ABSTRACT

DEVASTHALI, VINAYAK SUDHAKAR. Application of Body Biasing and Supply Voltage Scaling Techniques for Leakage Reduction and Performance Improvement of CMOS Circuits. (Under the direction of Professor Paul Franzon).

The efficiency of body biasing technique is evaluated in 90-nm process technology for regular and low threshold voltage devices. A new leakage monitor circuit for detecting an optimum reverse body bias voltage is designed. The simulation results shows that the monitor circuit accurately tracks the leakage currents within ±5% of the actual leakage current values. The standby leakage reduction in static CMOS circuits using reverse body biasing is presented. The results indicate that the reverse body biasing is more beneficial for high speed circuits using low threshold voltage devices. For circuits using nominal threshold voltage devices, the efficiency of reverse body biasing decreases due to the presence of gate leakage. Speed improvement in ring oscillator and ripple carry adder using forward body bias is measured. The results show that the forward body biasing is less effective due to the lower body effect parameter. Supply voltage scaling technique for active power reduction is implemented using 180-nm technology. Power savings up to 50% is achieved by scaling the supply voltage as per the operating frequency requirements.

Application of Body Biasing and Supply Voltage Scaling Techniques for Leakage Reduction and Performance Improvement of CMOS Circuits

by

Vinayak Sudhakar Devasthali

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Electrical and Computer Engineering

Raleigh, North Carolina

2007

Approved By:

Dr. W. Rhett Davis	Dr. Kevin Gard

Dr. Paul Franzon
Chair of Advisory Committee

Biography

Vinayak Devasthali was born on July 26, 1983 in Maharashtra, India. He obtained his bachelor's degree in Instrumentation Engineering at Vivekananda Education Society's Institute of Technology, an affiliate of the University of Mumbai, in June, 2004. Since August 2005, he has been a master's student, majoring in Electrical Engineering, at North Carolina State University, Raleigh, USA. After graduation, he plans to join Qimonda as a DRAM Design Engineer, at Burlington, Vermont.

Acknowledgments

I would like to take this opportunity to thank the people, who have directly or indirectly helped me in completing my thesis work.

To start with, I thank my family for supporting me throughout my academic career. Their faith has helped me to stay positive during tough times.

I am grateful to Dr. Paul Franzon for agreeing to be my thesis advisor. It has been a privilege to work under his guidance. I highly appreciate his advice and support when most needed. I thank my thesis committee members, Dr. Kevin Gard and Dr. W. Rhett Davis, for their valuable comments and suggestions regarding my thesis. They have been extremely patient with me during the thesis work.

A special round of thanks goes to Shep Pitts and Shrivatsan Parthasarathy for the enlightening technical discussions. It was in these discussions that I found the seeds of my thesis topic. I am thankful to Steve Lipa for providing me a workplace and for being tolerant of me. I would like to thank Dr. H Joel Trussell for offering me a teaching assistantship, which provided a huge financial relief and helped me concentrate on my thesis work. I also thank the entire staff of ECE department for their help throughout my graduate studies at NCSU. In particular, I acknowledge Elaine Hardin's help in completing the thesis paperwork and Kaye Bailey's assistance in arranging a conference room for my thesis defense.

I would also like to thank Jean Appleman and Michael Kleiner of Qimonda for their understanding and support during the completion of this work.

Finally, I express my sincere gratitude towards my friends at NCSU for all that I have learnt during my stay in Raleigh. They have been a huge support to me for the last two years. A special thanks to Manish Sachdev for providing me the LaTeX template for this thesis, to Saurabh Gupta and Jingzhen Hu for teaching me swimming and driving respectively, to Po-Chih Lin for numerous rides and to Ajit Gopalakrishnan for an infinite number of reasons. These people constitute a major motivating force for me in accomplishing whatever I could. I thank them all from the bottom of my heart for the memories that I shall cherish forever.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

With the continuous scaling of CMOS technology, the effects of fabrication process variations and operating temperature variations are becoming more prominent. Process variations affect the parameters such as transistor channel lengths, oxide thickness and dopant concentrations. Fluctuations in these parameters increase the delay and leakage of the circuit. Temperature variations cause the mobilities of electrons and holes and the threshold voltage to decrease. This leads to change in on-state current I_{on}, which, in turn, affects the speed of the circuit ([Kumar06]). In addition to leakage and speed, power consumption is also an important circuit design consideration. Rapid growth of portable electronic devices demands reduction in total power consumption to extend the battery lifetime. Therefore, it has become necessary for circuit designers to develop techniques for minimizing leakage and power dissipation while maintaining the optimum circuit speed.

[Kuroda96] proposed a variable threshold voltage (VT) scheme based on body bias control for leakage and active power reduction. Body biasing is a promising technique for improving performance of the circuit[Arnim05]. Forward body biasing reduces the threshold voltage of a transistor and thus increases the on state current. This improves speed of the circuit during active mode. Reverse body biasing during standby mode raises the threshold voltage, thereby reducing the subthreshold leakage in the circuit. Thus, dynamic body

biasing can be used to meet the performance requirements in low power, high speed CMOS technologies.

In active mode, dynamic power dissipation dominates the total power consumed by the digital circuit[Kuang03]. The dynamic power is proportional to the clock frequency and the square of the supply voltage. Lowering the clock frequency alone reduces the power, but it does not save energy consumed in every cycle[Wei99]. The fixed supply voltage means that every operation is carried out at the same voltage level[Rabaey04]. The clock frequency and the supply voltage are fixed depending on the maximum speed requirements while executing computationally intensive, low latency tasks. In portable battery-powered devices, the intensive tasks are few and sporadic[Soto03]. These devices are idle for large portion of their time and hence, the peak performance is not needed throughout all time. Significant energy saving is, therefore possible by scaling the supply voltage depending on the performance requirements.

1.2 Process Requirements

In bulk CMOS technology, the substrate (p-type) is common for all NMOS transistors. NMOS transistors can not be operated as four-terminal devices because of their shared body connection. In order to access the bulk terminal of individual NMOS transistor, each NMOS must have its own p-well. This requires an additional implant layer in fabrication process. Hence, body biasing can be implemented only in triple-well technology. Figure 1.1 shows the structure of a triple-well. The PMOS and NMOS transistors are fabricated in a separate n-well and p-well respectively. An extra thick N-isolation layer separates the individual p-wells from the common p-substrate and avoids any excess parasitic-bipolar current. The layer also breaks the resistive path from any digital noise source on chip into the analog circuits. This is particularly helpful when using separate V_{DD} and V_{SS} for the digital and analog circuits.

1.3 Research Goal

The goal of this research is to evaluate the efficiency of body biasing in 90nm CMOS technology and to implement a supply voltage regulator with a feedback control

loop, which can adjust the supply voltage based on the operating frequency of the circuit. The leakage in CMOS circuits arises from different mechanisms. Each leakage component has different dependence on the threshold voltage. Hence, a thorough understanding of each leakage mechanism is required for evaluating the effectiveness of body biasing in leakage reduction. An optimum reverse bias voltage can minimize the total standby leakage. The scope of this work includes the design of a leakage monitor circuit for detecting the optimum reverse bias, analysis of the performance improvement achieved using body biasing and the design of the supply voltage scaling scheme using feedback control and buck converter.

1.4 Thesis Organization

This thesis is organized as follows:

Chapter 2 discusses the important leakage mechanisms in CMOS circuits. Effects of threshold voltage, drain-to-source voltage and input sequences on leakage components are explained. The techniques for leakage reduction are briefly covered.

Chapter 3 discusses the body biasing technique. Effect of body biasing on the leakage components is explained. Design of leakage monitoring circuits for detecting an optimum reverse bias voltage has been discussed. Standby leakage reduction using the optimum reverse bias in various digital circuits is measured and analyzed. Improvement in

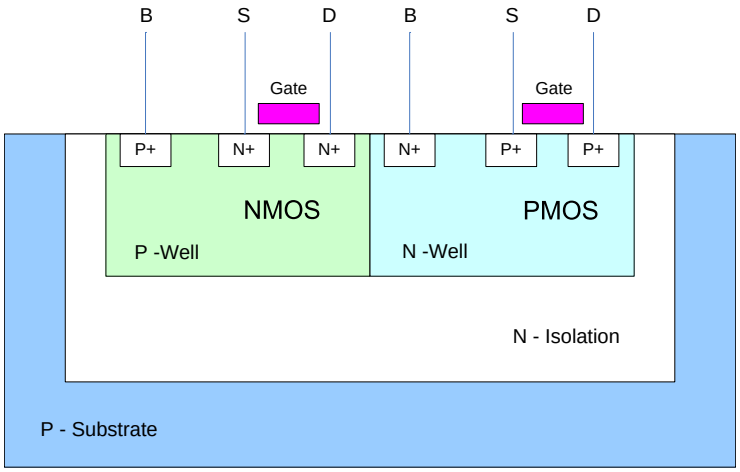

Figure 1.1: Triple-well Structure

on state current and circuit speed under forward body bias is also discussed. Regular and low threshold voltage devices (reg-V_{TH} and low-V_{TH}), available in the 90-nm process, are considered to investigate the efficiency of body biasing.

Chapter 4 describes the implementation of supply voltage scaling scheme. The operation of a synchronous buck converter is explained. The pulse width modulation technique, used for controlling the power switches in the converter stage, is briefly discussed. The supply voltage scaling technique requires a negative feedback control loop for adjusting the supply voltage at the desired level depending on the frequency of operation. The design of each component of such control loop is explained. Finally, the performance of a ring oscillator, used as a replica of the critical path in the circuit, under supply voltage scaling is discussed. The benefits of supply voltage scaling technique depend on the power conversion efficiency of the converter stage. Zero voltage switching (ZVS) technique for improving the efficiency of buck converter is discussed.

Chapter 5 concludes this thesis. The research contributions of this work and directions for future work are discussed.

Chapter 2

Leakage Mechanisms in CMOS circuits

CMOS technology has been scaled down to achieve higher integration density and performance. As technology scales down, leakage currents are becoming increasingly important and hence must be taken into account to minimize the total power consumption. Leakage currents are particularly important in design of portable battery-operated devices. In such devices intermittent computational activity is separated by long periods of inactivity. During the period of inactivity, ideally, the device should consume zero power, which is possible only if there are no leakage currents.

The major leakage mechanisms contributing to the overall leakage include subthreshold leakage, gate leakage and junction band-to-band tunneling leakage. The different leakage paths are shown in Figure 2.1. Each of these leakage components have different dependence on device properties, operating temperature and supply voltage.

2.1 Subthreshold Leakage

The current in MOS transistor does not drop to 0 at $V_{GS} = V_{TH}$. The diffusion current flowing between the source and drain terminals for $V_{GS} < V_{TH}$ is called the sub-

threshold current. Figure 2.2 shows a plot of I_D vs V_{GS} for 90-nm NMOS transistor on logarithmic scale.

From Figure 2.2, it is clear that the transistor is already conducting in the region $V_{GS} < V_{TH}$. This conduction for voltages below V_{TH} is known as subthreshold conduction or weak inversion. The drain current is exponential dependent on the gate-to-source voltage in weak inversion region. The expression for the subthreshold current is given by Eq.(2.1)[Kao02].

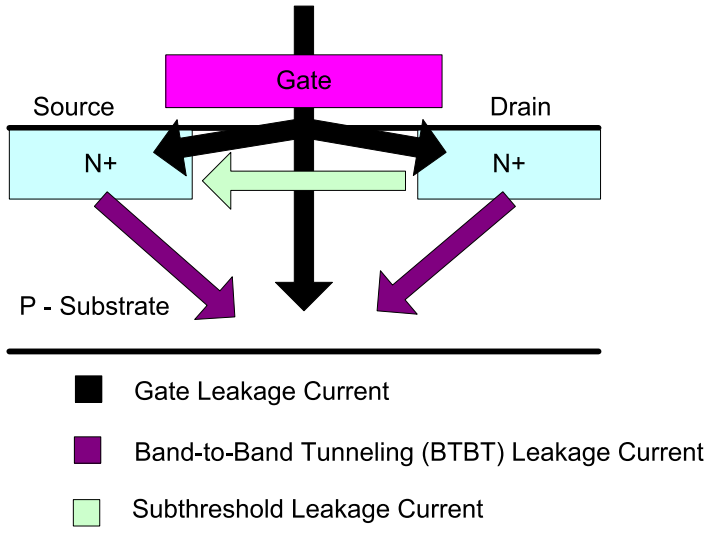

Figure 2.1: Different Leakage Paths in CMOS Circuits

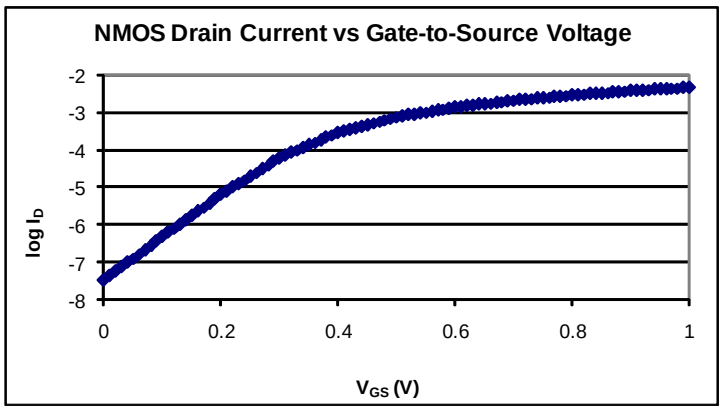

Figure 2.2: I_D vs V_{GS} plot(logarithmic scale) for NMOS ($W = 5\mu m$, $L = 80nm$, $V_{TH} = 245.2mV$) in 90-nm process

$$I_{SUB} = I_S \cdot 10^{\frac{V_{GS}-V_{TH}}{S}} \left(1 - 10^{-\frac{n \cdot V_{DS}}{S}}\right) \tag{2.1}$$

where I_S and n are empirical parameters. S is the *slope factor* of the device. It measures by how much V_{GS} has to be reduced below V_{TH} for I_{SUB} to drop by a factor of 10. S can be given in mV/decade as[Rabaey04]

$$S = n \left(\frac{kT}{q}\right) \cdot ln\,(10) \tag{2.2}$$

where kT/q is the thermal voltage V_T. For typical value of $n = 1.5$, $S = 90$mV/decade[Rabaey04].

Subthreshold current is undesirable in digital circuits since it prevents the MOS transistor from behaving like an ideal switch. Especially, subthreshold leakage can degrade the performance of dynamic circuits, which rely on the storage of charge on a capacitor.

2.1.1 Threshold Voltage Dependence of The Subthreshold Current

In order to meet the power and reliability requirements, it is necessary to reduce the supply voltage with technology scaling. Scaling the supply voltage while keeping the threshold voltage constant results in performance degradation, especially when V_{DD} approaches $2V_{TH}$. To maintain performance, threshold voltage needs to be scaled down as well. However, the subthreshold leakage increases exponentially with reduction in V_{TH}, as seen in Eq.(2.2). The 90-nm process provides regular threshold voltage devices (reg-V_{TH}) and devices with low threshold voltage (low-V_{TH}). The threshold voltages of reg-V_{TH} and low-V_{TH} NMOS transistors are $396.9mV$ and $245.2mV$ respectively. Figure 2.3 shows a logarithmic plot of the drain current of reg-V_{TH} and low-V_{TH} NMOS transistors against the gate-to-source voltage. It can be seen that the subthreshold leakage in low-V_{TH} transistor is almost 20% higher than that in an identical reg-V_{TH} transistor.

For low leakage during standby mode, higher threshold voltage is required, whereas better performance during active mode demands lower threshold. To satisfy these contradicting requirements, circuit techniques such as body biasing need to be implemented.

2.1.2 Stack Effect

Subthreshold leakage is also a function of the circuit topology and the combination of the input signals. For example, consider a 2-input NAND gate shown in Figure 2.4.

In single-well (common p-substrate) process, body terminals of both the NMOS

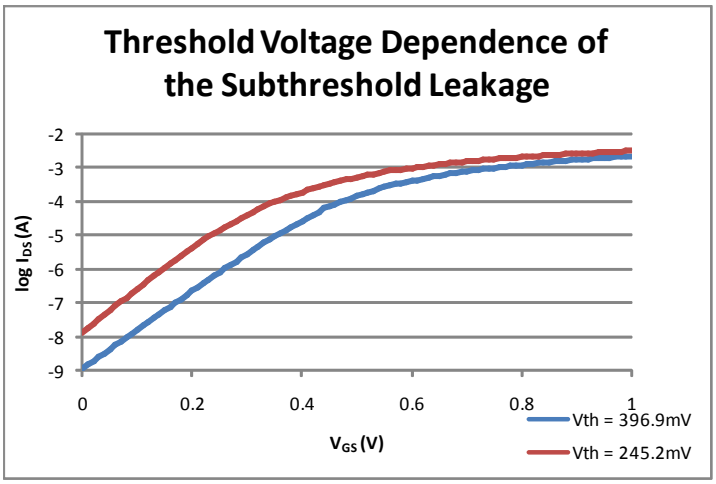

Figure 2.3: Threshold Voltage Dependence of Subthreshold Leakage. The drain current of NMOS transistor ($W = 5\mu m$, $L = 80nm$) in 90-nm process against the gate-to-source voltage on a logarithmic scale

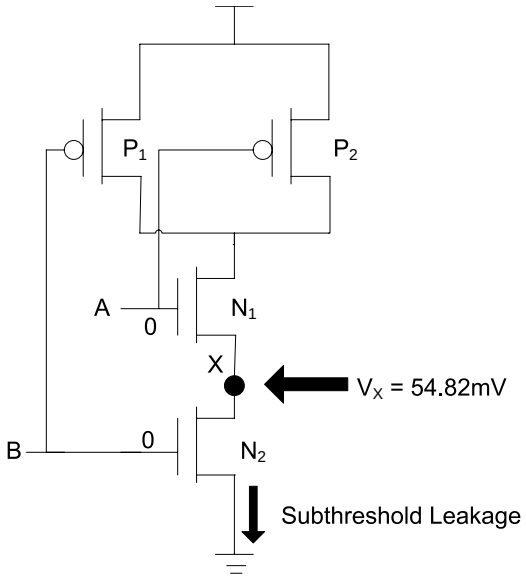

Figure 2.4: Schematic of 2-input NAND Gate

Table 2.1: Subthreshold Leakage Current in 2-input NAND Gate in 90-nm Technology

Input Sequence	V_X	I_{SUB}
00	55mV	326.5pA
01	990nV	19nA
10	820.2mV	1.594nA
11	4μV	57.65nA

transistors are connected to ground. Due to raised source potential, V_{BS} is negative for N_1 and therefore, N_1 is more reverse-biased that the bottom transistor, N_2. This increases the threshold voltage of N_1 transistor. In stacked transistors configuration, increase in the threshold voltage is caused by the "*self-reverse biasing effect*".

The maximum leakage reduction occurs when all the stacked transistors are off. The leakage current of the gate is then determined by the topmost transistor with the highest threshold voltage. In 2-input NAND gate, when A = B = 0V, the leakage current depends on the reverse bias potential ($V_{BS} = -V_X$) of N_1. The subthreshold leakage currents for 2-input NAND gate in 90-nm technology for all the input combinations are listed in Table 2.1.

2.2 Gate Leakage

In scaled technologies, the gate oxide thickness T_{OX} is reduced to provide sufficient current drive and to minimize short-channel effects. The resultant high electric field permits large number of electrons to tunnel through the insulating oxide layer. This tunneling results in significant gate leakage current, which has an exponential dependence on T_{OX} and the potential difference V_{OX} across the oxide. The leakage current density can be given by the following equation[Agarwal05].

$$J_{gate} = A \cdot \left(\frac{V_{OX}}{T_{OX}}\right)^2 \cdot exp\left[-B\frac{1-(1-V_{OX}/\phi_B)^{3/2}}{V_{OX}/T_{OX}}\right] \tag{2.3}$$

where $A = q^3/16\pi^2 h\phi_B$, $B = 4\sqrt{2m_{ox}}\phi_B^{3/2}/3hq$, ϕ_B is the tunneling barrier height in eV and m_{ox} is effective carrier mass in oxide.

