Variability-Aware Design of Subthreshold Devices

by

Rodrigo Jaramillo Ramirez

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering.

Waterloo, Ontario, Canada, 2007

©Rodrigo Jaramillo Ramirez, 2007

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

Rodrigo Jaramillo Ramirez.

I understand that my thesis may be made electronically available to the public.

Rodrigo Jaramillo Ramirez.

Abstract

Over the last 10 years, digital subthreshold logic circuits have been developed for applications in the ultra-low power design domain, where performance is not the priority. Recently, devices optimized for subthreshold operation have been introduced as potential construction blocks. However, for these devices, a strong sensitivity to process variations is expected due to the exponential relationship of the subthreshold drive current and the threshold voltage. In this thesis, a yield optimization technique is proposed to suppress the variability of a device optimized for subthreshold operation. The goal of this technique is to construct and inscribe a maximum yield cube in the 3-D feasible region composed of oxide thickness, gate length, and channel doping concentration. The center of this cube is chosen as the maximum yield design point with the highest immunity against variations. By using the technique, a transistor is optimized for subthreshold operation in terms of the desired total leakage current and intrinsic delay bounds. To develop the concept of the technique, sample devices are designed for $90nm$ and $65nm$ technologies. Monte Carlo simulations verify the accuracy of the technique for meeting power and delay constraints under technology-specific variances of the design parameters of the device.

Acknowledgements

First of all, I thank God for giving me his guidance and the great family I have. I express my deepest gratitude to my parents, Margarita and Fernando. They have supported me, as always, with their love, advice, and generosity. To them I owe my values and achievements. Chavitos, gracias totales.

I am grateful to Norma Serna, whose love and support during the last two years have inspired me and lighted up this process. Specially, for her confidence and patient all along the difficult times we have gotten over.

This research project would not have been possible without the support of many people in the University of Waterloo. I am deeply grateful to my advisor, Dr. Mohab Anis, as when I started my M.A.Sc. I was a completely beginner in VLSI research, but he gave me the opportunity and honor to have a position in his group. His guidance and encouragement throughout these two years are of paramount importance in my graduate studies. I would also like to thank Javid Jaffari for being a good friend, a coworker, and specially a second mentor of this project, and to all my colleagues in the VLSI Research Group for their friendship and useful technical support. In addition, thanks to Professors Manoj Sachdev and John Hamel, members of my committee, for their confidence and valuable suggestions.

Also I would like to say a special thanks to Jazmin Romero for her advice and support. Thanks to her friendship my stay in Waterloo has been more pleasant and unforgettable.

An important person in this process has been Gerardo Leyva, he was the mastermind who suggested and encouraged me to study abroad. Thanks for being my teacher and excellent friend.

This research work has been supported by the National Council of Science and Technology (Consejo Nacional de Ciencia y Tecnologia, CONACyT, Mexico). This support is greatly appreciated.

Contents

1 Introduction **1**

1.1 Contribution of this Work 3

1.2 Organization of the Dissertation 4

2 Subthreshold Design: Background and Related Work **5**

2.1 Power Consumption 7

2.1.1 Static Power: Leakage Mechanisms 9

2.1.2 Technology Scaling and Leakage Power 14

2.2 MOS Transistor in the Subthreshold Region 19

2.3 Subthreshold Circuits 23

2.3.1 Origin and Evolution of Subthreshold Circuits 25

2.4 Process Variations 26

2.4.1 Process Variations in Subthreshold Designs 28

2.5 Devices in Subthreshold Circuits 30

2.6 Summary 32

3 Problem Definition and Yield Maximization Technique **33**

3.1 Design Parameters and Variations 33

3.1.1 Device Structure and Constraints 34

3.1.2 Device Parameters and Variations 38
3.2 Problem Statement and Yield Maximization Technique 43
3.2.1 General Problem Statement 43
3.2.2 Qualitative Approach 44
3.2.3 Formal Problem Statement 45
3.3 Summary 50

4 Results and Discussion 51
4.1 Implementation 51
4.2 Results and Discussion 52
4.3 Optimizing a Subthreshold Transistor 62
4.4 Summary 62

5 Conclusion and Future Work 64

List of Tables

2.1 CE scaling of MOSFET device and circuit parameters. 17

4.1 Desired device constraints and their technology node for optimized transistors. 53

4.2 Design parameters and specifications of optimum devices. 54

4.3 Various worst case leakage components of optimum devices. 56

List of Figures

2.1 Architecture of an n-channel MOSFET 6

2.2 Projected leakage power as a fraction of the total power consumption according to the ITRS. 8

2.3 Short-channel transistor mechanisms: (I_1), reverse biased p-n junction; (I_2), subthreshold or weak inversion; (I_3), drain-induced barrier lowering; (I_4), punch-through; (I_5), gate-induced drain leakage; (I_6), gate oxide tunneling; and (I_7), hot-carrier injection. 9

