The body voltages converge to a steady state in figure 6.8 for both the nMOS and pMOS. The nMOS body voltages converge at a peak of approximately 200mV. The circuit delay can be predicted if a free running circuit is constantly switching with the same duty cycle. Most circuits do not remain at equilibrium for long periods of time and switch every cycle. More detailed explanation on the nMOS convergence is offered by Bernstein and Rohrer in “SOI Design Concepts” [6.4].

6.6.1 Delay vs. Effective gate length

Static circuits have an active output drive and hence are relatively immune to noise. Figure 6.9 shows a delay for an inverter with a fan out of 3 in bulk,

SOI with a body contact and floating body SOI. This SOI circuit gives a 15-25% performance improvement compared to bulk. Reduced depletion capacitance and inverse body effect improves the inverter performance. V_{BS} is positive resulting in lower threshold voltage in SOI technology and hence smaller delay compared to bulk inverter. Body contacts on SOI devices can increase gate capacitance, and any extensions to the gate subsequently increase input capacitance. The body contact is detrimental to the inverter's performance and impacts area. Body contacts can be used in static circuits if they are small or cause little performance degradation.

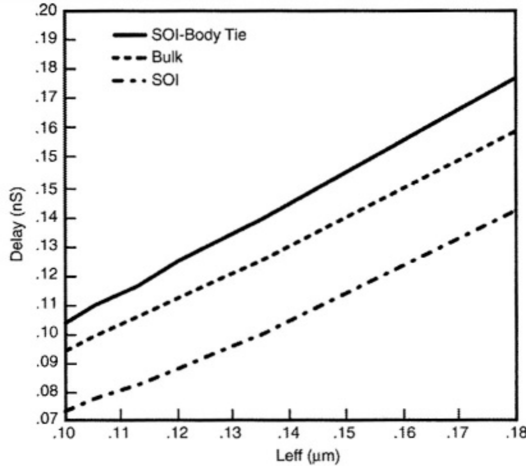


Figure 6.9: Delay Versus L_{eff} for SOI and bulk.

The body voltage depends on the previous history of circuit. The history mechanism is correlated to the inverter delay's dependence on switching history.

6.7 Noise Margin In Inverters

Noise margin is closely related to input/output characteristics of a SOI inverter. It determines the allowable noise voltage limit on the input of a gate so that the output will not be affected. Noise margin is specified in terms of two parameters: the low noise margin NM_L, and high noise margin, NM_H. NM_L is defined as the difference in magnitude between the maximum LOW output voltage of the driving gate and the maximum input voltage seen by the driven gate [6.8].

Noise mechanisms in a SOI inverter tend to be worse than in bulk due to lower threshold voltages of the devices. However, the inverter input noise can be suppressed due to higher current carrying characteristics of the

devices. These two competing factors determine the noise margin. Figure 6.10 shows the transfer curve of an inverter in bulk and SOI. Bulk has a steeper curve than SOI. As input voltages begin to rise, the first unity gain point is nearly identical for both curves, resulting in nearly identical input noise margins (NM_{IL}). However, the slope of the bulk curve is steeper and arrives at the second unity gain point at a lower input voltage than SOI resulting in the bulk inverter switching sooner than the SOI inverter. Thus, the input high noise margin ($NM_{IH} = V_{OH} - V_{IH}$) is lower for SOI than bulk. The output range of a CMOS inverter is not compromised and still switches from GND to V_{DD} [6.4].

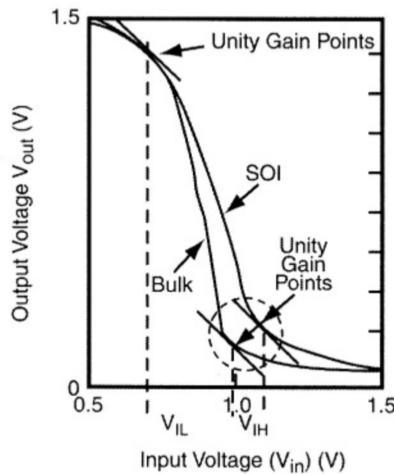


Figure 6.10: Transfer Curves for an inverter in bulk and SOI [6.9]

Wider pulse widths require shorter pulse heights to pass the input pulse to the output nodes and produce a glitch. Although the noise margin is lower for SOI than bulk, SOI's noise schmoo for this inverter is slightly higher and more resistant than bulk to wider input pulses, though they worsen as input pulses narrow below 90pS. A similar plot for a downward transition pulse shows very similar results.

6.8 Nand Gate Response

Speed advantages can be gained through SOI technology in NAND gates. The nMOS transistors closer to the output and away from the ground are prone to higher body effects and increased delays in bulk silicon. Due to the floating body, the MOSFET's are free to operate in the inverse body effect range. The body voltage of the closest device to the output will reach a steady state once the source and drain of the device reach V_{DD}.

Figure6.11shows a SOI CMOS NAND gate schematic. Simulations of a 2-input NAND are shown in figure 6.12 when input A in high and input B is switching. Input A is held initially at supply voltage V_{DD}, and input B is switched from low to high. Under steady state conditions, the body voltages of MN1 reaches a voltage between V_{DD} and V_{DD}-V_T and the body voltage of MN2 will be low.

When input B starts switching, the body couples to the gate and increases in voltage initially. After the input is fully switched high, body to drain coupling causes the body voltage of MN2 to fall to a steady state value. The initial value of the body voltages for MN1 and MN2 will not be V_{DD} and GND. The body voltage of MN2 is 0.207V and MN1 is 0.0492V after the input switch.

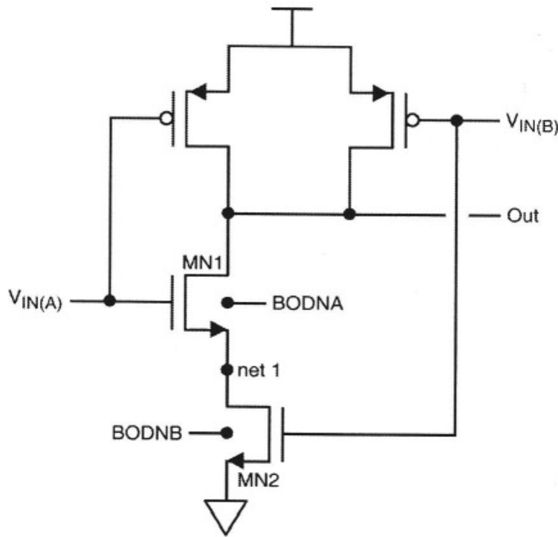


Figure 6.11: Nand Gate Schematic – MN1 is upper nMOS transistor with input A, and MN2 is the lower nMOS device with input B.

Figure 6.13 shows the waveforms when the transistor MN1 input is low. During steady state, the intermediate voltage balances the sub-threshold current between the two pull down nMOS transistors and hence the body voltages of MN1 is 0.268V, MN2 is 0.0287V and the intermediate node is 0.08V.

When the input to MN2 is low, body voltage of the two transistors are in equilibrium. Once the gate input of MN2 starts rising, the intermediate node rises due to the Miller capacitance from gate to drain of MN2.

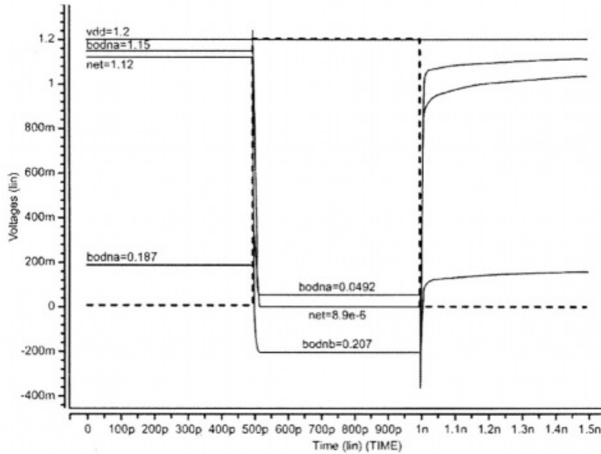


Figure 6.12: Simulation for a 2-input NAND gate. Input B is switched with input A tied to Vdd. Body voltages are shown for MN1 and MN2. The body voltage of MN2 before input of MN1 starts switching is 0.187V.

After the gate input of MN2 reaches supply voltage, the intermediate node voltage falls to ground as shown in the figure. Eventually over time, the body voltage of MN2 will drift toward ground since both source and drain are forward biased.

The body voltages of MN1 and MN2 are 0.232V and 0.0558V respectively after the input is switched high completely. The second transition on MN2 has no effect on the intermediate node or the body potentials.

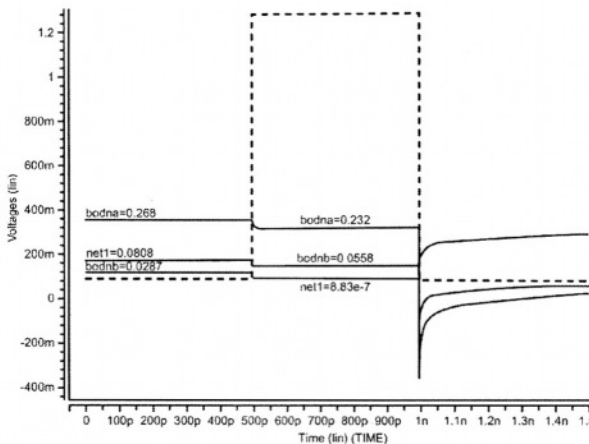


Figure 6.13: Simulation for a 2-input NAND gate. Input B is switched with Input A tied to GND. Body voltages are shown.

Figure 6.14 details the waveforms when both the NAND gate inputs are switched simultaneously. Figure 6.14(a) shows the input and the body voltages of the nMOS devices when both inputs are switching. The body potential of MN1 and MN2 during steady state is 0.268V and 0.0287V respectively. The body voltages of MN1 will rise due to the body to gate coupling of transistor MN1 initially, and during this time, the intermediate node voltage will rise due to the Miller capacitance between drain and gate of MN1, this in turn affects the body potential of MN1.

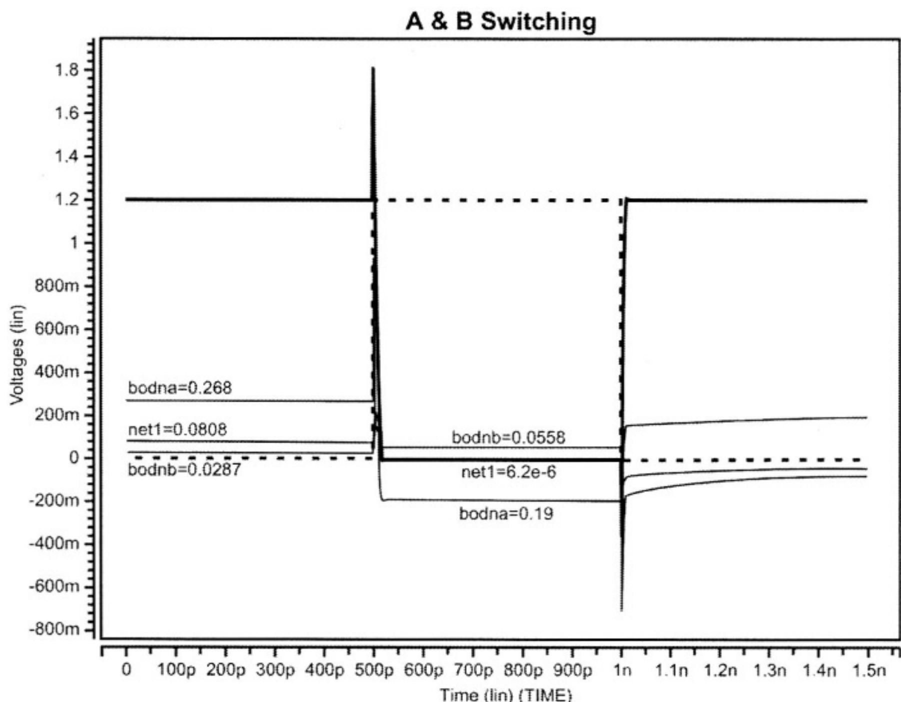


Figure 6.14(a): Simulation waveform of a 2-input Nand gate with both inputs A and B switching low to high.

When the input B is switching simultaneously, the body voltage of MN2 rises due to body-gate coupling and falls once MN2 is turned ON since the intermediate node voltage will collapse to ground when both the transistors are fully ON. The body voltage of MN1 remains higher than the body of MN2 due to body to source coupling during switching. After both the transistors are completely ON, the body voltage of both the nMOS devices are close to ground. The body voltage during the rising and falling edges of the inputs are shown in figure 6.14(b) & figure 6.14(c).

Arrival times of the signal at the two NAND gate inputs is critical. For static circuits, it is not necessary to track the body voltage of each device since their potentials are limited approximately to V_{BE} above source. Dynamically modulating the threshold voltage of the devices increases or decreases the device currents. The body potential is significantly different than bulk and the delay improves in SOI with more number of devices in the nMOS pull down stack.

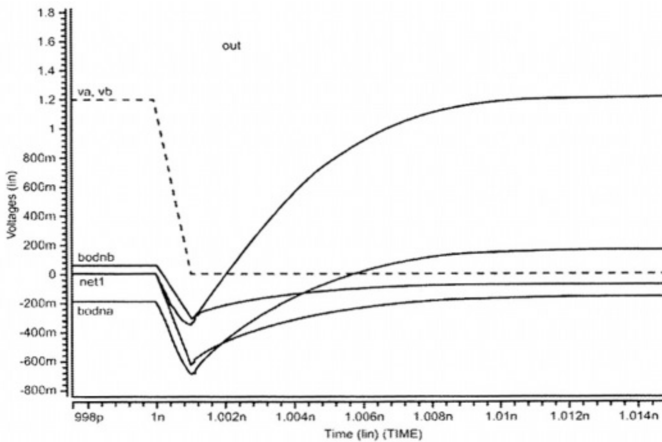


Figure 6.14(b): Body voltages of 2-input NAND at the input rising edges.

The last arriving signal should arrive at the gate input (top device in the stack) to obtain the fastest output transition. Figure 6.15 shows the delay improvement in SOI for varying stack heights without body contacts [6.4].

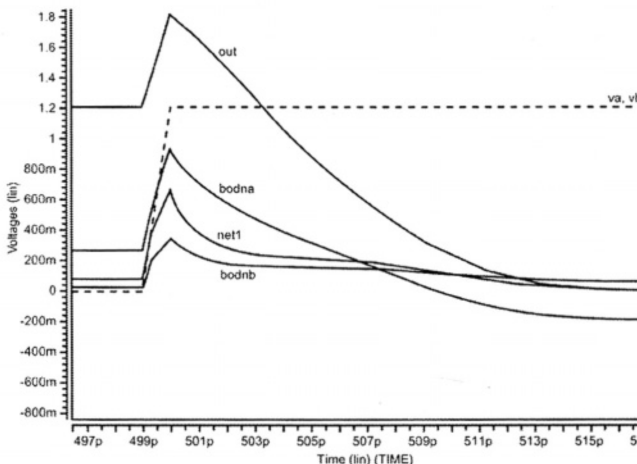


Figure 6.14(c): Body voltages during falling edges of the inputs.

6.8.1 Body Voltage Response in Nand Gates

Figure 6.16 shows the body voltages of nMOS and pMOS devices when input A is high and input B is transitioned low to high and vice-versa. The body voltages converge over a period of approximately 200μS. The body voltage convergence of MN2 and MP2 is important since input B switches.

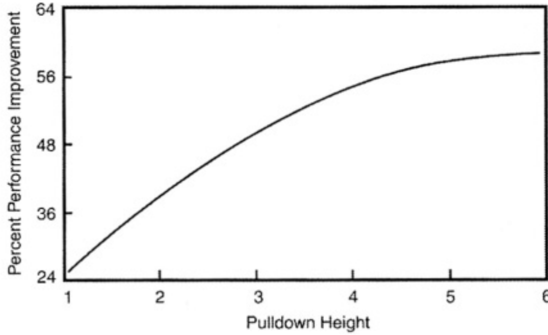


Figure 6.15: Delay Improvement vs. NAND Stack Height [6.13]

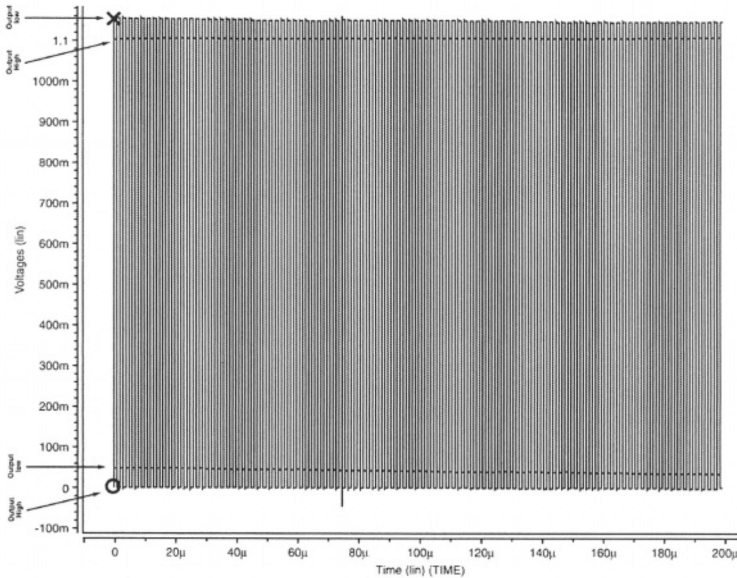


Figure 6.16(a): nMOS (MN1) Body voltage convergence over time. Note that the figures 6.16(a) through 6.16(d) show the various MOS body voltages of a two input NAND gate as switching occurs. Convergence towards a steady state occurs following an initial pulse, if a periodic input signal occurs.

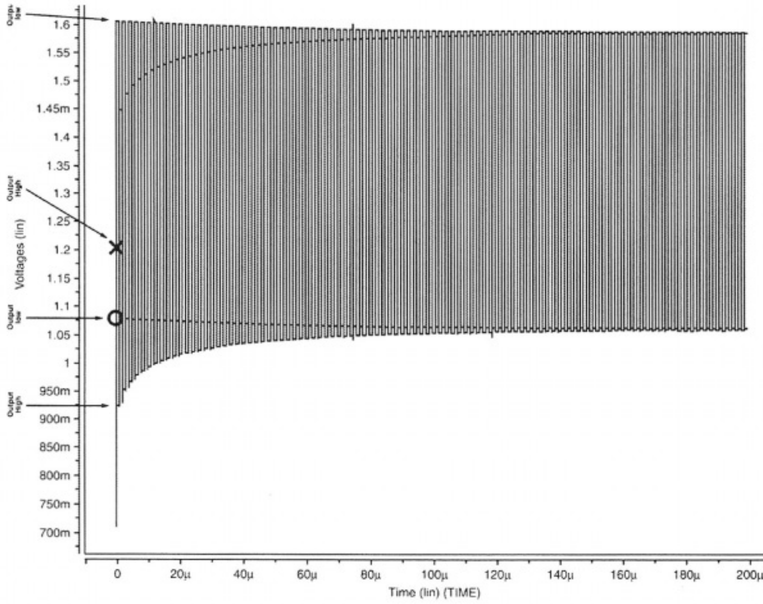


Figure 6.16(b): pMOS (MP1) Body voltage convergence over time.

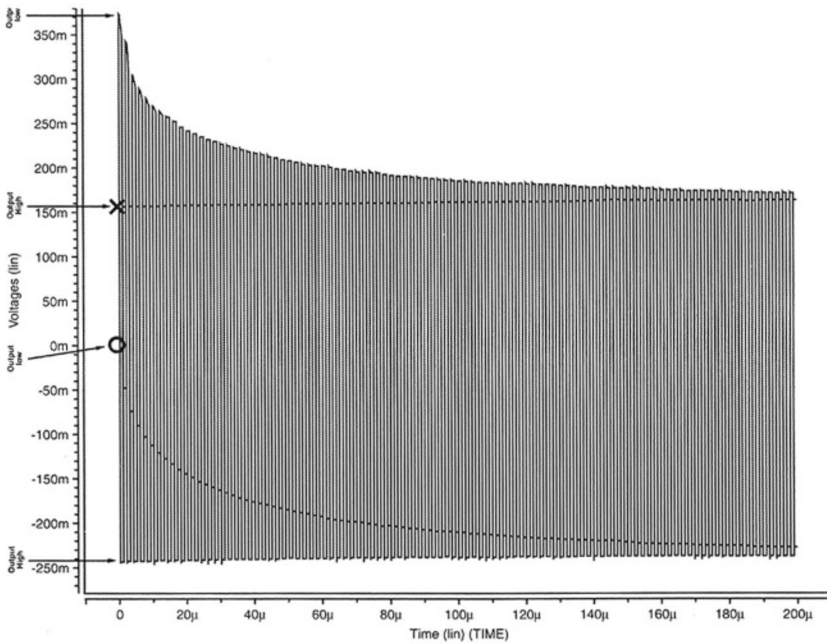


Figure 6.16(c): nMOS (MN2) body voltage convergence over time.

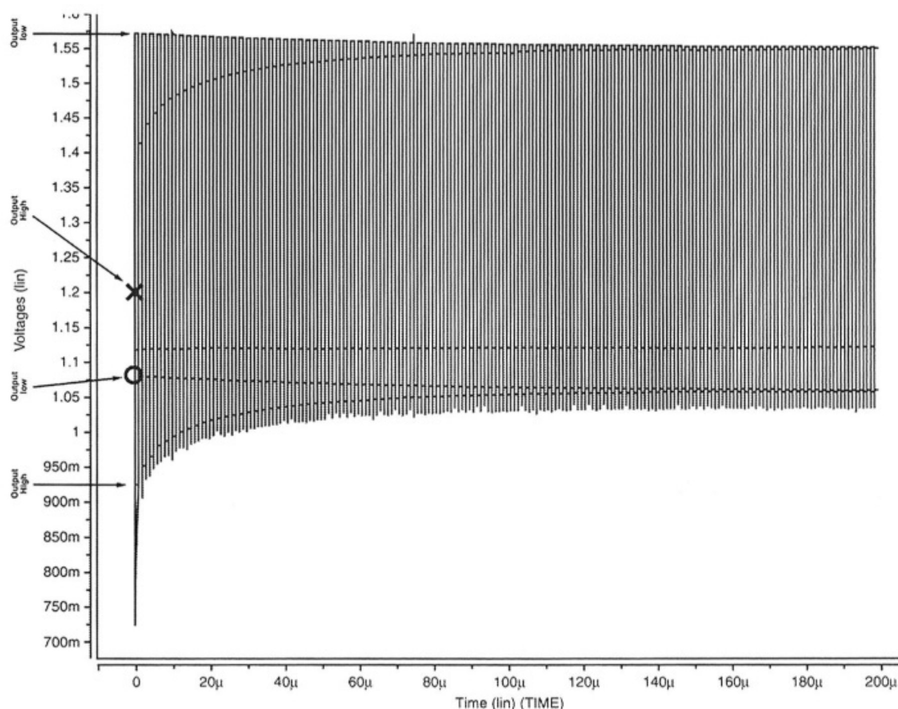


Figure 6.16(d): pMOS (MP2) body voltage convergence over time.

6.9 Nor Gate Response

Figure 6.17 shows the schematic of a NOR gate and associated simulation waveforms. When input to MP1 (B) is high, and input A switches from low to high (figure 6.17(c)), the body voltage of transistor MP0 rises for a short period due to body to gate coupling. Once the input has switched high, the body voltage settles at 1.28V, slightly higher than supply.

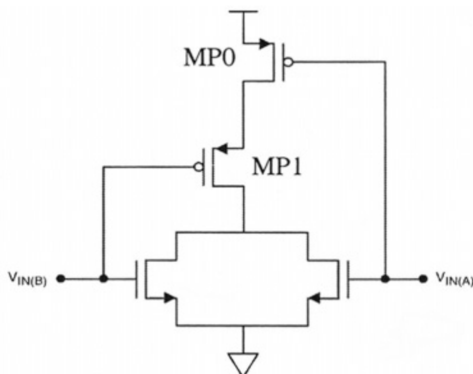


Figure 6.17(a): Two input static NOR gate schematic. MP0 and MP1 body voltages are pa and pb respectively.

When input B is low and input A is initially low (figures 6.17(b-d)), the output and the intermediate node are charged to V_{DD}. Once input A goes high, the body voltages of transistors MP1 and MP0 go low since there is a path to ground through transistor MP1 and MN1. The body voltage of MP0 goes low due to body to drain (intermediate) coupling. When both the inputs are switching, the body voltages do not drop as low as the case when input A is low [6.10].

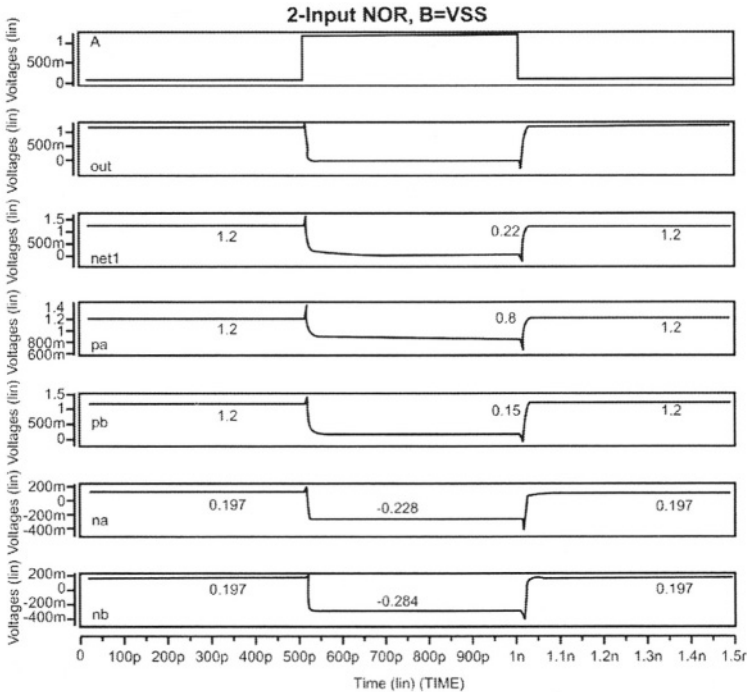


Figure 6.17(b): Waveforms when input B is at V_{SS} and A is switching.

6.10 Static OR-AND SOICMOS Circuit

Figure 6.18(a) shows a static stacked OR-AND CMOS circuit [6.11, 6.12]. Assume initially the input to MN1 is high and the inputs to MN2, MN3, & MN4 are low. Since MN4 is low, the pMOS device is 'ON' and output is at V_{DD}. The common source node (N1) to the three nMOS pass gates will be at V_{DD}-V_T since one of the pass gates is 'ON' (MN1 is ON and MN2, MN3 are OFF) and the drain side of the pass gates are at V_{DD}.

The body voltages of these devices settle between drain and source voltages. Bipolar current flows initially from gate to source for transistor MN1 and subsequently drain to source current flows after the nMOS devices are fully "ON".

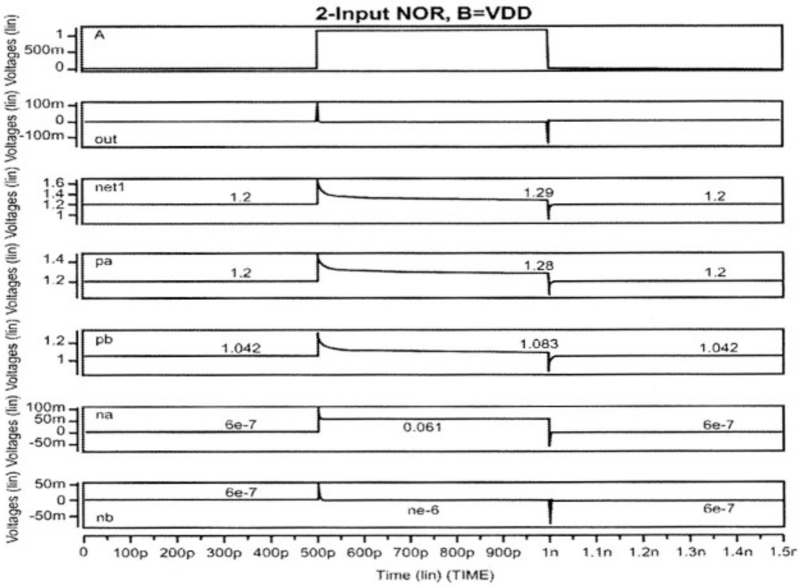


Figure 6.17(c): Waveforms when input B is at V_{DD} and input A switching.

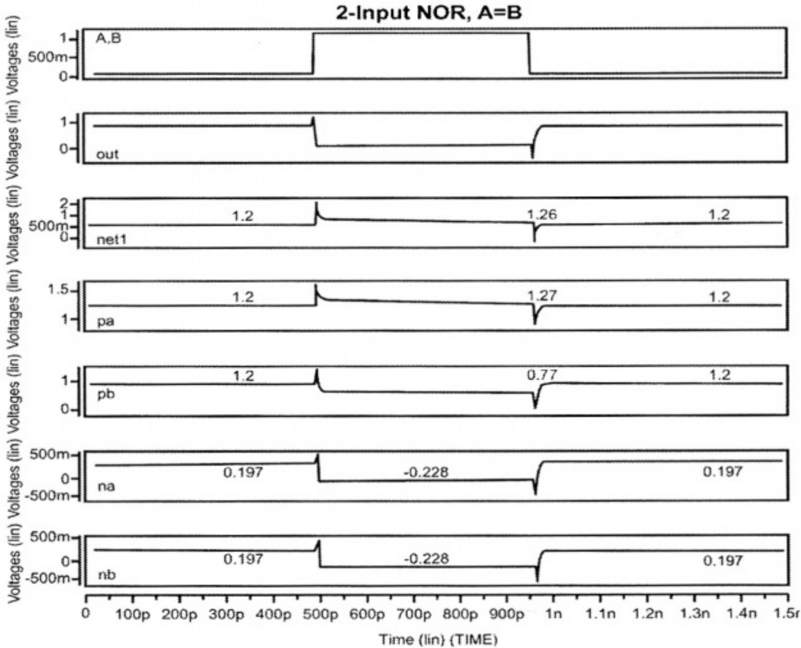


Figure 6.17(d): Waveforms showing inputs A and B switch simultaneously. The body voltage of MP0 is 0.8V and MP1 is 0.15V.

When the input to MN1 switches high to low, the body voltage of N1 is capacitively coupled down. The node voltages and parasitic currents are shown in figure 6.18(b).

Here, the body voltages of MN1 is capacitively coupled down by the large gate to body capacitance of MN1 and the gate to source capacitance. The body voltages of MN2 and MN3 are reduced slightly because their respective gate voltages remain unchanged and the voltage at source node is reduced slightly compared to MN1. When the input to MN4 switches to pull the node N1 to ground, high body source voltages are developed for MN2 and MN3 causing parasitic bipolar current.

The high body source voltage does not exist for MN1 because the body of this device is already coupled down. The body voltages of MN1 and MN2 discharge to ground after MN4 switches. For static CMOS circuit, the pMOS path restores and maintains the output. If the circuit has been properly sized, the net effect is only a very small dip in the output voltage waveform and the extra power consumption due to the parasitic bipolar current. The spikes in the output node voltage V_{OUT} is caused due to the transient switching noise.

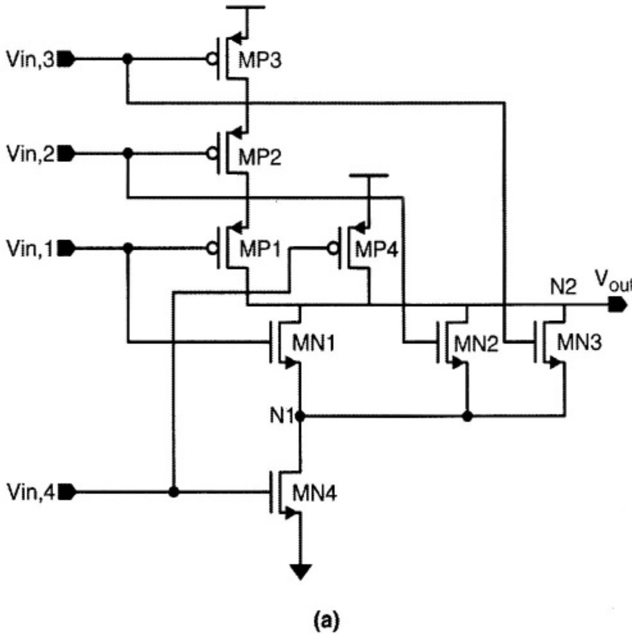


Figure 6.18(a): Static 3-way OR-AND schematic [6.11,6.12]. V1 is the node voltage at N1.

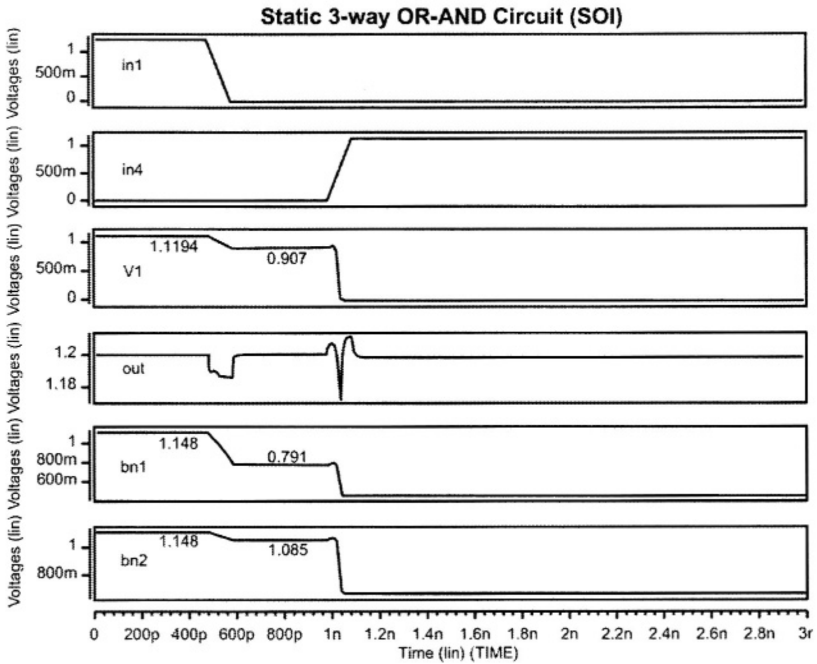


Figure 6.18(b): Switching waveforms for a static 3-way OR-AND circuit. The body voltage of MN1 drops to 0.907V after the input to MN1 switches low. The body voltage of MN2 is degraded from 1.15V to 1.085V.

6.11 XOR Gate Response in SOI

Figure 6.19(a) shows a simple static exclusive-OR gate schematic [6.10]. The node voltages and the body voltages are shown in figure 6.19(b). When input B is high, and A is switching low to high, the body voltage of the pMOS drops due to the body to drain capacitance between the pMOS MP1 and the output node.

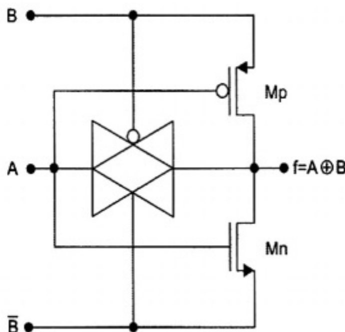


Figure 6.19(a): Schematic of a SOI CMOS XOR gate [6.10].

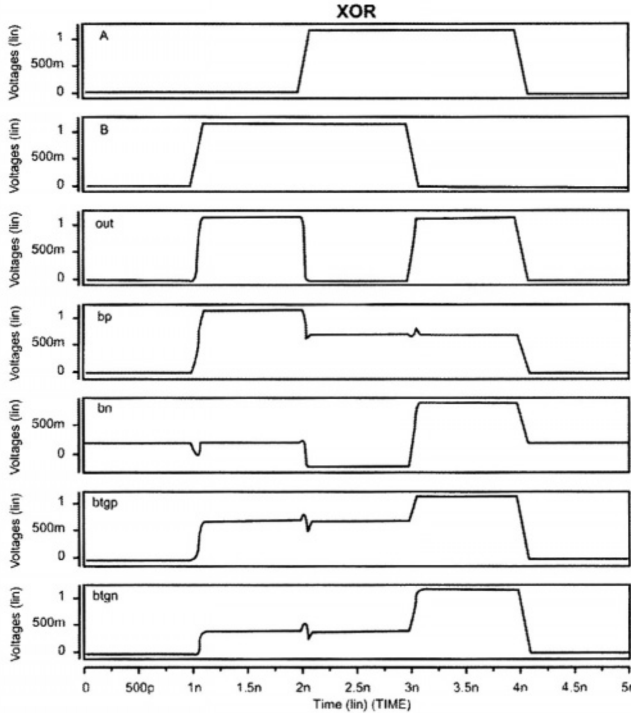


Figure 6.19(b): Node voltages and body voltages of a four-transistor XOR gate.

The body voltage of the pMOS devices is at an intermediate voltage between supply and ground once input 'A' switches to supply. Similarly for nMOS MN1, the body voltage drops below ground due to body-drain and body-gate capacitive coupling. Once the gate input to MN1 is above V_T of the device, the output node falls to ground [6.9, 6.14].

6.12 Ring Oscillator Performance

Ring oscillators have long been used as a guide to process performance. While information obtained from a ring oscillator is limited, it remains a significant comparison for figure of merits. The fastest ring oscillator is a minimally loaded fan-out of 1 inverter. More realistic in indicating actual circuit operation are ring oscillators containing NAND gates with larger fan-outs. Interconnect loaded ring oscillators are also useful in the determination of capacitive and resistive effects of the process. Figure 6.20 shows a basic ring-of-eleven oscillator circuit. The ring oscillator, with different fan-outs and interconnect loading are shown. It is common practice to also include a

'hold' function to stop the ring, and to include a frequency divider for more convenient frequency monitoring.

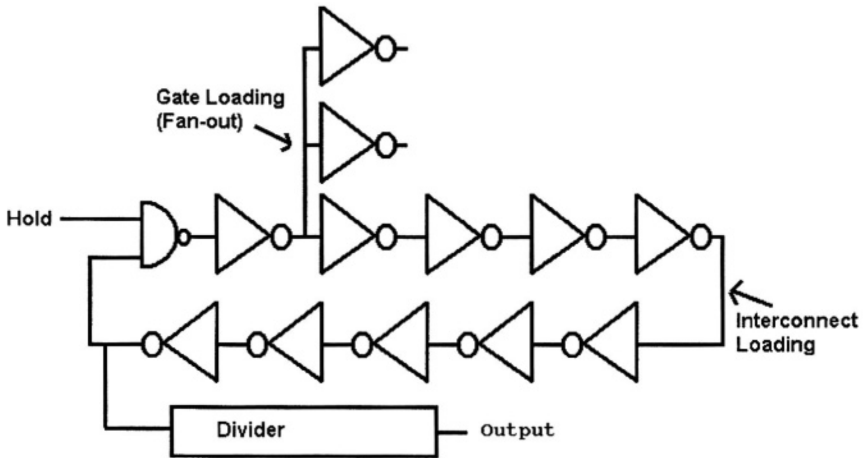


Figure 6.20: Ring oscillator circuit, indicating the different features of typical ring oscillators.

Possibly of most interest is the NAND gate ring oscillator with a fan-out of three or with interconnect loading, as this more closely simulates actual circuits than the fan-out of one inverter. The following section compares ring oscillator results from otherwise identical SOI and bulk processed material, using a layout optimized for bulk material performance.

6.12.1 Nand Fan-out of 3 ring – Performance vs. supply voltage

Figure 6.21 shows the room temperature divider output frequency of a fan-out of three Nand gate ring oscillator, for various supply voltages. This demonstrates a clear performance improvement, which varies with supply voltage as shown below in figure 6.22.

Comparative performance improvement (SOI output frequency / bulk substrate frequency)

	Supply Voltage					
	0.8	1	1.2	1.4	1.6	1.8
Nand3 - 25C	1.23	1.19	1.17	1.16	1.14	1.13

Figure 6.22: Comparison of performance improvement

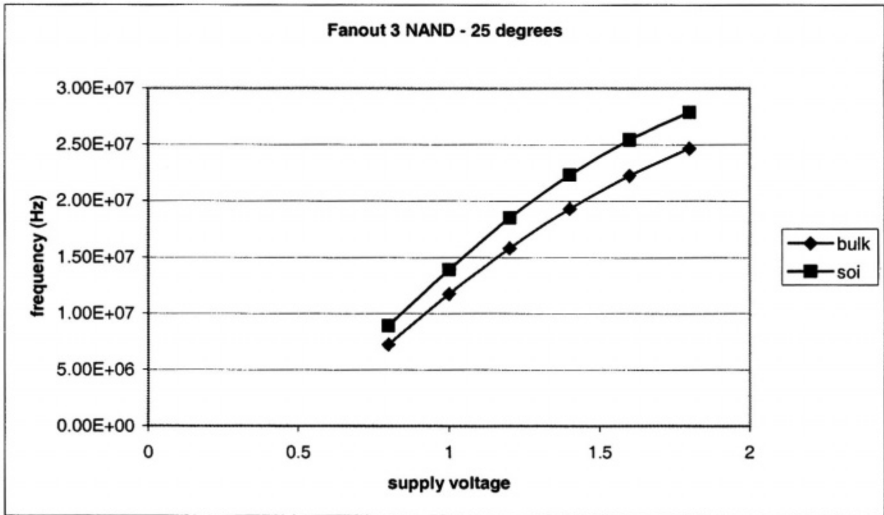


Figure 6.21: Plot of Nand fan-out of three output frequency with respect to supply voltage at room temperature.

6.12.2 Nand fanout of three – Performance vs supply operating current

One of SOI's major advantages is the reduction in power for a given operating frequency. This is used to advantage in low power systems such as digital watches, pagers etc. For the Nand fan-out of three oscillator, the room temperature results show a significant improvement reduction in power.

Figure 6.23 shows significant current reductions for a given operating frequency. At 20MHz, a 20% current reduction is observed. The data points for each curve represent, from left to right, supply voltages of 0.8, 1.0, 1.2, 1.4, 1.6, and 1.8V. When voltage is also considered, it can be determined that at the same 20MHz frequency, SOI power requirement is just 71% that of bulk silicon.

As operating frequency is reduced, current and power improvements are reduced, for example the two curves intersect at about 8MHz, indicating current requirements of the SOI and bulk rings are the same. Even here however, the voltage requirements are different, giving SOI a 12% power advantage.

6.12.3 Nand fan-out of one

We could expect the NAND fan-out of one ring to show more improvement than the NAND fan-out of three rings. This is because the fan-out of one has less baseline capacitance, thus capacitance reduction at the source and drain,

becomes a larger proportion of the capacitive effect. The room temperature NAND fan-out of one, ring oscillator results are shown in figure 6.24.

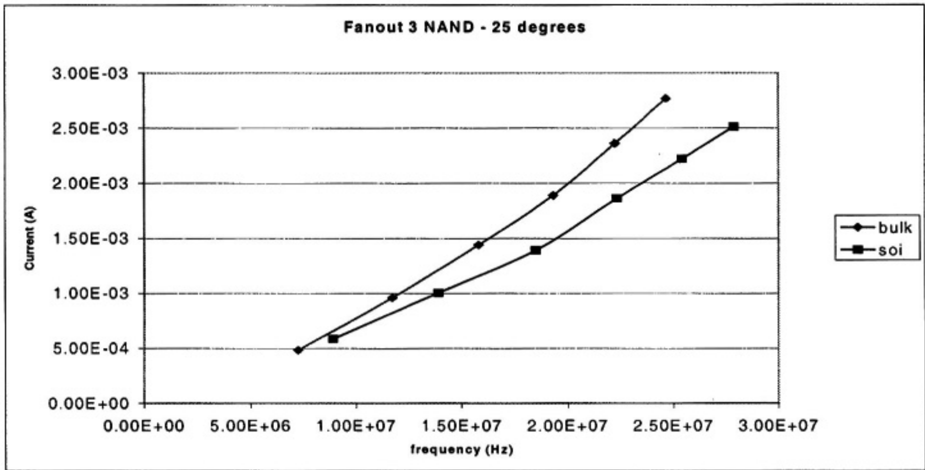


Figure 6.23: Frequency vs current. For a given frequency, current requirements are substantially reduced for SOI.

PD-SOI (frequency) performance improvement is superior to bulk. The SOI fan-out of 1 NAND gives more improvement over its corresponding bulk ring than obtained from the equivalent NAND fan-out of thee ring. At 85°C the improvement is still greater.

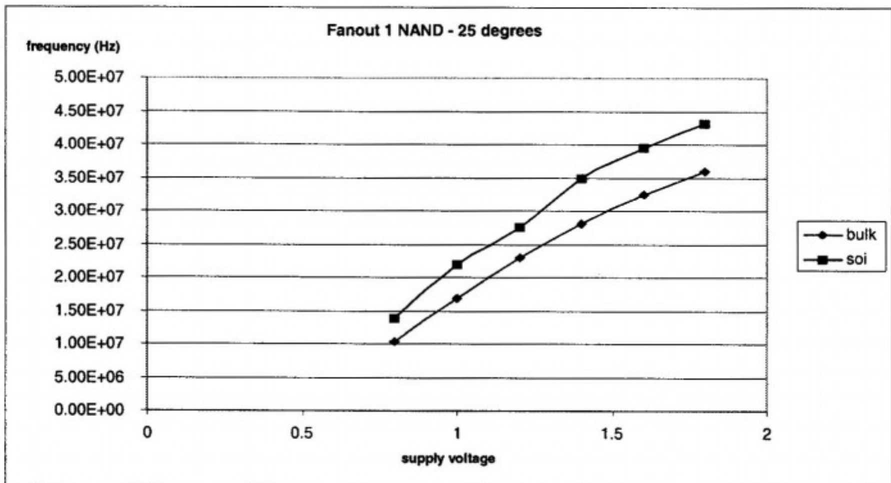


Figure 6.24: NAND fan-out of 1 structure. This shows the comparison of bulk and SOI over a range of supply voltages.

6.13 Pass Gate Response

The parasitic bipolar effect occurs when a large voltage is developed across the base-emitter junction of the parasitic bipolar transistor [6.11, 6.14]. Parasitic bipolar effects occur in floating body when the source node is pulled low due to body-to-gate or source/drain coupling. Pass gate devices (figures 6.25a,b) are susceptible to the parasitic bipolar leakage phenomenon [6.15 - 6.17].

If 'IN' and 'OUT' are high, and node C switches low, the body voltage will also be at V_{DD}. When the input switches low the body will couple onto the source, and will be pulled low. The output node "OUT" voltage degrades from 1.2V to 1.14V over a period of 5nS, and over long periods, degrades to 0V.

Similarly the body voltages of the nMOS decays over a period of several nanoseconds. The input switching "low" creates a forward biased junction diode from body to source junction, resulting in large parasitic current through the parasitic bipolar transistor. Since the body is discharged by the current flow, the parasitic bipolar current presents only a transient phenomenon.

The pertinent switching waveforms are shown in figure 6.25(c). Figure 6.25(d) shows the body voltage waveforms over longer periods. For the pMOS, the complementary situation is true. But the parasitic bipolar effect is less pronounced due to the lower impact ionization rate and smaller current gain of the parasitic PNP transistor.

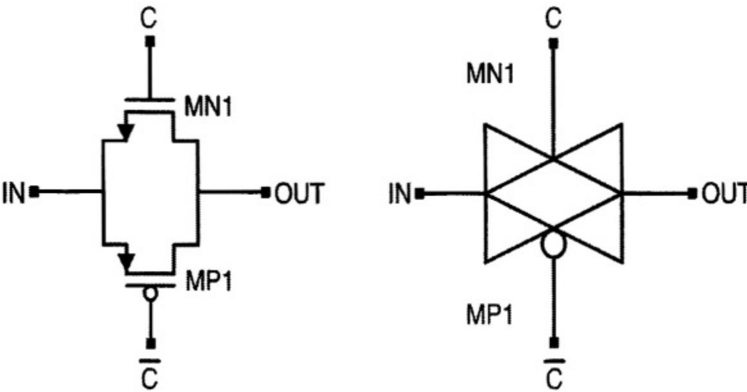


Figure 6.25: (a) Basic Pass Gate schematic (b) Circuit symbol [6.11, 6.14]

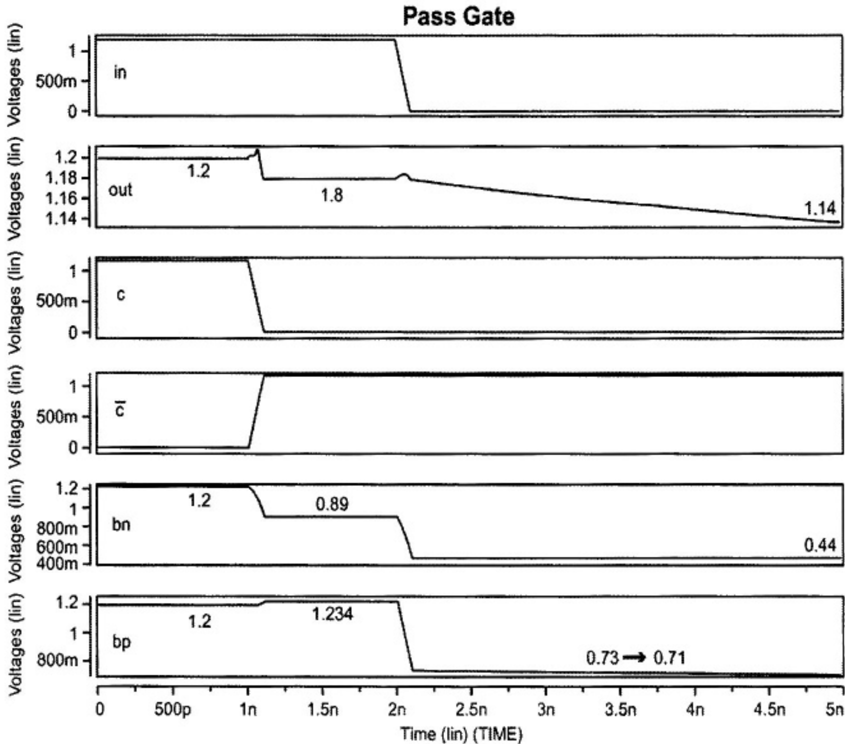


Figure 6.25(c): Switching waveforms detailing the degradation of pass-gate input signal. The output node "OUT" voltage degrades from 1.2V to 1.14V over a period of 5nS. Over an extended period this drops to 0V. The body voltages of the nMOS drops from 0.89V to 0.44V over a period of 5nS and decays much lower over longer periods

Current gain of the parasitic bipolar transistor can be suppressed with source/drain extensions. This reduces the emitter/collector area, as well as retrogade channel doping to increase the back interface doping and the effective bipolar Gummel number [6.18], the existence of the floating body require the circuit designers to examine circuit functionality and margin under various conditions.

Pass gate leakage is a device effect that can cause circuit failure. In PD-SOI nMOS devices, if both source and drain are maintained high, while the gate is low for greater than approximately 1μS, charge is accumulated in the body. Pulling down the source gives rise to a current pulse, even though the gate is OFF (figure 6.26). This current can lead to node discharge unless a keeper (or bleeder) device is used [6.19 - 6.22].

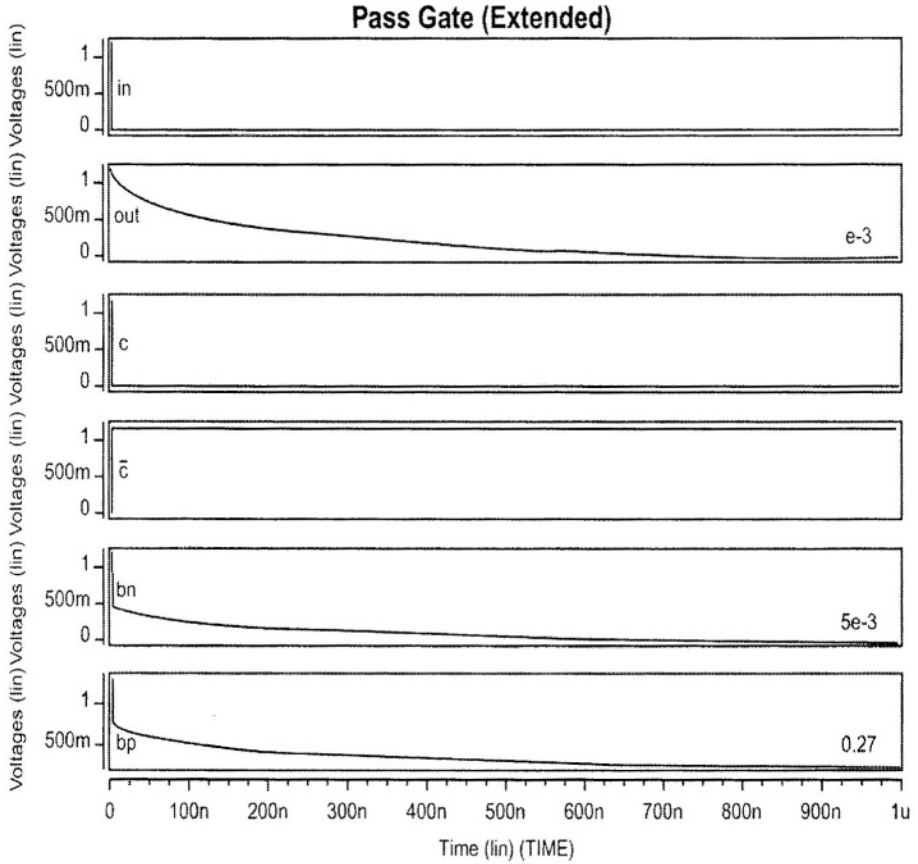


Figure 6.25(d): Pass gate body voltages over an extended time period. The body voltages decay to ground over a 1μS period, with input grounded.

6.13.1 Pass Transistor Based circuits

Pass transistors logic is area efficient. In addition, the reduced transistor count required to implement a given function improves delay and reduces power consumption.

Pass transistor logic is a common circuit choice for fast arithmetic operations such as ALU, multipliers, and processor data flow elements such as barrel shifter circuits and multiplexers. Pass transistor implementations in SOI are attractive because it potentially results in substantial power reduction and subsequently makes it useful for low power applications.

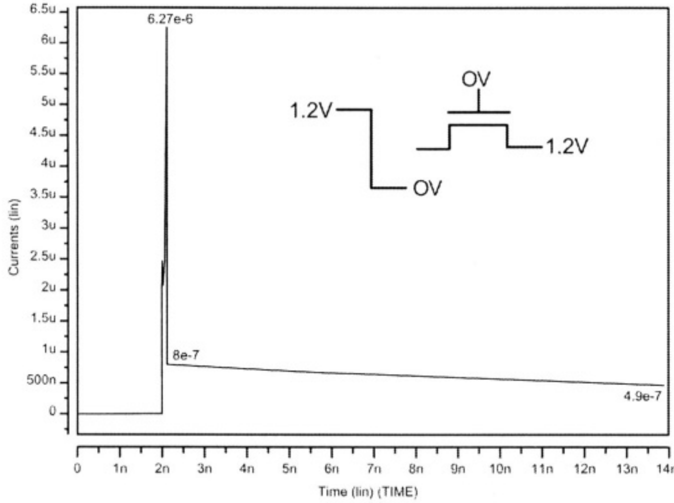


Figure 6.26: Pass gate leakage in SOI. The peak current when drain switches from high to low is 6.27μA [6.19].

6.13.2 Pass Transistors Based Multiplexers

Pass transistors based wide multiplexers are important for critical data flow elements such as rotators and shifters and, for the control portion of a microprocessor. The schematic of a 4-to-1 multiplexer is shown in figure 6.27(a).

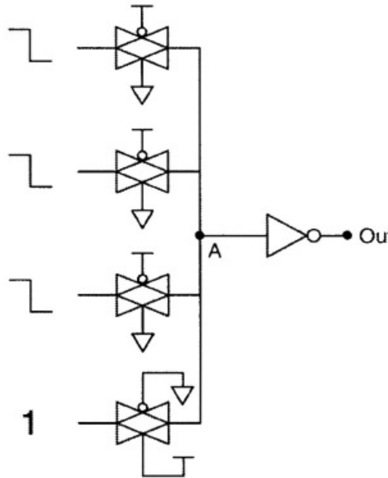


Figure 6.27(a): Schematic diagram of a 4-to-1pass-gate based multiplexer. Three inputs switch high to low, while one input remains high.

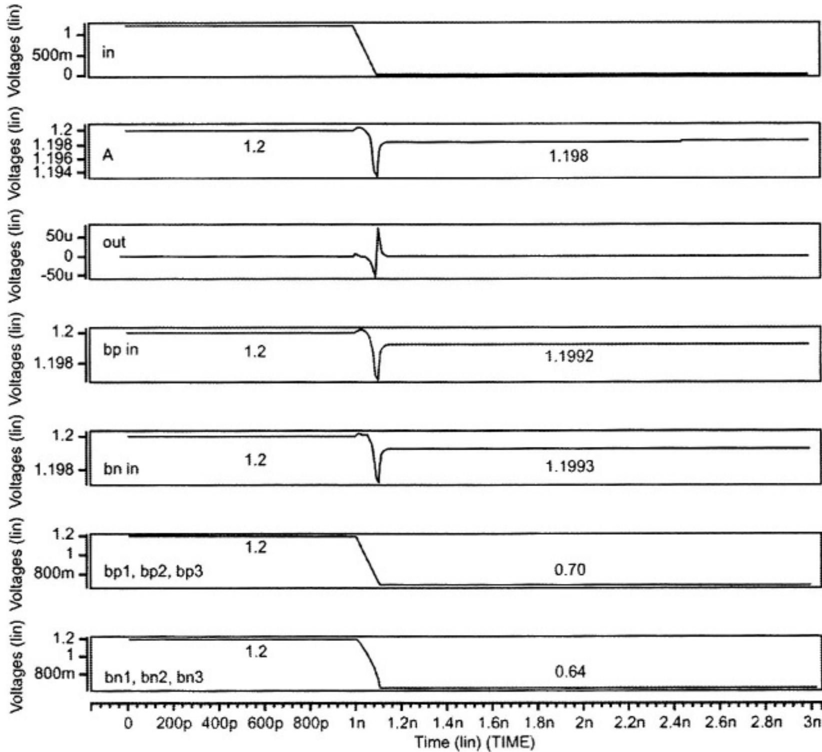


Figure 6.27(b): Switching Waveform of a 4 to 1 pass-gate based multiplexer. Body voltage is coupled down to 0.64V when the inputs switch low.

Usually the control signals are orthogonal, selecting only one input at a time. The worst case scenario for parasitic bipolar effect assumes all inputs are 'HIGH' initially with except the selected input. The selected input passes the 'HIGH' state to node N1. All three unselected inputs are then pulled low. As a result, parasitic bipolar currents flow through the three nMOS in the three unselected pass gates to pull down node N1, which is being held only by a single selected pass gate (figure 6.27(b)). For four pass gates connected in parallel, the node 'A' voltage drops to 1.198V from 2V. Body voltage is coupled down when the inputs starts going low. If more pass gates are connected in parallel, the voltage drop degradation is much higher. Since the parasitic bipolar current is a transient phenomenon, the selected pass gate eventually restores node 'A' voltage. If the output is sampled and latched into subsequent logic states when it is HIGH, an inaccurate state can be transmitted.

For static logic, reported performance improvements in inverters, two-input NAND and NOR gates are 15-20%, although this depends on the assumptions made [6.23, 6.24]. Three and four-input NAND gates give greater improvement. Dynamic circuits (chapter 7) have demonstrated improvements of the order of about 17-22%. SRAM access time is improved by approximately 15%, though this is highly dependent upon the circuit layout. Figure 6.28 shows comparative improvements for various SOI static logic gates.

Gate	Bulk	PD-SOI	% improvement
Inverter	17.7ps	15.9ps	10%
2 I/P Nand (low i/p)	33.0ps	24.9ps	25%
2 I/P Nand (top i/p)	28.8ps	22.3ps	23%
3 I/P Nand (low i/p)	51ps	35.9ps	30%
3 I/P Nand (top i/p)	40.6ps	29.7ps	27%
2 I/P Nor (top i/p)	41.7ps	30.4ps	27%
2 I/P Nor (low i/p)	33.9ps	26.9ps	21%
3 I/P Nor (top i/p)	70.8ps	48.4ps	32%
3 I/P Nor (low i/p)	50.1ps	38.4ps	25%

Figure 6.28: Comparative improvements of PD-SOI delay per stage times over bulk equivalent for a variety of logic gates.

6.14 History Dependence

"History Dependence" is the change in delay through a gate as a function of the switching history of the device (figure 6.30). As the period of the input pulse changes (figure 6.29(a)), the delay through the gate changes. Figure 6.29(b) shows the delay through an even number of NAND gates [6.25].

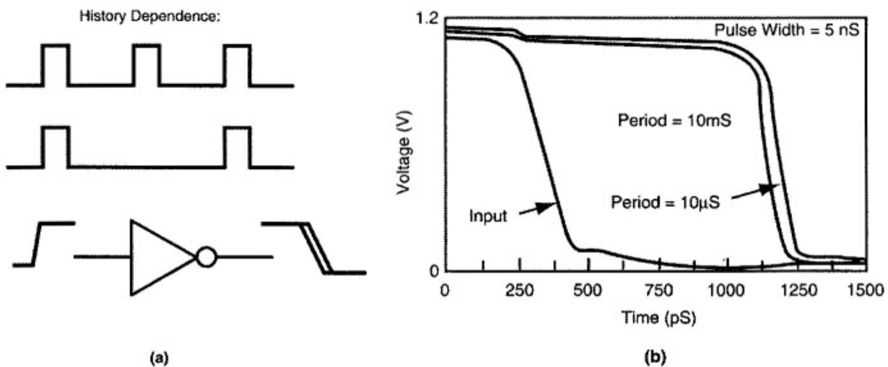


Figure 6.29: History Dependence of Delay [6.19, 6.25]

During fast switching, V_{SS} is determined by capacitive coupling to the gate, drain, source voltages and also, the gate tunneling currents in deep submicron technology. Capacitive coupling perturbs the steady-state value of the V_{BS}. During steady state, V_{BS} is determined by diode leakage and impact ionization (figure 6.30). When the nMOS in an inverter turns "ON", initially the body capacitively couples to the drain and is pulled low and subsequently converges to a value set by the drain, source and gate voltages as well as their leakages. Variations in delay through a gate are possible if another switching event occurs before the body returns to equilibrium and as the input period is changed. Figure 6.31 shows variation of V_{BS} when the device is turned ON [6.19].

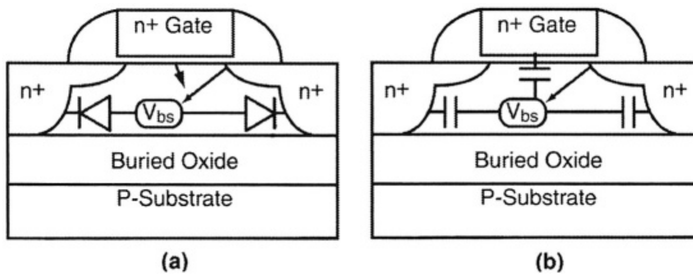


Figure 6.30: Body Biasing Elements [6.19]

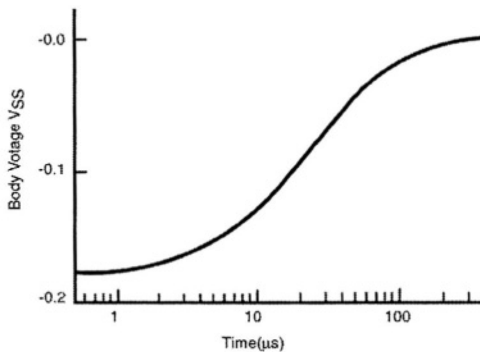


Figure 6.31: Variation of Vbs when the device turns ON [6.19]

6.15 SOI vs BULK : Performance Benefits in Digital Circuits

Because of the dynamic threshold voltage effect as well as the reduced junction capacitance, SOI CMOS circuits exhibit better performance than bulk silicon. When driving large loads, the drive currents are higher which is an advantage in SOI CMOS circuits, but the reduced junction capacitance benefit is reduced. SOI has the advantage that junction capacitance can be

scaled as supply voltage reduces. This is difficult with bulk. In addition, the floating body bias, combined with dynamic increase of V_{BS} due to capacitive coupling and the kink effect, increases the dynamic drive current. The reduced junction capacitance of SOI CMOS also implies superior SRAM performance, which translates to faster microprocessors. Furthermore, reduced SER inherent in SOI memory circuits provides a distinctive advantage of SOI over bulk silicon in SRAM applications.

Capacitive coupling in SOI can result in reduced transient off-state current. SOI offers possible benefits in circuits based on pass-gate logic. With the body floating, there is no increase in the pass transistor threshold voltage due to reverse bias of the body-source junction, which is a significant problem in bulk silicon pass gates. This feature also has implications in multiple-input gates (e.g., NAND) structures, where the number of gates can be increased considerably to realize functions more efficiently.

The performance advantage of SOI over bulk CMOS technology depends on several factors. In order to maintain the same transistor off state leakage current, SOI MOSFETs are typically designed to have higher threshold voltage than their bulk counterparts. The higher threshold voltage reduces available current drive and the performance leverage of SOI devices, especially at low supply voltages. Further, in typical logic application, interconnect accounts for 33% of the total chip load capacitance and can easily contribute to more than 60% of the delays in heavily loaded circuits. If the threshold voltage of SOI devices is set too high because of process technology and device design window concerns, the advantage of SOI diminishes and can even disappear completely [6.26].

6.16 Floating body and hysteresis effect

The floating body and hysteresis effects are intrinsic to partially depleted SOI. A thorough understanding and technology advances have allowed parasitic bipolar leakage to be accurately measured/modeled and suppressed [6.14, 6.27]. The parasitic bipolar effect can be suppressed by using non-ideal body source junctions to reduce the current gain of the parasitic bipolar transistor and the floating body voltage. It is important to confine the non-ideal body source junction leakage to well below the sub-threshold leakage of the MOSFET to avoid increasing leakage [6.13].

Scaling supply voltage reduces the parasitic bipolar effect significantly as there is less voltage across the body-source PN junction. Although, the parasitic bipolar effect might be assumed to be significant with channel

length reduction, it has been found experimentally that the parasitic bipolar effect decreases with decreasing channel length [6.26, 6.28].
The floating body induced hysteric threshold voltage variation becomes more serious as the supply voltage and threshold voltage are scaled [6.27, 6.29]. Figure 6.6 shows the body voltage of a nMOS in a CMOS inverter as a function of time through one switching period. Before the start of the switching, the gate input is HIGH and the drain output is LOW, causing the body voltage to be at 0V. As the gate voltage switches low the drain voltage rises to V_{DD}, and the body voltage rises by an amount ΔV_{BD} determined by the drain body capacitive coupling.

The body voltage is then charged by the diode current through the reverse biased drain body PN junction and off state impact ionization current and discharged by the diode current through the forward biased body source PN junction. When the gate switches high again, the body voltage follows the gate voltage rise initially, due to gate body capacitive coupling. This results in a voltage ΔV_{BD}.

Compared to earlier inverter characteristics figure 6.6 (input H-L-H), the body voltage spike increases the leakage through the forward biased body source PN junction, resulting in a loss of body charges. Similar characteristics are obtained in figure 6.5 (input L-H-L) due to drain body and gate-body coupling. The body is also charged by on-state impact ionization current. The body voltage decreases as the drain voltage collapses. To minimize the hysteretic threshold voltage variation, it is desirable to balance the gain and loss of body charges in different logic states. This can be achieved through process/device design with decreased gate-body capacitive coupling and leaky body source/drain diodes [6.29].

6.17 Non Ideal diode characterstics

Non-ideal diode characteristics result in higher diode currents. This reduces the significance of impact ionization during switching. It also increases the symmetry between forward and reverse biased diode currents, providing higher reverse diode current to replenish the loss of body charges due to the forward biased diode. Diode non-ideality also reduces the loss of charge through capacitive coupling by decreasing the change in forward current per change in voltage. Operating in the regime where the leaky body source/drain diodes dominate device operation to minimize hysteretic V_T variation implies a higher body voltage (when the gate is LOW), which increases the off-state current and degrades the transient sub-threshold leakage.

Scaled device design for low supply voltage tends to alleviate the hysteretic V_T variation since higher peak doping in devices with highly non uniform doping profiles result in more non ideal diode characteristics and less gate body capacitive coupling due to increased body source/drain capacitance. Experimentally, the variation of hysteretic delay in simple logic circuits like NAND/NOR, inverters etc, has been determined to causes approximately 5-6% variation in frequency [6.5, 6.13, 6.26, 6.28].

References

[6.1] J.P.Colinge, IEDM Technical Digest, 1989, pp. 817.

[6.2] H. Mogul, et. al., "Designing with PD-SOI", IEEE Dallas CAS Workshop on Low Power / Low Voltage Mixed-Signal Circuits and Systems (DCAS-01), March 2001, pp. p31-p35

[6.3] A. Marshall & S. Natarajan, "SOI Ring Oscillators" VLSI Design and Test Conference 2001, India, pp. 217 – 224.

[6.4] K. Bernstein & N. J. Rohrer, "SOI Circuit Design Concepts", Kluwer Academic Publishers, January 2000, ISBN 0-7923-7762-1

[6.5] J. P. Colinge, "SOI Technology, Materials to VLSI", 2nd edition, Kluwer Academic Publishers.

[6.6] G. Shahidi et. al., "Partially-Depleted SOI Technology for Digital Logic", ISSCC 1999, pp. 426 - 7.

[6.7] T. W. Houston & S. Unnikrishnan, "A guide to simulation of hysteretic gate delays based on physical understanding [SOI logic]", 1998 IEEE International SOI Conference, pp. 121 –122.

[6.8] W. Wolf, "Modern VLSI design - Systems on Silicon", Prentice Hall

[6.9] N. Weste & K. Eshragian, "Principles of CMOS VLSI Design – A system perspective", Addison-Wesley, Reading, MA 1988.

[6.10] J. Uyemura, "CMOS Logic Circuit Design", Kluwer Academic Publishers, 1999.

[6.11] P. F. Lu, et.al., "Floating body effects in partially depleted SOI CMOS circuits", Proceedings of the 1996 International Symposium on Low Power Electronics and Design, Monterey, CA, Aug 12-14, 1996, pp. 139-144.

[6.12] A. Wei & D. A. Antoniadia, "Measurement of transient effects in SOI DRAM/SRAM access transistors", IEEE Electron Device letters, Vol 17, May 1996, pp. 193-195.

[6.13] C. T. Chuang, et. al., "SOI for digital CMOS VLSI: Design considerations and Advances", IEEE Proceedings, Vol 86., No. 4, Apr 1998. pp. 689 – 720.

[6.14] M. M. Pellela, et. al., "Low voltage transient bipolar effect induced by dynamic body charging in PD-SOI MOSFETs", Proc. IEEE Int. SOI Conf., Oct 1995, pp. 8 – 9.

[6.15] F. Morishita et al., "Leakage mechanism due to floating body and countermeasure on dynamic retention mode of SOI DRAM", Symp. VLSI Technology, 1995, pp. 141-142.

[6.16] M. M. Pellela, et. al., "Low voltage transient bipolar effect induced by dynamic floating body charging in scaled PD-SOI MOSFETs", IEEE Electron Device letters, Vol 17, May 1996, pp. 196 – 198.

[6.17] P. F. Lu, et.al., "Floating body effects in PD-SOI CMOS circuits", IEEE JSSC, Vol 32, Aug 1997, pp. 1241-1253.

[6.18] "CMOS scaling in 0.1um, 1.x volt regime for high performance applications," IBM Research development, Vol 39, No. 1/2, Jan/Mar 1995, pp. 229 – 244.

[6.19] G. Shahidi, et. al., "Device and Circuit Design Issues in SOI technology", IEEE Custom Integrated Circuits Conference, 1998, pp. 339-42.

[6.20] D. Allen, et. al., "A 0.2um 1.8V 64b powerPC microprocessor", ISSCC 1999

[6.21] F. Assaderaghi et. al., "Transient pass transistor leakage current in SOI MOSFETs", IEEE Electron Device Letters, 1997, pp. 241.

[6.22] J. Gautier & J. Y. C. Sun, "On the transient operation of partially depleted SOI NMOSFETs", IEEE Electron Device Letters, Vol. 16, Nov 1995, pp. 497 - 499.

[6.23] G. Shahidi, et.al. "A room temperature 0.1um CMOS on SOI" IEEE Trans. Electron Devices, Dec 1994, pp. 2405.

[6.24] G. G. Shahidi, et.al., "Device and Circuit Design Issues in SOI Technology", ISSCC 2000.

[6.25] F. Assaderaghi, et. al., "History dependance of Non Fully Depleted (NFD) digital SOI circuits", VLSI Technical Digest, 1996, pp. 122.

[6.26] F. Assaderaghi, et. al., "A 7.9/5.5 pSec room/low temperature SOI CMOS", IEDM technical digest, 1997, pp. 415 – 418.

[6.27] A.Wei, et. al., "Minimizing floating body induced threshold voltage variation in partially depleted SOI CMOS", IEEE Electron Devices letters, Vol 17, Aug 1996, pp. 391 - 394.

[6.28] D. J. Schepis, et. al., "A 0.25µm CMOS SOI technology and its application to 4Mb SRAM", IEDM Technical Digest, 1997, pp. 587 - 590.

[6.29] A. Wei & D. Antoniadis, "Design methodology for minimizing hysteristic variation V_T variation in PD-SOI CMOS", IEDM Digest, 1997, pp.411 - 414.

Chapter 7: Dynamic SOI Digital Design

7.1 Introduction

This chapter focuses on dynamic circuits in partially depleted SOI technology. Charge sharing, Miller capacitance and self-heating are critical design aspects of a dynamic circuit in PD-SOI technology. Parasitic bipolar leakages can upset the functionality of dynamic circuits and are critical design parameters.

Design of dynamic logic circuit is more complex than the design considerations required for static logic circuits. However, the superior performance of dynamic logic over static is the motivation for identifying failure mechanisms. A common problem of dynamic circuits is loss of pre-charge, noise immunity, and variability in delay [7.1].

7.2 Dynamic Circuit Response

Dynamic logic is normally faster than static logic and is used frequently in digital signal processors, microprocessors and other high performance products. If charge loss occurs from a dynamic circuit node, its current logic state can be corrupted. Parasitic bipolar currents, which are of only moderate concern in static logic, can be very detrimental to the charge storage of dynamic logic circuits.

7.2.1 Dynamic History Effect

Figure 7.1 shows a dynamic CMOS circuit [7.1] susceptible to parasitic bipolar current. Consider the case when MN0 is high and MN1, MN2, and MN3 are all low. Node N1 is high and the output is low during precharge.

Since there is no path to ground, intermediate node N2 reaches V_{DD}-V_T, and over a period of time, the body voltages of MN0, MN1 and MN2 rise. Once input MN3 begins switching high, node N2 discharges to ground. The body voltage of MN3 will rise initially due to body to gate coupling. Once the transistor is completely ON, the body voltage drops.

After node N2 discharges to ground, the bodies of transistors MN0, MN1 and MN2 also discharge through their forward biased body-source diodes. The body to source current in each device causes bipolar drain-to-source current. The bipolar body to source current continues until the body is completely discharged. This current can affect the charge at intermediate node N2. Figure 7.1(b) shows the body voltage of nMOS MN3 falling as the input signal IN4 rises.

7.2.2 Dynamic Charge Sharing

Charge sharing is the capacitive voltage division between two nodes. Assuming node N2 is discharged from a previous precharge cycle and N1 is high. If three inputs MN0, MN1 and MN2 are high and MN3 is LOW, voltage is shared between the nodes N1 and N2, since there is no direct path to GND. Node N1 can fall to a voltage that is low enough to corrupt the output. The voltage on node N1 when MN0, MN1, MN2 are ‘high’, and MN3 ‘low’ can be given by:

$$V_{N1} = V_{DD}*C_{N1}/(C_{N1}+C_{N2})$$

The severity of charge sharing is lower in SOI than bulk. Source/drain capacitance is approximately a factor of four lower than bulk CMOS. Reduced capacitance in SOI helps reduce charge sharing. When the gate voltage rises, the source of the nMOS transistor rises due to capacitive coupling between the source and gate. A reduction in the voltage difference between N1 and N2 causes charge sharing.

7.2.3 Capacitive Coupling Effects

The gate overlap capacitance caused by source/drain region is called Miller capacitance. Assume node N2 (figure 7.1) precharged high from a previous cycle, when the input to MN0 rises, node N2 couples to the gate and increasing its potential, which in turn couples onto output node N1 and increases its charge.

If transistor MN0 is completely ON, Miller effect will be reduced due to the reduced voltage difference between N1 and N2 since there exists a current path to ground.

7.2.4 Keeper Devices or Bleeders

Subthreshold leakage current and other process related leakage currents are a major cause of circuit failures in SOI. To maintain functionality from leakage currents, keeper devices are used, which are normally weak pMOS devices with extended gate lengths. Keepers replace lost charge due to parasitic bipolar leakage currents. They require careful sizing to maintain proper functionality. High current keeper devices impact hysteresis and degrade performance. In SOI technology, parasitic bipolar response is controlled better with more effective design of the nMOS evaluate stack rather than with a strong pMOS keeper.

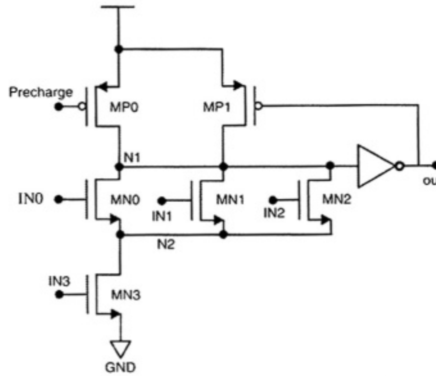


Figure 7.1(a): Dynamic Circuit vulnerability to parasitic bipolar current.

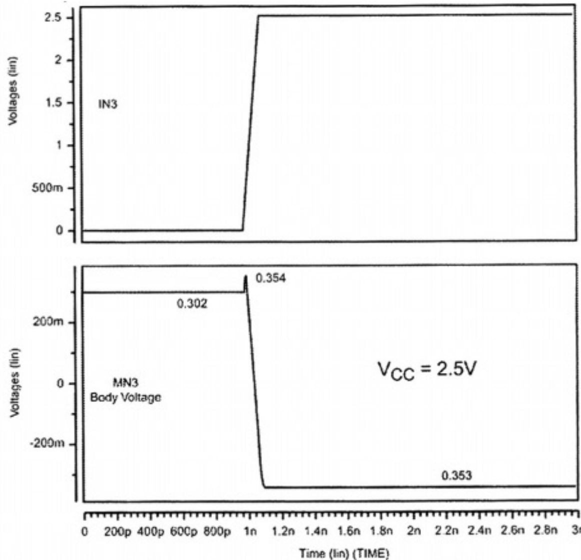


Figure 7.1(b) Input and body voltages of transistor MN3 falling from 0.354V(@V_{DD}=2.5V) to ground after the gate input raises. The body voltage dip for V_{DD}=1.2V is 0.244V.

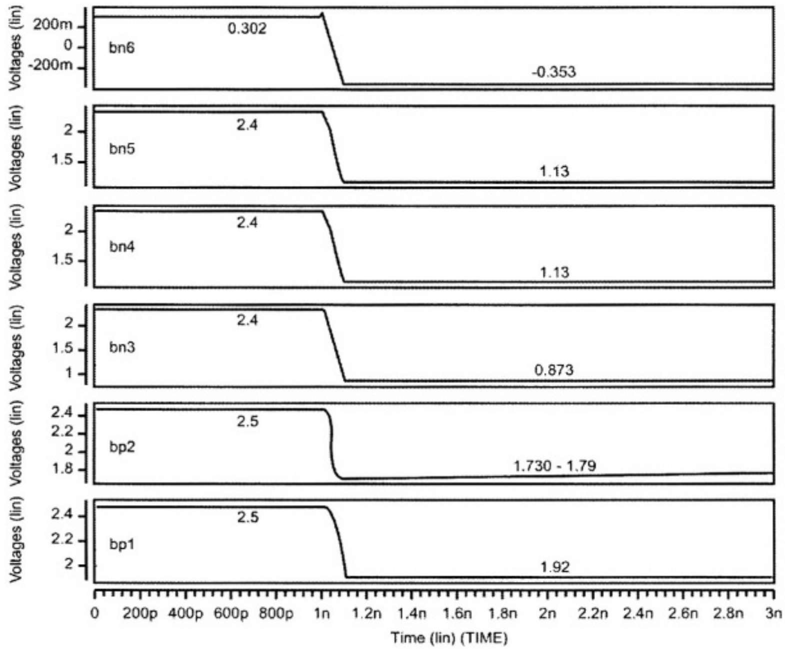


Figure 7.1(c): Node voltages associated with schematic of figure 7.1(a)

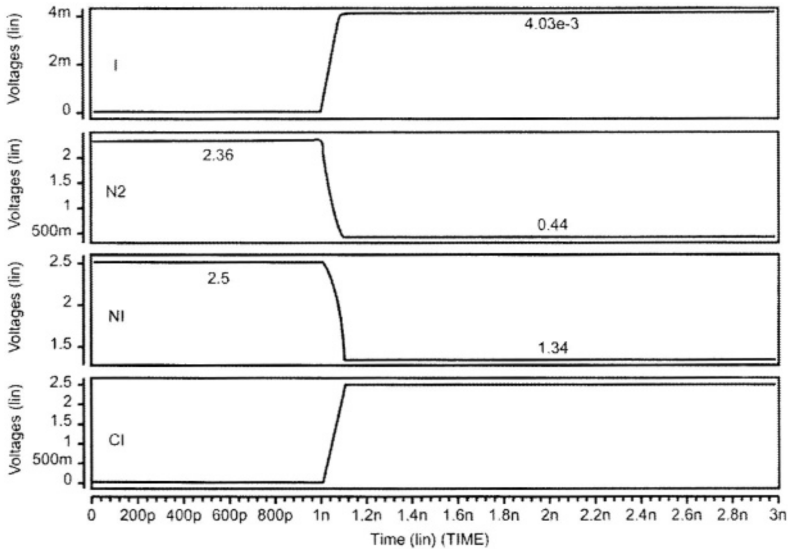


Figure 7.1(d): Body voltages associated with schematic of figure 7.1(a)

7.3 Dynamic Circuit Design Considerations

Pre-charging intermediate nodes in a domino stack reduces the loss caused by charge sharing and charge redistribution. There are three different precharge mechanisms to an intermediate node in a dynamic circuit. These are: (1) Precharge high, (2) Precharge Low (3) Precharge between supply rails. Precharging has different trade-offs in bulk and SOI. Refering to figure 7.1 we analyse some of the trade-offs [7.1, 7.2].

(1) Precharging N2 high results in worst case noise immunity as it charges the bodies of devices MN0-2 high. This causes bipolar devices to turn 'on' and causes increased device leakage. If intermediate node N2 is precharged high, with the source and drain of MN0-2 at V_{DD} and V_{DD}-V_T, the bodies charge to V_{DD} causing source-drain leakage. Body effect is minimal when the output is high, but results in higher drain junction capacitance. The drain body space charge region is small when the body and drain potentials are high helping noise immunity but hurting performance. Discharge performance is very slow, since the high induced charge must be discharged to get the evaluate stack to discharge.

(2) Precharging to ground results in discharging transistors bodies MN0-2 to GND, eliminating bipolar leakage, and improving noise immunity. Discharge is rapid. The diffusion capacitance is relatively low, easing charge-sharing concerns. The maximum body effect occurs when the bodies of transistors MN0-2 are low. Low drain junction capacitance occurs when the bodies of devices MN0-2 are low, since the drain body space charge region increases as the body and drain (pre-charge node) potential difference increases. This helps performance but degrades noise immunity.

(3) Intermediate charging achieves low parasitic bipolar and MOSFET leakage. Charge exists on the storage nodes, diffusion capacitance is small and charge sharing is minimal. Node N2 is not fully charged, hence the presence of body effect. Junction capacitance is not at a minimum, causing a trade off between noise immunity and performance. Discharge performance is balanced [7.3].