The gate leakage current is composed of following components:

- gate to substrate leakage current, I_{gb}

- gate to source/drain overlap region leakage current I_{gs0} and I_{gd0}
- gate to channel leakage current, I_{gc}. It can be divided into I_{gcs}, which flows to source and I_{gcd}, which flows to drain.

Each of the above gate leakage current component can be expressed using Eq.(2.3). The three major direct tunneling mechanisms in MOS gates are shown in Figure 2.5. In NMOS device, electron tunneling from the conduction band (ECB) governs I_{gc} in inversion mode and I_{gb} in accumulation mode. Electron tunneling from the valence band (EVB) controls I_{gb} in depletion mode. For PMOS devices, holes to valence band tunneling (HVB) controls I_{gc} in inversion mode. In the absence of channel, I_{gb} is controlled by EVB in depletion region and ECB in accumulation mode. As shown in Figure 2.5, the barrier height (ϕ_{HVB}= 4.5eV) for holes tunneling is much larger than the barrier height (ϕ_{ECB} = 3.1eV) for electrons tunneling. Therefore, holes require much higher energy to tunnel through the oxide layer. Hence, gate leakage current for PMOS transistor is one order of magnitude less than that of an NMOS transistor with identical T_{OX} [Agarwal05].

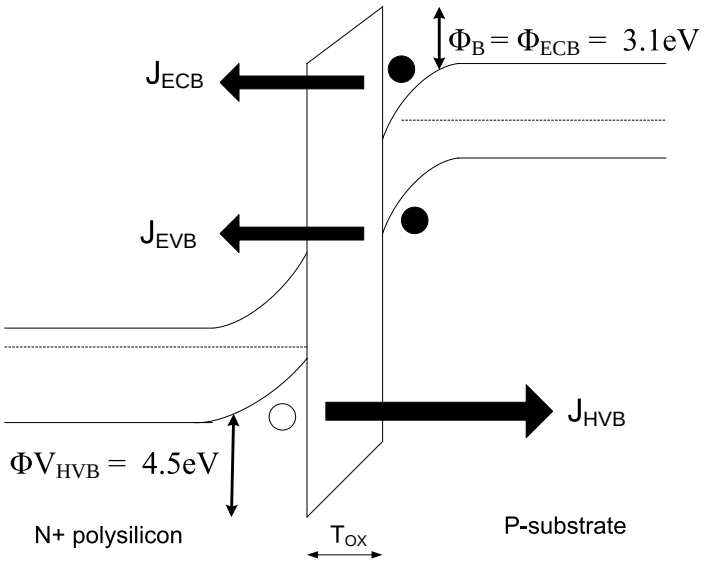

Figure 2.5: Gate Leakage Tunneling Mechanisms

2.2.1 Input Sequence Dependence of Gate Leakage Current

Gate leakage current has a strong, exponential dependence on V_{GS} and V_{GD}, which leads to input sequence dependence. In case of a simple inverter, the maximum leakage current occurs when gate is at V_{DD} and source and drain are at ground potential. As the input voltage is gradually decreased from V_{DD}, the gate leakage current falls rapidly. At the same time, drain voltage increases which results in negative V_{GD}. This leads to reverse gate tunneling from drain to gate, when the transistor is off. However, due to the absence of a channel, this tunneling is limited to the gate-drain overlap region. Since this overlap region is much smaller that the channel, the reverse tunneling leakage can be ignored[Lee04].

To illustrate dependence of gate leakage current on inputs, consider a stack of 3 NMOS transistors with 3 different input sequences as shown in Figure 2.6. The on-state gate leakage current through transistor N_M in each case is discussed in the following[Lee04].

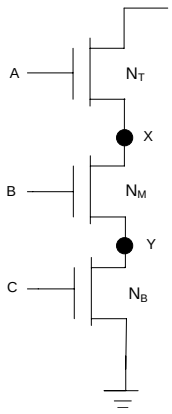

Figure 2.6: Stacked NMOS transistors in NAND3 Gate

- Case 1: Input Sequence - 011

In this case, transistor N_B is conducting and hence both the internal node voltages V_X and V_Y are close to ground potential. This makes $V_{GS} = V_{GD} = V_{DD}$ and equal amount of leakage current flows from gate to source and drain.

- Case 2: Input Sequence - 110

In this case, transistor N_T is on and both V_X and V_Y are close to $V_{DD} - V_{TH}$ potential.

Table 2.2: Gate Leakage Current in 3 Stacked NMOS Transistors for Different Input Sequences in 90-nm Technology. Leakage Currents are in $nA/\mu m$

ABC	$\|I_{gA}\|$	$\|I_{gB}\|$	$\|I_{gC}\|$	$\|I_{gT}\|$
000	9.06	5.18	3.89	18.13
001	9.06	5.18	8.00	22.24
010	9.16	4.88	3.99	18.03
011	9.06	8.06	8.00	25.12
100	0.01	5.77	3.89	9.67
101	0.01	5.74	8.00	13.75
110	0.01	0.02	4.37	4.40
111	10.5	9.59	9.16	29.25

The V_{GS} and V_{GD} voltages for N_M are much smaller in this case and hence gate leakage is significantly reduced.

- Case 3: Input Sequence - 010

Both N_T and N_B are off in this case. N_T has higher threshold voltage due to self-reverse biasing and is strongly turned off. The voltages V_X and V_Y are in the range of 100-200mV. The V_{GS} and V_{GD} voltages of N_M are close to V_{DD}, which results in significant gate leakage current.

Any conducting transistor in stacked transistor configuration will fall into one of the above three cases. For transistor in parallel with same gate input, the leakage current can be calculated by replacing with a single transistor with an equivalent size equal to the sum of sizes of all the parallel transistors. Table 2.2 lists the gate leakage currents in a stack of 3 90-nm NMOS transistors for all input sequences. The absolute values of gate leakage currents are considered.

2.2.2 Interdependence of Subthreshold Current and Gate Leakage

In a simple inverter, when input is high, the gate leakage of NMOS device can be combined with the subthreshold current of PMOS device to obtain the total leakage current. These two currents can be calculated separately and added together. However, in stacked transistor configurations, the subthreshold current and gate leakage current combine at the internal nodes and affect the node voltages. The leakage currents are interdependent in such cases. To study the interaction of gate leakage and subthreshold current, consider a 3-input

NAND gate. The total leakage current for different input sequences can be computed as follows.

- Case 1: Input Sequence - 000

All NMOS devices in series are nonconducting. The subthreshold leakage current is minimum in this case due to stack effect. The top NMOS transistor N_A, which has the highest threshold voltage, determines the subthreshold current. The PMOS transistors in parallel exhibit gate leakage current which can be added to the NMOS subthreshold current. As mentioned before, gate leakage of PMOS devices is much smaller than that of NMOS devices and therefore, can be ignored.

- Case 2: Input Sequence - 001

The internal node Y is connected to ground through N_C. $V_{GS} = V_{GD} = V_{DD}$ for N_C, which results in a large gate leakage current. This leakage current can be combined with the subthreshold leakage of nonconducting N_A and N_B transistors to obtain total leakage current.

- Case 3: Input Sequence - 010

The internal node voltages V_X and V_Y are in in the range of 100-200mV. Transistor N_B exhibits significant gate leakage current, which combines with the subthreshold current of the top N_A transistor and causes an increase in the node voltages V_X and V_Y. Increase in V_X reduces the subthreshold current of N_A and also the gate leakage current of N_B. However, the subthreshold current drops more rapidly as compared to the gate leakage current. Hence, the total leakage current is less than the sum of gate leakage current and the subthreshold current[Lee04].

- Case 4: Input Sequence - 011

NMOS devices N_B and N_C exhibit maximum gate leakage in this case. Both V_{GS} and V_{GD} voltages are close to V_{DD} for these transistors. The internal nodes X and Y are close to ground potential and the gate leakage current does not affect these voltages. Therefore, the total leakage current equals the sum of gate leakage and the subthreshold current of topmost transistor in stack, N_A[Lee04].

- Case 5: Input Sequence - 100

The internal node X is connected to the output, which is at nominal V_{DD} potential, through the conducting N_A transistor. The node voltage V_X would be close to $V_{DD} - V_{TH}$. The gate leakage of N_A in this case is negligible, since V_{GS} and V_{GD} are very small. The total leakage current in this case is dominated by the subthreshold current of the nonconducting transistors, which is determined by N_B transistor.

- Case 6: Input Sequence - 101

The internal nodes have a conducting path, in this case and the node voltages V_X and V_Y are close to $V_{DD} - V_{TH}$ and ground potential respectively. N_C shows significant leakage current compared to N_A, since both V_{GD} and V_{GS} are close to V_{DD}, which can be added to the subthreshold current of N_B.

- Case 7: Input Sequence - 110

In this case, both the internal nodes have a conducting path to output through N_A and N_B. The node voltages V_X and V_Y equal $V_{DD} - V_{TH}$. The gate leakage currents of both N_A and N_B are negligible in this case. The total leakage current is dominated by the subthreshold leakage of N_C[Lee04].

- Case 8: Input Sequence - 111

All NMOS transistors are conducting in this case. The output and the internal nodes are at $0V$, which results in maximum gate leakage from all stacked transistors. In addition, the PMOS transistors in parallel exhibit subthreshold leakage which must be added to the gate leakage to obtain total leakage current. This is the worst case leakage for NAND gate.

2.3 Junction Band-to-Band Tunneling (BTBT) Leakage

In MOS transistors p-n junctions are formed by drain-substrate and source-substrate, which are normally reverse biased. The high electric field across the p-n junction allows electrons to tunnel from the valence band of the p-region to the conduction band of the

n-region as shown in Figure 2.7. This band-to-band tunneling (BTBT) results in a minority diffusion/drift current across the junctions. The tunneling current density can be given by[Neau03]

$$J_{BTBT} = A\frac{EV_{RB}}{E_g^{1/2}} exp\left(-B\frac{E_g^{3/2}}{E}\right) \tag{2.4}$$

$$A = \frac{\sqrt{2m}q^3}{\pi h^2} B = \frac{8\pi\sqrt{2m}}{3qh}$$

where m is effective mass of electron; E_g is the energy band-gap; V_{RB} is the applied reverse bias voltage across the junction; h is Planck's constant; E is the electric field at the junction and q is the electronic charge. Assuming a step junction, the electric field across the junction is given by[Neau03]

$$E = \sqrt{\frac{2qN_AN_D\left(V_{RB} + \Psi_B\right)}{\epsilon_{si}\left(N_A + N_D\right)}} \tag{2.5}$$

where N_A and N_D are the doping concentrations on p and n side of the junction, respectively; Ψ_B is the built-in potential across the junction and ϵ_{si} is the permittivity of silicon. The tunneling occurs when the total voltage across the junction (V_{RB} + Ψ_B) is higher than the silicon band-gap energy (E_g).

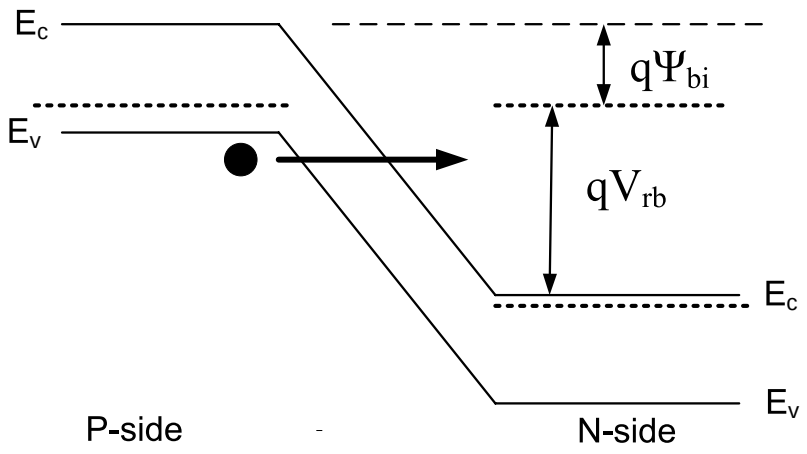

Figure 2.7: Band-to-Band Tunneling across a reverse-biased p-n junction

The BTBT leakage is a function of the junction area and the doping concentration. In scaled devices, heavily doped substrate and halo doping profiles are used to suppress the short-channel effects. This causes a significant increase in the BTBT leakage.

2.4 Leakage Reduction Techniques

Techniques for leakage reduction can be broadly classified into design-time techniques and run-time techniques as shown in Figure 2.8[Agarwal06].

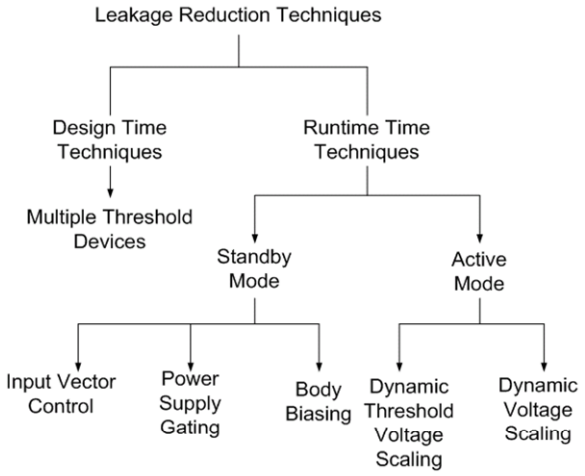

Figure 2.8: Classification of Leakage Reduction Techniques

2.4.1 Design-Time Techniques

Design-time techniques require modification of the fabrication process technology to achieve leakage reduction during the design step. Once implemented in the design, these techniques cannot be altered during the circuit operation.

Multiple Threshold Devices

Devices with higher threshold voltage exhibit leakage current that is one order of magnitude lesser than that of the devices with lower threshold voltages. Therefore, high-threshold devices can be used in noncritical paths to reduce leakage whereas transistors in

critical paths are assigned lower threshold to preserve performance. This technique does not require any additional circuitry for implementation. The high threshold devices are created by varying doping profile, increasing oxide thickness and using longer transistor lengths.

2.4.2 Run-Time Techniques

Run-time leakage reduction techniques are based on circuit-level optimization methods, which can be dynamically adjusted during circuit operation. Depending on the mode of operation of the circuit, these techniques can be classified into two categories. Standby mode techniques place the entire system, or certain modules of the circuitry into a standby state when computational activity is not required. During normal operation, dynamic power constitutes the major portion of the total power consumption. However, the circuit temperature increases during active mode due to transistors' switching activity. The subthreshold leakage increases exponentially with temperature and dominates the total leakage power. Active mode techniques focus on reducing leakage by dynamically changing the threshold voltage of transistors. When peak performance is not needed, these techniques slow down the circuit by lowering the supply voltage, which also helps in minimizing dynamic power dissipation.

Input Vector Control

In stacked transistors configuration, the voltages at the internal nodes depend on the input vectors applied to the stack. If transistors in stack are off, then the internal node voltages cause self-reverse biasing effect as explained in section 2.1.2. The topmost off transistor in stack will have the highest threshold voltage and will determine the subthreshold leakage. Hence, maximum number of stack transistors should be turned off to reduce the subthreshold leakage. By selecting the proper input vectors, leakage current can be substantially reduced in the standby mode.

Power Supply Gating

This scheme uses large transistors, called sleep transistors, in series with the pull-up and pull-down stacks to cut off the power supply rail from the circuit when the circuit is in standby mode. During normal operation, sleep transistors are on and must present

minimum resistance to maintain gate drive current and delay. The size of the sleep transistor should be very large for low on-resistance. Higher threshold voltage is also desirable for better leakage reduction.

Body Biasing

The threshold voltage of a transistor can be given by[Rabaey04]

$$V_{TH} = V_{THO} + \gamma \left(\sqrt{|\,(-2)\,\phi_F + V_{SB}|} - \sqrt{|2\phi_F|} \right) \tag{2.6}$$

where γ is called body-effect parameter that signifies the effect of changes in V_{SB} on threshold voltage. ϕ_F is Fermi potential and V_{THO} is the threshold voltage when $V_{SB} = 0V$. For NMOS, $V_{SB} < 0V$ (reverse body bias) increases the threshold voltage whereas positive V_{SB} (forward body bias) results in lower threshold. Reverse body biasing technique changes the body voltage, making it higher than V_{DD} for PMOS devices and lower than ground for NMOS devices, to increase the threshold voltage during standby mode. This results in significant reduction in the subthreshold leakage. Forward body bias can be used to improve circuit speed during normal operation. Forward bias voltage reduces the threshold voltage, which in turn results in higher on-state current.

Implementing body biasing poses certain difficulties. To control the body bias, devices need to be implemented in triple-well process. Generation of body bias voltages increases design complexity. Applying body bias needs routing grid, which increases layout area and hence cost of the chip.

Dynamic Threshold Voltage Scaling (DVTS)

The threshold voltage of a transistor can be changed dynamically by controlling the body bias voltage. DVTS technique adaptively changes the threshold voltage of the circuit depending on the performance requirements. When high speed and low latency are desired, threshold voltage is reduced using forward body biasing. For lower workload, this scheme slows down the circuit by increasing the threshold voltage through reverse body biasing. Hence, this technique allows the circuit to deliver optimal performance and to lower the leakage power during active mode by adaptively adjusting the threshold voltage.

To track the speed requirements and to adjust the threshold voltage accordingly, DVTS uses a feedback control loop. Replica of the critical path is used to measure the actual frequency of operation. This frequency is then compared with the desired reference frequency to generate an error signal. Based on the error signal, DVTS controller adjusts the body bias to change the threshold voltages.

Dynamic Voltage Scaling (DVS)

During active mode of operation, dynamic power constitutes major part of total power consumption. Dynamic power consumption in synchronous digital circuits strongly depends on supply voltage and operating frequency. Reducing the frequency alone reduces the power, but it does not save energy per operation. Reducing the supply voltage along with frequency results in substantial energy and power savings. This technique of adjusting the supply voltage and frequency as per the performance requirements is called as dynamic voltage scaling (DVS).

DVS technique exploits the fact that the computational activities in digital circuits, especially battery-powered devices, do not require maximum performance continuously. Such activities are often sparse, separated by long standby times. DVS relies on a feedback control loop along with voltage regulator to adjust the supply voltage to the optimal value necessary for desired operating frequency.

From the discussion of this chapter it has become clear that leakage power forms a significant portion of total power dissipation in the scaled technologies. Leakage during standby mode lowers the battery life of portable devices. Circuit techniques to reduce leakage are becoming increasingly important in design of low power digital circuits. During active mode, dynamic power consumption can be minimized by operating the circuit at lower supply voltage when peak performance is not required. In this thesis, focus is on the use of body biasing for standby leakage reduction and the supply voltage scaling for dynamic power savings. Both these techniques are discussed in the remaining chapters.

Chapter 3

Body Biasing

The increasing leakage currents of deep submicron CMOS devices is a major concern in designing high performance, low standby power circuits. As CMOS technology continues to scale, supply voltage is lowered to reduce the dynamic power consumption. In order to maintain the current drive capability and speed improvement, the threshold voltage and gate oxide thickness must also be reduced. This leads to an exponential increase in leakage currents as explained in previous chapter. In today's scaled technologies, leakage power constitutes almost 20% of the total power consumption. As the channel length continues to scale, leakage power is expected to become comparable to dynamic power in the nanometer regime[Martin02].

Reverse body biasing (RBB) can be used to dynamically raise the threshold voltage during standby mode, thereby reducing the leakage power. [Keshavarzi99] shows that RBB reduces leakage power by three orders of magnitude in 0.35μm technology. However, RBB is becoming less effective for recent nanoscale technologies due to increase in short channel effects (SCEs). RBB raises the drain-bulk voltage, which increases the drain-induced barrier lowering (DIBL). DIBL in turn lowers the threshold voltage. In heavily doped substrates, rise in the drain/source-bulk voltages widens the depletion regions around drain-bulk and source-bulk junctions, leading to higher BTBT leakage. This increase in tunneling currents can nullify any leakage reduction benefits from RBB. Therefore, new device engineering techniques are required to control the junction leakage and to maintain the effectiveness of RBB.

While RBB can reduce the standby leakage, forward body biasing (FBB) can improve performance of digital circuit in active mode. FBB lowers the threshold voltage and increases the on-state current drive. FBB can be combined with supply voltage scaling in active mode to reduce dynamic power consumption without performance degradation. However, FBB results in an increased gate capacitance and the junction capacitances. This reduces the speed improvement achieved through reducing the threshold voltage.