2.4 Band-to-band tunneling in an nMOS: valence band electron tunneling from the valence band of the p-side to the conduction band of the n-side; the total voltage drop across the junction, the reverse bias voltage and the built-in voltage is greater than the energy-band gap, $(V_{app} + \psi_{bi} > E_g)$. 10

2.5 Normalized channel length from the source to the drain versus the surface potential: *Curve A*: a long-channel MOSFET, *Curve B*: a short-channel MOSFET at a low drain bias, and *Curve C*: a short-channel MOSFET at a high drain bias; the gate voltage is constant for the three cases. 11

2.6 Tunneling of electrons: direct tunneling occurs when the potential drop across the gate oxide is lower than the barrier height of the tunneling electron $(V_{ox} < \phi_{ox})$. 13

2.7 Injection of hot electrons from the substrate to the oxide. 13
2.8 Leakage components of $0.35\mu m$ technology for a $20\mu m$ wide transistor; currents for various leakage mechanisms are accumulated for a given drain bias. 14
2.9 Variation of leakage components: (a) oxide thickness and channel length, and (b) doping profile in a 50nm device; "Doping-1" has a stronger halo profile than "Doping-2". 15
2.10 MOS constant electric field scaling by a factor of S. 16
2.11 Power breakdown trend over technology generations. 18
2.12 Drift and diffusion components of the total current, represented by the solid curve. 20
2.13 Capacitance in weak inversion. 21
2.14 Percentage of the total variations accounted for intra-die variation for different technology generations. 28

3.1 Architecture of the subthreshold device. 34
3.2 Total leakage (TL) estimation scheme. 37
3.3 Charge sharing model. 39
3.4 Top view of the transistor. 41
3.5 Simplified problem in 2-D. 45
3.6 Flowchart of the yield maximization technique. For each iteration, a set of design values are determined by the optimization engine, and directly evaluated by MEDICI 49

4.1 %Yield obtained from experimental $90nm$ devices. 56
4.2 Mean and standard deviation of total leakage. The bars correspond to mean values whereas the dotted vertical lines are the standard deviations (range of device variation). 57

4.3 Mean and standard deviation of intrinsic delay. The bars correspond to mean values whereas the dotted vertical lines are the standard deviations (range of device variation). 58
4.4 Monte Carlo simulation of the $90_{300,5.0}$ device, 98.9% of devices satisfy both constraints (relaxed TL bound). 59
4.5 Monte Carlo simulation of the $90_{300,2.5}$ device, 84.0% of devices satisfy both constraints (tight TL bound). 60
4.6 I-V characteristics of the $90_{300,5.0}$ device (more relaxed constraints, yield = 98.9%). 61
4.7 I-V characteristics of the $90_{300,2.5}$ device (tighter constraints, yield = 84.0%). 61

Chapter 1

Introduction

With the growing demand for system portability, power consumption has become an issue in VLSI applications such as biomedical devices, self-powered Radio Frequency IDentification (RFID), and wireless sensor networks. The progress of mobile electronics will depend on the development of inexpensive devices with complex functionality and long battery life [25]. In turn, the continuous increase in system necessities demands higher levels of integration and performance which have been solved by highly scaled CMOS technologies. However, MOS transistor scaling has resulted in unacceptable off-state leakage currents beyond 130nm process technologies [66]. Moreover, since the fabrication process tolerances have not scaled proportionally with the device dimensions, the relative impact of process variations has become more significant with each technology generation (especially beyond the 90nm technology) [67]. These two issues, power and variability, are the most challenging obstacles for modern IC design, according to the International Technology Roadmap for Semiconductors (ITRS) [7]. It is evident that new strategies for designing low-power systems under effects of parameter variations are crucial needed.

In recent years, dramatically scaled CMOS technologies have boosted the mass produc-

tion of portable battery-powered electronic devices, and an immense interest in energy-efficient designs. Nowadays, low-power IC design is an especially vibrant area of research and development, resulting in advances in low-power fabrication processes and energy-sensitive design techniques [24]. It has been proven that, in a CMOS circuit, the minimum energy operation occurs in the subthreshold region (or weak inversion) [12, 73]; thus, *subthreshold logic* [58, 59, 60] has emerged as a compelling approach to design energy-efficient systems.