The sequence in which the signals arrive at the gates of the transistors in the evaluate mode is very important and may need be different case from the bulk devices. One method to reduce the losses due to bipolar current and precharge, is to cross couple inputs in the nMOS stack. Figure 7.2(a) shows the conventional dynamic logic circuit, and, figure 7.2(b) shows the same circuit after re-ordering/cross coupling the inputs. This approach also achieves favorable performance while limiting charge loss [7.1].

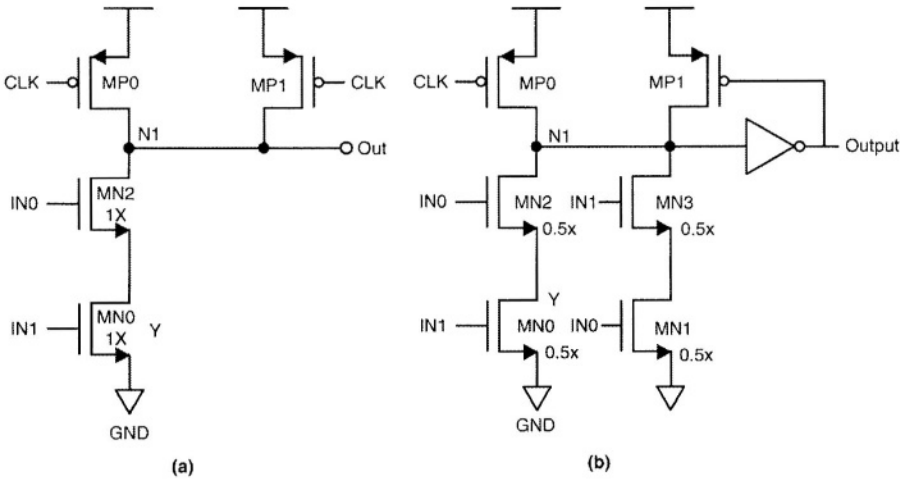


Figure 7.2: Cross-coupling for balancing performance and noise. (a) Original signal coupling as practiced in bulk and (b) Design practice in SOI. [7.1]

7.4 Re-ordering and Remapping

Latest arriving signals should be connected closer to the output in dynamic bulk circuits to prevent charge sharing, and charge redistribution. This technique proves useful for bulk CMOS circuits, however, in SOI technology, due to the parasitic bipolar leakage, it is of reduced benefit. As the number of devices in the nMOS stack reduces, the amount of charge lost to leakage increases.

Domino stages use a single path to GND for multiple inputs also reducing charge loss. Furthermore, it is useful to put a gating switch closer to the output to minimize charge loss. Figure 7.3(a) shows a dynamic CMOS circuit and figure 7.3(b) shows the circuit after re-ordering. The total number of devices developing high body potential and charge loss is reduced.

7.5 Logical Remapping

A frequently used technique is that of logical remapping. This involves, re-positioning of inputs from the dynamic portion of the circuit into the static circuits. This technique is been used in bulk dynamic circuit design frequently.

Logical re-mapping of inputs reduces the stack height, history magnitude and bipolar current loss in SOI technology. Figure 7.4 shows the

conventional dynamic logic circuit and the remapped circuit. The stack is reduced by one transistor.

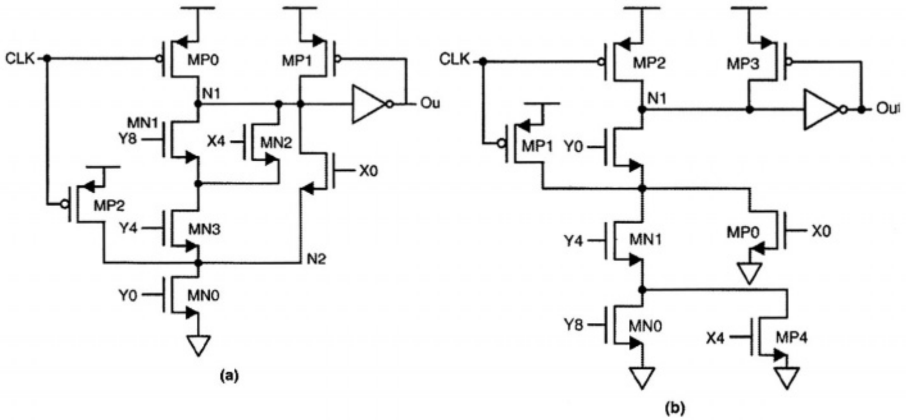


Figure 7.3: Reordered input sequence to mitigate bipolar and MOSFET leakage: (a) Original signal coupling as practiced in bulk and (b) Recommended design practice in SOI.

If the pre-charge cycle extends into the evaluate cycle, charge lost due to bipolar parasitic current is replaced by the pre-charge pMOS device. This method may be used on a selective basis, but since overlap allows DC currents to pass through the chip, care must be exercised. The technique usually requires an adjustable duty cycle clock generator.

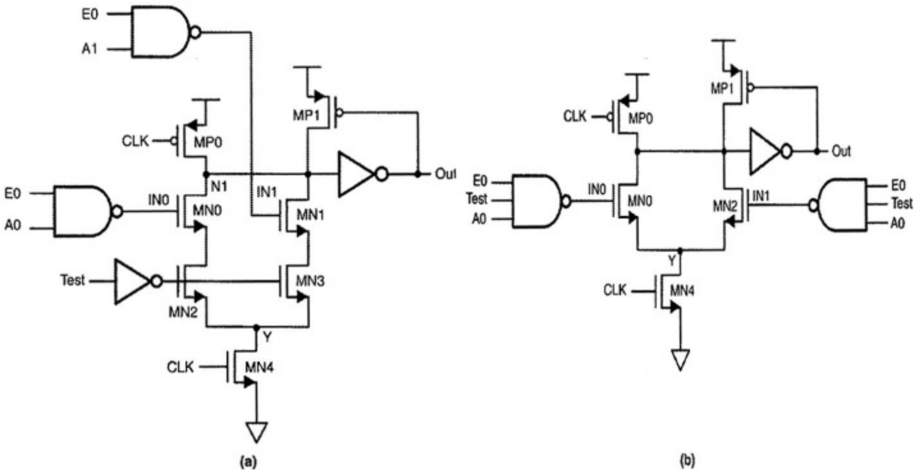


Figure 7.4: Logic Remapping in: (a) Bulk material and (b) partially depleted silicon on insulator

7.6 Complex Domino

Domino circuits are used to reduce leakage. Logically wide functions are designed by moving half the circuits on the precharge node to a secondary node. Both pre-charged node circuits are evaluated in parallel. The evaluated nodes are combined in the final static buffer producing a single bit output.

Such a combination avoids excessive pre-charge or evaluate delays in bulk, and reduces the total parasitic bipolar current on a single node in SOI. Limiting the bipolar current spike magnitude helps sizing keeper devices (figure 7.5).

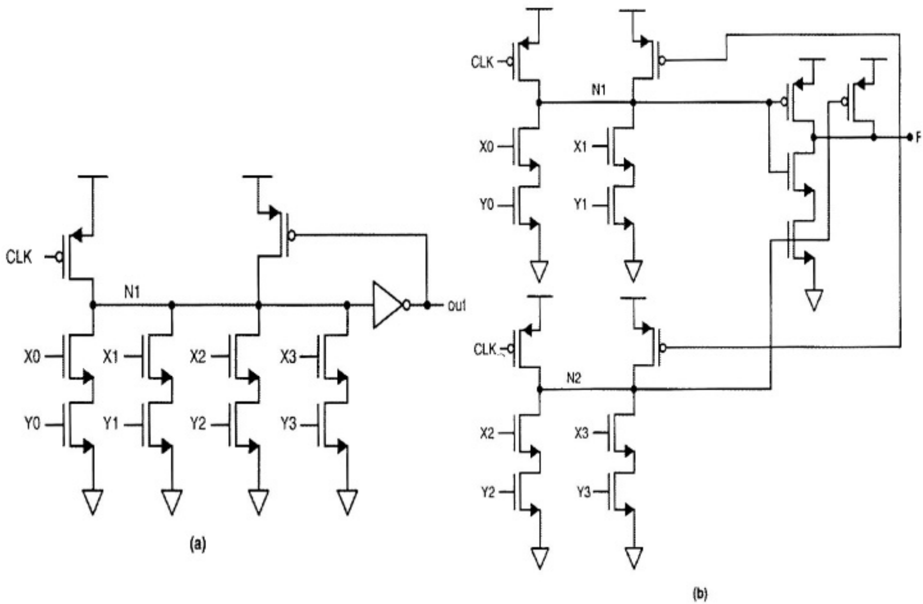


Figure 7.5: Complex domino logic induces substantial parallel leakage, and logical width reduction in SOI. This achieves a given function while cutting node leakage in half.

7.6.1 3-input Domino OR Gate

Domino is a system design style that eliminates nMOS-pMOS glitch problems without introducing pMOS type logic gates. Figure 7.6 shows a 3-input OR domino logic gate [7.4]. Precharge occurs when $\Phi=0$ with C_X charging to a voltage $V_X=V_{DD}$. If one or more of the IN0, IN1 or IN2 inputs are "1", when $\Phi=1$, C_X discharges, the output voltage then changes from a precharge value of 0V to a value of $V_{OUT}=V_{DD}$. When all nMOS transistors are "ON", parasitic bipolar current flows, since both source and drain were at V_{DD}. The output node discharges once Φ goes high. The intermediate node

will rise initially before falling low when Φ goes high due to gate-drain coupling. Device MN3's body voltage will fall once its turned 'ON'.

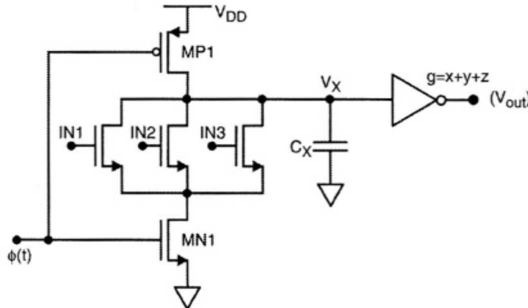


Figure 7.6: Schematic of a 3-input Domino OR Logic Gate.

7.6.2 Dynamic AND-OR Domino Gate

Figure 7.7(a) shows the AND-OR domino gate [7.4] and its simulation waveforms. Input X is maintained high initially and input Y is toggled low to high. With Φ=0, the output node N1 (V_x) is maintained high since MN1 is 'off'. The body voltage of MN1 will be high since the source and drain are high. Once input B is switched high, the body voltage of MN2 tends to go higher due to capacitive coupling.

The intermediate node voltage will be high and at threshold voltage below V_{DD} since there is no rapid discharge path to ground. Once Φ=1, intermediate node N2 discharges to ground through MN3. The body voltage initially couples high before finally settling low. The body voltages of MN1 settles lower once Φ=1. Figure 7.7(b) and (c) show the body and node voltages when input Z is low. Figure 7.7(d) and (e) shows the body and node voltages when Z is high [7.4].

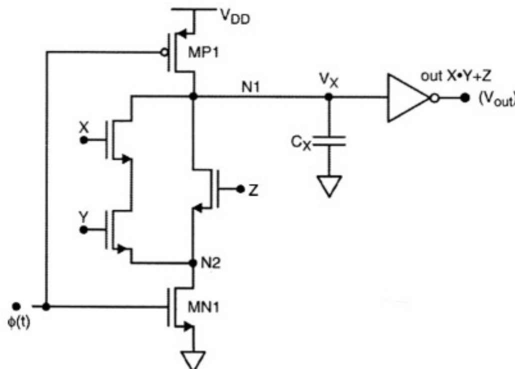


Figure 7.7(a): Dynamic AND-OR schematic.

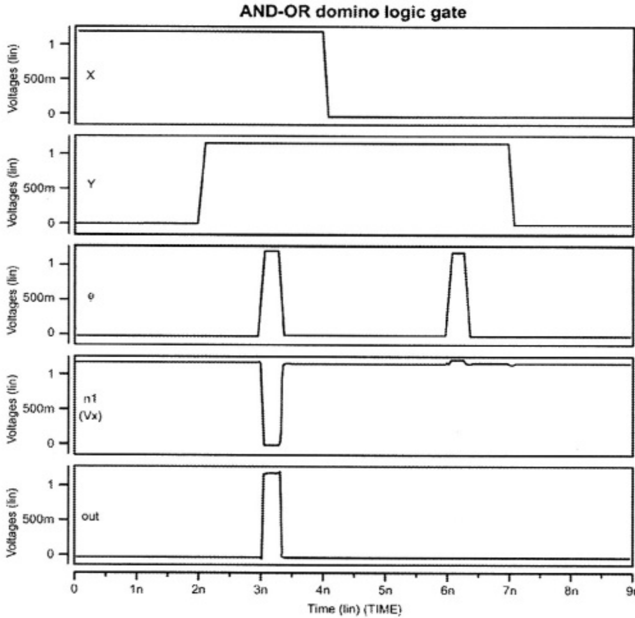


Figure 7.7(b): Waveforms when input A is high and input B is switching low to high. Body voltages with Z low.

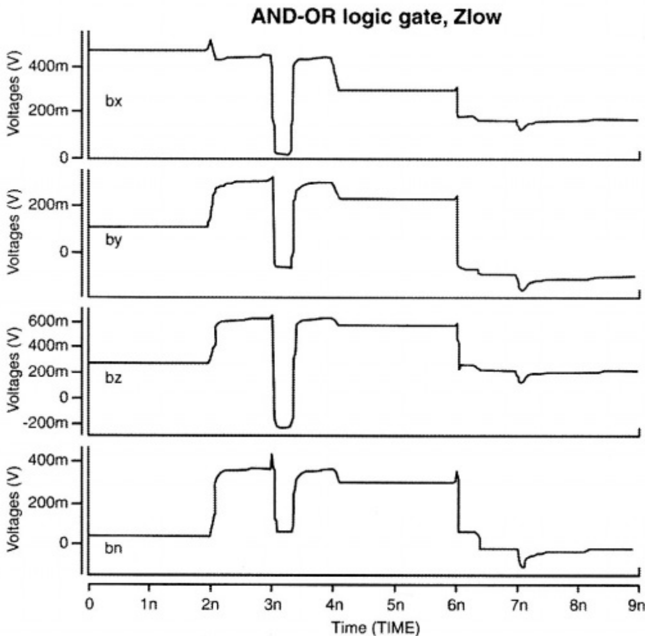


Figure 7.7(c): Waveforms when input A is high and input B is switching low to high node voltages. (Body voltages of devices with inputs X, Y, Z).

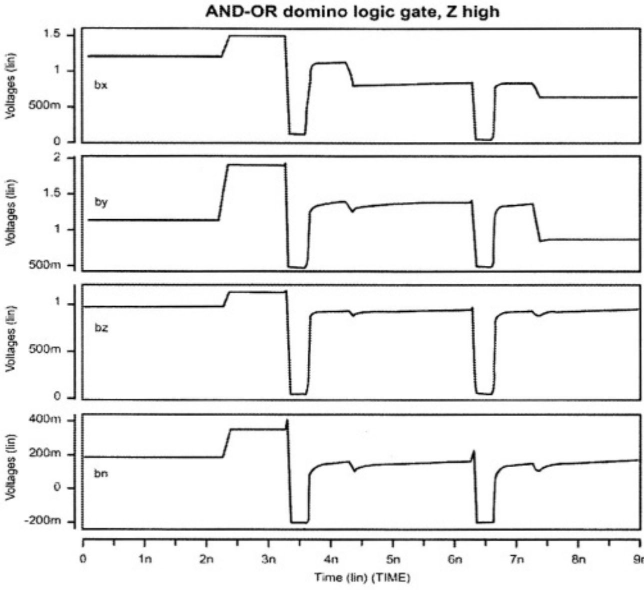


Figure 7.7(d): Node voltages for Dynamic AND-OR when input A is high and input B is switching low to high with Z high.

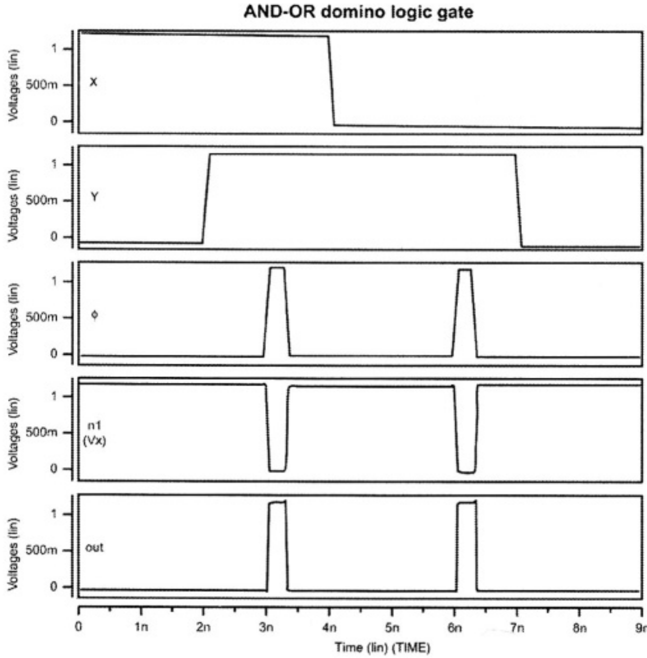


Figure 7.7(e): Node voltages for Dynamic AND-OR when input A is high and input B is switching low to high with Z high.

7.7 No-Race Logic (NORA)

No-Race logic was introduced to overcome signal race conditions associated with clocked MOSFETs and transmission gates to control data flow. A pass gate exists on input and output of a logic block to control the data flow. If the input to the pass gate is initially high, this signal passes through the logic and to the output. Once the input changes at the same time as the pass gate control changes, wrong data will be sent to the logic reflecting incorrect signal at the output.

Wrong data gets transmitted through the circuit reflecting a race condition between the data intended to be passed and the data passed through the logic circuitry. The signal race problem can be aggravated by the presence of clock skew in which the clock signals are displaced from each other. A large value of clock skew widens the race window, and enhances the possibility of race in the system. Skew is particularly troublesome in high-speed circuits as it limits the operating frequency.

The basic building block of NORA logic is the Φ-section logic network (Figure 7.8) [7.4]. This consists of a dynamic nMOS logic state that is cascaded into a dynamic pMOS logic gate: a C^2MOS latch is used as an output latch. Optional inverters are provided at the outputs of both logic gates for a glitch free domino nMOS-nMOS cascade. Moreover, the ordering of the logic gates may be reversed (pMOS to nMOS) without loss of generality [7.5].

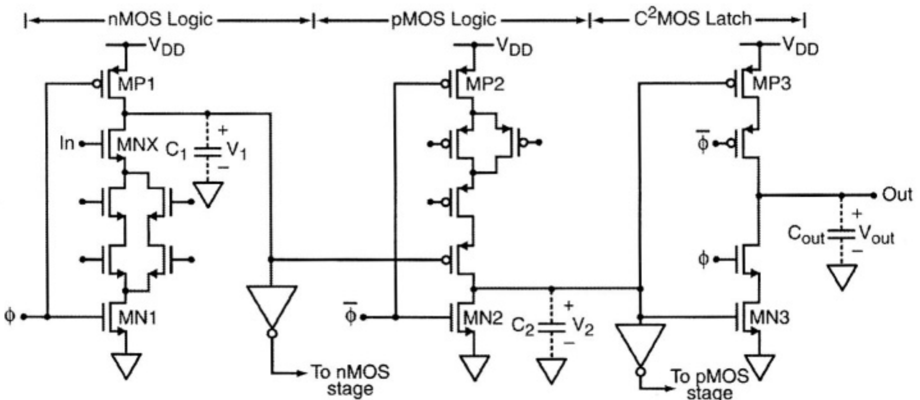


Figure 7.8(a): Schematic of a SOI NORA Φ-section logic network.

The main feature of NORA logic arises in the way the clocks are applied to the logic gates and the latch circuits. Φ=0 defines the precharge interval. The

output node V1 is precharged high and V2 is low. V_{OUT} is tristated during precharge. The output voltage depends on the amount of charge held in the capacitor (C_{LOAD}). Once the circuit switches to evaluation by Φ=1, the output node attains a voltage level based on the input stored (which is high in this case). Body voltage of the nMOS (MN3) also rises due to body to drain coupling since Φ switches high, as shown in figures 7.8 (b, c).

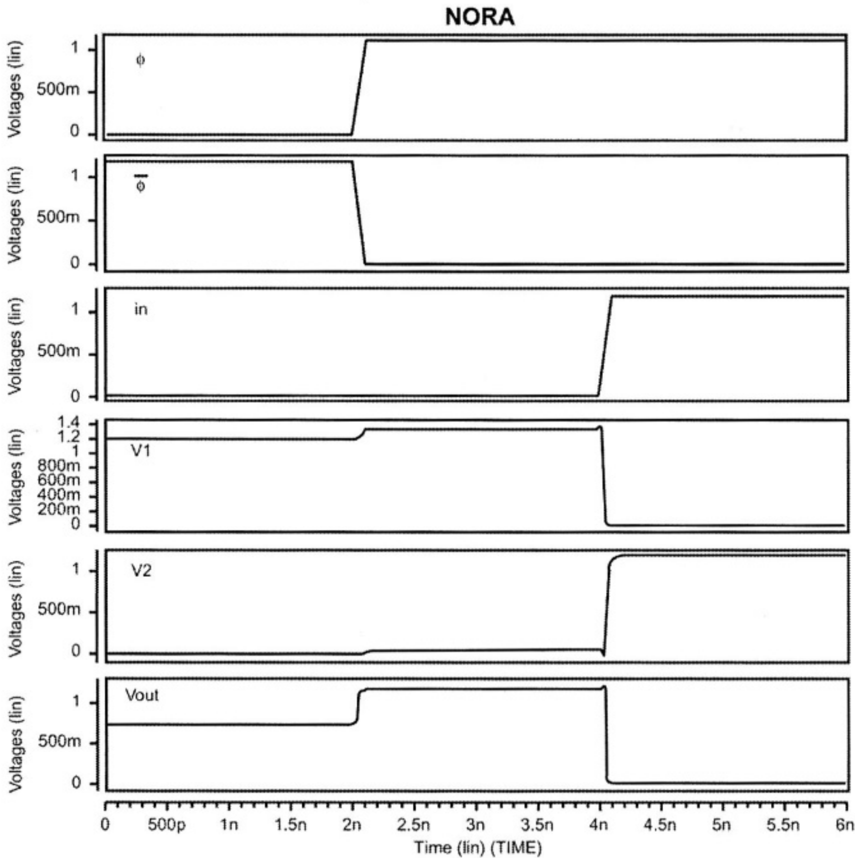


Figure 7.8(b): Additional waveforms of NORA Φ-section logic network

The NORA $\overline{\Phi}$ section may then be constructed (figure 7.9). Dynamic cascade logic gates are used with alternating polarities. This is the similar to the original Φ section except that Φ and $\overline{\Phi}$ are reversed throughout the circuit. The $\overline{\Phi}$ section precharges when Φ=1, and evaluates when Φ=0. The simulation waveforms are shown in figure 7.9(b). The body voltages of critical devices are shown in figure 7.9(c).

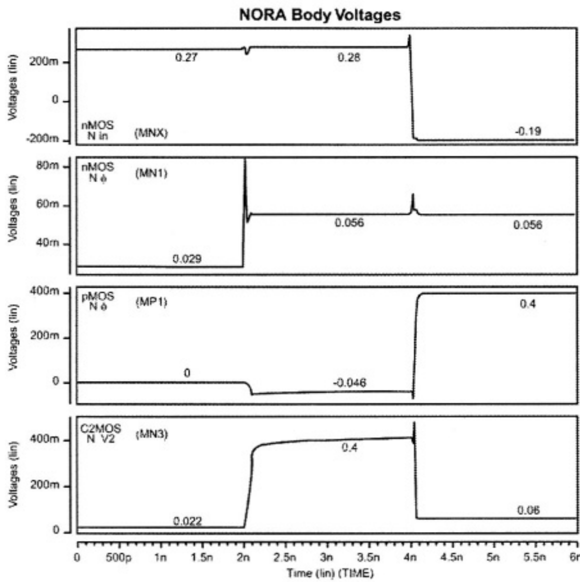


Figure 7.8(c): Body voltages of a NORA Φ-section logic network. Input MN switches low to high, and all the other nMOS are held HIGH.

When $\overline{\Phi}$ is low, the output voltage V1 will be high irrespective of the state of the input transistor MN1. The final chain output, V_{OUT}, will be tristated. Once input MN1 turns 'on', the body voltage of MN1 will rise due to body-gate coupling which in turn couples to the drain node (output of first stage V1). The output voltage V1 of the first stage droops down due to drain-gate capacitance.

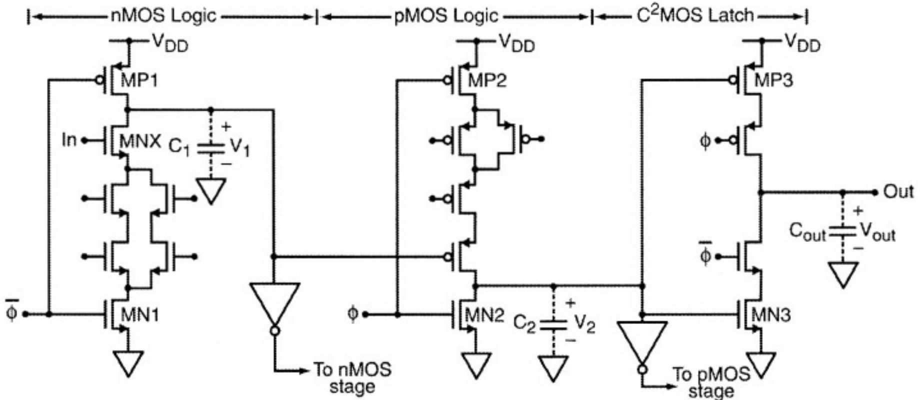


Figure 7.9 (a): NORA $\overline{\Phi}$ logic gates.

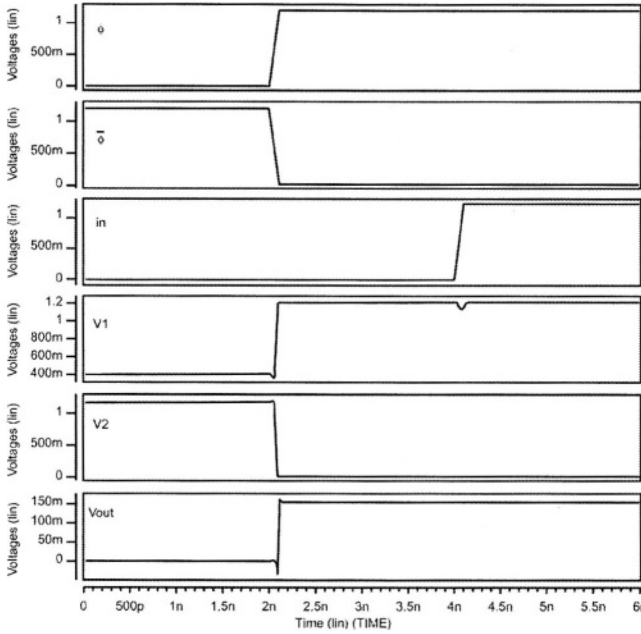


Figure 7.9 (b) NORA $\overline{\Phi}$ Section waveforms. Input N switching and all other nMOS maintained HIGH.

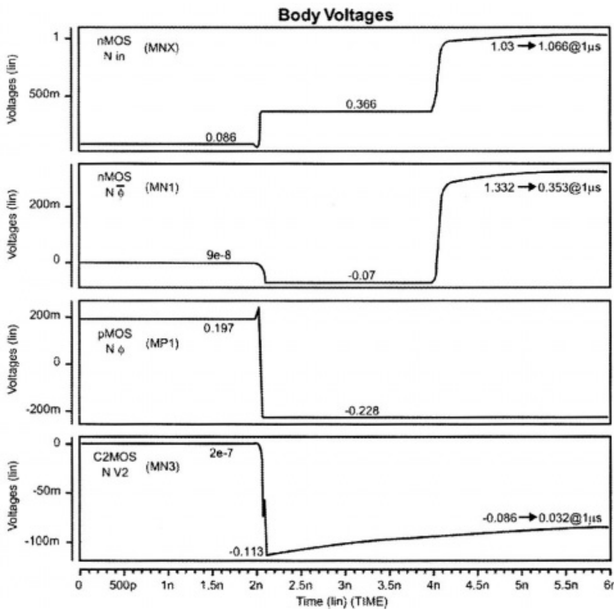


Figure 7.9(c): Body voltages of $\overline{\Phi}$ Section. Input IN switches low to high and other nMOS maintained HIGH.

The no-race characteristics are obtained by creating alternating cascade of Φ and $\overline{\Phi}$ sections as shown in figure 7.10. The timing of these two sections ensure that signal races do not occur through either section. During precharge Φ=1, the Φ-section are in evaluation and $\overline{\Phi}$ sections precharging. The inputs are valid during this period and produce results at the output. However, the next logic $\overline{\Phi}$ section is precharging since Φ=1, and does not accept input data values. This eliminates race conditions through the $\overline{\Phi}$ sections. Similarly, when Φ=0, the Φ-sections are in precharge and block data transmission when $\overline{\Phi}$ evaluates. The sections take turns evaluating the inputs and block data transmission as the clock frequency oscillates [7.5]. Figure 7.10(b) shows a pipelined NORA CMOS logic with C^2MOS logic.

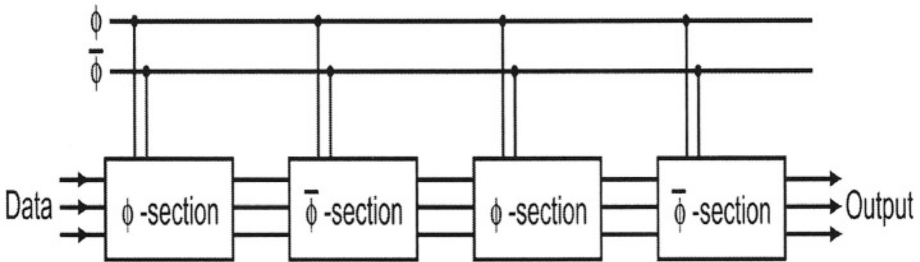


Figure 7.10(a): A NORA Pipeline cascade.

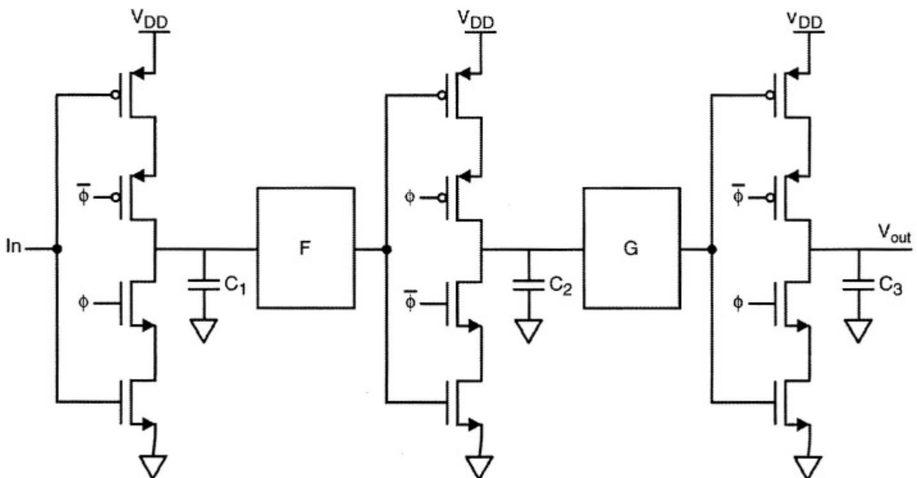


Figure 7.10 (b) Pipelined NORA CMOS logic with C^2MOS. Race free as long as all the logic functions F and G between latches are non-inverting.

7.8 Dynamic Noise Suppression

The parasitic bipolar response of a floating body in partially depleted SOI MOSFETs becomes more prominent at higher voltages. Most products undergo burn-in and/or voltage stress prior to shipment to determine reliability due to processing defects. Dynamic logic circuits maintain functionality but degrade in performance at elevated voltages.

Figure 7.11(a) shows a conventional domino circuit, and figure 7.11(b) shows the improved hysterisis free circuit. For robust functionality with speed trade-offs, the feedback transistor is turned ON to avoid hysteresis and is turned OFF when performance is critical. The feedback adds drain capacitance to the evaluate node slowing down the output. Figure 7.12 uses dynamic feedback to achieve the same functionality. Device MP1 should be large and keeper device MP5 should be small to minimize cycle time during functionality testing. During burn-in, MP1 must be small and MP5 large to prevent bipolar leakage.

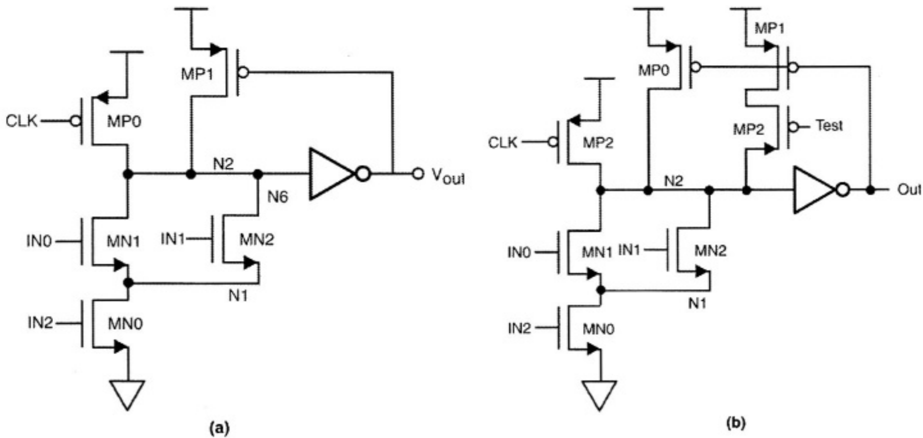


Figure 7.11: Dynamic noise suppression during high voltage stress (a) Bulk (b) additional feedback in SOI for supplement charge replacement.

During stress operation, the TEST signal dynamically increases pMOS drive strength of the parallel pMOS transistors by turning MN2 'OFF' and turning 'ON' MP4. This helps suppress noise at higher voltages.

During operational mode, MN2 is 'ON' and MP4 is 'OFF'. The sum of the device widths MP1, MP3, MP5 are the same as the precharge and the keeper devices in figure 7.5. The effective precharge device is maximized to reduce the delay for restore operation [7.1].

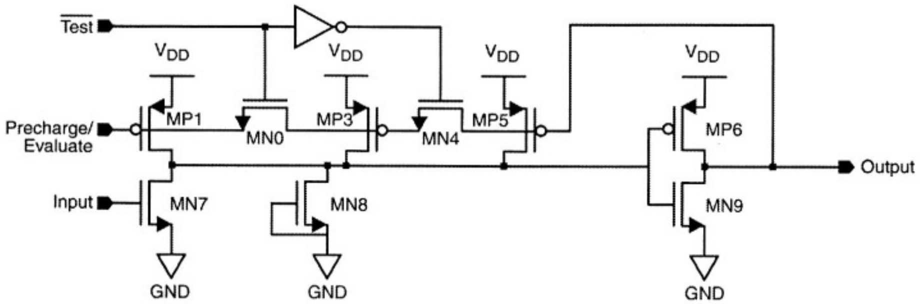


Figure 7.12: Dynamic feedback technique

7.9 Design Issues in Dynamic 2-way NAND Logic

Timing rules for modern VLSI designs are typically generated based on worst case process corners, supply, voltage, temperature, slew rate and switching patterns. For bulk CMOS, the pattern dependency of the stacked configuration degrades timing rules.

Figure 7.13(a) shows a simple dynamic 2-way NAND. Switching MN1 (when V_{IN1} switches from low to high and V_{IN2} stays High) to pull dynamic node N2 low is slower than switching MN2. This is because the voltage V1 at Node 1 is at V_{IN2}-V_T initially.

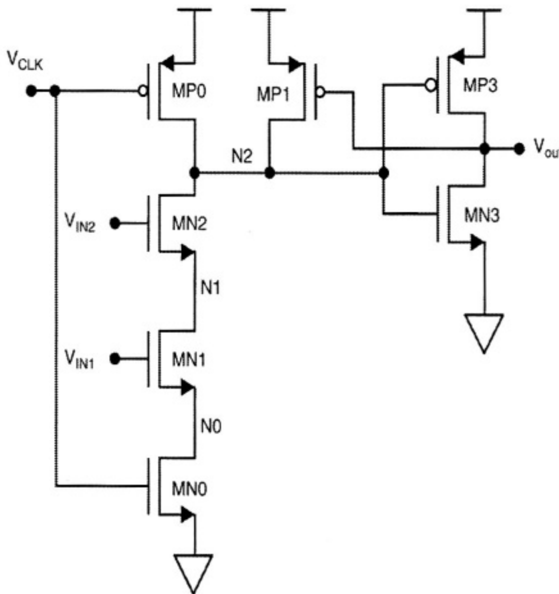


Figure 7.13(a): Schematic of a simple dynamic 2-way NAND Circuit [7.6].

MN2 suffers from the reverse body effect and thus has high threshold voltage initially. (With its source node N1 high and its body grounded, there is a large reverse bias between body and source of MN2, resulting in a high threshold voltage.)

Node N2 is at V_{DD} during clock low. Node N1 is high, but the body voltage of MN1 is low. Once the input to transistor MN1 turns 'ON', the body voltage of MN1 increases (0.87V) due to body to gate capacitance. The body voltage of MN0 is 0.389V when CLK is low and input high.

Waveforms are shown in figure 7.13(b)-(d). Once the clock transitions high, the nodes N0, N1 and N2 discharge to ground, and the respective nMOS body voltages also fall to ground [7.6].

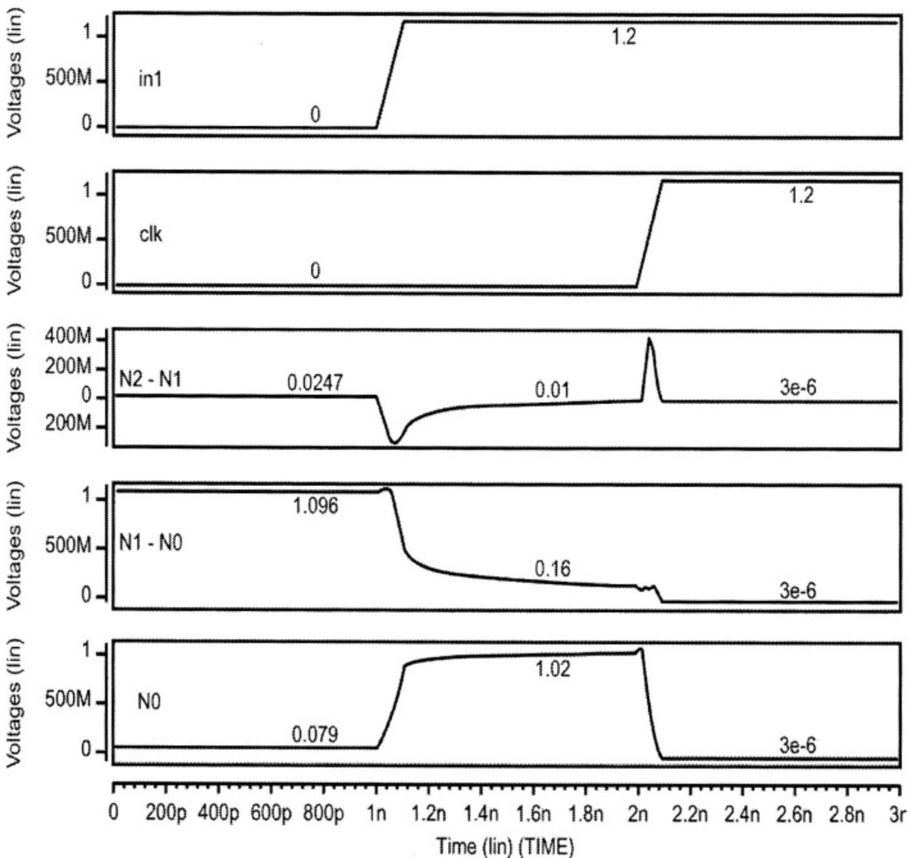


Figure 7.13(b): Dynamic 2-way Nand gate simulation waveforms. Vin2 = H.

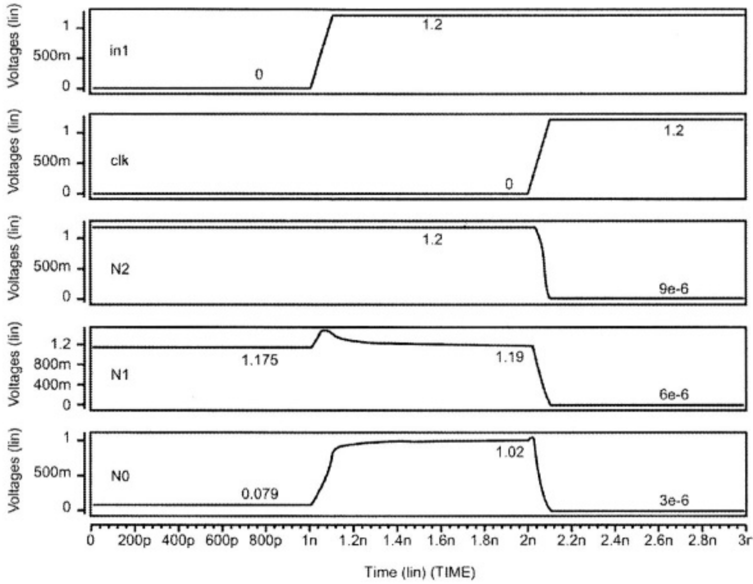


Figure 7.13(c): Intermediate node N0, N1, N2 voltages.

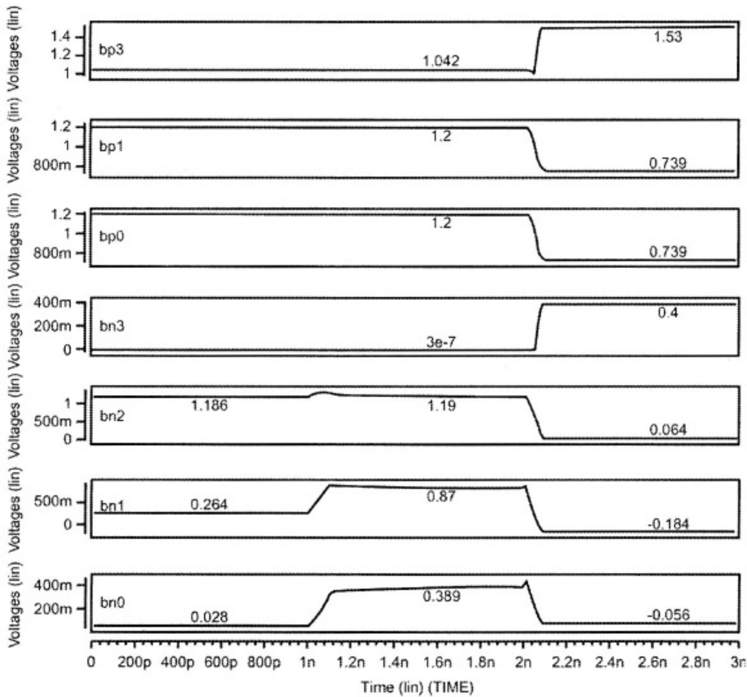


Figure 7.13(d): Body voltages of pull-down nMOS devices and pull–up pMOS devices

For the case of switching the top transistor MN2, full V_{DD} is present as V_{DS} across the transistor. Also, current to pull down node N2 is immediately available once MN2 switches 'ON'. While the reduced parasitic in SOI structure improves both switching cases, the improvement for the bottom transistor switching case is particularly significant.

The body voltage of MN2 lies between the voltages of nodes N2 and N1, hence, is very close to V_{DD} before N1 switches. Threshold voltage of MN2 is significantly reduced. Node N1 voltage is closer to V_{DD} than in bulk case, resulting in larger V_{DS} across MN1 to start with. The body voltage of MN1 is between N1 and ground, thus reducing V_T of MN1. The timing rules are therefore improved for the nMOS stack and pMOS stacked configurations in SOI.

Timing rules for pass transistor-based circuits, such as complementary pass gate logic devices, [7.7-7.9] are improved, primarily as a result of lack of reverse body effect in floating body SOI. The timing rules for some other circuit topologies, however, are degraded by the parasitic bipolar effect and transient threshold voltage variation. Sizing-up the holding devices, adding body contacts or widening design margin degrades circuit speed, area and timing rules.

7.10 Dynamic 2-Way OR Circuit

Parasitic bipolar leakage currents can be disastrous in dynamic logic circuits [7.10, 7.11]. Figure 7.14(a) shows a dynamic 2-way OR-AND circuit [7.12]. The dynamic circuit with nMOS stack and the evaluation is designed similar to the static OR-AND circuit discussed in chapter 6. Assume input to MN1 'high' and MN2 'low'. During precharge, when clock is low, node N2 is high, and V_{OUT} is low. Thus N1 is at V_{DD}-V_T. When MN1 switches low, due to gate source coupling node N1 couples down to a lower voltage. Once V_{CLK} goes high the circuit switches to the evaluate state.

The gate to source coupling at node N2 and parasitic bipolar current attempt to pull node N1 to ground. MP1 has to be sized very carefully, such that it is strong enough to maintain N1 high. Figure 7.14(b) shows the waveforms for the dynamic 2-way OR circuit.

When CLK is low and MN1 'high', node N1 voltage is at V_{DD}, once the input to MN1 switches low, the voltage at node N1 goes low due to the gate-body capacitance. The body voltage of node N1 is coupled down lower following switching of MN1.

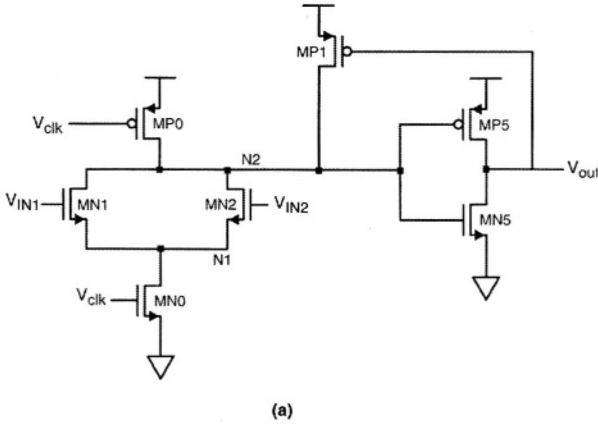


Figure 7.14(a): Schematic of 2-way dynamic-OR CMOS gate with floating bodies [7.12].

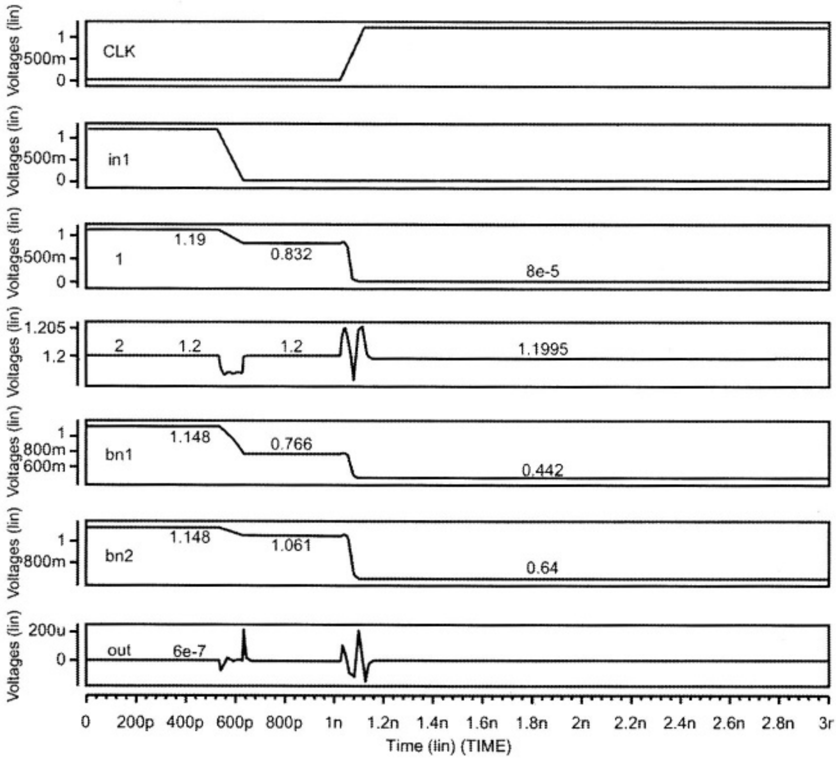


Figure 7.14(b): Dynamic OR gate switching wveforms. Node N1 voltage drops from 1.119V to 0.832V after input goes low and subsequently drops to ground after CLK high.

7.11 Dynamic Cascade Switch Logic

As with many other dynamic circuits, designing with SOI for cascade voltage switch logic (CVSL) circuits requires consideration of body effects. Multilevel voltage switch current steering type circuits may encounter much more complicated situations resulting in logic state errors [7.13]. Figure shows the schematic of a 3-input dynamic CVSL XOR gate [7.6]. In the pre-charge phase, all branches are inactive because the clocked evaluation transistor N0 is 'off ' (Figure 7.15). Node N1 and N2 are pre-charged to V_{DD}. Due to the differential input, all common source nodes in all cascade levels are 'HIGH'. Consequently, all logic transistors with LOW inputs are set up in a condition with both their drains and sources HIGH (thus, bodies are 'high' also). When the circuit evaluates, the common source nodes in the active branches are pulled 'low' and parasitic bipolar currents flow through these 'OFF' transistors. Because of the cross-coupled drain connection, each common source node is also a common drain node of the lower cascade level. Pulling down these common drain nodes causes 'inverse mode' parasitic bipolar current to flow from source to drain in the inactive branch 'off' transistors (figure 7.16) [7.6, 7.13].

Figure 7.15: Schematics of a 3-input Dynamic Cascade voltage switch logic (CVSL) XOR Circuit [7.13].

Evaluating figure 7.16, if the inputs to the XOR gate A, B and C are (0,1,1) respectively, during clock low, there is no path to ground with MN0 'OFF'. When clock switches high and since MN1, MN3 and MN7 are also "high", node N2 is maintained high. Since transistor MN1 is 'ON' and MN2 is

'OFF', the body voltage of MN1 falls to ground, while the body voltage of MN2 remains 'high'. The N1 and N2 node voltages are low and high respectively and, alternatively, nodes N6 and N7 are high and low. Transistors with inputs connected to $\{A$ (N1), $\overline{B}$(N5/N6), and C (N9/N10)$\}$ having zero gate bias are susceptible to parasitic bipolar effect during evaluation. The inverse mode parasitic bipolar currents flow through the 'OFF' transistors N5 and N9. Their voltages therefore droop much faster and earlier than the source nodes. The normal mode parasitic bipolar current pulls down high node N2, resulting in an erroneous state where both node N1 and node N2 are LOW.

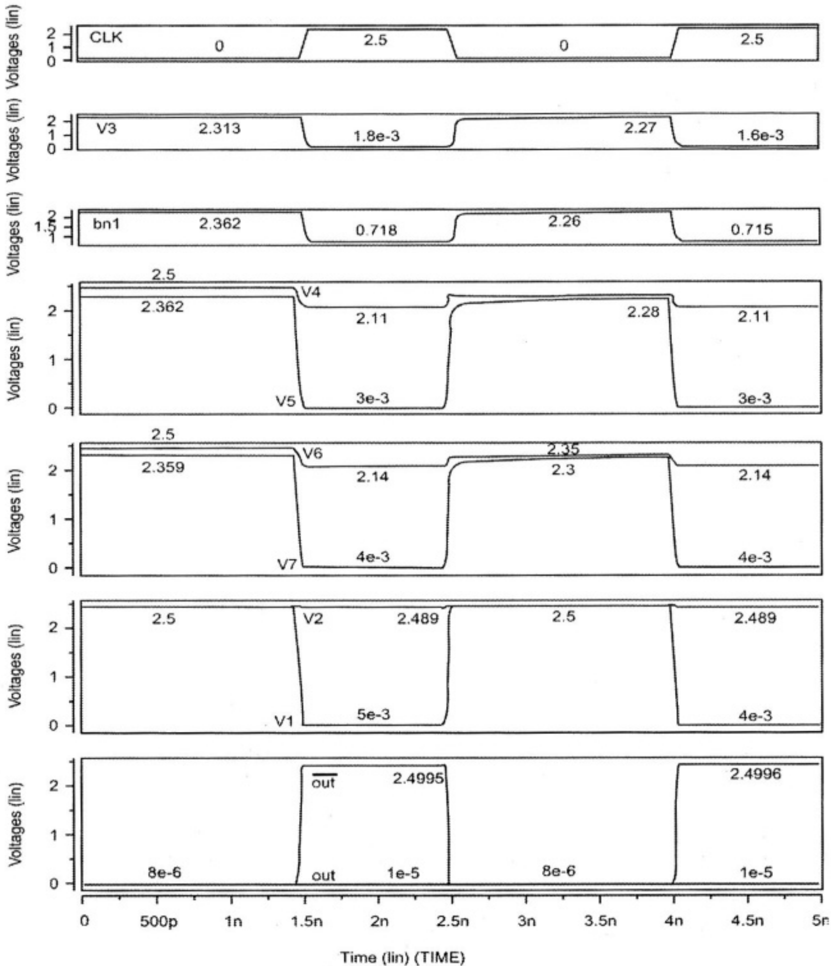


Figure 7.16: Switching Waveforms of dynamic CVSL XOR. Here, Vdd=2.5V. Input (A,B,C) = (0,1,1). The internal node voltages and the output are shown with there corresponding pairs.

Parasitic bipolar currents in the second evaluation cycle are substantially lower than those in the first evaluation cycle. This hysteretic behaviour stems from the time interval for the second pre-charge cycle being shorter than the body voltage charge up time. Cascade current switch logic are also normally used for faster logic performance trading off current consumption.

7.12 Clocked CMOS

Clocked CMOS (C^2MOS) is a logic family that combines static logic design with clock synchronization. Many small and medium scale integrated circuits have used C^2MOS logic. This technique is still implemented in modern VLSI designs as dynamic no-race (NORA) circuits explained earlier.

Figure 7.17 shows a C^2MOS gate. The inputs are connected as a static circuit, two pass gates, nMOS and pMOS are inserted between the output nodes with Φ and $\overline{\Phi}$ as the select signal. When Φ is high, both control pass-gates are 'ON', connecting the logic to the output node, and the gate behaves as its static equivalent circuit. Switching times are longer than static designs due to additional parasitics. When Φ=0, the output is isolated from the rest of the circuit [7.4].

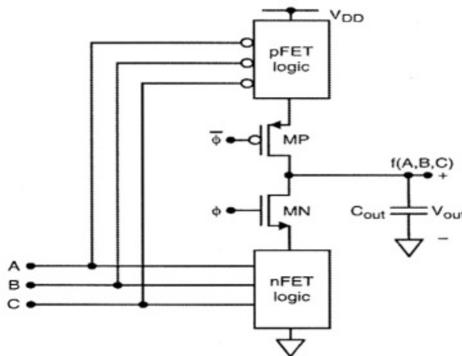


Figure 7.17: Clocked CMOS Gate Circuit [7.4].

Figure 7.18 shows a C^2MOS inverter and associated waveforms. When Φ=1, both clocked MOSFETs are active and the input controls the logic MOSFETs MP1 and MN1. If V_{IN}=0V, then $V_{OUT}=V_{DD}$, an input $V_{IN}=V_{DD}$ results in V_{OUT} = 0V. With the presence of C_P and C_N, both t_{LH} and t_{HL} are larger than for a static inverter. The charging time constants are given by:

$$\tau_P = (R_{P1}+R_P)\, C_{OUT} + R_{P1}C_P$$

R_{P1} and R_{P2} are the pMOS resistances. The output is driven to a high impedance state when the clock transitions low (Φ=0). The output voltage is maintained at V_{OUT} by charge storage on C_{OUT}. The voltages V_P and V_N

across the capacitors C_P and C_N, respectively, can be changed by input. This votlage can also be affected by the gate-drain and drain-body coupling at V_N. The voltages from V_P and V_N are isolated from the output by the passgate. The body voltage of MN1 couples high due to gate-body capacitance when Φ=1 and settles at an intermediate voltage after Φ switches low. Body effect is minimized in SOI. A spike in the output occurs with input switching high and Φ low. This is due to the combined capacitive coupling from gate-drain and gate-body of MN1 and V_N. Due to source/drain parasitic bipolar leakage of the nMOS (closer to output), V_N spike (trying to pull low) is reflected on the output. As nMOS devices are added to the pull down, body effect is reduced and less variation occurs in threshold voltages.

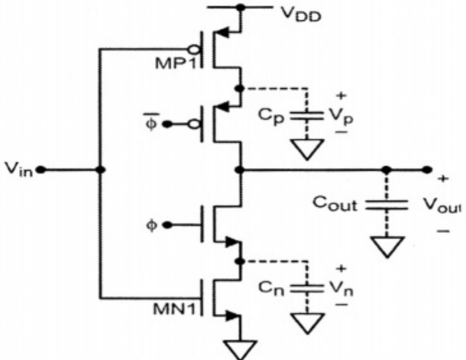


Figure 7.18(a): Clocked SOI CMOS inverter with floating bodies

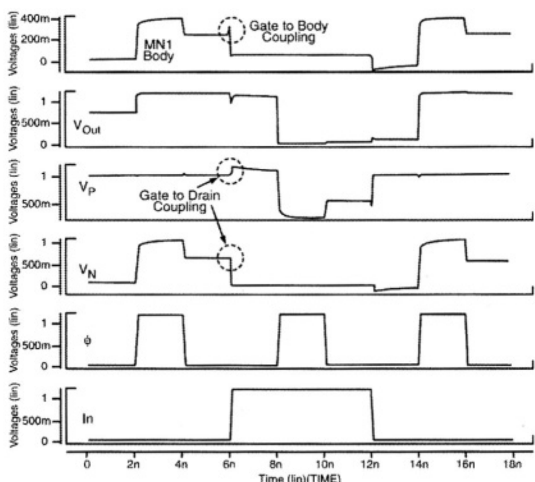


Figure 7.18(b): Waveforms of the Clocked CMOS inverter. The body voltage of MN1 raises with input for a short period of time before settling down. Once input (MN1) goes high, body switches low.

Figure 7.19(a) shows a clocked latch oppositely phased inverters. The timing is chosen to insure that when one inverter input is high, the other is tri-stated. The operation of this circuit is similar to the master slave D-flip flop. When Φ=1, the output of the first inverter is $\overline{Vin}$ and output of the second inverter is the same as input. Once the signal Φ switches low, the first stage output is tri-stated and the second stage accepts the $\overline{Vin}$ as input.

Figure 7.19(b) shows the waveforms. When Φ=0 and V_{IN} is low, output V_O is tri-stated, but pulls up when Φ switches high. When Φ switches low, output V_O couples low due to gate/drain coupling and drain/body capacitance. The body voltage is affected by switching the input voltage of the top nMOS device in the stack. While Φ=1, the output node is isolated from the input.

Output node V_X is V_T above ground during the next cycle before Φ goes low and then pulls high with Φ going high. V_O rises with Φ transitioninghigh due to drain-gate coupling and drain-body coupling. The voltage at the node between the two nMOS transistors in the pull down stack varies with gate-drain and body-drain coupling. Charge sharing affects are reduced in SOI nMOS stack compared to bulk.

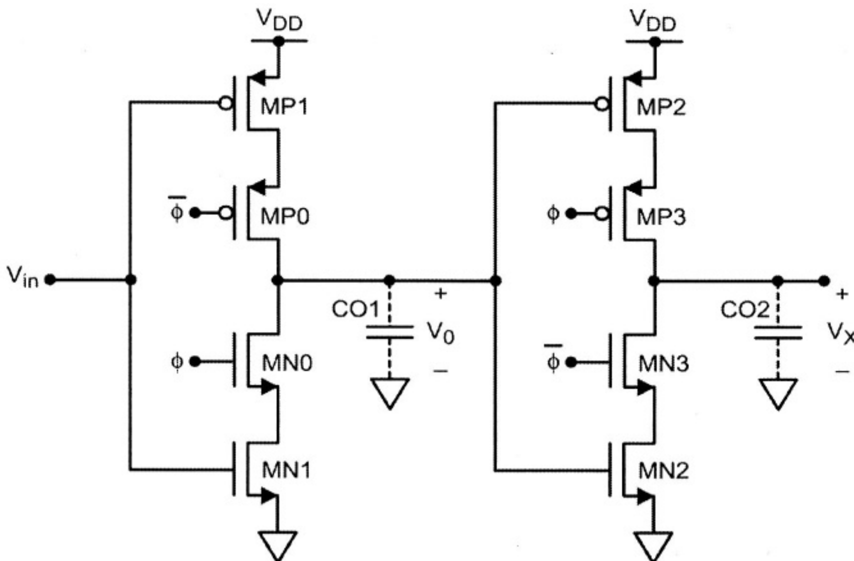


Figure 7.19(a): Clocked latch oppositely phased inverter schematic.

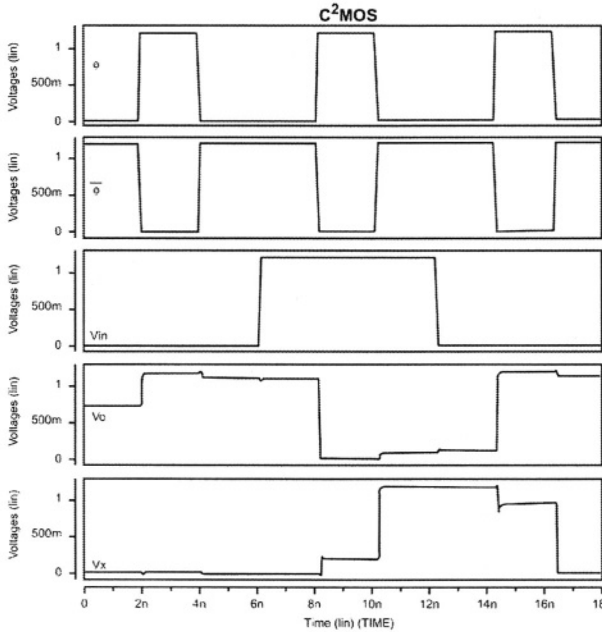


Figure 7.19(b): V_O and V_X for clocked Latch oppositely phased inverter.

Since C^2MOS is based on static logic, it is simpler to design an entire family of gates with the same characteristics. Examples are the C^2MOS NAND2 and NOR2 schematics shown in figures 7.20 and 7.21.

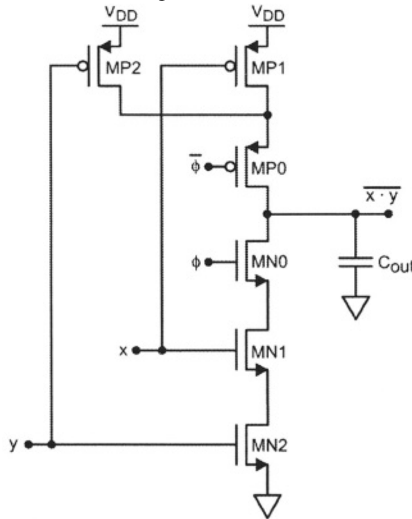


Figure 7.20: Schematic of C^2MOS NAND2 logic gate.

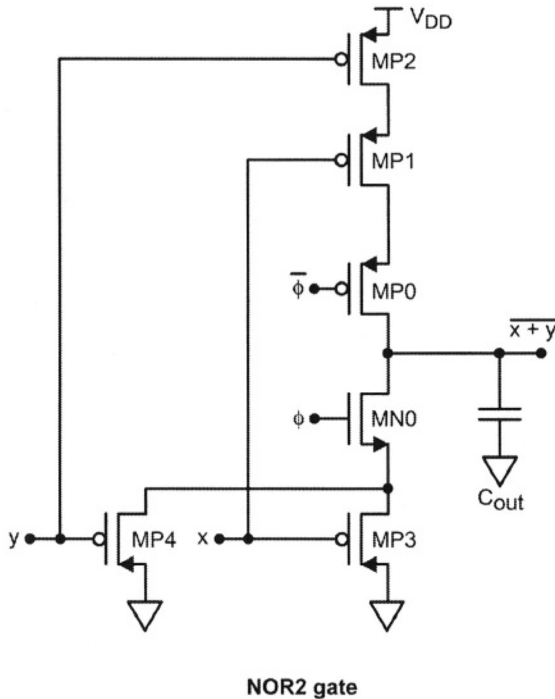


Figure 7.21: Schematic of C^2MOS NOR2 logic gate.

7.13 Pulse Stretching in Dynamic Circuits

The floating body induced transient threshold voltage exists even when the parasitic bipolar current is not significant enough to affect circuit operation. Threshold voltage variations cause a frequency dependent pulse stretching effect in partially depleted SOI CMOS inverter chains [7.14]. Such behaviour is attributed to the imbalance in charge between logic states during switching.

Figure 7.22 shows an inverter chain and depicts the pulse-stretching phenomenon. If the inverter is 'low' for a long period of time, majority carriers deplete the subsequent stage nMOS with a 'high' input. This causes low inverter inputs to accumulate most of the majority carriers, meaning that odd stages have more majority carriers than the even stages.

When the chain switches, the first stage switches with a higher current drive (lower nMOS V_T) than the second stage. The same trend extends down the entire chain. All odd stages nMOS gate inputs are LOW and even stages HIGH. Equilibrium body voltage is reached when charging and discharging currents are equal [7.4].

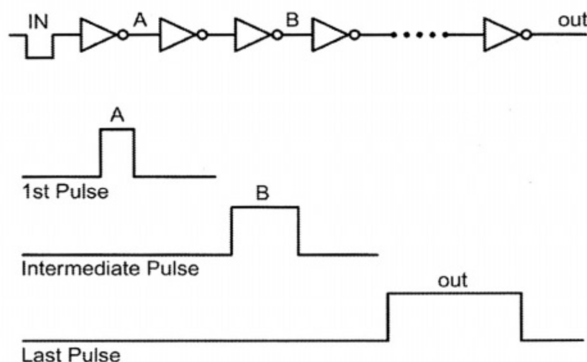


Figure 7.22: Partially depleted SOI inverter chain showing the sequence of input pulse stretching across the chain [7.14].

For even stage devices, with gate inputs HIGH and drains LOW, the body voltages are LOW, resulting in higher threshold voltage. The pMOS devices operate in a complementary fashion. Odd stage pMOS devices have reduced majority body charges and thus higher threshold voltages, while the even stage pMOS devices have more charge and lower V_T. Thus, when the input pulse rises from a sustained LOW period to HIGH, the rising edge of the input pulse propagates down the chain faster than the falling edge of the pulse. This is because all the devices involved with the propagation of the rising edge have lower threshold voltages, and the ones associated with falling stages have higher threshold voltages. The pulse, therefore, stretches as it propagates down the chain (Figure 7.22).

The pulse stretching depends on input frequency and supply voltage V_{DD}. As input frequency increases, the pulse stretching decreases because there is less time for the devices to switch and recover to equilibrium. Pulse stretching also decreases with increased V_{DD}. Drain induced body depletion reduces the equilibrium number of holes in the body and is more significant at higher supply voltages. Thus, the difference of equilibrium number of holes between the two states (figures 7.22(a) and 7.22(b)) diminishes with increasing V_{DD}.

Pulse stretching has several implications on the circuit operation. It affects the duty cycle and degrades clock skew and jitter in a clock distribution network. Circuit timing methodology is complicated and performance is degraded [7.15, 7.16]. Figure 7.23 [7.6] shows the pulse stretching per stage in a 480-stage PD-SOI CMOS inverter versus input pulse frequency and V_{DD}.

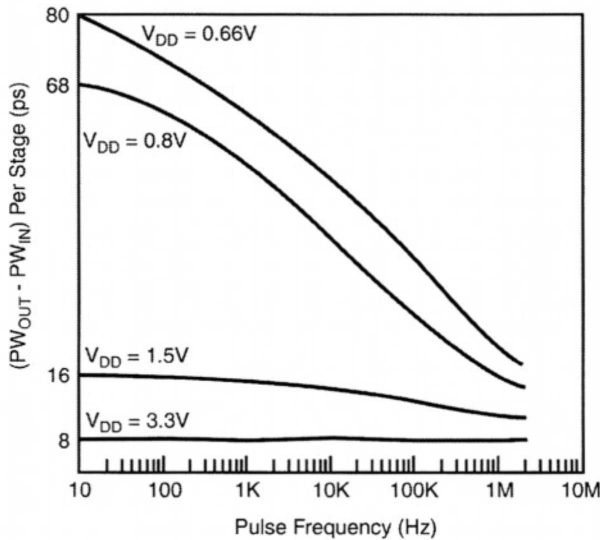


Figure 7.23: Measured pulse stretch per stage in a 480-stage PD-SOI CMOS inverter versus input pulse frequency and V_{DD}.

Figure 7.24 shows a self-resetting CMOS circuit with input and output pulses [7.14, 7.16]. The circuit uses delayed feedback loop from the output to reset the circuit in order to eliminate clock skew associated with the global clock.

Once the input signal arrives, node N1 is pulled low, and the output starts rising. Figure 7.24(b) shows the simulation waveforms when utilizing a nMOS pull- down transistor and an 11-stage inverter chain in the feedback loop.

As input IN goes high, node N1 falls and the output rises. This rising edge is fed back through a reset timing chain and then applied to precharge transistor MP0. Node N1 then falls, allowing the output to switch low. When node N1 is pre-charged HIGH, the output goes to LOW. The falling edge of the output triggers the turn 'ON' of MP0.

Pulse stretching broadens the pre-charge pulse V_{PCH}, since the rising edge of the output pulse propagates faster through the delay reset chain than the falling edge, thus restraining the evaluation cycle and degrading cycle time. The simulation waveforms shown here do not reflect much delay, but as the

number of inverter stages increases in the feedback loop, the feedback pMOS input pulse is widened.

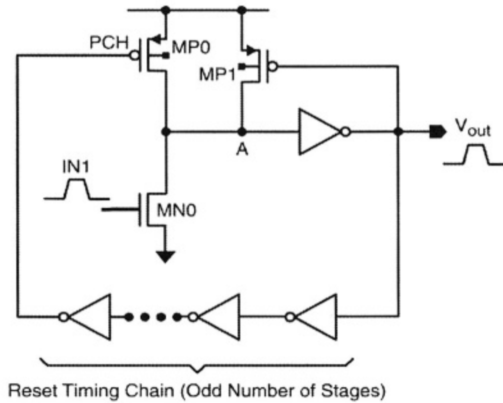


Figure 7.24(a): Schematics of a self-resetting CMOS circuit (SRCMOS) circuit. MPO is the precharge transistor, and MP1 is the feedback half latch to improve the noise margin of the dynamic node. 11-stage inverter chain in the feed back path and a single nMOS in the pulldown stack.

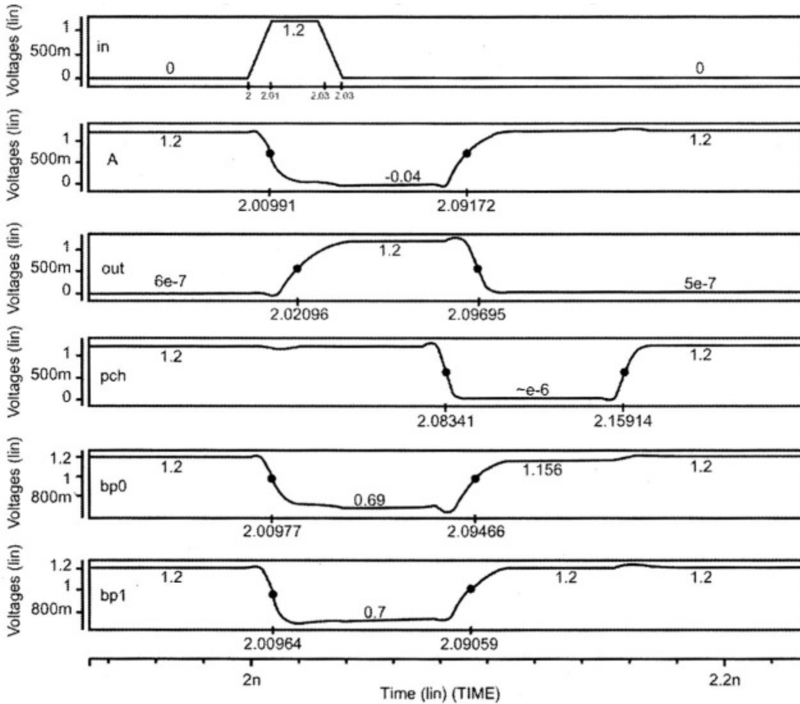


Figure 7.24(b): Simulated node and body voltages in SRCMOS circuit.

It is important that input pulses align properly to ensure sufficient overlap. Proper separations between pulses prevent them from overlapping. The pulse alignment becomes much more complicated in the presence of the pulse stretching effect [7.14].

7.14 Dynamic Wide-OR

Pass-gate leakage of a dynamic wide-OR (figure 7.25) is a function of the width of the OR devices. If the combined width of all the OR devices is large, then pass gate leakage can cause circuit failure at higher voltages. This is normally not a problem at nominal voltages or when the keeper is large, but at process and operating corners, problems may occur. Solutions to avoid this effect are to increase keeper device strength, eliminate MP3, or discharging node N2 periodically, e.g., during reset (figure 7.25(b)) [7.17].

Once the CLK is 'on', devices MN1 and MN3 are low, and the input to MN2 turns 'on' after a few nanoseconds, the body voltage of MN2 falls low, and the body voltage of MN_X reduces due to the presence of a discharge path to ground. The body voltage of MN2 rises (not shown) due to body to gate coupling and then settles low once the device MN2 is ON. It has been reported that [7.17], at Vdd=2.5V, parasitic leakages can cause failure if the combined widths of the OR gates are higher (>200μm). Our simulations do not show major failures until about Vcc=3.5V (and combined device widths of 500μm). The output node remains low with very small spikes caused due to parasitic leakages.

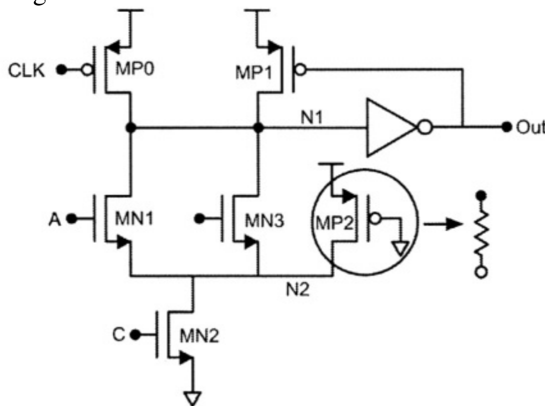


Figure 7.25(a): Wide-OR sensitive to pass-gate leakage [7.17]. nMOS MN2 has to be a big device in order to discharge to GND. MP2 has to be a weak device. Large combined width of OR nMOS gates can cause problem at very high supply voltages.

7.15 Non Overlapping Clocks

A string of pass transistors/inverters (figure 7.26) represents a dynamic shift register. When signal Φ1 goes HIGH, the first and third stages of the register are enabled. Data are passed from the input to point S0 and from point S1 to S2. If Φ2 is low while Φ1 is high, data cannot pass from S0 to S1 and from S2 to S3. If both Φ1 and Φ2 are high at the same time, the input of the shift register and the output are connected together, which is not acceptable in a shift register application.

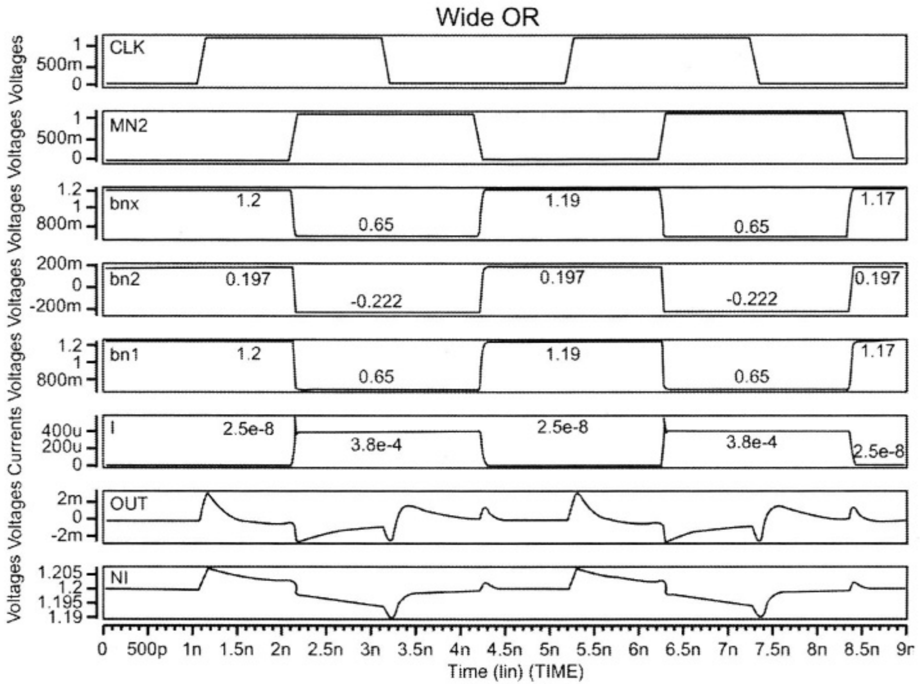


Figure 7.25(b): Simulation waveforms for dynamic Wide-OR gate.

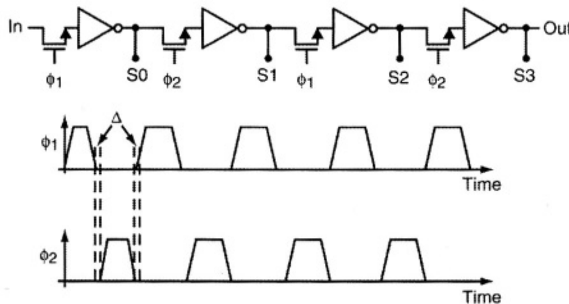


Figure 7.26: Dynamic Shift register with non-overlapping clocks.

The purpose of the inverter between pass transistors is to restore logic levels, since the nMOS pass transistor passes a high with a threshold voltage drop. A buffer instead of an inverter would eliminate logic inversion between stages. The clocks used in this dynamic circuit must be non-overlapping or logically Φ1.Φ2 = 0 [7.18].

A simple logic inversion will not generate non-overlapping clock signals. Figure 7.27 shows an example of non-overlapping clock generation. A clock signal is used to generate a two-phase non-overlapping clock. The gap between clocks is set by the delay through the NAND gate and the two inverters at the NAND gate output. If the input clock goes high, Φ1 is forced high and Φ2 low. After Φ1 goes low, Φ2 settles high.

When driving long transmission lines such as polysilicon interconnects, where the rise time of the signals can be significant, a large number of inverters may be required. Line drivers or a string of inverters can be used to drive large capacitances as part of the delay in the non-overlapping clock generation circuit.

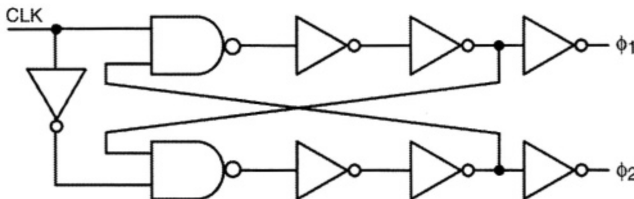


Figure 7.27: Non Overlapping Clock Generation Circuit

7.16 Pass transistor based non-overlapping clocks

A common and important application of pass gates is the lock/timing control in various latch designs [7.19]. Figure 7.28(a) depicts a two-level latch with two non-overlapping clocks, Φ1 and Φ2. After passing a high state, assume Φ1 switches low and input signal 'IN' switches low. When Φ1 is low, the pass-gate at the input to the second stage latch is supposed to be 'off', while the pass gate in the feedback loop of the first stage latch is 'on' to hold the state of the latch.

Significant parasitic bipolar current flows through the 'off' nMOS of the input pass gate to pull down node N1. Since the pass gate in the feedback loop is 'on', the pMOS in the feedback inverter fights the parasitic bipolar current to restore node N1. The result is a transient voltage dip of 'A' (figure 7.28(b)) [7.20]. The complementary situation is of lower concern as impact

ionization of the pMOS is reduced and current gain of the parasitic PNP transistor is lower [7.6]. The voltage dip on node N1 is lower at supply voltages of 1.5V. Figure 7.28(b) shows the node voltages when Φ1 and Φ2 switch.

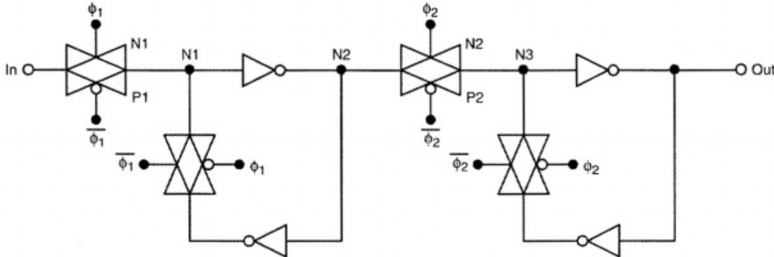


Figure 7.28(a): Pass transistor based 2-Φ Non Overlapping Clocks

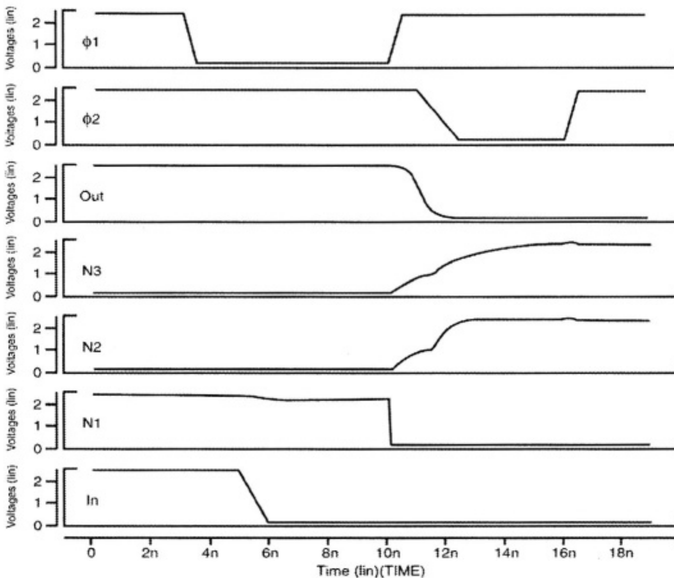


Figure 7.28(b): Waveforms of 2-phase Non Overlapping Clocks. The node voltage N1 dips slightly below Vdd once the input "IN' falls.

In some high-density designs, the feedback loop pass-gate is removed, and is replaced with a trickle inverter constructed of low gain devices. It is important to make sure that the trickle inverter has enough strength to overcome the parasitic bipolar current and restore node L1, to prevent the latch from settling in the wrong state.

The parasitic bipolar can cause errors in pseudo-two-phase dynamic logic (figure 7.29) [7.19, 7.20]. During the condition when C2 is LOW, the pass

gate to the second stage is 'off', and the second stage is evaluating. If data stored in the gate capacitance is HIGH and the first stage is evaluated LOW, then following the falling edge of C1, parasitic bipolar current flows through the nMOS in the OFF pass gate, discharging node N3.

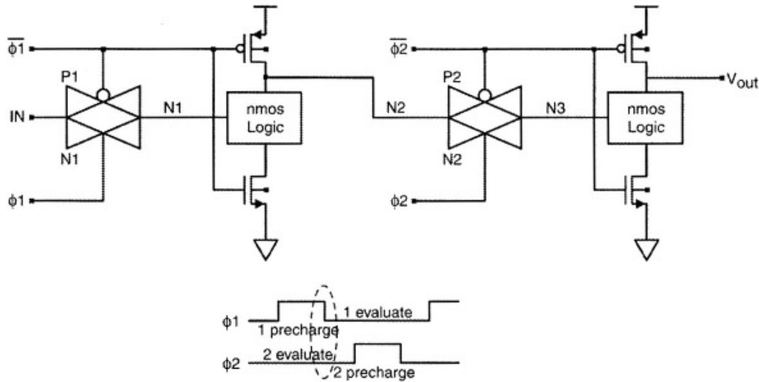


Figure 7.29: Schematic of a pseudo two-phase dynamic logic circuit.