Body biasing techniques modulate the threshold voltage of the transistor by controlling the substrate bias. The concept of threshold voltage is briefly explained in this chapter. As explained in chapter 2, the subthreshold leakage has an exponential dependence on threshold voltage. Naturally the factors affecting the threshold voltage influence the subthreshold leakage as well. In sub-micron CMOS technologies, many secondary effects impact the threshold voltage. Some of the most important effects are discussed in this chapter. For maintaining the efficiency of body biasing, it is important to take these secondary effects into consideration.

3.1 The Threshold Voltage

Figure 3.1 shows a cross-section of a long channel NMOS with source, drain and bulk terminals grounded. As the gate voltage is increased from 0V, depletion region is created below the gate. As the gate voltage is increased further, a condition of strong inversion is reached wherein the silicon surface inverts from p-type material to n-type. This phenomenon of strong inversion occurs at a critical value of gate-source voltage, which is termed as the threshold voltage, V_{TH}[Zhang05].

The threshold voltage is a function of following three voltage components:
- the difference in work function between gate and substrate (ϕ_{MS})
- the fixed oxide charge present at the $Si - SiO_2$ interface ($-Q_{OX}/C_{OX}$)
- the gate voltage required to bring the surface potential to the strong inversion condition ($2\phi_F$) and to offset the induced depletion region charge ($-Q_B/C_{OX}$). ϕ_F is called Fermi potential. Typical value of ϕ_F for p-type silicon substrate is -0.3V. C_{OX} is gate oxide capacitance per unit area.

Putting the above three components together, the threshold voltage under no body bias condition can be given by[Zhang05]

$$V_{TH} = V_{THO} = \phi_{MS} - \frac{Q_{OX}}{C_{OX}} + 2\phi_F - \frac{Q_B}{C_{OX}} \tag{3.1}$$

With no body bias ($V_{SB} = 0V$), the charge stored in depletion region under the strong inversion condition can be expressed as

$$Q_B = Q_{BO} = -\sqrt{2qN_A\epsilon_{si}|2\phi_F|} \tag{3.2}$$

where q is the electron charge, N_A is doping concentration and ϵ_{si} is the permittivity of silicon. A body effect coefficient γ is defined as $\gamma = \sqrt{2qN_A\epsilon_{si}}/C_{OX}$. Thus equation 3.1 can be simplified using equation 3.2 and γ as

$$V_{THO} = \phi_{MS} - \frac{Q_{OX}}{C_{OX}} + 2\phi_F + \gamma\sqrt{2\phi_F} \tag{3.3}$$

3.1.1 Body Effect

Under body biasing condition ($V_{SB} \neq 0V$), the surface potential required for strong inversion increases from $|2\phi_F|$ to $|2\phi_F + V_{SB}|$ and the charge stored in the depletion region is given by[Zhang05]

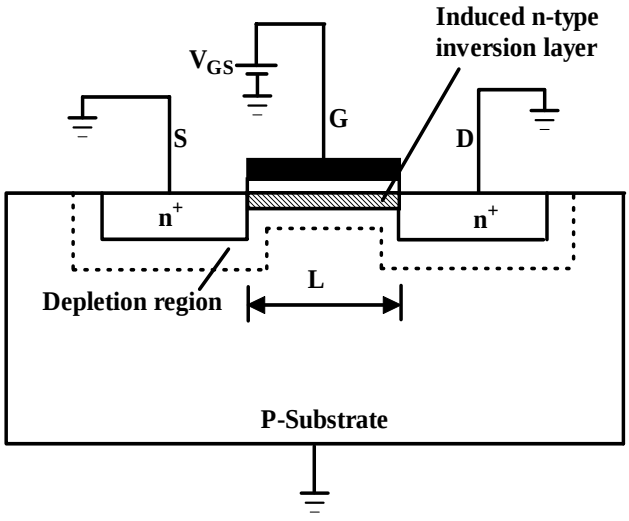

Figure 3.1: Cross-sectional view of NMOS showing depletion region and inversion layer formed below the gate

$$Q_B = -\sqrt{2qN_A\epsilon_{si}|2\phi_F + V_{SB}|} \tag{3.4}$$

The threshold voltage under different body biasing conditions can then be written as follows

$$V_{TH} = V_{THO} + \gamma\left(\sqrt{|2\phi_F + V_{SB}|} - \sqrt{|2\phi_F|}\right) \tag{3.5}$$

From the above equation, it is clear that for NMOS transistor, applying reverse bias voltage ($V_{SB} > 0V$) to substrate-source junction of a transistor widens the source-junction depletion region and leads to an increase in the threshold voltage. Forward body bias ($V_{SB} < 0V$) reduces the depletion region and lowers the threshold voltage. This change in threshold voltage from its nominal value V_{THO} due to non-zero source-substrate voltage is called as *body effect* and it can be given by

$$\Delta V_{TH} = V_{TH} - V_{THO} = \gamma\left(\sqrt{|2\phi_F + V_{SB}|} - \sqrt{|2\phi_F|}\right) \tag{3.6}$$

where $\gamma = \sqrt{2qN_A\epsilon_{si}}/C_{OX}$ and $C_{OX} = \epsilon_{OX}/T_{OX}$. Thus, body effect becomes less pronounced for thinner oxides, higher gate dielectric constants and lightly doped substrates. As a result, body biasing becomes less effective in every new technology generation. The efficiency of body biasing further decreases in scaled devices due to short channel effects, which are discussed in the following section.

3.2 Short Channel Effects

Equation (3.5) shows that the threshold voltage is a function of the manufacturing process and the body bias voltage V_{SB}. Threshold voltage is therefore, independent of the channel length and the drain bias. This holds true for long-channel devices, where drain and source depletion regions have no impact on the surface potentials or field patterns.

In short-channel devices, source and drain depletion regions penetrate more into the channel and deplete more region under the inversion layer. These deeper depletion regions make the channel more attractive for electrons. This is equivalent to lowering of

the surface potential. Hence, lesser gate voltage is sufficient to cause strong inversion and device can conduct more current than that predicted by the long-channel device equations. This phenomenon is known as short-channel effect[Rabaey04]. As a result of short-channel effect, the threshold voltage of a short-channel transistor depends on channel length and drain-source voltage. Variation in the threshold voltage due to these parameters is explained in the following subsections.

3.2.1 V_{TH} Roll-Off

In the derivation of threshold voltage equation, it is assumed that all the depletion region charge (Q_B) is induced by the applied gate voltage. This assumption is not true in short-channel devices, in which widths of source and drain depletion regions become comparable to the effective channel length. Fraction of the total depletion charge (ΔQ_B) is supplied by the drain and source depletion regions. As the channel length continues to shrink, ΔQ_B increases. Therefore, a smaller gate voltage is enough to offset the remaining depletion charge and reach strong inversion condition. As a result, threshold voltage of a short-channel device decreases with decrease in gate length. The reduction in the threshold voltage with decreasing device length in short-channel devices is called *V_{TH} roll-off*.

3.2.2 Drain Induced Barrier Lowering (DIBL)

For a fixed source-bulk voltage V_{SB}, raising the drain-source voltage V_{DS} increases the width of the depletion region around the drain-substrate junction. Hence, drain depletion region contributes more depletion charge (ΔQ_B) in short-channel devices. As a result, smaller threshold voltage is sufficient to reach the strong inversion condition[Rabaey04]. This reduction in threshold voltage with increasing drain bias is known as drain induced barrier lowering (DIBL) effect. DIBL effect increases the subthreshold leakage. It is a major problem in dynamic memory circuits, where it appears as a data-dependent noise source.

From the above discussion it is clear that the short-channel effects can be minimized by reducing the width of source and drain depletion regions. In deep submicron devices, this is achieved by inserting highly doped structures, called 'halo' implants, near the source-substrate and drain-substrate junctions.

3.3 Effect of Body Biasing on Leakage Components

In order to reduce leakage during standby mode using body biasing, it is important to understand the impact of body bias on major leakage components. The primary leakage mechanisms have been discussed in the previous chapter. In this section the variations of these leakage components with body bias is discussed.

Taking into account the body effect and DIBL effect, SPICE BSIM4 model for the subthreshold leakage current is as follows[Li05]

$$I_{SUB} = Aexp\left(\frac{V_{GS} - V_{THO} - \gamma\prime V_{SB} + \eta V_{DS}}{nV_T}\right)\left(1 - exp\left(\frac{V_{DS}}{V_T}\right)\right) \tag{3.7}$$

where $A = \mu_0 C_{ox}\frac{W}{L_{eff}}V_T^2 e^{1.8}$, V_T is thermal voltage, $\gamma\prime$ is the linearized body effect coefficient, η is DIBL coefficient and L_{eff} is effective channel length.

Under reverse bias conditions ($V_{SB} > 0V$), $\gamma\prime V_{SB}$ is positive and the subthreshold leakage decreases exponentially with increasing reverse bias. Hence, application of reverse body biasing in standby mode can significantly reduce the subthreshold leakage. On the other hand, forward bias ($V_{SB} < 0V$) makes the $\gamma\prime V_{SB}$ term negative thereby increasing the subthreshold current.

As mentioned in chapter 2, the gate leakage current density can be given by[Agarwal05]

$$J_{gate} = A \cdot \left(\frac{V_{OX}}{T_{OX}}\right)^2 \cdot exp\left[-B\frac{1-(1-V_{OX}/\phi_B)^{3/2}}{V_{OX}/T_{OX}}\right]$$

where $A = q^3/16\pi^2 h\phi_B$, $B = 4\sqrt{2m_{ox}}\phi_B^{3/2}/3hq$, ϕ_B is the tunneling barrier height in eV and m_{ox} is effective carrier mass in oxide.

Gate leakage has an exponential dependence on the thickness of the gate oxide (T_{OX}) and the voltage across it (V_{OX}). Both these parameters do not directly depend on substrate bias. The body bias only affects the gate-to-substrate component (I_{gb}) of total gate leakage. However, this component is much smaller in magnitude compared to gate-to-channel (I_{gc}) or gate-to-source/drain overlap (I_{gso}/I_{gdo}) components. Hence, gate leakage is almost insensitive to body biasing. However, gate leakage is becoming increasingly important in nanoscale devices due to reduced oxide thickness and needs to be controlled using technology alternatives such as high-κ dielectrics[Neau03].

BTBT leakage current density is given by[Neau03]

$$J_{BTBT} = A\frac{EV_{RB}}{E_g^{1/2}} exp\left(-B\frac{E_g^{3/2}}{E}\right)$$

where V_{RB} is the reverse bias voltage applied across the junction. The electric field E across the junction is given by[Neau03]

$$E = \sqrt{\frac{2qN_AN_D\left(V_{RB}+\Psi_B\right)}{\epsilon_{si}\left(N_A+N_D\right)}}$$

With increase in V_{RB} (reverse bias), BTBT leakage increases since the electric field E depends on reverse bias. BTBT leakage also depends on doping profiles. Strong halo implants located close to to the source and drain region cause significant rise in BTBT leakage with reverse bias.

Figure 3.2 shows the effect of body biasing on different leakage components at room temperature for a NMOS transistor in 90-nm technology. It can be observed that subthreshold leakage and BTBT leakage have exponential but opposite dependence on body bias.

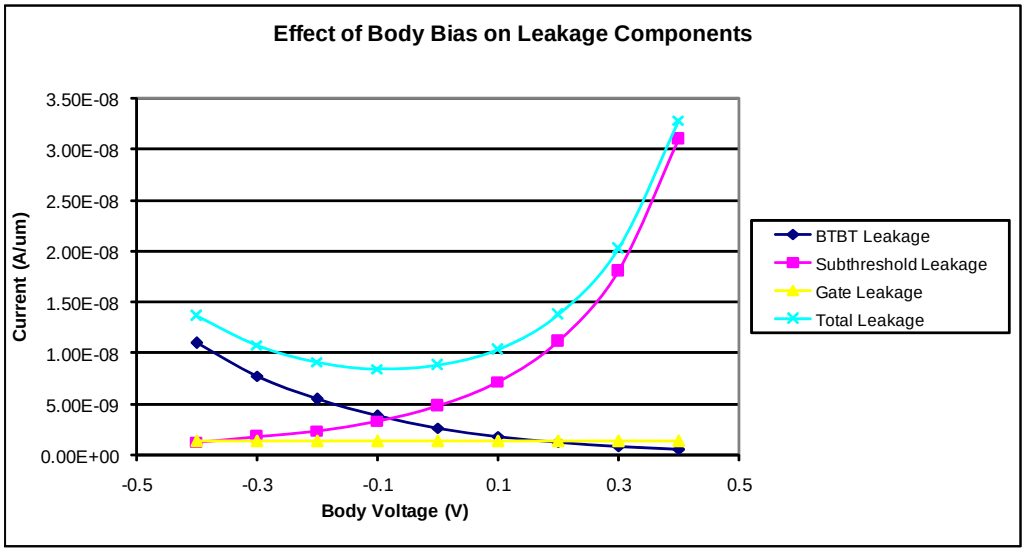

Figure 3.2: Variations in leakage components of a NMOS transistor with body bias in 90-nm technology

3.4 Optimum Reverse Body Bias for Minimum Standby Leakage

Figure 3.2 shows that there exists an optimum reverse body bias voltage for which the total standby leakage is minimized. The subthreshold leakage and BTBT leakage have opposite dependence on body bias. Gate leakage is almost insensitive to body bias. Hence the standby leakage is minimized at the reverse bias voltage for which the sum of BTBT leakage and subthreshold leakage is minimum. Since both types of leakage are exponentially dependent on body bias, the optimum reverse bias occurs when the subthreshold current and BTBT current are approximately equal[Keshavarzi99].

In any process technology, it is difficult to correctly determine the ratio of I_{SUB} and I_{BTBT}, due to process variations and complexity in calculating the electric field across the junction. In addition, the ratio varies with each technology and doping profile[Neau03]. As a result, the location of the optimum body bias is highly dependent on technology, doping profiles and temperature. Hence in order to find the optimum reverse body bias, it is necessary to develop a leakage monitor circuit, which can compare the subthreshold leakage and the BTBT leakage.

3.4.1 Previously Designed Leakage Comparator Circuits

Circuits for comparing I_{SUB} and I_{BTBT} have been designed in [Neau03] and [Nomura06]. The schematic of current mirror based comparator from [Neau03] is shown in Figure 3.3. The circuit detects body bias where the BTBT leakage is half of the total leakage. It mirrors half the drain current of left side NMOS N_1 into the right side using PMOS current mirror at the top. The drain current of transistor N_1 mainly consists of the subthreshold leakage (I_{SUB}), BTBT leakage (I_{BTBT}) and the drain-to-gate leakage component of gate leakage (I_{DG}). The subthreshold leakage through N_2 is significantly lower due to the *stack effect*. The N_2 leakage therefore, primarily consists of I_{BTBT} and I_{DG}. Assuming the gate leakage to be negligible and that the subthreshold leakage of N_1 and N_2 are identical, the voltage at nodes A and B, V_A and V_B, would depend on the BTBT currents through transistors N_1 and N_2 respectively. The two voltages would be equal when the N_2 leakage is half of the N_1 leakage.

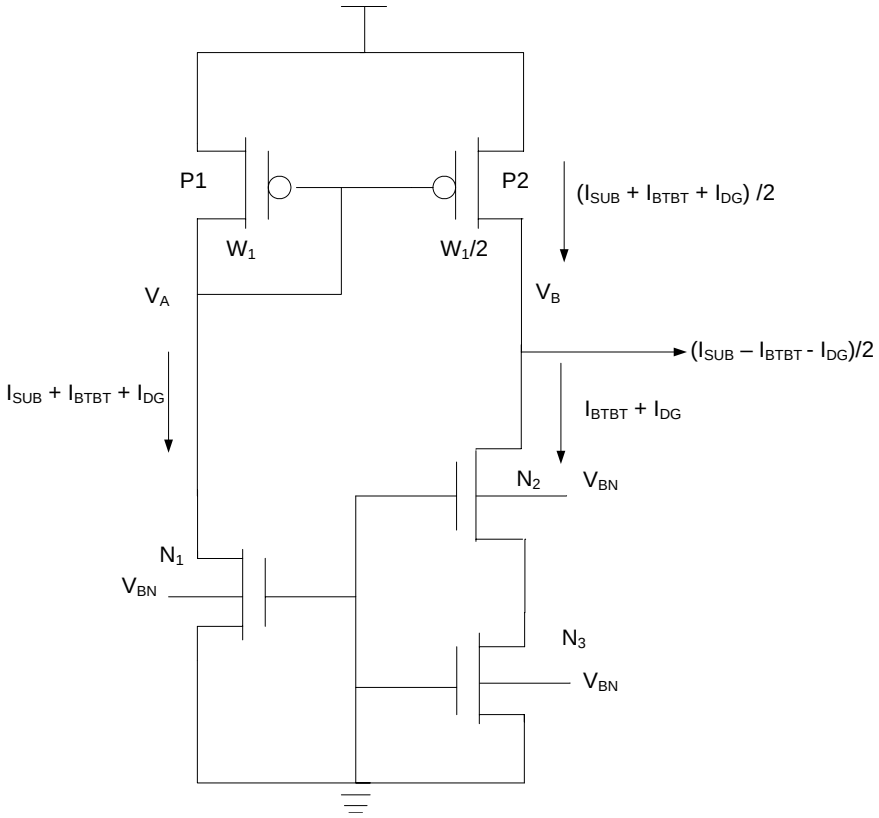

Figure 3.3: Leakage Monitor Circuit from [Neau03]

The circuit has a few problems that affect the accurate comparison of I_{SUB} and I_{BTBT}. Gate leakage has been ignored in the calculation. Considering gate leakage introduces error in the output current as shown in Figure 3.3. As it is clear from the discussion of previous chapter, gate leakage is becoming increasingly important in nanometer technologies and hence should be subtracted from the current mirrored into the right side of the circuit.

In short-channel devices, the accuracy of the current mirror depends on the drain-to-source voltages of PMOS devices due to channel length modulation. In this circuit, there is no provision to keep the voltages of V_A and V_B constant. Any drop in the drain voltage would result in an error in mirroring the total leakage current. To avoid this error, the drain voltages of current mirror transistors should be held constant. Finally, the circuit does not take into account gate-to-bulk leakage component (I_{GB}) while calculating I_{BTBT}.

[Nomura06] has proposed a better leakage comparator circuit, which takes into consideration I_{GB} and I_{DG}. The circuit , shown in Figure 3.4, compares the leakage current at the source terminal of N_1 with the leakage current at the drain terminal of N_4. NMOS and PMOS current mirrors are used to mirror leakage currents. Two operational amplifiers are used to maintain the drain voltages of current mirror transistors constant. The circuit compares the subthreshold leakage with BTBT leakage, which includes I_{GB}.

This leakage monitor improves accuracy by considering the gate leakage of NMOS transistors. However, it ignores the input leakage currents of amplifiers used to keep the drain voltages constant. Input leakage of amplifier reduces the leakage current on left side to $I_{SUB} + I_{DG} - I_{GB} - I_{IL}$, where I_{IL} is the input leakage of amplifier. On right side, the input leakage is added to the NMOS leakage ($I_{GIDL} + I_{DG}$). As a result, the comparator output current has an error of $2I_{IL}$.

Secondly, the two components of BTBT leakage, I_{GIDL} and I_{GB}, are obtained from two different NMOS transistors. The final value of I_{BTBT} is the summation of these two components mirrored through NMOS and PMOS current mirrors. This leads to an erroneous value of I_{BTBT} due to mismatches in individual transistors and the current mirrors. To avoid this error, it would be better if total BTBT leakage can be calculated from a single transistor.

3.4.2 Proposed Leakage Comparator Circuit

A new leakage comparator is proposed to improve the accuracy in estimating the difference between the subthreshold leakage and the BTBT leakage. The schematic of the proposed leakage comparator is shown in Figure 3.5. BTBT leakage is measured using a single transistor on left side with drain, source and gate connected together. This arrangement combines I_{GIDL} and I_{GB} to generate I_{BTBT}. The currents in different parts of the circuit are marked in the schematic.

As mentioned in chapter 2, BTBT leakage depends on the voltage across the junction (V_{RB}) and gate leakage depends on the voltage across the oxide (V_{OX}). Hence to find out the maximum BTBT and gate leakage, both V_{RB} and V_{OX} should be maintained at V_{DD}, the maximum voltage difference across the junction and the oxide. This is achieved using an amplifier, shown in Figure 3.6, with one input connected to V_{DD}. The input leakage of the two amplifiers cancel each other out, thus reducing the error in output current ($I_{SUB} - I_{BTBT}$). Figure 3.7 shows the results of this comparator for increasing reverse bias voltage. The percentage error between the simulated and the calculated leakage currents

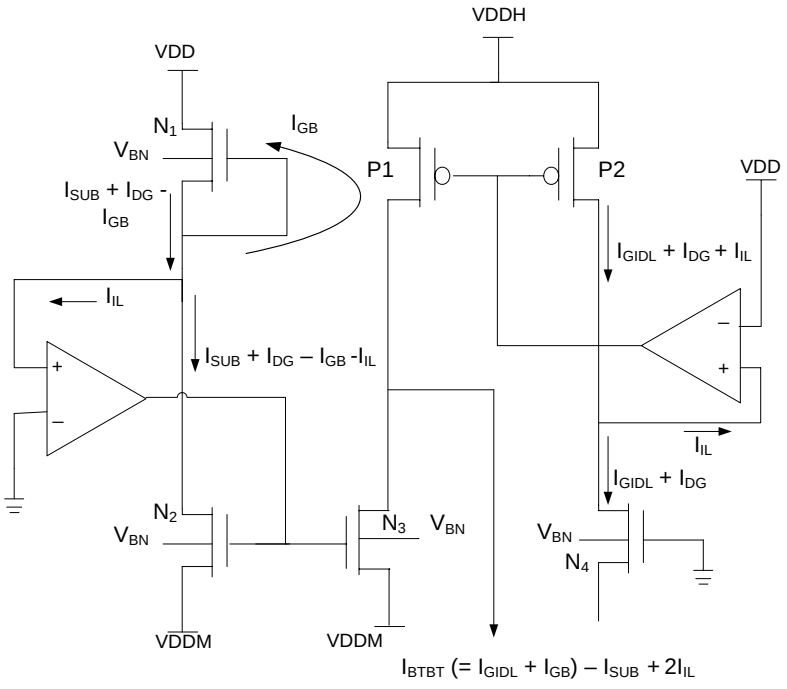

Figure 3.4: Leakage Monitor Circuit from [Nomura06]

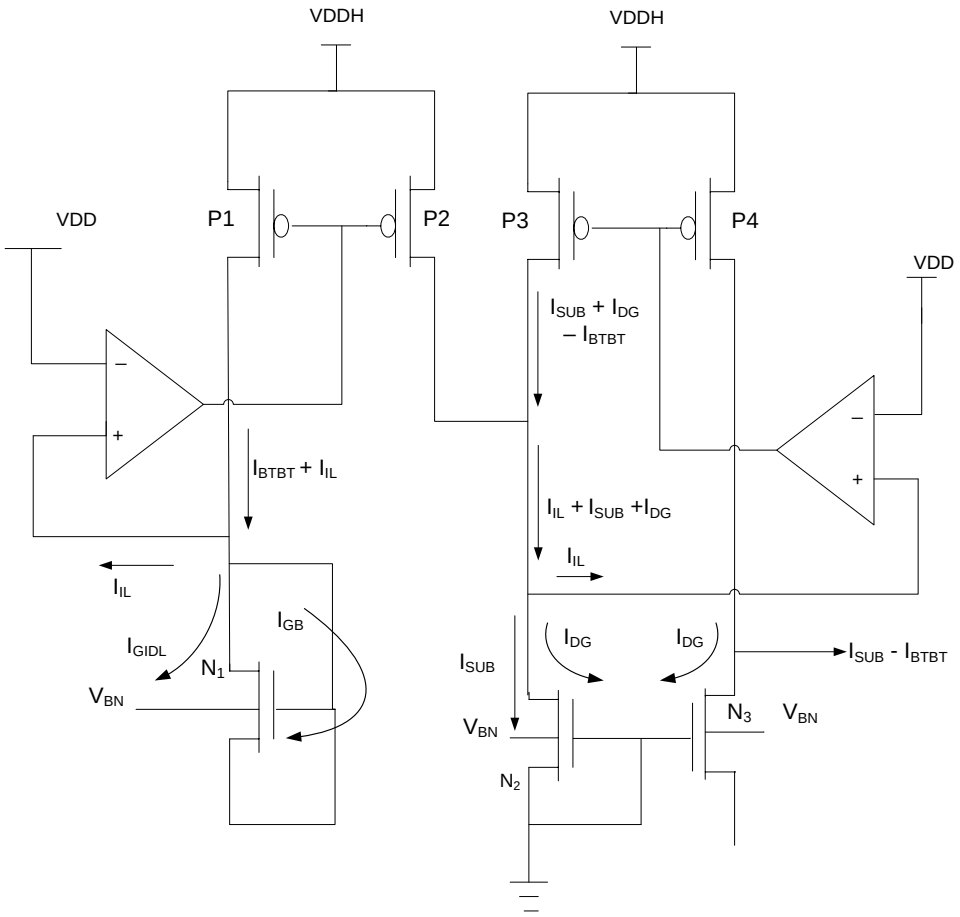

Figure 3.5: Proposed Leakage Monitor Circuit

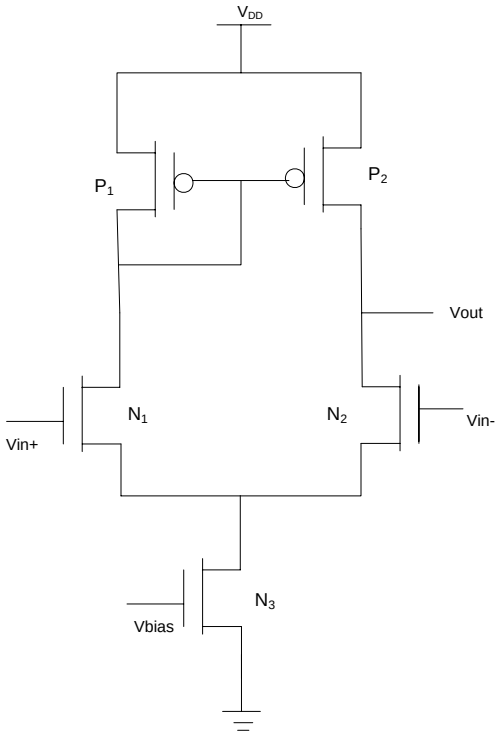

Figure 3.6: Amplifier used in Leakage Monitor Circuit

is always below $\pm 5\%$ for the range $-1V$ to $0V$ of V_{BS}. The results indicate that the point of minimum leakage ($I_{SUB} = I_{BTBT}$) occurs at $V_{BS} = -0.88V$. Table 3.1 compares the performance of the designed leakage monitor with the previously discussed leakage monitor circuits.

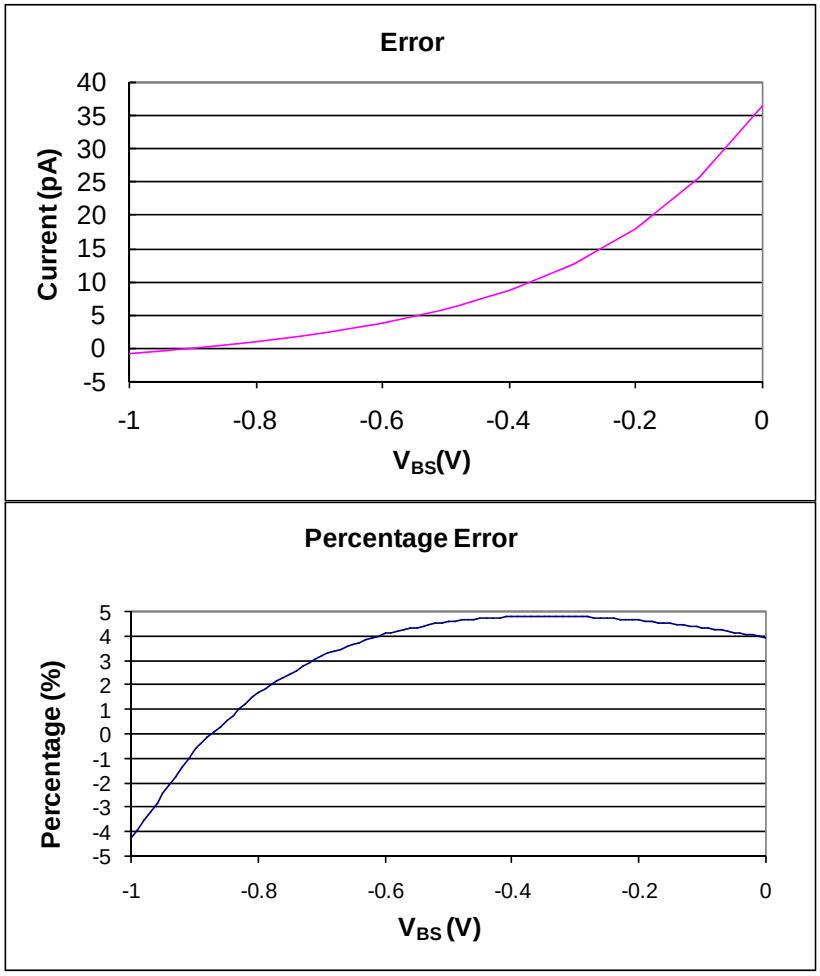

Figure 3.7: Leakage Monitor Results

3.5 Standby Leakage Reduction using Reverse Body Bias

90-nm process technology provides devices with nominal threshold voltage,(reg-V_{TH}), for general purpose and devices with lower threshold, low-V_{TH}, which can be used

Table 3.1: Performance of the leakage monitor circuits

Parameters	[Neau03]	[Nomura06]	This work
Range of V_{BS}	$-0.2V$ to $0.2V$	$-1.5V$ to $0V$	$-1V$ to $0V$
Percentage Error	35%	$< 20\%$	$< \pm 5\%$
Supply Voltage Range	$0.6V$ to $0.8V$	$0.4V$ to $1V$	$1V$ to $1.2V$
Device Technology Used	50-nm, 70-nm	90-nm	90-nm
Gate Leakage	ignored	considered	considered
BTBT Leakage Measured	only I_{GIDL}	both I_{GIDL} and I_{GB}	both I_{GIDL} and I_{GB}
Amplifier Input Leakage	-	ignored	considered

Table 3.2: Devices Used for Body Biasing

Parameters	Reg-V_{TH}	Low-V_{TH}
Gate Length	80nm	80nm
Oxide Thickness	2.2nm	2.2nm
V_{TH} (NMOS/PMOS)	(370mV/315mV)	(250mV/190mV)
Body Effect Coefficient (NMOS/PMOS))	(114mV/100mV)	(90mV/50mV)
Nominal Supply Voltage	1V	1V

for low power, high speed applications. Table 3.2 lists the important device parameters for both types of devices.

For a single NMOS or PMOS transistor, the off-state leakage is primarily governed by the subthreshold leakage. The low-V_{TH} devices exhibit considerably higher subthreshold leakage than the reg-V_{TH} devices. Due to its exponential dependence on reverse bias voltage, the off-state leakage decreases rapidly with increasing reverse bias voltage as shown in Figure 3.8 and 3.9.

In static CMOS circuits, on-state gate leakage is comparable to the total off-state leakage [Arnim05]. Hence, both on and off states of the circuit should be considered while calculating the total leakage of a CMOS circuit. Figure 3.10 shows the effect of reverse body bias on the total leakage of an inverter pair. Although, reverse body bias significantly reduces the total leakage of low-V_{TH} inverter pair, it is always higher than the reg-V_{TH} counterpart. The leakage of low-V_{TH} inverter pair is twice as much that of reg-V_{TH} even at the most negative body bias.

To check the efficiency of reverse body bias in leakage reduction, an optimum reverse body bias voltage, detected using the leakage monitor, is applied to several static circuits. Single devices, inverter pairs and static gates are considered to study leakage

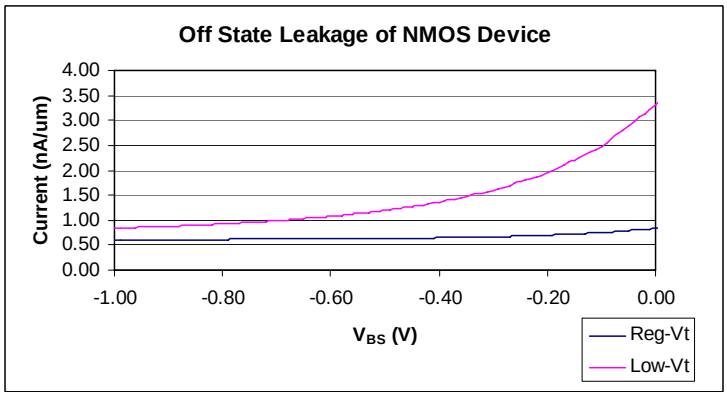

Figure 3.8: NMOS Off State Leakage Current

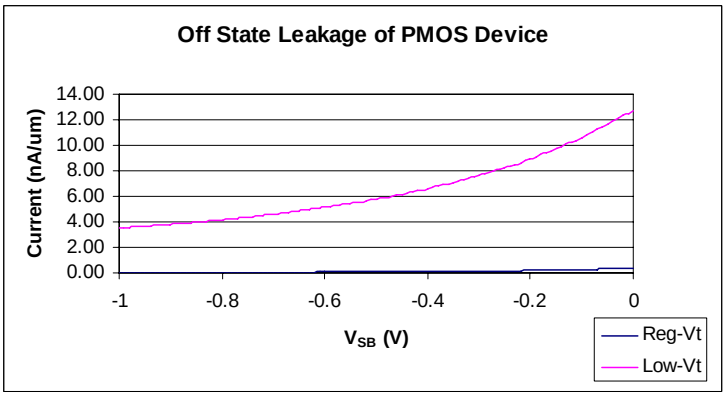

Figure 3.9: PMOS Off State Leakage Current

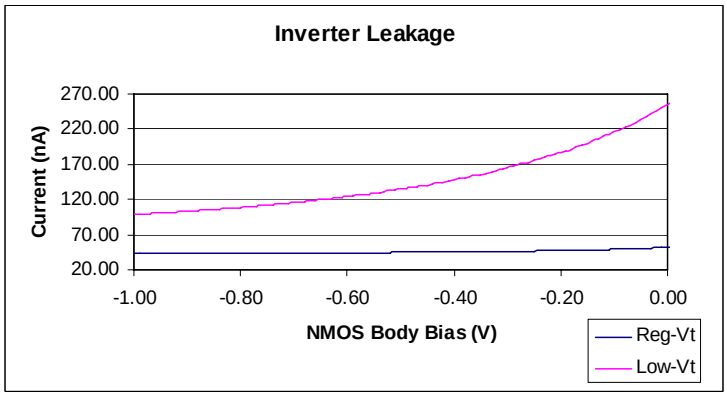

Figure 3.10: Total Leakage Current of an Inverter Pair

Table 3.3: Leakage Reduction Using Reverse Body Bias for Reg-Vt Devices

Circuit	V_{BS} = 0V	V_{BS} = -0.88V	Percentage Reduction
NMOS	1.13nA	46.67pA	95.87%
PMOS	5.482nA	460pA	91.6%
Inverter Pair	8.27nA	614.8pA	92.56%
2-input NAND	27.26nA	25.42nA	6.75%
2-input NOR	26.84nA	20.56nA	30.54%

Table 3.4: Leakage Reduction Using Reverse Body Bias for Low-Vt Devices

Circuit	V_{BS} = 0V	V_{BS} = -0.88V	Percentage Reduction
NMOS	13.16nA	1nA	92.4%
PMOS	190.9nA	57.88nA	69.68%
Inverter Pair	205.8nA	55.64nA	72.96%
2-input NAND	45nA	25.51nA	43.31%
2-input NOR	208.1nA	78.4nA	62.32%

reduction. Tables 3.3 and 3.4 list the reduction in leakage current due to applied optimum reverse bias. For NAND and NOR gates, the worst case standby leakage is considered depending on the input sequences.

The results indicate that the reverse body bias is less effective for the low-V_{TH} devices than the reg-V_{TH} devices, mainly due to the lower body effect coefficient. In case of PMOS transistors, the body effect parameter for low-V_{TH} device is 44.44% lower than the reg-V_{TH} counterpart (See Table 3.2). As a result, the leakage reduction achieved using reverse bias is considerably lower for low-V_{TH} transistor. To maintain the efficiency of body biasing, higher reverse bias voltage would be required for low-V_{TH} devices. However, at higher reverse bias voltages, the effectiveness of this technique decreases due to square root dependence of threshold voltage on body bias[Arnim05].

3.5.1 Leakage Reduction in Stacked Transistors Configurations

Table 3.3 and 3.4 show that the efficiency of reverse body bias decreases with transistor stacking. For NAND and NOR gates the worst case standby leakage occurs when all the transistors in the stack are conducting except the topmost transistor (See Table 2.1 for the standby leakage in a 2-input NAND gate). As discussed in chapter 2 (section 2.2.2), the conducting transistors in the stack exhibit large gate leakage currents because the

internal node voltages are close to ground potential. The total standby leakage is given by the sum of the dominant gate leakage and the subthreshold leakage of the nonconducting topmost transistor. Since gate leakage is independent of the reverse bias, the reduction in standby leakage is much less in the stacked transistors configuration. As mentioned in chapter 2, gate leakage of PMOS device is an order of magnitude lesser than that of NMOS device. As a result, percentage reduction in leakage is higher in a NOR gate than a NAND gate.

In case of low-V_{TH} circuits, the total leakage is dominated by the subthreshold leakage of the topmost transistor. Therefore, the percentage reduction in total leakage using reverse biasing is higher in low-V_{TH} NAND and NOR gates. From the results of Table 3.3 and 3.4, reverse biasing looks more effective in high speed circuits using smaller V_{TH} transistors[Arnim05]. In such circuits, reverse biasing reduces the dominant subthreshold leakage thereby decreasing the total standby leakage.

3.6 Dynamic Performance Improvement Using Forward Body Bias

Forward body bias in active mode increases the on-state current by reducing the threshold voltage. This results in higher speed and better dynamic performance. Figure 3.11 and 3.12 show the effect of body bias on the on-state current of NMOS and PMOS devices respectively.

5-stage ring oscillator and 16-bit ripple carry adder have been used as test structures to study the effect of forward body bias on speed. Table 3.5 and 3.6 show the speed improvement for a 5 stage ring oscillator and 16-bit ripple carry adder under 0.3V of forward body bias.