7.17 Low Power SOI Techniques

We here consider dynamic logic specific low power SOI structures. More general low power techniques are discussed in chapter 12. Isolated bodies in SOI offer a degree of freedom for the designers to enhance the power and delay performance. The body contact can be used to achieve high speed, low voltage operation in active mode and low leakage, low power operation in standby mode. Circuits with body tied to gate reduce threshold voltages and increase drive currents. Body tied to gate circuits are often the dynamic threshold voltage MOSFETs (DTMOS). Body potential is low and threshold voltages are high in standby mode to reduce leakage. DTMOS circuits are suitable for low voltage applications. The power supply voltage is limited to less than one diode drop to avoid turning 'on' of the parasitic bipolar transistor. A circuit limiter to limit the body to source junction voltage to less than a diode threshold is necessary at higher supply voltages.

7.17.1 Dynamic Threshold SOI CMOS

A dynamic threshold voltage MOSFET inverter (DTMOS) includes static, switching and short circuit power dissipation. For the switching power dissipation, both load and device parasitic capacitances are considered.

Body tied to gate SOI dynamic threshold voltage MOSFET do not exhibit floating body effect and have better sub-threshold characteristic than conventional SOI MOSFETs. As gate voltage is increased, the gate couples onto body and forms body-source forward biased diode, reducing its

threshold voltage as a result of body effect. Hence, threshold voltage can be changed dynamically without the application of additional control circuitry. DTMOS can be operated at low power supply while still maintaining a large drain current for high-speed application, without introducing any extra standby current. Further operational restrictions are discussed in chapter 3. The advantages of dynamic threshold MOS are partially offset with a number of drawbacks. DTMOS devices connected in the pull up/pull down configuration such as inverters have increased power dissipation (compared to a conventional SOI inverter). Static power dissipation increases due to leakage of the forward biased body-drain/source P-N junctions and capacitances associated with the parasitic P-N junction diodes. These add to power dissipation during switching, increasing short circuit power dissipation [7.14].

7.17.2 Dynamic Threshold Multithreshold CMOS Logic

Dynamic threshold voltage control may be used in multi-threshold CMOS (MTCMOS) circuit [7.21] combining low V_T-CMOS logic gates with variable V_T sleep mode. This type of threshold voltage control is suitable for sub 1V applications where low standby power is very important. Figure 7.30 shows circuit delays of bulk MTCMOS and PD-SOI MTCMOS as functions of the supply voltage. The SOI-MTCMOS offers greater than 2X delay improvement and can also operate down to 0.4V, while bulk MTCMOS loses functionality around 0.5V [7.6, 7.22].

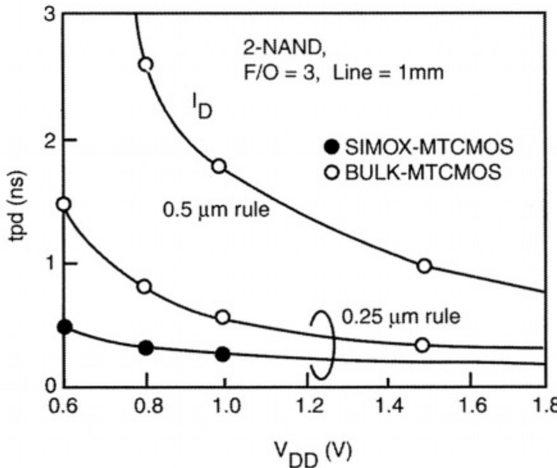


Figure 7.30: Comparison of basic logic gate delays of bulk MTCMOS and SIMOX-MTCMOS.

Figure 7.31 shows a multi-threshold CMOS logic operation at low voltages. Using a combination of fully depleted and partially depleted SOI transistors,

performance is achieved in active cycle and low power in standby mode. The fully depleted pMOS device is used to lower the body voltage of the partially depleted pMOS, resulting in decreased threshold voltage and lower operational supply voltage. The pMOS body voltage is raised during standby, increasing its threshold voltage and reducing standby current.

During the active cycle, when SL is low, if the pMOS body is tied to gate and the supply voltage is larger than one diode threshold, significant current flows from to the drain-body p-n junction to the gate. This can be suppressed by tying the pMOS body to its gate through a low V_T pMOS reverse biased diode (figure 7.31(b)). When SL is "high" (sleep mode), the body of the sleep mode control pMOS will sit at one (low) diode drop below SL.

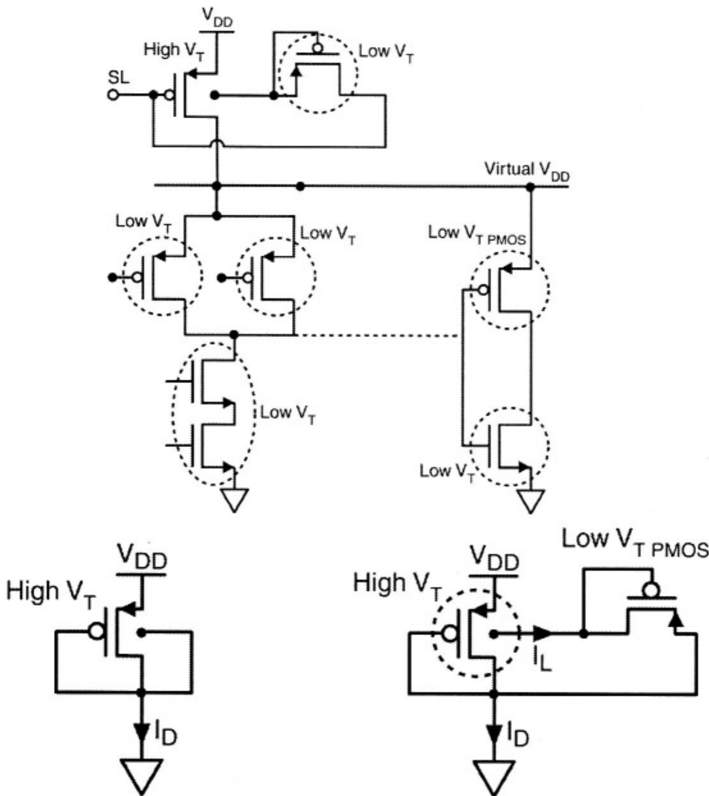


Figure 7.31: (a) SIMOX MTCMOS circuit, and (b) Current characteristics of DTCMOS and variable high V_T MOSFET with reverse biased MOS diode [7.6, 7.22].

7.17.3 Dynamic Threshold Pass Transistor Logic

Dynamic threshold techniques may also be applied to pass transistor logic [7.23]. Figure 7.32(a) shows a complementary pass transistor logic (CPL) network. Degradation in drive capability and threshold voltage loss in transmitting a "high" signal restrict the low voltage applications of this circuit. Figure 7.32(b) shows the pass gate bodies and the gate tied together. Body bias control can also be applied to the buffer (figure 7.32(b)).

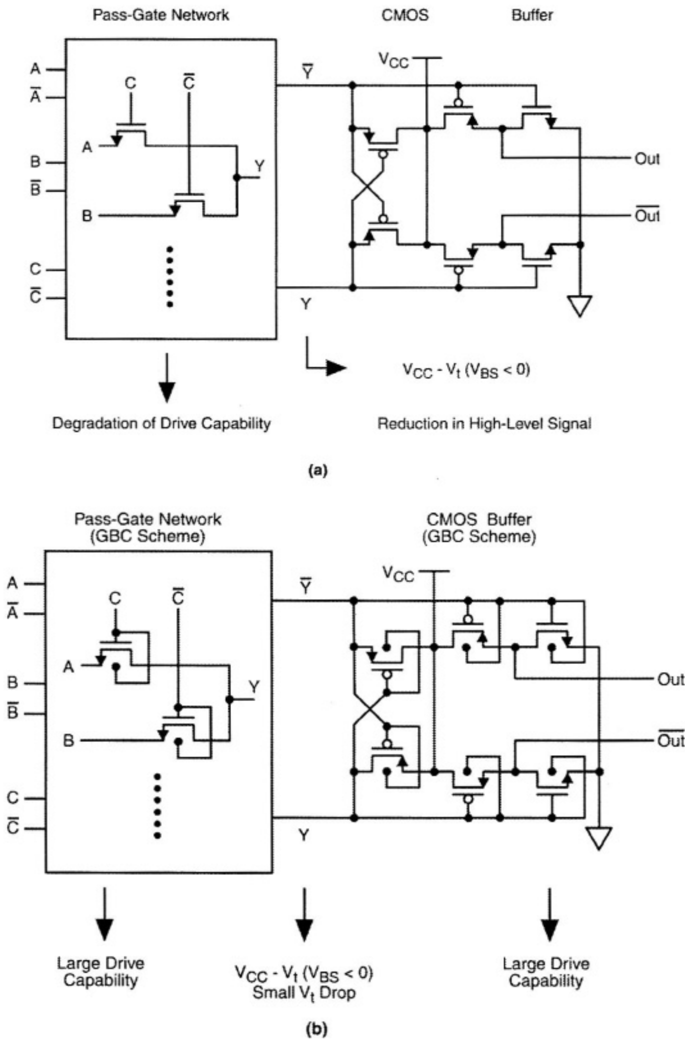


Figure 7.32: (a) Conventional complementary pass gate logic (b) Gate body connected SOI pass gate logic.

7.17.4 Dynamic threshold voltage Full Adder

Because of the low threshold voltage of the "ON" MOSFETs, the dynamic threshold voltage scheme reduces the delay of a full adder to approximately a third of the conventional SOI complementary pass gate logic at a supply voltage of 0.5V (figure 7.33) [7.6]. At a fixed delay of 2.5nS, the lowest operation voltage improved by 0.18V. The buffer section can be further enhanced by crosscoupling pull-ups (Figure 7.33). The bodies of the pull up pMOS can be connected to their respective gates or to the inputs (figure 7.34). With the bodies tied to gates and gates cross-coupled to the output nodes of the opposite phases, the threshold voltages of the "ON" pull up pMOS remains high until the output nodes change. It is therefore advantageous to connect the body to the input so the pMOS threshold voltage decreases as soon as the input changes, improving circuit speed and reducing short circuit current [7.23].

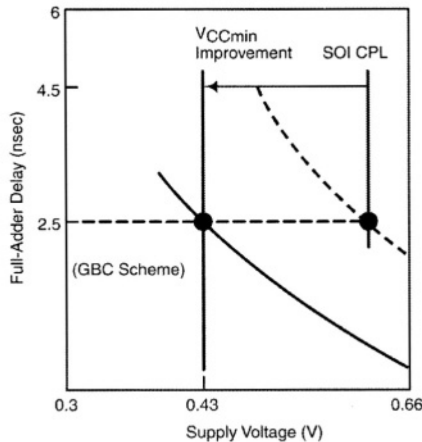


Figure 7.33: Simulated full adder delay of conventional complementary pass gate logic and gate body connected pass gate logic.

To extend the operating voltage of the body bias controlled SOI pass gate circuits to greater than one diode voltage, the boosted ground scheme as shown in figure 7.35(a) is a preferred method [7.24]. Here, a reference voltage generator, composed of drain body connected nMOS SOI transistor and a series resistor, generates reference voltage, V_{REF}. The bias applied to the main circuitry, V_{DD}-V_{REF}, is lower than the soft breakdown voltage of the body bias controlled SOI devices, thus reducing leakage. The boosted ground scheme is driven by a body source tied SOI nMOS transistor to avoid the floating body effect and to ensure high threshold voltages to suppress the standby leakage current of the main circuitry. A charge pump boosts the supply voltage for stable operation of the analog portion of the circuits.

Figure 7.35(b) shows the maximum frequency and active power as functions of the supply voltage for a one stage, 32-bit ALU using the boosted ground scheme. The operating voltage range increases while restricting IDDQ. The maximum frequency remains relatively constant as supply voltage increases over 0.5V since the highest voltage applied across the main circuitry is limited to 0.5V.

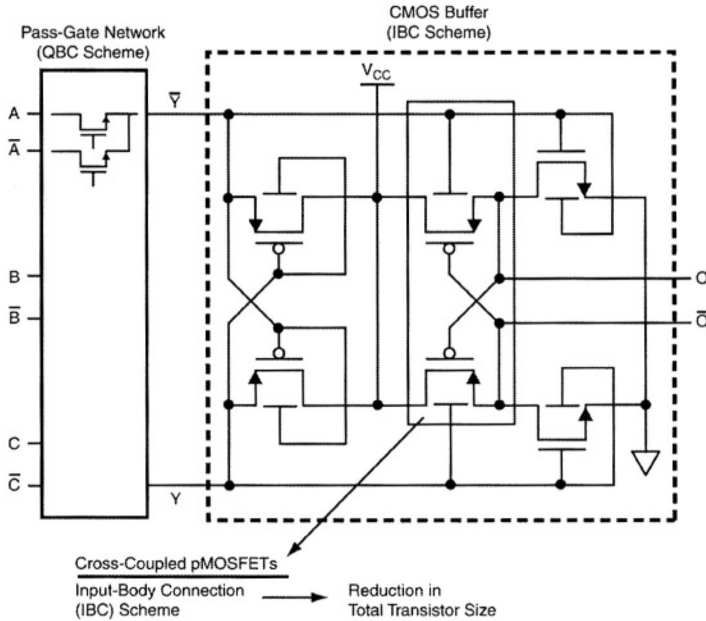


Figure 7.34: SOI CMOS buffer with cross coupled pMOS and body input connection.

Capacitive coupling can be used to maintain body bias control and extend operating voltage range. Figure 7.36(a) shows the schematic of a double gate driven CMOS (DGCMOS) circuit where the body of the switching device is dynamically connected to the gate through a capacitor [7.25]. Coupling capacitor C_B is added between the body of the pMOS and the input of the inverter and a p-n junction diode D1 is added between the pMOS and supply.

When the input switches from V_{DD} to ground, the pMOS is turned 'on' and the body voltage couples down (V_{BP}=~0.5V). This reduces its threshold voltage and increases drain current. When the pMOS body voltage falls below a diode drop from supply, the pMOS body to source junction is

charged. V_{BP}, settles at V_{DD}-V_{SBD} (~0.52V), where V_{SBD}, is the threshold voltage of the source body diode (figure 7.36(b)).

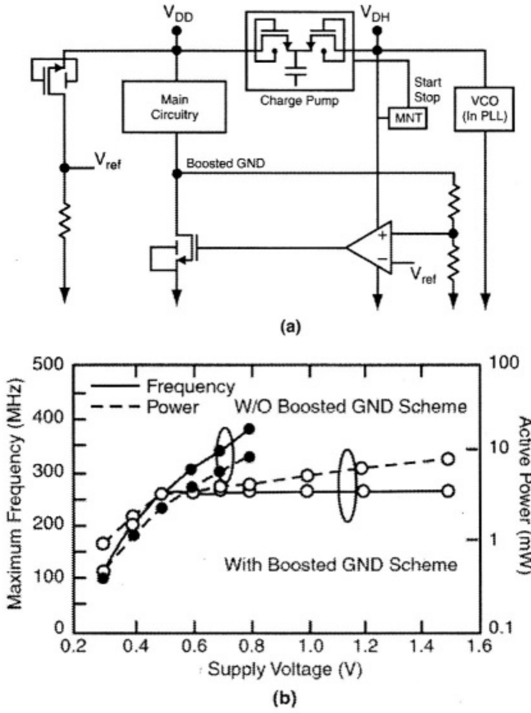


Figure 7.35: (a) Body Bias controlled SOI pas gate logic with boosted ground scheme (b) maximum operating frequency & active power versus supply voltage [7.6, 7.24].

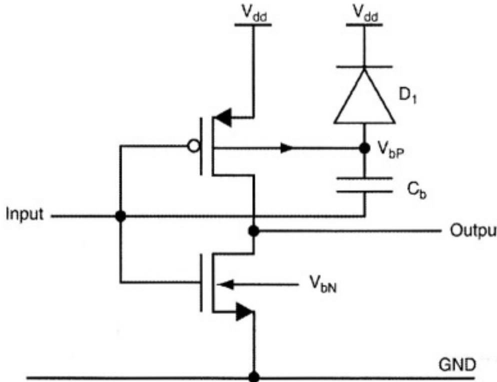


Figure 7.36:Double gate driven pMOS for an inverter [7.25].

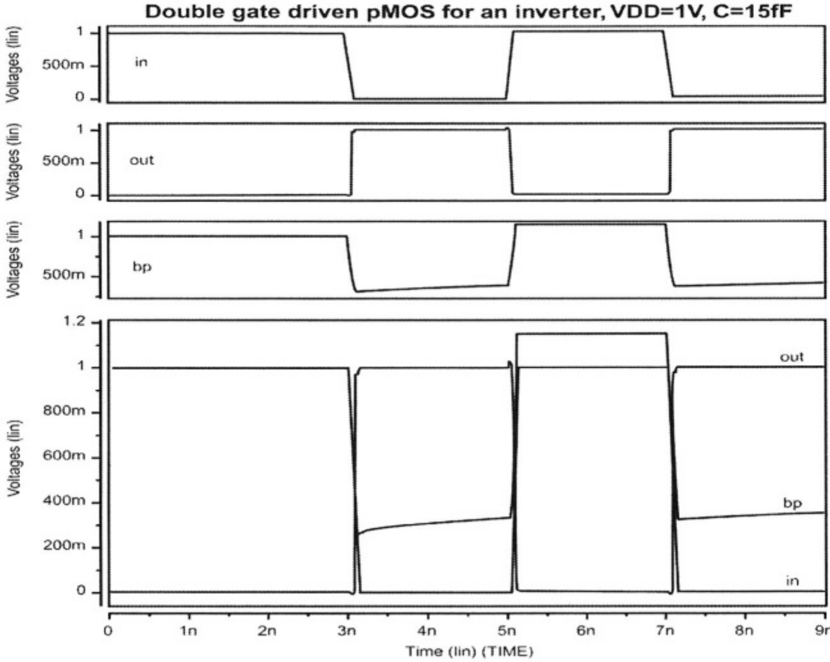


Figure 7.36(b): Inverter input, output and node waveforms. The pMOS body voltage rises from 0.15V to 1.4V during input switching low to high.

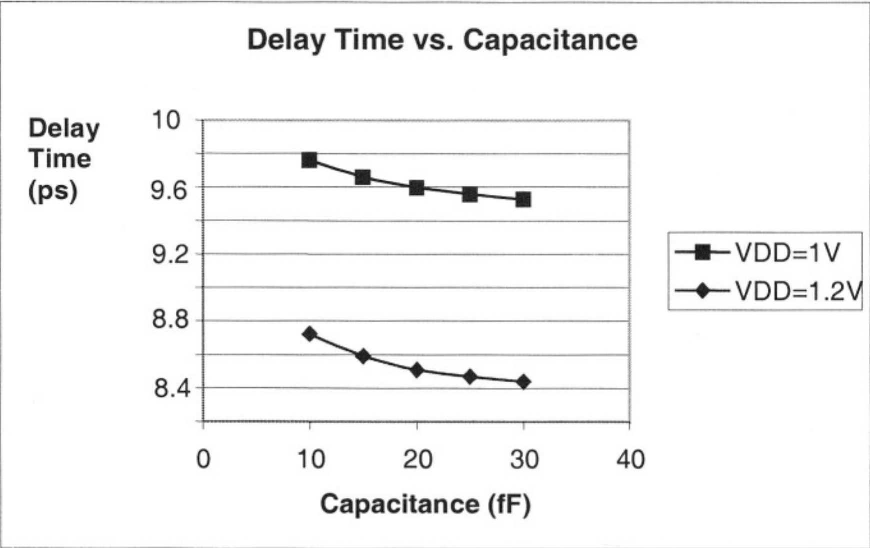


Figure 7.36(c): Delay versus capacitance for V_{DD}=1V and V_{DD}=1.2V. Delay time reduces with increased capacitance.

When the input switches from GND to V_{DD}, the pMOS is turned off. The pMOS body voltage V_{BP} rises from V_{DD}-V_{SBD} (~0.15V) to V_{DD}-V_{SBD}+V_{DD} (~1.4V) and eventually drops to V_{DD}+V_{DIODE} (where V_{DIODE} is the threshold voltage of the diode D1 and V_{SBD} is the threshold voltage of source-body diode) through the discharge diode D1.

This circuit can operate at any supply voltage compatible with the device design without the leakage of some other body bias controlled schemes. Larger C_B increases V_{BP} modulation and reduces leakage significantly. The drawback of this technique is the increased input capacitance, resulting in a larger load on the previous stage. C_B therefore requires a design tradeoff between the circuit speed and leakage. The delay time reduces with increased capacitance in SOI (figure 7.36c).

Device operation may become unreliable at high frequencies because of the feedback through C_B. Another alternative technique that may overcome this frequency limitation is the active body bias scheme [7.26]. Figure 7.37(a) shows two active body biased driver circuits.

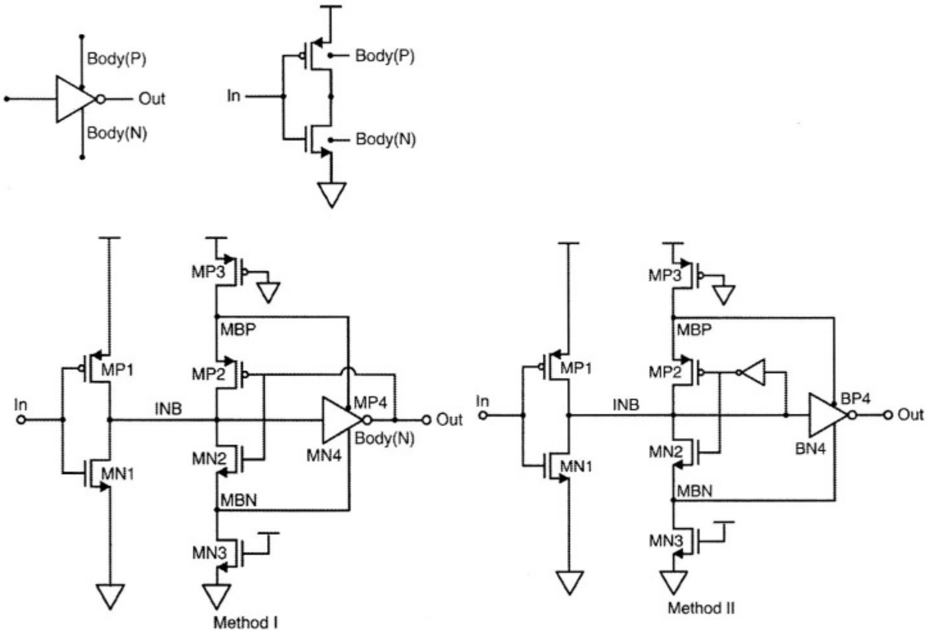


Figure 7.37(a): Active body bias SOI CMOS drivers.

The pMOS and nMOS transistors in the stack MP3/MP2/MN2/MN3 dynamically control the body voltage of the switching transistors MP4/MN4

in method I. If the input is "low", INT is high and the output is low (figure 7.37(b)). Then the PMOS MP4 drain is low and the gate of MP2 is low (MP2 ON and MN2 OFF). The body voltage of MP4 falls slightly after the input switches. MP2 functions like the feedback half latch in a dynamic circuit [7.19]. If the input then switches low to high, MP2 turns off and MN1 turns 'on'.

During initial switching, transistor MP2 is ON and the transistor MN1 is 'on', providing a current path from MP3 through MP2 and MN1 to ground. Transistor MP3 is sized such that its current is substantially smaller than that through MN1/MP2. Leakage current through MP3 pulls down the voltage at the intermediate node between MP3 and MP2 (body voltage of the MP4), reducing the threshold voltage of MP4, and speeding up the MP4 pull up. In standby, the output is 'high' turning OFF transistor MP2 and disconnecting the leakage path.

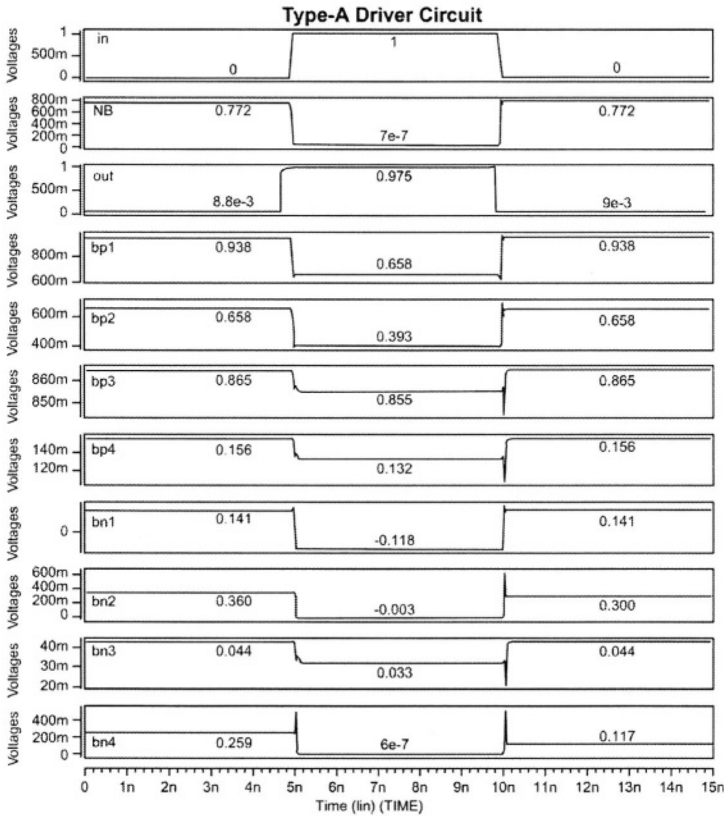


Figure 7.37(b): Active body bias SOI CMOS driver output voltages for the circuit shown as method I (figure 7.37(a)).

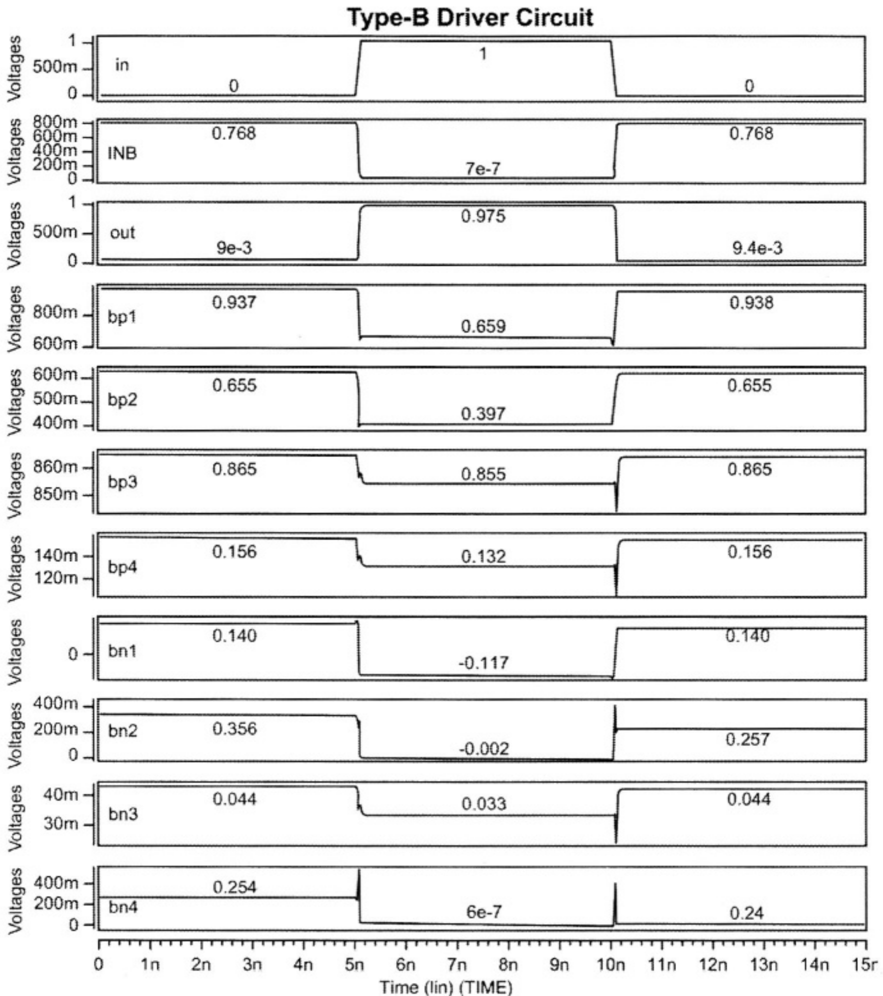


Figure 7.37(c): Active body bias SOI CMOS driver output voltages for the method II circuit of figure 7.37(a)

In the method II circuit, (figure 7.37(a)), feedback is through a small inverter. This provides long low threshold voltage periods for MP4 or MN4. Figure 7.37(b & c) shows the output voltages for the method I & II circuits.

Figure 7.37(d) shows the drain current through MN1 for method I and II circuits for input rising and falling pulses in bulk. High to low transition dissipates less power compared to low to high transitions for both the methods. Figure 7.37(e) compares the rising and fall times for the two methods in SOI as well as their standby currents and Figure 7.37(f) shows the same data in bulk designs [7.6].

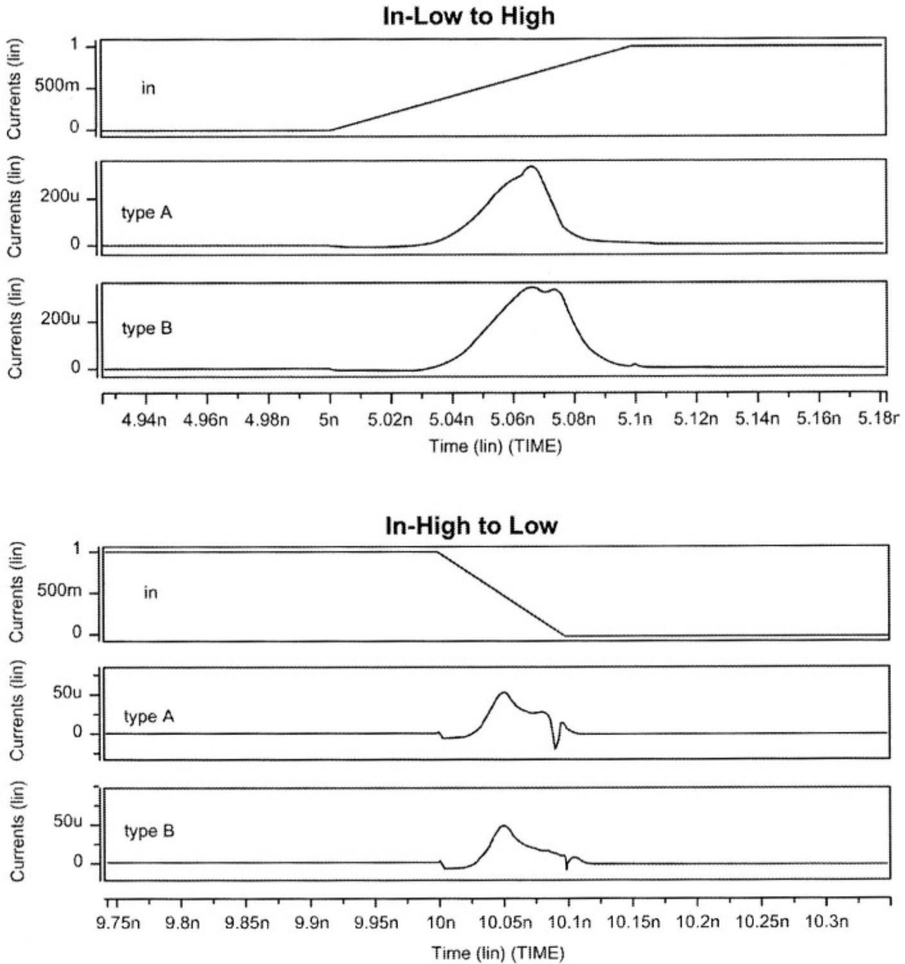


Figure 7.37(d): The drain current through MN1 for method I and II circuits for input rising and falling pulses in bulk.

The method II SOI is faster than method I in SOI and both the SOI methods are faster than bulk during rising outputs, and have faster pull downs during falling edges compared to bulk. Both the methods are also simulated in bulk (noted by bulk A and bulk B in the waveforms).

The delay for the method II circuit in SOI can be reduced by re-sizing transistors MP4 and MN4. Method II is faster than method I for rising outputs and vice versa for falling outputs. Figure 7.38 shows the comparison of various methods at different supply voltages.

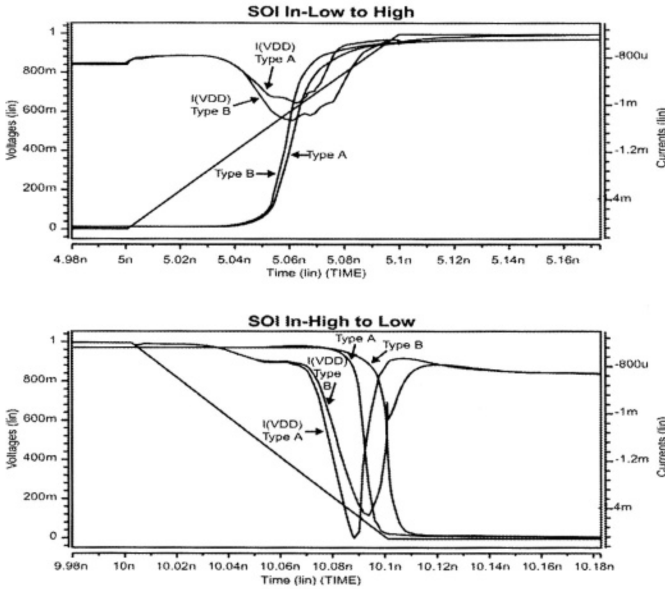


Figure 7.37(e): High to low and low to high transitionsand currents in SOI for methods I and II.

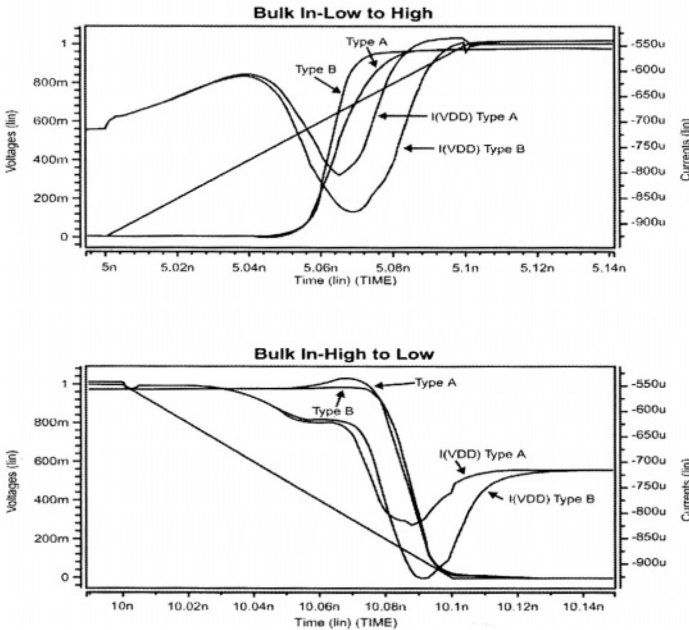


Figure 7.37(f): High to low and low to high transitions in bulk for methods I and II.

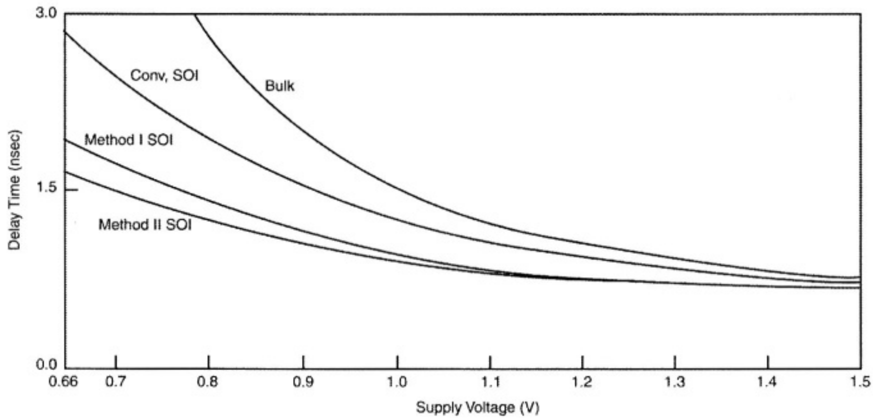


Figure 7.38: Delay time vs. supply voltage for various methods in bulk and SOI [7.26].

7.17.5 Dynamically Body Bias SOI CMOS Inverter

Figure 7.39(a) shows a dynamic body biased controlled inverter [7.27 - 7.29]. The bodies of the primary inverter transistors are controlled through another pair of alternate pMOS and nMOS transistors. The alternate transistor gates are driven from the primary gate inputs. The drains are connected to the output and the sources are connected to the bodies of the primary transistors.

As the input switches from low to high (Figure 7.39(b)) during the initial phase, the body voltage of the primary nMOS transistor increases through the alternate nMOS transistor in a source follower configuration. This reduces the threshold voltage of the primary nMOS transistor and improves pull down speed. Once the output switches low, the body of the primary transistor is reduced to zero. During the initial pull-down state, current through the alternate transistors M_{NA} charges the body of MN0 and helps pull down the output node directly. Conversely, once the output falls, capacitance of the primary transistor M_{ND} slows down the final part of the switching curve.

The threshold voltage of the subsidiary transistors can be designed to be substantially lower than the primary driving transistor since they do not provide a direct leakage current path during standby. This scheme works for any supply voltage. When the load capacitance is small, the delay improvement of the present scheme is not significant due to the increased capacitance caused by the alternate devices. As load capacitance increases, the improvement becomes significant in the dynamically body controlled inverter [7.6].

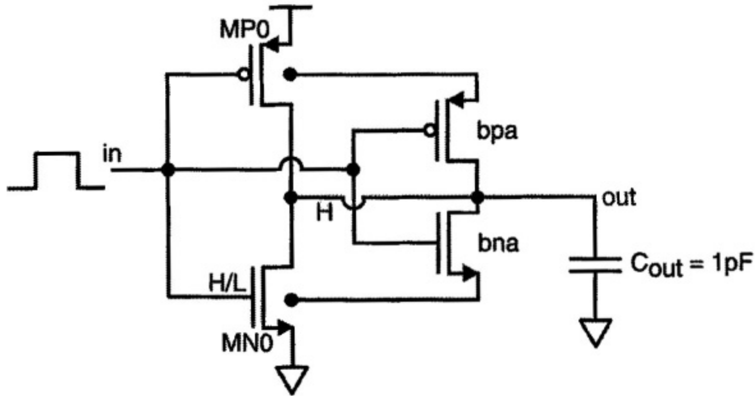


Figure 7.39(a): Schematic of a SOI CMOS dynamically body controlled inverter [7.27]

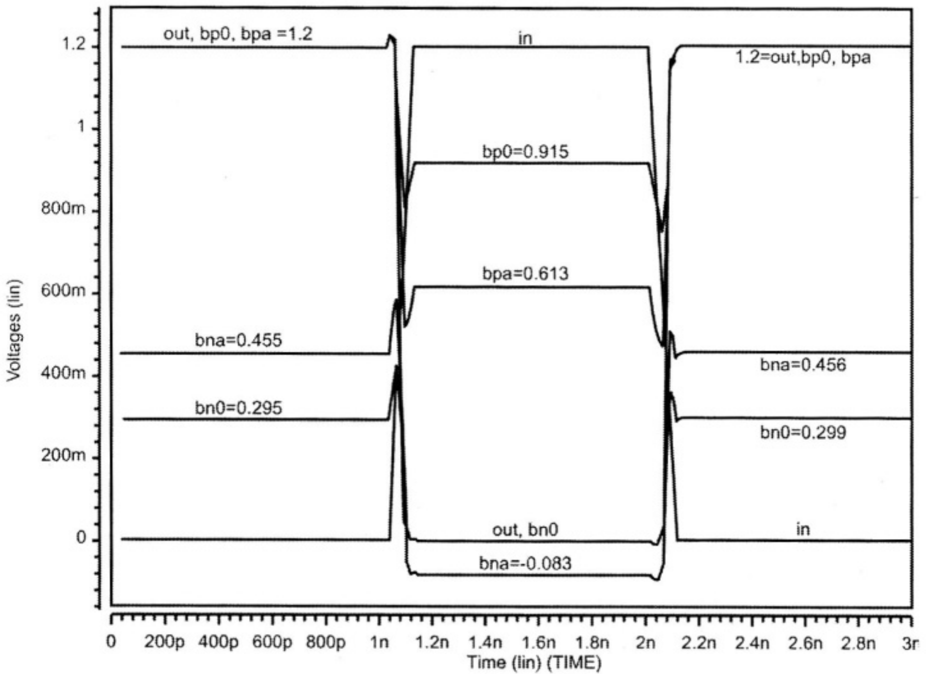


Figure 7.39(b): Waveforms for input switching low to high displaying the body voltages of the primary and alternate transistors.

Design strategy can also affect the leverage of SOI. In a dual design approach, requiring designs to work on both bulk and SOI technologies with only minor mask level adjustments, the performance in both the bulk and SOI versions is likely to be sacrificed.

The necessity to reserve area in the bulk version to accommodate body contacts in the SOI version increases the area in the bulk design. On the other hand, circuits in SOI may suffer from more severe supply and ground bounce, and degrade performance due to inadequate amount of decoupling capacitance. The long paths in the SOI may differ significantly from the bulk due to all the affects discussed above. Thus, timing optimization for the bulk version may not improve. The wire delay also degrades the advantage of the SOI version substantially.

In a silicon on insulator-only design, a global trade-off can be pursued to enhance the overall performance. For on-chip cache memory, decoupling capacitance can be used. Smart body contacts and driver configurations can be used to reduce the wire delay significantly, which tends to improves performance.

Circuit simulations and floating bodies require substantially more memory and simulation time. Simulation convergence for the initial DC solutions tends to be a problem since the initial DC potential of floating nodes in the circuits is set by small leakage currents with long time constants. The floating body related effects further complicate the identification and definition of various timing modes.

For memory applications, the circuits is more confined. The circuit design, timing, optimization and use of selective and/or smart body contacts to improve the performance appear more straightforward than for mainstream logic applications. Timing large scale logic circuits with all the floating body related effects represents the single most challenging task in bringing SOI into mainstream microprocessor application.

The device lifetime does not appear to be a problem for SOI. Adequate hot carrier lifetimes have been demonstrated for 0.35μm partially depleted SOI devices with both body floating and body grounded schemes [7.30]. Device lifetime (to 10% reduction in saturation current) similar to bulk technology has also been demonstrated in 0.25μm SOI-CMOS technology, and 10-year operation at 1.8V with effective channel length down to 0.1μm appears feasible [7.31].

References

[7.1] K. Bernstein & N. J. Rohrer, "SOI Circuit Design Concepts", Kluwer Academic Publishers, January 2000, ISBN 0-7923-7762-1

[7.2] D. H. Allen, et. al., "A 0.2um 1.8V SOI 550MHZ 64b PowerPC microprocessor with copper interconnects", Proceedings of 1999 IEEE ISSCC, 1999, pp. 438 – 439.

[7.3] A. Wei, et. al., "Minimizing floating body induced threshold voltage variation in partially depleted SOI CMOS" IEEE Electron Device Letters, Vol 17, Aug 1996, pp. 391-394.

[7.4] J. Uyemura, "CMOS Logic design style", Kluwer Academic Publishers.

[7.5] C. Kim at el., "Parallel Dynamic Logic with speed enhance skewed static logic", ISCAS 2000.

[7.6] C. T. Chuang, et. al., "SOI for Digital CMOS VLSI: Design considerations and Advances", IEEE Proceedings, Vol 86, No. 4, April 1988, pp. 689 – 720.

[7.7] K. Yano, et. al., "A 3.8nS CMOS 16x16 bit multiplier using complementary pass transistor logic", IEEE JSSC, Vol 25, Mar 1990, pp. 388 - 395.

[7.8] M. Suzuki, et. al., "A 1.5nS 32b CMOS ALUin double pass transistor logic", ISSCC Dig. Tech. Papers, 1993, pp. 90 – 91.

[7.9] Y. Sasaki, et. al., "Multilevel pass transistor logic for low power ULSI", Dig. Tech. Papers, Symp. Low power Electronics, 1995, pp. 14 - 17.

[7.10] C. T. Chuang, et. al., "Dual mode parasitic bipolar effect in dynamic CVSL XOR circuit with floating body PDSOI devices", VLSI Technology, Taipei, 1997, pp. 288 – 292.

[7.11] M. M. Pellela, et. al., "Low voltage transient bipolar effect induced by dynamic floating body charging in scaled PD-SOI MOSFET's" IEEE Electron Device Letters, May 1996, pp. 196 - 198.

[7.12] P. F. Lu, et. al., “Floating body effects in PD-SOI CMOS circuits”, Proc. 1996 Int. Symp. Low power Electronics and design, Monterey, CA, Aug 12-14, 1996, pp. 139 - 144.

[7.13] J. B. Kuo, et. al., "Compact MOS/Bipolar Charge-Control Models of Partially-Depleted SOI CMOS Devices for VLSI Circuit Simulation -- SOI - Technology (ST)-SPICE", ESSDERC, 1999.

[7.14] J. B. Kuo, “Modeling of Low-Voltage SOI CMOS VLSI Circuits”, IEEE SCV Solid State Circuits, April 2001 Meeting, (Univ. of Waterloo).

[7.15] A. Pellela, et. al., “A 2nS access, 500Mhz 288Kb SRAM Macro”, Symp. VLSI Circuits, 1996, pp. 128 – 129.

[7.16] T. I. Chappell, et. al., “A 2nS cycle, 4nS access 512Kb CMOS ECL SRAM”, ISSCC Dig. Technical papers, 1991, pp. 50 - 51.

[7.17] G. Shahidi, et. al., “Partially Depleted SOI Technology for digital logic” ISSCC 1999, pp. 426 –427.

[7.18] R. Jacob Baker, et. al., “CMOS Design, Layout and simulation”, IEEE Press, ISBN : 0-7803-3416-7

[7.19] N. Weste & K. Eshragian, “Principles of CMOS VLSI Design – A system perspective”, Reading, MA: Addison-Wesley, 1988.

[7.20] P. F. Lu, et. al., “Floating body effects in PD-SOI CMOS circuits”, IEEE JSSC, Vol 32., Aug 1997, pp. 1241 – 1253.

[7.21] S. Nutoh, et. al., “A 1V high speed digital circuit technology with 0.5um multithreshold CMOS”, Proc. IEEE Int. ASIC Conf., 1993, pp. 186 – 189.

[7.22] T. Douseki, et. al., “A 0.5V SIMOX-MTCMOS circuit with 200pS logic gate”, ISSCC Digest technical papers, 1996, pp. 84 - 85.

[7.23] T. Fuse, et. al., “0.5V SOI CMOS pass gate logic”, ISSCC Digest technical papers, 1996, pp. 88 - 89.

[7.24] T. Fuse, et. al., “A 0.5V 200Mhz, 1-stage 32b ALU using a body bias controlled SOI pass gate logic”, ISSCC Digest Technical papers, 1997, pp. 286 - 287.

[7.25] L. S. Y. Wong & G. A. Rigby, "A 1V CMOS digital circuit with double gate driven MOSFET", ISSCC Digest technical papers, 1997, pp. 292 - 293.

[7.26] Y. Wada, et. al., "Active body bias SOI-CMOS driver circuits", Symp. VLSI circuits, 1997, pp. 29 – 30.

[7.27] I. Y. Chung, et. al., "A new SOI inverter for low power applications", IEEE International SOI Conference, 1996, pp. 20 - 21.

[7.28] J. H. Lee & Y. J. Park, "High speed SOI buffer circuit with the efficient connection of subsidiary MOSFET's for dynamic threshold control", Proc. IEEE Int. SOI conference, 1997, pp. 152 - 153.

[7.29] T. W. Houston, "A novel dynamic Vt circuit configuration", Proc. IEEE SOI Conference, 1997, pp. 154 - 155.

[7.30] K. Mistry, et. al., "A 2V 0.35um PD-SOI SOI-CMOS technology" IEDM Technical digest, 1997, pp. 583 - 586.

[7.31] D. J. Schepis, et. al., "A 0.25um CMOS SOI technology and its application to 4Mb SRAM", IEDM technical digest, 1997, pp. 587 - 590.

[7.32] D. Suh & J. G. Fossum, "Dynamic floating body instabilities in PD-SOI CMOS circuits", IEDM Digest, 1994, pp. 661 – 664.

[7.33] C. Y. Chang & S. M. Sze, "ULSI Devices", John Wiley and Sons, 2000, ISBN 0-471-24067-2

This page intentionally left blank

Chapter 8: SOI SRAMs

8.1 Introduction

Low power large-scale integration of memory technology is an increasingly important and growing area of electronics. Although rapid progress has been made in the reduction of power for RAM subsystems there remains significant potential for future improvement. Low power RAM technology is achieved by a combination of low power chip technology, multi data bit chip configurations in which a large number of data bits are processed simultaneously, small package technology, and low voltage chip to chip interfaces. SOI technology is a relatively new introduction in memory design and its substitution can improve performance through reduced junction capacitance and immunity to soft error rates. Low power chip technology has so far mainly contributed to subsystem power reduction. Multi-data bit chip configurations have taken a stronger position with increased memory density due to the reduction of subsystem power and resulting in reduced chip count for a fixed subsystem memory capacity. Fewer chips and smaller packaging technologies combined with high-density chip technology have reduced AC power, with less capacitance for address lines, control and data I/O bus lines on memory boards. Reducing the voltage swing on the data bus lines has contributed to reduced AC power dissipation.

Ultra low voltage operation is crucial for modern SRAM designs especially for mobile and embedded applications. However, sub-threshold current and threshold voltage mismatches developed in the memory cell are major concerns for low voltage operation. Short gate lengths in SOI technology can

lead to short channel effects, and mismatch between devices in the memory cell can cause instability and disrupt functionality.

In addition to improved radiation hardness the feature sizes of CMOS SOI transistors are readily scalable to well below 0.1 μm, and appear, in principle, to be scalable below the limits of bulk CMOS bulk transistors. Scaling of CMOS SOI transistors increases the future potentials of CMOS SOI technology applications not only in radiation hardening, but also in general IC processing and manufacturing.

The high current drive capabilities of partially depleted devices and the low sub-threshold leakage currents and high saturation resistance of fully depleted devices can be combined by switching between FD and PD modes to obtain optimal performance and power, if a mixed FD/PD process is used.

8.2 SRAM Cell structures

Latched semiconductor storage cells are usually in the form of bi-stable transistor flip-flops. Semiconductor flip-flops can be made of either MOS or bipolar transistors. The CMOS SRAM cell consists of two load devices and two storage transistors together with two access transistors as shown in figure 8.1 [8.1]. Load devices may be enhancement or depletion mode transistors (in a nMOS cell), pMOS transistors (in a CMOS cell) or load resistors (in Mix-MOS or R-Load cell). The access and storage transistors are enhancement type nMOS.

The load device offsets the charge leakage at the drains of the storage and select transistors. When the load transistor is pMOS, the resulting six transistors CMOS cell (Figure 8.1(a)) has only leakage current through the cell except during switching. Either the nMOS or pMOS is always off. The standby power dissipation is higher for a resistive load cell than a full CMOS six transistor cell. One of the disadvantages using a pMOS load in a SRAM cell is, difficulty to write into the cell.

A depletion type load transistor is normally used in the nMOS flip-flop instead of a second enhancement transistor due to its better switching performance, higher impedance, and relative insensitivity to power supply variations. It is similar to a CMOS p-channel load as the internal voltage swings are essentially equal to the supply voltage. If a double polysilicon process is used, the load resistor cell (Figure 8.1(b)) offers approximately 30% reduction in cell size over that of the pMOS load cell with double polysilicon technology for the resistor. This is because the resistors can be stacked above the cell, reducing cell size by two transistors. This cell has

lower performance, but the reduced die area of this cell provides a major cost benefit.

Design technique can reduce the drawbacks of the R-load cell. Standby power consumption is reduced if all peripheral circuits outside the memory array can be powered down without disturbing the stored data. Sub-threshold leakages of 'OFF' transistors in the memory need to be at adequately low levels to maintain data during standby.

Figure 8.1 shows various common CMOS static RAM latched cells. Figure 8.1(a) shows the 6T SRAM cell and Figure 8.1(b) the 4T-2R static RAM cell with resistive load pull up. Figure 8.1(c) and (d) are eight-transistor double ended dual port static cell, and a nine transistor static content addressable memory (CAM) cell respectively. All cells shown in Figure 8.1 are static SRAM cells, and do not require refresh to maintain data. Figure 8.1(e) shows a single ended five-transistor static CMOS cell, which can replace the basic six-transistor full CMOS cell where a smaller array size is needed. The five-transistor cell contains one less component than the 6T cell and only bit line per cell. It is not widely used because of lower operating margins and difficulty in write operations. Writing a '0' is straightforward but writing a '1' is difficult since the nMOS pass-gate transistor operates in source follower mode limiting the voltage transmitted to the cell.

All MOS static memory cells have in common a basic cell consisting of two transistors and two load elements in a flip-flop configuration as shown in Figure 8.1(a). T1 and T2 are the driver transistors, T5, T6 are the pass gate nMOS devices and finally T3, T4 are the load pMOS devices. Information is stored in the form of voltage levels in the flop-flop, which is formed by the two cross-coupled inverters. The flip-flop has two stable states typically designated '1' and '0'. In figure 8.1(a), assume the memory cell storing a logic state '1', the node C_5 is 'high' and C_6 'low', the transistor T1 is 'off' and T2 is 'on' and T3 is 'on' and T4 is 'off'. The logic '0' state would be opposite with node C5 low and node C6 high. Both states would be stable with neither branch of the flip-flop conducting as long as the DC voltage is applied.

The memory cell is embedded in an array of similar memory cells which are accessed by a row decoder which selects the word line and a column decoder circuit which selects the appropriate *BIT* and $\overline{BIT}$ lines. The single ended dual ported cell (figure 8.1(f)) can replace the more common double ended dual port cell (figure 8.1(c)) with corresponding area improvement due to the elimination of two transistors.

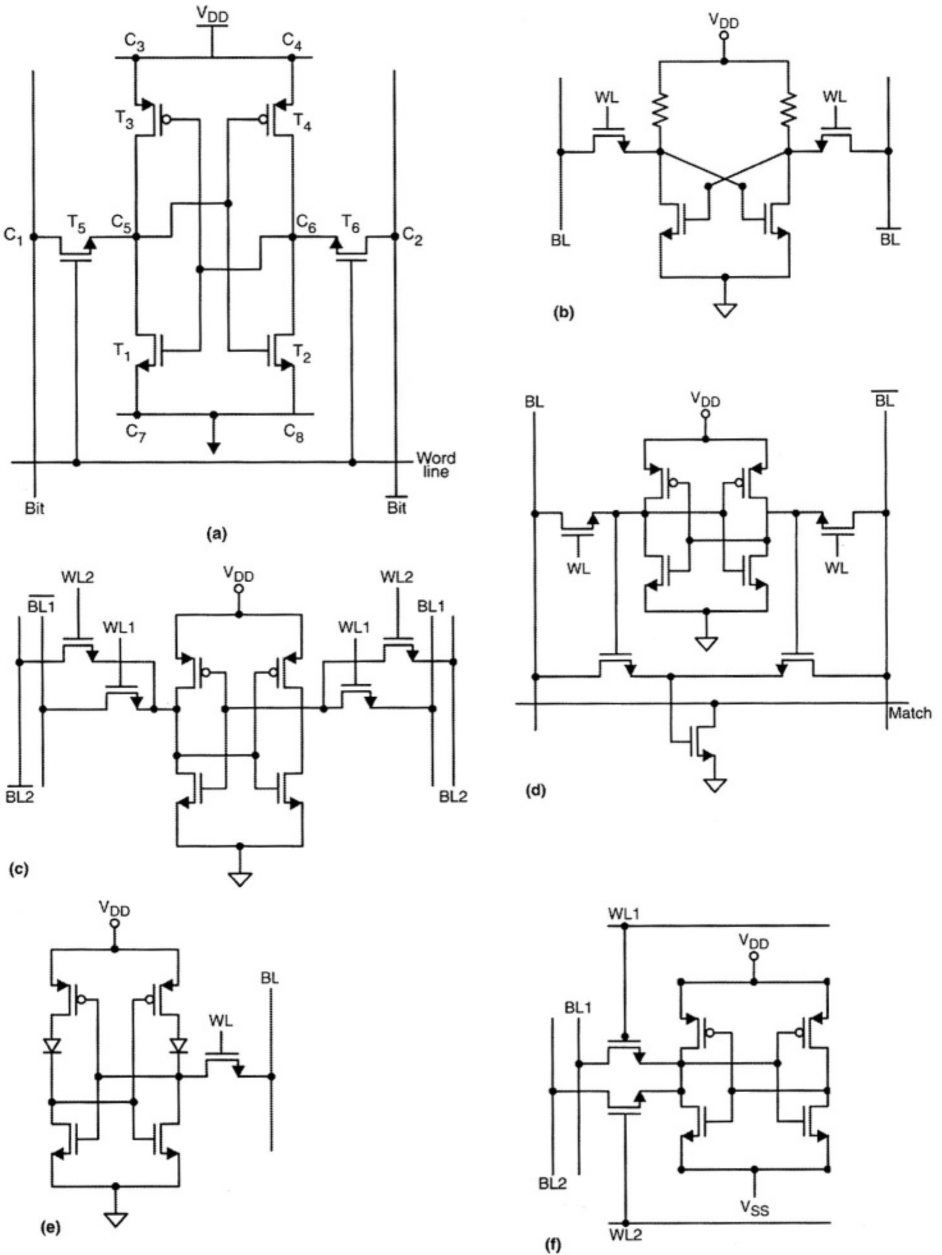


Figure 8.1: Static RAM Cells (see text for details)

The SRAM cells shown in Figure 8.1 are as follows:

(a) 6T SRAM cell with pMOS load.
(b) 4T with Resistive Load
(c) Dual port double ended access
(d) Content Addressable Memory
(e) Five transistor full CMOS having a single access transistor
(f) Dual port with single ended access

The dual port cell is useful as an embedded cache memory in microprocessors since it can be accessed simultaneously through both ports. Simultaneous access eliminates wait states for the microprocessor unless one port wants to read while the other is writing, or both ports intend to write simultaneously.

The content addressable, or associative memory cell is used in applications where both the contents and the location of the cell are required. During normal operation, read and write is performed like a normal cell. During comparison operation, *DATA* is placed on $\overline{BIT}$ and $\overline{DATA}$ is placed on the *BIT* lines. This is opposite of the normal read configuration. If the data match those in the cell, then the match transistor will stay OFF. If any cell has data that does not match, the match transistor pulls a previously pre-charged 'match' low.

While the content addressable cell shown has nine transistors, other applications require associative cells with even more transistors. Clearly the size of such a cell and the resulting memory size is larger than for standard RAM cells. The application must require the particular associative features of this cell to justify the additional cost. If body contacts are added in SOI to minimize variability in delays, the area penalty could be unacceptable for most designs.

<u>8.3 Design considerations and specifications for SRAM Cells</u>

The following are some of the designs consideration for SRAMs:

* Adequate cell stability for reliable operation
* Maximizing "drive" current to achieve high speed
* Minimum cell size for highest density
* Good manufacturability
* Minimum operating current for low power
* Minimum versus larger than minimum dimensions
  -- Use minimium dimensions where needed and non-minimum in other

circuits for robustness.
* Minimization of alignment induced mismatches
* "Non-uniform Geometries"

These are often conflicting requirements and compromises must be made specific to the application.

8.3.1 4T-2R Polysilicon resistor load SRAM

The 4T-2R polysilicon resistor load SRAM has the following specific attributes:

* High process complexity (will require 2 layers of Poly, buried contacts)
* Typical resistive load range is 10MΩ – 10GΩ
* Requires high cell ratio for stability (Cell Ratio = $W_{pulldown}/W_{passgate}$)

This is often the SRAM design of choice up to about 4Mb-16Mb. For SRAM densities of 16Mb and above, the standby current tradeoff versus data retention is a problem and the 6T cell is typically used.

8.3.2 SRAM cell with 2 thin-film transistor loads

SRAM cells with 2 thin-film transistor loads have the following properties:

* High process complexity: 3-4 layers of poly, buried contacts
* Roff/Ron = 10-100, Roff ~ 1-10GΩ
* Requires high cell ratio for stability.
* High process complexity, not widely commercialized.
* Modulated resistance allows improved trade-off between standby current (R_{OFF}) and cell leakage current (R_{ON}).
* Pull up still too weak to aid cell stability or soft error immunity.

8.3.3 6T- PMOS lOAD SRAM cells

6T- PMOS LOAD SRAM cells have the following properties:

* Slightly higher process complexity than 4T resistive load cells: 1 layer of Poly, 1 or more metal
* 1.5X – 2X larger cell size than 4T
* Allows lower cell ratio for stability
* Used mostly in embedded and ultra low power memories
* Active pull up aids cell stability and soft error immunity (higher capacitance and recovery current).
* Writing to the cell is difficult compared to 4T-2R [8.2].

8.4 4T SRAM cell using Self-body biased (SBB) SOI MOSFET structure

The body region of a PD-SOI MOSFET [8.3] is shown in figure 8.2, where an "H" shaped gate electrode is used as a resistor in the inverter pair of the SRAM cell [8.4]. Resistance of the body region is controlled by both physical structure and high resistivity region beneath the gate. This technique is applicable to low voltage applications.

The self body biased SOI MOSFET utilizes the gate depletion layer beneath the auxiliary T-Shaped gate electrode in order to modulate the body bias of the SOI MOSFET. There is a potential barrier between the body of the SBB SOI MOSFET and the body diffusion terminal [8.5]. Body voltages "Body1" and "Body2" are controlled by voltage sources. The H-Shaped structure is characteristic of a modified double gate structure. The body potential is controlled via "Body1" and "Body2" depending on the gate voltages [8.6].

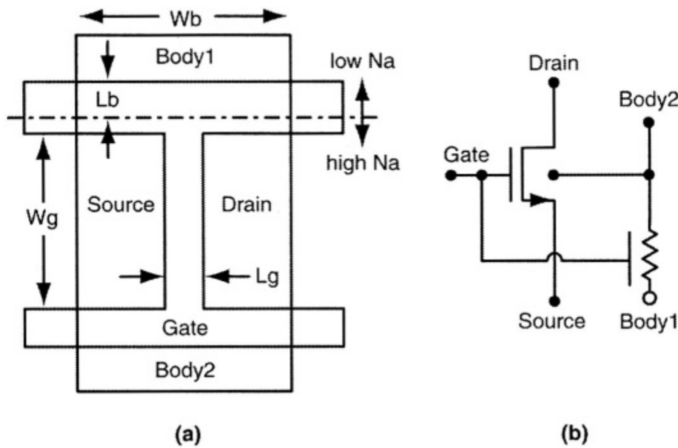


Figure 8.2: (a) *Modified Double Gate Structure showing the high and low concentration region. (b) Schematic symbol.*

Body2 diffusion is same as a conventional body tied SOI MOSFETs. The gate depletion layer extending into the low concentration region modulates conductivity between "Body1 and Body2". Thus changing the impurity concentration in the low concentration region and the auxiliary gate geometry (L_b and W_b), the desired resistance can be achieved.

Plots of V_{GS} versus I_{DS} and V_D versus I_D characteristics of the H-shaped SOI MOSFET shown in figures 8.3 and 8.4 indicate no significant difference between the conditions where $V_{Body1} = 0.5V$, $V_{Body2} = 0.5V$ and where $V_{Body1} = 0V$, $V_{Body2} = 0.5V$. This is because there is little influence of the current

flowing from "Body2" to "Body1" on the characteristics of the MOSFET in the body region. Thus the resistance between "Body1" and "Body2" can be treated as an independent resistor. In the T-gate MOSFET, the body potential is determined by V_{Body2} because the gate depletion layer extending in the low concentration region isolates the influence of V_{Body1} [8.4, 8.6].

Combining the I_{DS} versus V_G characteristics (figure 8.3) and I_{Body1}-I_{Body2} versus V_G (figure 8.5) characteristics into R_{DS}-V_G and $R_{Body1\text{-}Body2}$ versus V_G characteristics. The 4T SRAM cell with self body biased SOI MOSFET is shown in figure 8.6. This cell has two stable cross-points and a substantial operating margin at low voltages.

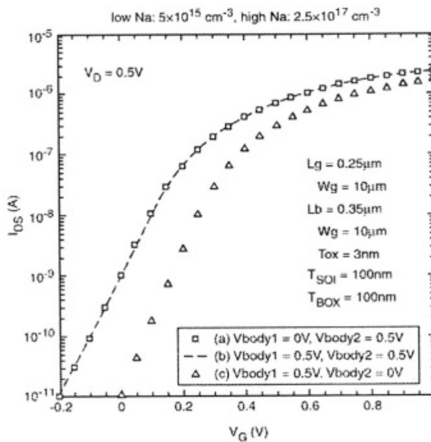


Figure 8.3: I_{DS} vs. V_G Characteristics

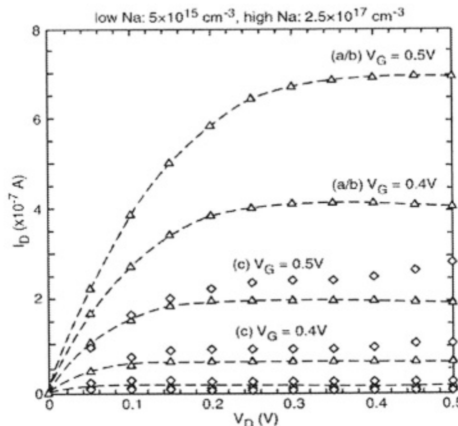


Figure 8.4: I_{DS} -V_{DS} relationship of devices with various (V_{BODY1}, V_{BODY2}) configurations [8.6].

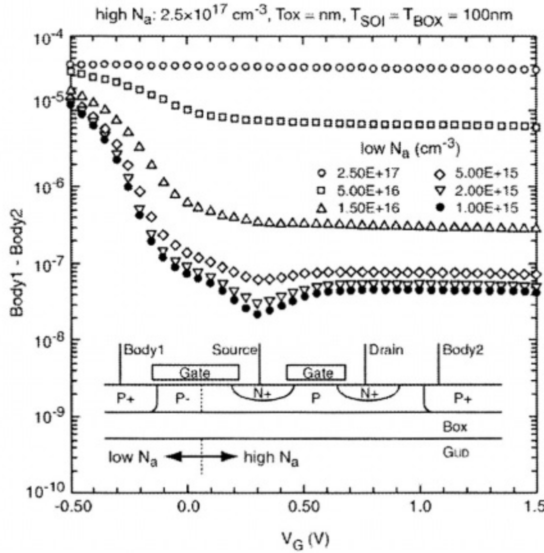


Figure 8.5: Body voltage versus gate voltage. The body voltages is the difference between I_{BODY1}-I_{BODY2} (Yaxis is $I_{BODY1} - I_{BODY2}$)

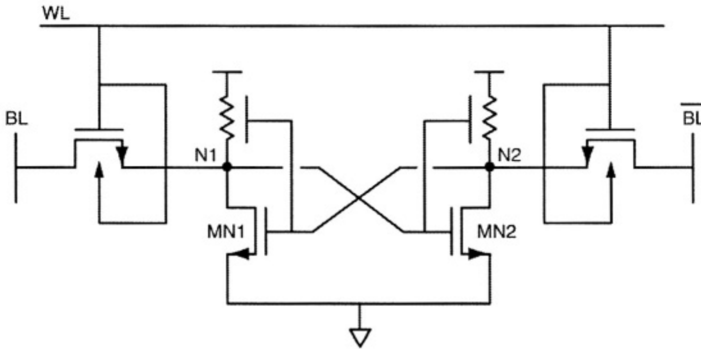


Figure 8.6: 4 transistor SRAM cell using H-gate self body biased SOI MOSFET.

8.5 Basic SOI SRAM Cell operation

Optimal SRAM cell designs should take into account the following factors: (1) High array efficiency (2) Fast array access – short bit lines and word lines (3) Simple decoding structure (4) Faster data path access (5) Low standby power consumption.

8.5.1 READ Operation in a SRAM Cell:

READ is typically performed by precharging both *bit* and $\overline{bit}$ lines high and selecting the desired word line. All cells in a row are selected. The selected

column pulls one of the bit lines low. This differential signal on the bitlines is detected by the sense amplifier, amplified and read out onto the output buffer.

Figure 8.1(a) demonstrates a read operation from the static SRAM cell. To initiate the READ, the bit line (*BL*) and the bit line complement ($\overline{BL}$) are released from precharge before the word lines are activated. Once the word line is enabled turning 'on' the SRAM cell passgates, the data from the cell is transferred onto the bit lines (*BL* and $\overline{BL}$) and fed to the sense amplifiers. Reading a '0' stored in the cell would be mirrored function with the current flowing through T5 and T1 to ground and with the bit line *BL* at a lower potential than $\overline{BL}$.

After the bit lines are released from precharge and data is read onto the bit lines, the minimum voltage differential required on the bit lines to ensure proper sensing and amplification by the sense amplifier is of the order of 30-200mV, depending on the technology and process parameters.

During READ operation, the off state pass-gate sub-threshold leakages hurt the development of the differential on the bit lines, because the bit lines are not discharged to low enough voltage in order to sink all the charge in the pass gate bodies. Leakage currents during the READ cycle are lower than WRITE cycle, since potential across the pass gates is only a fraction of the supply voltage. This leakage current changes with bias across the pass gate.

Reading the cell of a static RAM cell is non-destructive and after the READ operation the logic state of the cell remains unaltered. Read current is determined by the width of the pass gate. Typically to achieve low leakage, the pass gate is made minimum width and wider for higher driver currents. The pass gate has to be minimum for cell size and power and should be greater than minimum size for speed. The cell design must be a compromise of preferred cell size, power consumption and speed.

The gain (β) of a device is $\mu C_{ox}W/L$ where μ is the mobility parameter, C_{ox} is the oxide thickness and W/L are the ratio of width to length. The ratio between gain of the driver transistor and gain of the access transistor is called the SRAM cell ratio. The voltages on the storage nodes depend very critically on the cell ratio [8.7].

Figure 8.7(a) shows the basic elements of a SOI SRAM READ critical path and figures 8.7(b, c) show the critical nodes and their body voltages during a read '1' operation. Read operation begins by turning on the word lines and

the transferring the data from the memory cell onto the bit lines. Write ($\overline{WE}$) enable turns on immiediately following this operation and data is transferred from the bit lines to the sense amplifier. $\overline{WE}$ "high" signifies a read operation, and a "low" represents a write operation. The sense amplifier amplfiers the signal from the bitlines and transmits to the output buffers. The sense amplifier outputs are shown as *SAOUT* and $\overline{SAOUT}$. *SAOUT* will be high if reading a "1" and low if reading a "0". Sense amp enable 'SAE' in figure 8.7(a) is equivalent to signal 'SA' in figure 8.7(b).

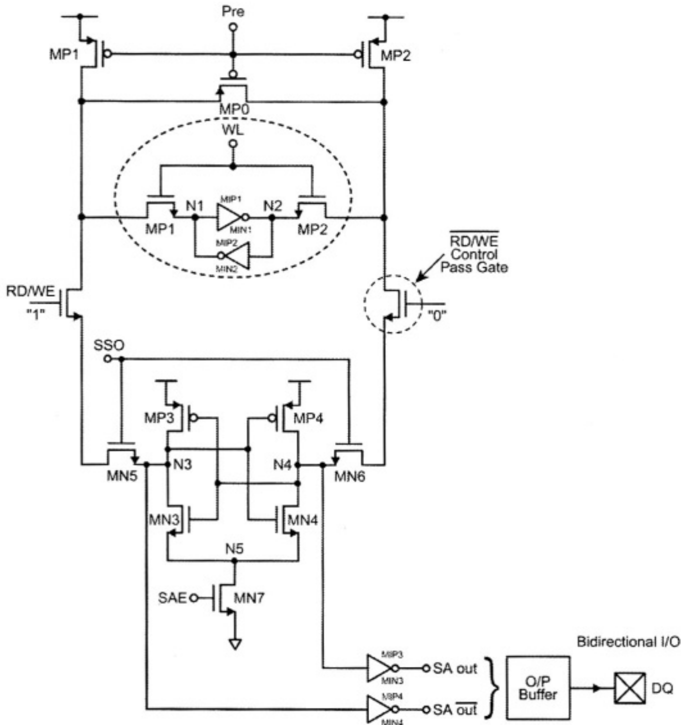


Figure 8.7*(a): Critical Path read operation in SRAMs.*

The waveforms show the sense amplifier enabled after adequate potential is developed on the bit-lines. The sense amplifier may be enabled once the potential difference between the two bit-lines is ~100mV. Once the sense amplifier is enabled the body voltage of the enable transistor MN7 drops from 0.177V to 0V due to body discharge as shown in figures 8.7(b, c). The body voltage of MN5 rises due to gate-body coupling once the word line is turned 'on'. When reading a "1", the body voltages of MN1 and MN2 also rise due to body to gate coupling. Subsequently, the body voltage of MN2 falls since the potential on the complementary bitlines discharges to zero.

The body voltages of the sense amplifier before and after sensing are shown in figure 8.7(c). Absence of body contacts for sense amplifiers introduces wide variability in achieving a steady state. Body contacts to supply voltage for sense amplifiers and SRAM cells ensure improved stability for the SRAM cell.

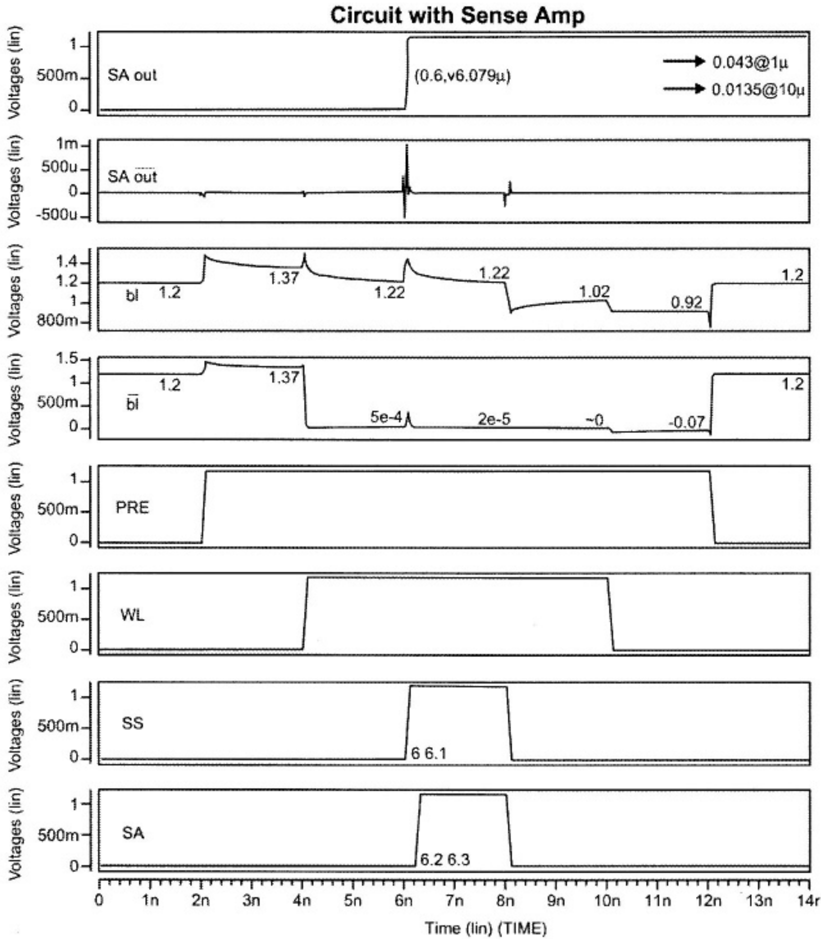


Figure 8.7(b): Waveforms during reading a '1' with floating bodies. BL reads a high and $\overline{BL}$ reads a low after the sense amplifier is enabled.

Adding body contacts impacts area and gate capacitance but reduces leakage. Standby current is measured during precharge for floating body and body tied to supply cases. Figure 8.7(d) compares standby currents. The standby current with body ties to supply is 140nA compared to 188nA with

floating bodies for 8K SRAM array. Standby current improves by 30% due to reduction in parasitic bipolar and transistor subthreshold currents.

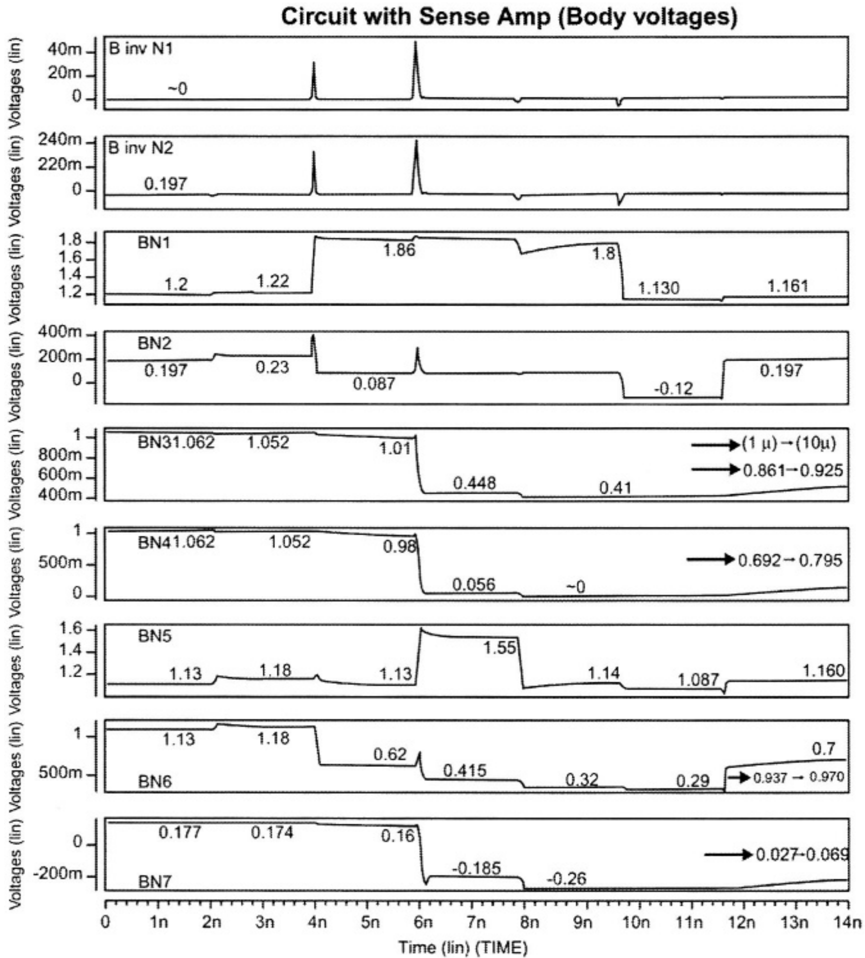


Figure 8.7(c): Body voltages during read '1'.

8.5.2 Write operation in SRAM Cell

Figure 8.8(a) shows the schematic critical path for a write operation. Bit lines are precharged high during standby. A 'write' into the SRAM cell starts by activating write enable (WE) signal which pulls one of the bit lines low based on the write driver (figure 8.8) data. Data "0" and "1" are placed on the *BL* and $\overline{BL}$ lines and the word line is activated. The cell flips to the appropriate state depending on the bit line data. To write a "0" *BL* will be low and $\overline{BL}$ high (figures 8.8(b,c)). Conversely, to write a '1', *BL* is raised

high and $\overline{BL}$ contains a low [8.7]. The minimum delay to write data into a cell depends on the condition of other memory cells on the bit-line. The slowest write is when all the other cells are logically low, and has been in that state for a long enough time. The bit line driver must sink current from several different circuit paths. SRAM designs usually require the bit-line to be fully discharged, thus the write inverter must sink the entire range of possible currents caused by the variability in SOI device currents. Figure 8.8(a) shows the simulated internal storage node voltages in a SRAM cell during 'write'.

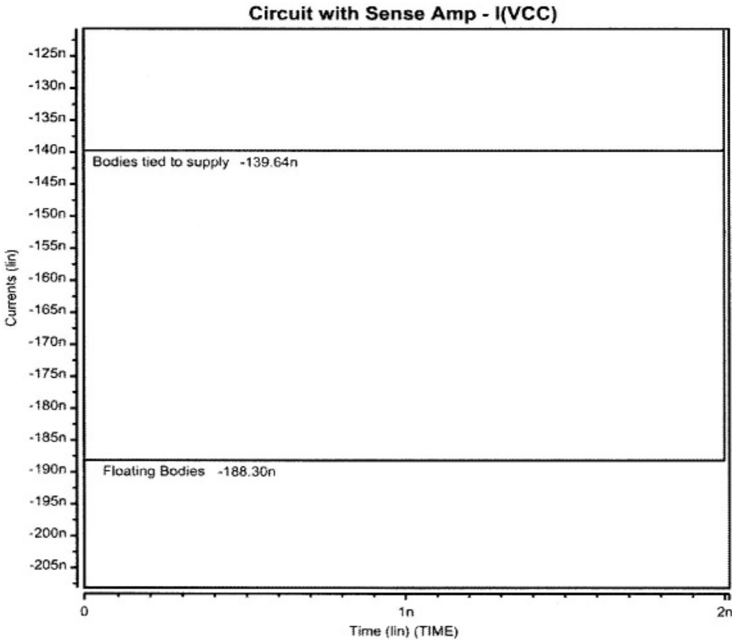


Figure 8.7(d): Standby Current comparison during Read between floating body and body tied to supply.

Capacitive discharge current

Bit-line capacitance and pass gate loads are major sources of current discharge in SRAM cells. Current due to the bit line capacitance and leakage through unaccessed cells can corrupt the data in the cell. The total capacitance on the bit lines and rate of capacitive discharge can vary since the bodies of the pass-gate transistors have different steady state voltages and are based on the data in the cells and the cell history. Pass-gate source/drain capacitance is lowered in SOI compared to bulk due to the absence of the area component of the pass gate diffusion capacitance. Hence, the relative value of capacitance between source and body is high. Both sides of the bodies converge to the same potential over a period of time. The

resulting depletion width will be small effectively resulting in higher capacitance on the bit lines and increased bit-line discharge times. High current capacity bit-line drivers can reduce the delay between reading a "0" and "1" and also reduce the variability in bit line write times [8.7, 8.8].

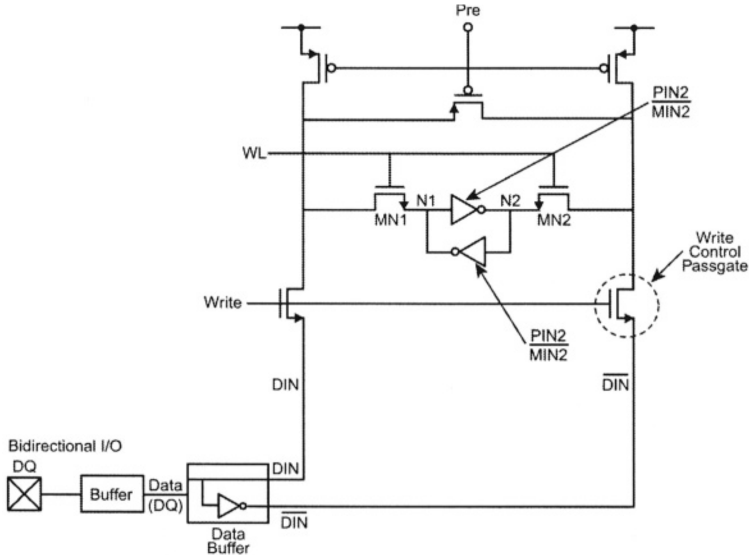


Figure 8.8(a): Schematic path for write operation in SRAMs.

The trip voltage (switching point) of the memory cell is set by the cell ratio. A high cell ratio leads to a low trip voltage and it becomes hard to write into the cell. A low cell ratio will cause a higher trip voltage and possible read-disturb. Writing into the 6T SRAM cell is more difficult than writing to the 4T-2R cell as discussed previously because of reduced load currents.