The results indicate that only slight performance improvement can be achieved using forward body bias. Therefore, body biasing is less effective in improving speed of the circuit in active mode. There are additional difficulties in implementing forward body bias. Latch-up has to be avoided over operating temperature and process variations. Leakage currents, especially the subthreshold leakage grows exponentially with the forward bias. Considering the design complexity involved in generating the body bias voltages, the forward

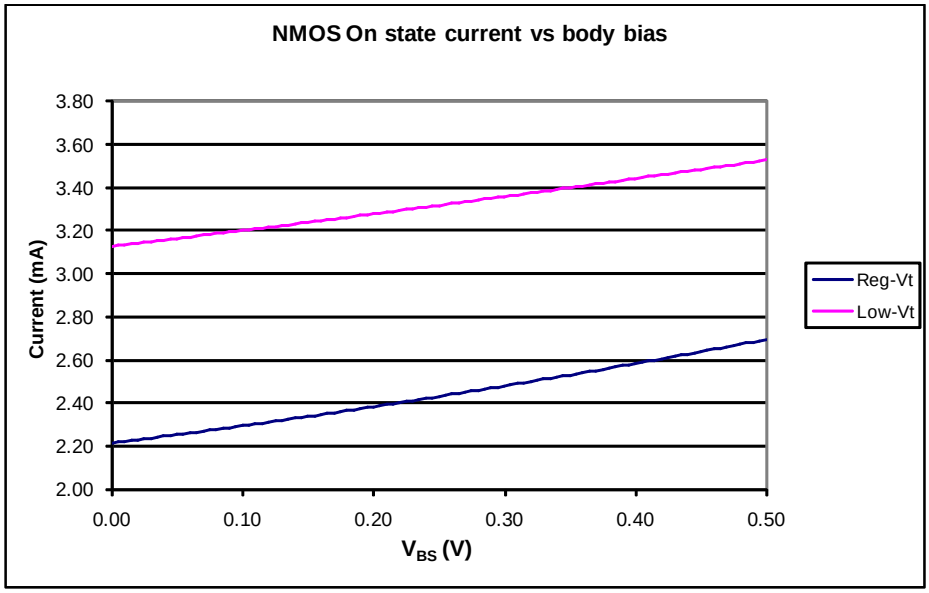

Figure 3.11: Increase in on-state current of NMOS device with forward body bias

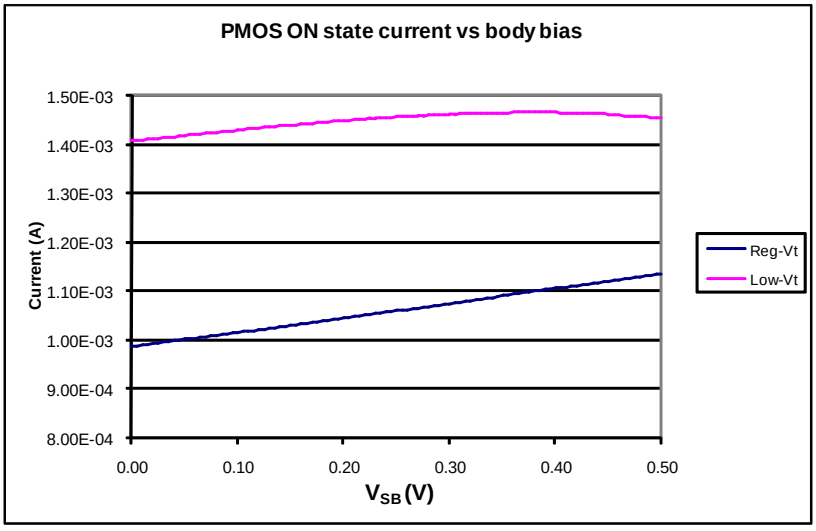

Figure 3.12: Increase in on-state current of PMOS with forward body bias

Table 3.5: Performance Improvement using Forward Body Bias for Reg-Vt Devices

Circuit	$V_{BS} = 0$V	$V_{BS} = 0.3$V	Percentage Improvement
NMOS	2.213mA	2.479mA	12.01%
PMOS	3mA	3.287mA	9.56%
Ripple Cary Adder	50ps	45.53ps	8.94%
Ring Oscillator	15.19MHz	16.59MHz	9.2%
16-bit Adder	1.428ns	1.316ns	7.84%

Table 3.6: Performance Improvement using Forward Body Bias for Low-Vt Devices

Circuit	$V_{BS} = 0$V	$V_{BS} = 0.3$V	Percentage Improvement
NMOS	3.128mA	3.358mA	7.35%
PMOS	4.372mA	4.56mA	4.3%
Ripple Cary Adder	34.5ps	32.9ps	4.63%
Ring Oscillator	33.39MHz	34.87MHz	4.43%
16-bit Adder	983.5ps	960.4ps	2.34%

body biasing techniques is less practical in large integrated circuits implemented in nanoscale devices. As a result, other circuit techniques should be used for improving the dynamic performance of the circuit. Supply voltage scaling, which can reduce the dynamic power consumption of the circuit without speed degradation is discussed in the following chapter.

Chapter 4

Supply Voltage Scaling

With the advent of portable, battery-powered electronic devices, it has become imperative to reduce the power consumption and extend the battery lifetime. During active mode of operation, the power dissipation is dominated by the dynamic power consumption, which is given by[Rabaey04]

$$P_D = \alpha C_L V_{DD}^2 f_{CLK} \tag{4.1}$$

where α is the switching activity factor, C_L is the output load capacitance, V_{DD} is the supply voltage and f_{CLK} is the operating frequency. The load capacitance and the switching activity can be reduced through logic and architectural optimizations. Lowering the clock frequency reduces the dynamic power but it does not save energy consumed in each switching event. Power-delay product (PDP) is a measure of the energy consumed per switching event. Assuming maximum switching activity ($\alpha = 1/2$) and ignoring the leakage and short-circuit power consumption, PDP can be given as

$$PDP = \frac{P_{total}}{f_{CLK}} = \frac{C_L V_{DD}^2}{2} \tag{4.2}$$

Lowering a supply voltage can result in significant energy savings, since PDP shows quadratic dependence on supply voltage. However, the delay of a circuit increases drastically with

V_{DD} reduction as seen in the following expression[Chandrakasan92]

$$T_d = \frac{C_L V_{DD}}{I} = \frac{C_L V_{DD}}{\mu C_{OX}(W/L)(V_{DD} - V_{TH})^2} \tag{4.3}$$

Although, circuit delays increase with reduced supply voltage, this performance penalty is still acceptable considering the fact that the peak performance is not always required. This is particularly true for the portable devices, in which the computational workload varies from time to time. There are very few intensive activities which require peak performance and hence the maximum supply voltage. For a large portion of time portable devices are idle, waiting for an external event to occur. Keeping the supply voltage fixed in such cases, obviously results in an excessive power dissipation.

A synchronous buck converter is an appropriate choice for implementation of supply voltage scaling scheme due to their high conversion efficiency[Wei99]. The buck converter provides the regulated supply voltage from an external power supply. The duty ratio D of the converter is controlled by monitoring the circuit delay such that the circuit meets the speed requirement while operating at minimum possible supply voltage. The controller generates an output signal to drive the buck converter based on the difference between the desired frequency of operation and the actual circuit frequency.

Digital implementation of the control loop has become popular due to its compatibility with the digital CMOS processes[Wei99][Trescases04]. Digital controller requires a high resolution, high speed analog to digital converter (ADC) for converting regulated output voltage into digital signal. Such ADC consumes a significant amount of power, occupies large chip area and limits the dynamic response of the converter. Analog implementation of the controller can minimize the overall power consumption and improve the conversion efficiency. In this chapter, charge pump based analog controller is implemented that provides the control signal for the buck converter. A ring oscillator is used to model the critical path of the circuit. The pulse width modulation (PWM) technique for achieving higher conversion efficiency and better control has also been discussed. The operation of buck converter is briefly explained in the following section.

4.1 Buck Converter

Figure 4.1 shows the schematic of a buck converter. For simplicity, the control loop and the gate drivers are not shown. The buck converter consists of a pair of switches, an external inductor and an output capacitor. PMOS transistor switch, also known as power switch, is used to control the flow of power from an external power supply to the converter output. The transistor should have fast switching times and minimum voltage drop across it. In addition it should be able to withstand high voltage spikes produced by the external inductor.

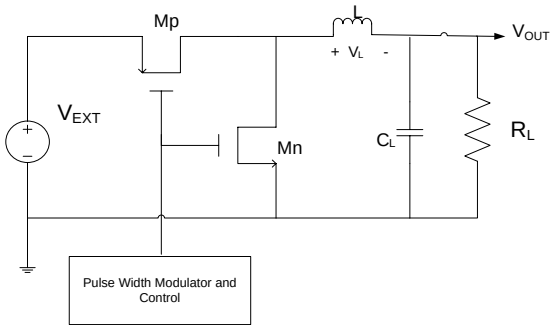

Figure 4.1: Buck Converter

The NMOS transistor switch acts as a freewheeling diode and provides a path for the inductor current when the PMOS switch is turned off. This switch enables the converter to transfer the energy stored in the inductor to the load. NMOS transistor is used (instead of diode) to achieve higher conversion efficiency, since the voltage drop across the NMOS in saturation region (typically 0.2-0.3V) would be smaller that drop across the diode (0.7V).

The external inductor L and capacitor C are used to store energy and transfer it to the load. Together they also filter the output voltage and current by providing bypass path for harmonic currents. Sizing of the inductor and capacitor determine output voltage ripple and overshoot. Small inductor values result in faster transient response but also require larger capacitor values to lower the output ripple.

4.1.1 Circuit Operation

The buck converter operates in two states, ON state and OFF state, depending on whether the power switch is on or off. When PMOS switch is turned on (ON state), The equivalent circuit during ON state is as shown in Figure 4.2. The inductor is connected to the external supply through the conducting PMOS switch. The inductor current i_L increases linearly and results in a positive voltage drop $v_L = V_{EXT} - V_{OUT}$ across the inductor. Energy is stored in the magnetic field of L and the electric field of C.

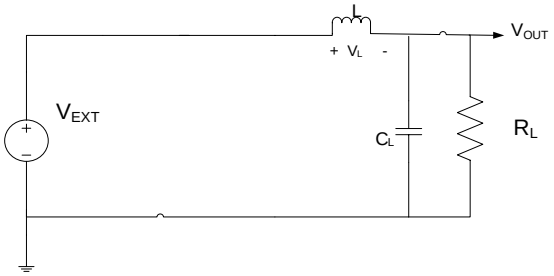

Figure 4.2: Buck Converter Equivalent Circuit During ON State

During OFF state, the inductor continues to carry current i_L as shown in Figure 4.3. The NMOS transistor is turned ON to provide path for circulating the inductor current. The inductor current linearly decreases with time, causing a negative voltage drop $v_L = -V_{OUT}$ across the inductor. The energy stored in L and C during ON state is transferred to the load. The waveforms of inductor and capacitor voltages and currents are shown in Figure 4.4.

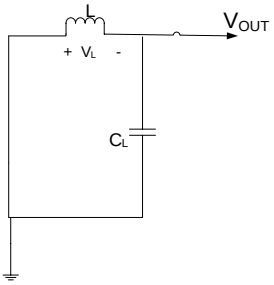

Figure 4.3: Buck Converter Equivalent Circuit During OFF State

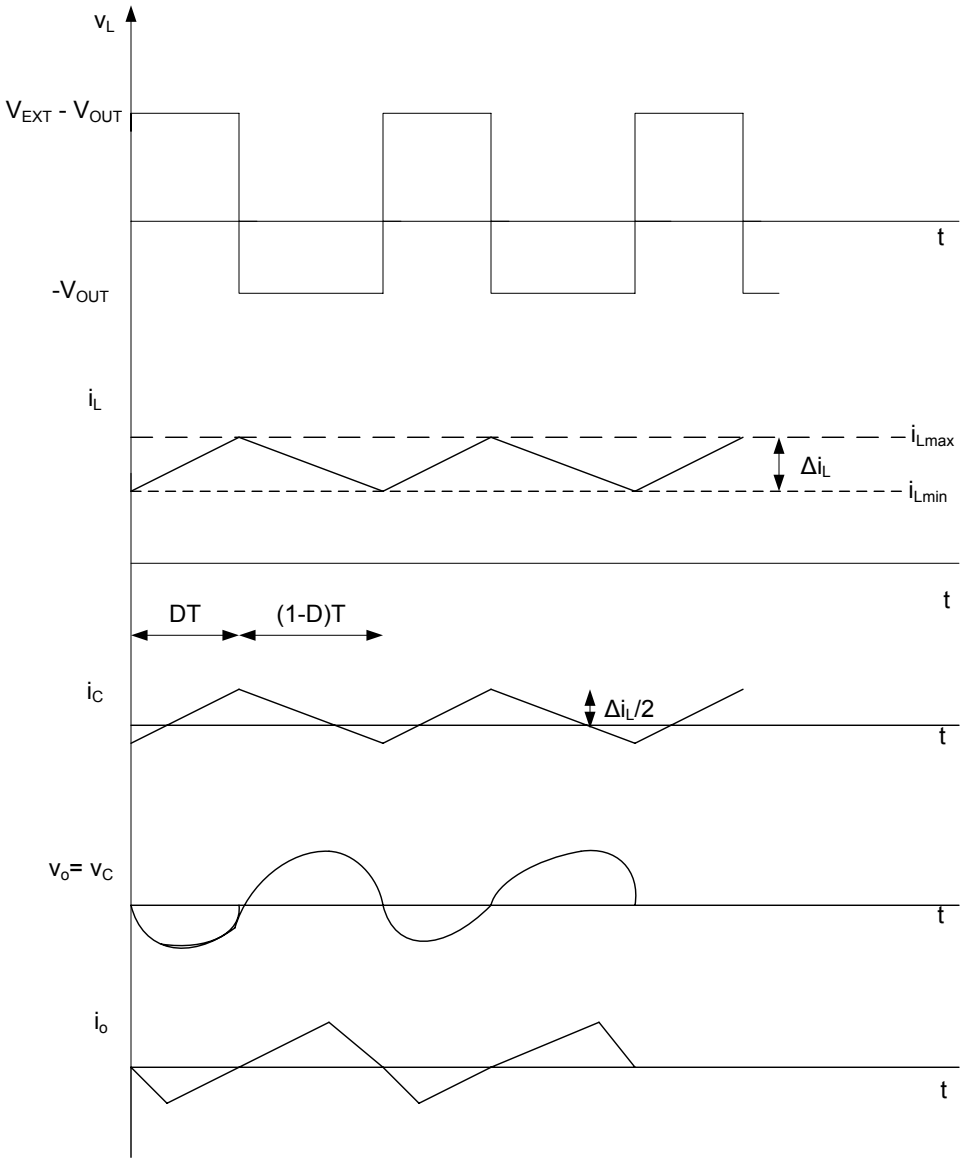

Figure 4.4: Current and Voltage Waveforms for the Buck Converter

4.1.2 Analysis of Buck Converter

To derive expressions for duty ratio D, output voltage ripple $\Delta V_{OUT}/V_{OUT}$, inductor current i_L and conversion efficiency η, it is assumed that the circuit operates in steady state continuous conduction mode.

Duty ratio can be found out by equating the integral of the inductor current over an entire switching period to zero. From waveforms in Figure 4.4

$$\int_0^T v_L dt = \int_0^{DT} v_L dt + \int_{DT}^T v_L dt = 0$$

$$(V_{EXT} - V_{OUT}) \times DT + (-V_{OUT}) \times (1 - D)T = 0$$

$$D = \frac{V_{OUT}}{V_{EXT}} \tag{4.4}$$

The difference between the minimum and maximum value of the inductor current, Δi_L can be calculated by integrating the voltage across the inductor during ON state, as shown in Figure 4.4.

$$\Delta i_L = \frac{1}{L} \int_0^{DT} v_L dt$$

$$\Delta i_L = \frac{V_{EXT} - V_{OUT}}{L} \times DT \tag{4.5}$$

The average value of inductor current equals the output load current, which is given by

$$I_L = I_{OUT} = \frac{V_{OUT}}{R} \tag{4.6}$$

The difference in peak-to-peak output voltage, ΔV_{OUT}, can be found by integrating the current through the capacitor as shown in Figure 4.4.

$$\Delta V_{OUT} = \Delta v_C = \frac{1}{C} \left(\int_{\frac{DT}{2}}^{DT} i_C dt + \int_{DT}^{\frac{(D+1)T}{2}} i_C dt \right)$$

Using the waveforms in Figure 4.4, the peak value of i_C is $\Delta i_L/2$. Therefore

$$\Delta V_{OUT} = \frac{1}{C} \times \frac{1}{2} \times \frac{T}{2} \times \frac{\Delta i_L}{2}$$

Using equations (4.4) and (4.5), the expression for output voltage ripple is given by

$$\frac{\Delta V_{OUT}}{V_{OUT}} = \frac{(1-D)T^2}{8LC} \tag{4.7}$$

For 50% duty cycle ($D = 0.5$), the above equation simplifies to

$$\frac{\Delta V_{OUT}}{V_{OUT}} = \frac{T^2}{16LC}$$

The power conversion efficiency η can be given by[Kuroda98]

$$\eta = \frac{V_{OUT} I_{OUT}}{V_{OUT} I_{OUT} + I_{OUT}^2 R_{switch} + P_X + P_{CTRL}} \tag{4.8}$$

where R_{switch} is the effective ON resistance of the switch transistor, P_X is the power consumed at node X, output of the power switch. P_X is caused by overshoot and undershoot in voltage V_X due to fluctuations in the inductor current. P_{CTRL} is the power dissipated in the control loop.

4.1.3 Modes of Operation of Buck Converter

The buck converter can operate either in continuous conduction mode (CCM) or discontinuous conduction mode (DCM) depending on whether the inductor current drops to zero at the end of the OFF period. Figure 4.5 shows the inductor current and voltage across it for both the conduction modes.

In CCM, current flows continuously through the inductor during the entire switching period. Overall performance of DC/DC converter is better using CCM, since it allows maximum power to be delivered to the load for a given input voltage. In DCM, the current through the inductor rises from zero to the peak value, drops to zero and then remains at

zero for some portion of the switching cycle. DCM is better suited for applications where the desired output load current is low. In such applications, DCM requires smaller inductor and hence results in smaller converter size.

The boundary condition between the continuous and the discontinuous mode is the condition when the inductor current goes to zero at the end of the OFF state. At the boundary condition,

$$I_L = \frac{\Delta i_L}{2}$$

$$i_{L,min} = 0$$

$$i_{L,max} = \Delta i_L$$

From equations (4.4), (4.5) and (4.6)

$$\frac{V_{OUT}}{R} = \frac{1}{2L} V_{OUT}(1-D)T$$

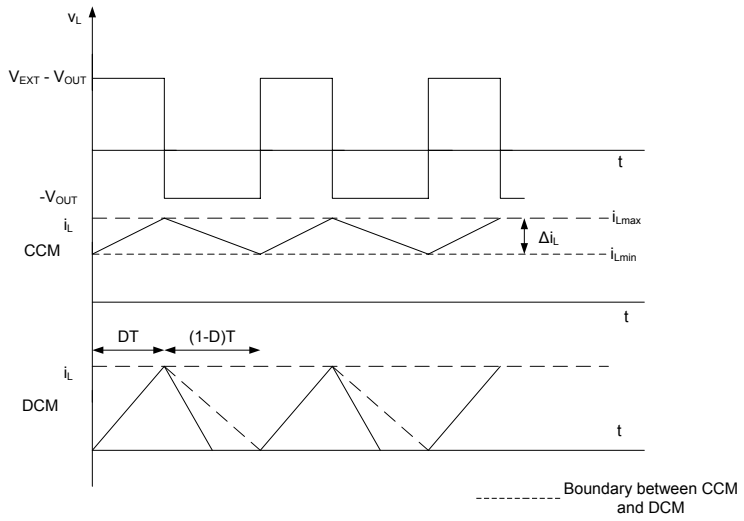

Figure 4.5: Inductor Current and Voltage waveforms for CCM and DCM

Therefore the minimum inductor value required for continuous conduction mode is calculated as

$$L_{min} = \frac{(1-D)T \times R}{2} \tag{4.9}$$

4.2 Pulse Width Modulation Technique

In DC/DC converters, output voltage regulation under varying supply and load conditions is an important design consideration. To maintain the average output voltage under different line and load conditions, a negative feedback control system is used in conjunction with the pulse width modulator to generate the control signal for the converter power switch. In pulse width modulation (PWM) technique, the duty ratio of the power switch is modulated by a pulse train to regulate the output voltage such that the load can only see the average value of the output voltage. The train of pulses charges and discharges the inductor in finer steps, thus lowering the peak inductor current and the output ripple. PWM control allows operation of the converter at a constant switching frequency and improves the power efficiency of the converter.

The PWM control signal is generated by comparing the control voltage from the feedback controller with a repetitive waveform as shown in Figure 4.6. The controller output is compared with a sawtooth waveform, which determines the switching frequency for the converter. When the controller output is higher that the sawtooth waveform, the PWM control signal becomes HIGH to turn on the power switch. Thus the threshold voltage of the comparator and hence, the duty cycle of the PWM signal are controlled by the controller output.

4.3 Feedback Control Loop Design

Figure 4.7 shows the block diagram of the proposed control loop for the supply voltage scaling scheme. The design has been implemented in 180-nm process. The synchronous buck converter provides the regulated output voltage from an external supply. A voltage controlled ring oscillator is used to model the delay of the critical path in a circuit.

The output frequency of the ring oscillator is directly proportional to the supply voltage. All the components of the control loop are discussed in the following subsections.

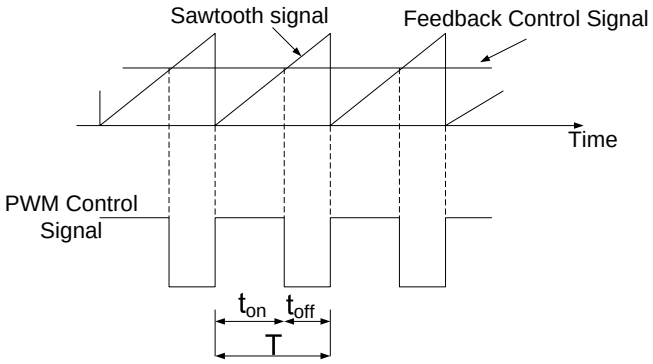

Figure 4.6: PWM Control Signal for Switches in Buck Converter

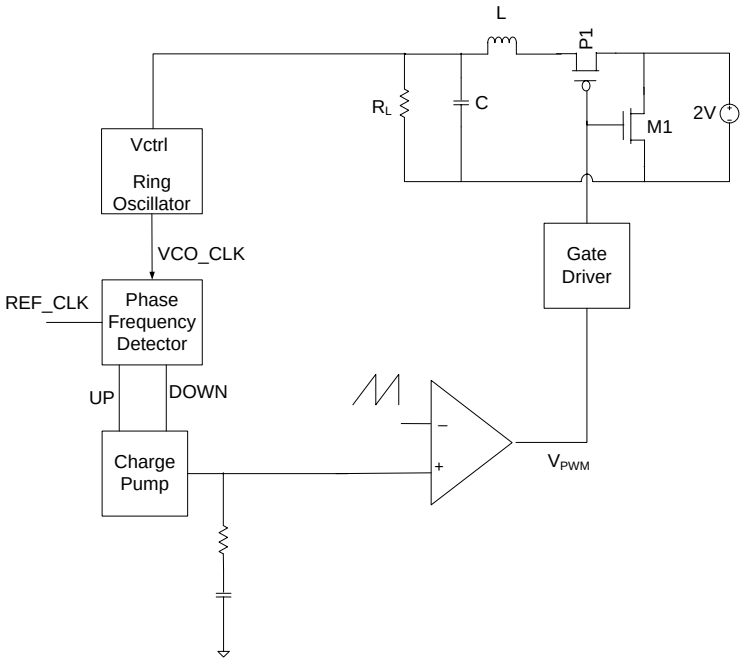

Figure 4.7: Control-Loop Block Diagram

4.3.1 Phase Frequency Detector(PFD)

The ring oscillator in the feedback path translates the regulated supply voltage into frequency. Phase frequency detector(PFD), shown in Figure 4.8, is used to compare the oscillator frequency (VCO_CLK) with the desired operating frequency (REF_CLK). It consists of two delay flip-flops with an asynchronous reset and the AND gate. The delay flip-flop is implemented using transmission gates and NOR gates.

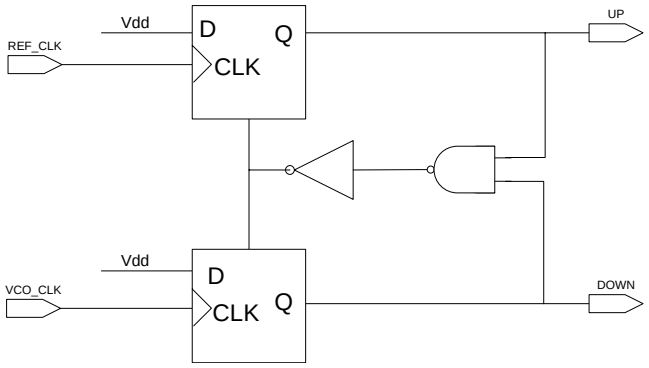

Figure 4.8: Phase Frequency Detector Schematic

The PFD provides UP and DOWN output pulses which are measure of the phase and frequency difference between the two inputs. It has three distinct states of operation. Both the outputs go from low to high on the rising edge of the REF_CLK and VCO_CLK respectively. When REF_CLK is higher than the VCO_CLK, output UP remains high till the low-to-high transition of the DOWN output. The asynchronous reset is asserted whenever both the outputs are high, causing the flip-flops to reset after a short duration equal to the propagation delay of the NAND gate, an inverter and flip-flop reset-to-Q delay. The sequence of events is reversed when VCO_CLK exceeds REF_CLK. In both cases, pulse width of the UP/DOWN pulses indicates the phase error between the two input frequencies. When the difference between REF_CLK and VCO_CLK is high, the circuit tracks the frequency difference, producing more UP or DOWN pulses for higher REF_CLK or VCO_CLK respectively. The average value of the pulse is proportional to the frequency error. Figure 4.9, 4.10 and 4.11 show the timing waveforms for the PFD for all the three cases.

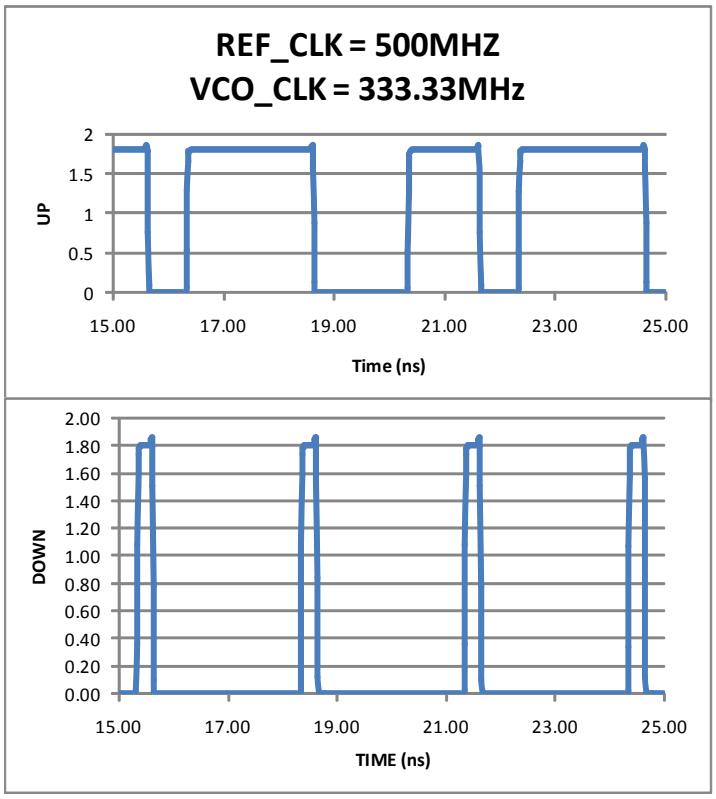

Figure 4.9: Timing Waveforms for PFD. $VCO\_CLK < REF\_CLK$

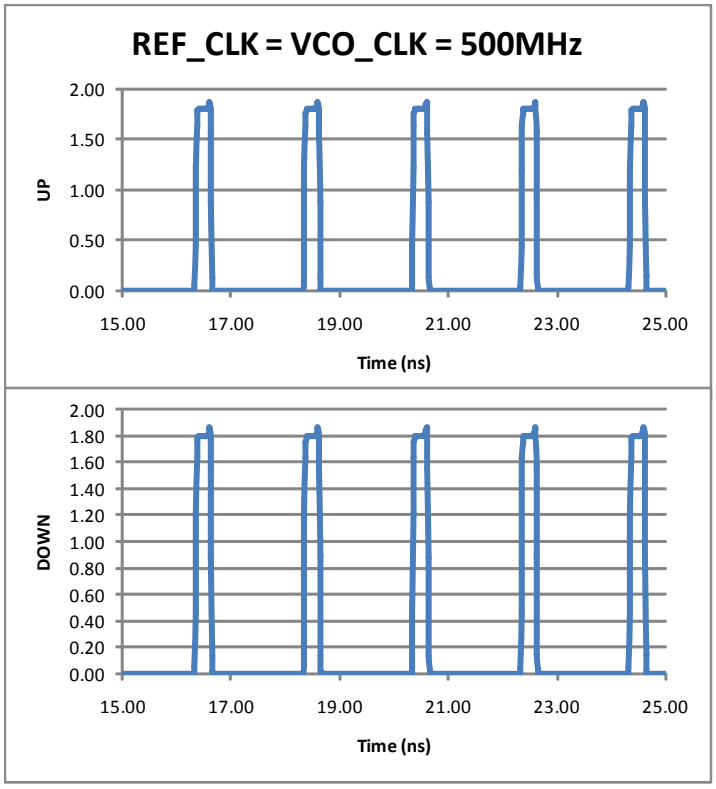

Figure 4.10: Timing Waveforms for PFD. $VCO\_CLK = REF\_CLK$

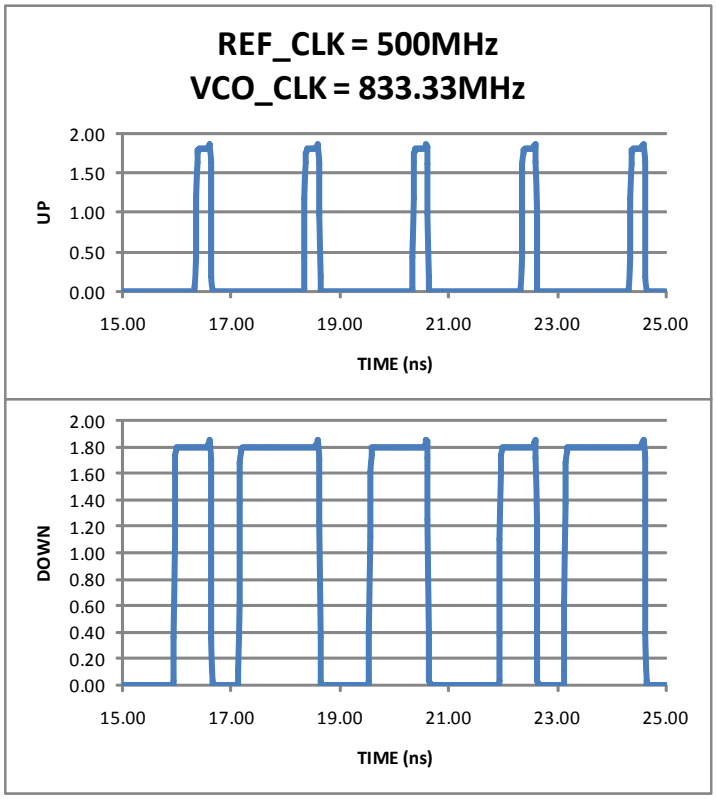

Figure 4.11: Timing Waveforms for PFD. $VCO\_CLK > REF\_CLK$

4.3.2 Charge Pump

The function of the charge pump is to convert UP and DOWN pulses into an analog voltage signal using current switches and loop filter. The UP signal controls the charge injection into the loop filter while DOWN signal removes charge from the loop filter. The amount of charge injected or removed is proportional to the pulse width of the UP or DOWN signal respectively.

As discussed earlier, the PFD will generate UP and DOWN pulses with equal duration when both REF_CLK and VCO_CLK are equal. In order to avoid any static offset error, the charge pump should not transfer any charge to the loop filter when both UP and DOWN signals have equal pulse widths. This requires that both I_{UP} and I_{DOWN} must be equal and independent of the output voltage[Maneatis96]. Using symmetric active loads and source coupled differential pairs, it is possible to match the two currents when UP and DOWN pulses have equal duration. The schematic of the charge pump is shown in Figure 4.12. The PMOS devices at the top form the symmetric resistive load. Two NMOS differential pairs are used to supply or remove charge from the loop filter. When the pulse width of UP signal is higher, PMOS load in the differential pair on the left will draw larger current. This current is mirrored into the PMOS load in the right side differential pair. However, the right differential pair will sink a current proportional to the pulse width of the DOWN signal. The difference in these two currents is sourced to the loop filter, thus increasing the output voltage. When the DOWN signal has higher pulse width, the right differential pair will sink more current than that supplied by the PMOS load. This results in removal of charge from the loop filter and hence the output voltage decreases. For equal duration UP and DOWN pulses the current flowing through the PMOS current mirror exactly matches the current sunk by the differential pair. As a result, no net charge is transferred to the loop filter. The output voltage waveform is as shown in Figure 4.13.

4.3.3 Loop Filter

The loop filter acts as a charge storage element for the charge pump. The simplest loop filter would consist of a single capacitor. However, it creates a pole at origin which causes instability. A series resistance R is often added to provide stability. In addition, an extra small capacitor C_2 as shown in Figure 4.14 is added for reducing the output ripple.

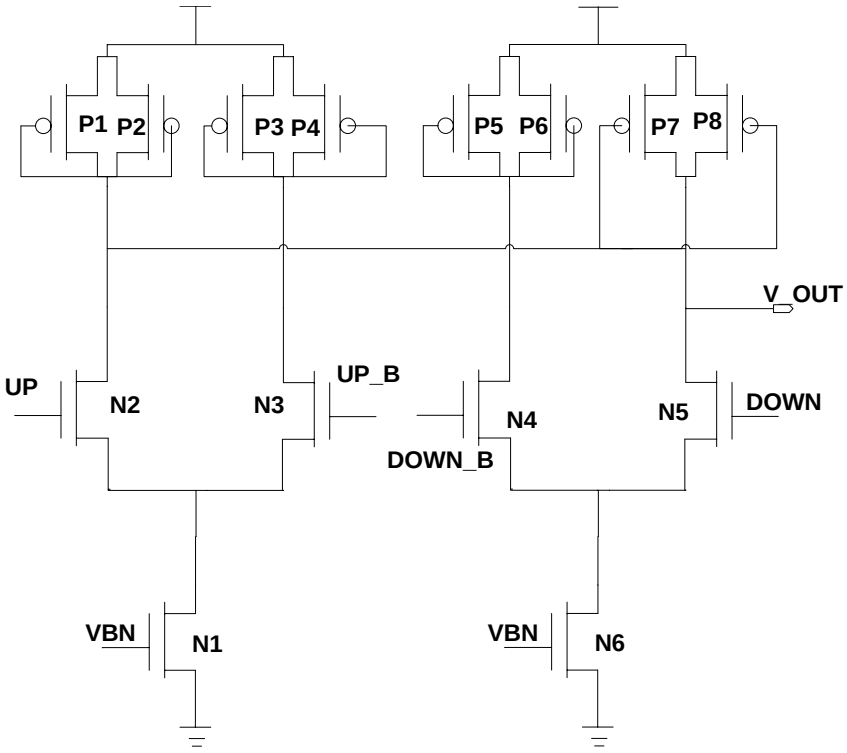

Figure 4.12: Charge Pump Schematic

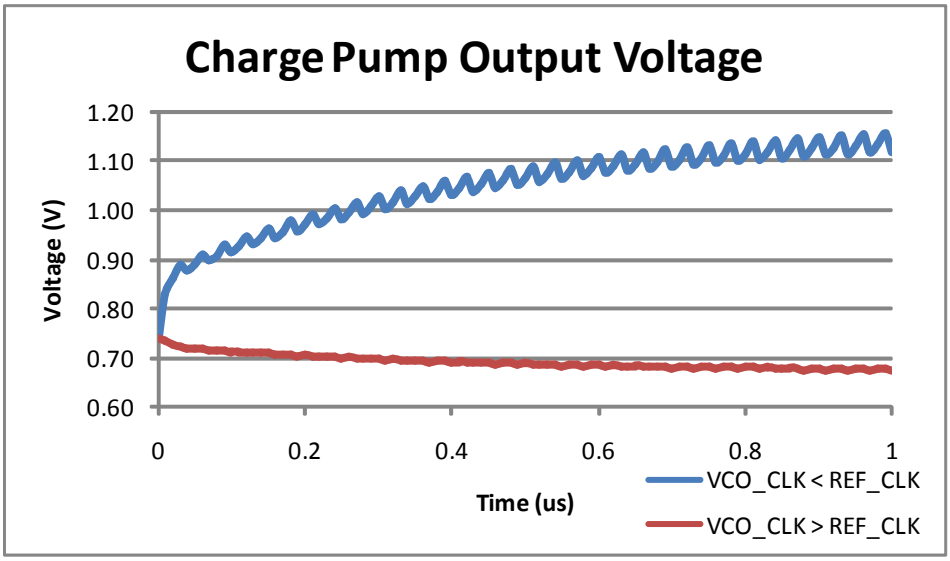

Figure 4.13: Charge Pump Output Voltage Waveforms

The loop filter plays an important role in the stability and the dynamic response of the control loop. It determines the closed loop bandwidth. Lower loop bandwidth results in faster dynamic response but higher jitter. On the other hand, higher loop bandwidth reduces input noise and jitter at the expense of longer settling time. For stability of the control loop, the loop bandwidth should be at least a decade below the minimum operating frequency[Maneatis96].

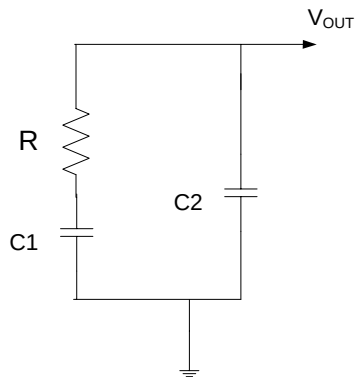

Figure 4.14: Second Order Loop Filter

4.3.4 Ring Oscillator

[Wei99] shows that the ring oscillator delay closely tracks the critical path delay of the circuit across process and temperature variations. Hence ring oscillator is a suitable circuit for detecting the effect of supply voltage variation on the speed of the circuit. Ring oscillator based on differential delay-cell, is chosen for better common mode noise rejection. The schematic of the differential delay cell is shown in Figure 4.15. The buffer stage consists of NMOS source coupled differential pair with symmetric load elements. Diode-connected PMOS transistor in parallel with an equally sized biased PMOS transistor forms the symmetric load[Maneatis96]. The effective resistance of the symmetric load changes linearly with the bias voltage (V_{BP}). The NMOS current source at the bottom, biased at V_{BN}, decides the delay cell tail current. Both the bias voltages (V_{BN} and V_{BP}) are generated using replica-feedback bias circuit, which will be discussed in the next subsection.

Due to the diode-connected PMOS transistor, the effective resistance R_{EFF} of the symmetric load element equals $1/g_m$, where g_m is the circuit transconductance. Hence, the

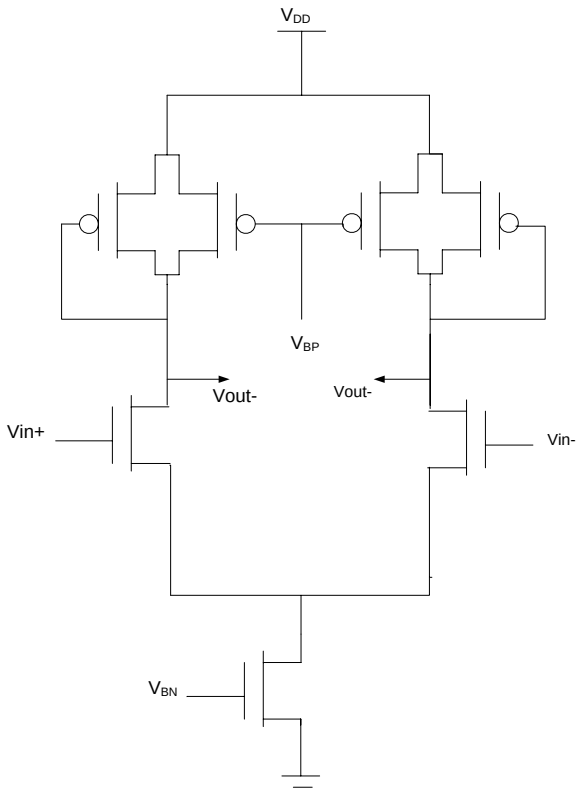

Figure 4.15: Schematic of Differential Delay Cell Used in Ring Oscillator Design

delay of a differential cell is given by

$$T_d = R_{EFF} \cdot C_{EFF} = \frac{1}{g_m} \cdot C_{EFF} \tag{4.10}$$

where C_{EFF} is the effective output capacitance[Maneatis96]. The transconductance g_m can be written in terms of bias voltage V_{BP} and bias current I_D as

$$g_m = k \cdot (V_{DD} - V_{BP} - V_{TH}) = \sqrt{2kI_D} \tag{4.11}$$

where k is PMOS device transconductance. From the above equations, delay of single cell is given by

$$T_d = \frac{C_{EFF}}{k \cdot (V_{DD} - V_{BP} - V_{TH})} \tag{4.12}$$

The complete ring oscillator configuration is shown in Figure 4.16. It consists of 4 stages of differential delay cells. The oscillation frequency is given by

$$F_{OSC} = \frac{1}{2 \cdot N \cdot T_d} = \frac{k \cdot (V_{DD} - V_{BP} - V_{TH})}{8 \cdot C_{EFF}} \tag{4.13}$$

where N is the number of delay stages. The oscillator gain K_V can be obtained by taking the derivative of F_{OSC} with respect to $V_{DD} - V_{BP}$.

$$K_V = \frac{k}{8 \cdot C_{EFF}} \tag{4.14}$$

The change in the oscillation frequency with $V_{DD} - V_{BP}$ is shown in Figure 4.17.