Diode and Bipolar Currents

Discharge currents can affect operation of the cell, once the bit line drops a diode voltage from V_{DD}. Since the pass gates bodies are charged to V_{DD}, and the bit lines are at V_{DD}, the body to bit-line forms a forward biased diode when the bit line is pulled low. Parasitic leakage current from the un-accessed cells flows from the bodies to the bit lines when the bit lines fall. This produces a current pulse, which cannot be reduced until the bodies are completely discharged.

Ideally, only leakage currents flow in standby mode however, there is some subthreshold current flow through the pass transistor. Sub-threshold Current is given by $I_S = I_T \exp(-qV_T/mkT)$, where I_T is the drain current at the threshold voltage, V_T is the threshold voltage, and m is the sub-threshold

slope multiplier and is related to sub-threshold slope parameter, S (mV/decade) by 0.001* qS/kTln10. The transistor sub-threshold current can be significant if the threshold voltage of the pass transistor is too low. The sub-threshold current in SOI devices in higher than bulk devices, although this is not true always. The sub-threshold leakage could be made lower by adjusting the threshold voltages [8.7].

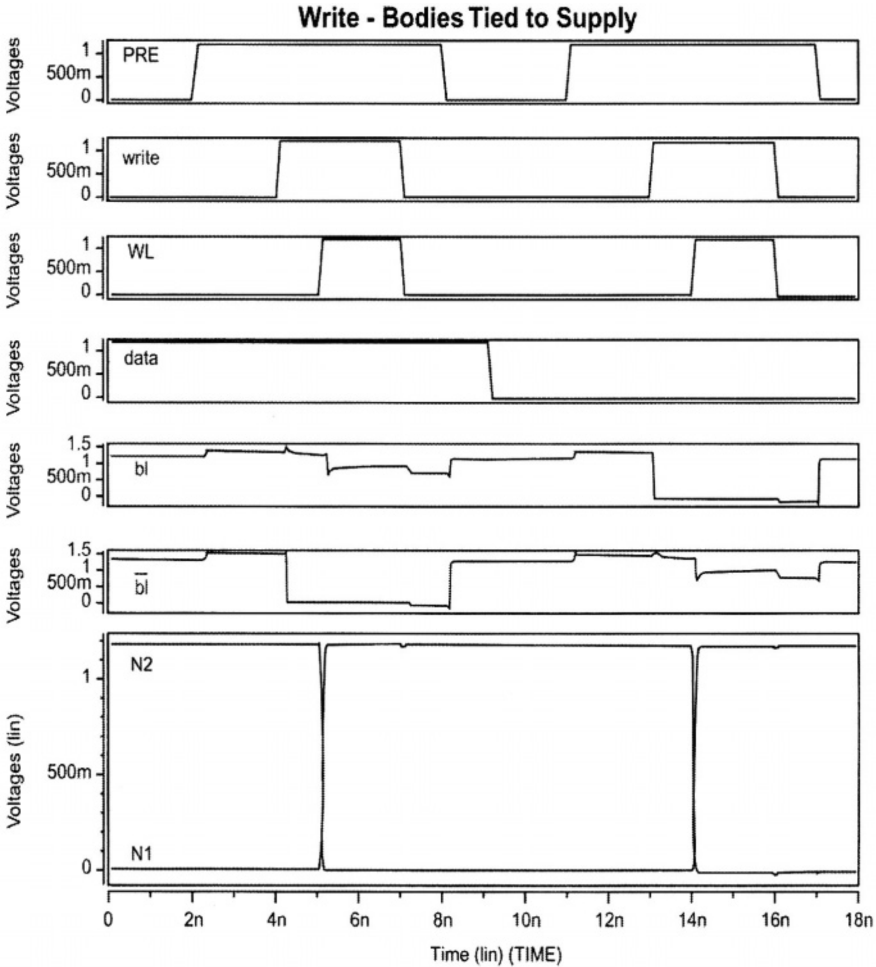


Figure 8.8(b): Simulation waveforms during WRITE operation in a SRAM with the bodies tied to supply voltages. The internal nodes N1 and N2 flip in opposite direction. The first cycle writes a '1' and the second cycle writes a '0'. The data in signal carries the input signal to be written in the cell, and the storage nodes and bit lines flip respectively. Here, the BL is high during first cycle and low during the second cycle.

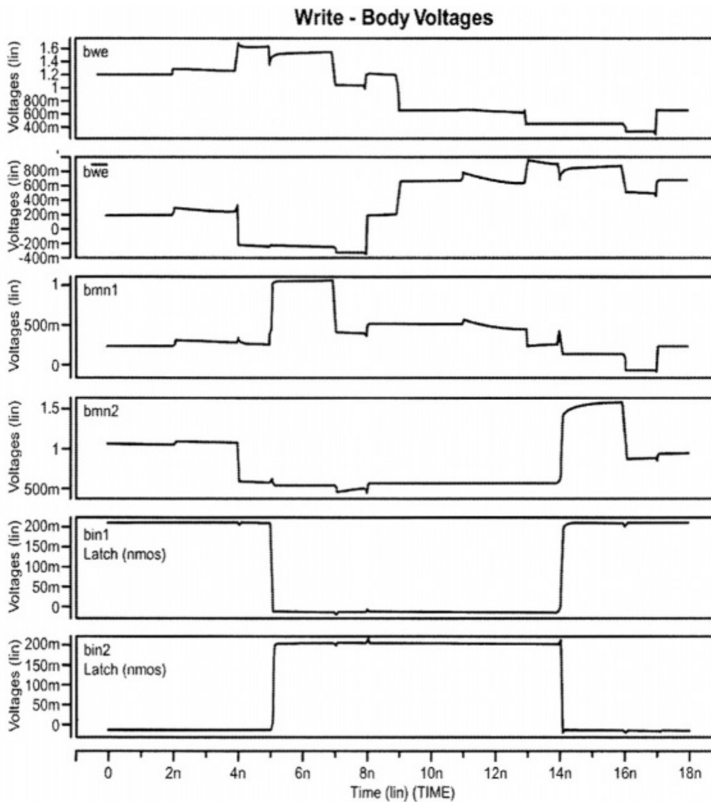


Figure 8.8(c) Pass gate and driver body voltages during a SRAM write operation. The body voltages of the two driver transistors (MN1, MN2) when node N1 goes high and N2 low, are low and high respectively.

CMOS Subthreshold Leakage

CMOS sub-threshold leakage from the un-accessed cells can leak through their respective pass-gates and from the pMOS on the high side to their bitlines (figure 8.8). The threshold voltages of the pass gates are lower because their bodies are charged to V_{DD} producing leakage currents an order of magnitude higher than bulk. This type of sub-threshold leakage current is produced as soon as the bit lines switch low, degrading the bit line differential needed to sense.

8.6 Cell Stability

Cell stability is a measure of the ability of the cell to retain data. It is a function of voltage, temperature, process variations, noise and other disturbing events. Soft error rates (SER) and sensitivity of memory to process tolerances and operating conditions are important for cell stability.

SER and the ability to retain charge are interdependent since designing a cell for improved stability requires larger cell area, and larger area cells are more susceptible to SER effects.

The ability of a SRAM cell to retain data most recently written while being read is called read stability. Inability to read the correct cell data is known as read disturb.

The worst case condition affecting the stability of SRAM cells occurs during read when the bit lines are pre-charged high, since the load devices are shunted by the access transistors, reducing the gain of the cell inverters. The static noise margin of the SRAM cell increases with increasing threshold voltage since the threshold voltages decreases with temperature.

One important parameter for the stability of the SRAM cell is the ratio between β (gain) of the driver transistor and access transistor (β is $\mu C_{ox} W/L$) and is usually called the cell ratio or beta ratio of the SRAM cell. During read, the storage node voltages between the read/driver transistor and the pass-gate transistor is determined by the gain (β) ratio of the two transistors. If β of the driver transistor is too small, it can bring the cross coupled low side node too high upseting the cell. Higher β ratio results in better stability but if the pass gate becomes too strong with respect to the pull down driver transistor, the cell being read may have a high bit-line charge and cause the cell to flip to a false state. If the pass gate is too weak compared to the pull-down driver transistor, the cell may not develop sufficient differential within the required time. Higher β values result in bigger cell size causing a tradeoff between the cell area and stability. The load devices are very important for the write operation. If the pMOS load devices are strong, writing into a 6T SRAM cell is difficult [8.9].

The cell ratio ($\beta_{PULLDOWN}/\beta_{PASSGATE}$) is also affected by statistical variations and asymmetries in the cell device characteristics. Supply noise and bit-line noise suppressiong are critical for successful SRAM read and write operation.

Both resistive and pMOS device loads are represented by a flip-flop comprised of two inverters (figure 8.9(a)). The voltage sources V_N are static noise sources. Static noise is caused by dc offsets, process mismatches and variations in operating conditions. The SNM of the flip-flop is defined as the maximum value of V_N that can be tolerated by the flip-flop before changing states. SRAM cell design should be toleratant to dynamic disturbances

caused by alpha particles, crosstalk, supply voltage ripples, and thermal noise.

Static noise margin of SRAM cell can be analyzed graphically from the inverter characteristics by finding the maximum possible square between the two waveforms [8.10-8.12] (Figure 8.9). A typical static noise margin calculation uses the basic MOS model with constant threshold voltage and a simple exponential sub-threshold current model. Cell ratio is often assumed to be 2 for 6T-Cell and 3.5 for R-load cell. The R-load inverter characteristics start at V_{DD} when V_{IN} =0v, and drop sharply as soon as the sub-threshold current is able to develop a noticeable voltage drop across the resistor. Subsequently, the output voltage is clamped by the access transistor and drops further with reduced slope for $V_{IN} > V_T$ when the driver transistor is turned on. When V_{DD} is decreased, the R-load cell will lose data before the 6T cell does. In addition, the R-load cell can only be used with V_{DD} larger than $2V_T$ during access to avoid write time problems.

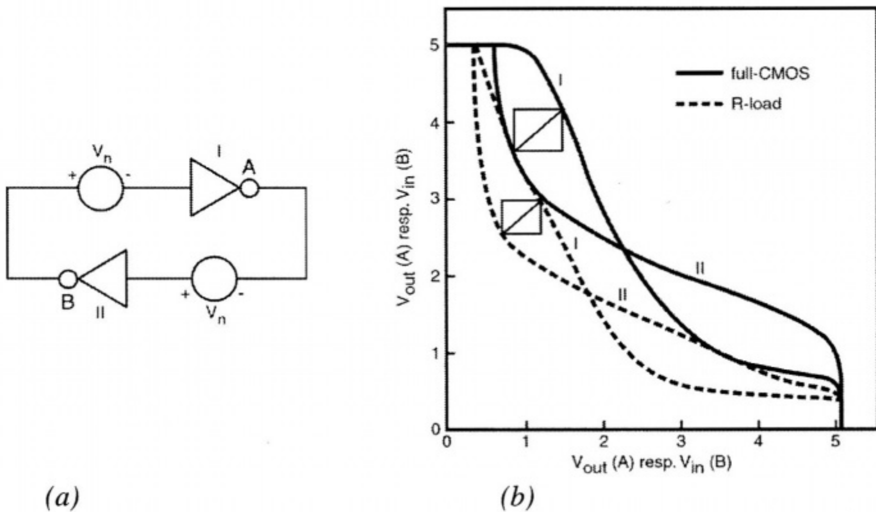


Figure 8.9. Comparison Of Static Noise Margin for R-load and Full CMOS SRAM cell: (a) Static Noise margin simulation circuit (b) SNM estimation based on maximium sqaures in a 45° rotated coordinate system.

To estimate the SNM values graphically, a procedure is needed to find values for the diagonals of the maximum squares in figure 8.9(b). Figure 8.9(c) shows the stylized version of figure 8.9(b), which are rotated 45 degrees relative to each other. In the *(u,v)* system, subtraction of the "*v*"

values of normal and mirrored inverter characteristics at given "*u*" yields curve "A", which is a measure of diagonal's length. The maximum and minimum of curve "A" represents the required maximum squares. Curves have been mirrored with respect to a line passing through the origin at 45 degrees from the horizontal.

Figure 8.9 shows that the full CMOS cell has a much larger SNM than the R-load cell despite its larger value of cell ratio. Another important difference between the two cells is the dependence of SNM on the power supply voltage. SNM for R-Load cells rapidly degrades at lower power supply voltages, whereas for full CMOS cell it stays fairly high even at low power supply voltage. As voltage scales below 1.8V for SRAMs, the full CMOS static cell is the preferred cell. Techniques such as tank biasing can be used to adjust threshold voltages for improved static noise margin. This affects drive current, and hence static noise margin versus drive current trade-off. The static noise margin curve for a SOI SRAM cell is shown in figure 8.10 [8.9].

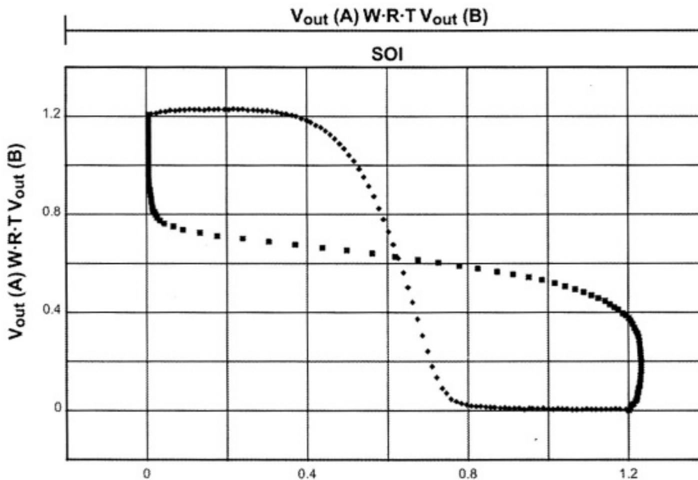


Figure 8.10: SOI SRAM cell static noise margin at 1.5V. The x-axis is input voltage and y-axis is output voltage with respect to each other.

Static noise margin is important factor in determining the stability of SRAM cell. Variation in process, length and width of the transistors are important and should be included in the simulations to obtain optimum SNM value. In addition statistical variation, such as local and global variations in width/length must be taken into account. In addition, dopant fluctuations and V_T variations are critical process parameters to obtain an accurate SNM value.

8.7 SRAM Junction & Bit line capacitance

SOI technology applied to SRAM offers a significant performance advantage due to junction capacitance reduction. The benefit is particularly striking in the differential bit line topology, which contains hundreds of source/drain junctions. The device junction capacitance constitutes a substantial portion of the total bit line nodal capacitance. Collective capacitance reduction of the pass-gate transistors in conjunction with higher current drive of the selected pass gate transistor results in substantial reduction of the cell access time.

The floating body potential of the pass gate transistors depends on cell content ("0" or "1"), dynamic coupling of all internal capacitive elements during switching and parasitic leakages. Hence, the internal source to body capacitances of the pass gate transistors become dependent on the array content causing an imbalance in the node capacitance between the true and complementary bit lines. Bit line capacitance disparity becomes more significant as supply voltage scales. Furthermore, during the WRITE operation, a disturbance of unselected cells may occur due to excess parasitic bipolar current when the bit line is pulled down [8.13 - 8.15]. This is especially likely to occur during the first cycle.

8.8 Decoders

Decoders are used to 'decode' input addresses and select a particular location in the memory array to read or write. Some of the types of decoders commonly used are shown here. Figure 8.11(a) shows the ground decoding structure. This is better than a 3-input NAND, but not quite as good as a 2-input NAND gate. Capacitance looking into the nMOS source is lower than gate if layout is shared with adjacent decoders. The series nMOS can contribute to charge sharing and the floating body will speed up switching. Weak pull-up transistors are required to assist in driving the nMOS high. Figure 8.11(b) shows the clocked decoding structure. This structure reduces loading on the address lines. The RC delay and skew requirements are critical since the clock lines have high capacitance. The pMOS transistors size (width/length) should be minimized for loading and width should be maximized for speed.

This type of decoder works well when the address lines are pulsed, and are OFF during pre-charge since it minimizes address line capacitance (including word line capacitance). Use of self-resetting decoders (figure 8.11(c)) is the preferred method for SOI SRAMs. Self-resetting circuits minimize address lines loading. Although, clocked decoding also offers the same reduction in address loading, no clock distribution is needed since

RESET is internal to the word line driver circuit. The weak pMOS should be strong enough to reduce offset leakage and noise. Loop delay must be longer than input pulse width. The pulse width extends down the chain and must reset with one shot when reset margin becomes too small. Pulse width management and noise bookkeeping is critical [8.2, 8.16].

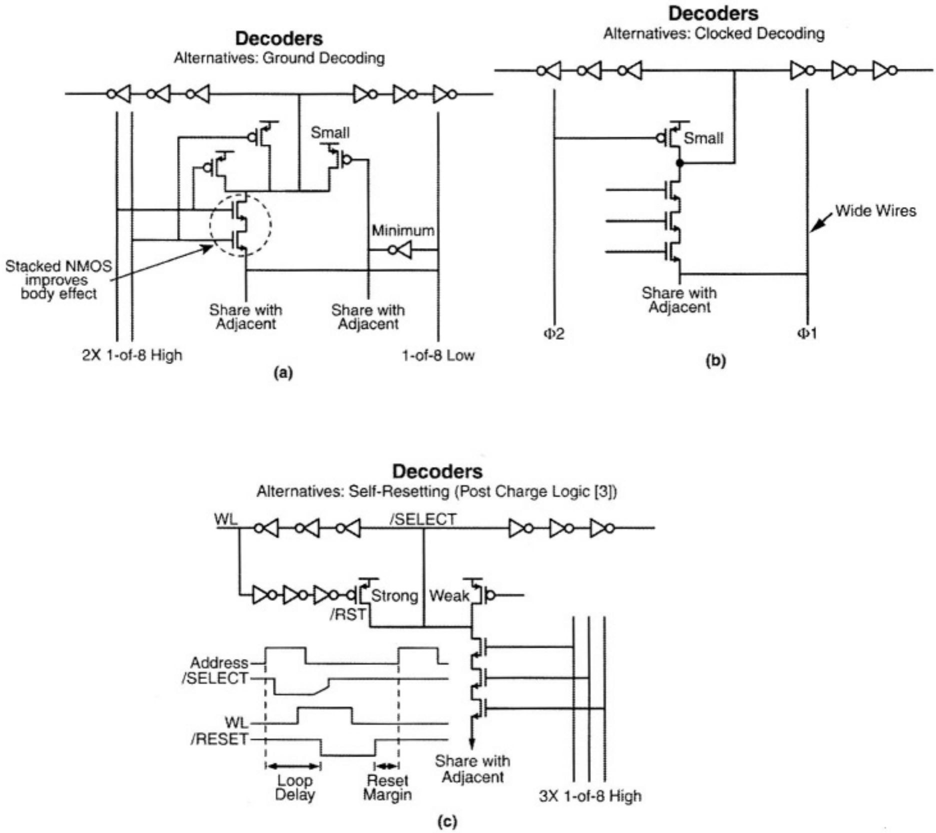


Figure 8.11: Different types of Decoders (a) Ground Decoding (b) Clocked Decoding (c) Self Resetting Logic Decoders.

Row decoder circuits decode the addresses to produce row-select signals (row address select signals and word line signals), and column decoders produce column select addresses. In design with just a few row address bits, row decoders can be implemented using CMOS and pseudo nMOS gates (figure 8.12). Pseudo nMOS decoders achieve higher speed trading at the expense of increased power consumption. Buffers are added to minimize clock skew across the chip and reduce delay caused by heavily loaded word lines.

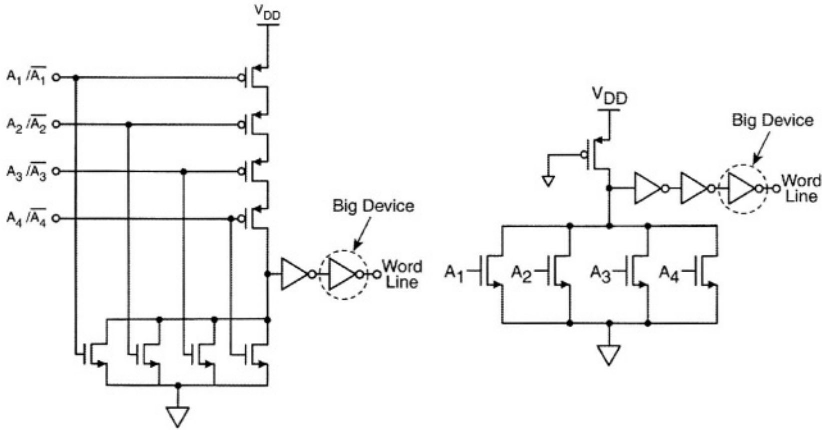


Figure 8.12: R*ow decoders in pseudo (a) CMOS and (b) nMOS configurations.*

Word lines are typically implemented in a metal level for SRAMs. The RC delay is high due to the length of the trace and the capacitance of the trace and loads attached to it. If the word line driver is placed on one end of the memory array from the output of the decoder, the word line has the longest path connecting all the cells in that row, causing the longest delays. Delay can be reduced by re-arrangement of the row decoder into the center of the chip architecture. This cuts the longest length word line in half. Another option is to equally divide the word line into several sub word lines with a hierarchical word line decoder. Global word lines are implemented with higher metal lines (higher levels of metal), are connected to the sub-word lines to reduce RC delays. Hierarchical word lines are most commonly used in the newer architecture SRAMs, SDRAMs and DDR SDRAMs [8.17].

8.9 SRAM Architecture

In conventional architectures when the selected word line is high, all cells connected to the word line in that row are activated, increasing delay and power consumption [8.18]. Figure 8.13 shows a divided word line (DWL) architecture [8.19]. The block select (BS) signals are gated with global word lines generated by the global row decoder to select a particular block. If n_B is the number of blocks, the load of the word line can be reduced to $1/n_B$ by using a divided word line architecture. In addition, the parasitic resistance and capacitance becomes smaller due to decreased word line lengths and hence propagation delays are reduced. During READ/WRITE access, only the word line of a selected block of memory cells is active causing a reduction in power dissipation [8.20, 8.21].

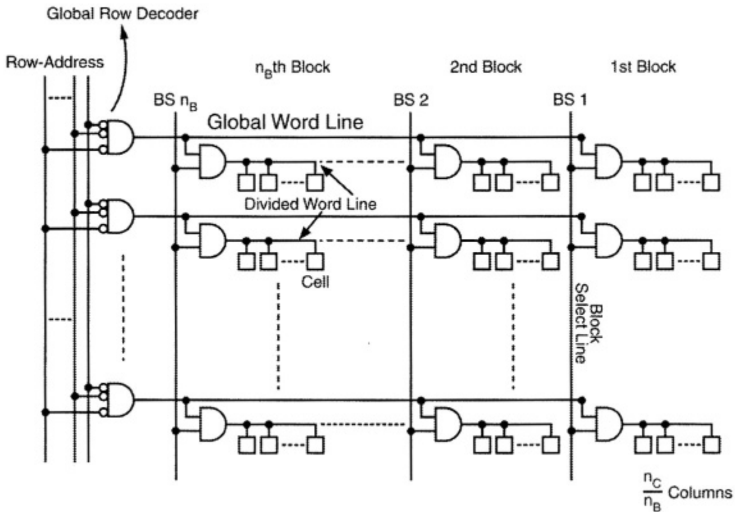


Figure 8.13: Divided Word Line structure in SRAM [8.20, 8.22]

In high density SRAMs, the decoding offered by DWL architecture is not adequate since propagation delay of the 2-stage decoder and increased word line load is still too large. Figure 8.14 shows a hierarchical word line architecture for high density SRAMs. The hierarcical word line (HWL) technique is a further extension of the DWL technique [8.23].

Unlike DWL, which is a 2-layer hierarchy, HWD is a 3-layer hierarchy, which decodes the global, sub global, and the local word lines. The signal produced by the global word decoder is gated with the block select signal to generate the sub global word line. The sub global word line is further gated with a local select to produce the local word line. Since the global word line needs to drive only sub-global row decoders, which are located in the center of the memory arrays, the length of the word line is reduced to that of half of the DWL technique. In addition, the sub global word lines drive one half of the local word line drivers. The local word-lines drive a small number of memory cells. The RC loads of the global and sub global decoders can be effectively reduced using the hierarchical approach, improving the rise and fall times of the word line [8.2, 8.21, 8.23].

One design technique used to guarantee optimum circuit performance is to initialize the bodies of the SRAM cell transistors at the worst case potentials, this exaggerates the cell bias allowing for a robust design differential margin.

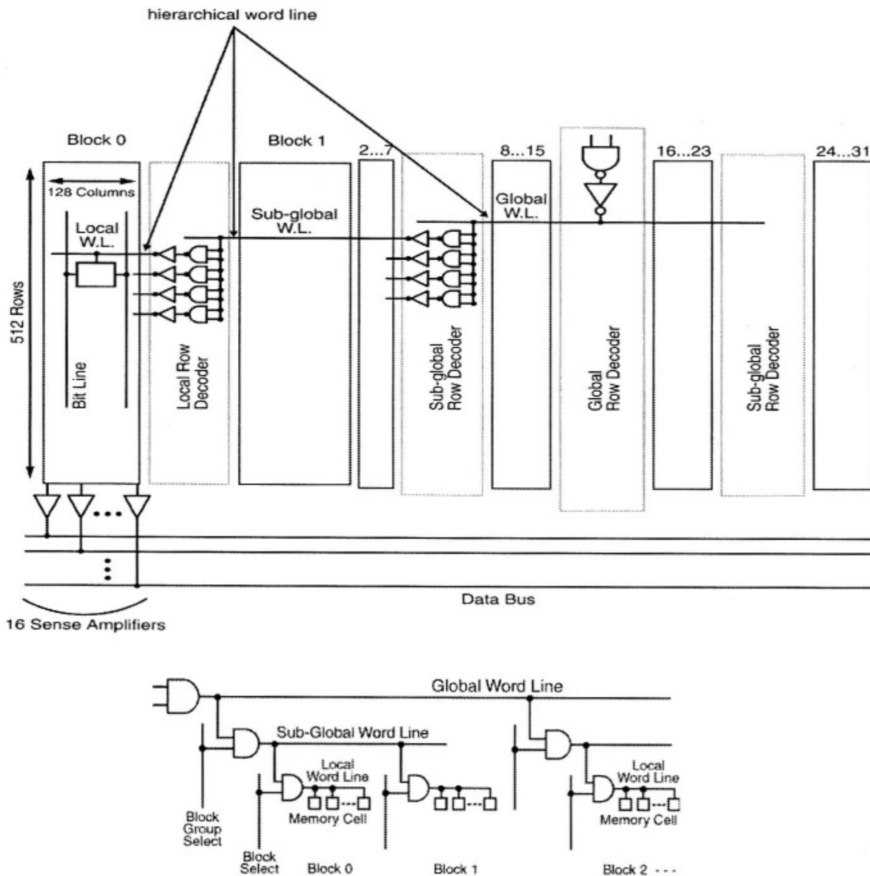


Figure 8.14: Hierarchical Word Line Structure [8.24]

8.10 Bit Line Related Archtecture

Several bitlines are connected to the column multiplexer and data lines in high density SRAM architectures. During write access, the bit line signals are at full swing. Figure 8.16 shows, bit line load circuits with various write control structures. The bit lines are precharged to V_{DD} before a read or write operation. Write recovery equalizes and pre-charges the data lines after the write cycle improving READ access times. Worst case read occurs when a read follows a write operation on the same bit line. Figure 8.16(a) shows the common bit line load circuit. During write operation, a substantial current flows from the bit line loads onto the bit lines dissipating short circuit power and a reduced write speed. A write recovery load is placed at the same end of the bit line from the column select gate (figure 8.16(b)). Pre-charge current can flow only from one end of the bit line during write recovery.

Charging times are slower due to longer bit-lines and RC delay. In figure 8.16(c) the write recovery load is placed at the opposite end of the bit-lines from the column select pass-gate. The load line currents flow from both ends of the bit lines of the column select pass gate and is more efficient than the methods of figure 8.16(a) and (b) [8.2].

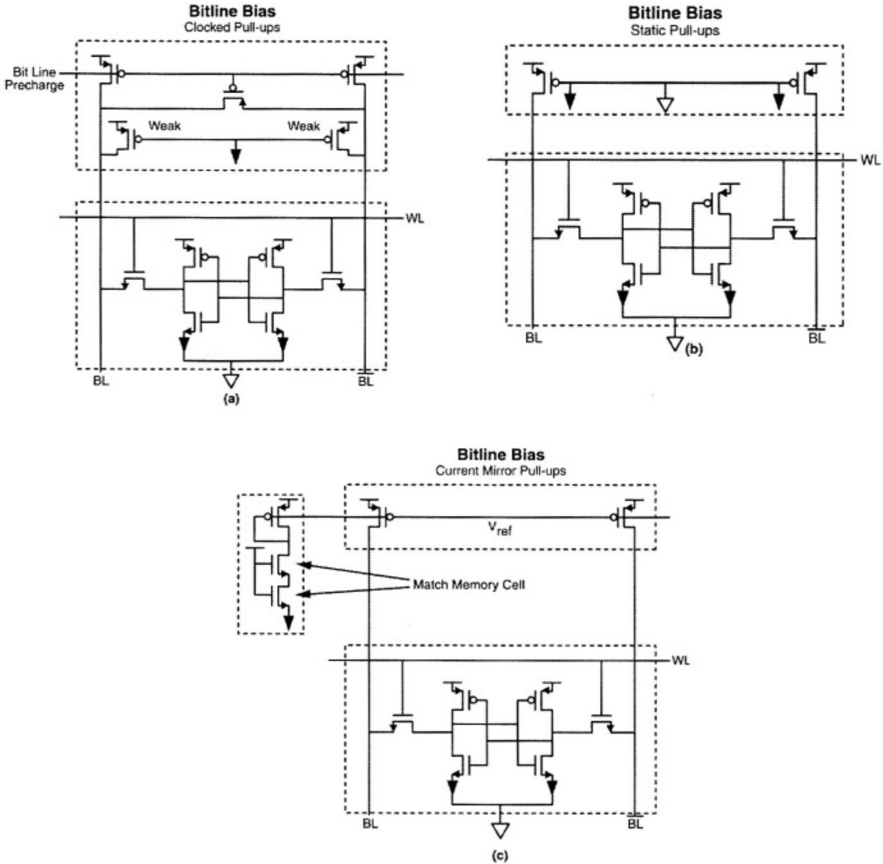


Figure 8.15: Bitline Pre-charge Techniques (a) Clocked pull up technique (b) Static pull up technique (c) Current mirror type bit line precharge

8.11 Sense Amplifiers

SRAM Sense amplifiers can be classified into two types: (a) voltage sensing latch type, and (b) current sensing. The voltage sense amplifier was used in older generation SRAMs but suffers from poor voltage gain at low voltage operation. It also requires a relatively high bit line voltage differential, which is a disadvantage for low power operation. The current sense amplifier operates with a small bit line swing and provides a large voltage gain even at

low supply voltages. It draws more current but is faster than the voltage latch type sense amplifier. In addition, the entire bit line capacitance is seen at the gate of the sense amplifier's pass transistor, which is a drawback for larger memory arrays since many cells are connected to the column lines.

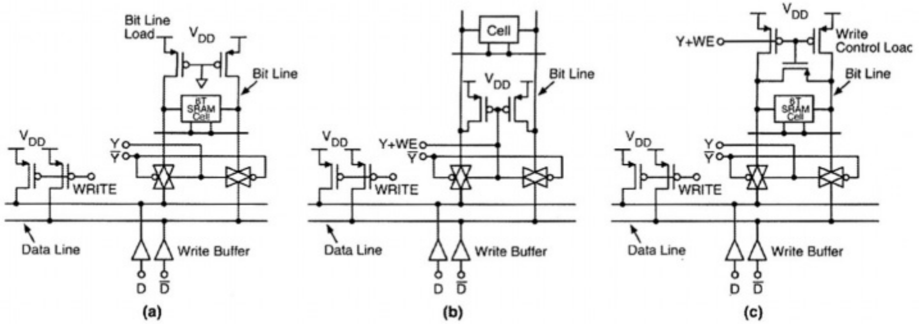


Figure 8.16: Bit Line precharge methods (a) Conventional bitline loads (b) Load placed on the same side of the bit line (c) Load placed on the opposite side of the bit line from the column select pass-gate.

8.11.1 Differential Amplifier

The differential amplifier produces an output voltage that depends on the difference between the two input voltages V_{IN1} and V_{IN2}:

$$\Delta V_D = V_{IN1} - V_{IN2}$$

This circuit converts a dual rail logic signal to single rail to interface to standard logic. MN0 and MN2 are source-coupled input transistors (figure 8.17). PMOS devices MP1 and MP2 are load devices. The logic level output voltage V_{OUT} is based on the differential input voltage and is determined by voltage drop across MP1 which due to current I_{D2}. Current I_{D1} equals I_{D2} at balance point if both nMOS devices have the same aspect ratio [8.25].

The total current I_{SS} is equal to the sum of I_{D1} and I_{D2} and is typically a few microamperes. As ΔV_D increases from a negative value ($V_{IN1} < V_{IN2}$) to a positive value ($V_{IN1} > V_{IN2}$), I_{D1} increases from 0 to a maximum value. The point of equilibrium where $V_{IN1}=V_{IN2}$ is the part where:

$$I_{D1}=I_{D2}=I_{D1\,(max)}/2.$$

At this point the output voltage V_{OUT} is equal to the votlage on the drain of MP1. This type of differential amplifiers is suitable in memory designs as a sense amplifier.

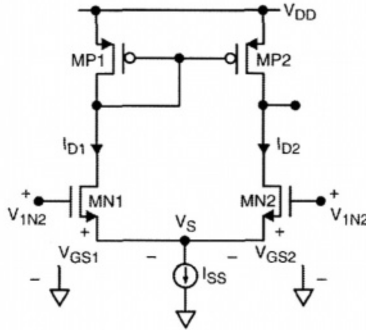


Figure 8.17: Basic differential amplifier [8.25]

The two inputs to the differential amplifier inputs arrive from the true and complementary bit-lines. The difference between the two inputs voltages could be as small as 30mV. The amplifier has to be designed to sense and successfully swing the output in the correct direction. Figure 8.18(a) shows the differential amplifier with the resistor replaced with a nMOS device.

Once the input to the nMOS MN1 turns 'on' (IN high), the sense amplifier (differential amplifier) is activated, and the drain nodes of both the nMOS devices are discharged to ground (figure 8.18(b)). The body voltage of the device MN3 rises initially due to body to gate coupling. The intermediate node N1 also couples higher when either input (V1 or V2) switches low to high due to coupling (see also chapter 10). The body voltages and internal node voltages are shown in the figure 8.18(b). The output true signal is measured at V_{O2} and the complementary side is measured at V_{O1}.

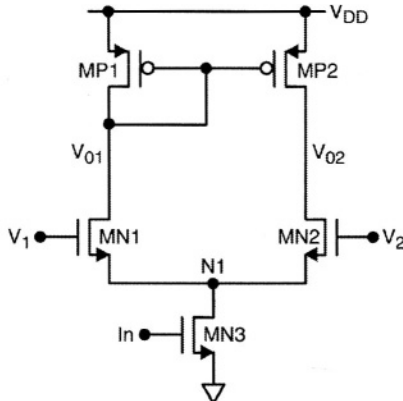


Figure 8.18(a): Differential Amplifier with current sources replaced with nMOS devices.

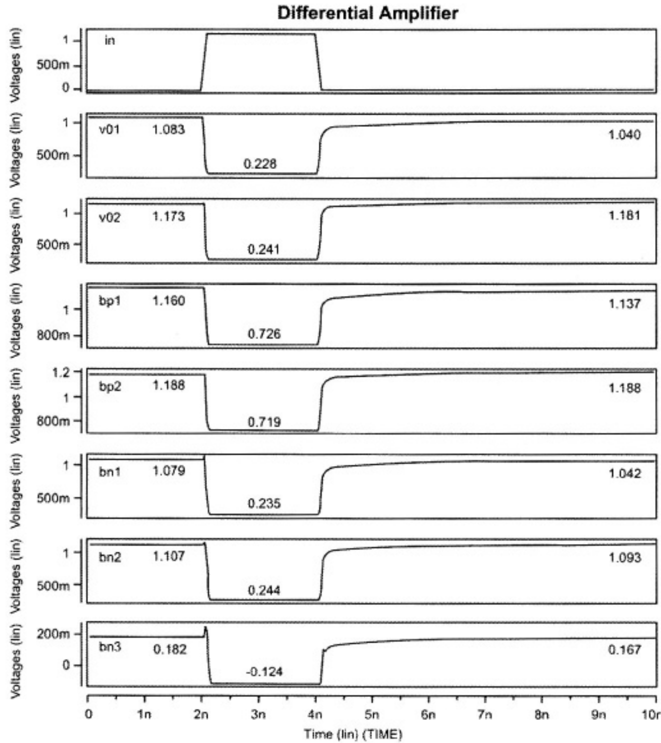


Figure 8.18(b): The body voltages of the nMOS and pMOS devices and the output node waveforms. V_{01} and V_{02} are the output node voltages of the differential amplifier, V_{IN} is the differential amplifier enable clock. The input voltages V1 and V2 of 1.2v and 1.1v respectively are used for simulation [V_{DD}=1.2v].

Figure 8.19 shows various sense amplifier designs. The classic differential sense amplifier is shown in figure 8.19(a). The advantages of this amplifier are as follows [8.26]:

* Symmetric differential output
* Good common mode rejection
* Well controlled gain proportional to I*R
* Simple biasing scheme
* Good supply rejection

Figure 8.19(b & c) shows different types of current mirror sense amplifiers [8.27]. Current mirror sense amplifier shown in figure 8.19(d) uses dual stages for differential output or a single stage for a single ended output.

When the bodies are floating, delay may occur in the amplification of the signal from the bit lines [8.28]. Circuit techniques like body tie to supply, body tied to source and body pulse methods are used to make an efficient and fast sense amplifer in SOI. Figure 8.19(e) [8.2] shows the classic DRAM type clocked sense amplifier.

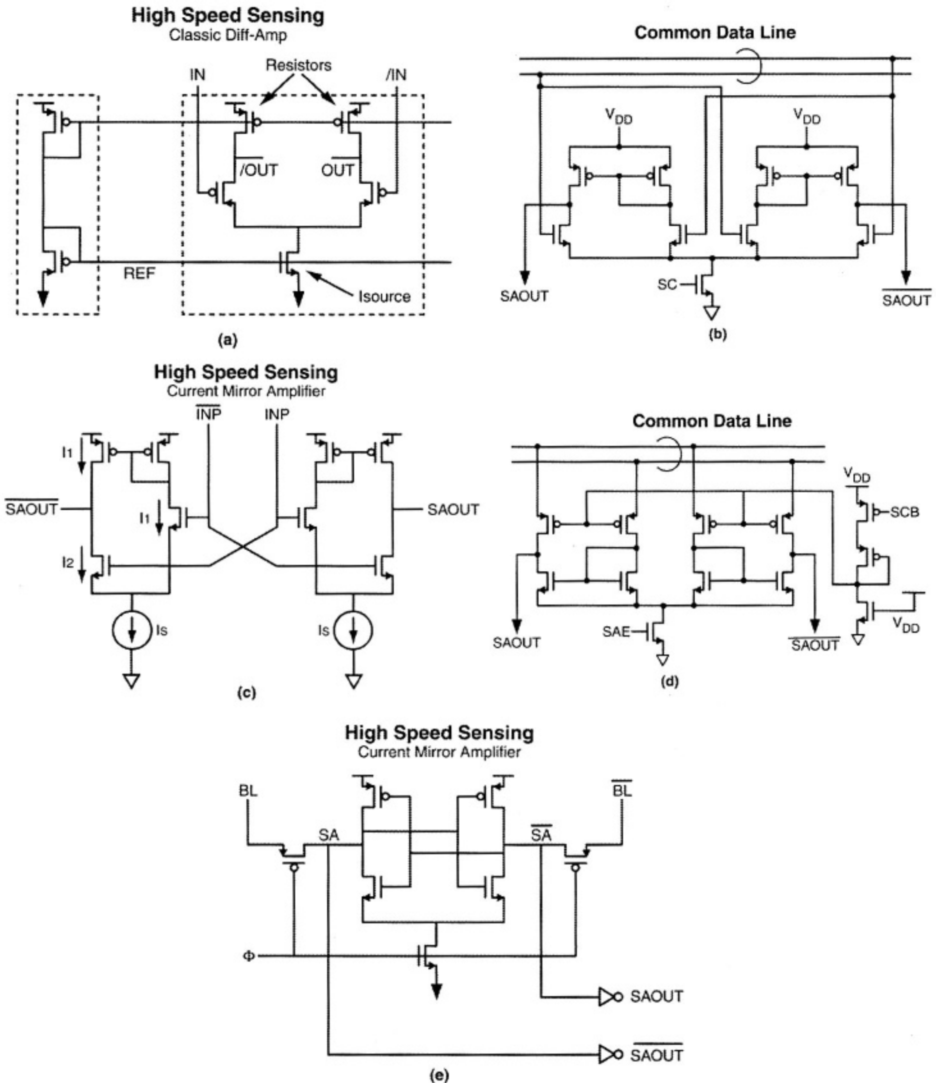


Figure 8.19: Different types of sense amplifiers (a) Classic differential sense amplifier (b,c) Current mirror sense amplifiers (d) Differential amplifier with dual stages (e) Clocked sense amplifiers

The input data is setup to the rising edge of the clock edge. When Φ is low, the bit-lines and sense amp differential nodes (*SA* and $\overline{SA}$) are pre-charged. When Φ makes the transition from high to low, *SA* and $\overline{SA}$ are isolated and the latch is enabled. The drawback of this type of sense amplifier is that when Φ is high, developing bit-line differential does not continue to drive the latch as the pMOS turns off and isolates the bit lines from the latch. The differential voltage developed at the rising edge of Φ must be adequate to reliably set the latch. This sense amplifier limits speed compared to static sense amplifier. The latch has a very high gain, and the capacitance should be kept small. The high side of the latch dips when the sense amplifier is enabled, and is a function of the pMOS size. The sense amplifier clock is often generated from dummy word lines / bit-lines.

8.11.2 Clocked Dual Slope Sense Amplifiers

Most high performance SRAMs use clocked dual slope sense amplifiers (Figure 8.20) [8.29 - 8.31]. During sensing, a narrow device MN0 (narrow transistor) is turned 'on' allowing current to increase slowly, and developing a differential voltage across the cross coupled pair. The wide device MN1 (wide transistor) is subsequently turned 'on' when adequate cell differential voltage is developed, to provide a fast pull down of the bit line. With older technologies the differential voltage before MN1 is turned 'on' is 150-200mV but is as low as 40-80mV as technology scales.

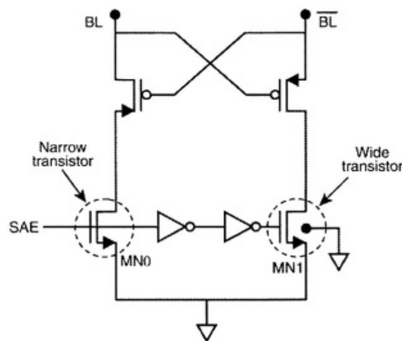


Figure 8.20(a): Schematic of a Dual Slope Sense Amplifier

Initial development of the differential voltage is critical. If the transistor is turned on with a large current, then the bit lines may not develop enough voltage differential, and the output voltage will collapse before the difference is amplified. In floating body SOI configuration, the uncertainity in the floating body potential translates into the uncertainity in the transistor threshold voltage. A lower threshold voltage in the narrow device may cause

the sense amplifier to turn on hard. The imbalance in the sense transistor threshold voltages further degrades the sense margin and sensing speed. Inadequate differential voltage on the bit lines can cause the sense amplifier to flip to the wrong state.

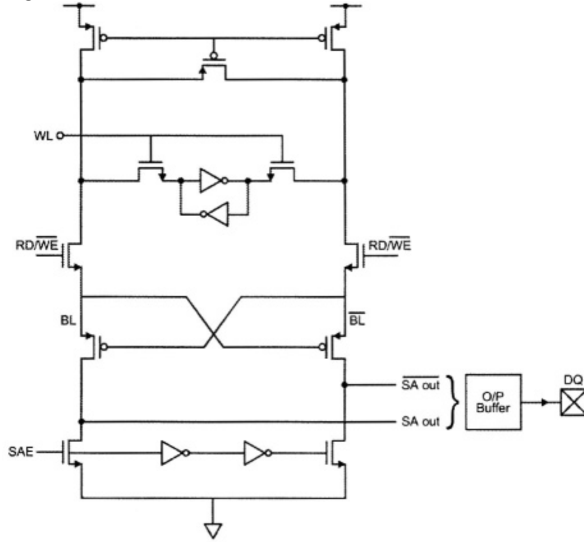


Figure 8.20(b) shows the critical path schematic for a read operation with the dual slope sense amplifier.

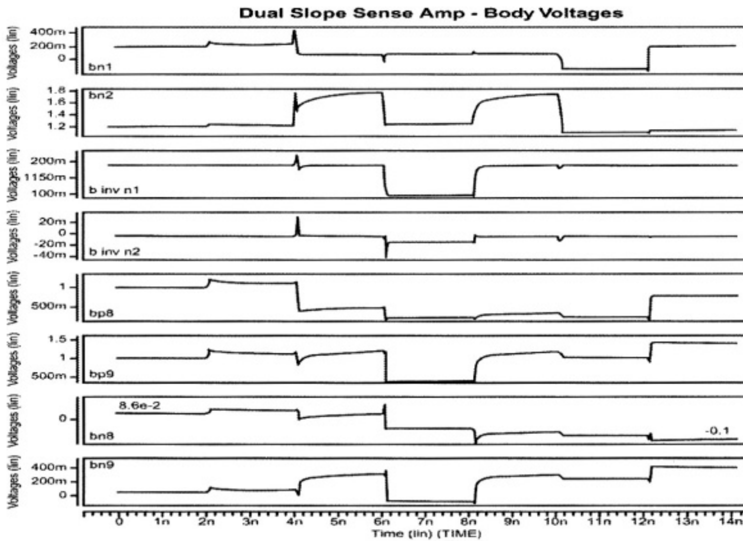


Figure 8.20(c): Body voltages of a SRAM critical path using dual slope sense amplifiers.

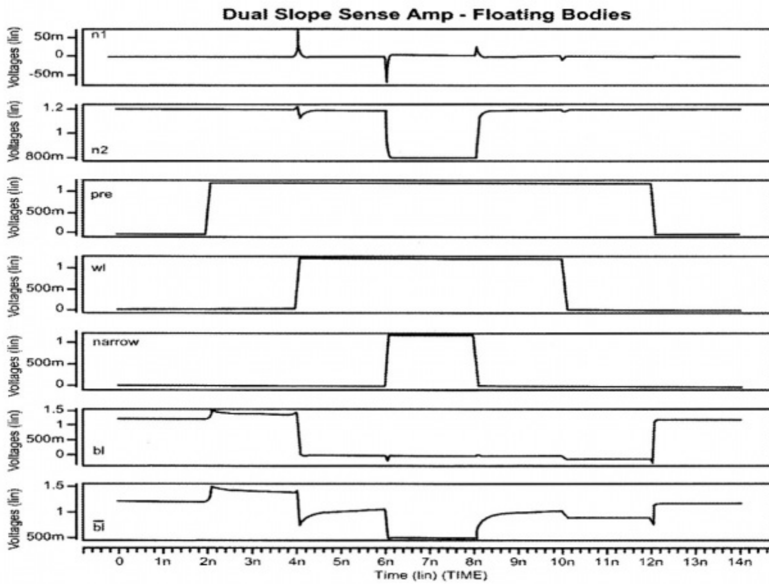


Figure 8.20(d): Internal critical node voltages of a critical path SRAM simulation with dual slope sense amplifier.

Figure 8.20(b) shows the critical path schematic for a read operation with the dual slope sense amplifier. Figure 8.20(c & d) shows the node and body voltages. Precharge is shut off before the word line turns 'ON'. The narrow passgate MN0 turns on first to enable to sense amplifier. The wide transistor MN1 turns on with a slide delay to help in developing adequate differential on the bit lines. The bit line signal is amplified by the sense amplifier. The waveforms show that as a result of the floating bodies bit line swings are noisy, though body ties can be used to suppress the noise and reduce variability in the signal timings.

8.11.3 Dynamic Body Charge Controlled Sense Amplifier

Dynamic sensing circuits in SOI SRAMs are particularly prone to performance variation and mismatch in their transfer characteristics [8.30, 8.32]. Such discrepancies are attributed to device body potential and threshold voltage differences due to operating history and asymmetrical time constants to charge and discharge SOI MOSFET bodies. Repetitive read operations performed by a sense amplifier over time can result in significant asymmetric charge accumulation, body voltage bias, sense point runaway, and consequentially functional failures in dual railed RAMs or register file circuits.

Figure 8.21 shows a dynamic body charge controlled sense amplifier technique to improve matching of CMOS device threshold voltage characteristics in a partially depleted SOI technology.

Symmetric long channel MOSFETs are provided to charge the bodies of the cross-coupled sensing devices to a steady voltage, minimizing the history dependent body potential mismatches. Matching device characteristics can be restored prior to each time the sense amplifier is activated. Body contacted MOSFETs are returned to their floating body states once charging is completed. This charge flooding mechanism, as opposed to functionally discharging of MOSFET body voltages, has the advantage of providing low threshold voltage operation.

The dynamic body charge control technique achieves low V_T and improves performance in RAM circuits. The body pulse signal does not have a stringent timing requirement. So, its effectiveness is relatively insensitive to body contact quality of the MOSFET. It also offers the advantages of speed and noise immunity in the low voltage, low power operating region. Floating MOSFET bodies influence the behaviour of partially depleted SOI CMOS circuits in number of ways [8.18, 8.33].

For SOI CMOS circuits, tracking the threshold voltage requires tight control of critical MOSFET body voltage in addition to matching of layout and electrical parameters commonly deemed essential for analog designs. The same method can be applied to other types of cross-coupled sensing or differential CMOS circuit topologies where mismatch and cycle to cycle variations cannot be tolerated.

Dynamic body charge control techniques may be used to regulate body potentials [8.31]. This is area efficient but possibly provides reduced precision matching in cross coupled dual railed circuits, due to long time constants to move floating body charges. Since the cross-coupled sensing MOSFET bodies are symmetrically charged to a known stable voltage whenever a read operation is anticipated, initial equalization is achieved for bit lines and cycle dependent timing uncertainities typically found in SOI operations are also minimized. By consistently elevating the sensing FET bodies to the same potential we always start a high performance read cycle with bodies well disposed. After equalization of the body potentials, sensing MOSFET's resume their floating body nature.

Figure 8.21(a) shows a basic implementation of this technique, where the body-contacted nMOS are charged to V_{DD} through the low current pMOS

devices. If the pMOS device floating body effects are non negligible, a complimentary charge to ground topology using long channel nMOS devices can be built.

The input to the flooding pMOS is a pulse signal, which is enabled prior to triggering of the sense amplifier. The leading edge of the pulse can be derived from the active low sense amplifier reset/restore signal. The pulse width is relatively non-critical if signal integrity is maintained at the highest body contact resistivity. The MOSFET bodies can be raised to the point where the diode between body and source is forward biased [8.33].

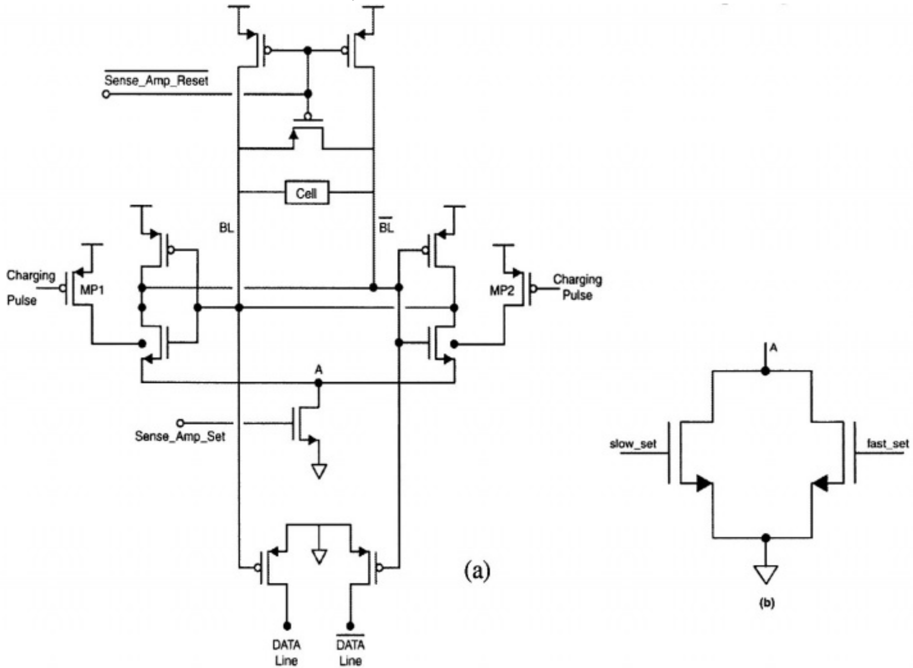


Figure 8.21: (a) *Dynamic body control for sense amplifiers in partially depleted SOI (b) Controlling the sense amplifier enabled signal by splitting into two devices and turning one of the transistors with a delay (split firing controls).*

Determination of the minimum required bit line offset voltage eliminates the need for device resizing in the sense amplifier, which is required for migration to process technologies. Since the charging pulse is turned ON only during part of the active read cycle, excessive leakage is not a major consideration. Body contacted nMOS devices with longer than minimum channel length can be used. Body contacts reduce threshold variations and minimize the influence of body contact resistivity. To initiate the operation

in a read cycle, the two nMOS bodies are flooded before the offset voltage differential starts to develop in the complementary bit lines. The nMOS bodies are released before the amplifier is triggered, i.e., the sense-amp-set signal goes high and the sense amplifier enters its high gain region. Both body contacted nMOS return to operate in the floating body mode at the completion of charge flooding action. The time constants of the body voltages are adequate to hold both body voltages at similar levels during sensing. Timing sequence for flooding and sensing actions is shown in figure 8.22. The charging pulse gate control to the flooding pMOS is active low.

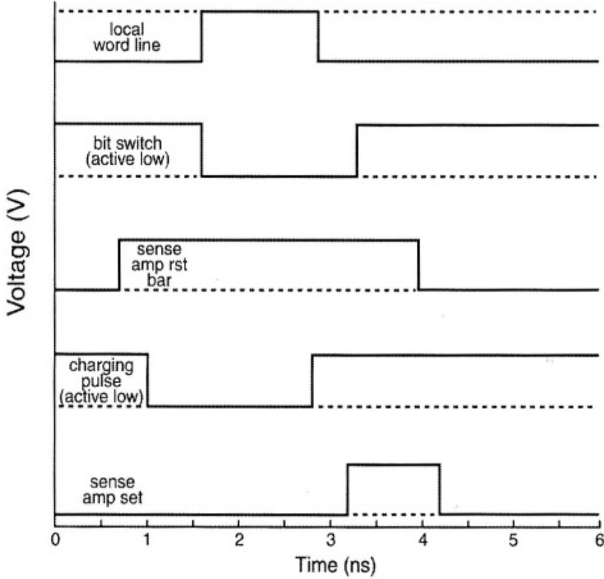


Figure 8.22. Timing sequence for flooding and sensing actions.

The charging pulse has a relaxed timing requirement. Such that even if a body contact resistance disparity exists, equalization for both sensing nFET bodies can be completed well before activation of the sense amplifier. Figure 8.23 shows the operation of this type of equalization, where the bit line timing difference is negligible compared with matched body contact. The equalization speed of the body voltages depends on the physical separation and resistivity of the body contacts.

The dynamic body control technique is useful in low power SOI SRAM applications while multiple threshold process technology can be used to improve sense amplifier performance. Low voltage operations may be

effectively realized with a single threshold voltage technology. The use of dynamic body control may permit process simplificiation.

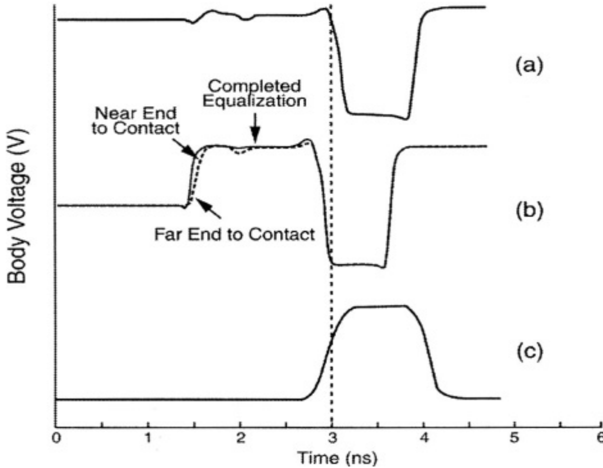


Figure 8.23: Body Voltage and Bit Line Offset Voltage Versus Time

A sense amplifier with split firing controls (figure 8.21(b)) has been used in the study of low operating voltages [8.30, 8.31]. The sense circuit slow set signal precedes the fast set signal with the relative timing arrangement for control signals unchanged. A body potential mismatch of 10% supply voltage is assumed on the sensing nMOS devices. Bit line offset voltages measured at 50% V_{DD} develop normally in the dynamic controlled case as well as the conventional design (Figure 8.24).

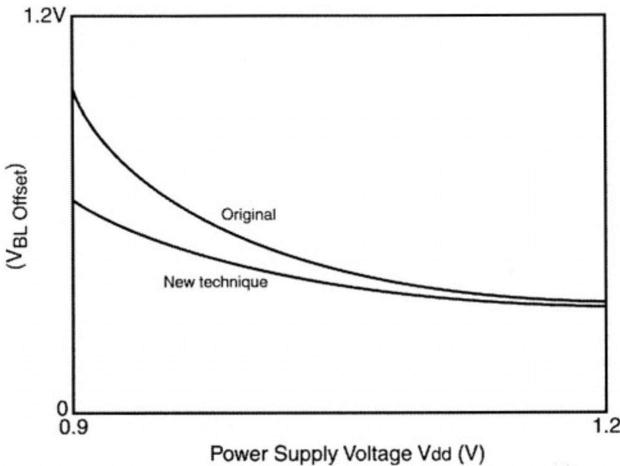


Figure 8.24: Bit line offset voltage versus power supply voltage [8.31, 8.33]

Offset voltage degradation becomes noticeable in the high gain region for the conventional design due to threshold voltage differences. Differential voltage separation for the array before sense amplifiers are enabled is determined by cell characteristics and bit line loading. Higher differential voltages result in improved noise margin and operating speed.

8.11.4 Sense Amplifier Techniques

Many sense amplifier techniques are used. Some of possible most common options are evaluated here. (1) and (2) control the threshold voltage by body ties to critical sense devices. (3), (4) and (5) manipulate threshold voltage dynamically by introducing additional circuit elements and control signals.

(1) The body to source connected configurations and the conventional body grounded configuration for the sense transistors are simple to implement. However, this suppresses the floating body characteristics resulting in increased threshold voltages and reduced speed. Although there is a choice between a body-to-source or body grounded configuration, it can be performance or layout based, typically the body to source configuration is usually the best choice for both performance and layout optimzation. This technique has been used in low voltage high density SOI DRAM design [8.34].

(2) The body to body connected configuration for sense transistors effectively suppress history induced body potential mismatch. However, it degrades performance in the large-signal operating region. Timing variation exists between cycles because the initial voltage of connected MOSFET bodies is subject to the history of the circuit. If threshold voltage in SOI technology is highly sensitive to elevated body voltages, a body to source connected configuration is a better solution to address timing uncertainties. Solutions (1) and (2) have simple timing requirements as no dynamic body charging or discharging is required.

(3) The dynamic body discharge circuit topology [8.35](Figure 8.25) has lower sense performance improvement compared to the body charge flooding scheme because of the higher threshold voltage. The gate control for dynamic body discharge can be implemented in a similar manner to that of the body charge flooding technique, using the same timing sequence. Performance of the circuit is similar to the body tied to source configuration.

(4) Body to drain modulated techniques may be used [8.36]. Voltages on the bit lines remain close to V_{DD}, though these drift apart as the bit line voltage offset develops. The flooding devices in the design can be turned OFF after

boosting the body potential, or can be left ON until before the sense amp set signal is enabled. During sensing, allowing the two nMOS bodies to float achieves similar performance as the body-to-drain modulated technique.

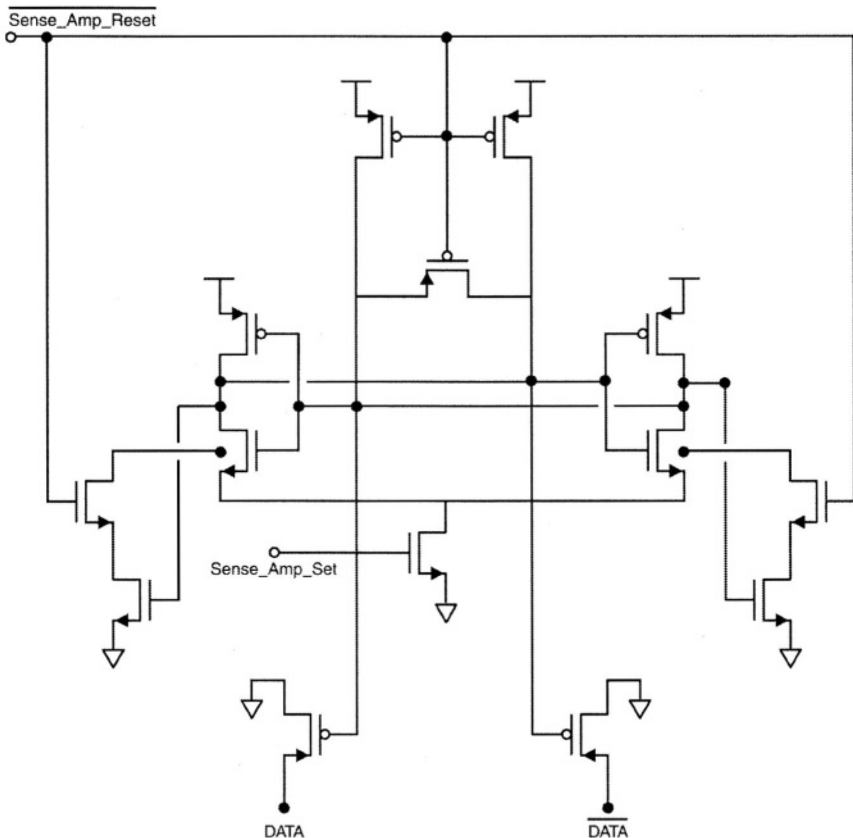


Figure 8.25: Dynamic Body sense amplifier with discharging nMOSFETs.

(5) The 'super body synchronous' sensing scheme [8.37], controls device threshold voltages dynamically. During equalization and sensing, both nMOS and pMOS bodies are biased at fixed voltage levels to achieve low threshold voltages.

After sensing, MOSFET bodies are biased at different voltage levels to achieve high threshold voltages. This method can be adapted for use in the SOI SRAM sense amplifier by selectively biasing the two sensing nMOS bodies towards V_{DD} for a portion of the read cycle and at ground for the remainder of the read and standby cycle.

The nMOS turn OFF times on one side of the latch can be slowed down in PD-SOI SRAMs with excessively low threshold voltages due to high body voltages. In addition, the voltage level on that side of the latch can degrade under extreme operating conditions. Low threshold voltage increases switching and DC power loss. Complete circuit analysis is required to determine the modulation of the body bias. Optimum switching time is difficult to determine, because even a calibrated SOI device model suffers some degree of inaccuracy in regions where transient terminal or body voltage gradients are large. Previous attempts to resolve this have changed the body bias state before the onset of active latching. However, this is a performance compromise.

The charge flooding circuit appears to offer the best performance among the circuits discussed here, but may be difficult to implement. Area saving is achieved due to the omission of discharge paths to ground. This method also outperforms the alternatives in operating voltage range and sense speed.

8.12 Mismatches in Sense Amplifiers

Some of the external design issues for sense amplifiers are: High speed, high packing density, ability to fit in the pitch of the bit lines, wide timing margin and stability. In addition to these, the sense amplifier electrical (and layout) characteristics should be closely matched to improve sensitivity to small differential voltages on the bit-lines. Floating body sense amplifier inputs are susceptible to floating body variations.

Mismatch in threshold voltage between cross-coupled nMOS devices in a sense amplifier introduces additional noise, which affects sensing reliability. Threshold voltage mismatches can arise from mismatched body charge state, which is dependent on the operating history of the MOSFETs [8.38]. Selective use of body ties allows closely matched electrical characteristics within groups of devices without seriously degrading the performance advantages in SOI. Device modeling has demonstrated the use of body ties as a means of reducing electrical mismatch between cross coupled SOI nMOS [8.39].

Figure 8.26 shows a basic differential sense amplifier connected to the bit lines and memory cell. If BL is logical low, and $\overline{BL}$ is logical high, the sense amplifier repeatedly senses a low. The sense amplifier then tends to build up a bias towards sensing low since the body potential of MN3 is lower. This increases MN3's threshold voltage and trip point at which the low (BL) side inverter will switch. Thus, makes it easier to read a '0'. The body potential of MN4 will be higher and will have a lower threshold voltage [8.2, 8.7].

When the sense amplifier has latched "0" and "1" to *BL* and $\overline{BL}$ respectively, the body potential of device MN3 is higher than the body potential of device MN4 since $V_{DS}=V_{DD}$ for device MN3, and $V_{DS}=0V$, for device MN4. Once the sense amplifier is disabled (SAE goes low), device MN3 has its source pre-charging and MN4 has its source and drain pre-charging. This increases device MN4's body potential compared to device MN3 body potential. The body potential of devices MN3 and MN4 eventually equalize if the sense amplifier is not enabled for a long period. Worst case mismatch for a PD-SOI sense amplifier is when it spends the largest percentage of its time holding the latched read data and is run at high frequency [8.7, 8.39].

The worst case condition is when a logical '1' is read for many cycles before a '0' is read. This results in an accumulation of charge in the body of the sense amplifier transistors on the 'high' input [8.40].

One means of reducing sense amplifier mismatch is to electrically connect the bodies of the critical devices to a known value. Cross-tied bodies share accumulated body potential. An alternative method connects the body of each critical transistor to its source (Figure 8.27). The sense amplifier nMOS devices MN0 and MN1 will have variable but higher body potentials and lower threshold voltages than if their bodies had each been tied to GND. This forces the sense amplifier's nMOS devices to permanently lower body potentials and raise collective threshold voltages [8.41].

The body contact technology adds capacitance, area and delay to the sense amplifier. A different approach is to wait for a larger differential to develop on the bit lines before enabling the sense amplifier, thus eliminating the need for avoid body contacts and trading off performance.

Enable timing of the sense amplifier is very critical. The sense amplifier strobe delay can be initiated by the same clock that enables the word lines and other critical signals once the differential has been developed on the bit lines. This delay circuit is active every cycle, and reaches a stable SOI steady state from cycle to cycle. The word line is activated only when the word line enable signal goes high which may be slow for the first switch and fast at the second switch, or also attain steady state delay [8.7].

Small arrays use inverters instead of sense amplifiers. This requires that the bit lines be fully discharged to switch the inverter. The parasitic bipolar leakage and the increased capacitance of the bit-lines limit the speed of the bit-lines.

The same clock triggers all the word lines. The word line decoder circuits develop a higher body voltage and lower threshold voltages if the same cell is read repeatedly. Conversely, if a cell is not read for a long time (milliseconds), the body potential of the devices in the cell are reduced and tend to have a higher threshold voltage causing delay in the word line timing.

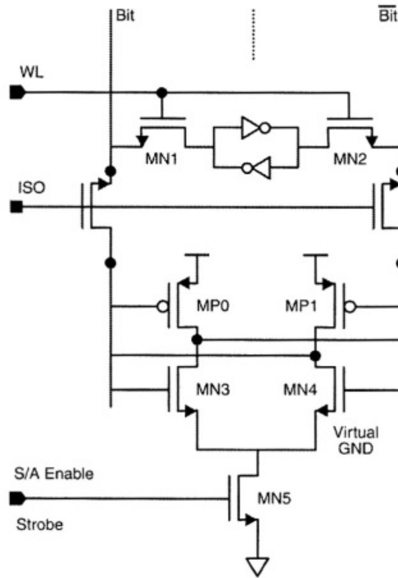


Figure 8.26: Basic Differential Sense amplifier.

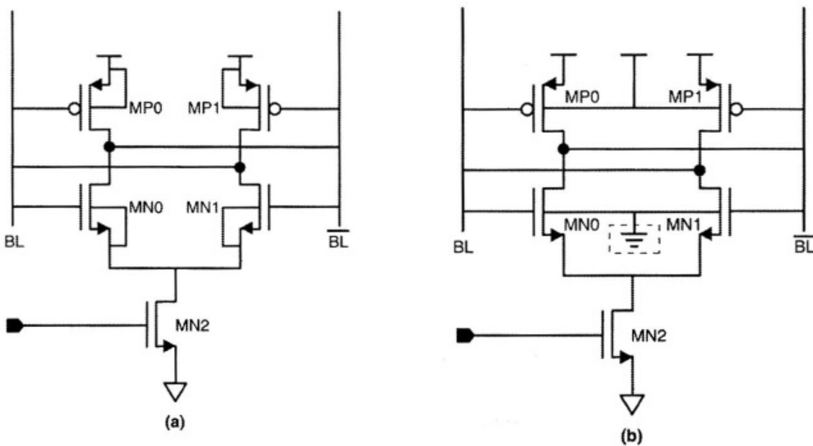


Figure 8.27: (a) Source Connection (b) Body-body connection.

The imbalance in the sense transistor threshold voltages further degrades the sense margin and sensing speed. The differential voltage may collapse in the early development stage, resulting in an incorrect state in the sense amplifier output.

To improve noise margin, it is possible to insert body contacts in the sense amplifier transistors or the bodies of the cross-coupled sense transistors can be tied together, forcing equal body potential and threshold voltage on the sense transistors. Tying bodies together saves area, but the overall sense margin does not improve since sense transistors on different differential bit line pairs still have different body potentials because of their dependency of the bit line capacitance on the cell contents. The only means to guarantee improvement in the sense margin is to equalize bit lines and all body potentials of the cross coupled MOSFETs within each sense amplifier.

Leakage currents due to parasitic bipolar current, charge sharing, sub-threshold currents, and other processing factors lead to insufficient bias margin, making the design susceptible to defect induced bias. In bulk CMOS, the substrate is tied to ground allowing defects to be screened during testing. Hence AC testing is a must for SOI devices to screen defects [8.7].

8.12.1 Offset Considerations for High Speed Sensing

It is critical to minimize offset voltages in the sense amplifier. The offset voltages between the internal sense amplifier storage nodes can cause false sensing and slower access times. The offsets have several sources:

Alignment induced offsets

* Transistor size modulation
* Parasitic resistance mismatch
* Capacitive mismatch

Optical and chemical proximity effects

* Critical dimensions influenced by adjacent geometries

Statistical variations in transistor characteristics

* Critical dimensions variations
* Doping variations causing threshold voltage changes.

Offsets can be minimized by careful layout of the sense amplifier. Alignment induced offsets are a major concern, and layout techniques that can minimize alignment induced offsets are shown in chapter 5. Some of the solutions that are used to minimize alignment induced offsets are [8.2]:

* Avoid diffusion and poly bends

* Add margin to critical design rules
    * Use polysilicon end-caps on devices
    * Bend polysilicon only after crossing a diffusion edge.

* Consideration of poly misalignment to diffusion edges (Figure 8.28(a))
    * As transistor (N1) becomes wider → Transistor (N2) Smaller
    * As Area (N1) decreases → Area (N2) increases
    * Leakage with minimal end-cap.
    * Transistor "L" increases at top edge.

* Contact misalignment to poly (Figure 8.28(b))
    * Rd (N1) decreases → Rd (N2) increases
    * Rs (N1) increases → Rs (N2) decreases.

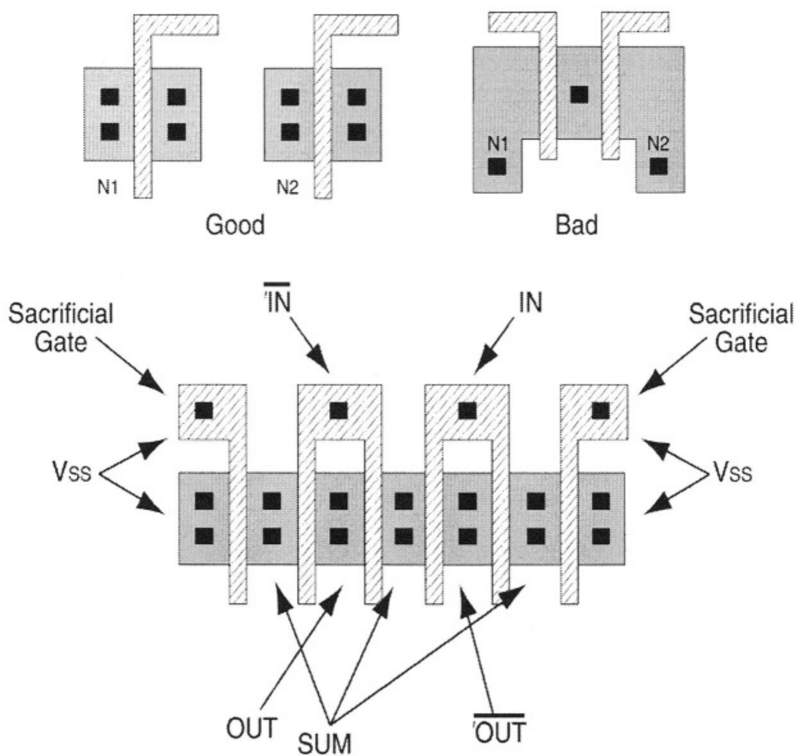


Figure 8.28: Misalignment in layout that causes mismatches in device characteristics

The following is a summary of proximity Effects:

Chemical proximity effects
* Out-flux of reaction biproducts influenced by “spacing”.
* Narrow spaces etch differently than wide spaces
* Etched CD influenced by line spacing

Optical proximity effects
* Light Scattering off adjacent mask edges effect printed CD
* CD is a function of line spacing.

Layout solutions (Figure 8.28(c))
* Use folded gates
* Add sacrificial dummy gates or field poly at edges of layout
* Enforce symmetry in adjacent geometries. (this uses contacts)
* Use multiple transistor legs, more is better

Statistical Effects
* Statistical Events (“defects”) cause local transistor mismatches
* Doping Variations → threshold voltage mismatch
* Photo resist non-uniformities
  - Blobs, Mouse Bites etc → CD Mismatches

8.13 Mismatch in SRAM Cells

8.13.1 Body Bias

As with sense amplifier transistors, the devices in the SRAM cells also accumulate body bias by remaining in one state until the body voltages reach equilibrium. Body ties in the SRAM array hurt performance due to increased capacitance and increases area. Body bias can accumulate on any of the memory cell transistors. The pass gate body bias is the most degrading because its source can reach supply voltages and induce more variability in the bias.

8.13.2 Supply Rail Droop

Power supply distribution contributes to low drive current and is a major cause for read disturbs in bulk and SOI SRAMs. Supply voltage variations can induce bias and degrade bit line differentials affecting sense margin. These problems exist in SOI and bulk CMOS, although, in SOI this problem is aggravated by the floating body. Bias sensitivity increases inversely with voltage, and depends also with temperature and the ratio of the forward biased to reverse biased junction diode currents in bulk [8.7, 8.17].

8.13.3 Body-to-Body Coupling

Cell stability and body charge variation both affect read disturb. A previous section discussed read disturbs with respect to cell stability. This section addresses the phenomenon with respect to body charges. Power dissipation in a SRAM cell pass-gate can couple into a neighbouring cell and can accumulate body charge. This reduces threshold voltage more than anticipated by its recent history and causes '*Read Disturb*'. Adjacent body coupling, cell instability, un-optimized cell ratio and bit-line coupling are all causes for read disturbs. SOI SRAM is very sensitive to defects than bulk CMOS SRAMs.

8.13.4 Common Mode Supply Rail

The elimination of N-well implanted in bulk substrate to establish pMOS in p-substrate wafers also removes the decoupling capacitance. The Nwell is tied to V_{DD} in bulk devices, forming a junction with its associated capacitance directly to substrate, which is tied to GND.

8.13.5 MOS Junction capacitance

The junction capacitance in SOI is a useful buffer to reduce supply noise. Reduced SOI junction capacitance removes the stabilizing capacitance within the SRAM cell. Because the cell response is a capacitive divider of the injected charge between junctions and the body, lower junction capacitance can cause upsets in the cell [8.7].

History effects occuring in the array are no different from those exhibited in other logic structures. Passive history effect occurs in array devices only when the drain or source voltage of a device changes.

Capacitive coupling to adjacent bit lines is another critical factor in designing high speed SOI SRAMs. If a read '1' is performed, the true bit line contains a 'high' and the complementary bitline a 'low'. The capacitive coupling from the adjacent bitlines (which may be switching low) can cause degradation of the high signal on bit line. This concern is aggravated in SOI due to the absence of the source and drain capacitances. This results in reduced total bitline capacitance. This adjacent bitline capacitive coupling can reduce the bit line differential in SOI drastically compared to bulk [8.42,8.43].

8.13.6 Self Heating

Devices in close proximity share common diffusion islands and transfer heat between each other. The pass gate sizes are designed for a fixed cell ratio and can be disturbed by heat transferred across the common diffusion island.

Reading the same data at the same location in the array repeatedly causes one side of the bit-line to discharge through the pass gate and pull down. Thus, the cell develops a tendency towards reading a "1" or a "0". At high frequencies, the duty factor of the devices are high, resulting in gradual degradation in transconductance due to heating. Self-heating is a function of the drive strength of the load. This degradation does not adversely affect cell stability, as weaker pass gate strength tends to stabilize the 6T cell, but it can affect read delay. Adjacent pass gates from neighboring cells can see a degradation in trans-conductance. SRAM cells suffer higher body induced variability and are stable only after the body reaches steady state.

8.14 SER Issues in SRAMs

An important advantage of SOI is improvement in soft error rate [8.43], mainly because of its long history in radiation-hardened applications and presence of buried oxide. Alpha particles from radioactive elements in packaging are known to induce soft errors and impose design constraints in six-transistor planar SRAM cells (figure 8.1 (a)). The net charge imbalance Q_{CRIT} necessary to upset the cell state is equal to $C\Delta V$, where C is the capacitance seen at the storage node of the SRAM cell and ΔV is the cell differential voltage (supply voltage in a 6T SRAM Cell). For a 0.5μm technology, the Q_{CRIT} is around 25-30fC [8.31]. Cell size shrinks with technology, and Q_{CRIT} also reduces. Due to the presence of buried oxide, it is very difficult for the alpha particles to get injected into the channel. The reduced cell size and storage node capacitance improves cell performance due to reduced parasitics. The sub-linear improvement in cell access time is traded against the almost exponential deterioration in SER. Alpha induced SER is expected to surpass the cosmic ray (which is relatively insensitive to Q_{CRIT} and SER remains relatively constant as technology scales) and become the major failure mechanisms for technologies 0.15μm and below [8.31].

In bulk CMOS, ∝-generated charges are collected mainly by the funneling effect [8.44], when particles collide with the drain diffusion layer. This is not significant in SOI MOSFET's due to presence of buried oxide [8.1]. Charge collection can only occur in SOI MOSFET's when an ∝-particle interacts with the channel region [8.45, 8.46]. The amount of ∝-generated charges in SOI MOSFET is less than bulk. The total charge collected at the cell storage node is significantly higher than ∝-generated charges due to the parasitic bipolar effect. Alpha induced bipolar current flows over extended time periods [8.46, 8.47]. Although, one would expect significant SER improvement in SOI, it is also been reported otherwise [8.48]. Although use of body contacts reduces parasitic bipolar leakage, density requirements may not permit such an option. The curve for SER (figure 8.29) of SOI SRAMs

behaves differently from bulk SRAM as the gate length is scaled. For bulk SRAM, the scaling of the cell and supply voltage reduces Q_{CRIT}, and the SER rises almost exponentially. For SOI SRAM, scaling the supply voltage reduces the parasitic bipolar effect since the body, which acts as the base of the parasitic, will be at a lower potential. The reduced parasitic bipolar effect compensates for the reduction in the Q_{CRIT} of SOI SRAM, and the SER remains relatively flat with technology scaling [8.31]. The noise currents for SOI and bulk behave differently over time.

The cell size and circuit performance of a SOI SRAM can be further improved by using a cell layout with abutted n^+ and p^+ drain regions [8.31]. The n^+ and p^+ drains of the inverter output node and the source/drain region of the nMOS access transistor are connected by abutting Ti-Silicided n^+ and p^+ regions. This removes layout constraint of well spacing in the bulk CMOS technology and allows a single contact layout for the cross-coupled inverters in the SOI SRAM 6T cell.

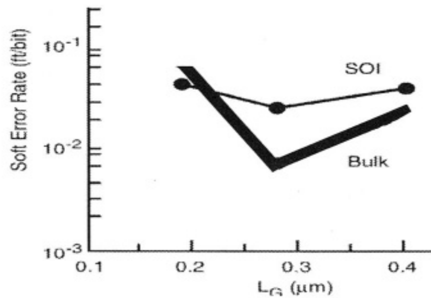


Figure 8.29: 6T SRAM cell SER's in SOI and bulk SRAM as a function of gate length [8.31]

8.15 SOI CMOS Memory Challenges

The general effects of pass-gate leakages on memory circuits comprises aggrandized number of read and write errors, lengthier access and cycle times, decreased radiation hardness, higher power dissipation, unreliable or impaired memory operation. CMOS PD-SOI memory have to be designed carefully due to floating body effects: Some of the issues are summarized:

(1) Timing failures in sub-circuit activations and logic gate functions, operation and noise margin degradations
(2) Read and write data pattern sensitivites in arrays
(3) Offset increases in sense amplifiers
(4) Gain reductions in sense and other amplifyng circuits

(5) Data losses in memory cells
(6) False addressing by NOR type of decoders other malfunctions

Memory circuit malfunctions may cause unexpected floating-body effects, which substantially degrade reliability, speed, power, and radiation hardness. In addition, the entire memory may be rendered non-functional [8.49].

8.16 Destructive read-out characteristics of SRAM

The non-destructive read out characteristics of SRAM do not require restoration, allowing the elimination of a sense amplifier on each data line. To obtain a fast read, the cell signal on the data line is reduced as much as possible, transmitted to the common I/O line through the column pass gate, and amplified. Since the cell signal is developed as a function of the ratio of data line load impedances to the cell transistors, a ratio current flows along the data line during the word line activation time. Here the data line charging current is small due to very small differential voltage. Active current paths are shown in figure 8.30 for SRAMs [8.50 - 8.53]. Total chip standby current increases with memory capacity.

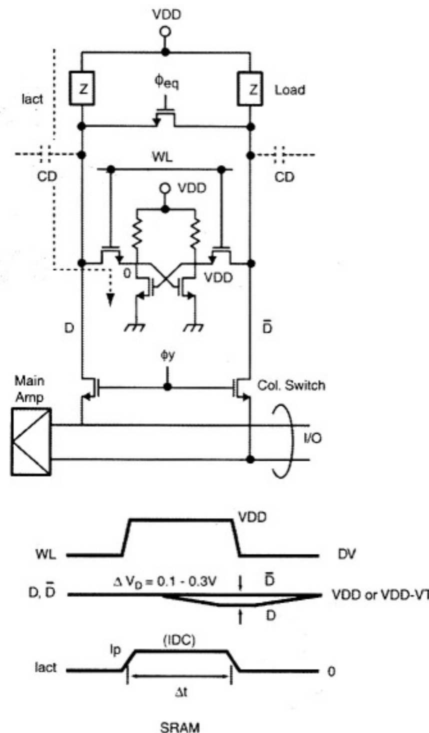


Figure 8.30: Critical path comparison between SRAM and DRAM during READ operation.

References

[8.1] B. Prince, "Semiconductor Memories", Wiley Publishers.

[8.2] B. Bateman, "High Speed SRAM Design", Tutorial, ISSCC 1998.

[8.3] M. Terauchi & K. Terada, IEEE International SOI Conference, 1999, pp. 36 – 37.