Differential-to-Single-Ended Converter

The ring oscillator produces differential outputs V_{OUT+} and V_{OUT-}. To compare the oscillator frequency with the reference frequency, it is necessary to convert these differential outputs to single ended output. In addition, the bias voltage V_{BP} limits the

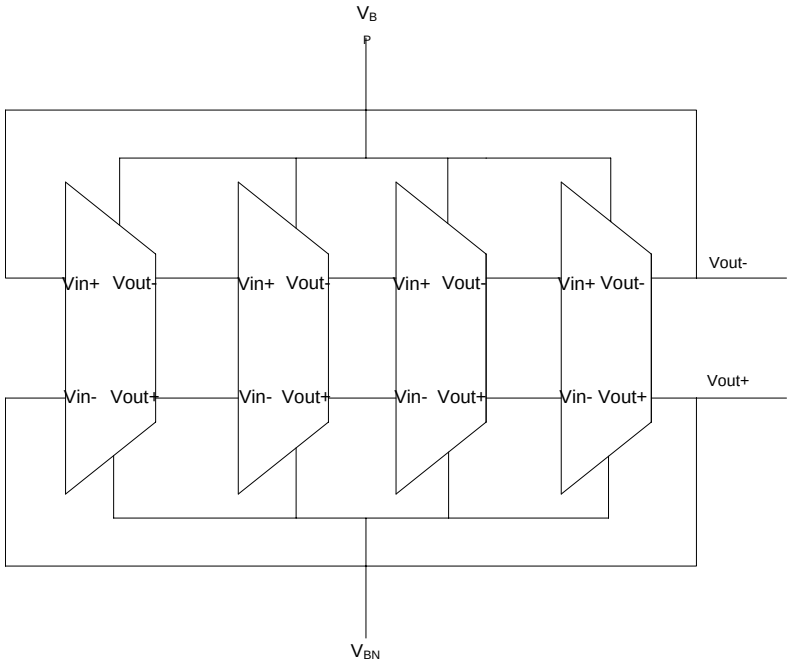

Figure 4.16: Schematic of 4-stage Ring Oscillator

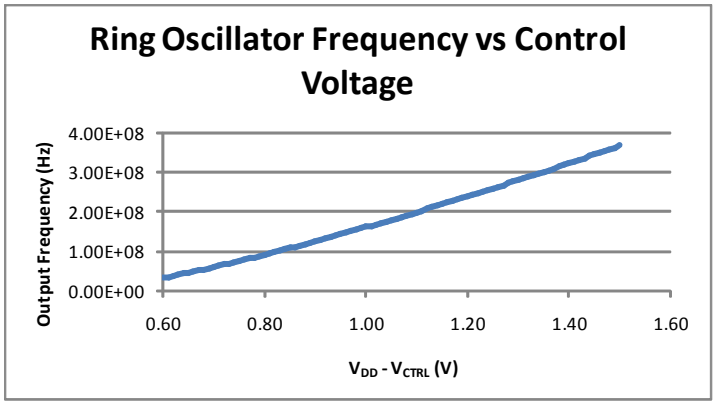

Figure 4.17: Ring Oscillator Frequency as a function of Bias Voltage V_{BP}

lower output voltage swing to $V_{DD} - V_{BP}$. A differential-to-single-ended converter from [Maneatis96], is used to obtain full rail to-rail swing single-ended output. As shown in Figure 4.18, it consists of two NMOS differential amplifiers with PMOS current mirrors. Both the differential amplifiers are driven by the same current source bias voltage V_{BN} as the oscillator delay cell to match the common-mode input voltage. The differential amplifiers pair convert the input voltage into differential current signals. Two PMOS common-source amplifiers provide signal amplification and conversion to single ended output through the NMOS current mirror.

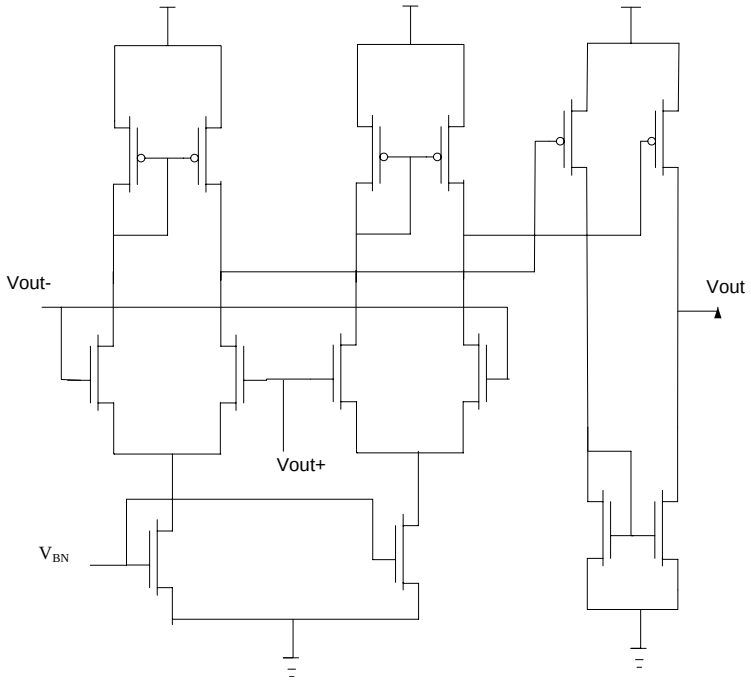

Figure 4.18: Schematic of Differential-to-Single-Ended Converter

4.3.5 Replica-Feedback Bias Generator

To make the delay of the ring oscillator dependent on the regulated output voltage of the buck converter, it is necessary that the delay cell bias voltages V_{BP} and V_{BN} are generated from the converter output voltage. Replica-feedback bias circuit is used to dynamically adjust these bias voltages depending on the regulated buck converter output. It sets the tail current of the delay cell independent of the supply voltage, thus improving

the supply noise rejection.

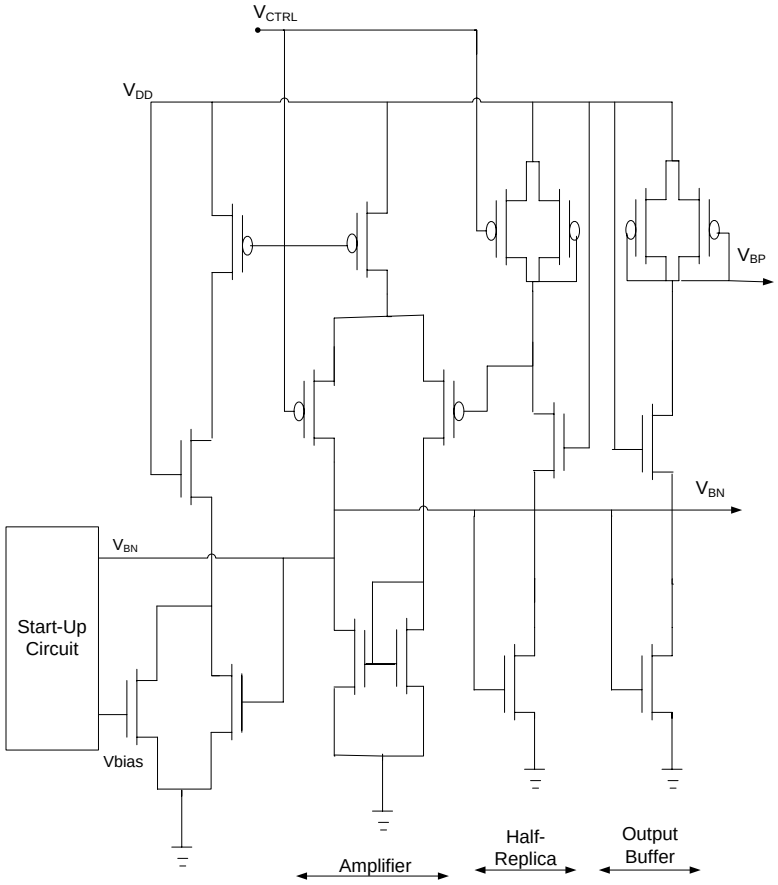

Figure 4.19: Replica-Feedback Bias Generator

As shown in Figure 4.19, the replica-bias circuit consists of differential amplifier, half-replica of the delay cell (hence the name, replica-bias) and an output buffer to provide V_{BP} and V_{BN}. V_{CTRL} is the control voltage for the bias generator, which in this case comes from the buck converter output. The amplifier inputs are connected to V_{CTRL} and the output of a half-replica of the delay cell. The half-replica of the delay cell consists of current source transistor, a differential NMOS transistor with input connected to supply voltage and the PMOS symmetric load. The amplifier adjusts V_{BN} such that the two inputs of the amplifier are virtually equal. In other words, amplifier forces the half-replica output voltage to equal V_{CTRL}, by dynamically adjusting the current source bias V_{BN}. Any changes in the supply voltage are also compensated by adjusting V_{BN} to keep the

bias current constant. The output buffer stage is an additional half-replica used to provide PMOS bias voltage V_{BP}, which is nominally equal to V_{CTRL}. A start-up circuit is needed to bias the amplifier at power-up. The start-up circuit, shown in Figure 4.20, prevents the amplifier output (V_{BN}) from turning off the NMOS tail current sources [Maneatis93].

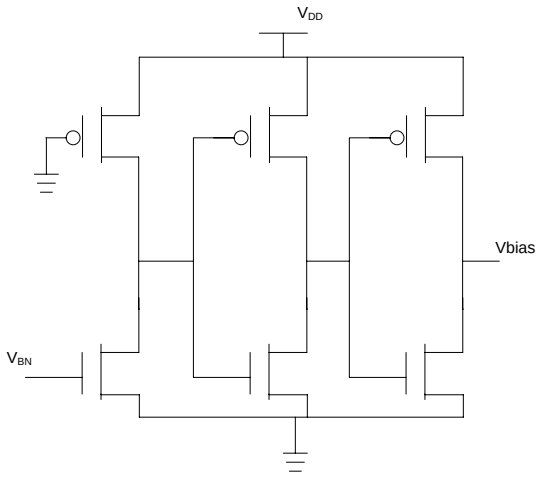

Figure 4.20: Start-up Circuit for Replica-Feedback Bias Generator

4.3.6 Pulse Width Modulator Design

The function of pulse width modulator (PWM) is to establish a switching frequency and to provide PWM control signal for the converter switches. Higher switching frequency improves the transient response of the converter but results in higher power loss in output inductor. On the other hand, if the switching frequency is too low, the converter output takes a long time to settle down. As a trade off, switching frequency of 1MHz is chosen in this design. The comparator in the pulse width modulator is implemented using 2-stage cascode amplifier. The schematic of the amplifier is shown in Figure 4.21. Frequency response of the amplifier is shown in Figure 4.22. The amplifier has a dc gain of 86.5dB and the unity gain bandwidth equal to 12MHz.

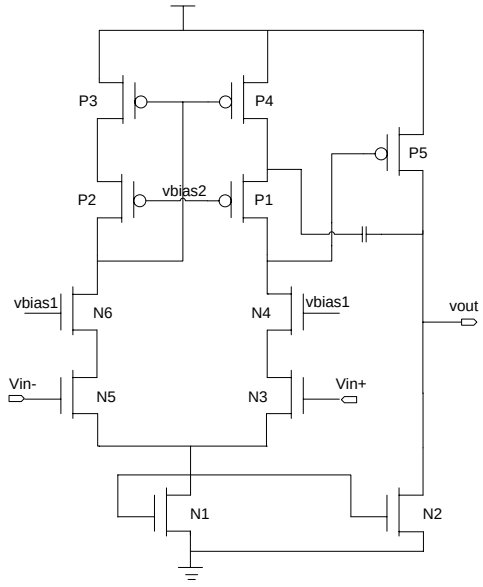

Figure 4.21: Schematic of the 2-stage amplifier used in PWM Comparator

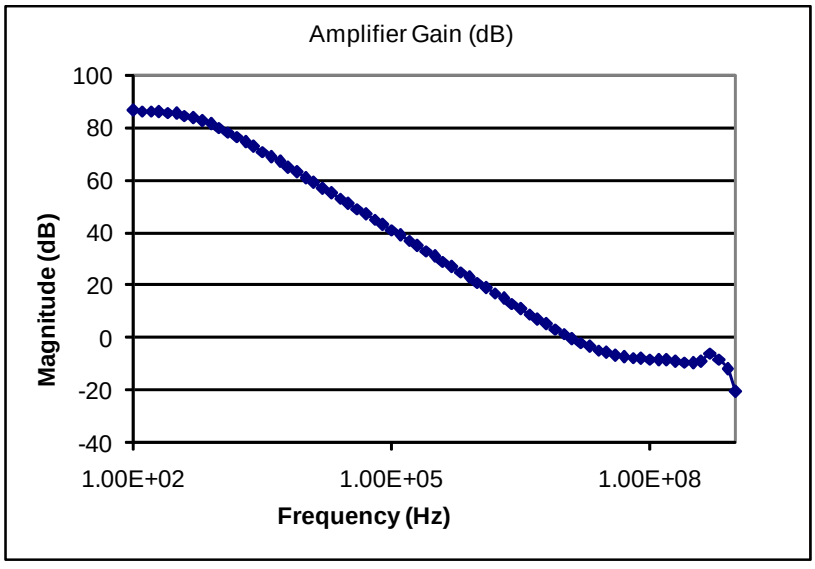

Figure 4.22: Comparator Frequency Response

Table 4.1: Design Specifications for Buck Converter

Output Voltage	0.6-1.5V
Output Ripple	250mV
Input Voltage	2V
Settling Time	100μ s
Switching Frequency	1MHz

4.4 Buck Converter Design

Design of buck converter involves output filter design and gate driver design. The design specifications considered for the converter are listed in Table 4.1.

4.4.1 Output Low Pass Filter Design

The external inductor L, capacitor C form a low pass filter together with the effective resistance R of the transistor switches used in the converter. The low pass filter should be designed considering the requirements of output voltage ripple and transient response. The damping factor of the filter is given by

$$\zeta = \frac{R}{2}\sqrt{\frac{C}{L}} \tag{4.15}$$

The filter should maintain the output ripple below 250mV. The filter resonant frequency of 10Hz is selected, well below the switching frequency. This provides the necessary attenuation at the switching frequency for output ripple requirement. The settling time of the output voltage is given by $\sqrt{LC}$. To minimize the winding loss in the inductor, smaller value of inductor should be selected. The inductor and capacitor values were selected to be $8\mu H$ and $32\mu F$ respectively. From Eq.(4.7), these values of L and C result in an output voltage ripple given by

$$\frac{\Delta V_{OUT}}{V_{OUT}} = 0.488 \times (1 - D)$$

For a control signal with 50% duty cycle, the ripple is 0.244. The transistor switches should have minimum resistance to reduce the voltage drop across them, which would reduce the power conversion efficiency. To achieve an effective resistance of 1Ω, PMOS and NMOS

transistor sizes are selected to be $8mm$ and $4mm$. Gate drivers are used to drive these large sized transistors.

4.4.2 Gate Driver Design

Four cascaded inverter stages are used to drive the large input capacitance of power converter switches. The first stage PMOS and NMOS transistors are sized to $8\mu m$ and $4\mu m$ respectively. The scale up factor x should be selected to minimize the power consumption in gate drivers. From [Kuroda98], expressions for x can be given by

$$x = 1 + \sqrt{1 + K} \tag{4.16}$$

where K is the ratio of dynamic power consumption to the power consumption due to direct path for an inverter loaded with an identical inverter. The typical value of K equals 8, which gives an optimum sizing factor of 4.

4.5 Simulation Results

The supply voltage controller was simulated in Cadence environment using SPECTRE circuit simulator. The reference frequency (REF_CLK) range used for simulation purposes is from 50MHz to 200MHz. The output voltage waveform for 100MHz operation is shown in Figure 4.23. The output voltage settles to final steady state value of 950mV within $100\mu s$. The corresponding PWM control signal is shown in Figure 4.24.

The ring oscillator output frequency waveform is shown in Figure 4.25. The oscillator frequency increases with time until it equals the reference frequency.

Figure 4.26 shows the output voltage waveforms for different reference frequencies. The output voltage takes longer time to settle for frequencies below 100MHz, which can be improved either by increasing the switching frequency or by decreasing the output filter time constant.

4.5.1 Converter Efficiency Calculation

The converter efficiency can be calculated from Eq.(4.8), which is repeated below.

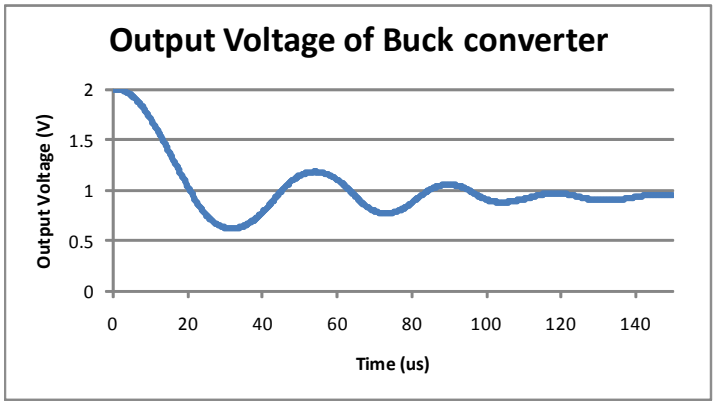

Figure 4.23: Output Voltage of Buck Converter for 100MHz operation

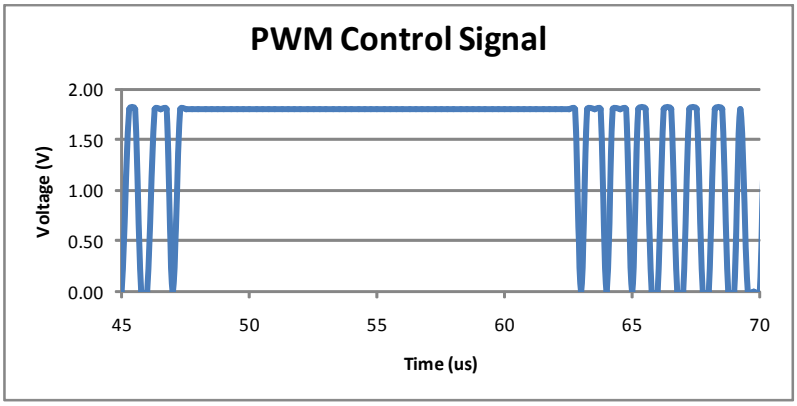

Figure 4.24: PWM Control Signal

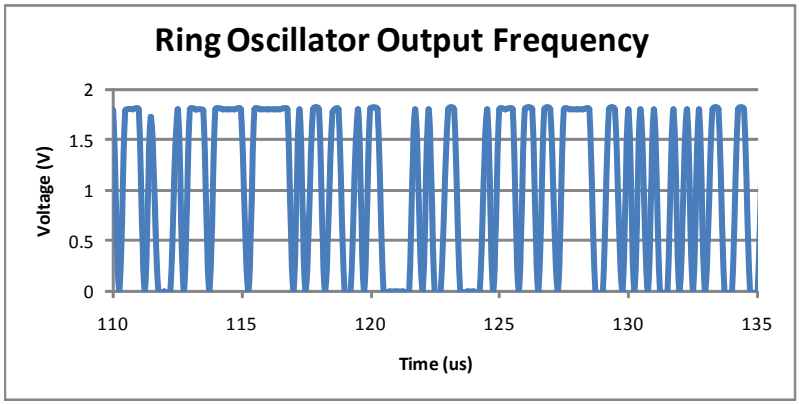

Figure 4.25: Ring Oscillator Output Frequency

$$\eta = \frac{V_{OUT} I_{OUT}}{V_{OUT} I_{OUT} + I_{OUT}^2 R_{switch} + P_X + P_{CTRL}}$$

where R_{switch} is the effective ON resistance of the switch transistor, P_X is the power consumed at node X, output of the power switch. P_X is caused by overshoot and undershoot in voltage V_X due to fluctuations in the inductor current. P_{CTRL} is the power dissipated in the control loop.