[8.4] M. Terauchi, "A novel 4T SRAM cell using self body biased SOI MOSFET structure operating at 0.5V", 2000 IEEE SOI Conference, pp. 108 –109.

[8.5] M.Terauchi, et. al., IEEE Trans. Electron Device, vol 44, 1997, pp. 2303.

[8.6] O. Le Neel & H. Haond, "Electrical transient study of negative resistance in SOI MOS transistors," Electron.Lett., vol. 26, no. 1, 1990, pp. 73 – 74.

[8.7] K. Bernstein & N. J. Rohrer, "SOI Circuit Design Concepts", Kluwer Academic Publishers, January 2000, ISBN 0-7923-7762-1

[8.8] G. Shahidi, et. al., "Partially depleted SOI Technology for digital logic", Proceedings of 1999 ISSCC, Feb 1999, pp. 426 – 427.

[8.9] E. Seevinck, "Static noise margin analysis of MOS SRAM cells",

[8.10] J. Lohsstroh, et. al., " Worst case static noise margin critiera for logic circuits and their mathematical equivalence", IEEE JSSC, Vol SC-18, no.6, Dec 1983, pp. 803 – 807.

[8.11] K.Anami, et. al., "Design considerations of SRAM cell" IEEE JSSC, Vol. SC-18, No.4, Aug 1983, pp. 414 - 418.

[8.12] C.F.Hill, "Noise margin and noise immunity in logic circuits", Microelectronics, Vol.1, Apr 1968, pp. 16 – 21.

[8.13] A. Wei & D. A. Antoniadis, "Measurement of transient effects in SOI DRAM/SRAM access transistors", IEEE Electron Devices, Vol 17, May 1996, pp. 193 – 195.

[8.14] M.M.Pallela et. al., “Low voltage transient bipolar effect induced by dynamic floating body charging in scaled PD/SOI MOSFETs”, IEEE Trans. Electron devices, Vol 17, May 1996, pp. 196 – 198.

[8.15] J.B.Kuang, et. al., “SRAM bitline circuits on PD SOI: Advanatages and concerns”, IEEE JSSC, Vol 32, June 1997, pp. 837 – 844.

[8.16] Chappell, et. al., “A 2ns cycle, 4ns access 512Kb CMOS ECL SRAM”, ISSCC 1991, Vol. 34, pp. 50 – 51.

[8.17] M.Yoshimoto, et. al., “A Divided Word line structure in SRAM and its application to a 64K Full CMOS RAM”, IEEE JSSC, 18, 1983, pp. 479 – 485.

[8.18] D.H.Allen, et. al., “A 0.2um 1.8V SOI 550Mhz 64b PowerPC Microprocessor with copper interconnects”, Proceedings of 1999 IEEE ISSCC, Feb 1999, pp. 438 – 439.

[8.19] A. Karandikar & K. K. Parhi, “Low Power SRAM Design using Hierarchical Divided Bit-Line Approac”, Proc. of the International Conference on Computer Design, 1998

[8.20] J. Kuo & J. Hong Lou, “Low voltage CMOS VLSI circuits”, John Wiley & Sons, ISBN: 0-471-32105-2, 1999.

[8.21] K. Mai, et. al., “Low power SRAM design using half swing pulse mode techniques”, IEEE Journal SSC., Vol 33, No. 11, Nov 1998, pp. 1659 – 1670.

[8.22] T.Tanizaki, et. al., “Practical Low power design architecture for 64Mb DRAM”, ESSCIRC 1997

[8.23] J. Juo & K. W. Su, “CMOS VLSI Engineering – SOI”, Kluwer Academic Publishers, ISBN: 0-7923-8272-2.

[8.24] T. Hirose, et. al., “A 20-ns 4Mb CMOS SRAM with Hierarchical Word Decoding Architecture”, IEEE JSSC 25(5), 1990, pp. 1068 – 1074.

[8.25] J. Uyemura, “CMOS Logic Circuit Design”, Kluwer Publishers. ISBN 0-7923-8452-0

[8.26] K. Bernstein, et. al., "High Speed CMOS Design styles", Kluwer Academic Publishers, 1998, ISBN 0-7923-8220-X.

[8.27] E. Seevinck, et. al., "Current Mode techniques for high speed VLSI circuits with application to current sense amplifier for CMOS SRAM", IEEE Journal SSC, Vol. 26., No. 4, April 1991, pp. 525 – 535.

[8.28] B.Trenbek, et. al., "Characterisation of layout dependent thermal coupling in SOI CMOS Current mirrors", IEEE Trans. Electron Devices, Vol 43, No. 12, Dec 1996

[8.29] A.Pellela, et. al., "A 2ns access 500Mhz 288Kb SRAM Macro", VLSI circuits, 1996, pp. 128 - 129.

[8.30] L. A. Glasser & D. W. Dobberpuhl, "The design and analysis of VLSI circuits", Addison-Wesley, 1988, p288.

[8.31] C.T.Chuang, et. al., "SOI for Digital CMOS VLSI:Design Considerations and Advances", IEEE Proc., Vol 86, No. 4, April 1988.

[8.32] J. Kuang, et. al., "Dynamic body charge modulation for sense amplifiers in partially depleted SOI technology", IEEE SOI Conference 2000.

[8.33] J. B. Kuang, et. al., "A high performance body charge modulated SOI sense amplifier", International SOI Conference 2000, pp. 100 – 101.

[8.34] D. Scepis et. al., "A 0.25um CMOS SOI technology and its application to 4Mb SRAM", IEDM Digest, Dec 1997, pp. 587 - 590.

[8.35] J. B. Kuang, et. al., "A dynamic body discharge technique for SOI circuit applications", IEEE International SOI Conference, 1999, pp. 77 – 78.

[8.36] T. W. Houston, "A novel dynamic V_T circuit configuration", IEEE International SOI Conf. Proc. 1997, pp. 154 – 155.

[8.37] S. Kuge et. al., "SOI DRAM circuit technologies for low power high speed multi giga scale memories", Dig. Tech. Papers, symp. VLSI Circuits, 1995, pp. 103 - 104.

[8.38] J. Gautier, et. al., IEEE IEDM Technical Digest, 1995, pp. 623 – 625.

[8.39] J. A. Mandelman, et. al., "SOI MOSFET mismatches due to floating body effects", Proceedings 1997 IEEE International SOI Conference, Oct 1997, pp. 164 – 165.

[8.40] C. F. Edwards & W. Redman-White, "The effect of body contact serios resistance on SOI CMOS stages", IEEE Electron devices, Vol. 44, Dec. 1997, pp. 2290 – 2294.

[8.41] M. Shoji, "Theory of digital CMOS circuits and circuit failures", Princeton University Press, ISBM 0-0691-08763-6, 1992.

[8.42] K. Sasaki, "High Speed low voltage design for high performance SRAMs", Proc. Tech Papers, VLSI Tech. Systems and Applications, May 1993, pp. 292 – 296.

[8.43] J. P. Colinge, "SOI Technology, Materials to VLSI, 2nd Edition", Kluwer Academic publishers, ISBN: 0-7923-8007-X.

[8.44] C. M. Hsieh, et. al., "A field funneling effect on the collection of alpha particle generated carriers in silicon devices", Electron Devices, April 1981, pp. 104 – 106.

[8.45] S. W. Kerns, et. al., "Model for CMOS/SOI single event vulnerability" IEEE Trans., Nucl. Sci., Vol 36, Dec 1989, pp. 2305 – 2310.

[8.46] H. Iwata, Ohzone, "Numerical analysis of alpha particle induced soft errors in SOI MOS devices", IEEE Trans. Electron Device, Vol 39, May 1992, pp. 1184 – 1190.

[8.47] Y.Tosaka, et. al., "Alpha particle induced soft errors in submicron SOI SRAM" Symp. VLSI Technology, 1995, pp. 39 - 40.

[8.48] T. Karnik, et. al., "Scaling trends of Cosmic Rays induced soft errors in static latches beyond 0.18um", VLSI Symp. Circuits, 2001, pp. 61 - 62.

[8.49] T. P. Haraszti, "CMOS Memory Circuits", Kluwer Academic Publishers, ISBN 0-7923-7950-0

[8.50] J. M. Rabaey & M. Pedram, "Low power design methodologies", Kluwer Academic publishers. ISBN:0-7923-9630-8

[8.51] M. Takada et. al., “Reviews and prospects of SRAM technology”, IEICE Transactions., Vol E74, No. 4, April 1991, pp. 827 - 838.

[8.52] H. Kawamato, et. al., “A 288K CMOS pseudo static RAM”, IEEE JSSC., vol SC-19, Oct 1984, pp. 619 – 623.

[8.53] J. J. J Bastian & W. C. H Gubbels, “The 256k SRAM: An important step on the way to submicron IC technology”, Philips Technical Review, 44, No.2, April 1988, pp. 33.

Chapter 9: SOI DRAMs

9.1 Introduction

Dynamic random access memory (DRAM) is the most widely used semiconductor memory type due to its low cost and high density. Initially only SRAM designs were demonstrated in SOI until high quality of SIMOX material became available, thereafter DRAM chips were also demonstrated. Table 9.1 shows the evolution of SOI memories [9.1 - 9.10].

Table 9.1: Partial list of DRAM memory design evolution, 1993-2001

Year	Company	RAM Type	Ref
2001	Micron	0.15um SOI DRAM	9.1
2000	Mitsubishi	Stressless SOI DRAM	9.2
1999	Samsung	1Mb, 0.3um, SOI DRAM	9.3
1999	IBM	1Gbit DDR SDRAM, 2.5V	9.4
1998	Mitsubishi	Multi Gbit PF Nand DRAM	9.5
1997	Hyundai	1Gb DRAM, 2V	9.6
1996	Mitsubushi	16Mb DRAM, 0.9V	9.7,8
1995	Samsung	16Mb DRAM	9.9
1993	Mitsubishi	64Kb DRAM, 1.5V	9.10

SOI DRAMs have several advantages over bulk DRAMs:

a) Reduction of the bit-line capacitance (by 25% compared to bulk)
b) Reduction of the access transistor leakage current
c) Reduction of soft error sensitivity.

Storage capacitance in SOI is around 55% lower than in bulk. Storage capacitance of 12fF is sufficient for a 256Mb, 1.5V DRAM in SOI, while 34fF is required for the equivalent bulk device making SOI DRAMs more attractive [9.11, 9.12]. For DRAMs to operate below 1V, a reduced bit line to storage capacitance must be achieved. Reduction of bit line capacitance is a natural convergence of the transfer to SOI, while increase in storage capacitance is achieved through cell design. Thin-film SOI MOSFET's have much lower leakage than bulk or thicker SOI devices. Off-state leakage lower than 1fA/μm is observed in full depleted SOI. This is important as retention time is a function of storage capacitance and leakage. The lower the leakage, the longer the retention time.

DRAM architectures frequently take advantage of the low body effect of SOI devices. Bulk pass transistors suffer from body effect during the storage capacitor charging state. The lost charge is compensated by boosting the word line voltage to $V_{WL}=V_T+V_{DS}+\Delta V$, where V_T is the threshold voltage of the MOS device when the body voltage equals source voltage, V_{DS}, the drain to source voltage and ΔV is the differential voltage which compensates for the threshold voltage increase in the bulk transistor design. The additional ΔV required is a critical design issue in bulk passgate transistors [9.13]. The use of fully depleted SOI significantly improves the performance of pass gates because of low body effect found in SOI devices.

The body potential of a partially depleted SOI transistor follows the source node potential during charging, in order to maintain the body to source potential difference constant. Figure 9.1 shows the charging efficiency [9.14] in DRAMs for bulk, FD and PD-SOI. SOI transistors have a higher charging efficiency than bulk, resulting in better programming speed for the DRAM cell. In order for a bulk transistor to achieve similar charging efficiency as a SOI transistor, the word line voltage has to be increased during charging.

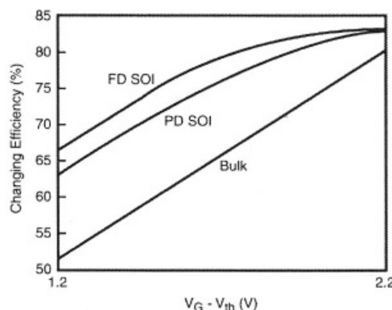


Figure 9.1: Charging efficiency vs. gate voltage overdrive.

9.2 DRAM structure and Operation

Figure 9.2 shows the architecture of a DRAM chip, with data and control signal paths. DRAM signal timing during read operation is shown in figure 9.3. Row and column addresses decoded from the input determine the X and Y co-ordinates of the memory cell to be addressed. These addresses are time multiplexed to minimize the number of address pins. The *RAS* (Row address strobe) and *CAS* (column address strobe) are typically active low signals. The row address is input first, and is latched by the falling edge of the $\overline{RAS}$, then the column address is latched by the falling edge of $\overline{CAS}$. $\overline{CAS}$ also enables the output buffer. Row address selects a word line, which connects a memory cell to the bit line, establishing a signal on the bit line. The sense amplifier detects the signal and the column decoder connects it to an I/O line. The amplified signal from the sense amplifier is transferred to the data output buffer and subsequently data appears at the output at the falling edge of $\overline{\text{CAS}}$. Time from the falling edge of $\overline{CAS}$ to data out is called *CAS access time* and the time from the falling edge of *RAS* to data out is called the *RAS access time.* During the *write* operation, the data paths of the I/O lines are reversed. The row path, which selects the word lines are the same as for the read operation.

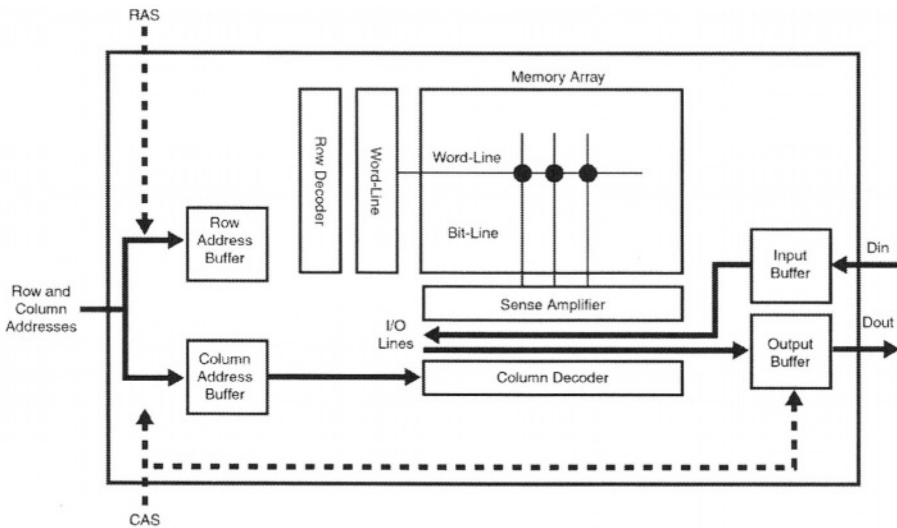


Figure 9.2: Architectural block diagram of a DRAM critical path [9.15]

Both read and write operations use the same path. However, the write operation bypasses the sense amplifier. Data to be written to the memory cell passes through the input buffers and onto the data bus. Signals are transferred to the bit-lines through a write select pass gate and are written to

the memory cell by activating the word lines. The write operation is signaled by $\overline{WE}$ low and is high for 'READ'. Input and output buffer circuitry provide ESD protection by clamping the highly capacitive I/O lines. These are considered critical to avoid undershoots and overshoots on the I/O signal.

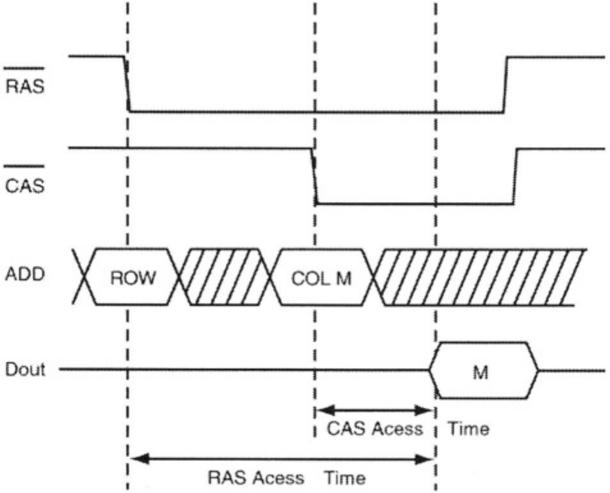


Figure 9.3: Signal timing in DRAMs.

Once a word line is activated, all sense amplifiers associated with the memory cells connected to that word line are activated regardless of the data line connected to the I/O line. Depending on the memory density and division of blocks in the array, the number of sense amplifiers activated simultaneously varies depending on the number of banks and sub-blocks within a bank. Data is available on the sense amplifier outputs and transferred to the I/O lines with a select signal. Different architectures may be used to minimize the number of active sense amplifiers and cells in a sub-block [9.15].

9.3 Memory Array

The simplest array architecture forms a memory cell wherever a word line and bit line cross each other. Since a pair of bit lines must be connected to a sense amplifier for data comparison, this architecture results in an open bit line configuration (figure 9.4). When one of the word lines on the right side is selected, none of the word lines on the left side is selected. The left-side bitline with no connected memory cell serves as a reference for the right side bit line. There are disadvantages with the open bit line configuration. The sense amplifier circuit layout must fit into the narrow pitch of a cell bit line (sum of line width and spacing). This is difficult, because memory cell size

and cell pitch decrease with successive generations of DRAM. A solution is to use a relaxed sense amplifier pitch open bit line architecture (figure 9.4(b)). Open bit line architecture also exhibits poor noise immunity. Localized noise in the array couples onto one side of the bit line pairs and degrades the sense amplifier differential input signal.

Localized noise coupling in the array may be avoided by the use of a folded bit line approach (figure 9.4(c)). *BL* and $\overline{BL}$ extend parallel from the sense amplifier. Localized noise couples onto both bit lines and is canceled by the differential operation of the sense amplifier. Thus, the sense amplifier can detect a smaller differential signal from the memory cell in a folded bit line configuration [9.15].

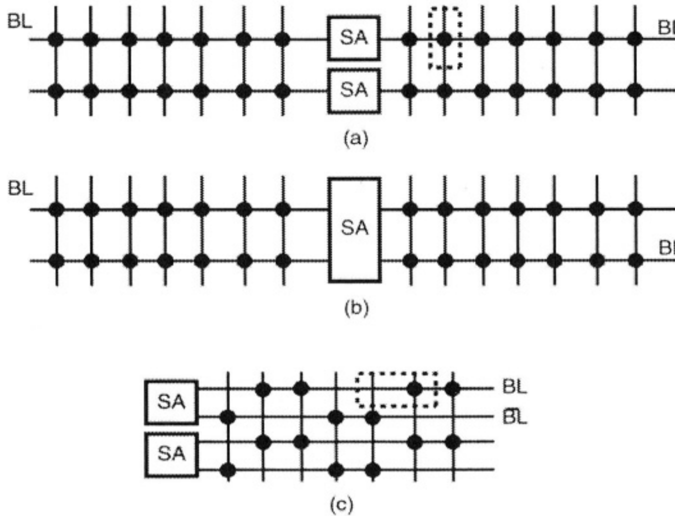


Figure 9.4: Bit Line configurations in DRAMs [9.16] (a) Open Bit Line (b) Relaxed sense amplifier pitch open bit line (c) Folded Bit line. The dotted regions represent the memory cell.

A memory cell is formed every time the word line crosses the bit line in the open bit line architecture,. Conversely, in the folded bitline architecture, each memory cell unit contains a word line that passes through, and a word line that is used by the DRAM cell (figure 9.5(a)). It is unusual for the bit density in the folded BL architecture to be as high as in open bit lines architecture.

Memory cell size is usually measured as a multiple of the feature size, F. The minimum possible cell area in open bitline architectures is $2F*2F = 4F^2$

since the word line and bit line pitches are both 2F (F for a line and F for space). In the folded bit line architecture, due to the existence of passing word line, the memory cell dimension is 4F in the direction parallel to bit line. Therefore, the minimum possible memory cell size is 2F x 4F = $8F^2$. The $8F^2$ limit places a severe barrier to the scaling of the DRAM cell and chip size. Various architectures may achieve lower than $8F^2$ limit without sacrificing the benefits of folded bit line sensing. Figure 9.5(c) shows how this can be achieved [9.17]. Here, only one passing word line exists between two active word lines compared to two passing word lines in conventional folded bit line configuration. Minimum memory cell size is therefore, $6F^2$. The sensing for this array requires combining folded and open bit line configurations. This may require an additional transistor between the sense amplifier and bit lines to select the bit line connection appriopriate to the selected word line. Another technique to attain $6F^2$ in a folded bit line configuration is to run the bit line diagonally rather than horizontally [9.18].

A circuit technique has been proposed for folded bit line sensing requiring only a $4F^2$ memory cell, the so-called cross point cell. This configuration uses a switch similar to $6F^2$ folded/open sensing approach [9.19]. The use of switches, however, requires additional timing and better timing control, which is difficult to maintain as DRAM access time decreases [9.15].

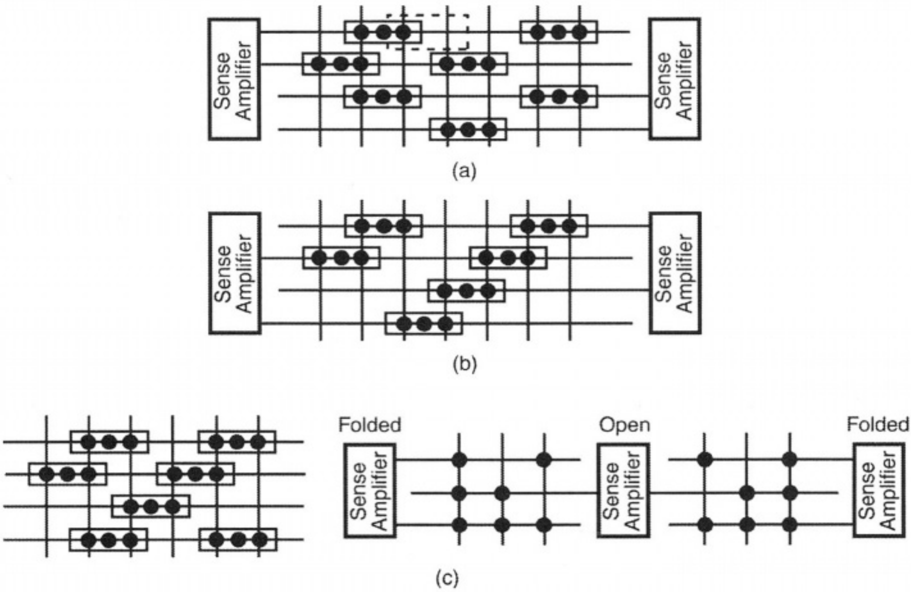


Figure 9.5: (a) Half Pitch Array architecture (b) Quarter pitch array architecture (c) $6F^2$ Cell array architecture.

9.4 Dynamic Random Access Memory cell storage

A variety of dynamic RAM cells are shown in figure 9.6. Most of these use a capacitor as the storage element, although gate/source capacitance is used to store charge in the circuits shown in figures 9.6(a) and (e). Stored charge tends to leak away with time. As a result, data must be refreshed periodically with the original '0' or '1' level stored in the cell [9.20].

The evolution of these cells began with the four-transistor dynamic memory cell (figure 9.6(a)). This cell is derived from the six-transistor static memory cell by removing the load devices. Since current is no longer supplied to the storage nodes to replace that lost to leakage, the cell must be periodically refreshed. This cell is activated using a clock pulse to the two access transistors. Since the charge eventually leaks off, the cell must be re-activated to refresh the data. Replacing the nMOS access passgates shown in figure 9.6(a) with pMOS results in the 4T loadless SRAM cell.

The four-transistor cell is compatible with MOS logic circuitry since it contains only transistors. It is simply used as an embedded memory cell in logic if refreshed frequently. A clock running in the system for other purposes can be used to provide the refresh timing. This type of DRAM cell is not power and area efficient compared to the one-transistor DRAM cell.

Figure 9.6(f) shows the conventional one transistor dynamic memory cell, which consists of a transistor and storage capacitor. The gate of the transistor is controlled by a word-line. When the word line is enabled the select transistor turns 'on' and the charge stored is transferred onto the bit lines and then to a sense amplifier.

The sense amplifier compares it with a reference cell and determines if the DRAM cell contains a stored '1' or '0'. The read out from the cell discharges the capacitor so that the data is no longer stored in the cell. Before the read operation is complete, the sense amplifier circuit must restore the original charge to the cell.

Although area efficient, the major problems of the 1T DRAM cell have been to ensure sufficient charge storage in the cell and designing small sense amplifiers to fit into the area between the bit-lines. The presence of buried oxide in SOI reduces the susceptibility of the DRAM cell to SER.

Figure 9.7 shows the cross-section of a DRAM cell [9.21] and the growth of epitaxial layer in which the transistor is stacked horizontally over trench.

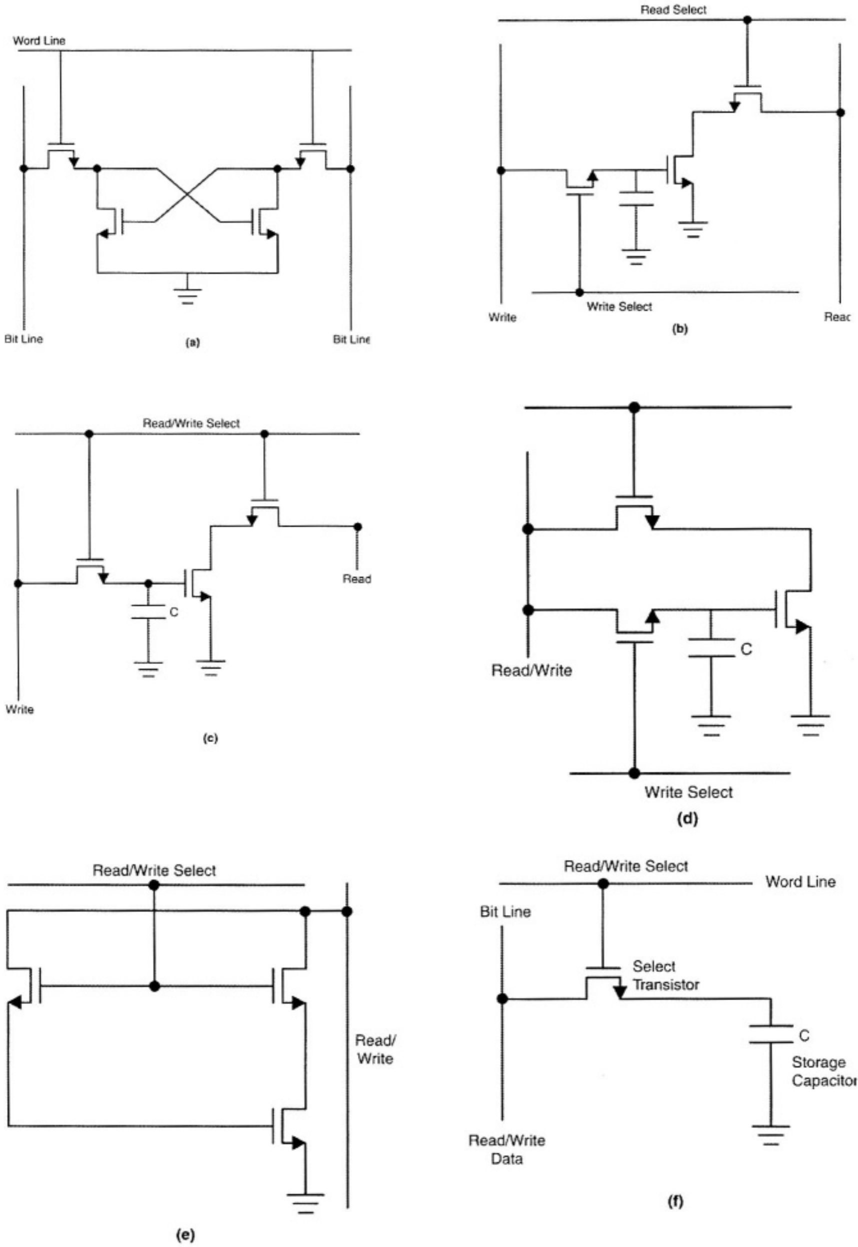


Figure 9.6: Historical Evolution of DRAM cell (a) Basic Bistable flip flop without load devices (b)2X-2Y (c) 1X-2Y(d) 2X-1Y(e)1X-1Y(f) 1X-1Y, where X is the number of control lines and Y the number of I/O lines.

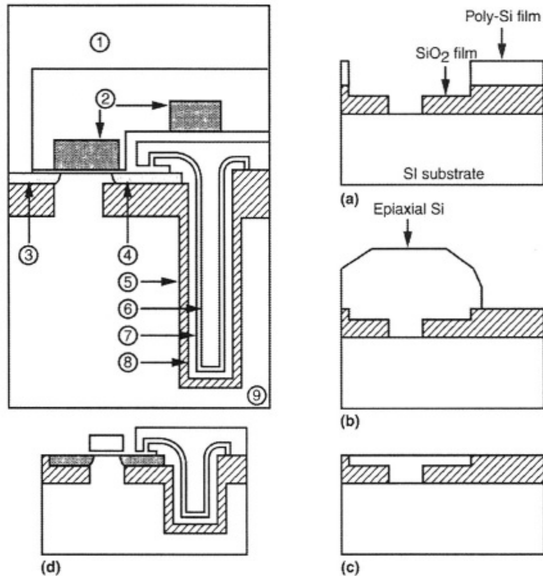


Figure 9.7: DRAM cross-section: 1. Bit Line, 2. Word line, 3. Drain Area, 4. Source Area, 5. SiO2 film, 6. capacitor area, 7. capacitor insulator film, 8. charge storage electrode, 9. Si Substrate [9.21].

9.4.1 Storage to Bit Line Capacitance

The storage capacitance to bit line capacitance ratio is one of the most important characteristics of the 1T DRAM. Sufficient charge must be stored on the capacitor to provide a readable signal to the sense amplifier.

The parasitic bit line capacitance (C_b), is typically larger than the capacitance of the storage cell (C_s) due to a large number of cells present on a single bit-line. When the cell is selected and the signal stored in the cell capacitor (V_s) is read out onto the bit line, it is reduced by the ratio of the storage capacitance to the bit line parasitic capacitance. The signal magnitude on the bit lines (V_b) is, therefore, frequently described by the ratio, i.e., $C_s/C_b = V_b/V_s$, or $V_b = V_s\,(C_s/C_b)$. The bit line capacitance includes the junction and the source/drain capacitance of all the inactive cells on the same column. Fortunately, for SOI technology, the reduction of junction capacitance and the dielectric isolation help achieve a lower bit line capacitance. Hence, the ratio of the parasitic bit line capacitance to the storage capacitance (C_b/C_s) in SOI will be lower, aiding in charge storage. The signal stored in the cell capacitor (V_s) will be closer to V_b in SOI technology compared to bulk.

The bit line capacitance C_b/C_s is determined by technology. The total value of C_b is strongly influenced by the number of storage cells on a bit line. C_b degrades the storage cell signal and hence the bitline signal has to be amplified by a sense amplifier, which must also be able to refresh the information of the selected cell. The more sensitive the amplifier, the smaller the storage capacitance. If the bit line capacitance is large relative to the capacitance of the memory cell then, the signal across the bit lines and the signal available to the sense amplifier will be too small to be read. For example, if the C_b to C_s is 10, and the voltage stored on the storage capacitor has decayed to 2V at the time the cell is read, the signal which is read out onto the bit line will be reduced by 1/10 to 200mV. The sense amplifier design is critical to DRAMs and must be sensitive enough to read this signal. Mismatches in the devices have to be taken into consideration to obtain efficient, fast and reliable sense amplifiers. The C_b/C_s ratio for SOI is less than bulk due to reduced junction capacitance in SOI. Signal degradation due to bit line capacitance, and the parasitic source/drain leakage from the unaccessed cells is less than bulk [9.20].

9.5 SOI DRAM Process

Low parasitic capacitance and good radiation hardness properties makes SOI ideal for integrating DRAMs. However, SOI DRAM processing is slightly different from SOI CMOS. Figure 9.8 shows the processing sequence of SOI DRAM technology for a memory cell [9.22].

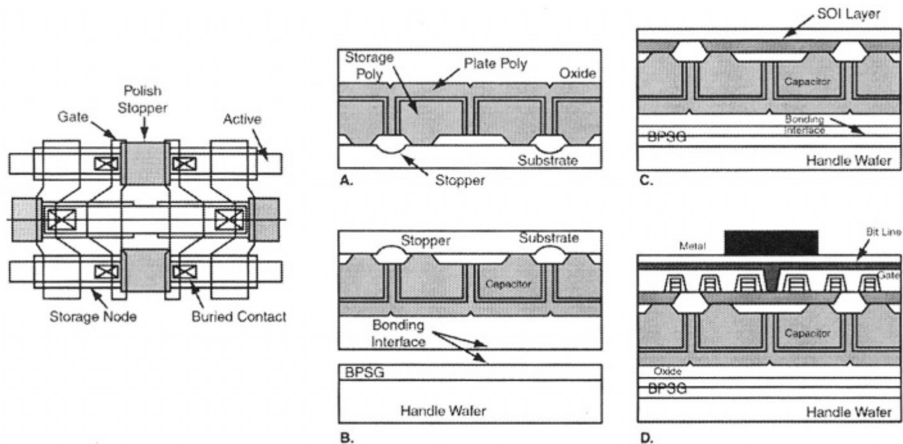


Figure 9.8: Process Sequence of an SOI DRAM technology and memory cell layout [9.23].

Localized field oxidation is followed by a thermal oxidation step resulting in buried oxide. Subsequently, contact etching, polysilicon deposition, and a

further thermal oxidation form the storage capacitor with a layer of oxide. The surface is then planarized. Using BESOI (Back Etched SOI) technology, the device wafer is bonded with the handle wafer, which is covered with BPSG (Boro-Phoso-Silicate Glass). The redundant silicon layer on the top of the device wafer is removed by grinding and chemical mechanical polishing (CMP). The field oxide is used after polishing and only the thin film remains after CMP. Shallow trench isolation (STI) can be used to isolate thin film regions. Poly-gate source/drain implant and metallization as for a regular CMOS fabrication procedure are performed subsequently [9.23]. In this SOI DRAM process, after building the storage capacitor, wafer bonding is initiated followed by fabrication of the MOS transistors.

An alternative method to produce SOI DRAMs involves not initiating wafer bonding until the storage capacitance and MOS transistors are fabricated (figure 9.9). Subsequently, a handle wafer is bonded [9.24]. The redundant silicon on the device wafer is removed by grinding and chemical mechanical polishing (CMP) and the bonded wafers are covered with an oxide layer for metallization.

The advantages of this SOI DRAM technology are that bit lines are planarized. In addition, bit lines and word lines (poly-gate) are located at the top and the bottom of the thin film active device region. Hence, the distance between bit line and source/drain region is smaller than for the conventional structure. This improves device density. Low cost glass substrates can be used as the handle wafer since no high temperature steps are required after wafer bonding.

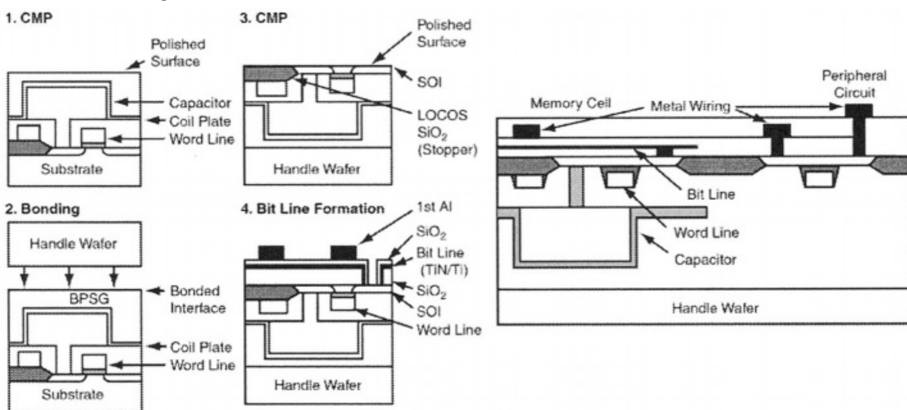


Figure 9.9: Alternative method to fabricate SOI DRAM's is without initiating wafer bonding until the storage capacitance and MOS transistors are fabricated.

9.5.1 Smart-cut for DRAM

Smart cut technology has been discussed for SOI in chapter 2 [9.25]. We here discuss the technology for use with DRAMs. Figure 9.10 shows the manufacturing steps involved with smart-cut SOI technology. After the storage capacitor and the MOS devices are built, a hydrogen implant is initiated followed by wafer bonding and wafer splitting. Grinding and CMP processing times can be substantially reduced by using smart cut SOI technology and the split wafer can be re-used.

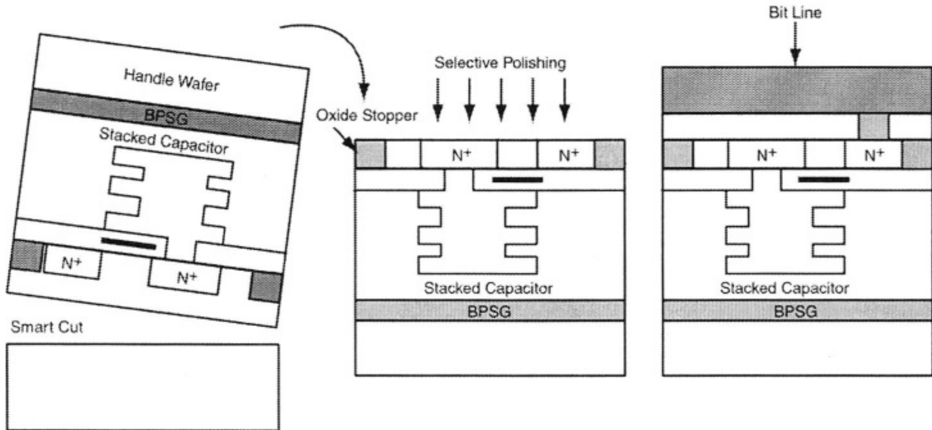


Figure 9.10: Application of smart cut SOI technology to realize the DRAM buried capacitor structure.

9.5.2 Quasi- SOI technology

Quasi-SOI technology has also been demonstrated for integration of DRAM. A thermal oxidation procedure is carried out (figure 9.11), followed by a reactive ion etch to generate the pedestal area and seed hole in the oxide [9.26]. Using an epitaxial overgrowth technique, the seed hole and pedestal area are filled with single crystal silicon. A polishing step removes the redundant silicon layer. The process is completed after building the storage capacitor and the MOS device in thin-film silicon.

9.6 Influence Of SER on SOI DRAMs

DRAM density constrains the cell storage capacitance required to maintain SER within acceptable levels. It also affects static and dynamic retention time, and sense amplifier sensitivity. Compared to bulk, SOI DRAM's have superior SER and static data retention time, inferring higher integration density may be possible. Figure 9.12 [9.27] shows the bit line sense signals plotted against cell storage capacitance for both bulk-Si and SOI DRAM 's. SER, static and dynamic data retention are major limitations for 256Mb

DRAM and beyond. The bit line capacitance C_b for SOI DRAM is approximately 25% smaller than that of a bulk-DRAM. In bulk DRAM, a sense amplifier sensitivity limit of 30mV requires cell storage capacitance higher than 4.1fF. The data retention requirement requires that the mean retention time be longer than 5s at 80°C placing a lower bound of 24fF on the cell storage capacitance. Efficient SER requirement of less than 100 FIT (1 FIT = 10e-09 failures per hour) dictates Q_{CRIT} be larger than 100fC, and a cell storage capacitance larger than 67fF at 1.5V. Therefore in bulk DRAM, the density is limited by SER [9.28, 9.29].

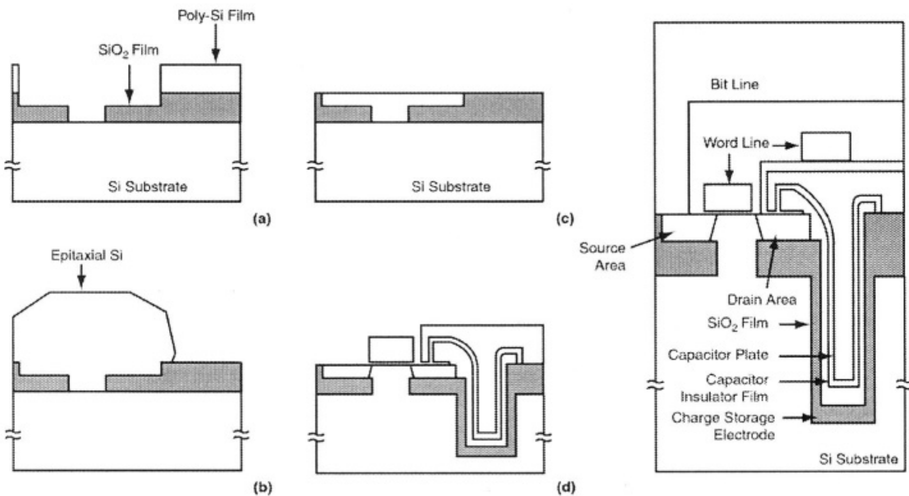


Figure 9.11: Transistor on a lateral epitaxial silicon layer.

The limit resulting from sense amplifier sensitivity is 3.1fF in SOI DRAM's which is due to reduced bit line capacitance. Static data retention limit on the cell storage capacitance is 4.5fF (about 20% that for bulk). This is due to smaller p-n junction leakage area. Experimentally, SOI DRAM's have been found to be essentially soft-error free, permitting the cell storage capacitance limit (on a SOI DRAM) to be set by the static data retention time requirement [9.30].

9.7 Cosmic Ray induced Soft Errors in SOI DRAMs

SOI is expected to reduce cosmic ray induced soft errors in high density DRAM's. Cosmic rays interact with the atmosphere generating neutrons. There is a probability that a neutron passing through silicon will interact with a nuclei. Recoiling heavy ions generate a large number of electron hole pairs over a short path. The resulting events while few in number, have a high probability of causing an error [9.31 - 9.33]. It is difficult to observe a neutron event due to the small neutron-nucleus cross-section for silicon.

An experimental approach to studying neutron-like induced SER is to directly strike devices with heavy ions that produce a charge track similar to that of the silicon ions recoiling from neutrons. Fluorine ions have been used for this purpose. These produce a track charge density greater than half that of silicon ions and about 10x that of α-particles (figure 9.13). Diodes have been used to collect induced charge. Table 9.2 indicates that for a fluorine ion strike, the charge collected in the SOI structure is more than an order of magnitude lower than bulk. The difference is primarily due to silicon film thickness, and shows the advantage of the SOI structure in avoiding cosmic ray neutron induced soft errors [9.28, 9.34].

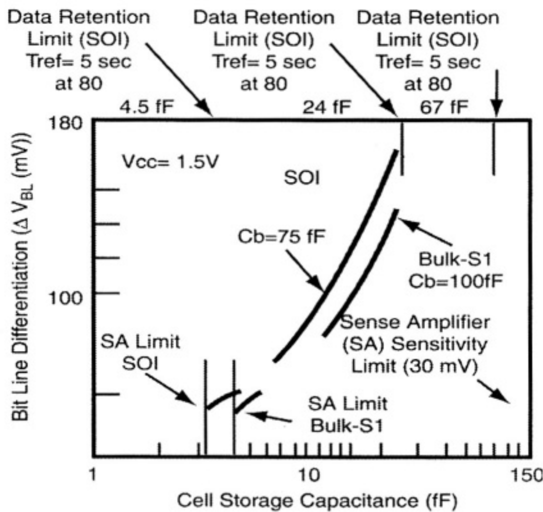


Figure 9.12: *Limitation of SOI DRAM and Bulk-Si DRAM [9.28]. Data retention limit refers to the static data retention characteristic limit [9.27].*

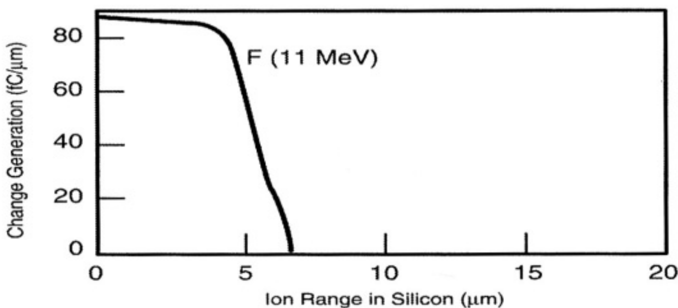


Figure 9.13: Charge generation for ions travelling into the silicon surface.

Table 9.2: Charge collection for α-particles for fluorine ion strikes of 11MeV energy.

Diode Size	Collected Charge (fC) SOI	Bulk
2 × 2 mm	6.7	81
5 × 5 mm	7.7	107
10 × 10 mm	9.0	140
20 × 20 mm	8.8	155

9.8 DRAM Refresh and Data Retention

A logic high voltage stored on a memory cell slowly degrades to logic low due to various leakages. Some of the leakage mechanisms and methods to reduce them have been discussed in chapter 8. A dynamic RAM can only retain stored information by periodically reading and rewriting the data (refresh) before the data is corrupted. The cell refresh time requirement has increased from 2mS for 64K DRAM to 128mS for 64Mb DRAM. The SOI body is tied down in circuits like the sense amplifier to avoid variability in body voltage and contribute more leakage, and to ensure accurate sensing. The cell storage capacitances play an important role in determining refresh time. The higher the capacitance, the lower the refresh time and vice versa. There are two types of data retention techniques, (1) static data retention (2) dynamic data retention.

9.8.1 Static Data Retention

Some common leakage mechanisms that affect charge storage in static data retention schemes or circuits are:

- Junction Leakage
- Pass transistor sub-threshold leakage
- Leakage through capacitor dielectric
- Other parasitic leakage paths

Junction Leakage

The p-n junction that makes the storage capacitor in the DRAM memory cell is always reverse biased or biased at ground. Leakage current through a reverse biased P-N junction can be expressed by the sum of the diffusion and the generation-recombination current [9.34] as:

$$J_R = qD_P\, p_{no}/L_P + qD_N\, n_{p0}/L_N + qn_I W/t_e$$

The first two components are the diffusion current, and the third term is the generation-recombination current. Here, D_P and D_N are the diffusion coefficients of holes and electrons, respectively; and L_P, L_N are the diffusion length of holes and electrons respectively. p_{no} and n_{po} are the equilibrium hole density on the 'n' side and electron density on the 'p' side; n_i the intrinsic carrier concentration.. W is the depletion width and t_e is the effective carrier lifetime in the depletion region. The generation-recombination current is dominant in most cases, however, at zero bias conditions and high temperatures, diffusion current is higher. Diffusion current typically has an activation energy of E_g, the band-gap energy of silicon, whereas the generation-recombination current has an activation energy of about $E_g/2$.

Thermionic field emission of electrons from deep levels in the depletion region is another leakage mechanism that impacts DRAM retention [9.35]. This current can be distinguished from diffusion and generation-recombination current by its dependence on the well doping concentration since both diffusion and generation-recombination current decrease with the doping concentration.

Leakage through the Capacitor Dielectric

Although a capacitor dielectric is normally considered an insulating film, it conducts some current especially at high bias voltage or when very thin. When the bias voltage exceeds the breakdown voltage of the dielectric film, the film becomes conductive and leakage current increases significantly. Below the breakdown voltage, leakage is either Frenkel-Poole or Fowler-Nordheim. Frenkel-Poole leakage is due to field enhanced thermal excitation of trapped electrons into the conduction band, whereas, the Fowler-Nordheim current is caused by electron tunneling [9.15, 9.34].

The static data retention limit occurs during refresh when the bit line of the unselected cell is precharged to $V_{DD}/2$, and is primarily due to junction leakage. Leakage mechanisms are shown in figure 9.14 for bulk and SOI DRAM cells [9.30]. Due to significantly reduced p-n junction area in SOI DRAMs, leakage is typically an order of magnitude smaller than in bulk-. Static data retention time has been experimentally verified to improve with silicon film thickness, and reduced p-n junction area [9.36].

9.8.2 Dynamic Data Retention

Dynamic data retention characteristics impose limitations on design. Here, the bit line of a selected cell swings from high to low ($V_{DD}/2$ to GND), increasing gate-to-source voltage of the cell transistor, thus causing sub-

threshold leakage [9.30]. In SOI DRAMs, holes are injected into the floating body. Body potential rises as a result of hole injection/redistribution and capacitive coupling, reducing the threshold voltage and degrading sub-threshold leakage. A particularly severe case is when dynamic data retention occurs after a long static data retention. The leakage current charges the body over a period of time, and enables dynamic data retention [9.37]. Poor dynamic retention characteristics have been experimentally observed and pose a major design challenge for SOI DRAM's [9.30, 9.38, 9.39]. The use of body contact in the DRAM cell is not an option due to tighter cell density, thus process/device modifications and/or circuit techniques are required.

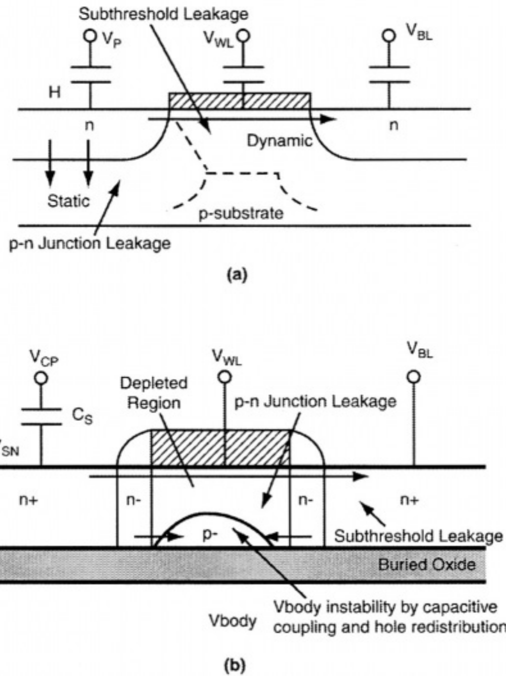


Figure 9.14: Leakage Mechanism for (a) bulk-Si DRAM cell and (b) SOI-DRAM cell.

Process/device modifications such as adding a lightly doped source/drain region (to reduce both current gain of the parasitic bipolar transistor and impact ionization near the drain) and the use of pMOS cell transistor (lower impact ionization rate and parasitic current gain) [9.38] are options. An effective technique is use of boosted sense ground (BSG) scheme [9.40-9.42], where a low bit line level is raised above the unselected word line level to suppress sub-threshold leakage. BSG improves dynamic retention time in a SOI DRAM. Figure 9.15 shows the simulated data retention at room temperature as a function of junction leakage.

Dynamic retention time can be further improved with a 'body refresh' scheme [9.37]. Here the bit line level is lowered momentarily (from $V_{DD}/2$ to ground in the BSG scheme) to remove accumulated holes from the body region to the bit line through the forward biased body source p-n junction and decreases the body potential. In the static retention condition (figure 9.16(a)), the body potential increases due to the junction leakage. It is crucial in this scheme to ensure that the body refresh/discharge current does not destroy the stored data in the cell [9.30].

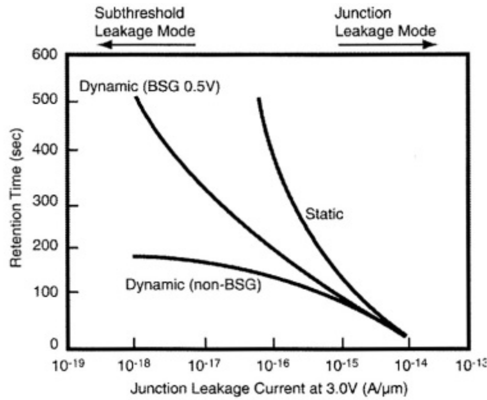


Figure 9.15: Data-Retention Time as a function of junction leakage current

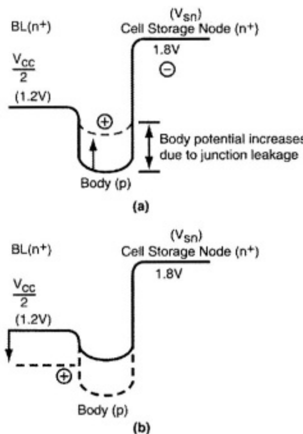


Figure 9.16: Potential of SOI DRAM (a) Static Retention Condition (b) Dynamic Retention Condition.

In the example described the voltage drop of the stored high data is only 40mV, thus alleviating concerns of degradation. The body refresh function is embedded in the normal DRAM operations and timings.

Figure 9.17(a) shows the dynamic data-retention characteristics after writing a 'high' into the SOI DRAM cell. Figure 9.17(b) shows the cross section.

After a few operations the body refresh maintains sub-threshold leakage at a low level improving dynamic data retention time. The dynamic retention characteristics for worst case scenario is when the dynamic data retention occurs after a long static data retention (figure 9.18). Body refresh prior to dynamic data retention suppresses sub-threshold leakage, improving data-retention time. If the body refresh operations (before dynamic data retention and during dynamic data retention) are combined, a factor of approximately three can be obtained in retention characteristics [9.43]. Floating body induced degradation in dynamic retention can be reduced by modulating threshold voltages in order to reduce sub-threshold leakage and trade-off cell performance.

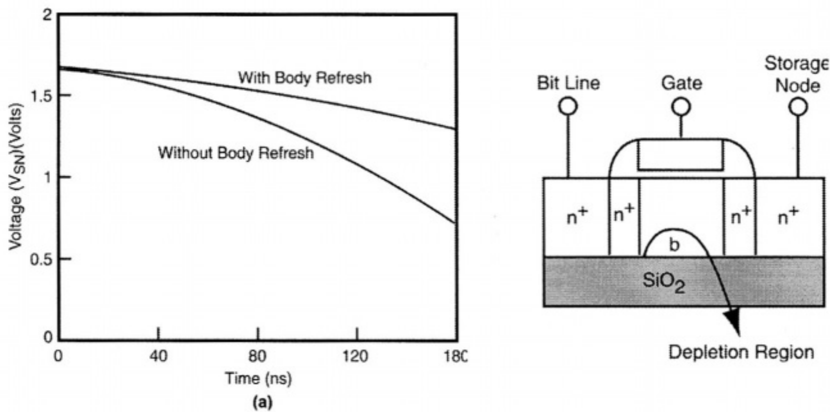


Figure 9.17:(a) Estimated dynamic data retention for SOI DRAM (b) cross-section.

The degradation in dynamic retention time induced by floating body effects, can be alleviated by raising the MOS threshold voltage to reduce sub-threshold leakage at the expense of cell performance. An alternative is to create a leaky diode between body and source. This lowers current gain of the parasitic bipolar transistor.

A further option is to reduce drain-body coupling to lower the floating body voltage [9.44]. Most of these options require transistor level design and process window tradeoffs, especially for the leaky body source junction,

where the junction leakage must remain substantially lower than the sub-threshold leakage of the MOSFET.

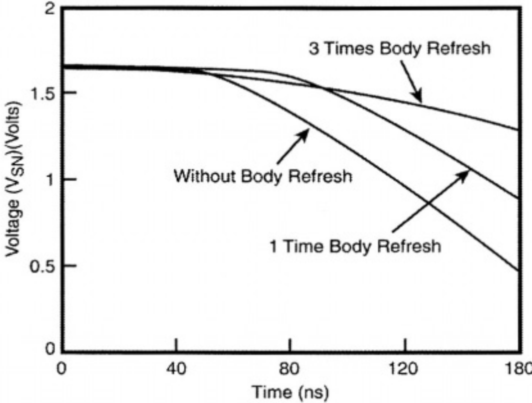


Figure 9.18: Data retention for SOI DRAMs. Storage node voltage for different body refresh operations. Worst case (dynamic data retention after long static data retention).

9.8.3 Parasitic Leakage in DRAMs

Parasitic leakage current due to floating body effect in a partially depleted DRAM affects performance. Figure 9.19 shows transient leakage mechanism in the SOI DRAM array [9.45]. There are three leakage mechanisms in a DRAM cell that can degrade the data stored in the cell.

(1) During write '1', the memory cell pass-gates are turned ON pulling the storage node high (V_{DD}).

(2) Cells on the same column are not accessed in the latency period (figure 9.19(a)). Due to diode leakage current and thermal generation effects in the body, a substantial amount of hole accumulation occurs, which effectively increases body potential. The body voltage rises from the initial value to approximately supply voltage over a 100mS period.

(3) If a cell on a column is accessed and 'write 0' is initiated, the bit line will be pulled low (figure 9.19(b)). Thus, although the pass transistor of this un-accessed cell is OFF, the sudden drop in the source voltage may turn-on the body-to-source diode and parasitic BJT. The body voltage of the nMOS pass gate when the bit lines are low is in this example 0.67V (in steady state conditions). Consequently, a large leakage current flows between drain and source. This continues until the holes in the body region are depleted and the body potential decreases causing a change in the storage node potential and

disturbing the bit-lines. If the latency period is too long, sub-threshold current of the pass transistor increases due to the rising body voltage and degrades the data retention time.

The leakage current issue of SOI DRAM can be improved by using PD-SOI CMOS with a non-uniform doping profile. If the bottom of the silicon film is more heavily doped, the current gain of the parasitic BJT is reduced.

Figure 9.20 shows current gain versus thin film doping density of a SOI MOS device with uniform and non-uniform P/P+ doping structures used in the SOI DRAM memory cell [9.46]. This indicates that a (figure 9.51) non-uniform doping sructure provides an effective means to reduce parasitic bipolar current gain.

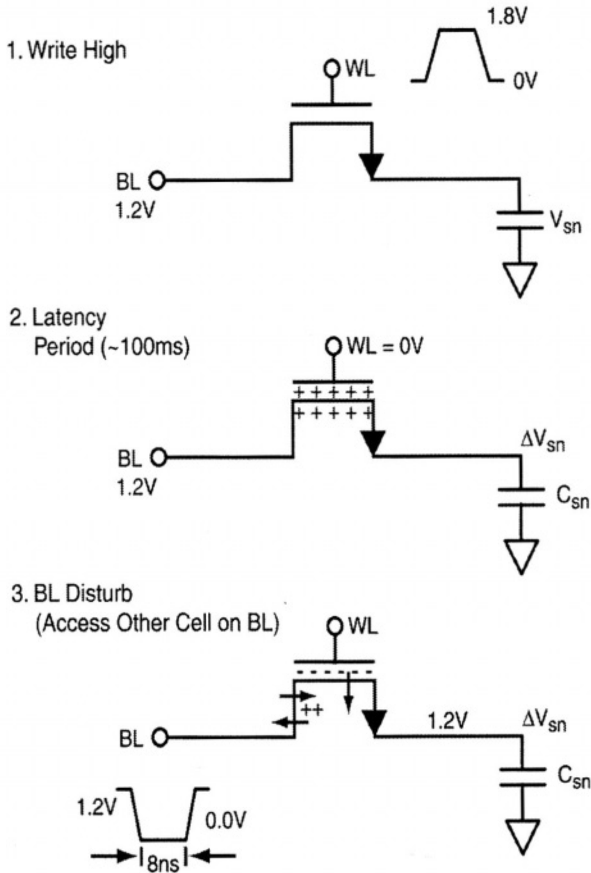


Figure 9.19(a): Schematic of a SOI DRAM cell under a variety of operating conditions.

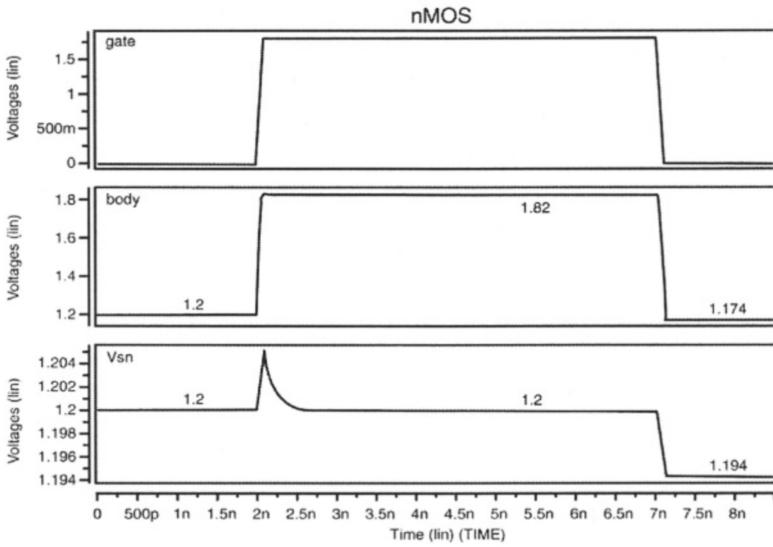


Figure 9.19(b) Simulation waveforms showing the storage node V_{SN} decaying from 1.2v to 1.194v over a period of 8nS after writing a "1".

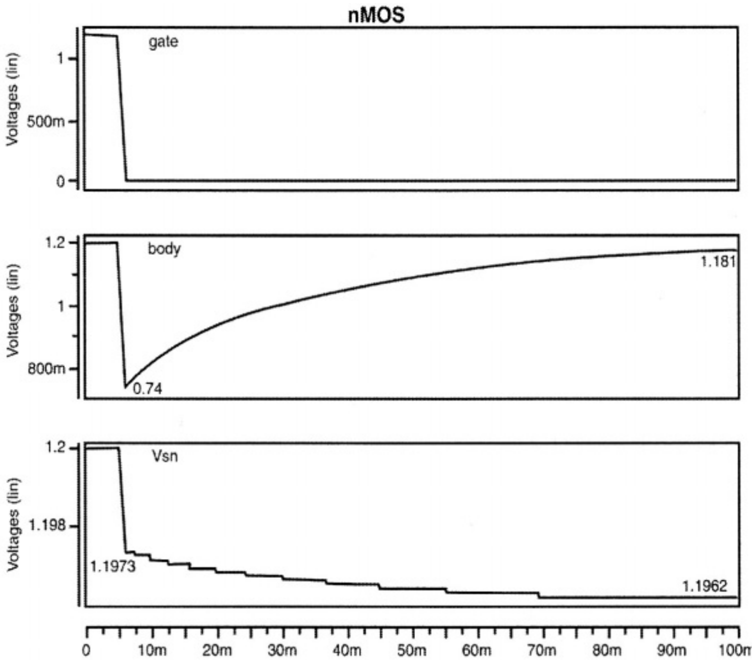


Figure 9.19(c): Latency Period 100mS, Gate is the WL.

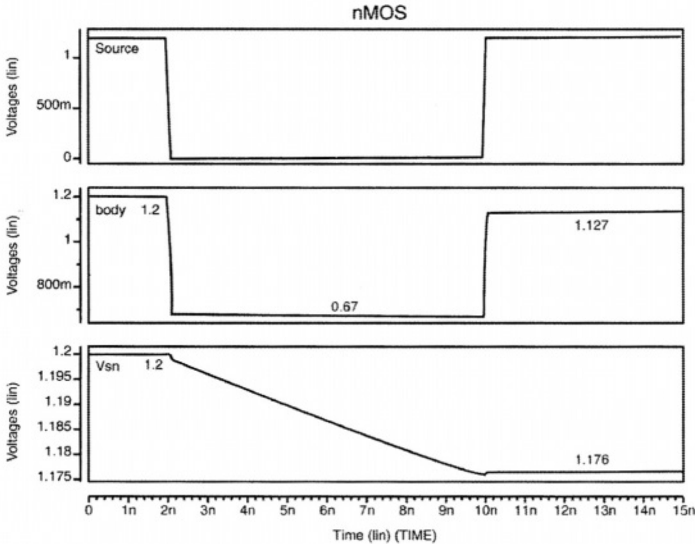


Figure 9.19(d): BL Disturb simulations of a SOI DRAM cell. Illustration of transient leakage mechanism in SOI DRAM

Leakage problems in a SOI DRAM memory cell can also be reduced using a body refresh operation (as discussed previously for dynamic data retention). For example if the body potential of a pass transistor in a SOI DRAM memory cell does not reach a designated value, the bit lines are pulled low intentionally to deplete from the body [9.47].

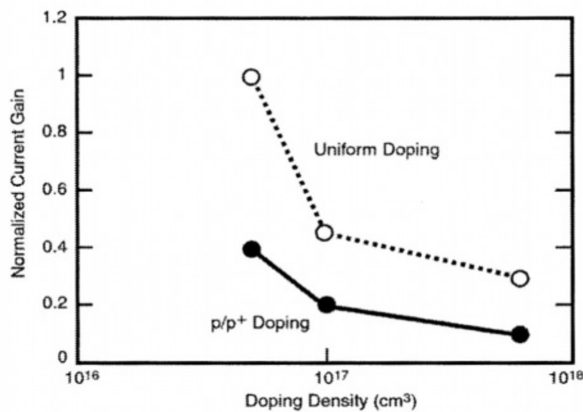


Figure 9.20: Current gain versus thin film doping density of the SOI MOS device with uniform and non-uniform p/p+ doping structure used in DRAM memory cell [9.46].

Leakage current can be suppressed using body contacts (chapter 3) [9.48]. One method to achieve this is use of partial trench contacting body contacts to the silicon thin film (figure 9.21). The synchronous operation period of an example 64Mb SOI DRAM built using these principles is 5.8 nS, which is 10% faster than bulk DRAM [9.23]. Adding body contacts may lower device density, but using body-tied SOI CMOS technology allows circuit design to be unaltered from bulk CMOS.

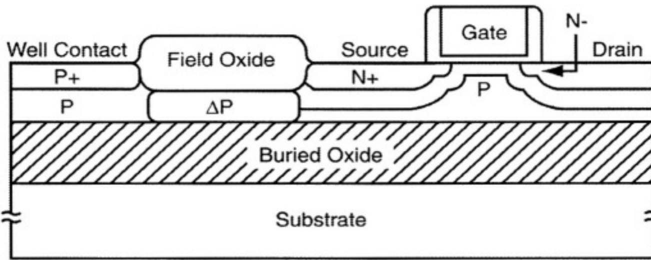


Figure 9.21: Schematic Diagram of a body contacted SOI MOS device used in 64Mb SOI DRAM.

9.9 High Density DRAMs with Body Contacts

Body tied SOI devices have been used to make 256Mb SOI DRAMs [9.49]. These exhibit improved refresh characteristics compared to bulk. In common with bulk devices the cell capacitance is cylindrical with a conventional NO capacitor dielectric film with equivalent SiO_2 thickness.

Storage node junction capacitance reduces with applied voltage in SOI compared to bulk (figure 9.22). This gives SOI DRAM the advantage of faster access time characteristics by 20% over bulk DRAM. SOI DRAM also exhibit good refresh characteristics due to their inherent low junction leakage [9.49-9.51].

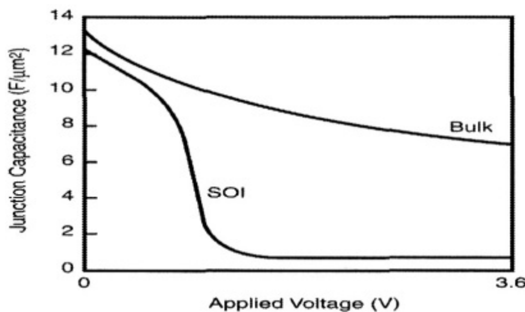


Figure 9.22: Junction capacitance versus applied voltage.

9.10 Operating Voltage Reduction

9.10.1 Half-Vdd Data Line Pre-charge

Low power operation is essential for good signal to noise ratio. Application of half V_{DD} data line pre-charge reduces the array operating current [9.52]. Figure 9.23, shows full V_{DD} pre-charge with a nMOS sense amplifier (SA). Half V_{DD} data line pre-charge is more favorable to SOI CMOS sense amplifiers. In full V_{DD} pre-charge, the voltage difference on the data lines is amplified by enabling pulse Φ_A. The degraded high level is restored to V_{DD} by applying a pulse Φ_R to the active restore circuit. By contrast, in half V_{DD} precharge, the restore operation is simply achieved by applying pulse Φ_R to the cross coupled pMOS. In principle, half V_{DD} pre-charge halves the data line power of full V_{DD} pre-charge with halved data line voltage swing. A large current spike caused by restoring or pre-charging periods is also halved with less noise generation, resulting in a less noisy array [9.53].

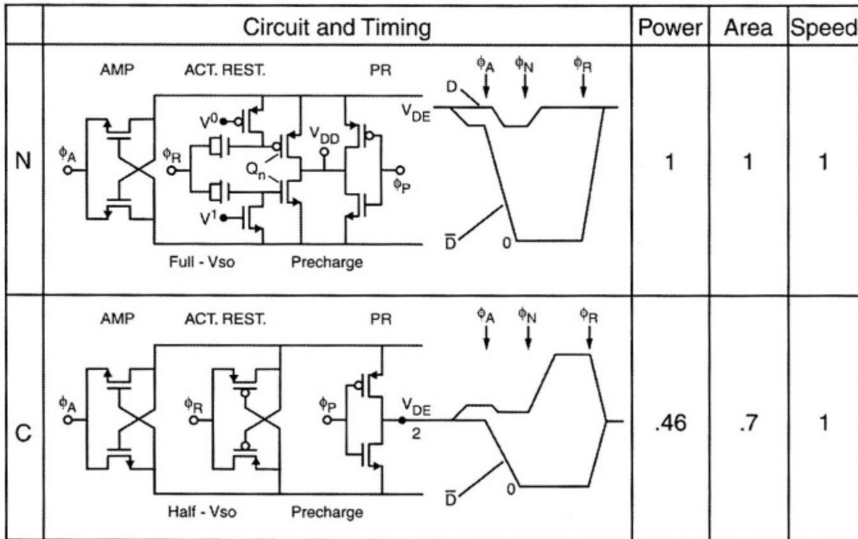


Figure 9.23: Half Vdd precharge technique.

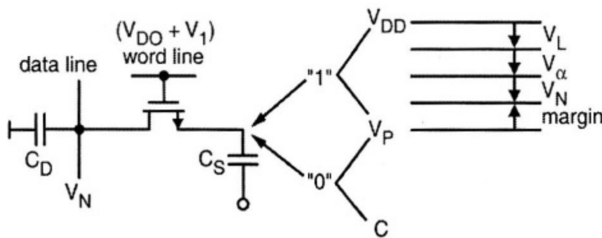


Figure 9.24: DRAM Signal to Noise ratio.

9.10.2 Signal To Noise Ratio

Low voltage operation requires a high S/N ratio design, since it reduces cell margin with reduced signal charge capability [9.54]. The operating margin of the cell (figure 9.24) can be expressed as:

$Q_S > Q_L + Q_C + Q_N$
Q_S (signal charge) = $(V_{DD}-V_P)C_S$
Q_L (leakage charge) = i_L* tREFmax
Q_C (maximum collected charge at the cell node by alpha particles)
Q_N (data line noise charge) = Cd * v_N

Where the available signal voltage for levels '1' and '0' is $V_{DD}-V_P$ for reference voltage $V_P=V_{DD}/2$., i_L is the cell leakage current and v_N is the noise caused by capacitive coupling to the data line from other conductors and the electrical imbalance between a pair of data lines and the amplifier offset voltage. It is essential to maximize the signal charge, C_S (V_D-V_P), while minimizing the effective noise charges in order to achieve stable memory operation [9.55].

9.11 Sense Amplifier Operation

The sense amplifier detects a small signal difference, between the active bit line and a reference bit line. Figure 9.25 shows a conventional DRAM sense amplifier and operation waveforms in a $V_{DD}/2$ precharge scheme [9.56]. Before the word line is activated, the bit line pairs are pre-charged to $V_{DD}/2$ through a pre-charge circuit. Once the word line is activated, the voltage on the bit line with a memory cell connected rises if the memory cell has stored "1" state (figure 9.25) and drops if it has stored "0". After a short time, the signal SN is taken high to turn 'on' the enable transistor and initiate sensing.

Figure 9.25(a) shows the schematic of a SOI DRAM critical path. Both bit lines fall low initially when SN goes high due to coupling effects (figure 9.25(b)). The internal node and body voltages are shown in figure 9.25(c). This turns off Q_M preventing the bitline from falling further. Figure 9.25(d) shows the waveforms for a bulk DRAM. The $\overline{BL}$ continues to decay as SN goes high. When sufficient voltage is developed between BL and $\overline{BL}$, signal SP falls low to restore the storage node to V_{DD}. Since the word line is still high, this "1" state is written back into the cell. The voltage on the storage node at this point is actually $V_{DD}-V_T$, where V_T is the threshold voltage of the pass transistor and since the WL voltage is at V_{DD}. To restore full voltage to the memory cell, the WL voltage must be greater than V_{DD} to compensate for the threshold voltage. SN and SP may be activated simultaneously at the

start of the sensing to allow simultaneous sensing by nMOS and pMOS latches in the sense amplifier [9.57].

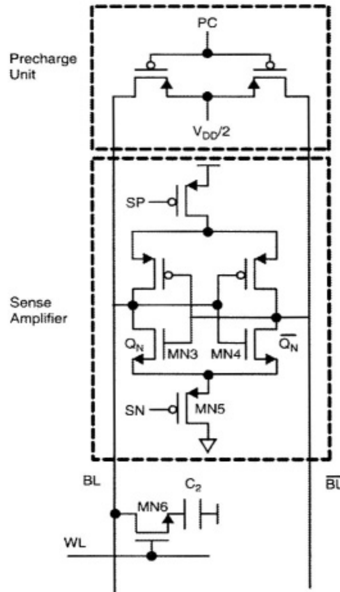


Figure 9.25(a): Schematic of SOI DRAM critical path and waveforms.

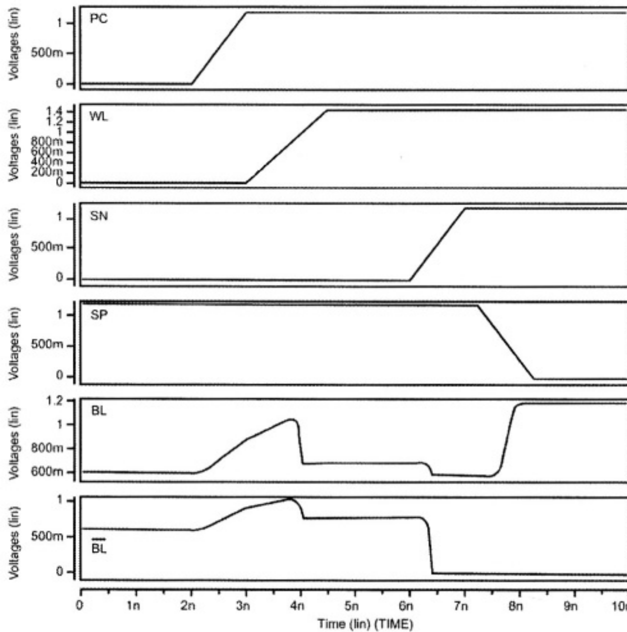


Figure 9.25(b): BL and WL waveforms during read operation in a DRAM, showing delayed activation of signal 'SP'

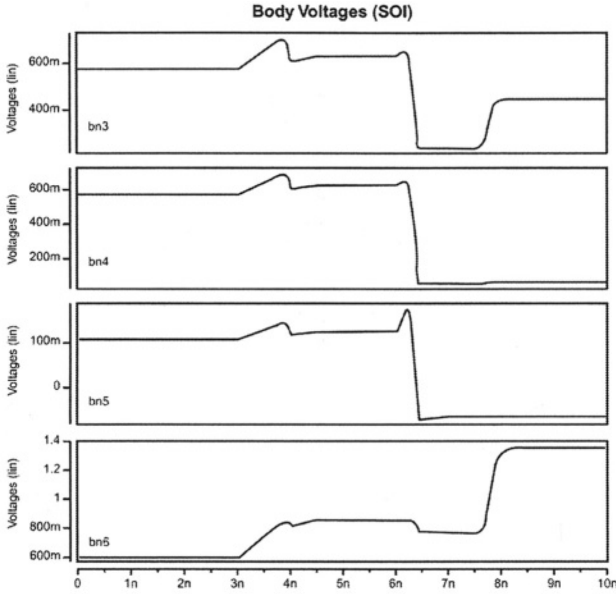


Figure 9.25(c): Body voltages of the internal sense amplifier transistors.

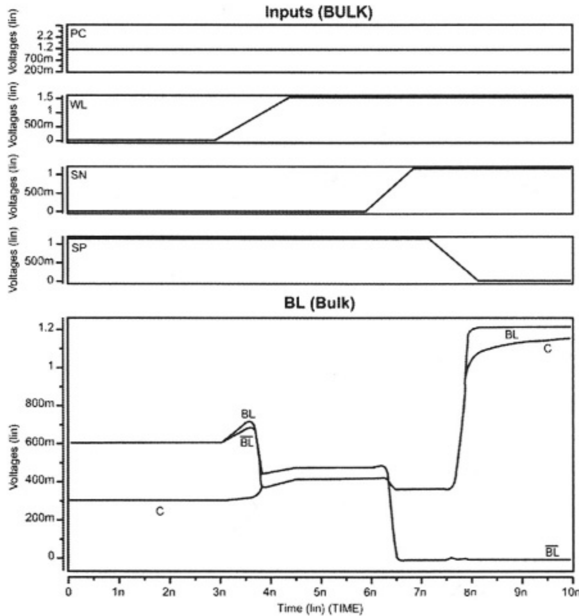


Figure 9.25(d): Bulk DRAM critical path read operation simulation for comparison.

Sensitivity of the sense amplifier depends on the rate at which the enable signal SN is fired. The slower the SN, the smaller signal the sense amplifier can detect. The relationship between time for signal SN to go low, or the latch time, T, and the sensitivity of the sense amplifier V, is given by

$$T = 2C_B(V_{TN} - f)/(\beta)\, f^2\, V_{TN}.V \quad [9.58]$$

Wher C_B is total bit line capacitance, V_{TN} the nMOS threshold voltage, and $f = C_B/(C_B+C_G)$, C_G is the gate oxide capacitance of transistor Q_N, and β is the gain of Q_N.

9.11.1 Sensing with Dummy Cell Structure

A sense amplifier with a dummy cell structure is often used to provide a more sensitive reference level. Figure 9.26 shows a sense amplifier with dummy cell structures. The bit lines are pre-charged to a common reference level before reading from the cell. Simultaneously the reference voltage level is written into the dummy cells on either side of the inverter. When a memory cell is selected, the dummy cells are released from the pre-charged state. As the word line is selected, data from a storage cell is dumped onto one side of the sense amplifier at the same time as the reference level from the dummy cell is dumped on the other side. This causes a voltage differential across the inputs of the sense amplifier. The load clock θ_L turns 'on' the load transistors. The storage nodes rise until the high side turns on the feedback (opposite side) transistor. Regenerative action then results in amplification of the signal to full logic levels, restorating a '0' or '1' level back into the storage cell. Since the bodies are floating, there is a tendency for history effect to slow the regenerative effect. The bodies can be tied together to avoid history effect and obtain uniform delays [9.20].

Figure 9.26: Sense Amplifier with bit lines shorted or common reference level [9.20].

Dynamic sense amplifiers are used in fast memory circuitry. During the reset period, the bit lines are pre-charged to the power supply. The common source node is also pre-charged to $V_{DD}-V_T$. Such a sense amplifier can provide very fast signal detection since it is sensitive to low voltage differential signals if the bodies are tied together. Without a body tie, the bit lines have to separate to a higher differential voltage requiring extra time for the body voltages to settle.

9.11.2 Sensing with body contacts

Body contacts can be used in sense amplifiers to minimize floating body effects and suppress leakage. SOI DRAM sensing requires a minimum bit line differential voltage of 50-100mV for successful read operation @0.18μm. Newer SDRAMs have a bit line differential of ~30mV for a successful READ operations. The higher the differential required for successful read, the slower the read access times. Hence there is a trade-off between performance and sense amplifier robustness.

Due to the stringent sensing requirements and smaller initial bit line signal of DRAMs compared to SRAMs, it is almost inevitable that body contacts be used in the sense transistors. Fitting the sense transistors with body contacts within the cell pitch is more challenging in DRAM's than SRAMs due to the tighter cell pitch. Figure 9.27(a) shows a SOI nMOSFET with a body contact and figure 9.27(b) shows a sense amplifier with body contacts [9.59].

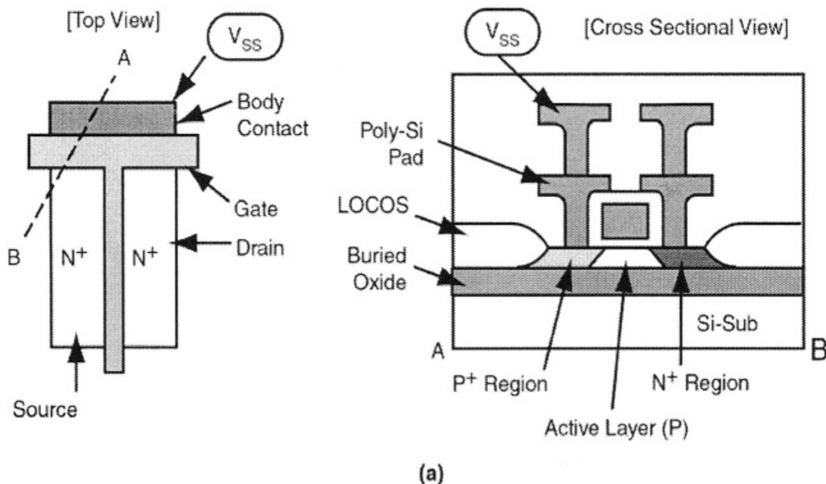


Figure 9.27(a): SOI MOSFET with body contact [9.23, 9.59]

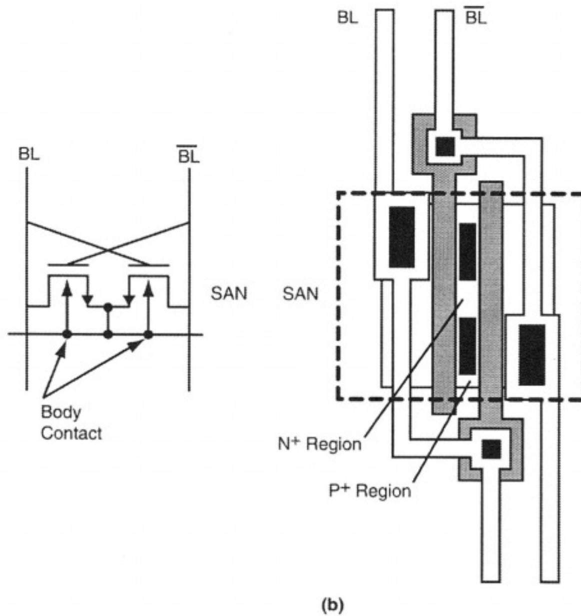


Figure 9.27(b): Sense Amplifier with body contact.

9.11.3 Body-Pulse Sense Amplifier (BPS)

Figure 9.28 shows a body pulse sense amplifier using body contacts for low voltage operation [9.60]. The body-pulse sense amplifier works on the principle of adjusting body potentials dynamically to enhance conducting capability when needed and maintains the original states at other times at the expense of increased power consumption.

During the first period (P1) the word line (WL) is activated and the bit line (BL) begins to change. During the second period (P2), SON goes HIGH and the sense amplifier is activated. Once the sense amplifier is activated the body voltage (SBN) of MN1 and MN2 is raised and the body voltage (SBP) of MN3 and MN4 is lowered momentarily. The magnitude of the threshold voltage of MN1-MN4 is thus reduced and performance of the sense amplifier improved. During period P3, the restoring period, body voltages (SWP, SWN) of MN5 and MP6 are modulated to increase there conducting capability resulting in a faster restoring process.

Based on a similar body-driven principle, figure 9.29 shows the body driven equalizer (BDEQ) circuit [9.60]. When the equalization signal BLEQ is high, the body voltages of the nMOS devices are raised. Hence, the nMOS threshold voltage is reduced enhancing the current drive capability of the nMOS sense transistors and reducing equalizing time.