For 100MHz operation, the average values of V_{OUT} and I_{OUT} are $0.95V$ and $224.1mA$ respectively. From [Kuroda98], the power consumption at node X is given by

$$P_X = \frac{V_{EXT}^2}{R_{switch}} \frac{1}{24 F_S^2 LC} \tag{4.17}$$

where F_S is the switching frequency. P_{CTRL} depends on the supply voltage ($V_{DD} = 1.8V$)used for control loop and the current supplied by V_{DD}. From the simulation results, this power equals $7.7mW$. The converter efficiency can be then calculated as

$$\eta = \frac{0.95V \times 224.1mA}{0.95V \times 224.1mA + (224mA)^2 \times 1\Omega) + 0.65mW + 7.7mW}$$

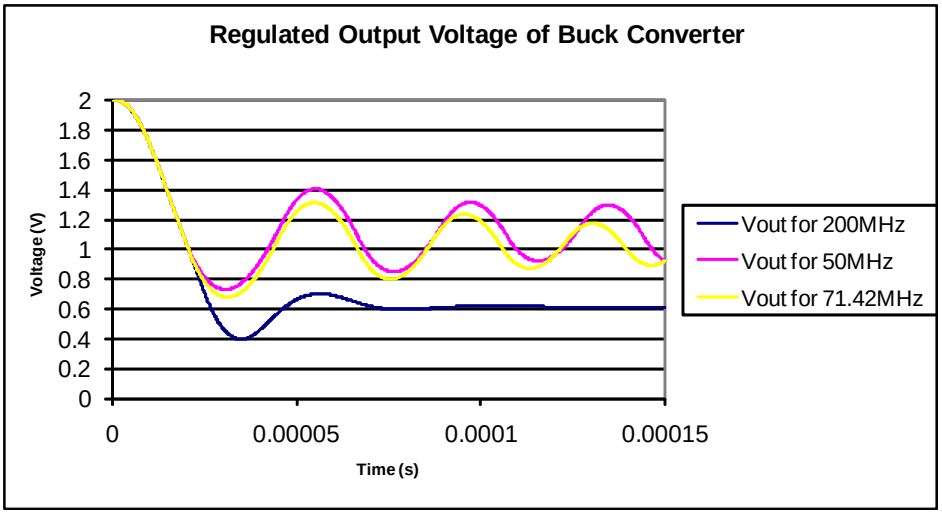

Figure 4.26: Output Voltage for Different Reference Frequencies

Table 4.2: Performance of Designed Buck Converter

Switching Frequency	1MHz
Output Voltage	$0.6V - 1.2V$
Output Ripple	50mV
Settling Time	$100\mu s$
Conversion Efficiency	$64\% - 80\%$

$$\eta = 78.42\%$$

The efficiency of the buck converter depends on the frequency of operation. At lower operating frequency (50MHz), the converter efficiency is 80.5%. With the increase in frequency, the efficiency drops to 64.85% at 200MHz. The performance of the designed buck converter is summarized in Table 4.2.

4.5.2 Power Savings using Supply Voltage Scaling

The supply voltage scaling using synchronous buck converter reduces the dynamic power consumption of the ring oscillator. Figure 4.27 shows the plot of dynamic power consumption against the operating frequency for fixed and variable supply voltage. At 100MHz frequency, the power consumption using adaptive supply voltage is 0.547mW, which is 52.77% lower than 1.16mW power consumped using fixed supply voltage. Even at 200MHz operation, the power consumption under variable supply is 33.33% lower than that under fixed supply voltage condition. At higher frequencies, power savings decreases due to the higher winding loss in inductor and the capacitive switching loss. Additional power savings are possible by reducing the switching frequency at the expense of longer transient response.

The efficiency of the buck converter greatly affects the performance of the entire control scheme. The winding losses in the inductor core, power consumed by the gate drivers and the voltage drop across the power switches limit the maximum achievable efficiency. Zero voltage switching (ZVS) technique can improve the efficiency by controlling the on/off timing of the power switches. In ZVS, the output inductor is used to charge and discharge the output capacitance of power transistors in a lossless manner. This allows the transistors to be switched at zero drain-to-source potential[Stratakos94]. The condition for achieving

ZVS can be explained with the help of Figure 4.28. V_X is the voltage at the output of power transistors and i_L is an instantaneous current through the inductor.

The switching cycle is divided into four time segments. During T_2 and T_4, the PMOS and NMOS power transistors are on respectively. Without ZVS, both the transistors are conducting during T_1 and T_3. These time fragments, T_1 and T_3, are called

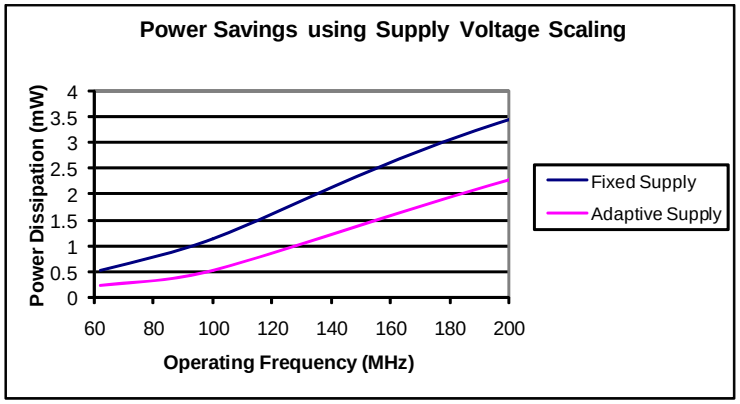

Figure 4.27: Power Savings using Supply Voltage Scaling

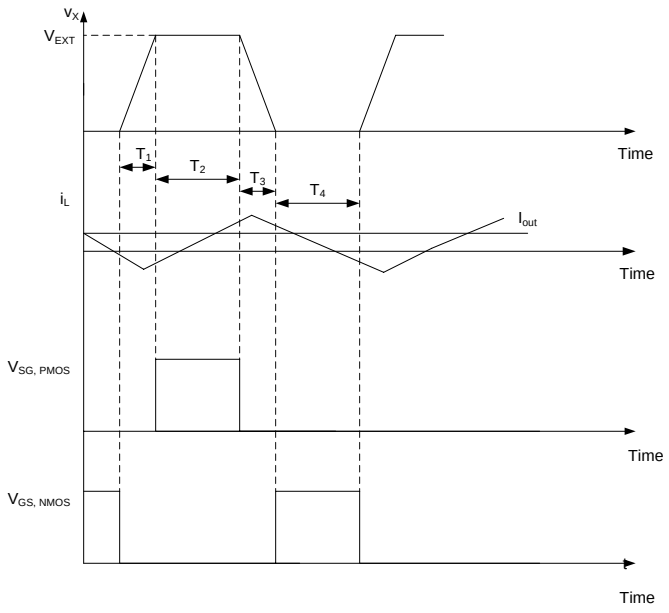

Figure 4.28: Waveforms of buck converter under Zero Voltage Switching Condition

deadtimes[Long06].

At the beginning, NMOS is on and the voltage across the inductor is negative. The inductor current decreases linearly. If NMOS is turned off after the inductor current reverses, then the inductor acts as a current source charging the output node[Stratakos94]. As a result, V_X charges linearly to V_{EXT}, where V_{EXT} is the external input voltage of the buck converter. To achieve lossless low-to-high transition, PMOS power switch should be turned on when V_X equals V_{EXT}. In other words, PMOS device turns on when its drain-to-source voltage is $0V$. For lossless high-to-low transition, PMOS switch is first turned off. The output inductor once again acts as a current source and discharges the output node capacitance. The NMOS switch is turned on when V_X drops to $0V$. Thus, both the power transistors are turned on with zero V_{DS} and ZVS is achieved. ZVS eliminates the switching loss in power transistors and allows use of higher switching frequency in low power systems[Stratakos94].

Other techniques such as variable switching frequency have been proposed in literature [Wei99]. The transient response of the buck converter can be improved by using PI or PID controller in the control loop[Trescases04]. An efficient buck converter together with such compensator would result in power savings up to 75% [Kuroda98].

4.5.3 Power and Area Trade Off

The synchronous buck converter, used for stepping down the supply voltage, requires large sized PMOS and NMOS devices to handle the high output current values. Large sizes are also necessary to minimize the body diode conduction losses and to reduce the voltage drop across the switches. Secondly, large values of inductor and capacitor are used to minimize the output ripple. These factors increase the total chip silicon area, which offsets the power savings achieved. However, the voltage converter has a fixed area overhead and using this scheme for a large circuit blocks, such as microprocessors would reduce the percentage increase in total area. [Kuroda98] mentions an area penalty of less than 1% for a 32 bit processor with 440,000 transistors.

For small, less complex systems, V_{DD} hopping scheme using PMOS switches would be more beneficial. The schematic of the V_{DD} hopping scheme is shown in Figure 4.29. The V_{DD} controller selects one of the external supply voltages depending on the reference frequency and the actual frequency of operation. Use of PMOS switches eliminates voltage

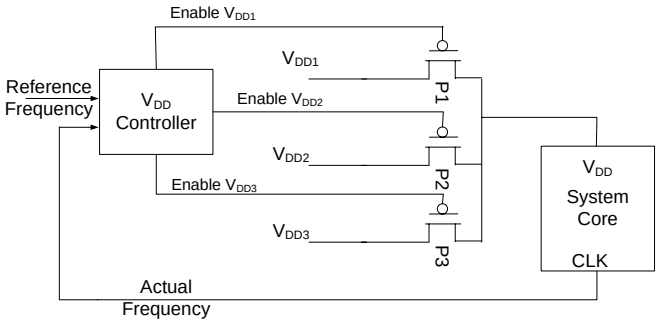

Figure 4.29: V_{DD} hopping scheme

drop due to body effect and lowers the gate leakage. This method does not require an inductor and a capacitor as in buck converter. As a result, the area overhead would be smaller than the supply voltage scaling scheme.

Chapter 5

Conclusion

Increasing leakage currents in nanoscale regime is an area of concern for high performance digital systems. The subthreshold leakage current is increasing exponentially with threshold voltage lowering. The reduced oxide thickness allows direct tunneling across the dielectric leading to higher gate leakage. The short channel effects (SCEs) are becoming more prominent with every technology scaling. To suppress SCEs, nonuniform doping profiles are used near source and drain junctions. However, the junction leakage current grows with higher doping concentrations. All these factors lead to an increase in total leakage power, which is becoming a major component of total power consumption. Hence, leakage reduction techniques are important in future high performance circuits.

Important leakage currents in CMOS circuits have been discussed in this thesis. The physical mechanisms governing leakage components have been briefly covered. In complex logic circuits, the subthreshold leakage and the on-state gate leakage interact with each other at the internal nodes. In stacked transistors configuration, gate leakage varies significantly depending on the input sequences. Methods for leakage reduction have been described. Basic principle of each method, relative advantages and limitations have been discussed.

The leakage monitor circuit has been designed to detect an optimum reverse bias voltage that minimizes the total standby leakage. The optimum reverse body bias voltage has been applied to single CMOS devices, an inverter pair and static NAND and NOR gates with nominal (reg-V_{TH}) and low (low-V_{TH}) threshold voltages. The results indicate

that the optimum reverse bias voltage reduces the standby leakage of both reg-V_{TH} and low-V_{TH} NMOS transistors by more than 90%. For PMOS transistors, the percentage of leakage reduction drops from 91.6% for reg-V_{TH} device to 69.6% for low-V_{TH} device due to the decrease in body effect parameter by 50%. For NAND and NOR gates, it has been found that the reverse biasing more effectively reduces the total leakage in low-V_{TH} circuits. In reg-V_{TH} circuits, the on-state gate leakage dominates the total standby leakage. Hence, the percentage reduction in total leakage is approximately 30% lower in reg-V_{TH} circuits than in low-V_{TH} circuits.

The efficiency of forward body biasing is evaluated using single devices, 5-stage ring oscillator and a 16-bit ripple carry adder. The improvement in on-state current under forward body bias condition is less than 15% for both NMOS and PMOS devices. For ring oscillator and 16-bit adder, only marginal improvement of less 10% has been observed in frequency. Since the subthreshold leakage increases exponentially with forward biasing, use of forward body bias for performance improvement is less beneficial in 90-nm process.

The supply voltage scaling technique for reducing the active power consumption has been implemented using a feedback control loop and a synchronous buck converter. The feedback control loop consists of a phase frequency detector and a charge pump to generate a control voltage based on the difference between the reference frequency and the actual frequency. The pulse width modulation (PWM) technique has been used for controlling the power switches in the buck converter stage. The amount of power savings depends largely on the efficiency of the buck converter. The designed buck converter efficiency lies in the range of $64-80\%$ for the desired operating frequencies. Zero voltage switching (ZVS) technique for reducing losses in the converter stage has been discussed.

5.1 Contribution of This Work

To detect an optimum reverse body bias for minimum standby leakage, an improved leakage monitor circuit has been designed. The circuit compares the subthreshold leakage and the BTBT leakage and detects an optimum reverse bias voltage where both these currents are equal in magnitude. The simulation results show that the circuit accurately tracks the leakage currents within $\pm 5\%$ of the actual leakage current values. An analog feedback control loop has been designed to implement supply voltage scaling. The

simulation results show that up to 50% power savings is possible using the variable supply voltage scheme.

5.2 Future Work

Leakage monitor circuit for detecting optimum reverse bias has been proposed in this work. An adaptive V_{TH} control loop can be designed to dynamically track the operating conditions and process variations. Such a dynamic threshold voltage scaling scheme (DVTS) would track the system variations and adjust the body bias to minimize both dynamic and leakage power consumption. The control loop of the supply voltage scaling scheme can be redesigned to include a proportional-integral (PI) or proportional-integral-derivative (PID) controller. Such controller would react quickly to the frequency error and improve the transient response of the output voltage. Zero voltage switching (ZVS) technique can be used to reduce the capacitive switching loss and the body-diode conduction losses in the power switches. This would increase the efficiency of buck converter, which in turn would improve the overall control system performance.

Bibliography

[Agarwal05] Agarwal, A., Mukhopadhyay, S., Kim, C.H., Raychowdhury, A., Roy, K. 2005. Leakage Power Analysis and Reduction: Models, Estimation and Tools. IEE Proceedings of Computers and Digital Techniques, Vol. 152, No. 3, (pp 353-368).

[Arnim05] Arnim, K.v., Borinski, E., Seegebrecht, P., Fiedler, H., Brederlow, R., Thewes, R., Berthold, J., Pacha, C. 2005. Efficiency of Body Biasing in 90-nm CMOS for Low-Power Digital Circuits. IEEE Journal of Solid-State Circuits, Vol. 40, No.7 (pp 1549-1556).

[Chandrakasan92] Chandrakasan, A.P., Sheng, S., Brodersen, R.W. 1992. Low-Power CMOS Digital Design. IEEE Journal of Solid-State Circuits, Vol.27, No.4 (pp 473-482).

[Kao02] Kao, J.T., Miyazaki, M., Chandrakasan, A.P. 2002. A 175-mV Multiply-Accumulate Unit Using an Adaptive Supply Voltage and Body Bias Architecture. IEEE Journal of Solid-State Circuits, Vol. 37, No. 11 (pp 1545-1554).

[Keshavarzi99] Keshavarzi, A., Narendra, S., Borkar, S., Hawkins, C., Roy, K., De, V. 1999. Technology Scaling Behavior of Optimum Reverse Body Bias for Standby Leakage Power Reduction in CMOS IC's. International Symposium on Low Power Electronics and Design, (pp 252-254).

[Kuang03] Kuang, W., Yuan, J.S. 2003. An Adaptive Supply-Voltage Scheme for Low Power Self-Timed CMOS Digital Design. Proceedings of IEEE 16th International Conference on VLSI Design. (pp 315-319).

[Kumar06] Kumar, S. V., Kim, C. H., Sapatnekar, S. S. 2006. Mathematically Assisted Adaptive Body Bias (ABB) for Temperature Compensation in Gigascale LSI Systems. Asia and South Pacific Conference on Design Automation.

[Kuroda98] Kuroda, T., Suzuki, K., Mita, S., Fujita, T., Yamane, F., Sano, F., Chiba, A., Watanabe, Y., Matsuda, K., Maeda, T., Sakurai, T., Furuyama, T. 1998. Variable Supply-Voltage Scheme for Low-Power High-Speed CMOS Digital Design. IEEE Journal of Solid-State Circuits, Vol. 33, No. 3 (pp 454-462).

[Lee04] Lee, D., Blaauw, D., Sylvester, D. 2004. Gate Oxide Leakage Current Analysis and Reduction for VLSI Circuits. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 12, No. 2, (pp 155-166).

[Li05] Li, F., He, L., Basile, J.M., Patel, R., Ramamurthy, H. 2005. Leakage Current Aware High-level Estimation for VLSI Circuits. IEE Proceedings of Computers and Digital Techniques, Vol. 152, No. 6, (pp 747-755).

[Long06] Long, C., Reddy, S., Pamarti, S., He, L., Karnik, T. 2006. Proceedings of International Symposium on Low Power Electronics and Design, (pp 326-329).

[Maneatis93] Maneatis, J.G., Horowitz, M.A. 1993. Precise Delay Generation Using Coupled Oscillators. IEEE Journal of Solid-State Circuits, Vol. 28, No.12 (pp 1273-1282).

[Maneatis96] Maneatis, J.G. 1996. Low-Jitter Process-Independent DLL and PLL Based on Self-Biased Techniques. IEEE Journal of Solid-State Circuits, Vol. 31, No.11 (pp 1723-1732).

[Martin02] Martin, S. M., Flautner, K., Mudge, T., Blaauw, D. 2002. Combined Dynamic Voltage Scaling and Adaptive Body Biasing for Low Power Microprocessors under Dynamic Workloads. IEEE/ACM International Conference on Computer Aided Design, (pp 721-725).

[Neau03] Neau, C., Roy, K. 2003. Optimal Body Bias Selection for Leakage Improvement and Process Compensation over Different Technology Generation. Proceedings of IEEE International Symposium on Low Power Electronics and Design, (pp 116-121).

[Nomura06] Nomura, M., Ikenaga, Y., Takeda, K., Nakazawa, Y., Aimota, Y., Hagihara, Y. 2006. Delay and Power Monitoring Schemes for Minimizing Power Consumption by Means of Supply and Threshold Voltage Control in Active and Standby Modes. IEEE Journal of Solid-State Circuits, Vol.41, No.4 (pp 805-814).

[Rabaey04] Rabaey, J. M., Chandrakasan, A., Nikolic, B. 2004. Digital Integrated Circuits - A Design Perspective, Pearson Education.

[Rahman07] Rahman, M.S. 2007. Buck Converter Design Issues. Master Thesis performed in *division of Electronic Devices*, Linköping Institute of Technology.

[Trescases04] Trescases, O., Ng, W.T. 2004. Variable Output, Soft-Switching DC/DC Converter for VLSI Dynamic Voltage Scaling Power Supply Applications. 35th Annual IEEE Power Electronics Specialists Conference, Aachen, Germany (pp 4149-4155).

[Wei99] Wei, G., Horowitz, M. 1999. A Fully Digital, Energy-Efficient, Adaptive Power-Supply Regulator. IEEE Journal of Solid-State Circuits, Vol. 34, No.4 (pp 520-528).

[Zhang05] Zhang, C. 2005. Techniques for Low Power Analog, Digital and Mixed Signal CMOS Integrated Circuit Design. PhD Dissertation, Louisiana State University.