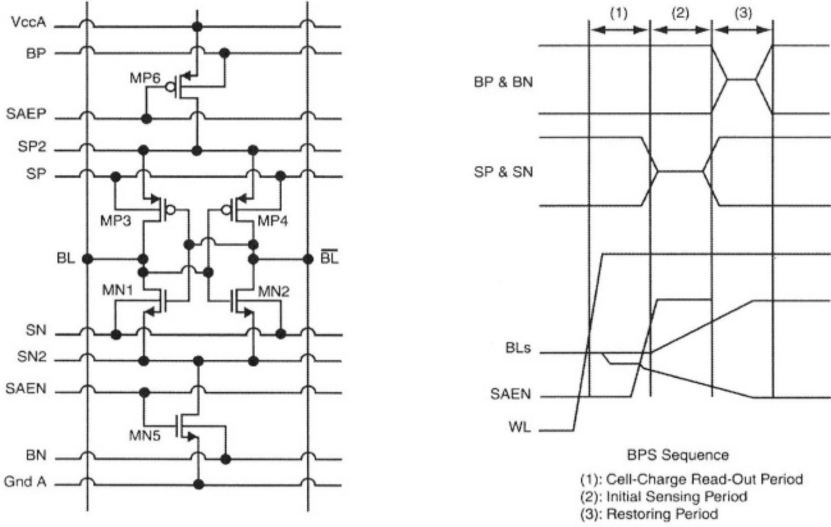


Figure 9.28: Body Pulse Sense Amplifier

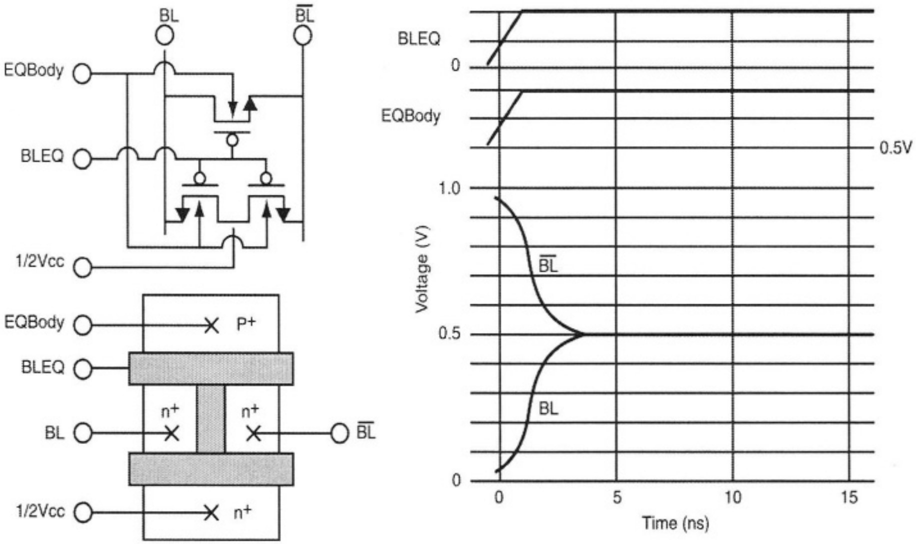


Figure 9.29: Body Driven Equalizer (BDEQ) Circuit

Figure 9.30 shows the body current clamper circuit [9.23, 9.60] used to adjust the body voltage of the sense amplifier transistors. To prevent the body-source diodes from turning 'on' during measurement, (which can increase current flow and affect precision) diodes may be inserted between the body-bias line and both V_{DDB} and V_{SSB}.

9.12 Word Line Boosting

When writing a '1' into the DRAM cell, the voltage is degraded due to the threshold voltage drop across the nMOS pass transistor. The word line voltage must be boosted above V_{DD} to write a full V_{DD} signal to the memory cell. Figure 9.31 shows the voltage condition for boosted word lines. Since the source of the pass transisor (BL) is at V_{DD}, the pass transistor sees an effective back bias voltage of $-(|V_{BB}| + V_{DD})$, V_{BB} is the normal back bias with the source at ground in bulk. Back gate bias technique is not a problem in FD-SOI or with floating body and/or body tied to source PD-SOI. The threshold voltage (V_T) of the pass transistor with back bias at V_{SS} is much higher than the threshold voltage with normal V_{BB}=0V. The word line voltage V_{WL} must be above the source voltage, which is V_{DD} in this case, by at least the threshold voltage. Therefore, the required word line voltage $V_{WL}=V_{DD}+V_T$. A rule of thumb in DRAM design, is that word line boost voltage is at least 1.5 times the power supply voltage for bulk designs and 1X-1.2X for SOI designs [9.61].

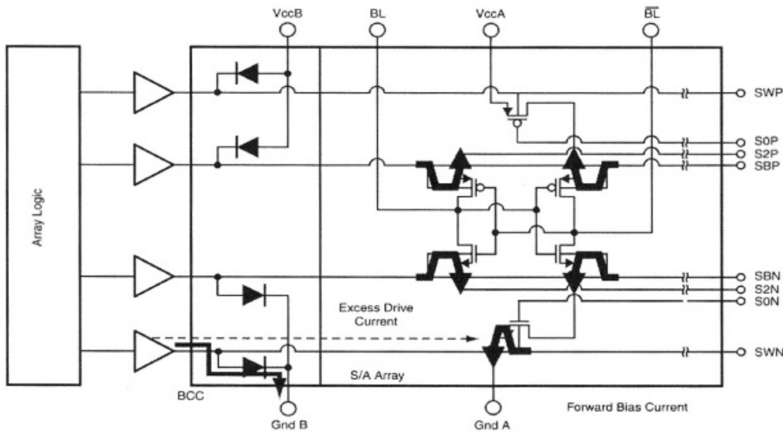


Figure 9.30: Body Current Clamper (BCC) Circuit.

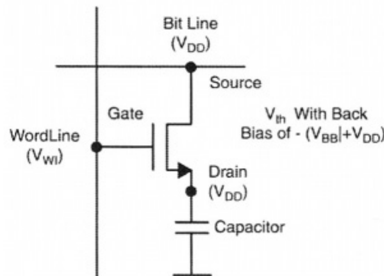


Figure 9.31: Voltage condition for word line boosting.

9.12.1 Boosted Word Line with body contacts
Circuitry that utilizes a boosted word line and the word line driver itself must either have higher drain to source breakdown voltages or be cascoded. Body contacts are usually used for these circuits to suppress early breakdown due to the floating body. Body contacts are also needed in the output drivers to counter supply/ground bounce. Improvements in t_{RAC} by using SOI technology typically range between 20 and 35%. Word line drivers have to be carefully laid-out to fit within the row pitch.

9.12.2 Bootstraped Wordline Driver
Figure 9.32 shows a simple nMOS bootstraped word line driver. This circuit relies on bootstrapping principles to bias the output transistor gates. This bias voltage must be high enough to allow the nMOS transistors to drive the word-line to the boosted voltage [9.61].

Operation of this circuit is as follows and the simulation waveforms are shown in figure 9.32(b) [9.62]. Assume Φ_0 is low and driver transistor OFF initially. $\overline{PDEC}$ is high and PDEC low correspondingly holding the word line low. Once DEC transitions high, the bootstrap node N1 is at V_{DD}-V_T as shown in figure 9.32(b) and transistor MN2 will be 'off' (since $\overline{PDEC}$ is low). The word line will remain at ground until Φ_0 transitions high.

The body voltage of MN2 will couple down initially when the $\overline{PDEC}$ goes low due to body gate coupling and pulling the word line voltage lower initially. The body voltage of MN1 will couple slightly higher than ground due to the body-gate (bootstrap node) capacitance. The body voltage of MN3 will be charged close to supply voltage since the source and gate are at V_{DD} and drain at V_{DD}-V_T. Once the signal Φ_0 goes high, the bootstrap node voltage raises above the supply voltage.

The bootstrap voltage depends on the parasitic capacitance of node N1, C_{GS1}, C_{GD1}, V_{DD}, initial voltage of node N1 (V_{DD}-V_T) and the body voltage of MN3 and MN1. The bootstrap voltage is given by:

$$V_{Bootstrap} = (V_{CC} * C_{GD1}) / [C_{GS1}+C_{GD1}+C_{P1} + C_{GD} + C_{GB}] + (Vcc - V_{TN})$$

C_{GD1} = Gate to drain capacitance of MN1
C_{GS1} = Gate to source capacitance of MN1
C_{P1} = Parasitic node capacitance of node N1.
C_{GB} = Body to Gate capacitance of MN1.
C_{GD} = Body to drain capacitance of MN1

Once the word line voltage rises from ground to V_{DD}, the gate to source capacitance of device MN1 provides a secondary boost to the bootstrap node. The secondary boost ensures that the bootstrap voltage is adequate to drive the word-line to full V_{DD} level.

Layout of the bootstrap node is important. Parasitic capacitance, parasitic leakages and overlap capacitances at node N1 must be minimized to achieve maximum efficiency. Absence of junction capacitance in SOI is helpful for this circuit to ensure adequate gate voltage at the boost node.

Charge leakage from the boost node must be minimized to ensure adequate V_{GS} for transistor MN1 so the word-line remains low for maximum $\overline{RAS}$ period. Low leakage is often achieved by minimizing the MN3 source area or using a donut gate structure [chapter 5] that surrounds the source area.

The source and drain voltages should not exceed limits, else, forward biased diodes are created causing parasitic leakage currents and degrading the word-line voltage at the output and the bootstrap transistor gate voltage. The bootstrap is turned OFF by driving Φ_0 low,and discharging the output node low i.e., the word line is discharged ground.

CMOS drivers are preferred since they dissipate least power. Higher density DRAMs require more rows per bank, and hence more wordline drivers. Leakage current can easily exceed the entire standby current unless care is exercised in the design of the row driver output transistors [9.61].

9.13 Charge pumps and generators

On chip generation of boosted wordline voltages and negative substrate bias techniques are usually achieved with charge pump circuits. In bulk design it is typical to pump V_{DD} using nMOS devices to generate boosted word lines and pMOS to pump V_{SS} to generate negative substrate bias voltages.

NMOS and PMOS transistor charge pumps are required to prevent latch-up and current injection into the arrays. NMOS transistors are typically used in V_{DD} pumps since the intermediate active nodes can go negative with respect to substrate voltage. Any n-diffusion regions connected to these active nodes would forward bias potentially causing latch-up and injection. Similar conditions exist for pMOS devices in V_{BB} pumps.

Figure 9.33 shows a simple voltage pump circuit. The two transistors MN1 and MN2 are configured as diodes. Switches improve the pump efficiency and operating range by eliminating the V_T drops associated with diodes.

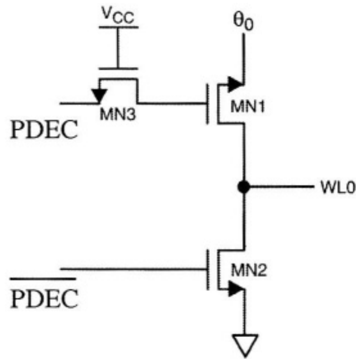


Figure 9.32(a): schematic of the SOI Bootstrap driver with floating bodies.

Figure 9.32(b): Waveforms of the internal nodes and body voltages with nMOS devices of a SOI Bootstrap driver.

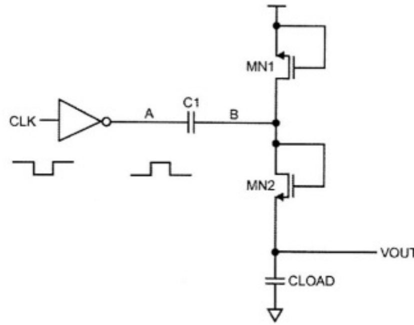


Figure 9.33(a): Schematic of a basic voltage pump circuit [9.61].

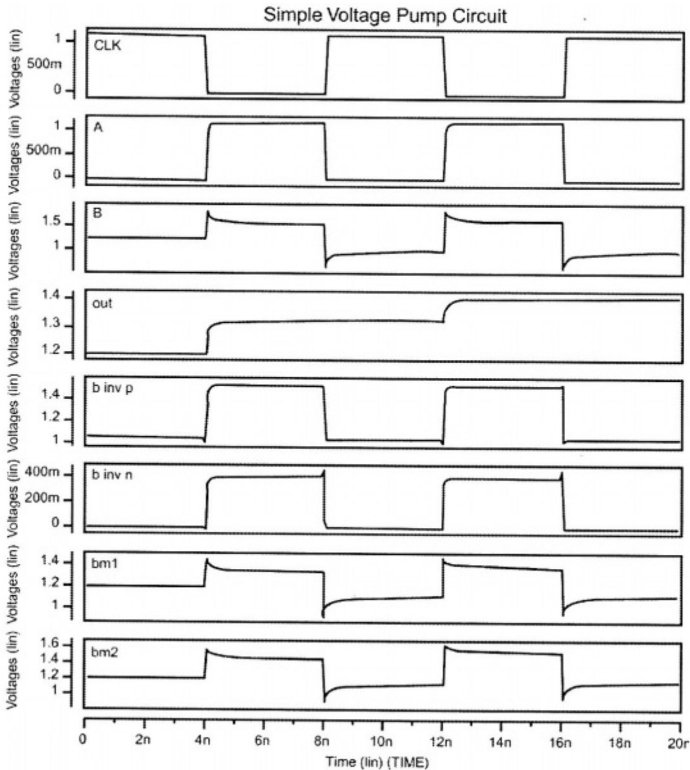


Figure 9.33(b): Simulation waveform of a basic voltage pump circuit.

When the CLK is high, the node A is low and B is high. The charge at node B is $C^*(V_{DD} - V_T)$. Once the clock switches low, the node A rises high, and the node B will raise to $2V_{DD}$-V_T ~1.65v shutting off transistor MN1. The inverter pMOS and nMOS body voltages will initially couple low due to body-gate coupling before they start their respective high and low

transitions. The body voltage of MN1 will couple higher initially due to the body-drain coupling from node B. Once the node B reaches V_T of the MN2, the transistor MN2 will start conducting and the charge from node B gets transferred to the output capacitor C_{OUT}, effectively raising the output node.

During subsequent cycles, the voltage pump continues to charge the output load capacitance until the voltage reaches $2V_{DD}$-V_{T1}-V_{T2}, V_{T1} and V_{T2} are the threshold voltages of MN1 and MN2 respectively. The body voltage of MN2 will couple higher due to body-source/gate and body to drain capacitance and will settle at a higher voltage than V_{DD}. This is because node B is at a higher voltage than supply voltage. A negative pump circuit could be built from figure 9.33 by using pMOS device for the two nMOS transistors [9.61]. More details on charge pumps are provided in chapter 10.

Figure 9.34 and figure 9.35 [9.61] show some common voltage pumps used in DRAMs. The pumps have similar logic functions except for the nMOS replaced by the pMOS. Figure 9.34(b) and figure 9.35(b) show waveforms for the nMOS and pMOS V_{DD} and V_{BB} pumps. As CLK goes low, the pump voltage rises above supply (figure 9.34(b)). Figure 9.35(b) shows the negative pump voltage generated by the clock going high.

Capacity and efficiency are important characteristics of charge pumps. Capacity is a measure of the operating current a charge pump can supply and is determined by the capacitor size, MOS characteristics and its operating frequency. Frequency is limited by the rate at which pump capacitor C_P can be charged and discharged. Efficiency, is a measure of charge wastage during each pump cycle. A typical DRAM voltage pump might be 30-50% efficient. In addition to pump circuits, regulators and oscillators complete the voltage pump design (chapter 10).

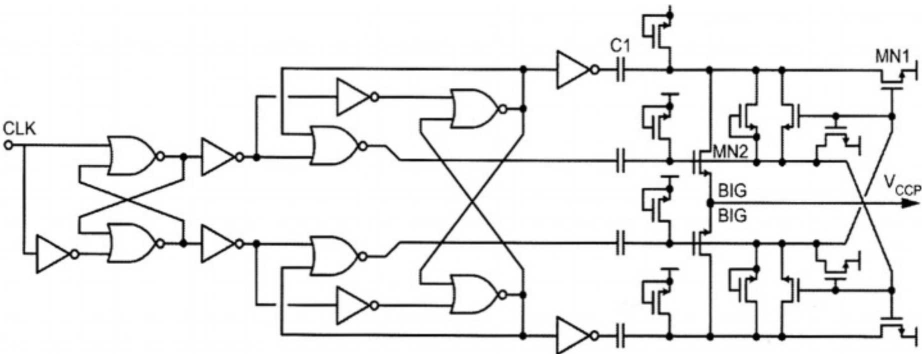


Figure 9.34(a): Schematic of a Vccp Pump.

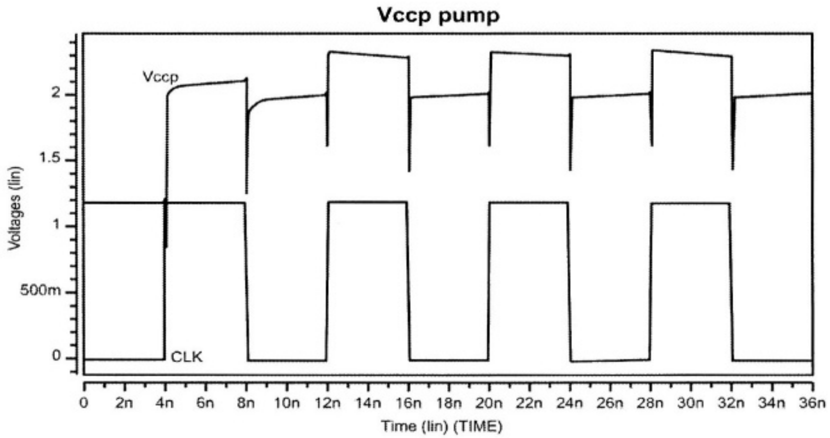


Figure 9.34(b): Vccp Pump simulation waveform, Vcc=1.2V.Pump voltage generated above supply voltage.

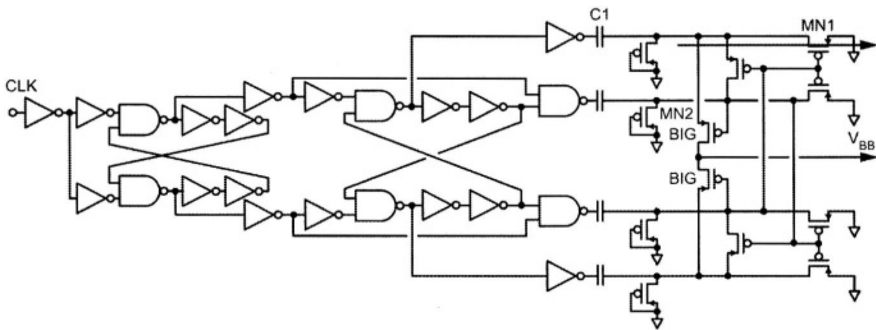


Figure 9.35(a): Vbbp Pump Schematic.

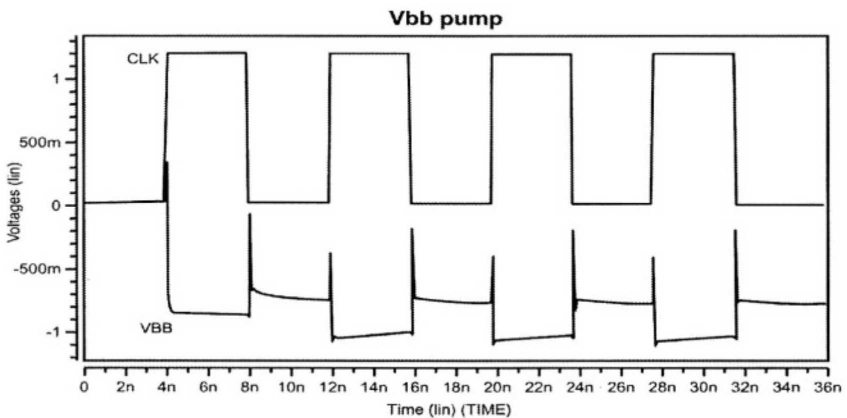


Figure 9.35(b): Vbb pump waveforms, the negative pump voltage generated by clock high.

9.14 Embedded DRAMS in SOI

Merging DRAM and logic large scale integration will be important for future systems. The performance gap between discrete DRAM and the microprocessor can be resolved by using embedded DRAMs. As operating frequency increases, the number of I/O pins also increases, effectively increasing total power consumption. DRAM integration can reduce the number of I/O's and help achieve low power from the smaller load capacitance of the internal I/O bus than that of the external I/O bus. Embedded DRAM can reduce wait states and electro magnetic interference. Many major manufacturers have sought to embedded DRAMs for their mobile and graphics applications [9.63, 9.64].

The periphery circuits in embedded DRAM should work with low power supplies and high clock frequencies for reduced power consumption and improved performance. SOI transistors have reduced junction capacitance which will help charge storage in DRAMs. Negative substrate bias effects help achieve low power. The SOI structure has been experimentally applied to DRAM and an improvement in low voltage operation and SER is observed. Combination of embedded DRAM and SOI transistors could enhance the system performance.

Various techniques have been proposed to help achieve low voltage operation with acceptable performance. The stressless DRAM array [9.65] has proven useful for low voltage operation. This technique could help reduce manufacturing cost while maintaining reliability. The stressless DRAM array takes advantage of the high performance of logic transistors, while the transitional DRAM process degrades the transistor performance compared to logic process. The isolated body in SOI transistor has to be forward biased for low voltage operation. The DTMOS [9.66] technique with gate tied to body (discussed earlier in this chapter) uses the isolated body to dynamically modulate the threshold voltage of the device.

9.15 DRAM operation problems

There are difficulties to be overcome when SOI transistors are used in a DRAM operating at low supply voltages. The gate oxide must be thicker than the logic circuits, effectively increasing the performance gap between the DRAM and logic as gate lengths shrink. To guarantee the stored charge in the DRAM cell, the word line voltage may be boosted. The threshold voltages cannot be scaled for adequate data retention as the sub-threshold current increases exponentially at low voltages with decreasing threshold voltages. To avoid this, one may maintain a negative word line voltage during standby, causing a negative V_{GS} for un-accessed memory cells. The

sub-threshold current decreases exponentially if the word line gate voltage is made more negative, improving the refresh characteristic of the SOI DRAM cell without affecting the threshold voltage of the device at low voltages. However, the gate to drain overlap region is stressed in terms of time dependant dielectric breakdown (TDDB).

The stressless DRAM array architecture [9.67] relaxes the electric field for both channel and fringe of the memory cell transistors so the threshold voltage and the gate oxide thickness of the memory cell transistor can be reduced with scaled voltage.

The voltage level of the unselected word-line is lower than ground level. However, the electric field of the memory cell architecture is completely relaxed by controlling the voltage at the storage node through the memory cell plate coupling. This architecture improves the data retention despite the lower threshold voltage of the memory cell transistor at lower power supplies.

Stored charge can be maintained without boosting the voltage level of the active word-line the same way it is enhanced by the memory cell capacitor coupling in the stored charge boosted architecture [9.67]. Therefore, the applied electric field to the channel region of the memory cell may be relaxed. Conventional architectures divide the cell plate at each wordline to prevent destruction of data in unselected memory cells. This causes an area penalty. The stressless architecture reduces the cell plate division and minimizes area penalty. Thus performance of the embedded DRAM is enhanced because both the threshold voltage and gate oxide thickness can be scaled.

The electric field at the fringe of the memory cell transistor after the restore operation may be relaxed to improve reliability. Once restore is completed, the storage node is boosted to a negative voltage by the memory cell plate coupling to reduce the applied voltage at the fringe region of the unselected memory cell transistor. The read-out signal from the selected memory cell is differentiated with full size memory cell. The storage node of the dummy memory cell is boosted by half the memory cell capacitance to generate a reference voltage level [9.65].

9.16 SOI DRAM READ Critical Path Body Contacts

Figure 9.36 shows typical SOI DRAM read path critical circuits like word line drivers, buffers, and sense amplifiers. The circuits that typically use body contacts are shown inside the dashed lines [9.27, 9.28]. This is to

reduce their susceptibility to floating body effects. The memory cell and logic gates may or may not use floating body configurations, depending on packing density, and may lead to performance trade-offs.

To increase cell density, body contacts should not be used in the DRAM cell. However, without the body contact, the SOI device is susceptible to floating body effects. Due to junction leakage and thermal generation, the body potential may rise since holes are accumulated in the floating body. This reduces threshold voltage, further increasing sub-threshold leakage.

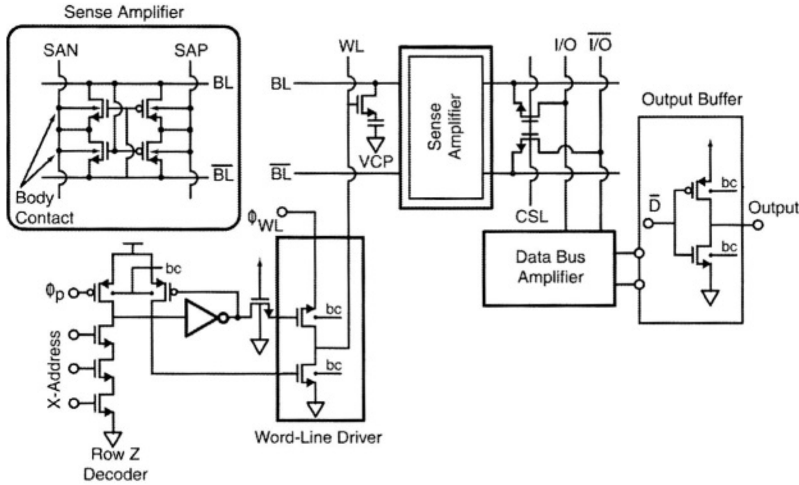


Figure 9.36: DRAM critical path circuits in read path with body contacts.

<u>9.17 Synchronous Interface on DRAMs</u>

Historically DRAMs have been controlled asynchronously by the processor. Asynchronous control places addresses on the DRAM inputs and strobes them by using the $\overline{RAS}$ and $\overline{CAS}$ pins. The addresses are held for a required minimum time, during which the DRAM accesses the address locations in memory. After a maximum delay (access time) the controller either writes new data from the processor into memory or provides data from the memory to its outputs for the processor to be read. During this time, the processor must wait for the DRAM to perform various internal functions such as pre-charging the bit lines, decoding addresses, sense data, and route the data out through the outputs buffers. This creates a wait state during which the high speed processor waits for the DRAM to respond, slowing down the entire system.

In case of an asynchronous DRAM, the processor must apply the row address and hold it active while strobing with $\overline{RAS}$ pin. This is followed by (t_{RCD}), the column address which must be held valid and strobed with $\overline{CAS}$

pin. The processor must then wait for the data to appear on the outputs later (t_{CAC}), stabilize and be read. Figure 9.37 shows the details of asynchronous and synchronous DRAMs.

An alternate strategy is to make the memory circuit synchronous. Input and output address latches on the DRAM, which can hold the data. Inputs latches store the addresses, data and control signals on the inputs of the DRAM, freeing the processor for other tasks. After a preset number of clock cycles, data can be available on the output latches of a DRAM with synchronous control for read or write. Synchronous control means that the DRAM latches information from the processor in and out under the control of the system clock. The processor can be told how many clock cycles it takes for the DRAM to complete its task, so it can safely go off an do other tasks while the DRAM is processing its request. The synchronous DRAM functions as a digital data storage device [9.68]. One advantage of synchronous DRAMs is that the system clock is the only timing edge required by the memory. This reduces propagation of multiple timing strobes around the system.

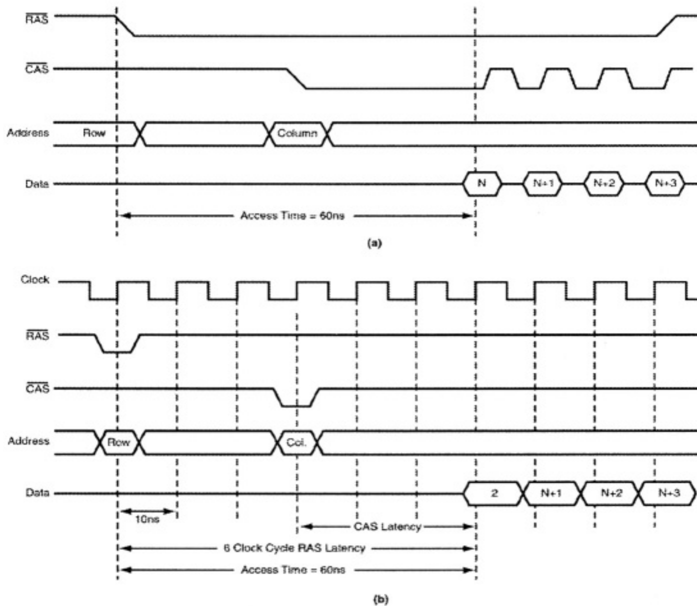


Figure 9.37: Comparison of DRAM (a) Asynchronous (b) Synchronous [9.68]

9.18 High Speed Modes for Synchronous DRAMs

Hiding internal timings in a synchronous DRAM has proved to be a challenge and many techniques have been proposed. The active command to

read/write, t_{RCD}, and pre-charge time, t_{RP}, can be hidden after the first access through the use of a page mode or burst mode, in which a series of data bits from the page currently on the sense amplifiers can be clocked out rapidly following access of the first bit. For operating in page mode, a new column address is needed for each access.

During burst mode, following an initial address input, subsequent addresses are either internally generated in rapid succession without inputting new address information to the DRAM or generated with linear or interleaved address sequences. Burst can be combined with a "wrap" feature which allows the string of bits stored before and after the initial bit location in the DRAM to be accessed in rapid succession after the initial access. This feature is useful for high speed cache fill applications since the most likely bits to be accessed following an initial access, are those physically close to the given bit.

Pipelining addresses is another way to speed up average DRAM access time. The input latch may be used to store subsequent addresses for the processor while the DRAM is operating on the previous address. Normally addresses to be accessed are known several cycles in advance so the processor can send the second address to the input address latch of the DRAM as soon as the first address has moved to the next stage of processing. This eliminates the need for the processor to wait for a full access cycle before starting the next DRAM access. Figure 9.38 shows the read operation with a pipelined architecture, and the staggering of stages with multiple tasks running in parallel.

The column address A1 is clocked into the address buffer on one cycle and is decoded. On the second cycle, the column switch transfers the corresponding data D1 from the sense amplifier to the read bus, and the column address A2 is clocked into the address buffer. At the third clock cycle, data D1 is clocked into the output buffer, D2 is transferred to the read bus, and A3 is clocked into the column address buffer. When D1 appears on the output, D2 and D3 are in the pipeline behind it.

While it takes three cycles for the data from A1 to appear on the output buffer, it only takes one more cycle for the data from A2 to appear. Thus pipelining can reduced the CAS cycle time if addresses are known several cycles in advance. This is particularly useful for burst mode operations.
Pipelined architectures do not generally increase chip size. They use the conventional SOI DRAM cell array and the normal operation fits well into the pipeline structure requiring no additional address or read buses. It also

permits interruptions of a burst on any address since both external and internally generated addresses are transmitted through the pipeline in the same manner. This permits random column access with a different external address on every clock cycle.

A drawback of pipelining is that addressing occurs every cycle so all stages of the SOI DRAM are active simultaneously, consuming power. Pipelined architectures are run at a slightly lower frequency due to the need to pre-charge long I/O lines during read. The column select line has to remain high long enough to write data from the I/O lines to the bit-lines but stay low long enough to pre-charge the I/O lines without contention to read quickly.

9.19 Prefetch Architecture

Prefetch techniques are used for increasing speed of a synchronous SOI DRAM (figure 9.39). Here, more than one data word is fetched from the memory on each address cycle and transferred to a data selector on the output buffer. Multiple data words can then be sequentially clocked out for each address access. This has the advantage that data access can be at higher clock rates than the internal DRAM timing. Thus, both power dissipation and noise can be reduced for long sequential bursts.

There are advantages and drawbacks to the pre-fetch architecture. An output register must be added to the chip to hold multiple prefetched words, adding to chip size. If more than two address bits are pre-fetched, chip size increases considerably, at the expense of fast data stream that can hide pre-charge between *RAS* and *CAS* accesses.

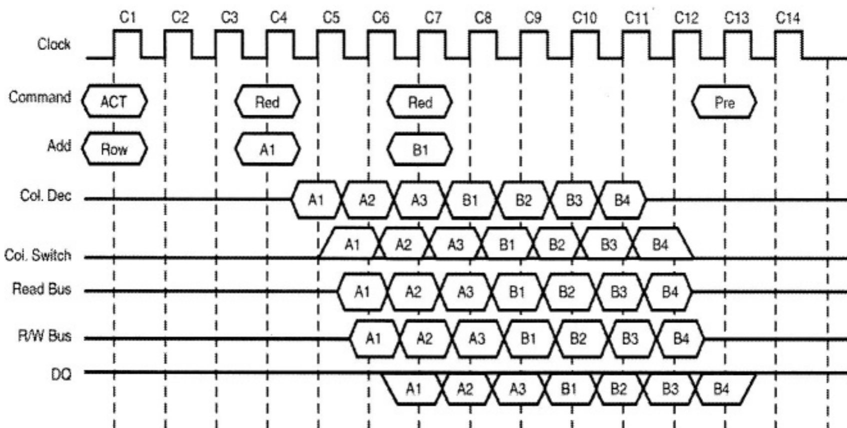


Figure 9.38: Read operation in a pipelined architecture [9.68]

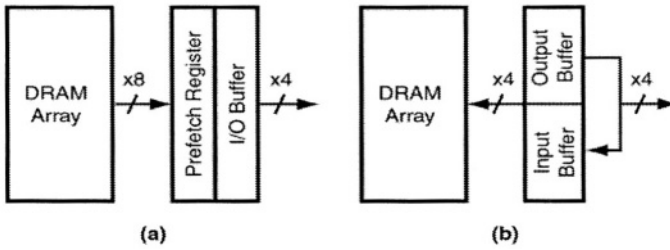


Figure 9.39: Block Diagrams of Asynchronous DRAM outputs (a) Prefetch (b) Pipelined [9.68].

9.20 Other Architectures

A combination of prefetch and pipelining may be used to enhance DRAM performance. Multiple internal banks can also be used to speed SOI DRAMs. These permit faster random access by allowing one bank to pre-charge or be refreshed while another bank is being accessed. Higher speeds can be achieved by switching back and forth between active and inactive banks (ping-pong) on one chip rather than by interleaving multiple banks in the system. There is, however, a power penalty since both banks must be kept active. Figure 9.40 shows the DRAM architecture for multibank operation [9.68].

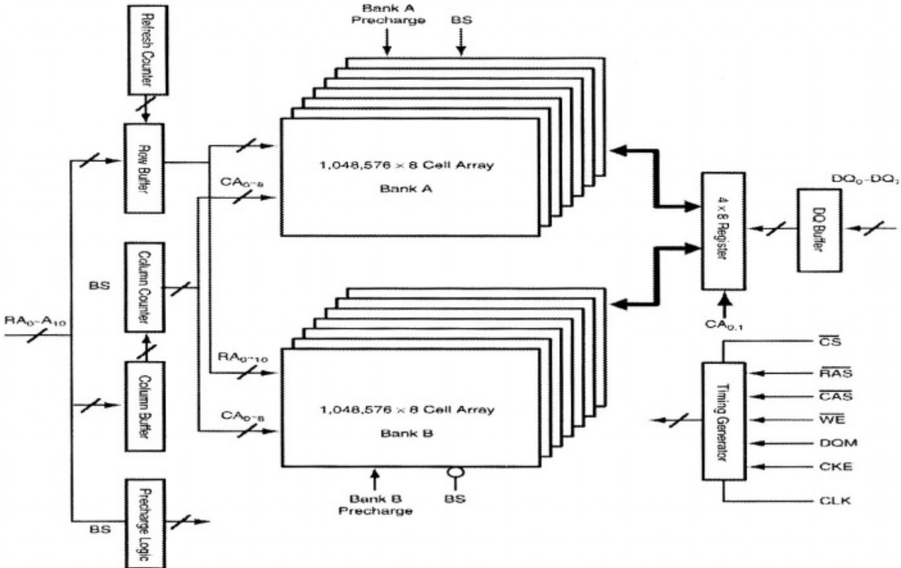


Figure 9.40: DRAM architecture with multibank operation. Multibanks improve DRAM access times by precharging one bank when the other is performing memory intensive tasks [9.68].

9.21 Destructive read out characteristics of DRAM

The destructive READ of a DRAM cell necessitates successive operations of amplification and restoration. This is performed by a latch type CMOS sense amplifier on each data line. Consequently, a data line is charged and discharged with a large voltage swing ΔV_D and with a charging current of $C_D \Delta V_D$ where C_D, is the data line capacitance. To reduce active power for a fixed cycle time it is necessary to (1) reduce charging capacitance (2) lowering external and internal voltages (or) (3) reduce static current. In particular, emphasis must be placed on reduction of total data line charge since it dominates total active power, while maintaining an acceptable signal-to-noise ratio (figure 9.41) [9.69].

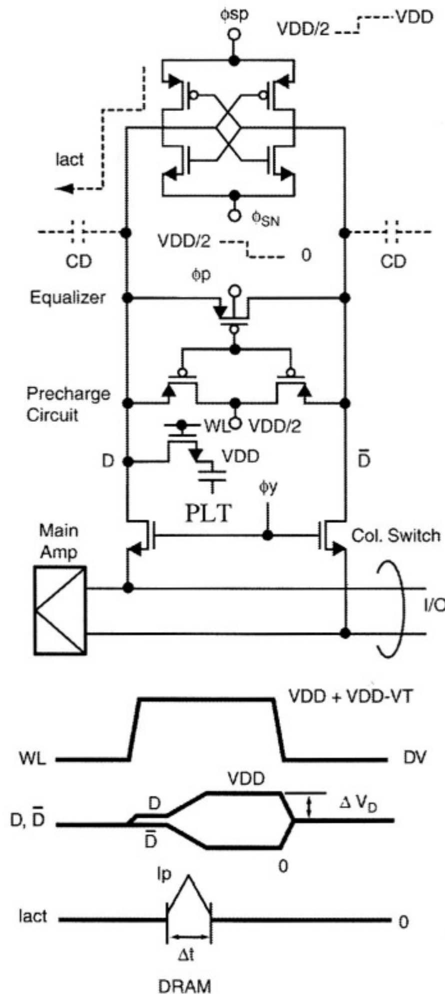


Figure 9.41: Critical Path and standby current for DRAM [9.69]

DRAMs and SRAMs have evolved to use similar circuit techniques. In addition to active current, DC sub-threshold current is a source of active power dissipation. As threshold voltages reduces, the sub-threshold current increases. Temperature, floating-body and threshold variations offset leakage currents, causing cell stability degradation. Sub-threshold current is proportional to total channel width. The sub-threshold current generated from an overwhelming large number if inactive circuits would eventually be larger than AC current from small number of active circuits. As technology advances leakage will be very important [9.69, 9.70].

Active current paths are shown in figure 9.41 DRAMs. Total chip standby current increases with memory capacity.

References

[9.1] D. Goldman et. al., "0.15um SOI DRAM technology incorporating sub-Volt dynamic threshold devices for embedded mixed signal & RF circuits". Proc. of SOI Conference, 2001.

[9.2] T. Yamauchi, et. al., "High performance embedded SOI DRAM architecture for the low power supply", IEEE Journal of Solid State circuits, Aug 2000, Vol.35, No.8, pp. 1169 – 1178.

[9.3] J.W.Park, et. al., "Performance characteristics of SOI DRAM for low power applications", ISSCC 1999, pp. 370.

[9.4] T. Kirihata, et. al., "A 390mm sq. 16 bank 1Gb DDR SDRAM with Hybrid bitline architecture" ISSCC 1999.

[9.5] T. Fujno, et. al., "Multi Gbit Scale Partially Frozen (PF) Nand DRAM with SDRAM compatible interface", VLSI symp., 1998, pp. 96 - 99.

[9.6] C.S.Kim (Semiconductors Asia/Pacific), Dataquest, - Korea, May 1997

[9.7] T.Oashi, et. al., Technical Digest of IEDM, 1996, pp. 609.

[9.8] K.Shimomura, et. al., Digest of technical papers, ISSCC, 1997, pp. 68.

[9.9] H.S.Kim, et. al., Digest of technical papers of symposium of VLSI technology, 1995, pp. 143.

[9.10] T.Eimori, et. al., Technical digest of IEDM, 1993, pp. 45.

[9.11] J. P. Colinge, "SOI Technology, Materials to VLSI, 2nd Edition", Kluwer Academic Press

[9.12] Y. Hu, et. al., VLSI Technology, 1996, pp. 128 - 129.

[9.13] A. Chatterjee, et. al., Technical Digest of IEDM, 1994, pp. 87.

[9.14] H.Lu, et. al., Proceedings of the international SOI Conference, 1993, pp. 182 – 183.

[9.15] C. Y. Chang & S. M. Sze, "ULSI Devices", John Wiley & Sons, ISBN:0-471-24067-2

[9.16] M.Inoue, et. al., "A 16Mb DRAM with a relaxed sense amplifier pitch open bit line architecture", IEEE JSSC , 23, Oct 1988, pp. 1104.

[9.17] D. Takashima, et. al., "Open/folded bitline arrangenet for ultra high density DRAMs", Symposium VLSI circuits, May 1993, pp. 89-90.

[9.18] K.Shibahara, et. al., "1GDRAM cell with diagonal bit line configuration and edge operation MOSFET", IEEE Electron devices, Dec 1994, pp. 639.

[9.19] A. H. Shah, et. al., "A 4Mb DRAM with crosspoint trench transistor cell", ISSCC, Feb 1986, pp. 268 – 269.

[9.20] B. Prince, "Semiconductor Memories, 2nd Edition", John Wiley & Sons, ISBN:0-471-94295-2

[9.21] Kubota, et. al., "A New soft error immune DRAM cell characterised by storage capacitor on a bit line structure", IEDM, 1987 , pp. 596.

[9.22] I. K. Kim, et. al., "Advanced Integration technology for a highly scalable SOI DRAM with SOC (Silicon on Capacitors)", IEDM Dig., 1996, pp. 605 – 608.

[9.23] J. Kuo & K. W. Su, "CMOS VLSI Engineering – SOI", Kluwer Academic Publishers, ISBN: 0-7923-8272-2

[9.24] S. Nakamura, et. al., "Gigabit DRAM cells with low capacitance and low resistance bitlines on buried MOSFETs and capacitors by using bonded SOI technology – Reversed stacked capacitor(RSTC) cell", IEDM Technical Digest, 1995, pp. 889 – 892.

[9.25] A. J. Auberton-Herve, "SOI: Materials to Systems", IEDM Digest, 1996, pp. 3 – 10.

[9.26] K. Terada, et. al., "A new DRAM cell with a transistor on a lateral epitaxial siliconlayer (TOLE) cell", IEEE transactions on electron devices, Vol 37., No. 9, sept 1990, pp. 2052 - 2057.

[9.27] K. Suma, et. al., "An SOI DRAM with wide operating voltage range by CMOS/SIMOX technology", ISSCC Digest Tech. Papers, 1994, pp. 138 – 139.

[9.28] C.T.Chuang, et. al., “SOI for Digital CMOS VLSI: Design considerations and Advances”, IEEE Journal of Solid State Circuits., Vol. 86, No.4, April 1988.

[9.29] Y. Tosaka, et. al, ”Alpha particle induced soft errors in submicron SRAM”, Dig. Tech. Papers, VLSI Technology Symposium, 1995, pp. 39 - 40.

[9.30] F. Morishita et. al., “Leakage mechanism due to floating body and countermeasure on dynamic retention mode of SOI DRAM”, Dig. Tec. Papers, VLSI tech. Symp., 1995, pp. 141 - 142.

[9.31] G. R. Srinivasan, et. al., “Accurate, predictive modeling of soft error rate due to cosmic rays and chip alpha radiations”, Proc. IRPS, 1994, pp. 12 - 16.

[9.32] E. Normond, et. al., “Single event upset and charge collection measurements during high energy protons and neutrons”, IEEE Trans. Nucl. Sci, Vol 41, 1994, pp. 2203 – 2209.

[9.33] T. Aton, et. al., ”Direct measurement for SOI and bulk diodes of single event upset charge collection from energetic ions and alpha particles” Dig. Tech. Papers. Symp., VLSI Technology, 1996, pp. 98 - 99.

[9.34] S. M. Sze, “Physics of Semiconductor Devices”, Wiley.

[9.35] T. Hamamoto, et. al., “Well concentration: A Novel scaling limitation factor derived from DRAM retention time and its modelling”, IEEE Int. Electron Devices Meeting, Dec. 1995, pp. 915.

[9.36] T. Tanigawa, et. al., “Enhancement of data retention time for giga-bit DRAM’s using SIMOX technology”, Dig. Tec. Papers, VLSI Syposium, 1994, pp. 37 – 38.

[9.37] S. Tomishima et. al., “A long data retention SOI DRAM with body refresh function”, Dig. Tech. papers, VLSI Symposium, 1996, pp. 198 – 199.

[9.38] H. S. Kim, et. al., ”A high performance 16M DRAM on thin film SOI”, Dig. Tech. Papers, Symposium VLSI Technology, 1995, pp. 143 – 144.

[9.39] T. Nishihara, et. al., ”Data retention in ultra thin film SOI DRAM with buried capacitor cell”, Dig. Tech. Papers, Symp, VLSI Technology, 1994, pp. 39 - 40.

[9.40] M. Asakura, et. al., ”A 34nS 256 Mb DRAM with boosted sense ground scheme”, ISSCC Digest, 1994, pp. 140 - 141.

[9.41] “An experimental 256 Mb DRAM with boosted sense ground scheme”, IEEE Journal of solid state circuits, Vol 29, Nov 1994, pp. 1303-1309.

[9.42] T. Ooishi, et. al., “An automatic temperature compensation of internal sense ground for sub-quarter micron DRAMs”, Dig. Tech. Papers, VLSI Circuits symposium, 1994, pp. 77 – 78.

[9.43] Y. H. Koh, et. al., “64Mb SOI DRAM technologies using body contacted structure”, IEEE SOI Conf, 1997, pp 170-171.

[9.44] M. Terauchi & M. Yoshimi, “Analysis of floating body induced leakage current in 0.15um SOI DRAM”, Proc. IEEE International SOI Conference, 1996, pp. 138 - 139.

[9.45] J. A. Mandelman, et. al., “Floating body concern for SOI DRAM”, IEEE SOI Conference, 1996, pp. 136 – 137.

[9.46] Z. Yang & S. Mourad., “Crosstalk in Deep Submicron DRAMs”., Proceedings of the 2000 IEEE International Workshop on Memory Technology, Design and Testing (MTDT'00).

[9.47] F. Assaderaghi, et. al., “A 7.9/5.c psec room/low temperature SOI CMOS,” IEDM Tech. Dig., Dec 1997, pp. 415 – 418.

[9.48] H. S. Kim, et. al., “Data retention times in SOI DRAMs” Symp. VLSI Tech, 1996, pp. 126 - 127.

[9.49] J. W. Lee, et. al., “Performance improvements in high density DRAM application using 0.15um body contacted SOI technology”, SOI Conference, Oct 2000.

[9.50] T. Eimori, et. al., "Approaches to extra low voltage DRAM operation by SOI-DRAM", IEEE Transactions on Electron Devices, May 1998, pp. 1000 – 1009.

[9.51] Y. H. Kon, et. al., IEEE Trans Electron Dev., Vol 45, 1998, pp. 1063.

[9.52] "Power Reduction Techniques in megabit DRAMs" IEEE JSSC, Vol SC-21, June 1986, pp. 381 - 389.

[9.53] "Half Vdd Bitline sensing scheme in CMOS DRAM", IEEE JSSC, Vol SC-19, Aug 1984, pp. 451 – 454.

[9.54] "A 288K CMOS Pseudo SRAM" IEEE JSSC, Vol SC-19, 1984, pp. 619 – 623.

[9.55] J. Rabaey & M. Pedram, "Low power design methodologies", Kluwer Academic Publishers.

[9.56] N. C. Lu & H. H. Chao, "Half Vdd Bitline sensing scheme in CMOS DRAMs", IEEE JSSC, 19, Aug 1984, pp. 451.

[9.57] K. Itoh, et. al., "Trends in Low power RAM Circuit technologies", Proc. IEEE, April 1995.

[9.58] W. T. Lynch & H. J. Boll, "Optimization of the latching pulse for dynamic flip flop sensors", IEEE JSSC, 9, Apr 1974, pp. 49.

[9.59] Y. Wada, et. al., "Active body bias SOI CMOS driver circuits" Symp. VLSI Circuits 1997, pp. 29 - 30.

[9.60] K. Shimomura, et. al., "A 1V 46nS 16Mb SOI DRAM with body control technique", IEEE JSSC, Vol 32, No. 11, Nov 1997, pp. 1712 – 1720.

[9.61] B. Keeth & J. Baker, "DRAM Circuit Design", IEEE Press, ISBN:0-7803-6014-1

[9.62] K. Noda, et. al., "A Boosted Dual Word line decoding scheme for 256Mb DRAMs" 1992 Symposium on VLSI Circuits, Digest of technical papers, pp. 112 – 113.

[9.63] K. Suma, et. al., “An SOI DRAM with wide operating range by CMOS/SIMOX technology” IEEE J. Solid State Circuits, Vol 29, Nov. 1994, pp.1323 – 1329.

[9.64] K. Shimomura, et. al., “A 1V 46nS 16Mb SOI DRAMwith body control technique”, ISSCC Dig. Tec. Papers, 1997, pp. 68 – 69.

[9.65] T. Yamauchi, et. al., “High performance embedded SOI DRAM architecture for the low power supply”, IEEE Journal of Solid State circuits, Aug 2000, Vol.35, No.8, pp. 1169 – 1178.

[9.66] F. Assaderaghi, et. al., “A dynamic threshold voltage MOSFET for ultra low voltage operation”, IEEE IEDM Tech. Dig. 1994, pp. 809 – 812.

[9.67] K. Fujishima, et. al., “A storage node boosted RAM with wordline delay compensation”, ISSCC Dig. Tech. Papers, 1982, pp. 66 - 67.

[9.68] B. Prince, “High Performance memories”, John Wiley & Sons, ISBN:0-471-95646-5

[9.69] M. Aoki & K. Itoh, “Low voltage low power ULSI circuit technique”, IEICE, Vol E77-C, No.8, Aug 1994, pp. 1351-1360.

[9.70] Baker & Boyce “CMOS Circuit design, layout and simulation”, IEEE Press

Chapter 10: Analog Design

10.1 Introduction

Digital circuit designers are comfortable with PD-SOI and are taking advantage of the benefits of this technology. Somewhat less attention has been paid to analog applications in PD-SOI. Slow development of analog CAD tools for SOI applications may have contributed to this. Analog design requires consideration of threshold voltage variations, body-related effects and thermal management. Body ties can in some cases be used to stabilize threshold voltage and body related effects, but they may not always be the optimal solution.

10.1.1 Benefits of SOI for Analog Design

Improved frequency performance through reduced capacitance, higher drive current and a reduction in interconnect length are significant advantages of using SOI for analog applications. Noise and latchup are minimized through reduced substrate coupling. Silicon resistors have improved linearity with respect to absolute voltage in many cases, since they do not form reverse biased diodes to substrate (as resistor voltage increases the reverse biased diode depletion region increases in bulk material, resulting in higher resistances for the same resistor layout). Inductor Q can be enhanced through the use of very high resistivity substrates.

10.1.2 Drawbacks of Analog Design on SOI

The most significant drawbacks for analog designs on SOI are due to the floating body, and especially the kink effect. Body ties can be effective at low frequencies, however, at RF frequencies it becomes more of a challenge due to the RC time constant of the gate to body coupling. Another drawback

is poor thermal response due to buried oxide and trench isolation. Device noise is an issue when using PD-SOI devices, especially during operation near the kink region.

The Kink Region

Kink free operation of PD-SOI with a floating body is generally desirable for analog circuits. A cascode configuration (figure 10.1) makes it possible to bypass the kink effect by operating outside the voltage region in which the kink occurs. A reference voltage is provided to node V_{in}. Device M_2 is used to define the drain voltage of device M1 more precisely, and thus the output sink current of M2 is more accurately mirrored than a non-cascoded pair.

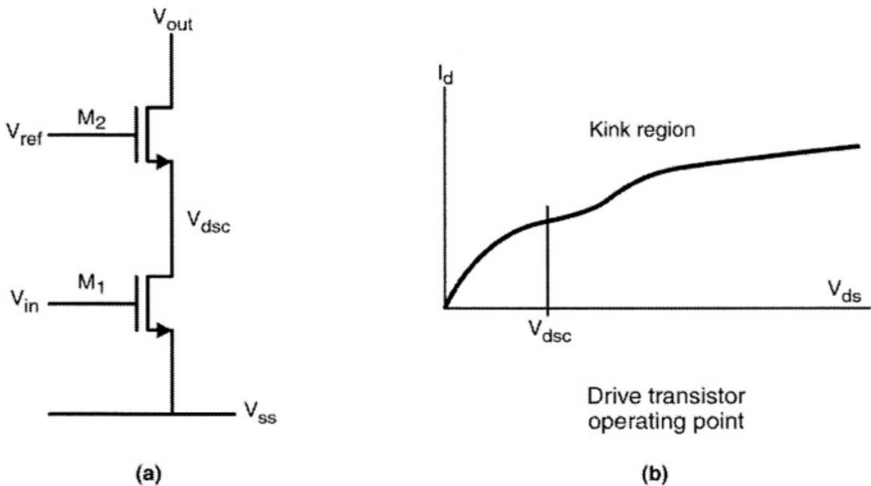


Figure 10.1: Demonstration of the operating point that can be maintained, from the PD-MOS device when used in a cascode configuration.

Figure 10.2(a) shows one example of how the circuit of figure 10.1 may be achieved in practice.

An additional benefit of operating the devices in this region is that noise resulting from impact ionization in the cascodes is reduced from the non-cascoded device configuration, as it appears in series with the high impedance of the driver device.

Figures 10.2(b) and 10.2(c) show the current and voltage waveforms as supply is ramped, for both the non-cascoded current mirror, and cascoded current mirror schemes.

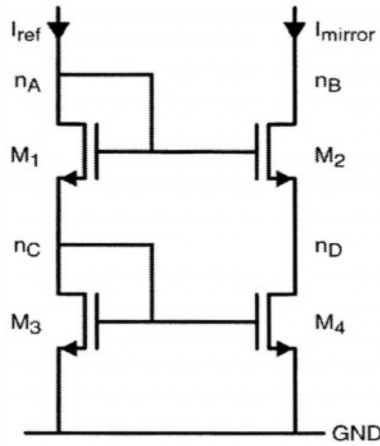


Figure 10.2(a): Basic cascode circuit design. This circuit allows the operating conditions of M4 to be restricted and keep away from the kink effect region of the MOS operating curve.

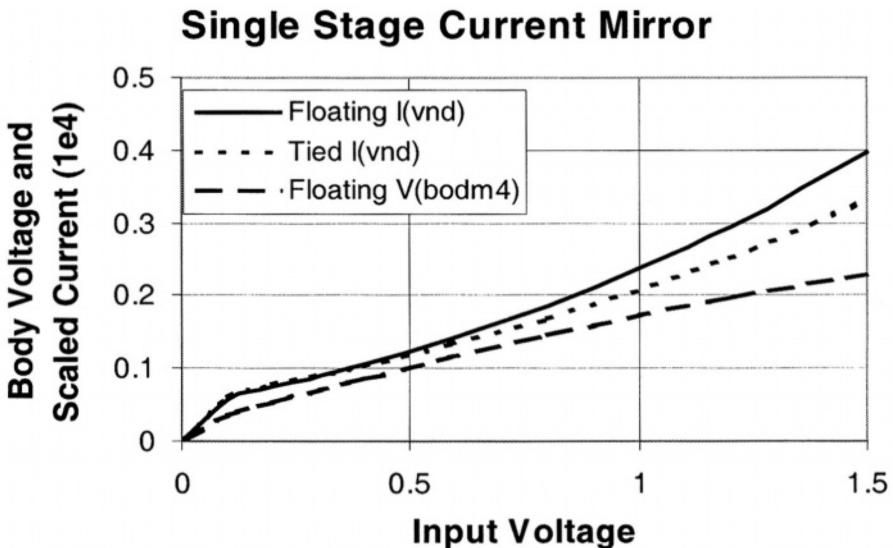


Figure 10.2(b): Current and voltage waveforms of non-cascoded current mirror configuration. Note that the SOI curve is steeper, representing poorer matching over supply voltage. However, curves for both floating body and tied body (bulk equivalent) devices have very poor mirroring.

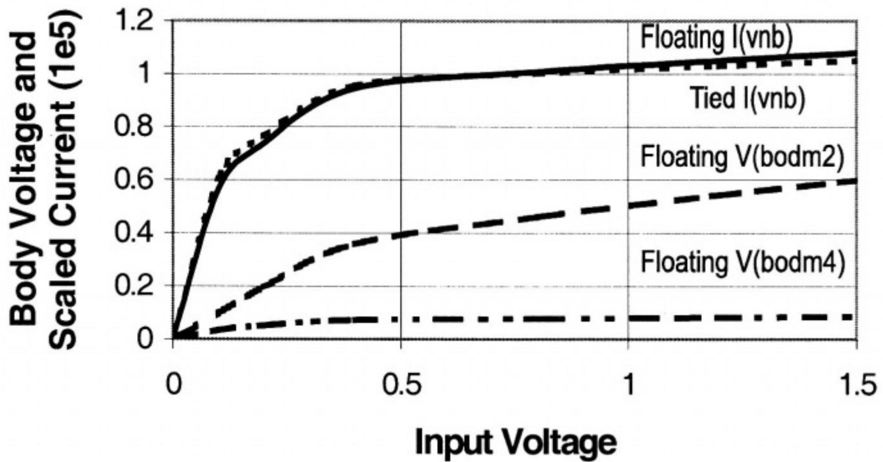


Figure 10.2(c): Current and voltage waveforms of cascoded current mirror configurations.

Cascoded circuits usually need relatively high supply voltages, or conversely output swings or input ranges are restricted. Figure 10.3 demonstrates the limitations of cascode usage in terms of supply voltage. The minimum supply voltage for accurate current mirroring is two NMOS threshold voltages (Vtn) and one PMOS threshold voltage (Vtp). This is the same as for bulk circuits with the same configuration.

10.2 Body Voltage Regulation

10.2.1 Dynamic Body

Dynamic body voltage control of PD-SOI MOS devices can be achieved through a simple connection between body and gate. However, the bias voltage limits its applicability. Active body-biasing circuits use additional circuitry to enable the increase in gate bias beyond the Vbe of the parasitic bipolar [10.1].

Figure 10.4 shows two circuit techniques for dynamic body biasing. One option is to add extra voltage, for example a forward bias diode drop, allowing the gate voltage to increase further before drawing current. No current will flow through the body path provided that the combined forward voltage is not reached. A disadvantage of the circuit is that the body voltage does not necessarily rise with the initial ramping of the gate voltage. Thus this scheme is not very efficient in fast switching environments. The other

option shown uses a current limiting component (in this case a resistor), permitting increase in the gate voltage even after the body voltage reaches the diode forward voltage drop. This has the advantage that the body voltage initially rises with the gate until it is high enough to turn on the diode. However, above this point, current flows through the current limiting circuitry and body diode, increasing operating current, and possibly limiting gate voltages.

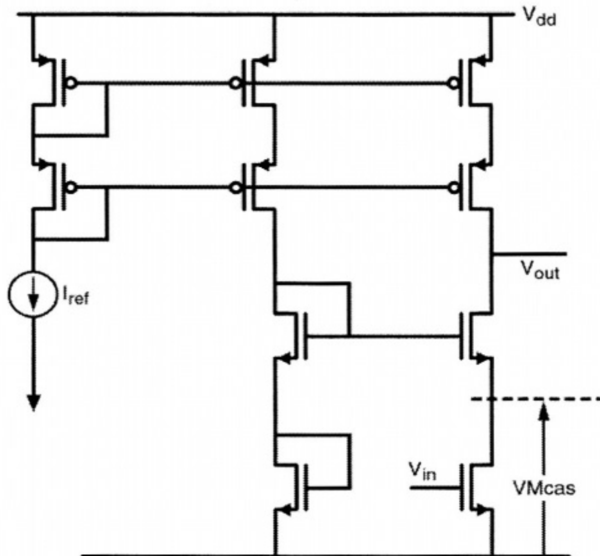


Figure 10.3: Demonstration of cascoding both PMOS and NMOS structures as a technique to avoid the kink effect.

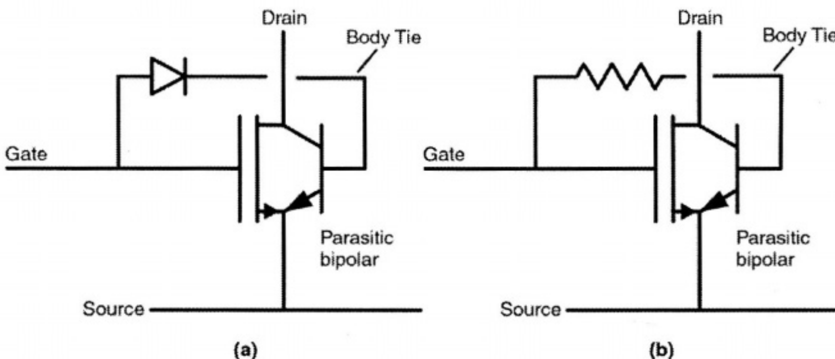


Figure 10.4: Dynamic body configurations (a) using a voltage threshold shifter (in this case a diode), and (b) utilizing a current limiting scheme (here a resistor).

Numerous other forms of dynamic body ties are possible, including a combination diode/resistor configuration between gate and body. The circuit of choice largely depends upon the application.

10.3 Circuit Thermal Coupling Effects

10.3.1 DC Thermal Coupling in Current Mirrors

Current mirror circuits rely on matched thermal and electrical conditions, usually through accurate layout techniques. The largest errors in bulk technology current mirrors are caused by electrical variations, particularly drain voltage mismatch between reference and mirror. Gate length reduction caused by technology scaling impacts matching of devices. Design changes such as increasing gate lengths to reduce channel modulation percentage, have minimized mismatch, but caused reduced circuit bandwidth. Channel modulation may also be reduced with cascode techniques (described above). This works well at voltages above 1.5V, but is more difficult to apply as supply voltages reduce. The current mirror in SOI has two further areas where mismatch may be introduced. These are the floating body and thermal mismatch.

External thermal effects can be minimized through adequate cross coupling and the application of common-centroid techniques. If the devices to be coupled can be combined within the same silicon mesa, thermal matching tends to be better in SOI than bulk material. However, thermal mismatch can still occur within the subcircuit cell and disturb the expected operation of the cell when the power dissipated on one side of a basic cascoded current mirror differs significantly from that of the other. For example, a 100μA DC current mirror dissipates approximately 100μA * Vt = 0.2V (20μW). The mirror device, meanwhile, may have a drain voltage of 1V, and with modulation effects this might be at close to 200μA current, thus dissipating 200μW. A local thermal resistance of 25000° C/W (chapter 11) would result in a local temperature rise of about 5° C. This compares to an insignificant rise in temperature in bulk material.

10.3.2 Transient Thermal Coupling in Current Mirrors

Thermal time constants for nominal size adjacent devices on SOI are around a microsecond [10.2, 10.3]. Transient switching of the mirroring device affects thermal dissipation. This in turn can couple to the reference device if the two are adjacent. Figure 10.5 shows a mirroring structure for determining transient thermal coupling. Repeating this structure with various device spacings permits the determination of short and long-range coupling.

Thermal gradients can introduce a feedback mechanism with time constants which are relatively long compared to the electrical time constants.

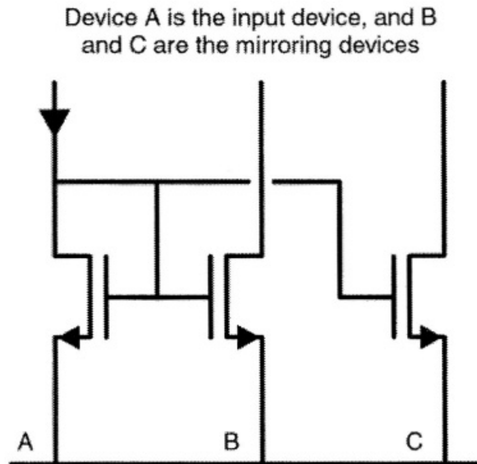


Figure 10.5: Dynamic thermal coupling configuration. This configuration can be used to determine coupling due to thermal effects, by placing multiple mirroring devices at different spacings.

A long thermal time constant introduces the possibility of unsimulated instability at frequencies less than 10-100MHz. Thus it is important to fully simulate thermal coupling in SOI circuits. In bulk silicon this effect is usually only seen in high current or high voltage structures such as power ICs. Spacing the devices apart further reduces intracell thermal coupling. However, when attempting to match for external thermal effects and process variation, increased separation is undesirable. It is worth considering placing several devices in the same tank using one of the layout schemes discussed in chapter 5 to minimize thermal mismatch, while maintaining close layout proximity.

10.4 Band-gap Designs

Most bandgap circuits use bipolar transistors or diodes as the reference element. Usually this is part of a so-called delta V_{be}, where the difference in V_{be} between two devices operating at different current densities is used to generate a reference voltage. Variations on this type of scheme, using MOS devices have been proposed, but generally have reduced accuracy. The NPNs and PNPs of a typical PD-SOI process have gains in the range of <1 to 10. Although low, it provides enough gain for some bandgap designs [10.4]. A simple bandgap structure, used frequently as a design example, is shown in figure 10.6.

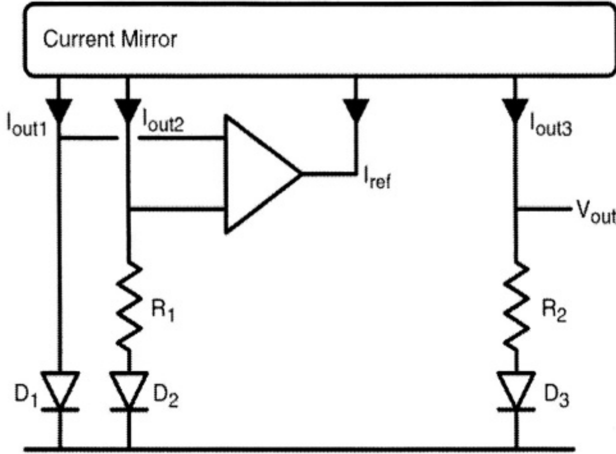


Figure 10.6: Basic conceptual bandgap circuit. Iout1 and Iout2 are usually equal. D2 is several times larger than D1, and current through R1 equalizes the amplifier input voltages. Iout1, Iout2 and Iout3 are mirrored values of the Iref current. Vout is normally the bandgap voltage of 1.2V.

The forward bias diode voltage of device D1 equals that of the larger D2 device, plus the resistor R1. Usually there are two stable solutions, the trivial state of zero current, and a current, which maintains the inputs to the amplifier at the same voltage (about 0.7V at room temperature). Circuitry (not shown) prevents the zero current option in normal operation. A typical diode area for D2 is 4x that of Dl. The operating condition is based on the following assumptions:

$V_{d1} = V_{d2} + V_{r1}$ . . . . (1)

Thus:

$V_{d1} - V_{d2} = \Delta V_d = V_{r1}$ . . . . (2)

Since $\Delta V_d = \ln 4 * KTq^{-1}$ [10.5], then V_{r1} = 36mV (at room temperature).

Therefore, for a 10μA diode current, the resistor must be approximately 3.6KΩ. A low temperature coefficient current can be obtained by tuning the resistor temperature coefficient, although this is not necessary if only a bandgap output voltage is required.

Most PD-SOI processes have optional bipolar devices available. By not using MOS devices floating body effects are clearly eliminated. However,

while they are available, most MOS SOI processes are not optimized for bipolar structures. If the beta of the bipolars being used is less than 10, some compensation for low gain operation may be required, possibly including the use of MOS beta helper devices. Figure 10.7 shows a simple bandgap circuit which uses bipolar devices. This circuit functions as follows: At equilibrium conditions the voltage across Ql and Q4 base-emitters equals the voltage across R1 and the base-emitter nodes of Q3 and Q2. Assuming Q2 is 4x the area of Ql and Q3 and Q4 are of equal size, and if the base current is small with respect to the collector current, the following equation applies:

VQ1be + VQ4be = VQ3be + VQ2be + VR1 . . . . . (3)

Neglecting base current, Ql and Q3 are part of one current path, and Q2 and Q4 are part of a different current path. The two currents are not necessarily (or even usually) equal. Re-arranging (3):

VQ3be – VQ1be = VQ4be – VQ2be – VR1 . . . . . (4)

If Q3 and Ql are the same size, then (4) simplifies to:

VR1 = VQ4be-VQ2be . . . . . (5)

This can be shown to be the same as equation (2) above.

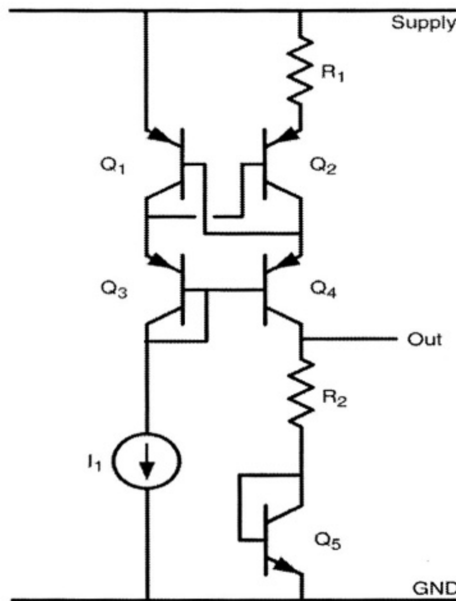


Figure 10.7: Bipolar based bandgap reference circuit

If bipolar gain is less than about 10, it is better to use MOS devices in most parts of the circuit. MOS devices can replace bipolar devices in current mirror structures, and can either be floating body devices, possibly with cascoding, or body tied devices. The bandgap circuit is a good candidate for body ties, since the circuit operates under DC conditions. Depending upon accuracy requirements, without any special design considerations, accuracy of approximately 20% is achieved with floating body SOI designs. With cascoding this improves to better than 10%. The use of body ties brings this to between 2-5%, similar to that of circuits built on bulk material. Trimming techniques allow for better than 1% accuracy.

When designing bandgap circuits for SOI it is important to account for the start-up condition of any floating body MOS devices. This is because the time constant of the body voltage is of the order of milliseconds. During startup, the bandgap itself could be expected to be fully operational within a few microseconds. However, floating bodies within the bandgap circuitry may affect the output voltage for several milliseconds.

10.4.1 β helper circuitry

Figure 10.8 shows a simple bandgap design [19.6]. This is effective if transistor gain is greater than about 50, but loses accuracy rapidly as gain reduces. This is the result of current being diverted from the collector of Ql into Ql and Q2 bases. Thus current mirrored into the output device M3 is lower than expected, and so the bandgap output voltage, BG_{out}, is low.

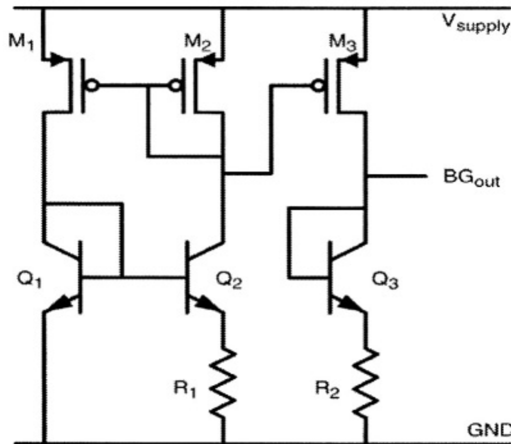


Figure 10.8: Simple bandgap circuit. This loses accuracy as the transistor gain reduces, and is thus only useful when a well behaved, high gain bipolar is available (M1, M2 and M3 are body tied to source structures).

This circuit can be adapted for low gain bipolar devices, such as those typical of PD-SOI processing, with the addition of a β-helper. β-helper devices compensate for low β bipolar transistors by providing base current from a low impedance supply (figure 10.9) [10.7]. Here, components M4, M5 and M6 have been added. Hence transistor M4 eliminates the base current losses but inaccuracy may still persist. If Q2 gain is very low significantly more current flows through Rl than through M2. To compensate for this it is possible to mirror 50% of the M4 current back into Q3 and R2. This circuit works with very low gain bipolar devices.

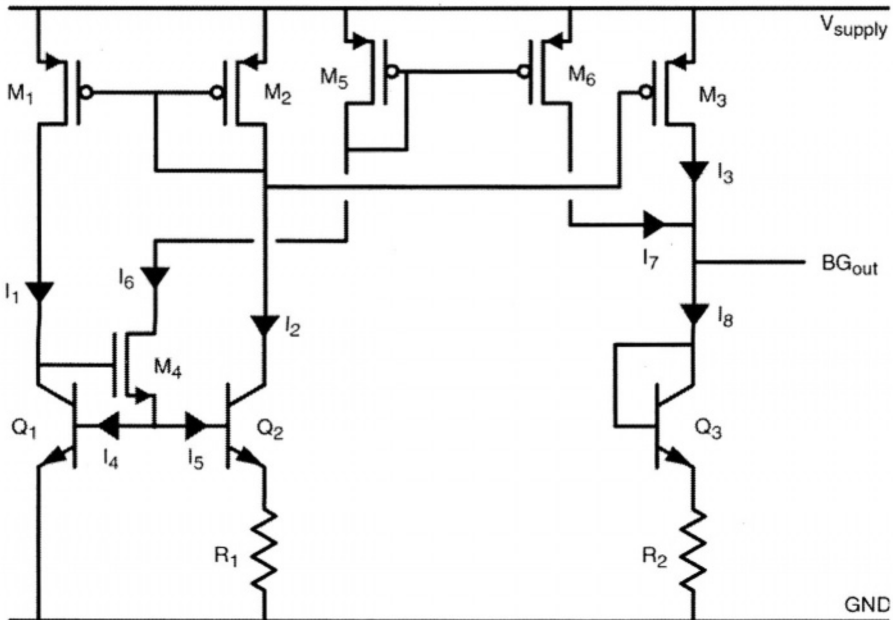


Figure 10.9: Compensation scheme for low gain bipolar structures such as those typically available in PD-SOI processes (here M1, M2, M3, M5 and M6 are body tied to source structures, M4 can float).

In the simplest case we assume that base currents I4 and I5 are equal

Therefore, I6 = 2*I5 = 2*I7

With an ideal current mirror, I1 = I2 = I3

Therefore Vbgout = Vbeq3 + (R2 / R1) * Δ Vbe

With appropriate sizing of R2 and Rl the bandgap output voltage, V_{bgout}, can be made virtually constant over typical operating conditions. Figure 10.10 shows output voltage over temperature for the circuit in figure 10.9.

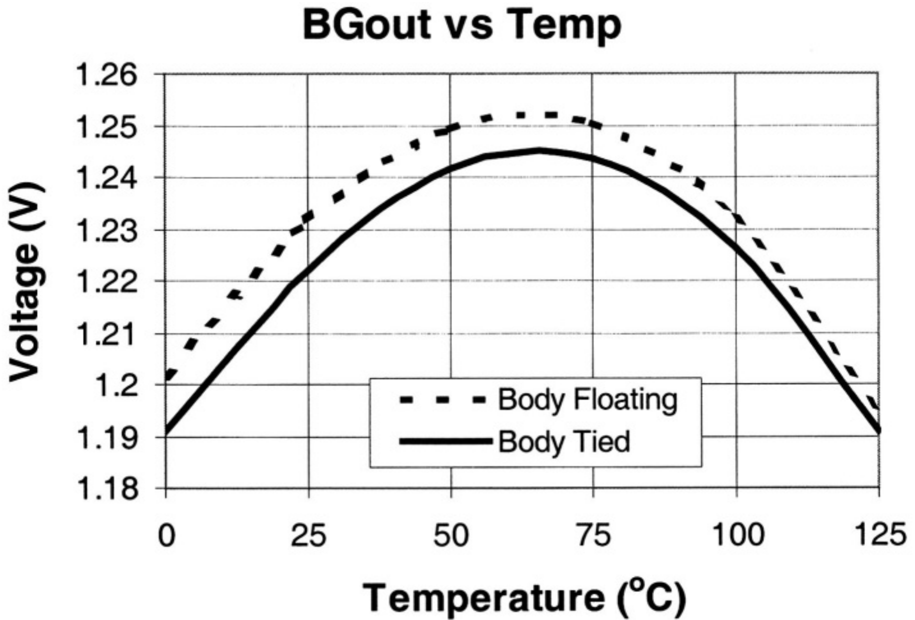


Figure 10.10: Output voltage waveforms for floating body and body-tied versions of the bandgap circuit shown in figure 10.9.

10.4.2 Voltage Reference using "Threshold Voltage Difference"

The "threshold voltage difference" principle makes use of the difference between gate-source threshold voltages of two MOSFET devices with different Vt implants [10.8]. This difference remains almost constant, independent of temperature, current or body voltage. The threshold voltage difference principle can be used for creating voltage reference circuits that operate up to 275°C [10.9].

The circuit is shown in figure 10.11, and is based on an operational transconductance amplifier (section 10.7.4). Device Ml has a higher threshold voltage than M2. Feedback through the resistor divider (Rl and R2) enables the generation of a bandgap voltage that is relatively constant. Optimum performance is achieved with body tied components, particularly for Ml and M2, although body tying M3 through M6 is advisable. No significant benefit is gained from body tying M7 through M11.

The output voltage V_{bg} is defined as:

$$V_{bg} = ((R2+R1)/R1) * (VtM1 - VtM2)$$

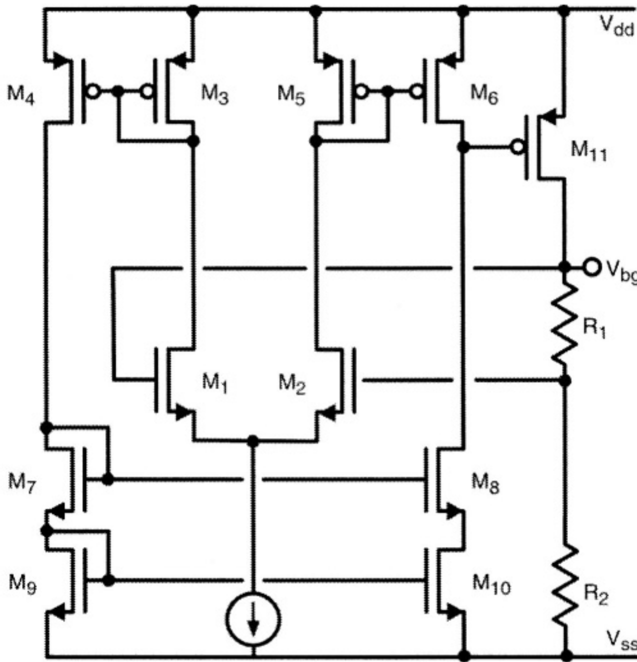


Figure 10.11: Threshold Voltage Difference type of bandgap reference circuit. This structure can be built on bulk silicon or SOI. Resistor matching may be better in some cases in SOI, but body voltage of the input pair M1 and M2 needs to be adequately grounded, or tied to source for optimal SOI performance.

10.5 Charge Pump Circuitry

Charge pump circuits are used to provide a voltage outside the normally available power supply rails, for example a voltage higher than supply, or a reversed polarity voltage. A common application is biasing the pass gates of DRAM devices (chapter 9). Figure 10.12 shows a typical charge pump design for generating a voltage above the supply. A 50% duty cycle clock provides the switching input, often operating at a frequency around 1-50 MHz. A signal 180° out of phase with the clock input is generated with an inverter.

Each of the MOS devices Ml through M4 acts as a unidirectional switch. An unloaded output voltage of over 3V is attainable, with a supply of 1V and a MOS threshold voltage of 0.2V [10.10]. This type of circuit is often easier to implement in SOI than in bulk, as there is complete isolation of the MOS devices. Isolation of the devices eliminates the possibility of latchup and reduces the number of leakage paths. No body ties are required in this circuit

design; indeed adding them may cause unwanted leakage in some modes of operation (refer also to chapter 9).

Figure 10.12: Voltage quadrupler charge pump configuration. The MOS devices in this configuration are completely isolated in SOI, thus unlike their counterparts on bulk silicon there is no risk of latchup.

10.6 Amplifiers

10.6.1 Sense Amplifier

Dynamic SOI sense amplifiers for memory applications are discussed in detail in chapter 8, but are considered here briefly in the context of general analog design. Sense amplifiers on PD-SOI are susceptible to body voltage induced variation and mismatch, especially if the read operation is repeatedly performed in one state [10.11]. This can cause asymmetric charge accumulation, resulting in read failure. One solution is to tie the bodies to source or supply (Vss for NMOS, Vdd for PMOS). This, however can lead to undesirably high threshold voltages, and reduce the operating range of the amplifier.

An option for equalizing body voltages is to flood the input devices with charge prior to read [10.11]. Body voltage rises, limiting at the voltage where the parasitic bipolar turns on, which is a similar voltage for both devices. Small PMOS devices are used to charge the body-contacted NMOS devices, as in figure 10.13 (This can also be applied to PMOS input pairs, by

using NMOS charging devices). The charging devices (PM1 and PM2) are turned on at some time before the amplifier is to read. After both bodies have reached equilibrium PM1 and PM2 are turned off. This has to occur just long enough before the 'read' for the parasitic bipolar to turn-off. The minimum time between turning off current to PM1 and PM2, and the 'read' is defined by the base (body) minority carrier recombination time. Minority carrier recombination time is typically less than $1\mu s$. Providing the read is done while the input pair body voltages remain in a similarly charged state, improved matching is achieved at reduced threshold voltage. The body modulation technique can be used for any comparator input, providing continual read of the input is not required, as there is the potential for the charging phase to affect the amplifier output state.

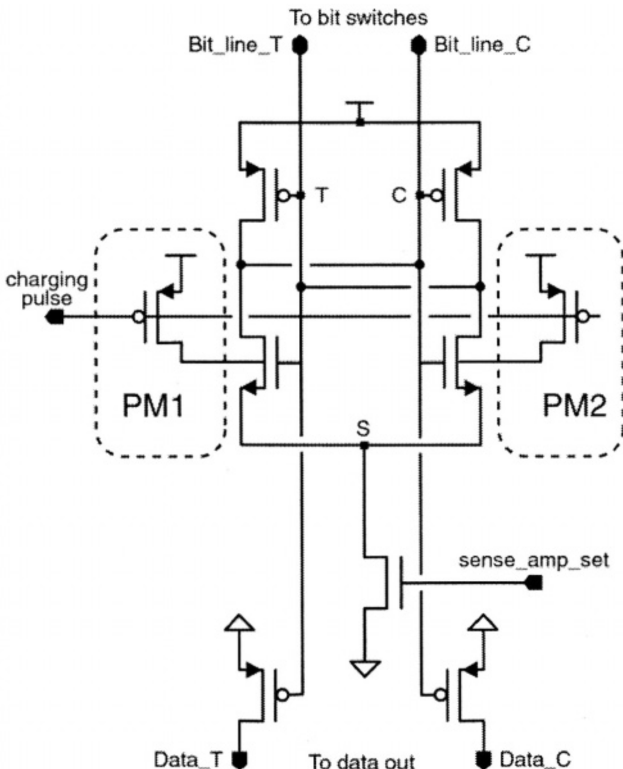


Figure 10.13: Schematic showing the use of body-charge modulated concept for a RAM sense amplifier.

10.6.2 Operational Amplifiers

A simple topology used for a CMOS operational amplifier (op-amp) is shown in figure 10.14(a). This uses a matched pair of input devices, with the output brought being a simple class A design. The CMOS op-amp does not

have the gain, offset or frequency of its bipolar counterpart, but has gained acceptance due to ease of integration into mixed signal IC designs.

The voltage gain of the basic operational amplifier can be determined by considering the input and output stages individually. Assuming NMOS devices Ml and M2, and PMOS devices M3 and M4 are matched, the small-signal voltage gain is [10.5]:

$$A = - gm1gm6/((go2+go4)(go6+go7))$$

Where gm is the device transconductance and go is the device small-signal output conductance. In PD-SOI there is no guarantee that an assumption of matched devices are, in fact, matched. Even with bulk processes variation can occur due to channel modulation, and with PD-SOI that is compounded by the addition of the kink effect. Thus this circuit is of limited applicability with SOI technology.

Figure 10.14(a): Schematic for a PMOS input, CMOS operational amplifier. This circuit is prone to offset problems when used with SOI processes.

Figure 10.14(b) shows a simulation of the body bias of the circuit shown in figure 10.14(a). Inputs in1 and in2 represent the input voltages, and bv1 and bv2 the corresponding input pair body voltages. As expected the body voltages vary depending upon the applied input voltage. In particular, body voltage 2 follows the rising large-signal input, until the switching threshold is reached. After the input MOS2 turns on, the body voltage is clamped at the source/drain voltage, collapsing the body voltage. Close to the crossover threshold the body voltages are closer, but even here there is an offset, due to the non-symmetry of the circuit and input ramp rate.

Figure 10.14(c) shows the same circuitry, simulated with a very slow ramp. It is evident that the ramp speed significantly affects the body voltage of the floating input pair. Analysis of the crossover threshold also shows that this is significantly impacted by the transition speed, which can alter the threshold point by up to 25mV in this circuit.

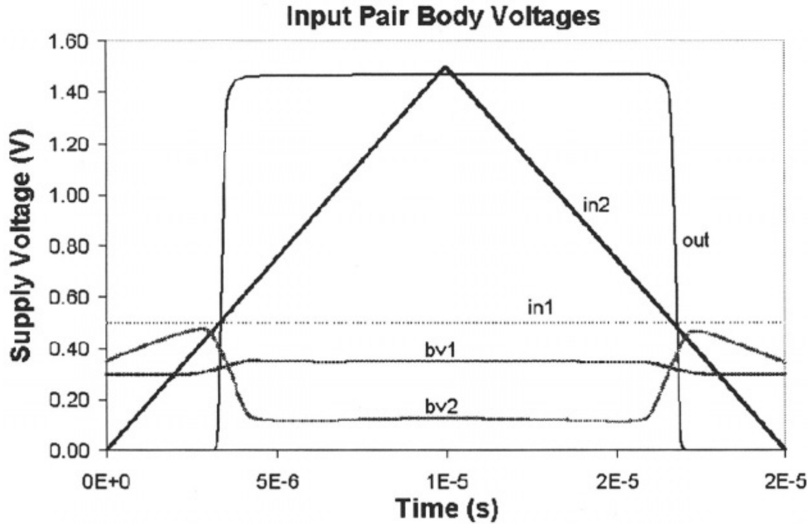


Figure 10.14(b): Simulation of the body voltage variation for the circuit shown in figure 10.14(a).

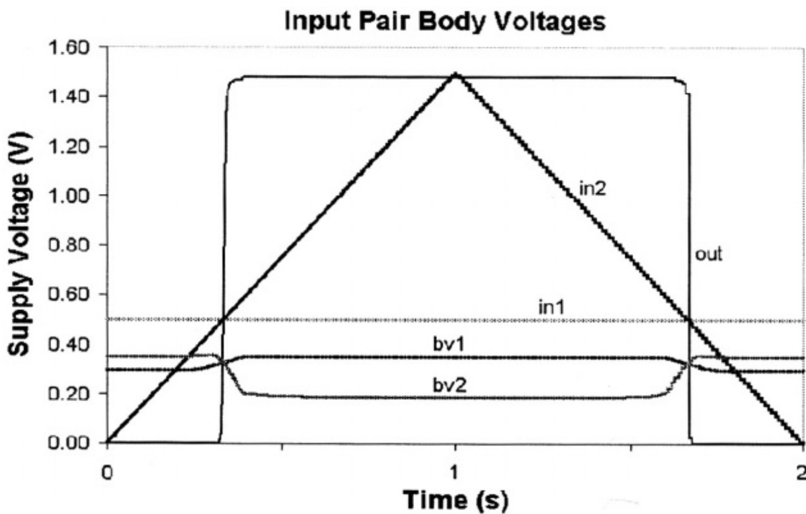


Figure 10.14(c): Simulation of the body voltage variation induced in the circuit shown in figure 10.14(a), for a very slow ramp. This is slow enough that capacitive coupling of the input ramp does not affect body voltage.

Cascoding within the op-amp has some advantages: Voltage gain can be improved, as a result of increased output resistance of the configuration. High frequency capacitive feedback is also reduced. Cascoding also helps to maintain drain voltages, which stabilizes the body-voltage, and minimizes the gain slope with respect to variations in supply voltage. A cascoded op-amp is shown in figure 10.15. The PMOS input transistors M1, and M2 are cascoded using NMOS devices M3 and M4. This circuit has a higher gain than the circuit shown in figure 10.14, and is also less susceptible to body effects. However, the output voltage swing is degraded compared to non-cascoded schemes. Additional cascoding of the input pair further restricts the voltage variation at the source-drains of the input pair (figure 10.16(a)), limiting the leakage induced body voltage drift.

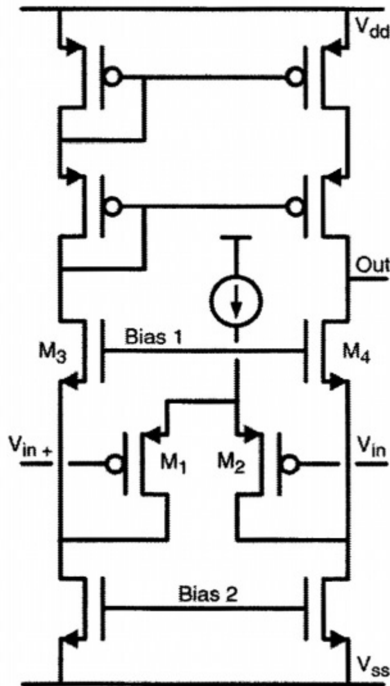


Figure 10.15: Cascoded opamp design that holds the input pair output voltages at approximately a constant voltage.

Simulated waveforms showing input, output and body conditions for the circuit are shown in figure 10.16(b) and figure 10.16(c). These indicate the improvement that can be made to body voltage stability, resulting in improvements to the offset over all conditions.

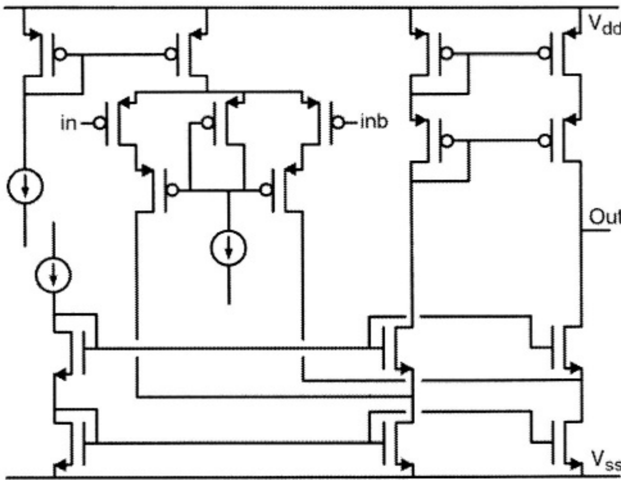


Figure 10.16(a): Further circuit modifications restrict the drain-source voltages of the input pair, inhibiting leakage dependent voltage drift.

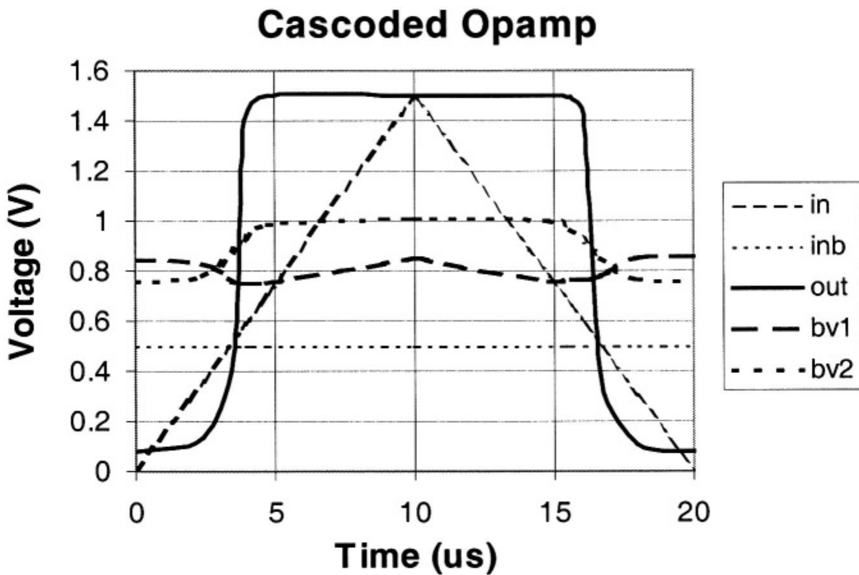


Figure 10.16(b): HSpice Simulation of the body voltage variation for a cascoded input operational amplifier, with a fast input ramp.

If a high enough power supply voltage is available telescopic designs help maintain the DC body voltage characteristics. For example, the circuit in figure 10.17 requires a voltage of approximately three NMOS threshold voltages and two PMOS threshold voltages as the minimum supply voltage.

Even this circuit is prone to some variation due to the non-cascoded M9 and M10 current mirror.

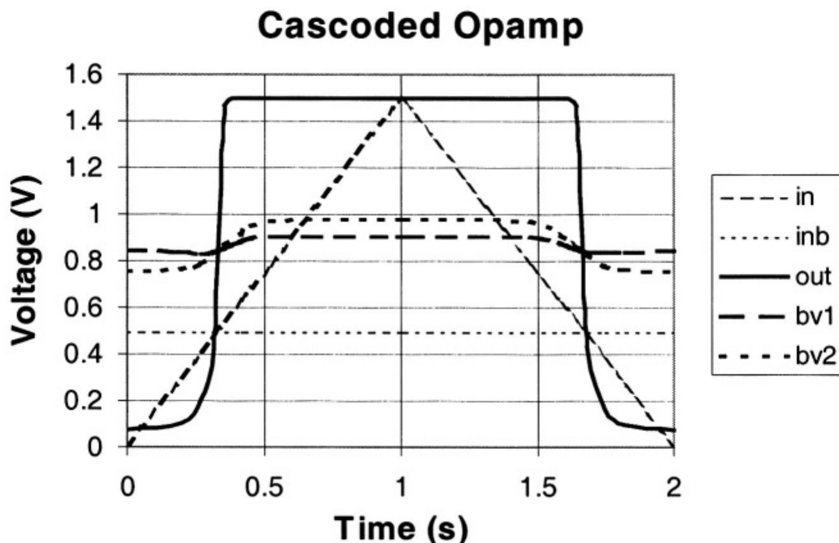


Figure 10.16(c): HSpice Simulation of the body voltage variation induced in a cascoded input operational amplifier, for a very slow ramp of 1.5V/second.

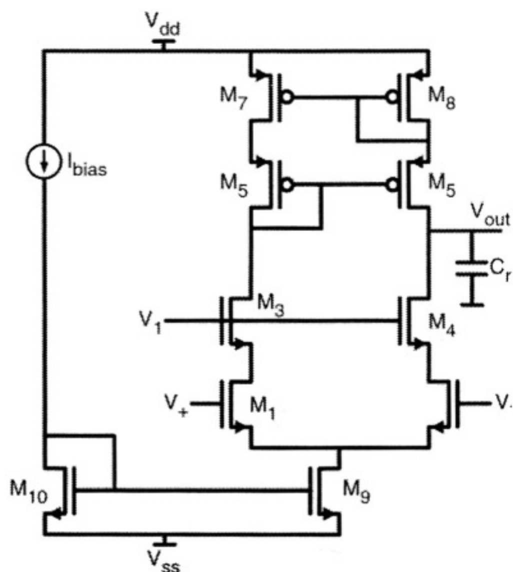


Figure 10.17: Example of a telescopic op-amp design. The drawback of this type of design is its requirement for a supply voltage higher than permitted in many SOI processes.

10.6.3 Operational Transconductance Amplifier (OTA) Design

Even the basic OTA design has reasonable component matching to ensure balanced operation (figure 10.18). This is particularly useful in SOI applications. As a result high frequency OTAs have been successfully produced in SOI [10.12]. Technology improvements can boost the operating frequency of transconductance amplifiers above that normally obtainable from bulk material, through reduction of source/drain-to-substrate capacitances [10.13]. Folded-cascode OTAs have also been implemented in SOI.

OTAs for switched-capacitor applications drive a capacitive load and require a high-impedance output. In addition they have to be stable in unity-gain feedback mode. This requires good modeling capability. To ensure stability the second pole of the circuit performance must be higher than the maximum threshold frequency, to provide adequate phase margin and ensure stability. For frequencies above about 100 MHz, transconductance and capacitances are highly frequency dependent. Most circuit simulations substitute a lumped gate resistance model for the physically distributed gate resistance. For high gate resistance systems operating at high frequencies and where the device transconductance frequency limit is included, the lumped gate resistance model is not valid.

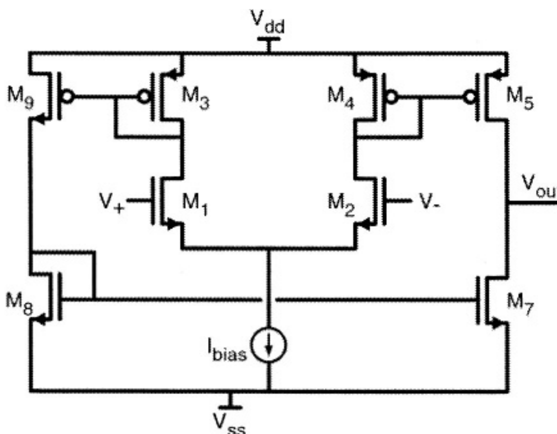


Figure 10.18: Basic, single-stage Operational Transconductance Amplifier.

Examples also exist of OTA designs for radiation-hard environments [10.14]. This structure uses a CMOS SOI process using "gate-all-around" (GAA) devices [10.15]. Most GAA processes require two additional process steps over standard SOI to create a cavity under the transistor areas (chapter

13). The single stage OTA is reasonably good for SOI applications, since the voltages are matched on all stages except the output stage. Further improvement can be achieved through cascoding of the NMOS current mirror.

10.7 Matching

Some component matching techniques in SOI technology are similar to those in bulk silicon. SOI devices, in certain applications, are less sensitive to supply voltage variation, which can cause problems with circuits using bulk material. This is mostly in components requiring specific non-supply or ground body biasing, or tank resistor biasing, which, when diffused into bulk silicon, may cause device differences.

Matching of MOS devices uses the same techniques in SOI and bulk (figure 10.19). A cross-coupled pair of nMOS devices is used, all being orientated similarly. Device '1' is a combination of structures 'a' and 'c', and device '2' of the matched pair is a combination of structures 'b' and 'd' (body ties not shown). This structure has good immunity to external thermal and process related gradients, but does not account for local device heating within the cross coupled configurations.

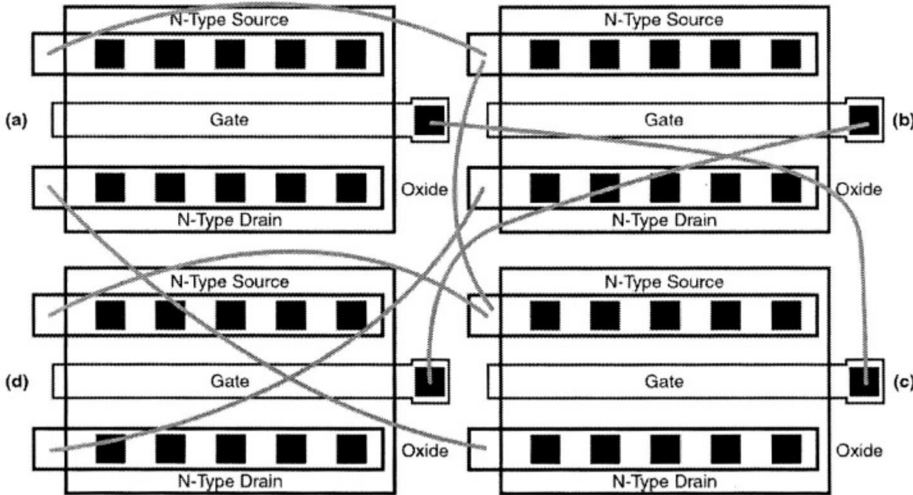


Figure 10.19: Cross coupling technique, showing coupling method for source, drain and gate, for a common source input pair.

Drain conductance varies with frequency as a result of self-heating [10.16]. This implies self heating impacts the behavior of small-signal amplifying

stages. Gain versus frequency plots show self-heating causes a gain change at low frequencies. In addition simulations demonstrate early roll-off, introduced by thermal time constants.

10.8 Output Stages / Buffers

Output stages may be hundreds or thousands of times larger than the width of a device used for core logic. It is worthwhile allowing the body to float, in order to achieve a higher current density capability during switching. This, however, can add to quiescent current leakage.

While it is not universally applicable, due to specific circuitry requirements, to reduce leakage in output structures it is often worth using a body tie configuration that is at ground potential in standby conditions. This achieves low standby leakage, but permits the body to float during operation. Positive body modulation during operation improves current density (figure 10.20).

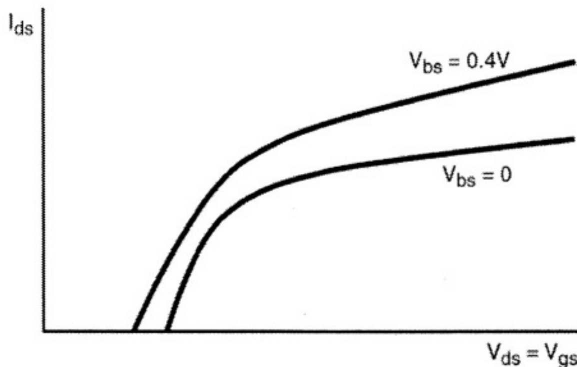


Figure 10.20: Graph showing the Ids variation with body modulation, illustrating improved performance in the operating state.

The two effects can be combined by using a body-tied-to-source state for standby and during operation tying the body to a somewhat positive potential for an NMOS (or negative with respect to supply for a PMOS). This gives the required drive current capability and the potential for standby current minimization through reduction of the output device size. An example of this combination of a PD-SOI buffer circuit with dynamic body biasing is shown in figure 10.21.

When V_{in} is switched low, the MN1 output device is switched on, and its body is biased high by the low current MP2 device. When V_{in} is switched high, the MN1 body floats. The reverse mechanism applies for the MP1 output driver.

Figure 10.21: Buffer output stage with dynamic body biasing.

10.9 High Voltage and Power Applications

SOI has significant advantages over bulk silicon when high voltage circuitry is required. Due to component isolation in SOI material, it is easy to design high voltage devices without adding excessive process complexity or cost. It is also simpler to avoid high voltage parasitics caused by inversion of the active region. High voltage structures, including the DMOS power transistor, have been discussed in chapter 3.

Output stages usually drive high voltages and/or currents when they interface to the outside world. Power dissipation can significantly increase device temperature, reducing performance and efficiency. It is generally considered important in SOI design that simulations of output structures are completed with the thermal dissipation (temperature) node activated (chapter 5).

10.10 Sample and Hold Circuitry

The design of sample and hold circuitry on SOI poses some specific design problems (figure 10.22). The input and output buffers require the same considerations as operational amplifier designs. The electronic switch is usually a MOSFET. In the case of PD-SOI the risk of leakage due to the floating body dictates that the body should be tied to the appropriate power supply rail in most cases. SOI based sample and hold circuits have reduced stray capacitance associated with the electronic switch and therefore

contribute reduced capacitive coupling offset. This can permit smaller size storage elements. Conversely, with thinner gate oxides, gate leakage may become an important consideration in circuit design (in both bulk and SOI designs) (refer also to chapter 6).

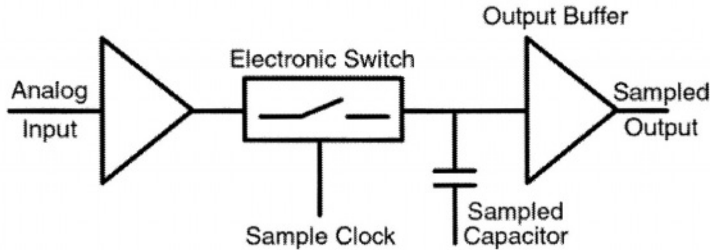


Figure 10.22: Block diagram of a sample and hold circuit. The input signal is buffered before passing through an electronic switch (usually a MOSFET). A storage capacitor maintains the voltage, which is buffered through a high- input-impedance unity-gain op-amp.

10.11 Circuits for RF/Wireless Applications

SOI is used for RF applications in the low GHz range [10.17], and the need for low-cost, low-power system-on-a-chip applications has led to investigation of SOI as a substrate for wireless applications [10.18]. High-resistivity substrates improve inductor Q values to a point where they are adequate for RF applications.

The RF-transmitter may use SiGe bipolar, silicon bipolar or sometimes MOS devices, thus combined bipolar and CMOS SOI is useful for full integration of analog and digital functions on one chip. SOI also provides isolation between the digital and analog circuitry, and input and power output sections of the chip. MOS devices on SOI have superior RF performance because of the inherent device oxide isolation [10.19, 10.20].

A direct-conversion architecture for receiver and transmitter is typical. [10.21]. Delta-Sigma modulators have been constructed in SOI [10.22], creating a mixed signal SOI CMOS process. Direct up-conversion transmitter architectures are commonly used. The baseband signal is mixed with a local oscillator (LO) at the transmission frequency (figure 10.23).

The route between antenna and low noise amplifier is termed the receive path. Once amplified, this signal is mixed with a locally adjusted oscillation frequency (shown here as a VCO). If an intermediate frequency is used it is often achieved with a fixed frequency local oscillator.

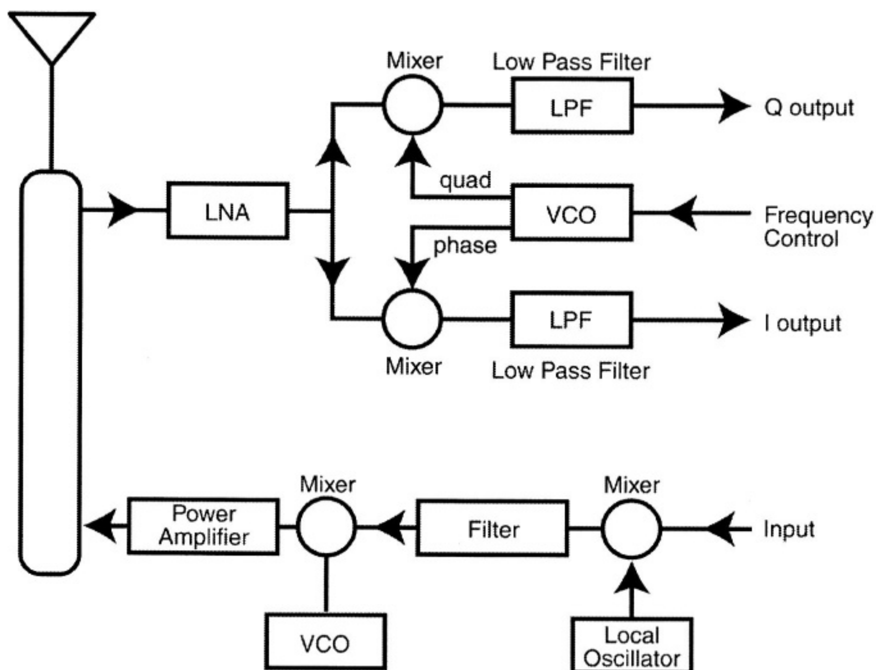


Figure 10.23: Wireless transceiver architecture. This block schematic shows that most circuit blocks are made of simpler elements, described elsewhere in this chapter.

Wireless architectures are made from many simple building blocks, such as low noise amplifiers, mixers and voltage controlled oscillators. Most of these are described within this chapter.

10.11.1 Radio Frequency Low Noise Amplifier (LNA)

SOI makes the task of integration of a single chip wireless transceiver very much simpler. The reduced capacitance of SOI over bulk MOS devices results in higher frequency capability. SOI provides the option of isolated high frequency bipolar transistors and also permits isolation between all types of devices. Furthermore, high resistivity, low-loss dielectric substrates enhance the quality factor of inductors. The frequency limit of operation in monolithic integrated circuits are defined by both the frequency capabilities of the transistors and the properties of the passive components [10.23].

A two-stage radio-frequency amplifier topology designed in SOI is shown in figure 10.24 [10.24]. On-chip inductive source degeneration is used for improved impedance matching and optimized noise figure. This topology has been demonstrated to operate with a 1V supply.

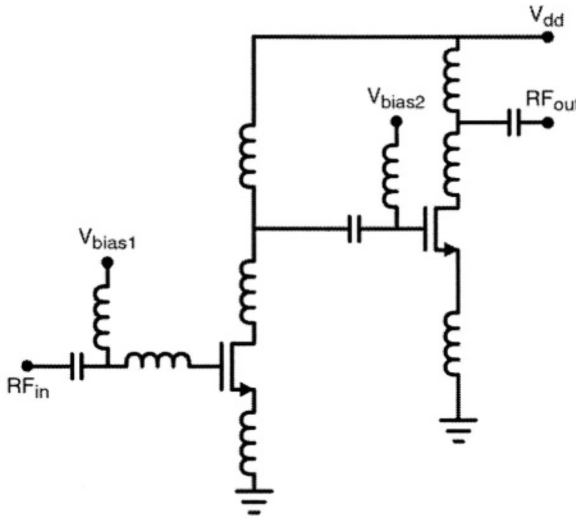


Figure 10.24: Two transistor LNA, designed for SOI. The passive components, in particular the inductors, benefit from a high resistivity substrate.

A single stage LNA design is shown in figure 10.25. This comprises a similar input configuration to the two stage structure shown above, with a cascode device in place of the second gain stage used in the previous example. Input inductor L_g is directly connected to the gate. Parasitic capacitances of the bond-pad and inductor impact operating frequency limit and noise figure, and must be considered in simulations. In addition, third-order intermodulation is observed as a result of non-linearity in the current-voltage characteristics.

Distortion sources in MOSFETs are transconductance non-linearity, voltage dependence of the drain junction to body, and output conductance non-linearity [10.17]. While transconductance non-linearity is common to circuits built on both SOI and Bulk, those on SOI can be designed to minimize the voltage-dependent drain capacitance however the "kink" effect is significant in DC output conductance non-linearity.

The cascode configuration increases output impedance and reduces the SOI MOS output conductance non-linearity [10.17]. Use of a high resistivity substrate for the LNA improves performance by increasing the inductor Q factor, and shifts the peak frequency of the Q factor up from just over 1GHz with 10 Ω cm material to over 2GHz with 1000 Ω-cm material.

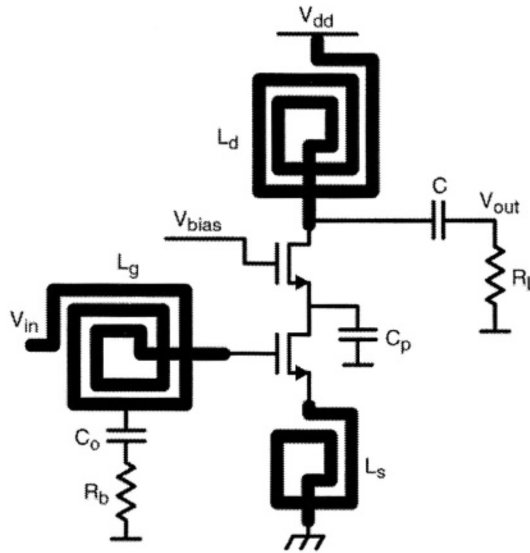


Figure 10.25: Radio Frequency Low Noise Amplifier, with cascoded output stage.

10.11.2 Mixers and Analog Multipliers

Analog multipliers combine or generate non-linear signals, such as multiplication, division, square, square root etc. They are useful as building blocks in a variety of circuits, including adaptive filters, function generators and wireless systems. They are particularly prone to mobility reduction effects and body effect [10.25]. Body effect mismatch is reduced in PD-SOI, since the NMOS body is not automatically tied to substrate, but can instead be tied to source, maintaining low offset between devices.

Gilbert Mixer

Most wireless systems use some variant of the familiar Gilbert mixer (figure 10.26)[10.26]. Mixers are among the highest power dissipation circuits in most wireless systems. However, for wireless systems, it is also a requirement to minimize power, since this allows reduced battery size, improved operating capability and lower standby current. It is also important to maximize the level of circuit integration for reduced cost and improved performance. CMOS on SOI is particularly suitable for helping achieve all of these requirements. For wireless systems, mixers are required on both the receiver, to convert the RF signal to lower frequencies, and on the transmitter, to mix low frequencies with the RF carrier. Frequently, the most important characteristics of mixers in wireless applications are high linearity, low noise and high gain.

Figure 10.26: Schematic of a CMOS-based Gilbert mixer.

The conventional Gilbert mixer structure on PD-SOI has the drawback that local matching is usually inadequate, and that supply voltage headroom may be marginal for many low voltage SOI processes. SOI mixers benefit from reduced capacitance, aiding high frequency applications while reducing power and providing improvements in passive components.

The simple dual gate mixer shown in figure 10.27 will operate to 1V or less, and in SOI the low substrate parasitics result in high performance [10.27]. To operate at 1V, this scheme uses zero threshold voltage devices. If higher supply voltages can be tolerated, higher threshold voltages may be used. In figure 10.27, the lower device (M1) is the mixing device, and is biased in the linear region. The upper MOS device is biased in saturation and provides amplification. Careful biasing is required to avoid kink-region operation, where noise and non-linearity are increased.

10.11.3 Delay Locked Loop

Delay Locked Loop (DLL) designs are often used for frequency multiplication applications. This is achieved by taking advantage of an inherently low-jitter frequency reference (such as a crystal oscillator), and generating between each reference pulse a burst of evenly spaced edges. The bursts are adjusted in frequency to form a continuous higher frequency, with a higher long-term stability than possible without the reference signal.

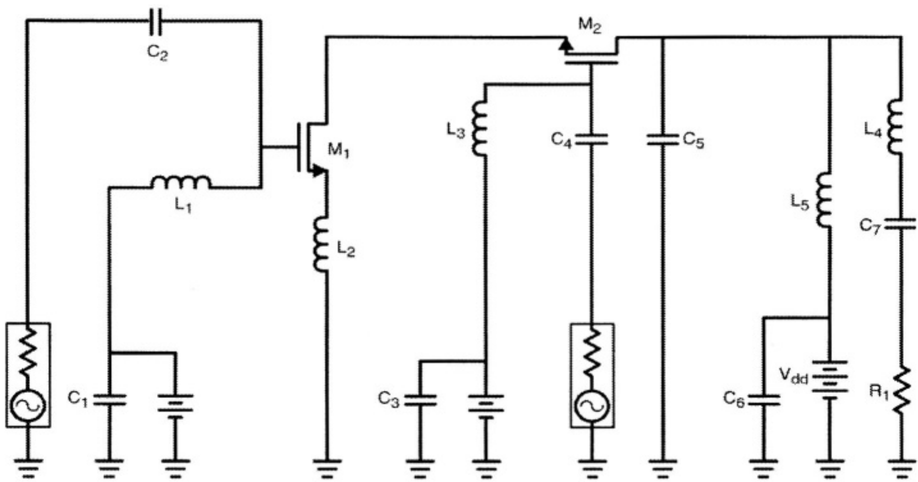


Figure 10.27: Two-transistor dual gate mixer, designed for a PD-SOI process, and operation at 1V.

To generate controlled, evenly spaced pulse edges a voltage controlled delay line is formed, using as its basis a voltage controlled ring oscillator that has been split to form a chain. A phase detector senses the phase shift between the input and output, generating an error signal, which is used to adjust the delay per stage through the chain (figure 10.28). DLL systems have better long-term jitter performance than PLL designs as timing error is reset with each reference pulse.

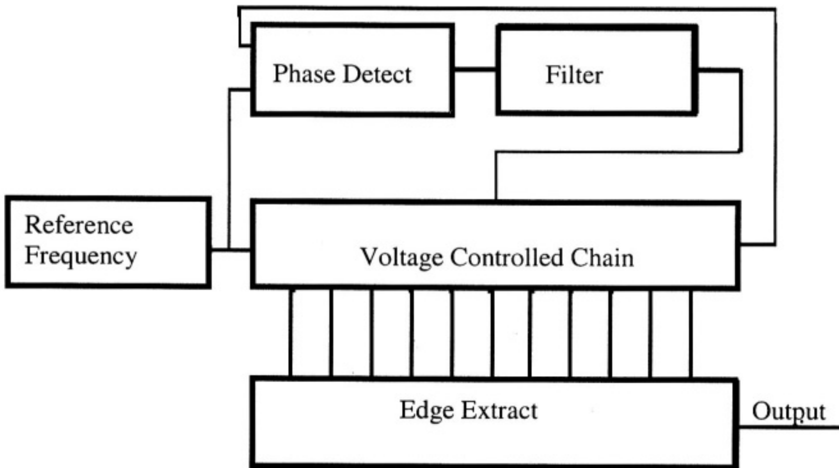


Figure 10.28: Typical Delay Locked Loop.

A DLL for graphics display applications has been compared for performance in bulk, PD-SOI and FD-SOI [10.28]. The FD version displays the shortest minimum delay and highest precision per stage, and bulk the longest minimum delay per stage. History effect of the SOI material did not cause a significant increase in jitter.

10.11.4 Phase Locked Loop

Phase Locked loop (PLL) systems are used extensively in wireless and communications systems particularly frequency synthesis and demodulation. The PLL consists of a filter and amplifier, with a phase detector comparing the input signal, and a frequency dependent signal derived from the output voltage (figure 10.29).

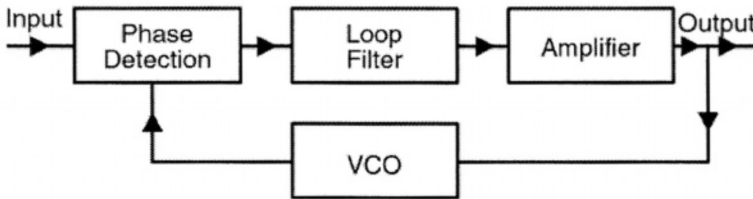


Figure 10.29: Typical Phase Locked Loop block diagram.

The reference and feedback are compared using a phase detector. The input of the loop filter and amplifier is an error signal from the phase detector. Synchronization is maintained as the output amplification completes the feedback loop through the VCO. The ideal steady state condition is a zero phase difference between the input and VCO output.

Various forms of jitter, or cycle differences, caused by noise or spurious signals limit PLL performance. Difference between successive periods causes cycle-to-cycle jitter. Phase jitter is related to modulation of successive pulses referenced to a continuous oscillator. Period jitter relates to the variation of period during non-successive periods. Random jitter is the result of noise components such as flicker noise or thermal noise and spurious signals. Additional device noise coupling from floating body structures, and spurious signals from the supply are critical issues to be addressed in SOI design. However, there is also a reduction of noise coupling through the substrate.

10.11.5 Phase Detector

The phase detector compares the phase of the reference signal to the feedback signal from the output, and generates a signal that is proportional

to the phase error. Digital phase detectors include edge-triggered flip-flops and exclusive OR gates. It is necessary to ensure that the phase detector accounts for the memory effect of PD-SOI. If the phase detect is operating in a stable state, any memory effect should also stabilize, however, this may take several milliseconds, especially at circuit power-up.

10.11.6 Loop Filter

The simplest form of filter used in phase lock loop systems is the passive analog type. This includes resistor-capacitor (RC) and inductor-capacitor (LC) filters. Most passive filters require some internal damping to prevent ringing following a frequency step. Many fully integrated filters have been of the RC type, due to the ease of construction and the low 'Q' factor associated with integrated inductors on bulk material. The potential for improved 'Q' with SOI shifts the point where it becomes economical or even possible to integrate the inductor. Off-chip filters are often of the LC type. On chip active analog filters are also widely used, and digital filter schemes are gaining popularity due to their capability for simple re-programmation for multi-standard wireless specifications. Hybrid filters, consisting of a reprogrammable digital filter and a relaxed specification for the analog filtering, may also be used.

10.11.7 Oscillators

Three types of oscillator in common use for VCO applications are the LC resonant circuit (for example the Colpitts oscillator), oscillators based on capacitive charging/discharging (often referred to as relaxation oscillators) and delay-based oscillators (such as the ring oscillator). A merged voltage-controlled oscillator and RF receiver have been demonstrated on the "silicon-on-anything" version of SOI [10.29].

Inductive resonant circuits such as the Colpitts or Clapp oscillator [figure 10.30] are favored in many applications due to their frequency stability. The circuit operating point is defined by the LCR resonant frequency. Inductive resonant circuits have proved difficult to integrate successfully in bulk material due to limitations of integrated inductor design. New techniques, applicable to SOI, have improved the on-chip inductor characteristics (chapter 3). These oscillators can be converted into VCOs by replacing the fixed capacitor with a varactor. Depending on circuit requirements, these can be optimized for linearity, Q and phase noise. Most varactors have a limited tuning range, unless several volts can be applied. One method to circumvent the tuning range limits, is to use several smaller tuning ranges with switchable varactors.

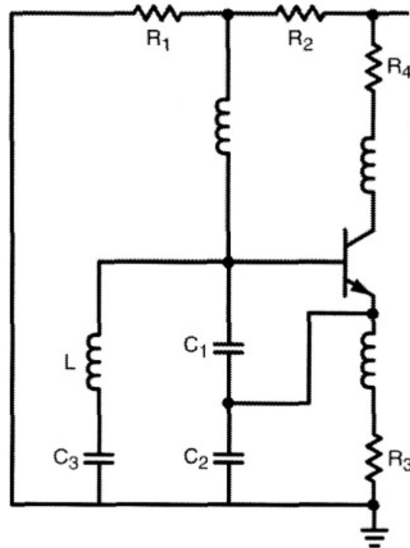


Figure 10.30: Typical Clapp Oscillator. This type of oscillator is more stable than ring or relaxation oscillators, a requirement in some wireless systems.

Relaxation oscillators (or multivibrators) are common oscillator designs in ICs. These operate through a continual charging and discharging of one or more capacitors between two voltage levels. A multivibrator is readily controllable using a voltage-controlled current source that supplies the charging and discharging currents (Figure 10.31). Basic relaxation oscillators tend to be sensitive to thermal effects, supply voltage variation, and phase noise. Design techniques such as temperature stabilized current supplies and controlled voltage thresholds aid in minimizing these effects.

The resistors in this circuit may be implemented as diffused components or polysilicon. For large resistances it is usual to use diffused resistors. In bulk these are fabricated straight into the substrate. Isolation from the substrate is achieved through a reverse biased diode junction. As reverse bias increases (voltage on the resistor increases) the resistance per unit length increases. This can lead to unexpectedly low switching threshold voltages. This is not a problem in SOI, since resistors are electrically isolated from the substrate.

Assume the capacitor is discharged, voltage at the top plate of the capacitor, V_c, is then close to ground. The output will then be close to V_{DD}. This permits charging of the capacitor through the controlled current source "I-charge". When the capacitor has charged above the level of V_{Low}, defined by a resistive divider, comprising R1, R2 and R3, the comparator output

switches to a level 1, thus inputs to the latch are both at a level '1'. When the capacitor has charged further, its potential reaches V_{high}. This forces comparator 1 to switch to level 0, changing the output state of the oscillator. The capacitor then begins to discharge. As soon as the capacitor voltage reduces below Vhigh the output of comparator 1 again switches high, without affecting the output voltage. When the capacitor voltage falls to Vlow, comparator 2 switches, changing the output state back to high. The cycle is repeated.

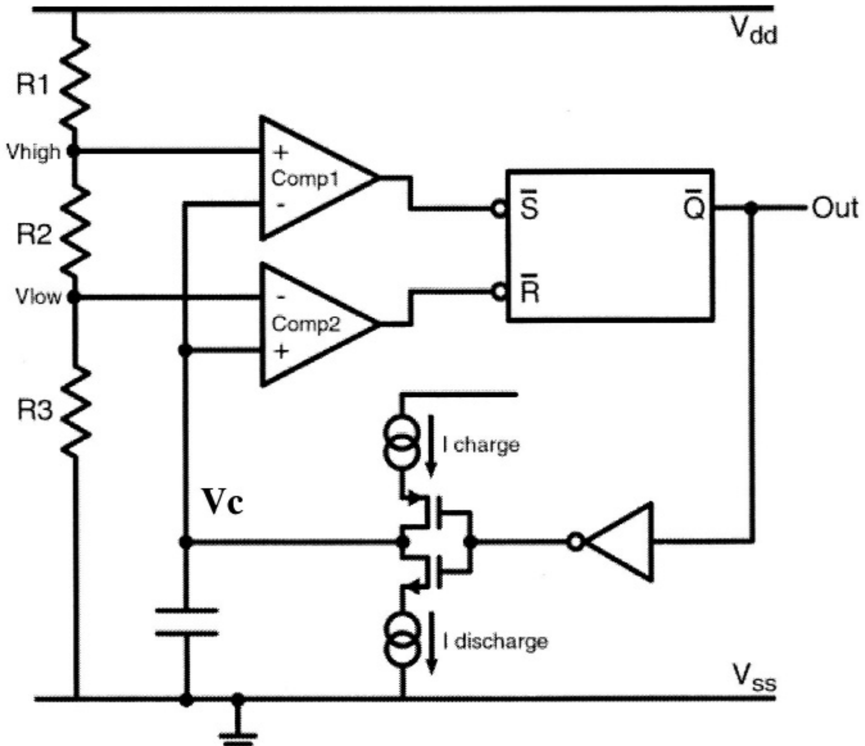


Figure 10.31: Example of a relaxation oscillator design.

There are a number of advantages and drawbacks to the use of this type of oscillator in SOI compared to bulk. When properly designed, the parasitic capacitances in SOI will be reduced. This improves accuracy and speed and furthermore reduces power if the charging capacitor size is reduced. Resistors diffused into the silicon (as opposed to polysilicon resistors) have more linear characteristics in SOI. When using this type of circuit, the same precautions should be taken in comparator design as would be taken for an operational amplifier design.

Ring oscillators are constructed with a series of gate or comparator delays. To create a VCO from this type of oscillator the delay cell has to be converted into a variable delay cell. This is readily achieved in the case of a comparator unit delay by adjusting the supply current, the supply current being approximately proportional to the inverse of gate delay. The circuit oscillates with a period equal to the cell delay times twice the number of delay stages. An example of this type of VCO is shown in figures 10.32 and 10.33. An advantage of most ring oscillator designs is that they can readily be adapted to provide quadrature outputs, useful in phase lock loop applications. The frequency is controlled through the I_{ss} current. This basic design has potential problems in PD-SOI due to floating body variability, and is most accurate as a body tied implementation.

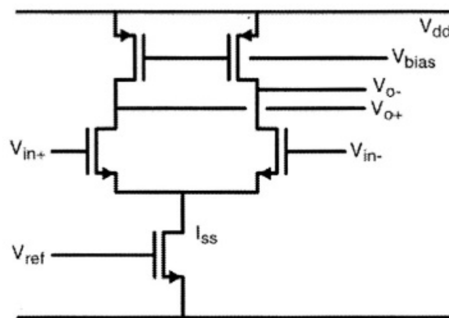


Figure 10.32: One stage of a ring oscillator type VCO. To alter frequency, I_{ss} is adjusted, which affects the delay timing.

Each stage is combined in a ring of many stages, as shown in figure 10.33. This represents a ring of 'n' oscillator. The structure has a common control for current, and a single inversion of the chain interconnect ensures oscillation (chapter 6).

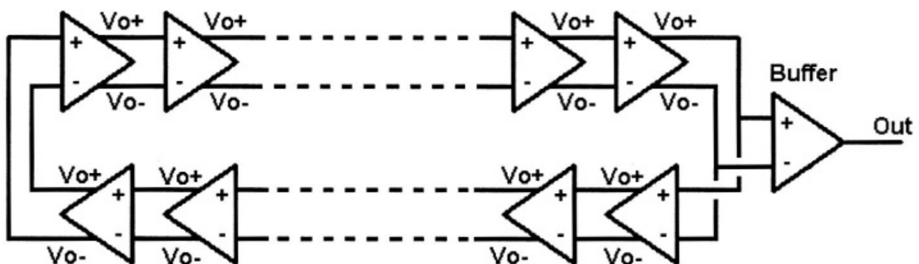


Figure 10.33: Block diagram of a voltage controlled ring oscillator. Each stage consists of the circuit in figure 10.32.

Self-heating in current controlled ring oscillators has been studied [10.16]. It has been found that no significant differences were observed between long and short chains. This indicates that self-heating is not important in this circuit.

10.12 Microwave Applications

Solid state microwave applications cover approximately the 1-30 GHz frequency range. At higher microwave frequencies circuit applications are fairly limited, typically being restricted to, for example, oscillators, amplifiers, frequency multipliers and mixers. Until recently many microwave amplifiers used GaAs as the semiconductor material. More recently Si implementations in both bulk and SOI have been developed [10.23].

Although currently only applicable to the lower frequencies of the microwave band, there are a number of advantages to SOI, including higher performance active components and improvements to passive components. In addition, some microwave devices are used in high radiation applications, where the potential to integrate microwave applications with rad-hard logic and memory is attractive.

Single FET mixers, similar in design to the circuits described elsewhere in this chapter, and other microwave circuits have been developed for board level and monolithic applications [10.30].

Silicon substrates used in SOI usually have relatively low resistivity (often around 20 ohm cm), and may vary substantially from wafer to wafer. When transmission lines are fabricated on silicon, variability in resistivity can cause a wide variation to the characteristic impedance of the transmission lines. The process dependency, and frequency dependent characteristics have to be considered in the designed. If variability of resistivity or absolute resistivity are issues, these can be minimized through additional process controls. Bonded wafer methods of SOI preparation have the advantage that the substrate can be tailored to specific applications. Silicon substrates can be chosen for optimum performance. In some cases alternative substrates (usually insulating quartz or glass) can be substituted.

Simulation procedures can be used which permit the determination of equivalent circuit parameters for small-signal MOS models at individual bias points. However, difficulties exist in extracting characteristics at higher frequencies. Channel charging delays can significantly impact the phase-margin of very high-frequency operational transconductance amplifier (OTA) designs, and can also affect input matching for low noise amplifiers.

Other high frequency circuits are also affected, such as mixers and high speed switching circuitry. With physical microwave models for SOI, circuitry can be optimized.

10.13 Voltage Regulation

Voltage regulators are very common circuits. They accept a somewhat variable DC input voltage and output a higher precision voltage that can be used as a power supply for other circuits. Two common types of voltage regulator are the "series" regulator and low-drop out (LDO) regulator.

10.13.1 Series Regulator

The output voltage of the series regulator is maintained by a power transistor (in this case nMOS) in series with the output. Figure 10.34 shows the components of a series regulator. A reference voltage is generated, typically using a bandgap reference. This reference voltage is fed to the non-inverting input of an operational amplifier, where it is compared with a divided sample of the output.

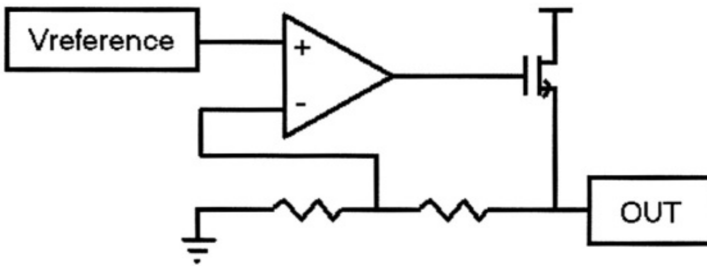


Figure 10.34: Schematic of a series voltage regulator.

Particular precautions required for the series regulator design as applied to PD-SOI processing have largely been covered in the sections on bandgap design and operational amplifiers. The power output transistor can be a floating body MOS device, but for highest precision a body tied device may be substituted.

10.13.2 LDO Regulator

The LDO regulator is a form of series regulator, but with the capability to retain regulation very close to the supply voltage. This type of regulator is much more difficult to stabilize than the series regulator, but is necessary in the compendium of low voltage circuits. The basic components of the LDO are similar to those of the series regulator, and are shown in figure 10.35. Bandgap, operational amplifier and voltage divider are usually the same or similar to those used for the series regulator. The controlling transistor is here a pMOS, which should be body-tied for optimum performance.

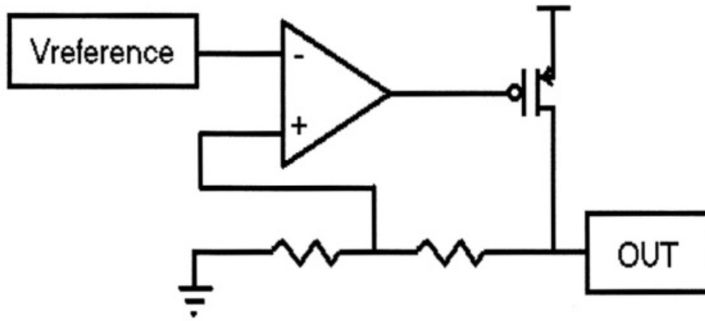


Figure 10.35: Schematic of a LDO regulator.

10.14 Analog to Digital Converters (ADC)

ADCs are vital components in converting real world inputs into digital signals that can be manipulated by digital signal processors or microprocessors. Requirements of the system dictate the type of ADC used, and is based on input voltage range, resolution, acquisition time, linearity, offset, power dissipation, conversion time and die area. The input voltage range refers to the maximum and minimum voltage to be digitized by the ADC. The ADC resolution is the number of steps into which the input range is divided, and is usually expressed as number of bits (n), where the number of steps is 2^n-1. Acquisition time is the time the input voltage must be sampled to get a reliable output voltage. To reduce this time when slow ADCs are being used, it is common to add a sample-and-hold circuit to the ADC input to freeze an otherwise varying analog voltage at the moment the sample is required. This voltage is held constant while the ADC digitizes the signal. Linearity is a measure of the equivalency of the input to output voltage. In practice each quantized ADC output step is offset from ideal, which leads to non-linearity in a plot of ADC output against input signal. Offset error affects all bits, and is a measure of the difference of the zero voltage input (or other specified input) to the output.

10.14.1 Successive Approximation ADC

A relatively slow but area efficient ADC uses the successive approximation topology (figure 10.36). The output of the comparator is a signal that is either high or low depending on the input voltage which can be higher or lower than the reference voltage from the digital to analog converter (DAC), described elsewhere in this chapter. Control logic increments or decrements the digital word, using one of numerous algorithms, and the digital output is converted to an analog signal by the DAC. Once the reference threshold of the DAC crosses the input voltage, the comparator output switches and the digital output can be read. This type of ADC can be free running, or can

perform a single conversion and then await a further signal from inputs to the control logic before repeating the sequence.

For SOI designs, no particular precautions are required for the digital portions of the circuitry, however care must be exercised for operational amplifier designs to ensure the comparator does not introduce offset into the converter. Refer to the section on DACs in this chapter.

Marginally improved performance can be expected with the successive approximation ADC built in SOI, compared to bulk, due to frequency improvements of the individual blocks. These improvements are mainly from reduced capacitance of the analog and digital structures, and higher drive strength of the digital control logic.

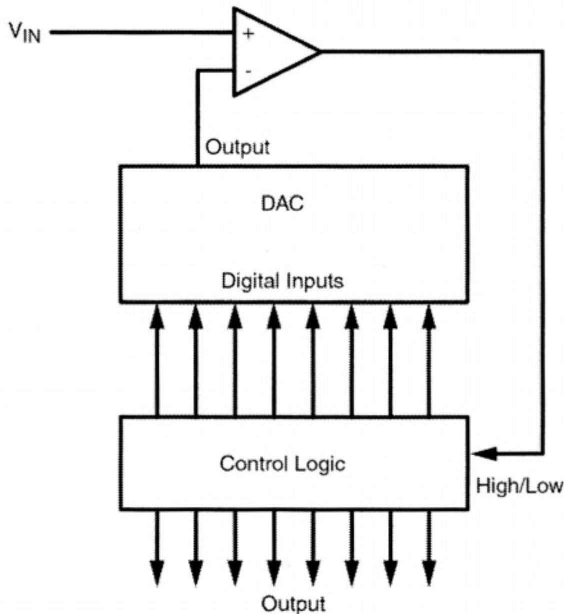


Figure 10.36: Successive Approximation ADC implementation.

10.14.2 Flash Converters

The flash (or Parallel-Encoding) ADC provides the fastest operation of all ADC architectures. However, for n-bit converters (n-1) comparators are required. This is expensive in terms of die area, the very large input capacitance of all the paralleled inputs and high supply power requirements. The large die area consumed can result in inaccuracies being introduced as a result of thermal mismatch and physical separation of the comparators. A resistor network sets discrete thresholds for the comparators. Decode logic

converts the comparator outputs to a usable digital sequence. Figure 10.37 shows a three bit flash implementation.

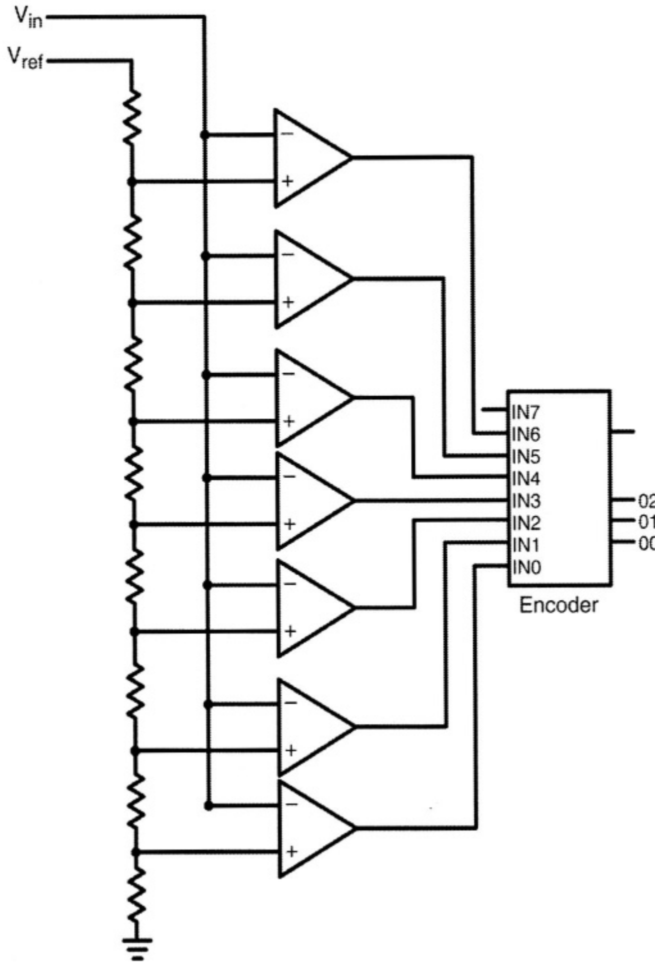


Figure 10.37: Implementation of a 3 bit flash analog to digital converter. This simplified diagram shows that the basic building blocks for the flash are comparators and digital logic.

A reference signal is applied to the resistive divider network. The outputs from the resistor ladder are the reference inputs for the comparators. When the input signal is higher than the reference the comparator output is low. When the input signal is lower than the reference the comparator output is high. All comparator outputs are fed to the encoding logic, which converts the signals into a digital output that can be used by the subsequent digital decode circuitry.

Comparative performance of designs for a flash ADC on bulk and PD-SOI, for the most part, favors PD-SOI. The resistor ladder can be built more accurately in SOI. The decoding logic is faster and smaller in SOI. In addition, comparators are faster (or alternatively can operate at lower power) in SOI. However, body ties or modified comparator design is required to equalize the body potential of the input pair. The simplest floating body comparators are prone to high and variable offsets which are unacceptable.

10.14.3 Pipelined ADCs

To minimize die area when high resolution and relatively high speed are required, pipelined architectures may be used. Figure 10.38 shows a typical multiple stage pipeline ADC architecture. Stages are typically scaled smaller at higher stage numbers, optimizing noise performance and power requirement.

Each stage quantizes the output from the previous stage into one or more bits using a flash ADC architecture. The quantized result is converted to an analog signal using a digital to analog converter and subtracted from the sampled signal. The resulting difference signal, which is smaller than the LSB of the stage is then amplified and is available for the next stage.

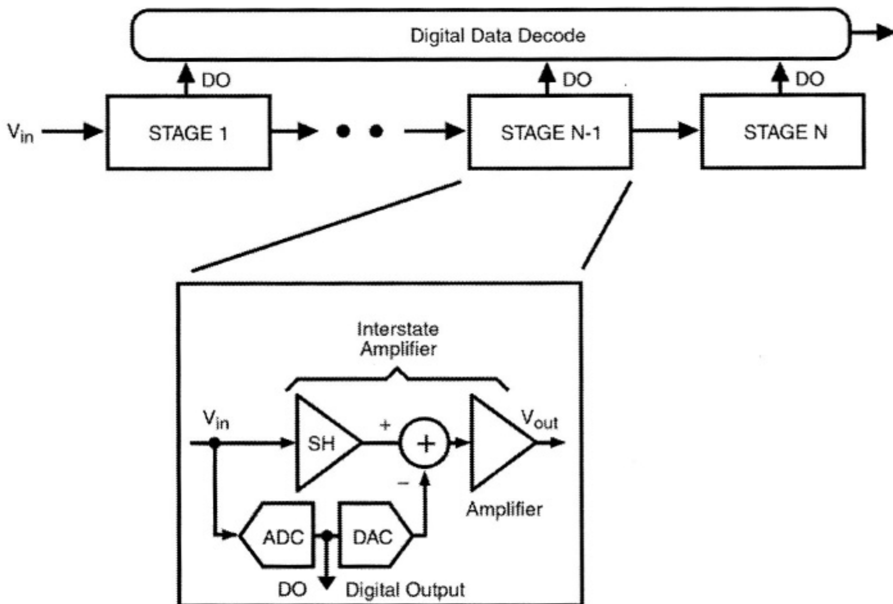


Figure 10.38: Pipelined architecture for an ADC. This is more efficient in terms of die area requirements than the Flash ADC architecture.

For PD-SOI operation of pipeline structures all precautions discussed in flash ADCs and DACs are applicable. It is particularly important to avoid errors, which will be amplified as the signal passes from stage to stage. Errors in the opamp and comparator thresholds can lead to errors in the difference signal, causing missing or spurious codes. Capacitor mismatch, an important source of gain error, is lower in SOI material than in bulk. Digital error correction techniques used for pipelined ADC architectures in bulk material are also applicable to pipelined architectures on SOI.

10.15 Digital-to-Analog Converters (DAC)

Accurate current sources and device mismatch minimization are critical in DAC design. In addition to the mismatch seen in bulk material, mismatch in SOI can cause thermal differences. The impact of localized heating on the design of DACs in SOI has been previously studied [10.16], where it was demonstrated that thermal behavior influenced linearity. However, it was also shown that this could be minimized to the point of being insignificant with careful design practice. A 10-bit DAC has been designed based on a PD-SOI technology. This uses the technique of biasing the operating region outside the kink region [10.31].

Most DACs are required to output a linear scale, with voltage or current weighted according to the binary value of the digital signal. Perhaps the simplest DAC in general use is based on a resistive ladder. Figure 10.39 shows the basic implementation of this configuration, which is termed the R-2R ladder DAC.

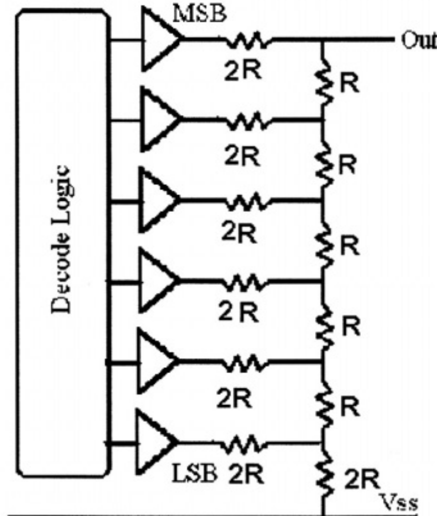


Figure 10.39: R-2R type DAC

The mode of operation of the circuit in figure 10.39 is as follows: The decode logic determines the output pattern to the ladder. If all decoded outputs are at logic level zero, then the output is at 0V. If the output is at a higher level, then one or more decode logic outputs is at a logic level 1. The output voltage from the ladder can be determined following an analysis of the currents through and voltages across the network.

This circuit benefits significantly from the use of SOI. The buffers can be smaller and decode logic is faster and smaller. Resistor area may be reduced for the same linearity and resistance. This leads to reduced capacitance and faster DAC operation. No particular precautions need be taken in the conversion of this structure to SOI from bulk unless the output is buffered using an op-amp.

Figure 10.40 shows an alternative configuration for a DAC. This is based on combining currents. This does not generally give a DAC output that swings rail to rail, as achieved with the DAC shown in figure 10.37, but it is more versatile in terms of the range of adjustment that can be made for different conditions such as supply voltage and thermal effects.

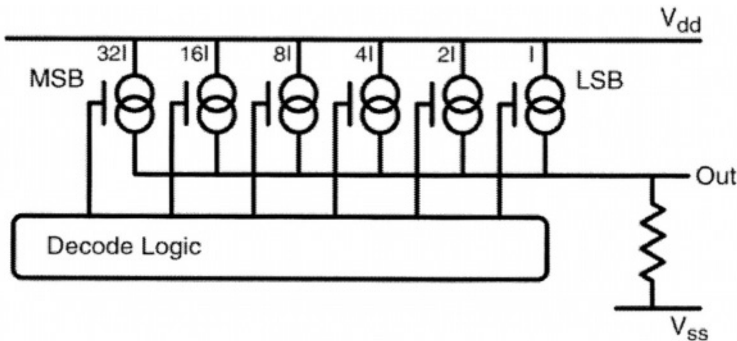


Figure 10.40: Binary weighted current type DAC

The mode of operation of the current scheme of DAC is straightforward. Binary weighted current sources are used to create a voltage across a resistor. These currents are matched to the characteristics of the resistor. The currents are compensated for temperature and possibly supply voltage. As previously discussed matching of resistors is easier in most cases for SOI material, but care must be taken in matching the current mirrors required for the current sources.

10.16 Sigma Delta Modulator

Sigma-Delta modulators are widely used in filters, phase-locked loops and as a method to convert between analog and digital circuitry, and are also

important in mobile communications applications. Basic operation of a sigma-delta modulator is as follows: The forward path consists of a filter and quantizer (figure 10.41). The low pass filter input is the difference between the input and feedback loop. Thus the average value of the filter output equals the average value of the input signal. The filter provides an analog representation of the digital output. The filter normally includes one or more operational amplifiers. The simplest way of dealing with a PD-SOI low pass filter is to make use of body ties, to avoid integration of any body-voltage-induced input offset. In addition the op-amps of the filter should be low-noise, again, using body ties keeps the body away from the higher noise kink region.

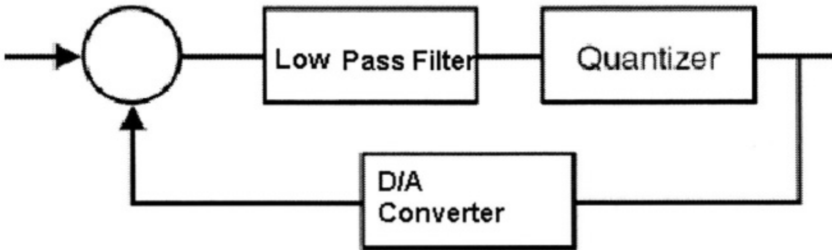


Figure 10.41: Block diagram for a sigma-delta modulator.

10.17 Interface between Digital and Analog Circuitry

An important area of circuit design that is frequently overlooked is the interface region between analog and digital circuitry. A common method to interface a signal between the two types of circuitry is the use of a level shifter. This is particularly appropriate when the logic operates at a different voltage to the analog circuitry. Such a situation is common with low voltage processes, where logic operates at a low voltage for power consumption reasons, but analog and interface circuitry has to operate at higher voltages. An example of interfacing to digital circuitry from an analog circuitry might be from a comparator output to a digital input. In this case an interface block such as figure 10.42, might be used. This converts the comparator output to a digital level signal. Circuit design for SOI processes remains similar to bulk, although fewer precautions need be observed for SOI material, due to the inherent PMOS tank isolation of SOI.

Some design and operating conditions may permit voltage V1 to be zero volts. An initial analysis of this might lead to the expectation that the output stage latch connected to V2 will maintain the same state as defined prior to the removal of the V1 supply. However, leakage effects do not guarantee this, and with the uncertainty of the leakage due to SOI body effects it is

advisable to use bypass circuitry to define the state in the event that voltage V1 falls below the threshold voltage of the V2 stage NMOS components.

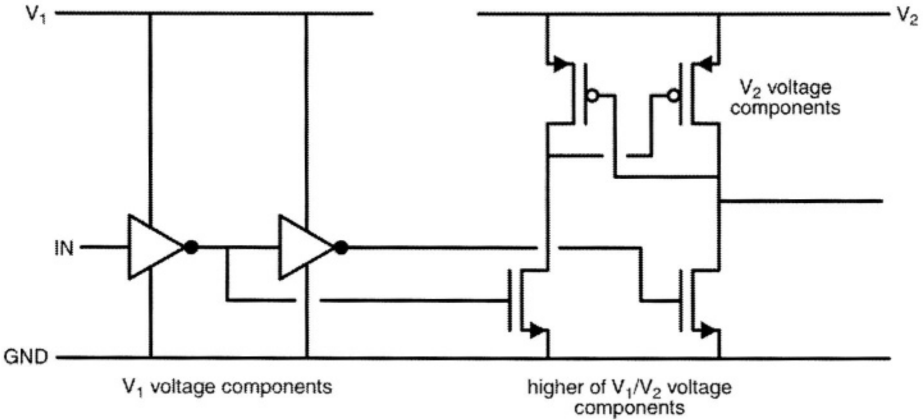


Figure 10.42: Example of a level shifting circuit used to interface between analog and digital circuitry, with different voltage supplies.

10.18 Power Amplifiers

Power amplifiers are used in many applications, often using DMOS (Double Diffused MOS) transistors. Power dissipation issues associated with DMOS and other high current density components have been discussed in chapter 3. High frequency integrated SOI power amplifiers have also been described [10.32]. Wireless power amplifier operates at 900MHz, and includes integrated MOS transistors, inductors and capacitors on a 1.5μm SOI BiMOS technology. The amplifier output stage is an LDMOS, in a class E configuration. The ability to integrate the RF power output onto SOI material improves the performance to cost ratio of the material.

10.19 Sensors and Actuators

Since the early 1990s a wide variety of integrated sensors and actuators have been developed. These range from Hall-effect sensors to Digital Light Processing technology. Most of these have been built on bulk silicon, but increasingly, a variety of Micro-mechanical (MEMS) devices, such as gyroscopes and accelerometers have been demonstrated in SOI. SOI provides the advantage of simple construction of electrically isolated, three-dimensional micro-mechanical structures, and enhances the variety, shape and layer thickness that can be fabricated.

Infrared sensors have experienced a recurring problem with sensitivity and precision due to high thermal dissipation through the silicon. The thermal

isolation afforded by surrounding oxide is expected to improve sensor sensitivity. In addition, the low capacitance to substrate should benefit thermal sensor performance.

SOI design of Hall Effect sensors has also been demonstrated for high temperature automotive applications [10.33]. To improve amplifier offset performance a bipolar chopper amplifier was used. The use of SOI enabled both elevated temperature and low leakage performance.

References

[10.1] M.R. Casu, et. al., “Comparative Analysis of PD-SOI Active Body-Biasing Circuits”, 2000 IEEE International SOI Conference, Oct. 2000, pp. 94-95.

[10.2] 0. L. Neel & H. Haond, "Electrical transient study of negative resistance in SOI MOS transistors", Electron. Lett., vol. 26, no. 1, 1990, pp. 73-74.

[10.3] B. M. Tenbroek, et.al., "Self-heating effects in SOI MOSFET’s and their measurement by small signal conductance techniques", IEEE Trans. Electron Devices, vol. 43, Dec. 1996, pp. 2240-2248.

[10.4] A. Marshall, “Beta Helper for Voltage and Current Reference Circuits” US patent number 5684394, November 4th, 1997

[10.5] P. R. Gray & R. G. Meyer, “Analysis and Design of Analog Integrated Circuits, third edition”, Wiley, 1993, ISBN 0-471-57495-3.

[10.6] A. P. Brokaw, “A Simple Three-Terminal IC Bandgap Reference”, IEEE Journal of Solid State Circuits, Vol SC-9, 1974, pp. 388-394.

[10.7] A. Marshall, et. al., “Compensation for low gain bipolar transistors in voltage and current reference circuits”, United States Patent #5,349,286, September 20, 1994

[10.8] R.A. Blauschild, et. al., “A New NMOS Temperature-Stable Voltage Reference”, IEEE Journal of Solid-State Circuits, Vol. SC-13, No. 6, December 1978, pp. 767-773.

[10.9] C. Eisenhut & J. W. Klein, "SIMOX Voltage References for Applications up to 275°C using the Threshold Difference Principle", Proceedings 1997 IEEE International SOI Conference, Oct 1997, pp. 110-1.

[10.10] A Nève, et. al., “Smart Card Circuits in SOI Technology”, 2000 IEEE International SOI Conference, Oct. 2000, pp.48-49.

[10.11] J.B. Kuang, et. al., ”Dynamic Body Charge Modulation for Sense Amplifiers in Partially Depleted SOI Technology", IEEE Journal of Solid-State Circuits, Vol. 36, No. 4, April 2001, pp. 597-604.

[10.12] J-P. Eggermont, et. al., "Potential and Modeling of 1μm SOI CMOS Operational Transconductance Amplifiers for Applications up to 1 Ghz" IEEE Journal of Solid State Circuits, Vol. 33, No. 4, April 1998, pp 640-643.

[10.13] D. Flandre, et.al., "Fully-depleted SOI CMOS technology for low-voltage low-power digital/analog/microwave circuits", Analog Integrated Circuits and Signal Processing. New York: Kluwer, 1998.

[10.14] A. Vandooren, et. al., "Performance of Irradiated Gate-All-Around SOI MOS OTA Amplifiers" Proceedings 1997 IEEE International SOI Conference, Oct 1997, p62-3.

[10.15] P. Francis, et. al., “Radiation-hard design for SOI MOS inverters”, IEEE Transactions on Nuclear Science, Volume 41, Issue 2 , April 1994, pp. 402-407.

[10.16] S. H. Voldman, "The State of the Art of Electrostatic Discharge Protection: Physics, Technology, Circuits, Design, Simulation, and Scaling", IEEE Journal of Solid State Circuits, Vol 34, No. 9, Sept. 1999, pp. 1272-1282.

[10.17] A. O. Adan, et. al., "Linearity and Low-Noise Performance of SOI MOSFETs for RF Applications ", 2000 IEEE International SOI Conference, Oct. 2000, pp.30-31.

[10.18] M. Kumar, et.al. “A 900MHz SOI Fully Integrated RF Power Amplifier for Wireless Transceivers” ISSCC 2000, pp.382-3.

[10.19] D. Buss, “Device Issues in the Integration of Analog/RF Functions in Deep Submicron Digital CMOS”, ISSCC 2000 (short course) ISBN 0-7803-6241-1

[10.20] J. P. Colinge, et. al., “A low-voltage, low-power microwave SOI MOSFET”, IEEE International SOI Conference Proceedings, Oct. 1996, pp. 128-9.

[10.21] S. V. Kishore, et. al., "A SOI-BiCMOS RF-transmitter for personal digital cellular communication (PDC)" Symposium on VLSI Circuits, 2000. pp. 62-65.

[10.22] C. F. Edwards, et. al., "A Multibit Modulator in Floating-Body SOS/SOI CMOS for Extreme Radiation Environments", IEEE Journal of Solid State Circuits, Vol. 34, No. 7, July 1999, pp. 937-948.

[10.23] R. Kokozinski, et. al., "Microwave wideband in bulk-CMOS and CMOS/SIMOX technologies", Proc. IEEE ISSCC Conf., San-Francisco, CA, Feb. 1995, pp. 188-189.

[10.24] H. Jin & C. A. T. Salama, “A I-V, 1.9-GHz CDMA, CMOS on SOI, Low Noise Amplifier”, 2000 IEEE International SOI Conference, Oct. 2000, pp. 102-103.

[10.25] E. Ibaragi, et. al., “A CMOS Analog Multiplier Free from Mobility Reduction and Body Effect”, IEICE Transactions Fundamentals, Vol. E82-A, No. 2, Feb. 1999, pp. 327-334.

[10.26] B. Gilbert, “A Precise Four-Quadrant Multiplier with Subnanosecond Response”, IEEE Journal of Solid State Circuits, Vol. SC-3, Dec 1968, pp. 365-373.

[10.27] S. Ye & C. A. T. Salama, “A 1 V, 1.9 GHz, Low Distortion Dual-Gate CMOS on SOI Mixer”, 2000 IEEE International SOI Conference, Oct. 2000, pp. 104-5.

[10.28] P. Delatte, et. al., “Comparison of 0.25 μm Bulk, PD and FD SOI CMOS implementations of a Low-Voltage Low-Power Programmable DLL for Linear Delay Generation”, 2000 IEEE International SOI Conference, Oct. 2000, pp. 92-93.

[10.29] R. Dekker, et. al., “An Ultra Low-Power RF bipolar technology on Glass” IEEE International Electron Devices Meeting, 1997, pp. 921-923.

[10.30] G. D. Vendelin et.al., “Microwave Circuit Design Using Linear and Nonlinear Techniques” Wiley Interscience 1990, ISBN 0-471-58060-0

[10.31] D. Chang, et. al., "Kink-Free Analog Circuit Design with Floating-Body NFD/SOI CMOS: A Current-Steering D-A Converter", Proceedings 1997 IEEE International SOI Conference, Oct. 1997, pp. 158-9.

[10.32] Y. Tan et. al., “A 900 MHz Fully Integrated SOI Power Amplifier of Single-Chip Wireless Transceiver Applications”, IEEE Journal of Solid-State Circuits, Vol. 35, No. 10, October 2000, pp. 1481 – 1486.

[10.33] N. Kordas et.al., “An SOI 0.6mV Offset Temperature-Compensated Hall Sensor Readout IC for Automotive Applications up to 200 degrees C” ISSCC 1999, pp. 134-5.

Chapter 11: Global Design Issues

11.1 Introduction to Global Design Issues

Some issues of IC design are common to multiple types of design, rather than as purely analog, digital or memory. In this chapter we discuss questions related to SOI that span individual design disciplines, including temperature effects, clock distribution, and parasitic resistance and capacitance.

11.1.1 Cell Libraries

Attempts to use a cell library modified from a bulk library, for SOI designs has proven to be inefficient, since it does not account for all the performance enhancements resulting from floating body effects. As a result, the modified bulk library has to assume worst case delays.

To make full use of the benefits of SOI designs with floating body and/or dynamic body contacts require detailed circuit characterization. This is a design resource issue rather than a fundamental roadblock, and should be given priority for optimum performance simulations.

11.1.2 Clock Distribution

Clock distribution and skew / jitter minimization have been challenges in high performance designs. In floating body SOI configuration, uncertainty of the floating body potential and hysteresis effect translates into clock skew and jitter, degrading performance. It is difficult to design the clock distribution tree to balance out these effects, thus body contacts should be considered in critical paths.

11.1.3 Decoupling Capacitance and Series Resistance in SOI

SOI material has significantly reduced on-chip decoupling capacitance. This can lead to increased supply bounce. In bulk material, about 50% of the decoupling capacitance is from the junction to substrate and well to substrate capacitances.

In SOI, with faster circuit speed and absence of the built-in decoupling capacitance, supply and ground bounce are more severe. Thus decoupling capacitance must be specifically added, and distributed around the chip. However, although more severe, the isolated nature of the MOS devices on SOI often reduces susceptibility to supply bounce.

11.1.4 High Temperature Operation

Circuits operate at much higher temperatures using SOI material than possible using bulk material. SOI devices in general, and FD-SOI devices in particular have leakage currents two to three orders of magnitude lower than bulk, particularly at elevated temperatures. Due to lower forward voltage drop of diodes at high temperatures, typical bulk silicon processes display reduced immunity to noise, and an increased tendency to latch-up. SOI circuit design avoids leakage to substrate including leakage through the PMOS tank.

Figure 11.1 shows leakage paths for a NAND in both bulk and SOI. Neglecting MOS gate leakage (which, at high temperatures, is insignificant compared to drain-source leakage) overall leakage is defined by the lowest leakage component.

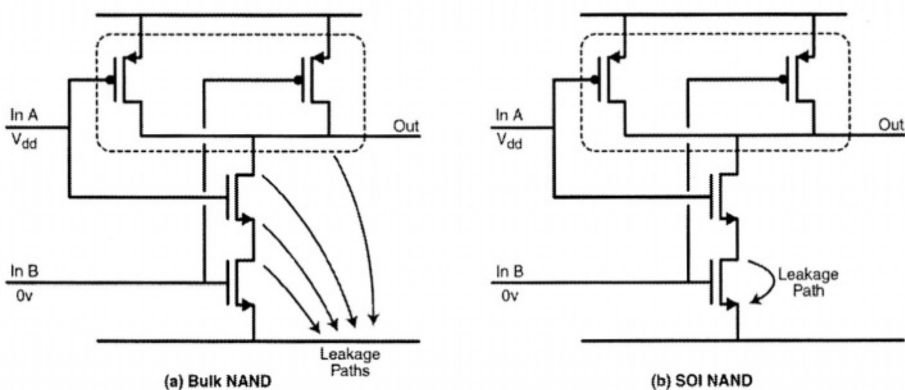


Figure 11.1: Leakage paths associated with a NAND (a) in bulk silicon, and (b) in SOI material, showing leakage path through the off-state SOI device, and reverse biased diodes, including the PMOS tank in bulk material.

For most applications PD-SOI can operate to about 250°C [11.1], while bulk is marginal even at 200°C [11.2]. However, very high temperature systems often benefit from the use of fully-depleted silicon, when operation is possible up to 300°C or higher [11.1, 11.3].

11.1.5 Gate Leakage

MOS transistor gate length has shrunk with process technology improvements. To maintain performance at lower voltages gate oxide thickness has also been forced to reduce [11.4]. Gate leakage is a significant proportion of the total leakage, especially at lower temperatures. Gate leakages can affect operation in memory cells and analog circuits even in bulk material. In circuits built on PD-SOI material gate tunneling modifies the history effect [11.4]. The gate leakage component can be included in some transistor models, such as BSIM4.

11.2 Noise Immunity

Noise issues are similar in bulk and SOI. Noise effects, however are aggravated in PD-SOI due to the floating body, and faster circuit switching. The most common types of noise are due to crosstalk, simultaneous switching, delay and logic noise. Synchronous noise may occur at the clock edges when latches are triggered. When a noise event occurs without any timing relation, this is termed asynchronous noise. While absolute noise may be higher, the isolated nature of many components means a higher level of noise can be tolerated, particularly in static logic applications.

11.2.1 Circuit Noise

Removal of junction isolation in favor of dielectric isolation relieves the designer of certain thyristor latchup issues related to MOS designs built on bulk processes. Thyristor latchup is of particular concern in power integrated circuits, automotive applications and other noisy environments. Design for automotive applications often includes a variety of tests to verify latchup immunity. This testing is administered to the chip as part of qualification and ongoing testing [11.5]. In addition, switch-mode regulators, and drivers for motors are specifically prone to latchup from inductive spiking [11.6]. Many high power chips include additional highly doped buried layers and deep high concentration diffusions to minimize such effects [11.6].

The replacement of junction isolation with dielectric isolation would, therefore appear to be beneficial. However junction isolated parts have one distinct advantage that SOI lacks, that of significant built-in capacitance from supply to ground, through the PMOS body contact to substrate (figure 11.2). Usually the PMOS body is created using an N-type tank diffusion into

a P-type substrate. The substrate is electrically connected to ground, as is the NMOS body contact. The N-well diffused into the substrate creates a reverse biased diode between supply and ground. This diode can be a quite large proportion of the chip area, and can add several tens of picofarads of capacitance between supply and ground. This helps minimize voltage switching transients caused by resistive and inductive wire bonds within the package.

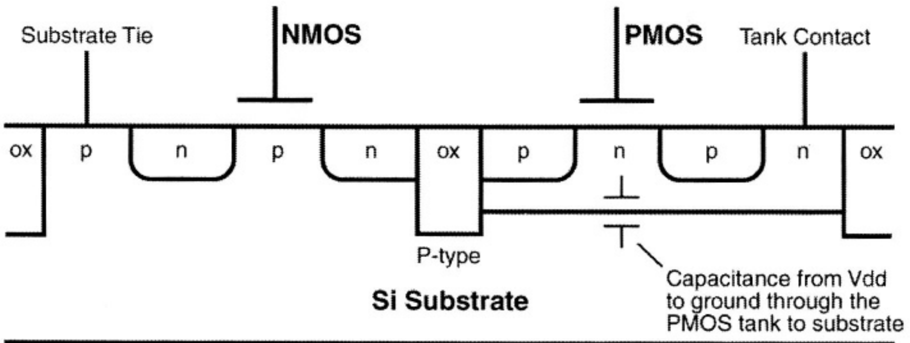


Figure 11.2: Demonstration of the junction capacitance that is commonplace in bulk IC designs, but absent in SOI. The cross-section shown is the capacitance in bulk material that is dielectric isolated. Capacitance in bulk material that is junction isolated is similar.

11.2.2 Capacitive Coupling Noise

Adjacent metal leads are capacitively coupled. The current coupling is given by:

I=C*dV/dT

where C is the capacitance between the wires and dV/dT the rate of change of voltage between the two wires.

Voltage switching of a wire induces a pulse on the adjacent signal. Crosstalk noise is caused by the coupling of the two signals. Memory arrays are susceptible to this type of noise. If the bit is maintaining a HIGH logic level, and the complementary bit switches LOW, the bit line develops a tendency to pull LOW due to coupling from the adjacent bitlines. Many methods have been used to reduce bit line to bit line coupling, including the use of twist lines forming pairs of signals. In general, however, the effects of capacitive coupling between adjacent metal traces may be more severe in SOI due to

the reduction in source and drain capacitance, and possible reductions in the interconnect to substrate capacitance.

11.2.3 Delay Noise

Delay noise refers to crosstalk coupling generated by one signal such that the adjacent coupling signal swing in either the same direction or in opposing directions. If both signals swing in the same direction, capacitive coupling speeds up the signal. If the signals switch in opposing directions capacitive couplings slows down the signal. This type of delay noise can therefore cause a race condition, as signals arrive sooner or later than expected, potentially causing chip failures.

11.2.4 Logic Noise

When a neighbouring signal pulls down a signal that is supposed to maintain a steady voltage functional failures may result. A memory circuit may have parallel bit lines. If one of two adjacent bit lines switches from HIGH to LOW, if the coupling capacitance is too high, then the bit line trying to maintain a HIGH level will tend to be pulled lower, possibly reading corrupted data out to the sense amplifiers. Dynamic and pass gates are more susceptible to logic noise fails than static logic circuits.

11.2.5 Decoupling Capacitors

In the absence of the supply-ground junction capacitance of bulk designs, SOI designers have explored different methods for achieving immunity from transient switching effects. One option is the addition of thin-oxide decoupling capacitors [11.7]. These have been used to supply surges in current from the chip, bypassing series impedance with the power pads. Integrated circuits that are substantially digital in nature can use distributed thin-oxide capacitors [11.8], to achieve a measure of noise immunity. In most cases it is adequate to add decoupling capacitors in all available space.

11.3 Latchup Immunity

Figure 11.3 shows the potential for latchup of conventional CMOS, through turning ON the intrinsic thyristor [11.1]. Latchup can occur as a result of the transient switch triggering a PNPN (thyristor or SCR) structure between supply and ground. The PMOS tank may provide enough resistance (R1) that the N-well tank voltage is at V_{be} below supply. This turns on the parasitic PNP (Q2) from Vdd to substrate. Current flowing through the substrate causes a voltage gradient, which can turn ON the NPN structure comprising the N-well, NMOS source and p-substrate. Careful layout eliminates this possibility, but clearly these parasitics cannot activate in SOI, where no direct path exists to cause a thyristor latch. It is still possible to

create a thyristor action in SOI, leading to uncontrolled current and/or circuit damage due to extremes in current or excessive thermal dissipation, however, it is much more difficult, and must be part of the circuit design (whether intented or not).

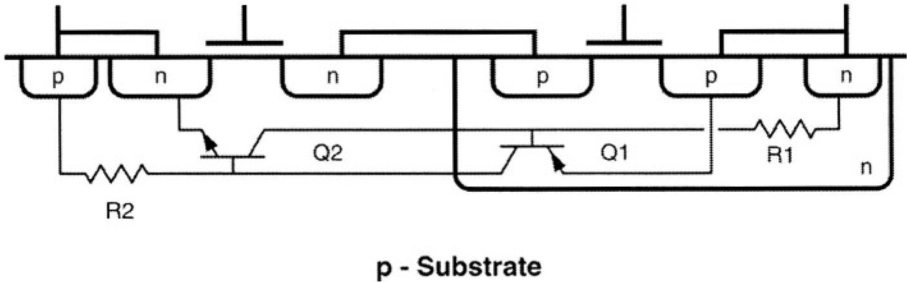


Figure 11.3: Cross sectional view of a bulk layout that can result in latchup. If resistances are not kept low through careful layout, the simple bulk CMOS inverter gate structure can be the source of a thyristor action. Resistance shown is formed between the Vdd or ground connection and the region that acts as a bipolar base.

Figure 11.4 shows the simplified electrical schematic of the latchup path of figure 11.3. To attain latchup of this circuit the overall gain of the circuit must be greater than unity. Process changes can be implemented to reduce the gains of transistors Q1 and Q2. Layout affects all components.

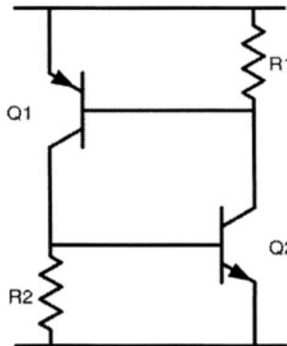


Figure 11.4: Electrical representation of the thyristor capable of causing latchup that is extractable from conventional bulk CMOS. The resistances and transistors are distributed, and as a result great care must be taken in positioning of body and tank ties, to keep the thyristor resistances low. This gives a desirable high hold current. (Note that the transistor and resistor identifiers correspond to the same devices numbers in figure 11.3)

The latchup immunity of SOI is a significant benefit to certain applications. In principle, however, PD SOI can still provide a multi-device latchup path, even though parasitic bipolars to substrate are absent. Latchup in SOI is possible where body tie structures are utilized [11.9]. Figure 11.5 shows how the use of body ties can create a thyristor function from the bipolar structure that is a parasitic of the MOS device.

Figure 11.5(a) is electrically equivalent to figure 11.5(b), and may represent two different ways to lay out the circuit. In the case of figure 11.5(b), with input LOW, the inverter output is HIGH, thus the PMOS output driver is off, and the NMOS body driver is also off. The NMOS output driver is on. Thus point 'C' is a low resistance to ground. If the body tie between the NMOS body driver and node C (shown as lead "nBres") is resistive, a latchable thyristor structure can be created. In most cases this structure would be far more robust against thyristor latchup than its bulk counterpart, but in the presence of certain transients on the output latchup cannot automatically be ruled out.

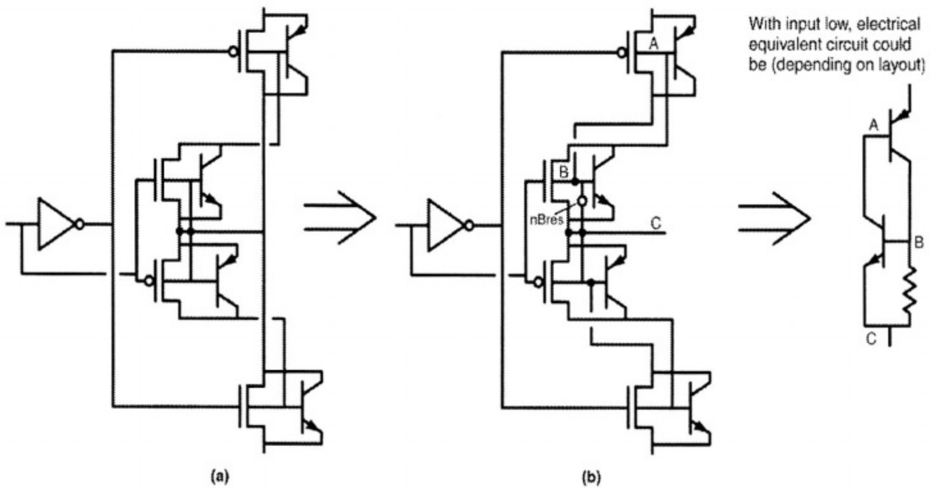


Figure 11.5: Demonstration of how a body tied PD-SOI structure can be prone to a thyristor type latchup, if layout is not carefully considered In this case when the input is low, the output 'C' would also be low. This turns off the PMOS output device and the NMOS body biasing device, resulting in a thyristor structure between supply and the output.

Latch-up can even occur in single SOI transistor if carrier generation due to impact ionization is higher than the source recombination (typically at high Vds) [11.10, 11.11].

11.4 Self Heating

The higher thermal resistance of SOI compared to bulk material is often not a problem, but care should be taken to ensure matched components are not at different temperatures, or that a high temperature element is not adjacent to temperature sensitive circuitry. Self heating is a problem even in some circuit designs on bulk material, and techniques have been developed to improve heat removal from the active silicon. There are various options for circumventing poor thermal characteristics. Perhaps the simplest and least costly is the use of metal layers above the heat generating component as a lateral heat spreader. Heat spreading can perform two functions. It can be used either to spread heat for dissipation outside the device area, or to spread heat to adjacent components that are required to maintain a similar temperature [11.1]. Another technique uses the heat-dissipating properties of bump technology [11.12, 11.13]. Bump technology combined with flip-chip packaging can be as efficient in removing heat from a package as conventional schemes. Much of the heat passes through the bumps (usually gold or solder plated copper), and through the electrically conducting traces on the substrate, or through the substrate itself. Figure 11.6 shows copper bump flip chip technology on a power IC.

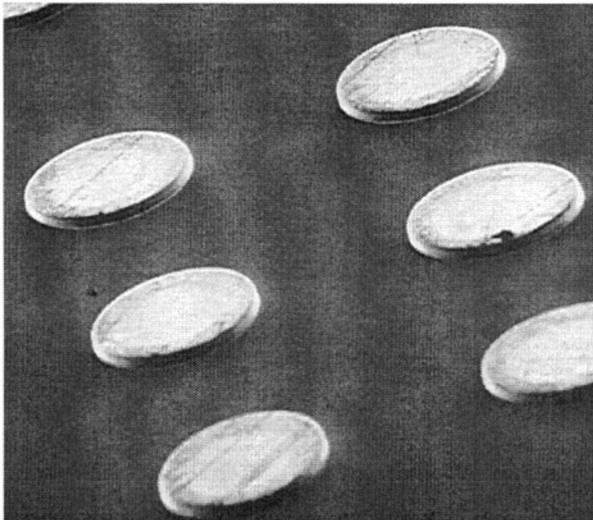


Figure 11.6: Photograph of copper bumps on a high thermal dissipation copper bump technology used for flip-chip packaging.

Figure 11.7 shows how a bumped IC mounts on to a flip-chip package. The bumps act as a low thermal resistance path to the leadframe. In many cases the ground leads of the package are positioned for maximized heat transfer.

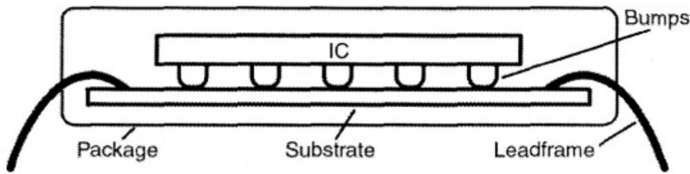


Figure 11.7: Improvement to thermal dissipation issues by the use of bump technology. The substrate is capable of high thermal dissipation, and contains patterned copper to connect to the leadframe.

A further development combines flip chip dissipation capabilities with that of conventional heat sink (Figure 11.8), enabling thermal dissipation from both sides of the silicon [11.14]. While not as effective as applying the same heat sink to bulk material, it improves the capabilities to levels adequate for most designs.

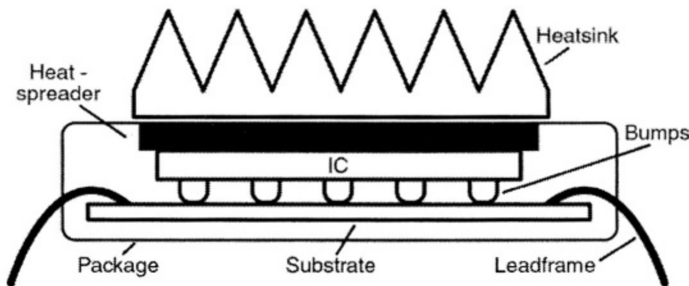


Figure 11.8: Package that permits heat removal from both sides of the silicon. This works by the use of bump technology to remove heat from the IC surface, and a heat spreader with heatsink to remove heat from the non-circuitry side of the chip.

11.5 Electrostatic Discharge (ESD)

On-chip ESD protection is standard on practically all integrated circuits. This is incorporated to improve reliability and yield. The three major ESD test methods are Human Body Model (HBM), Machine Model (MM) and Charged Device Model (CDM). The HBM simulates the action of a static discharge from a human body to ground. It uses a charged 100pF capacitor discharging through a 1500Ω resistor. The MM simulates static discharge from a machine to ground. It uses a charged 200pF capacitor discharging through a resistance of less than 1Ω. The CDM simulates charging and discharging events within production equipment. CDM is tested by charging an IC, then discharging through the package leads.

Damage due to ESD is usually thermal in nature although oxide breakdown may occur if the protection clamps are not properly designed. Therefore, typical protection design requires consideration of both the current path that can cause thermal damage and the subsequent voltage drop to avoid gate oxide damage. ESD protection elements have to dissipate currents that are considered high by the standards of operation of the IC, without failure. ESD circuits on SOI have the disadvantage that the oxide does not allow rapid thermal dissipation of heat generated in the element. Electrical and thermal models are routinely used to develop ESD circuits, and these have been modified for use with SOI [11.15, 11.16]. Studies indicate that the SOI diode, when compared with a bulk diode, has reduced junction area and is thermally isolated from the substrate heat sink resulting in significant self heating.

ESD protection networks and simulation have had to be modified for SOI applications. It was originally thought that ESD protection for SOI would be more difficult than for bulk. In part this was due to reduced capacitances and the fact that the substrate was no-longer available as an integral part of the ESD protection scheme. It has been demonstrated, however, that superior ESD robustness can be achieved without any increase in die area [11.17 - 11.19]. Industry standard ESD protection is generally achieved with gated and Zener diodes, and gate-coupled or body-coupled MOSFETS [11.20 - 11.26].

11.5.1 ESD Protection in Output Structures

When voltage is applied from an ESD transient to a logic output, a current path is created either to supply or ground, usually through the drain to body diode (figure 11.9). Additional, protection is afforded from the bipolar action of the structure, and from the turn-on of the MOS device in the reverse direction to supply. A vertical PNP to substrate can also turn-on directly to ground. Capacitive coupling can cause the parastic NPN of the output nMOS to turn-on momentarily.

With PD-SOI material, assuming no body ties are present, the diode is not present, and the bipolar devices can only turn on momentarily, since they are only capacitively coupled. For an ESD transient this may not matter, but for longer voltage transients, such as the automotive load dump, bipolars should not be considered as part of the protection capability. Protection is still obtained, however, from turn-on of the pMOS device. Figure 11.10 uses the example of a positive voltage spike, although a negative going voltage transient can be analyzed in a similar manner. Here, as the pMOS drain voltage raises above the source and gate, the device turns on in reverse (i.e.

the roles of source and drain switch). When the output to gate voltage rises above the Vt, the pMOS can turn on, sinking current into the supply. (For FD-SOI, the available paths are further limited, as there is no contribution from the capacitor-connected bipolar structure).

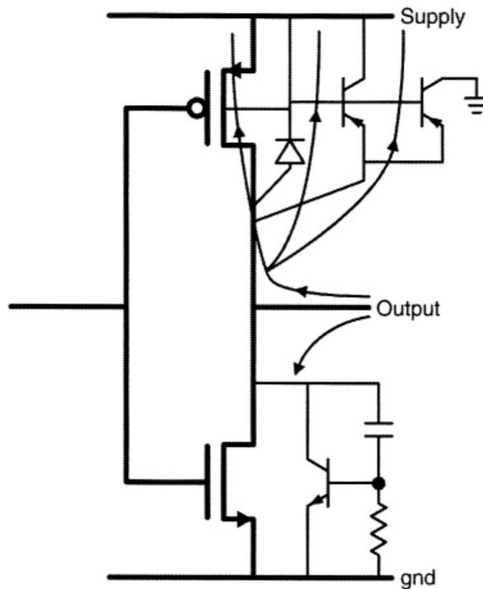


Figure 11.9: Bulk inverter output stage, showing available, current paths to ground, during a positive transient voltage spike.

While it may appear that SOI devices do not provide the same level of protection as bulk, there are two benefits which help offset the reduction in available paths. In SOI structures additional protection is provided through the fact that the output is not connected to ground or supply through a low resistance path. This allows the gate and body to float to some extent with respect to ground, providing a measure of protection not available from bulk material. In addition, the risk of destructive thyristor latchup is significantly reduced.

11.6 Radiation Hard (Rad-Hard) Circuits

Radiation affects circuitry in different ways, to a greater or lesser degree. Memory circuitry is highly susceptible, in particular because an entire memory array has to maintain radiation insensitivity even as cell areas reduce and array sizes increase. Memory elements, particularly those of dynamic RAM are particularly sensitive to the radiation effects, which can charge or discharge storage capacitors.

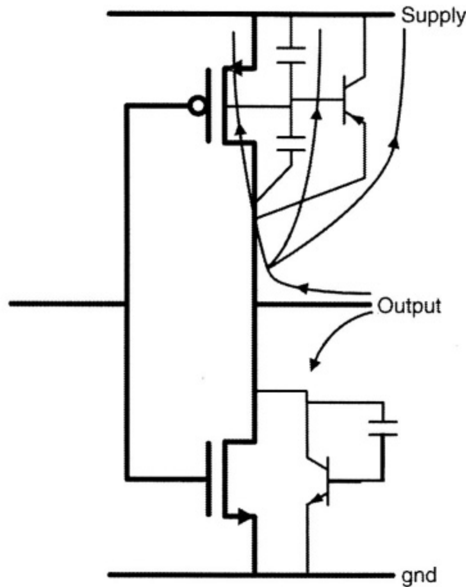


Figure 11.10: PD-SOI inverter output stage, showing available, current paths to ground, during a positive transient voltage spike.

Radiation insensitive components are required for computer systems in satellites, spacebourne systems and avionics. Numerous companies manufacture SOI components aimed specifically at rad-hard applications. Products available include CMOS, ASICs, SRAMs, ROMs, FIFOs, Non-Volatile Memories (MRAM), data bus interfaces and a limited number of analog circuits.

Threshold voltage shifts, caused by ionization and subsequent hole trapping in the gate, can affect all types of circuitry, through this is possibly less significant in digital than in analog designs, where a very small threshold voltage shifts can affect circuit operation. Within a digital system a small threshold voltage shift may not cause an affect on the device operation, particularly along a non-critical path. However, for analog circuits, a small change in Vt on a sensitive node can lead to a large change in circuit operating point [11.27], and cause parametric failure.

11.7 Reliability

There are several SOI specific aspects to reliability. The body voltage can go beyond the rail as a result of voltage pumping. This can potentially cause reliability issues, as the gate to body voltage may exceed the design limits for a substantial proportion of the operational condition. Gate oxide has been

demonstrated to be most susceptible to destruction at the same location as buried oxide leakage spots. Degradation of the gate oxide is the result of mechanical damage associated with deformation of the buried oxide [11.28].

Techniques for detecting reliability failures can be broadly split into "At-frequency" and "DC methods". "At-frequency" testing determines which ICs fall outside expected norms at given supply voltages, and temperature conditions. This determines failure due to certain abnormal resistive and leakage paths. These techniques tend not to be significantly different in their application from bulk. DC methods for reliability monitoring to detect failure mechanisms include quiescent supply current (IDDQ) testing. IDDQ analyses the DC leakage of an IC and variations of at-frequency test.

11.7.1 IDDQ (Quiescent Idd)

Excessive IDDQ may be caused by poor design practices, for example when a tristate node is allowed to float or when a bus conflict occurs. This becomes particularly problematic when the error-state occurs in a device sleep mode. IDDQ can also be high when process related problems occur, leading to excessive junction leakage or shorted transistors. Typically a design level IDDQ estimator is used. This determines leakage for static test cases. Excessive leakage is a signal that an unexpected leakage path exists, which is often a sign of a chip defect, which could cause failures later. Such units are screened out at test. IDDQ for SOI is similar, but some leakage paths are different. Clearly direct leakage paths to substrate no longer exist. Usually a leakage manifests itself from the body to source, and could be detected that way, but it might require more specific patterns for the IDDQ test vectors. Iddq is the current measured from Vdd with the device in a static or quiescent state. In a defect free circuit, the only current measured should be the device's normal off-state quiescent current. This should take the form of a distribution that should lie well below the customer requirement. If the circuit IDDQ is low enough, the IDDQ of outlier units susceptible to early failure is detectable with an IDDQ test as shown in figure 11.11.

Even with bulk material, due to node charging and capacitive effects, it takes tens of microseconds for a chip to settle into a state of true quiescence. Until the settled time is reached, the quiescent current is usually higher than the true IDDQ value. This is shown in figure 11.12. For floating body PD-SOI this time constant is extended, and can be milliseconds. This can cause excessively long test times, which is expensive. To avoid this, it is speculated that it may be possible to test IDDQ as a delta measurement,

while current is settling, by measuring IDDQ at two or more points during the settling time, and extrapolating the current trend.

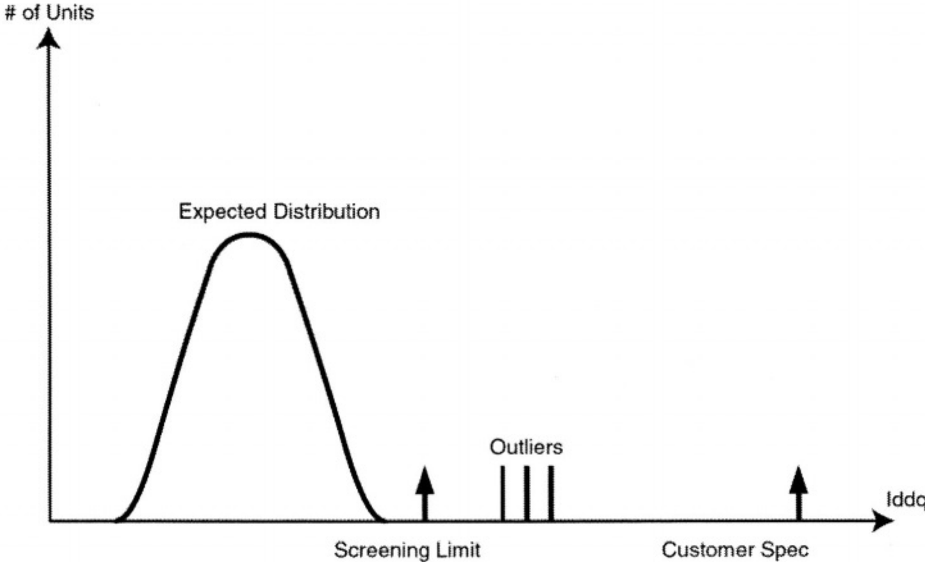


Figure 11.11: Graph of IDDQ showing expected distribution of IDDQ when the chip is in a quiescent state. The test limit is chosen to screen outliers, which are prone to early life failure.

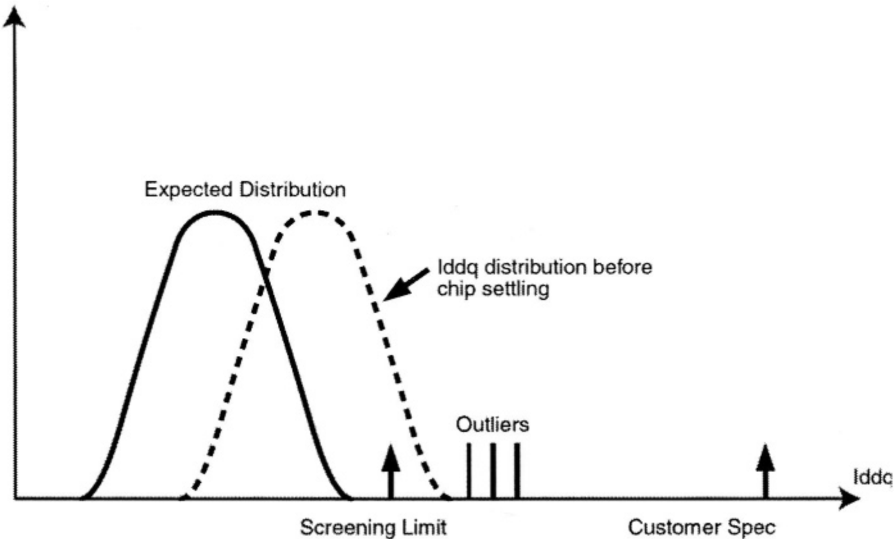


Figure 11.12: IDDQ limits before complete settling of the chip has occurred. This shows the potential to lose yield due to false detection of IDDQ failures.

Circuits on SOI can hide some faults that are easily tested in bulk CMOS, such as gate-oxide pinholes, since the current path is not to the substrate but to the body, then the source, and then possibly through another transistor that is off. As in the case of channel leakage in bulk-CMOS circuits, Iddq fault detection would require test patterns that turn on the transistors to complete the current path when detecting such failures. In contrast to channel leakage, for gate-oxide holes on SOI, the faulty transistor itself must be turned on rather than off to detect the leakage.

IDDQ is a useful measure of how likely a part may be to pass burn-in testing, or fail early in the product life. Various IDDQ tests are available and these are developed in detail in other works [11.29, 11.30]. It is worth noting, however, that the floating body of SOI material requires additional considerations when planning IDDQ testing. IDDQ usually detects excess oxide or channel leakage, or bridging faults in polysilicon or metal. Leakage current of SOI material may take several milliseconds to settle into a DC condition, due to the time constant of the body voltages. This can increase testing time.

As described above, SOI devices are not susceptible to diode leakage to substrate, or from body to source or drain, except at the gate edges because most of the device is oxide isolated. Figure 11.13 shows this for a NAND gate. Here a defect in the upper NMOS gate does not show up if the lower NMOS is off (figure 11.13a), whereas in conventional bulk CMOS the same fault would show up as a leakage to substrate even when the lower NMOS device is off (figure 11.13b). The same type of defect in a comparator input pair may manifest itself as a variation in offset, whereas in bulk material there would be a high gate to ground leakage.

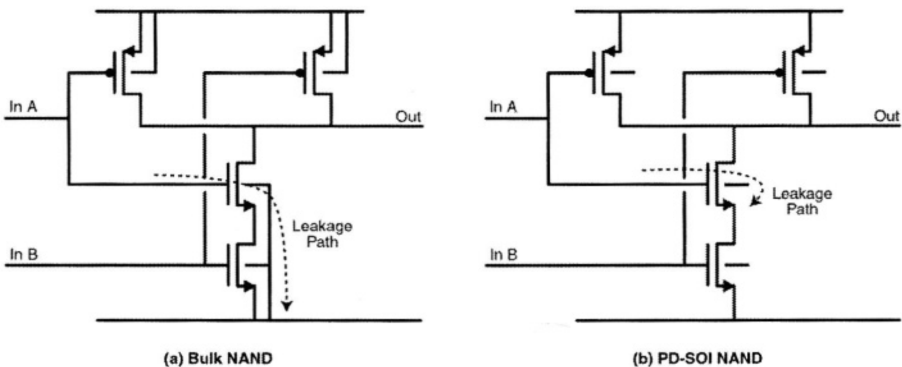


Figure 11.13: Nand gate configurations for bulk and SOI, showing defect leakage path for a gate fault in the upper NMOS structure. Under some conditions this will be observed as an IDDQ failure in bulk, but not in SOI.

11.7.2 Delay fault testing

The aim of delay fault reliability testing is to determine failure mechanisms, which can cause out-of-specification switching at either reduced or increased voltages [11.31, 11.32]. Often several conditions are tested beyond the normal operating limits, usually at a voltage just above the minimum theoretical operating point, and also at a high voltage, just below the maximum voltage sustainable by the gate oxides [11.30]. The expected variation of frequency with supply voltage depends on circuit type, but can be represented by a supply voltage vs frequency graph, for weak, nominal and strong device characteristics (figure 11.14).

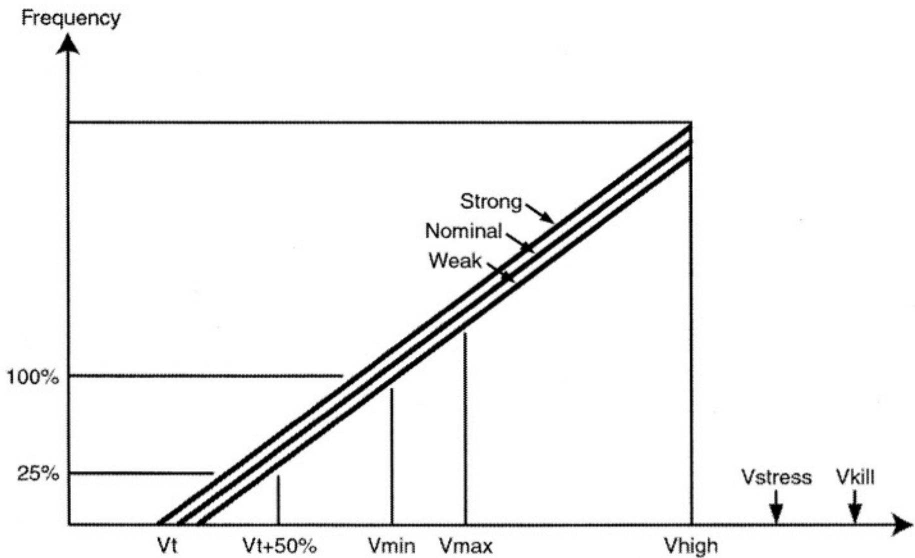


Figure 11.14: Graph showing conceptually how frequency of a critical path may change with frequency. To detect failures caused by bridging faults and resistive contacts, testing is done at voltages below and above normal operation.

The low voltage test point is applicable to SOI, and should not cause added testing difficulty, although the variation of critical path delays due to body voltage and threshold voltage variation needs consideration. At lower voltages this is not normally a problem but at the high voltage extreme, it may be an issue. At high voltages there is the possibility of a gate being pumped beyond the source, below ground for an NMOS and above the supply rail for a PMOS. Unless the at-frequency high voltage testing accounts for this there is a significant danger of exceeding gate breakdown

conditions. In addition, higher voltages lead to an increased variation of the path delay time, further increasing frequency uncertainty.

11.8 Package and Bond wire

Bond wires, package pins, printed circuit board traces, and IC metal all have parasitic inductances. When a SOI CMOS circuit switches, a current spike flows through the power bus. When current changes through an inductor, a voltage is generated across the inductor which is equal to $L*dI/dt$. This causes internal power rail ringing, which may cause errors. Ground bounce can adversely affect performance of the circuit, especially input-output circuits.

References

[11.1] C. Eisenhut & J. W. Klein, "SIMOX Voltage References for Applications up to 275°C using the Threshold Difference Principle", Proceedings 1997 IEEE International SOI Conference, Oct 1997, pp. 110-1.

[11.2] A. Marshall, “Operating Power ICs at 200 degrees”, IEEE Power Electronics Specialists Conference, Madrid, Spain 1992, pp. 1033-9

[11.3] Appendix 1: (11.3)

[11.4] S. K. H. Fung, et. al., “Impact of the Gate-to-Body Tunneling Current on SOI History Effect”, 2000 IEEE International SOI Conference, pp. 122-3.

[11.5] T. Efland, et. al., “The load dump (automobiles)”, IEEE Workshop on Electronic Applications in Transportation, 1990, pp. 73-78.

[11.6] A. Marshall & J. Devore, “Dual switch-mode regulator IC” International Solid-State Circuits Conference (ISSCC), 1995. pp. 52–53.

[11.7] D. W. Dobberpuld, et. al., "A 200-MHz 64-b dual-issue CMOS microprocessor", IEEE J. Solid-State Circuits, vol. 27, Nov. 1992, pp. 1555-1567.

[11.8] R. Berger, et. al., "A 1.3-GHz SOI CMOS Test Chip for Low-Power High-Speed Pulse Processing", IEEE Journal of Solid State Circuits, Vol. 33, No. 8, Aug 1998, pp. 1259-6.

[11.9] J-H. Lee & Y-J. Park, "High speed SOI buffer Circuit with the Efficient Connection of Subsidiary MOSFET's for Dynamic Threshold Control", Proceedings 1997 IEEE International SOI Conference, Oct 1997, pp. 152-3.

[11.10] C.Chen, et. al., "Single-transistor latch in SOI MOSFET's", IEEE Electron Device Letters, vol.9, 1988, pp.636-8.

[11.11] J. Gautier & A.J.Auberton-Herve, "A Latch Pheonomenon in buried N-body SOI NMOSFET's", IEEE Electron Device Letter, vol.12, 1991, pp.372-4.

[11.12] A. Marshall & F. Carvajal, “Power+ Logic Methodology Applied to a Six Output Power Driver”, IEEE 1993 Bipolar/BiCMOS Circuits and

Technology Meeting, Minneapolis, Minnesota, Oct 4-5, 1993. IEEE Cat No. 93CH3315-9, pp. 72-75.

[11.13] "Flip-chip bumping on 300mm wafers", European Semiconductor, Feb 2000, p38

[11.14] J. Klein, "Bottomless SO-8 Package Boosts MOSFET Performance", Power Conversion and Intelligent Motion, May 2000, pp. 110.

[11.15] Y. Wang et.al., "Electrothermal Modeling of ESD Diodes in Bulk-Si and SOI Technologies," EOS/ESD Symp. Proc., pp. 430-436, 2000.

[11.16] S. Ramaswamy et. al., "EOS/ESD Protection Circuit Design for Submicron SOI Technology," EOS/ESD Symp. Proc., pp. 212-217, 1995.

[11.17] G. G. Shahidi, et. al., "CMOS scaling for high performance applications" IBM J. Res. Development, vol. 39, Jan./Mar. 1995, pp. 229-244.

[11.18] S. H. Voldman, "The State of the Art of Electrostatic Discharge Protection: Physics, Technology, Circuits, Design, Simulation, and Scaling", IEEE Journal of Solid State Circuits, Vol 34, No. 9, Sept. 1999, pp. 1272-1282.

[11.19] P. Raha, et. al., "EOS/ESD Reliability of Partially Depleted SOI Technology", IEEE Trans. Electron Device., Vol. 46, No. 2, Feb. 1999, pp. 429-431.

[11.20] S. Voldman, et. al., "Dynamic Threshold Body and Gate-Coupled SOI ESD Protection Networks," Jounral of Electrostatics, 44, pp. 239-255, 1998.

[11.21] C. Duvvury, et. al., "ESD design for deep submicron SOI technology,"Symp. VLSI Technology Dig. Tech. Papers, 1996, pp. 194-195.

[11.22] J. C. Smith, et. al., "ESD protection circuits for TFSOI technology," Proc. 1996 IEEE Int. SOI Conf., 1996, pp. 170-171.

[11.23] S. Voldman et. al., "CMOS-on-SOI ESD protection networks", Proc. EOSIESD Symp., Sept. 1996, pp. 291-301.

[11.24] M. Chan, et. al., "Comparison of ESD protection capability of SOI and bulk CMOS output buffers" Proc. IRPS, 1994.

[11.25] K. Verhaege, et. al., "Double snapback in SOI NMOSFET's and its application for SOI ESD protection", IEEE Electron Device Lett., vol. 14, July 1993, pp. 326-328.

[11.26] "Analysis of snapback in SOI NMOSFET's and its use for an SOI ESD protection circuit", Proc. of the IEEE SOI Conf., 1992, pp. 140-141.

[11.27] T. P. Ma & P. V. Dressendorfer, Editors, "Ionizing Radiation Effects in MOS Devices and Circuits", New York: Wiley, 1989.

[11.28] Y. Minami, et. al., "Direct Observation of Gate Oxide Destruction due to BOX Breakdown in SOI" 2000 IEEE International SOI Conference, Oct. 2000, pp. 130-131.

[11.29] S. Chakravarty & P.J. Thadikaran, "Introduction to IDDQ Testing", Kluwer Academic Publishers, 1997

[11.30] B. T. Mitchell, "Operating-Extremes Test Improves Reliability", Test and Measurement World, Nov 2000, pp. 45-50.

[11.31] A. Krstic & K-T Cheng, "Delay Fault Testing for VLSI Circuits", Kluwer Academic Publishers, Boston, 1998.

[11.32] K. Baker, et. al., "Defect-based delay testing of Resistive Vias-contacts, a critical evaluation", International Test Conference ITC 1999, pp. 467-476

Chapter 12: Low Power Design

12.1 Introduction

Silicon on Insulator has a variety of advantages over bulk silicon for low power applications. These are mostly related to the reduced capacitance and higher current capabilities of SOI, at lower off-state currents. Power dissipation in a logic gate is approximately CV_{dd}^2. Load capacitance is typically 15% less in SOI than bulk, resulting in a 15% power reduction at constant Vdd if frequency is maintained constant. Further gains can be achieved by reducing supply voltage and maintaining switching speed, since the charge required is proportional to supply voltage in static logic circuits. Applied to 2.5V and 1.5V technologies, power can be calculated as:

$0.85*(1.5/2.5)^2 = 0.306$; i.e., power is reduced by 3.27x at constant delay.

Attempts to simultaneous reduce power supply and threshold voltages for low power designs and maintain performance reaches a point of diminishing returns as static power dissipation becomes a significant portion of total power. Dynamic threshold voltage control (modulating the back-gate bias to control gate thresholds) is sometimes used to extend the range of high performance and low power.

Temperature variations as a result of switching in SOI causes less than 2% delay-variation due to self-heating, and as supply voltage and power dissipation per device reduces, this becomes less important. Analog or high-power outputs should not be placed in close proximity to high frequency nodes as they adversely affect operating speed.

12.2 Clocking

As clock frequency increases, they must be distributed across a microprocessor, resulting in substantial power increase. Clock generation and distribution should be designed to reduce clock and system power demands.

12.2.1 Clock Generation

Low power microprocessors often use a phase lock loop (PLL) to achieve edge alignment between external clocks and the processor's internal clock. Increased wire widths achieve reduced resistance but have increased capacitance and power dissipation. A global clock is driven across a processor and generates the local clocks for the processor circuitry. Regenerating or buffering clocks locally controls global clock loading and reduces power.

Microprocessor designs often employ a single type of clock regenerator or buffer throughout the design, adding "dummy" loads to balance loading and control clock skew. Clock regenerators with different output drive strengths that drive predefined values of clock line capacitance should be used for optimal power dissipation. Clock regenerators are used to drive a group of latches, based on the local drive strength requirements. These measures help clock skew without redundant power dissipation.

12.2.2 Clock Distribution

An important factor in clock distribution is 'clock skew', which is defined as the difference in clock arrival times between signals across the chip. Distributed clock buffers dissipate more active power than a passive clock distribution network. However, passive clock distributions produce slower rise and fall times at the local clock regenerator inputs, degrading performance. Grid distribution networks can minimize clock skew but are prone to high routing capacitance, which increases power. A method that is effective in minimizing skew and controlling power is the H-tree clock network (figure 12.1). Although SOI technology offers improved performance and power over bulk CMOS, the design methodologies for bulk and SOI are similar. Overall speed improvement is higher in SOI than bulk CMOS due to the reduced junction capacitance and interconnect to substrate capacitance.

This network uses a recursive H structure to distribute the clock signal to the sub-blocks to maintain uniform synchronous clocks. The H structure is duplicated until the sub-blocks are small enough to produce tolerable skew within the sub-blocks [12.1]. A wider line is used for the main H structure to

minimize resistance, while a narrower line with successive branching points minimizes capacitance. Carefully controlled routing of R-C minimizes clock skew and power dissipation. H-tree distribution is very effective for processor designs as long as the depth of the H-tree does not lead to excessive routing capacitance.

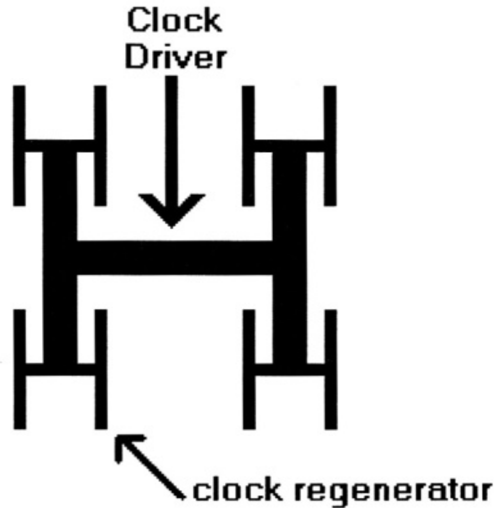


Figure 12.1: H-tree clock distribution network.

12.2.3 Clock Gating

On any given clock cycle, only some circuits in a microprocessor are active and dissipate power. In addition, only a subset of these circuits are active depending on the type of instructions being executed. Static circuits optimize power dissipation by disabling clocks to inactive areas of the processor. Efficient design involves optimizing static and dynamic circuitry throughout the chip. Disabling clocks to inactive circuits can be done dynamically.

Central control logic blocks or distributed logic can be implemented to dynamically control power distribution on chip. The distributed approach can be more efficient if logic to control clock gating and instruction execution are combined. Clock gating may also be used to implement low power modes used during processor idle times.

12.3 Options for Low Power

Various logic, circuit and floor-planning techniques have been used to achieve low power in CMOS digital design. Similar techniques can be used

for SOI digital and analog designs. Most of the low power design techniques can also be extended to microprocessor designs.

12.3.1 Static vs. Dynamic Logic

Both static and dynamic implementations have been used for lower power microprocessor designs. Static logic generally averages less switching activity, because dynamic logic nodes need pre-charging every cycle causing the nodes to switch frequently. Pre-charging internal nodes in SOI can deteriorate the dynamic operation of the circuit (chapter 7), and subsequently, driving a pre-charge signal can cause considerable power dissipation.

Static logic does not have the pre-charge requirement, but it can generate spurious transitions or "glitches" during a cycle as the function evaluates. Careful design techniques can minimize these glitches during transient operation. Static Logic is also susceptible to short circuit current. Dynamic logic does not generate glitches since each node can transition at most once every cycle. Depending on the implementation, it may be possible to have fewer intermediate nodes with dynamic circuitry, than with a static implementation [12.2].

12.3.2 Gate Sizing

Gates along a critical path are often sized based on to timing constraints and area considerations. Area reductions can reduce power by decreasing the load capacitance. All non-critical paths should be sized to match the delay in critical paths. Dynamic or fixed body ties can be used in critical parts of the design to enable faster and / or balanced switching in SOI.

In microprocessor designs, the critical path is frequently the worst-case pipeline stage delay, which determines the maximium operating frequency. In SRAMs, the critical path is from the word line through the sense amplifier and the output paths to the I/O. The word-line rise-time, bitline pre-charge, differential voltage at the sense amplifier and sense amplifier enable are the critical timing signals determining access through the memory array.

Floating bodies in the sense amplifier can hurt performance due to variability in body voltages. One option is to tie the body to the source. Another approach is to connect the body to the appropriate supply rail. Body ties will reduce variability in delay and noise in the sense amplifier. Another option, particularly applicable to low voltage operation is body pre-charging (chapters 8, 10). Bitline capacitance is somewhat reduced in SOI compared to bulk CMOS SRAM arrays. The ratio of bitline capacitance to gate

capacitance is important in determining bitline speed and differential voltage.

12.3.3 Minimizing Switching

Spurious switching during logic transitions can be minimized with logic. For example, re-ordering may reduce average switching activity. History effect can be used to hold logic outputs until a steady state has been reached. Instruction sets of an execution unit in a superscalar processor design is an example where latches can be used [12.2].

12.3.4 Interconnect

Interconnects can be a substantial percentage of the capacitance loading in microprocessors. Power related to interconnect may account for over 40% of the total chip power [12.3, 12.4]. Power and performance are critical issues in determining floorplan of a chip. The capacitance from interconnect to substrate is lower in SOI than bulk, and interconnect length can be reduced by optimal path layout.

12.3.5 Low Voltage Swing

An efficient way to reduce power consumption is to reduce the voltage swing required on high capacitance nodes (figure 12.2). Power consumption can be reduced by appropriate conversion of the signal between low voltage swing and full swing. This can lower both current and rate of charge / discharge of high capacitance nodes, and has the additional benefit of reducing driver size.

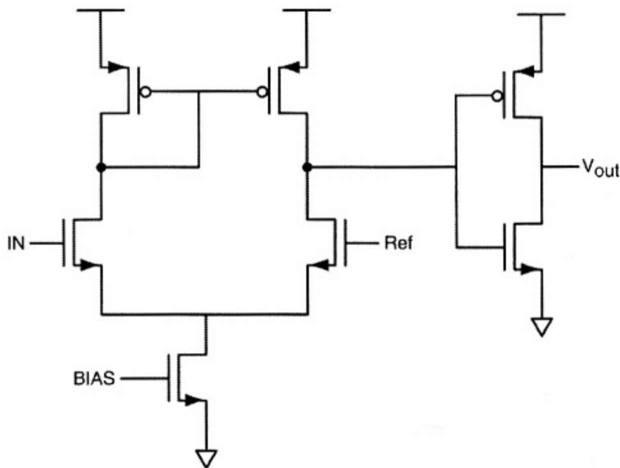


Figure 12.2: Circuit used to reduce capacitance voltage swing.

12.4 Analog Low Voltage Operation

Most SOI processes are designed for operation at 2.5V and below, and often down to around 1V. Specific low voltage design techniques are necessary at low voltages, and these have been discussed in some detail for bulk applications [12.5 – 12.7]. Many techniques for low voltage bulk design can be applied to SOI. Some circuit design schemes are covered here, although other techniques, requiring floating gates [12.8] may also be applicable, but usually require additional processing steps.

12.5 Floating Voltage Schemes

Floating voltage sources provide more efficient control of the voltage nodes in a circuit. A conventional current mirror (figure 12.3(a)) can work to relatively low voltages, but a floating voltage enabled current mirror (figure 12.3(b) is able to operate to somewhat lower voltages. This is achieved by adding a voltage source between gate and drain of MN1, so that the gate voltage of MN1 is higher than its drain voltage. In figure 12.3(a), the drain voltage of MN1 is close to the threshold voltage of the transistor. In figure 12.3(b), the drain of MN1 is below the threshold voltage of the gate by the potential of the voltage source. Cascoded SOI current mirrors benefit from reduced operation voltages as it further limits drain voltage swing of the current defining node, which in turn limits body voltage range. Limiting the body voltage reduces threshold voltage variation for the PD-SOI MOS devices [12.9].

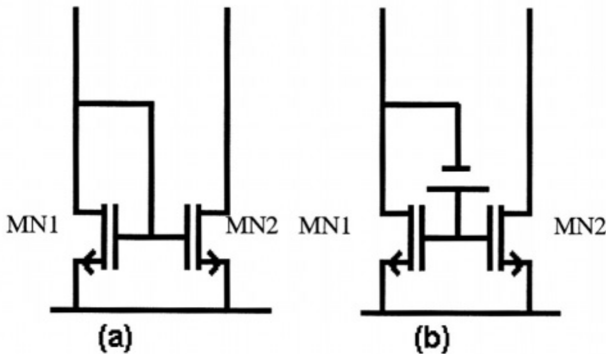


Figure 12.3: Conventional current mirror (a). The voltage on the gate of both mirrors is approximately the same, but the voltage headroom required on the input of the mirror is reduced in (b).

A floating voltage source is usually generated using two matched current sources and a resistance (figure 12.4). The current generated should not

interfere with circuit operation. In practice a node connected to the floating voltage source must be low impedance. Additional current from the floating voltage source results in higher circuit dissipation, however, this technique permits a simpler, lower power circuit, to be used in place of a more complex low voltage circuit, resulting in overall reduced power. The floating voltage source resistors may be more accurate over supply than bulk if tank resistors are used. Conversely, the current source and sink may be more susceptible to variation in SOI.

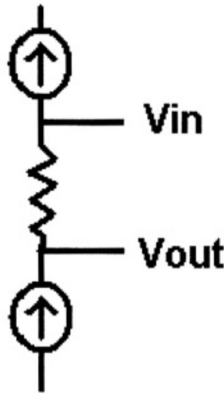


Figure 12.4: Basic configuration for floating voltage source. The current sources are matched, and current through the resistor defines the voltage. A low impedance tied to one end of the floating voltage source defines the voltage at the other end.

Floating voltage sources can also be applied to amplifiers and op-amp designs, with reduced bandwidth. A conventional PMOS op-amp will operate with a zero voltage input to about approximately a threshold voltage from supply. The input voltages can be artificially lowered by using floating voltage sources to operate the amplifier close to supply. In practice a second reference voltage is used rather than adding two floating voltage sources. Figure 12.5 shows the progression from the conventional form of the amplifier (a), to the form where both inputs are adjusted with floating voltage sources (b). Figure 12.5(c) shows a separate voltage source as a reference. Floating voltage sources are adjusted to optimize operation over the range required.

12.5.1 Low Voltage Output Operation

Conventional class A output stages can normally be made to operate without significant changes to topology, but more efficient class A-B operation is

often designed in the form of a linear inverter. Figure 12.6 (a) shows a conventional class A-B output. In order for the output stage to be operational in the required mode, the current circuit can be modified to include a floating voltage source. In the case of output stages it is standard practice to modulate the floating voltage source with supply voltage to avoid high current shoot-through at high voltages. The floating voltage source technique can be used in conjunction with other techniques used with output structures in SOI, for example modulated body circuitry (chapter 3).

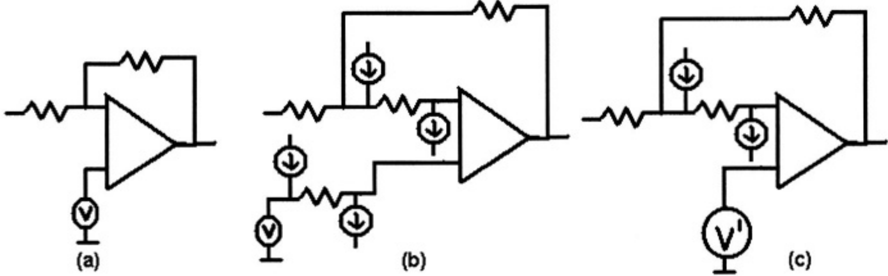


Figure 12.5: (a) shows a conventional op-amp. (b) shows input voltage adjustment using matching floating voltages on both inputs, and (c) shows the same implementation as (b), but with a modified reference supply to replace the original reference supply and floating source.

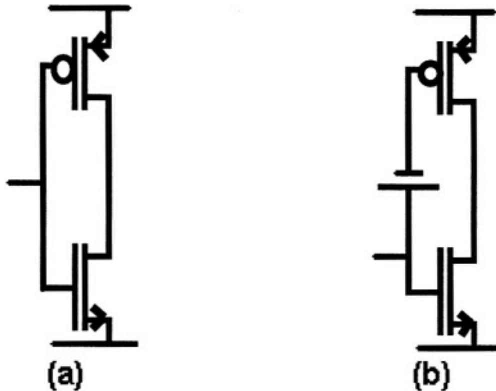


Figure 12.6: Class A-B output structure, (a) conventional structure and (b) structure with a floating voltage element to optimize the output performance

12.6 System Performance

SOI has performance advantages at circuit and chip levels. However, at chip level additional gains over the circuit level advantages are limited to improvements achieved from interconnect. Memory circuitry therefore has

little additional advantage at chip level, but for logic, significant reduction in interconnect can be obtained. Use of high resistivity substrates, not available to many bulk processes also improves performance, especially in high frequency applications.

When designing for low power, microprocessor performance is limited by system constraints. In PCs and workstations these constraints include energy conservation regulations and limitations to providing adequate cooling. Despite power limitations, notebook microprocessors must still approach desktop standards. Microprocessor performance therefore must be maximized for given power constraints. The development of low power technologies like SOI and scaling of device features sizes is important in reaching low power goals.

12.7 System Power Management

12.7.1 Low Power Standby Modes

Power management can be implemented at system level, with system logic or software controlling the power consumption of the processor and peripherals. Low power microprocessor designs have included a variety of 'sleep' modes to support system power management. These modes usually reduce the processor clock frequency during standby. This can be very effective, reducing standby power to a fraction of that consumed during normal operation.

During reduced clock frequency operation functionality is maintained and the processor can still service tasks that do not require full operating frequency. Processor clocks may only be stopped if the machine is maintained static, since dynamic designs lose the machine-state without clocking. Clocks are gated only in portions of the processor that need to be active or gated at the global level so no functions are active and maximum power is conserved.

12.7.2 Supply Voltage during Standby

During standby conditions it is often possible to reduce the power supply. This degrades operating performance, but is an acceptable practice when performance is not critical. It is also possible to reduce supply voltage to 0V if any necessary states can be stored off-chip. These can then be reloaded when activity resumes. It is economic to do this when the power and time required to save the processor-state is lower than the power reduction achieved.

12.7.3 Trade-Off for Power

It is very often difficult to estimate the trade-off between chip architectures, time to market and minimization of power. Standard cell libraries and re-engineering between process nodes limit power optimization, at the expense of delivery schedule. Architectural optimizations most commonly used in microprocessor designs are instruction execution pipelining, data-path design, caching and memory management all contribute to power trade-offs.

12.8 Instruction Set Architecture

A microprocessor instruction set may be defined to minimize the memory required to store a program. Since energy is required to fetch instructions, reduced program encoding, is generally more energy efficient. Reducing program size, or critical code loop, to fit into on-chip cache can cut the energy required to fetch an instruction by 50%. Reducing code size also reduces system power and cost by reducing memory requirements.

12.8.1 Instruction Complexity : RISC or CISC

A way of improving code density is to create more complex instructions so fewer instructions are required to code a function. Complex instruction sets computer (CISC) architectures were originally created to alleviate slow memory speeds, but were so slow that designers created more complex instructions to more efficiently use the time between instruction fetches. With higher speed memory, reduced instruction set computer (RISC) architectures with simpler instruction set can provide further performance improvement. The architecture design for SOI technology are similar to bulk except for power optimization circuits.

12.9 Reduction of Voltage below 3.3V

SOI technology offers power and performance advantages at lower supply voltages. As supply voltage is reduced to values closer to the MOS transistor threshold voltage, device current decreases with the square of the difference between the supply voltage and threshold voltage. Threshold voltage cannot be arbitarily reduced, since this causes leakage current to increase. For a fixed threshold voltage, as supply voltage is reduced, leakage current decreases. There are therefore advantages to reducing voltage. As a result the power saved by choosing a very low supply voltage must be weighed against possible performance losses and increased leakage currents affected by the choice of both the supply and threshold voltages.

References

[12.1] H. B. Bakaglu, "Circuits, Interconnections and Packaging for VLSI", Addision-Wesley, 1990

[12.2] Rabaey & Pedram, "Low power design methodologies", Kluwer Academic Publishers, ISBN: 0-7923-9630-8

[12.3] G. Gerosa, et. al., "A 2.2W 80MHz Superscalar RISC Microprocessor", IEEE Journal of Solid State Circuits, Vol 29, No. 12, Dec 1994, pp. 1440 - 1454.

[12.4] D. Liu & C. Svensson, "Power consumption Estimatino of CMOS VLSI Chips", IEEE Journal of Solid State Circuits, Vol 29, No. 6, June 1994, pp. 663 - 670.

[12.5] A. P. Chandrakasan, et. al., "Low-power CMOS digital design", IEEE J. Solid-State Circuits, vol. 27, Apr. 1992, pp. 473 - 484.

[12.6] E. A. Vittoz, "Low-power design: Ways to approach the limits", ISSCC 1994, pp.14 -18.

[12.7] P. Kinget & M. Steyaert, "Analog VLSI integration of massive parallel signal processing systems", pp.21-45, Kluwer Academic Publishers, ISBN 0-7923-9823-8, 1997.

[12.8] R. Gonzalez, et. al., "Supply and threshold voltage scaling for low power cmos", IEEE Journal of Solid-State Circuits, 32(8), August 1997, pp. 1210 –1216.

[12.9] J. Ramírez-Angulo, et. al., "A CMOS op-amps for a single supply close to a threshold voltage and with almost rail-to-rail signal swing", IEEE International Symposium on Circuits and Systems, 1999, pp. 408 – 411.

Chapter 13: SOI in Development

13.1 SOI Technology Roadmap

This final section is dedicated to circuits that are not yet proved commercially, and a collection of largely experimental or not yet mass-producible process or device concepts that may hold promise for improved performance, lower cost, higher density devices, or improved levels of integration. Circuit advances and revolutionary device structures, including ultra-thin channel SOI have the potential of making a breakthrough into the deep sub-0.1 µm channel length regime.

13.2 Device Enhancements

13.2.1 Enhanced-gate SOI MOSFET

Enhanced-gate SOI MOSFETs have gates on more than one side of the device. They take advantage of volume inversion, which helps to achieve higher current, increased trans-conductance, improved subthreshold swing, and attenuated short-channel effects [13.1]. These devices are more efficient than their single-gate counterparts, since most carrier flow is in the middle of the body, where the electric field is low and electron scattering caused by surface roughness is minimized. The gate-all-around structure (figure 13.1) is formed with a gate that completely surrounds the body region. Electrically, this is efficient, but is difficult to manufacture.

A simpler structure to the gate-all-around device is the wrap-around gate (figure 13.2). This provides a gate on three sides of the body, for an increased channel width. The multi-sided gate creates a more effective

channel, without adding proportionally to the drain or source capacitance, and is easier to fabricate than a complete wrap-around gate, since it does not require a gate region to be formed underneath the device.

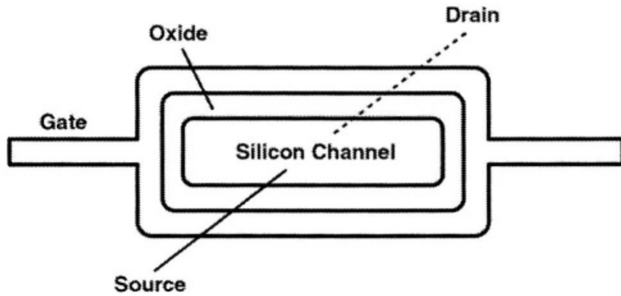


Figure 13.1: Cross section of gate-all-around device.

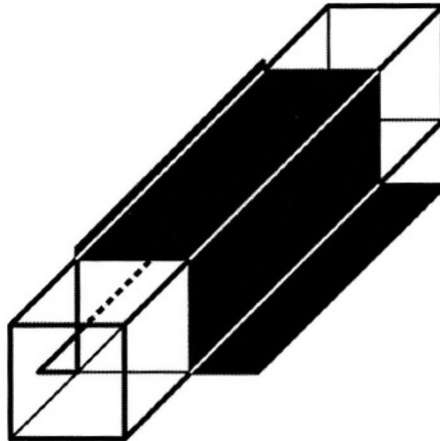


Figure 13.2: Example of wrap-around gate. This can be created in SOI by etching a trench around the active region, and filling the area with polysilicon, to effectively increase the gate width.

Processing of wrap-around gates is prone to create additional capacitances caused by difficulties in achieving accurate alignment of the gate. Excess gate to source or drain overlap can negate the advantages of improved device drive current, reducing the advantages of this process.

SOI technologies are being developed that include dual-gate and high-mobility channels, using Silicon Germanium alloy layers. These are aimed at increased operating frequency combined with decreased leakage [13.2]. “All-around-gate” and "dual-gate", as well as "thin-body" SOI devices

provide good control of leakage, which is useful for battery operated systems [13.2].

13.2.2 FinFET

The FinFET is a vertical silicon fin structure, with a self-aligned double-gate [13.3 - 13.4]. It is a CMOS process, and is capable of high current density, due to the stacked nature of the devices, which give a higher effective current density per unit area of silicon. The process has dual self-aligned gates, and devices provide good short channel performance down to 17nm gate length. The FinFET uses multiple-fins to achieve wide channels, and is capable of a higher width density than is possible with conventional planar structures.

13.3 Quantum Devices

Quantum effect devices, where one and two-dimensional carrier confinement has been demonstrated, have been available experimentally for a number of years [13.5, 13.6], but have not yet seen commercial development. They are structures in which one or more dimensions are short enough for the quantum-mechanical effects of electrons to be manifested. With extremely short active regions measuring 0.01-0.02 microns, these devices offer the possibility of switching speeds approaching 1,000 times faster than today's best semiconductor devices. Silicon-on-insulator technology can be used to demonstrate the Coulomb blockade, and lends itself to the fabrication of quantum devices because of its oxide isolation.

13.4 Stacked SOI

Multiple layers of SOI material can be stacked, allowing for an increase in density of gates per square micron. This is also known as three-dimensional (3D) SOI. Enhanced packing potential and reduced interconnect from stacked SOI layers have resulted in 3D SOI being a source of intense study. Multiple layers of SOI using a layer transfer process can be achieved in a variety of ways [13.7, 13.8]. Applications include advanced process integration, optical filters, micro-machining and sensors integrated with logic or DSP modules. It is difficult to produce stacked material, but if it can be manufactured, it has the advantage of reducing interconnect distances, and thus interconnect resistance and capacitance, resulting in reduced delay time.

The additional problems with this approach are even more severe thermal dissipation difficulties, matching between layers, the potential for electromagnetic or capacitive coupling between layers, and the probability of more severe bond pad limited designs.

Three-dimensional ICs composed of stacked, interconnected active circuit layers give higher circuit packing density (figure 13.3) [13.9, 13.10]. A 3D image sensor has utilized this technology. For this device two levels are processed together, as follows: One SOI wafer (the stacked wafer) is inverted on the patterned support wafer. The stacked wafer is ground down to the SOI oxide. Vias are etched through the stacked wafer to (a) the support wafer metal, (deep vias) and (b) the stacked wafer metal (shallow vias). A further level of metal is used to interconnect the vias being retrieved from within the stacked chip. Conceptually various other options exist for interconnecting stacked SOI.

Bulk material can substitute for the substrate and SOI layer 1. Additional layers are expected to be SOI, but can be PD, FD, or optimized for power devices, logic or analog.

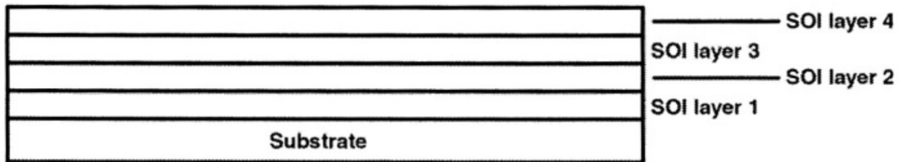


Figure 13.3: Stacked SOI cross section

13.5 Reduced Temperature Operation

One option to maintain the performance-scaling trend of CMOS is to reduce the chip ambient temperature. The behavior of floating-body PD-SOI has been evaluated to minus 100°C (173 Kelvin) [13.11], and found to give a higher than expected leakage. This is due to the floating body (or base) of the parasitic bipolar. This can be reduced through device design that minimizes the bipolar leakage, or by the use of body ties for critical components. Reduced temperature operation is desirable because at reduced temperatures MOS gain of MOS devices is higher, and interconnect resistance is reduced. This is also an approach being discussed for improving the performance of bulk material.

13.6 High Temperature Operation

An area where SOI appears poised to gain a niche market is for high temperature operation. This has been described briefly in chapter 10, which discussed the development of high temperature circuitry in applications up to at least 275°C [13.12]. High temperature operation is made possible by the reduction in leakage paths. For most applications partially depleted silicon can be used up to about 250°C, over 50 degrees higher than many bulk

applications [13.13]. ASICs have been demonstrated to 225°C [13.14]. Very high temperature systems appear likely to benefit from the use of fully-depleted silicon, where operation is considered possible up to 300°C or even higher.

13.7 New Circuit Designs for SOI.

13.7.1 Merged Bipolar / MOS

The isolated bipolar/MOS device, which has a common drain-collector and emitter-source is an interesting and novel device available to the user of PD-SOI. The gain curve is known to be similar to figure 13.4, which shows a graph of Ic Vs Ib as a function of Vgs. This permits the option of utilizing both sections of the device interdependently. This may be used to produce a super beta transistor configuration, or two independent inputs for a NOR gate or mixer, or as a novel input structure for an operational amplifier.

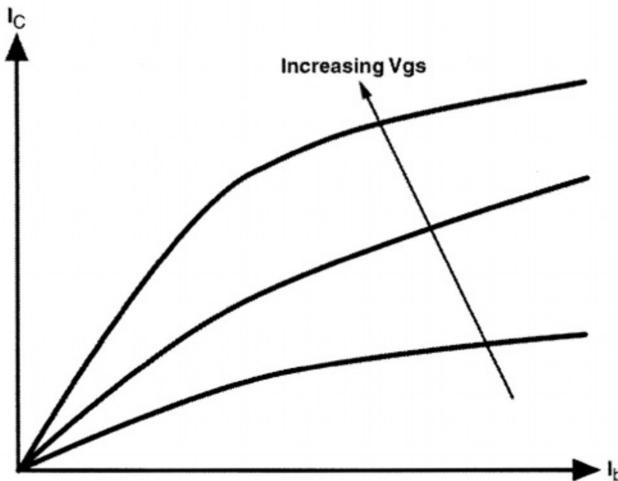


Figure 13.4: Graph of gain of merged MOS/bipolar device

13.7.2 Body Driven Operational Amplifier

Although there has been a great deal of proposed use of body connected components for logic circuits, there are also applications for analog circuits. Several body driven op-amps have been reported [13.15, 13.16]. One of these, a low-power CMOS operational amplifier can operate down to 1V (figure 13.5). This circuit was designed for use with bulk CMOS processes, based on the assumption that standard digital CMOS technology is expected to be a significant general purpose low-cost process for the foreseeable future [13.17].

The circuit is applicable to PD-SOI, as it uses the body as the differential inputs. There are several advantages of SOI for this circuit, over the same circuit in bulk. Input capacitance is reduced, problems associated with floating body nodes on the input pair are eliminated and either N-channel or P-channel devices can be manufacturered (compared to a bulk process, where the body of the NMOS device is shorted to substrate unless special processing is employed).

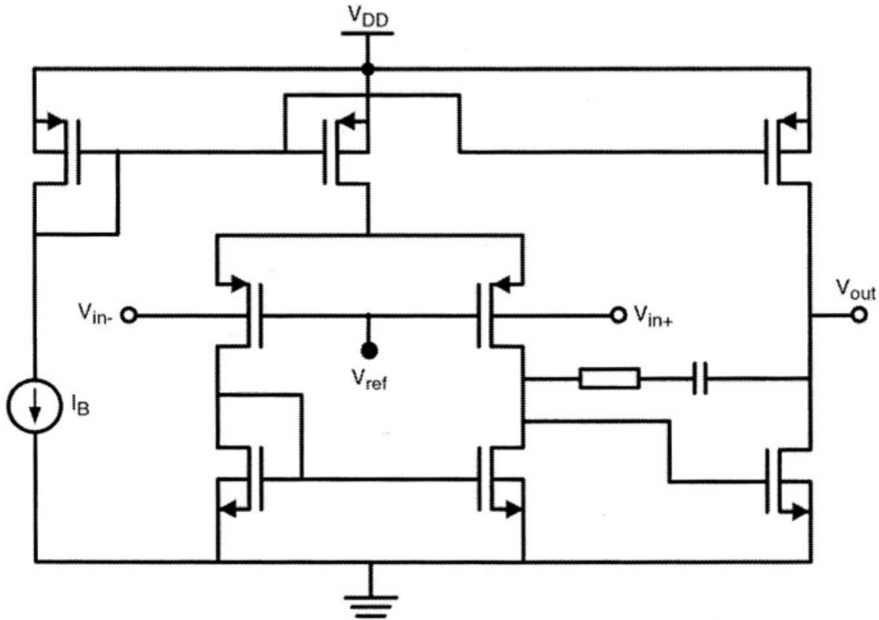


Figure 13.5: Operational Amplifier design using the MOS body contacts as input components. By varying the gate voltage of the input pair the input voltage range of the input pair can be adjusted.

As with conventional operational amplifiers, the stage most critical for matching is the input stage. The input range of a conventional opamp is limited near the positive or negative supply rail depending on whether a PMOS or NMOS input stage is used. By using bulk-driven FETs [13.18], the input voltage range can be adjusted depending on the voltage of the gates of the input pair. Voltage is applied to the bulk to modulate the transistor current flow.

With careful biasing and design this scheme can achieve rail-to-rail operation. In addition the input is not gate-connected, which may be an advantage for ESD protection. A benefit of this structure built in PD-SOI over the identical circuit for bulk is that both its NMOS and PMOS devices

can be body-driven, whereas in bulk CMOS typically only the PMOS can be body-driven.

Disadvantages of the circuit are that the voltage of the bulk to source junction must not exceed the diode forward voltage of the node. Trans-conductance of the devices operating in this regime, along with higher input capacitance, negatively impacts noise.

13.7.3 Body-input D-flip-flop

A body-input source-coupled-logic (SCL) D-type flip-flop circuit has been demonstrated, with a minimum operation voltage of 1.1V. This is achieved by applying an input signal to the SOI body [13.19]. Figure 13.6 shows the schematic for the flip-flop. The clk and $\overline{clk}$ signals are applied to the PD-SOI transistor bodies. The threshold voltages are modulated in accordance with the clock voltages, biasing the data input signal. The data evaluation and latch are thus switched by the difference of the drive capability between M1-M2 and M3-M4.

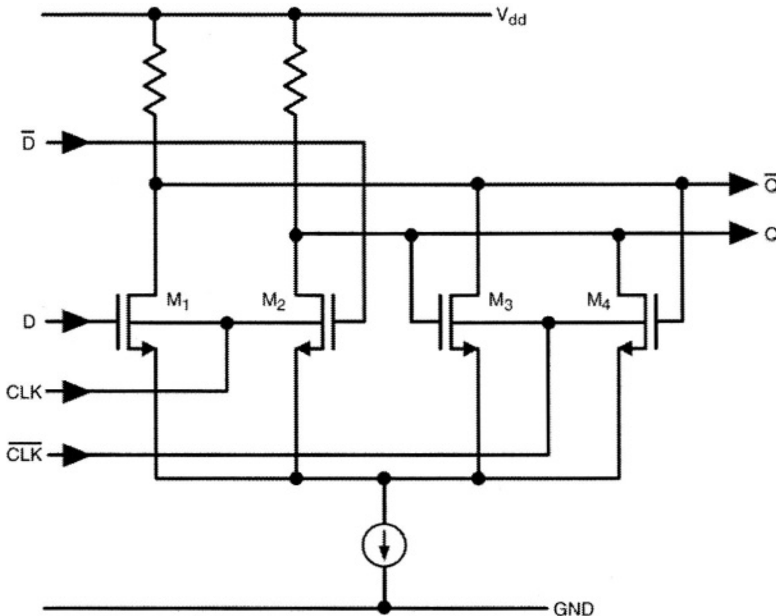


Figure 13.6: Schematic of a body-biased D-type flip-flop

13.7.4 SOI transistor as a DRAM

It has been proposed [13.20] that the floating body of PD-SOI may be used as a charge storage device similar to a DRAM. The floating body voltage affects the threshold voltage (Vt). If sufficient Vt shift occurs, it may in

principle be possible to make use of this technique as a dynamic storage technique, since in some states the body voltage can take hundreds of microseconds to relax to a steady state.

The structure described is half the area of a traditional 1T/1C DRAM structure, and does not require integration of a storage capacitor. Body bias manipulation relies on the presence or absence of a negative bias pulse on the PD-SOI transistor drain.

References

[13.1] S. Cristoloveanu, "Architecture of SOI Transistors: What's Next?", 2000 IEEE International SOI Conference, Oct. 2000, pp. 1-3.

[13.2] M.I. Current et. al., "Ultrashallow Junctions or Ultrathin SOI", Solid State Technology, Sept. 2000, pp. 66-77.

[13.3] D. Hisamoto et. al., " A Folded-channel MOSFET for Deep-sub-tenth Micron Era", IEDM Technical Digest 1998, pp. 1032-1034.

[13.4] X. Huang, et. al., "Sub 50-nm FinFET: PMOS" Int. Electron Devices Mtg., (IEDM)1999, pp. 67 –70.

[13.5] X. Baie et. al., "Quantum-wire effects in thin and narrow SOI MOSFETs", IEEE International SOI Conference, 1995, pp. 66 –67.

[13.6] J.P. Colinge, "Dual-gate SOI MOSFETs: Physics and potential", in "Silicon-on-Insulator Technology and Devices", Electrochemical Society Proceedings, Vol. 96-3, 1996, pp.271-286.

[13.7] R. M. Finnila, "Process of Manufacturing a Three Dimensional Integrated Circuit from Stacked-SOI Wafers using a temporary Silicon Substrate", United States Patent number 5426072, June 20th, 1995

[13.8] C. Maleville, et. al., "Multiple SOI layers by multiple Smart-Cut transfers", 2000 IEEE International SOI Conference, Oct. 2000, pp. 134-135.

[13.9] J. Burns et. al., "An SOI-Based Three-Dimensional Integrated Circuit Technology", 2000 IEEE International SOI Conference, Oct. 2000, pp. 20-21 .

[13.10] V. W. C. Chan, et. al., "Low-Power Stacked CMOS SRAM Cell with Large Grain Polysilicon PMOSFET Load", 2001 IEEE International SOI Conference, Oct. 2001.

[13.11] M. M. Pelella, et. al., "Low-Temperature Off-State Bipolar Effect in Floating-Body PD/SOI MOSFETs", 2000 IEEE International SOI Conference, Oct. 2000, pp. 84-85

[13.12] C. Eisenhut & J. W. Klein, "SIMOX Voltage References for Applications up to 275°C using the Threshold Difference Principle", Proceedings 1997 IEEE International SOI Conference, Oct 1997, pp. 110-1.

[13.13]. A. Marshall, "Operating Power ICs at 200 degrees", IEEE Power Electronics Specialists Conference, Madrid, Spain 1992, pp. 1033-9

[13.14] J. M. O'Connor et. al., "225C High Temperature Silicon-On-Insulator (SOI) ASICs for Harsh Environments", Proceedings of the IEEE International Workshop on Integrated Power Packaging, 1998, pp. 2-5

[13.15] B.J. Blalock, et. al., "Body Driving as a Low Voltage Analog Design Technique for CMOS Technology", 0-7803-5975-5/00 2000, IEEE

[13.16] K. Lasanen, et-al, "A 1V 5uW CMOS opamp with Bulk Driven Input Transistors", 43rd IEEE Midwest Symposium on Circuits and Systems, Lansing, Michigan, USA, August 2000.

[13.17] Semiconductor Industry Association, "National Technology Roadmap for Semiconductors," 1997.

[13.18] B. J. Blalock, et. al., "Designing I-V Op Amps Using Standard Digital CMOS Technology", IEEE Transactions on Circuit and Systems, Vol. 45, No. 7, July 1998, pp. 769-780.

[13.19] T. Fuse, et. al., "A 1.1V SOI CMOS Frequency Divider using Body-Inputting SCL Circuit Technology", 2000 IEEE International SOI Conference, Oct. 2000, pp. 106-7.

[13.20] S. Okhonin et. al., "A SOI Capacitor-less lT-DRAM Concept", 2001 IEEE International SOI Conference, Oct. 2001.

Appendix 1: Internet Sites (issue 1.0)

Due to the nature of this rapidly expanding and advancing field, in some cases we have made reference to web sites. While current at the time of publication, web-site information may change at any time. If the site you wish to visit is no-longer available please check on the website for this book. The website associated with this book is:

http://www.soidesign.info

(2.5) http://www.hitachi.com/semieqipment/simox.html

(2.11) http://www.soitec.com/pr33.htm

(3.18) D. Garner et. al., "Partial Silicon-on-Insulator for Power Integrated Circuits", http://www2.eng.cam.ac.uk/~www-edm/projects/partial/1.htm

(4.1) http://www-device.EECS.Berkeley.EDU/~bsimsoi/

(4.3) http://www.ecs.soton.ac.uk/~wrw/stag

(11.3) http://www.dice.ucl.ac.be/soi/activities/hightemp.html

Appendix 2: Trade Mark / Technology Information (issue 1.0)

This appendix points the reader to websites with more information on various processes, software etc. associated with SOI. For additional information and updates please visit the soidesign website (see appendix 1 for details).

BSIM	UC Berkeley	www-device.eecs.berkeley.eud/bsim_soi.html
Eldo (tm)	Mentor	www.mentor.com/eldo/overview.html
Eltran®	CANON	www.canon.co.jp/eltran
MATLAB®	MathWorks	www.mathworks.com
MINIMOS 6.1	U. of Vienna	www.iue.tuwien.ac.at/software/minimos%206.1/
Nanocleave(tm)	SiGen	www.sigen.com/wafers.html
Pisces	StanfordU	gloworm.stanford.edu
PROMIS 1.6	U. of Vienna	www.iue.tuwien.ac.at/software/promis%201.6/
PSPICE®	Cadencee	www.pspice.com
RAPHAEL	Avant!	www.avanticorp.com
SAP 1.4	U. of Vienna	www.iue.tuwien.ac.at/software/sap%201.4/
Fire&Ice®	Simplex	www.simplex.com
Sonnet	Sonnet	www.sonnetusa.com/products/index.asp
Smart Cut®	SOITEC	www.soitec.com/about_solution.htm
SmartSpice (tm)	Silvaco	www.silvaco.com
STAG	U. of Southhampton	www.micro.ecs.soton.ac.uk/stag
Star-Hspice(tm)	Avant!	www.avanticorp.com
Suprem	Stanford U	gloworm.stanford.edu
UFSOI	U. of Florida	www.soi.tec.ufl.edu/soispice.html
Unibond®	SOITEC	www.soitec.com/about_solution.htm
UTMOST IV(tm)	Silvaco	www.silvaco.com

INDEX

A

actuators 337
AcuThin 10
ADC 330
alpha particles 199, 227, 262
alternative models 70
amplifier 331, 336, 347, 366, 379
analog circuits 2, 25, 38, 290, 341, 350
analog multiplier 316

B

bandgap 277, 341
bipolar 25, 104, 221, 292, 344
bipolar effect 23, 116, 227
bleeders 23, 112, 119, 145, 227, 379
body bias 118, 169, 187, 345, 378
body capacitance 27, 69, 90, 151, 201
body charge sense amplifier 34, 265
body contact 192, 214, 227, 251, 264, 268, 339
body coupling 27, 120, 150, 226, 289
body diode 32, 69,167, 293, 348
body driven operational amplifier 375
body pre-charging 34, 362
body resistance 14, 26, 31, 51, 69
body tie 2,14,29, 76, 161, 187, 210, 345, 362, 374
body tied to gate 32, 161
body tied to substrate 31
body tied to unlike body 34
body voltage 66, 92
body voltage convergence 94, 101
body voltage response 101
BSIM 37, 63, 68, 341
BSIM-PD 64, 70
buried oxide 6, 14, 37, 227, 244, 290, 351
burn-in 141, 353

C

16, 18, 110, 201
capacitive body effects 26

capacitive coupling 24, 108, 166, 226, 251, 298
capacitor 29, 58
Cascode 81, 290, 306
cascoded mirror structure 81
cell capacitance 258, 275
cell library 339
cell stability 185, 197, 226
charge pump 269, 272
charge redistribution 129, 130
charge sharing 129, 151, 201
circuit noise 341
CISC 368
clock skew 136, 155, 202, 339, 360
clocked CMOS 149,150
clocked dual slope amplifiers 211
common mode 209, 226
components 250, 299, 314, 324, 350
compound structures 56
converting from bulk to SOI 49,77
cosmic ray 17, 227, 247
critical path 191, 260, 275, 339
cross coupling 81
current mirror 39, 80, 206, 209, 290, 308
CVSL XOR 147, 148

D

DAC 326, 327, 330, 331
decoders 201, 204
delay locked loop 317, 318
delta-sigma 331
destructive read out 229,281
device noise 34, 290, 319
device scaling 22
device temperature 39, 71, 312
differential amplifiers 207, 208, 210
digital circuits 37, 38, 118
digital signal processsors (DSP) 1, 125, 326, 373
diodes 34, 49, 77, 126, 289, 340, 348
distributed output stages 83
domino logic 132
doping profiles 14, 121
drain extended devices 53, 55

drain extended MOS 55
drain-source leakage 12, 16, 64, 340
dummy cell 263
dynamic 2way NAND 142
dynamic body 174, 213, 218, 292, 294, 311
dynamic circuits 117, 125, 153
dynamic data retention 246, 250, 253, 257
dynamic noise suppresssion 141
dynamic random access memory 218, 235
dynamic self-heating 37, 38, 46
dynamic wide OR 157

E

ELDO 63, 71
electrical isolation 13, 53
Eltran 9
embedded DRAMs 274
ESD 238, 347, 376
evaluate 127

F

fan-out 86, 108
FD-SOI 2, 19, 267, 319, 340, 349
filter 319, 331, 373
FINFET 373
flash converter 327
flip flop 182, 198, 320, 377
flip-chip 346
floating body 16, 119
floating gate 47, 75, 364
folded cascode 309
four-transistor (4T) SRAM 187
full adder 165

G

gate dielectric 22, 29
gate induced drain leakage 68
gate leakage 22, 68, 88, 313, 340
gate thickness 27
gate tunneling 27, 118, 341
gated diode structures 77, 78

Gilbert mixer 316
glass substrate 257
glitches 362

H

half Vdd precharge 259
Hall effect 334
heat spreading 53, 346
Heteroepitaxial Substrate 5
high current 49
high side drive 50
high voltage 49, 67
high-resistivity substrates 289, 367
historical perspective 3
history dependence 29, 117
history effect 17, 89, 125, 226, 319

I

IDDQ 11, 64, 166, 351
impact ionization 251, 290, 345
inductance 59
inductive loads 52
inductor 6,59
input pair 77
interconnect 323, 363
inverter 22, 75, 89, 366
isolation 243, 257

J

junction capacitance 16, 85, 258, 360

K

keeper 113, 141
kink effect 18, 25, 76, 302

L

latch 148, 151, 160
latch-up 25, 61, 289, 302, 341
latency 254, 267
lateral DMOS 49, 51, 333
lateral PNP 79, 80

layout 49, 75, 80
LDMOS 49, 333
LDO 325
leakage currents 27, 127, 145, 176, 190, 195, 235, 340
leakage mechanisms 27, 69, 90, 249
LNA 314
local oscillator 313
local thermal heating 37
logic noise 341
loop filter 319
low device threshold 16

M

matching 11, 25, 77, 214, 291, 373
material quality 31
merged bipolar 375
microprocessors 1, 85, 119,185,372
microwave 2, 324
mixer 314, 316, 375
mobility 4, 21, 190
models 63, 64, 68
MOS device characteristics 272
MOS I-V curve 37
MOS transistor noise 34
MOSFET models 71
multiplexers 115
multithreshold CMOS 162

N

NAND gate 96
Nanocleave 11
noise analysis 36, 95
non-overlapping clocks 158
NOR gate 23, 117
NORA 136

O

operational amplifier 303, 322, 375
OR gate 132, 146
OR-AND 104, 106
oscillator 37, 66, 272, 314, 320

OTA 36, 300, 309, 321
output drivers 37, 280
output Stages 83, 304, 311, 365
oxide isolation 13, 18, 58, 313, 373

P

parasitic bipolar 21, 49, 112, 125, 149, 374
parasitic capacitance 59, 69, 161
parasitic gain 64
parasitic NPN 23, 51
parasitic resistance 69, 203, 223, 339
parasitic thyristors 50
partial SOI 53
partial trench 14, 81, 258
pass gate 19, 113, 194, 254
passive Components 18, 57, 314
patterned Insulators 6,7
PD-SOI 2, 11, 35, 79, 125, 236, 290, 311
PD-SOI process 14, 295, 318
phase detector 319
phase locked loop 319
physical thinning and etchback 8
pipelined ADC 329
power amplifier 333
power device 7, 374
power IC 49, 53, 295, 346
precharge 206
process cross-section 13, 14
pulse stretching 92, 153

Q

Q factor 315
quantum devices 373
quartz 5, 60, 324
quasi SOI 246

R

radiation hardness 17, 18, 228, 244
reactive ion etch (RIE) 246
reduced short channel effect 16
refresh 249

regulator 272, 325
remapping 130
re-ordering 130
resistive Body Tie 2, 33
resistive load 182
resistors 57, 182, 289, 321, 365
RESURF LDMOS 52
RF 2, 49, 289, 313
ring oscillator 37, 67, 85, 318
RISC 368

S

S/D-to-substrate capacitance 22, 309
sample and hold 312, 313
scaling 14, 33, 119, 240, 367
selective SIMOX 7, 53, 83
self body biased 187, 188
self Heating 18, 36-39, 70, 226, 310, 346
sense amplifier 206
sense amplifier mismatch 220
sensors 333
shallow trench 14
short channel 16, 182
signal to noise ratio 260
silicide block 49, 78
silicon on anything 6, 60
silicon on Insulator - introduction 13
silicon-on-diamond 5
silicon-on-quartz 61
silicon-on-sapphire 5
silicon-oxide-silicon 6
SIMOX 6
Small signal equivalent circuit 71
SmartCut 246, 247
soft error rate (SER) 181, 197, 198, 227, 241
SOI Device Properties 22
SOI Materials 5
SOI Modeling 63
SOI simulations 63, 64
SOI Technology Advantages 16
SOI wafer 6, 8, 9, 10
SPICE 38

SRAM architecture 203
SRAM cell mismatch 225
SRAM read operation 261
stacked SOI 373
STAG 38,71
static circuits 94, 130, 361
static data retention 246, 249
static inverter 89
static noise 198
static random access memory 181
static timing 36
substrate capacitance modeling 22
substrate contacts 13
substrate ties 15
supply noise 16, 198, 226
supply rail 18, 77, 129, 225, 301, 354
switching effects 23,343
synchronous 258, 276

T

technology Scaling 21, 294
temperature effects 71,339
thermal coupling 81
thermal coupling 73, 81
thermal dissipation 37, 294, 312, 333, 347
thermal effects 21, 294, 321
thermal model 37
thermal response 29
threshold Voltage 236
transconductance 36, 300
transmitter 313, 316
tunneling current 27,43, 118

U

UFSOI 64, 68, 71
ultra-thin' film SOI 21

V

Vbb pump 273
Vccp pump 272
VCO 313

voltage reference 300
Vt variation 120, 124

W

wafer-bonding 245
wireless 313
word line boost 267
word line drivers 268, 275
write operation 191

X

XOR 107

This page intentionally left blank

About the Authors

Andrew Marshall is a Distinguished Member of Technical Staff at Texas Instruments Incorporated, Dallas, TX. Since joining Texas Instruments in 1984 he has worked on power transistor design, bipolar and CMOS mixed signal IC design and power integrated circuit development. He is currently involved in performance evaluation of advanced process development, including circuit design for SOI in the Silicon Technology Development Group. Dr. Marshall holds more than 25 US patents and has 35 publications. He is Chairman of the Dallas chapter of the IEEE Circuits and Systems group. He is also a senior member of the IEEE, a fellow of the Institute of Physics and a Professional Engineer.

Sreedhar Natarajan is a Member of the Group Technical Staff at Texas Instruments Incorporated, Dallas, TX. Since joining Texas Instruments in 1997, he worked on 64Mb DRAMs and embedded SRAMs designs. Prior to joining TI, Mr. Natarajan worked as a Design Manager responsible for SRAM development at Paradigm Technologies. He has various publications and over 20 filed and issued patents. His interests are low power and high speed memory designs. Mr. Natarajan obtained his M.S in Computer Engineering from University of South Western Louisiana, Lafayette. Mr.Natarajan is an active TI representative at the Joint Electron Device Engineering Council (JEDEC). Mr. Natarajan is currently serving on the ISSCC memory Sub-Committee, CICC eMemory committee and ICECS program committee. Mr. Natarajan has been an invited speaker at various IEEE international conferences. He is also on the technical advisory council for many universities on memory related projects. Mr. Natarajan is the recipient of the IEEE Circuits and Systems Outstanding Service Award'01. He is also the chairman for the Dallas Chapter of the IEEE-Solid State Circuits. Mr. Natarajan is a senior member for the Institute of Electrical and Electrical Engineers.