Subthreshold logic operates solely in the weak inversion region of transistors, that is, the power supply voltage is below the threshold voltage ($V_{dd} < V_{th}$) such that load capacitances are charged/discharged by subthreshold leakage current. On the one hand, driving CMOS circuits with the subthreshold leakage current can provide orders of magnitude power reduction over standard strong inversion (superthreshold) CMOS circuits [46]. On the other hand, the minute operating leakage current is orders of magnitude lower than the saturation drain currents in the strong inversion regime, and there are significant limits on the maximum performance of subthreshold circuits [73]. By pursuing subthreshold design, it is expected that an energy efficiency in the range of 1pJ/instruction can be achieved [44], enabling applications which require ultra-low-power dissipation and low-to-moderate circuit performance. Some successful applications include an ultra-low-power adaptive filter for hearing aid devices [31], a 180-mV subthreshold FFT processor [71], a 256k subthreshold SRAM [13], and a 2.60pJ/instruction subthreshold sensor processor [78].

Researchers have demonstrated that it is possible to implement subthreshold logic circuits by using standard CMOS transistors, however; such devices are not optimum in terms of power and performance in the subthreshold regime. Indeed, a device should be specifically optimized for subthreshold operation to improve the device's Power-Delay Product (PDP) as shown by Paul et al. [46]. They have introduced an optimized transistor for sub-

threshold operation that improves the PDP in comparison to that of a standard transistor in the subthreshold region. However, authors have not accounted for process variations. It is well known that subthreshold current is exponentially dependent on the transistor's threshold voltage, where V_{th} is strongly related to various device parameters; they in turn vary considerably in the Deep Sub-Micron (DSM) regime [64]. Therefore, subthreshold designs are expected to be prone to process variations [77]. Consequently, this dissertation focuses on a device optimization technique which includes process variations on subthreshold transistor design.

1.1 Contribution of this Work

The objective of this investigation is to provide an automatic optimization technique to design a transistor for subthreshold operation that addresses process variations and device yield, power and performance constraint satisfaction, in a simple framework. To the best of our knowledge, this is the first variability-aware design approach at the device level for subthreshold devices.

By using the technique, a transistor can be optimized for subthreshold operation in terms of the desired total leakage current and intrinsic delay bounds, taking into account device design-parameter variations. In addition, designers can still apply circuit and architecture level techniques to further mitigate the variation effects.

The design technique exploits the 3-D space generated by device's design parameters oxide thickness, gate length, and channel doping concentration, to maximize the device yield in the presence of variations in the three parameters; that is, a feasible region is constructed and bound by the desired total leakage current and intrinsic device delay. As a result, any point in the feasible region represents a satisfactory device under the stated power and delay restrictions. Finally, a cube is formed and inscribed in the region such

that the center of the cube represents the properties of a device that is the most robust to process variations.

1.2 Organization of the Dissertation

The remainder of this thesis is organized as follows.

Chapter 2 provides an overview of the subthreshold regime. After the basic characteristics of MOS transistors in the subthreshold region are summarized, subthreshold circuits, and the state-of-the-art of subthreshold designs are discussed with the emphasis on works, addressing process variations. Finally, the considerations for constructing a subthreshold optimized transistor are outlined.

Chapter 3 defines the problem and develops the yield maximization technique. An analytical discussion of device parameters and variations is presented. Subsequently, the problem is defined in terms of the selected parameters and constraints, and the yield maximization technique is formalized.

Chapter 4 describes the implementation, results, and discussion. Sample devices for two technologies are optimized. The experimental results are assessed by examining several characteristics of the new subthreshold optimized devices. Finally, guideline is introduced to optimize transistors.

Chapter 5 not only concludes this thesis but also outlines future work for subthreshold designs.

Chapter 2

Subthreshold Design: Background and Related Work

This chapter introduces the problem of power consumption in modern DSM devices. Subsequently, a description of a transistor's leakage mechanisms is given, and the obstacle to technology scaling, by exponential leakage power increases, is exposed. Later, the paradigms of subthreshold design at several levels of hierarchy are presented. First, the characteristics and behavior of MOS transistors in the subthreshold region are described to comprehend subthreshold logic. In addition, the properties and evolution of subthreshold designs, including the state-of-the-art are presented. Then, the challenges of process variations and its impact on circuit behavior are addressed with an emphasis on research in the subthreshold regime. Lastly, considerations for construct a subthreshold optimized transistor are discussed.

In this thesis, the object of discussion, analysis, and optimization is the Metal-Oxide-Semiconductor Field-Effect (MOSFET, or MOS for simplicity) transistor. Such a transistor is the dominant device in integrated circuits such as processors and memories. The tran-

sistor's current is transported by electrons in n-channel devices (nMOS) or by holes in p-channel devices (pMOS). A basic nMOS channel structure is depicted in Figure 2.1, the substrate (bulk or body) is composed of p-type silicon in which two heavily doped n-type silicon regions, the drain and the source, are formed. Typically, the gate consists of heavily doped or salicide polysilicon, and is separated from the substrate by a thin silicon dioxide film, the gate oxide. The main device parameters are gate oxide insulator thickness (T_{ox}), physical gate length (L_g), channel doping concentration (N_{ch}), source/drain junction depth (Y_j), and transistor width (W).

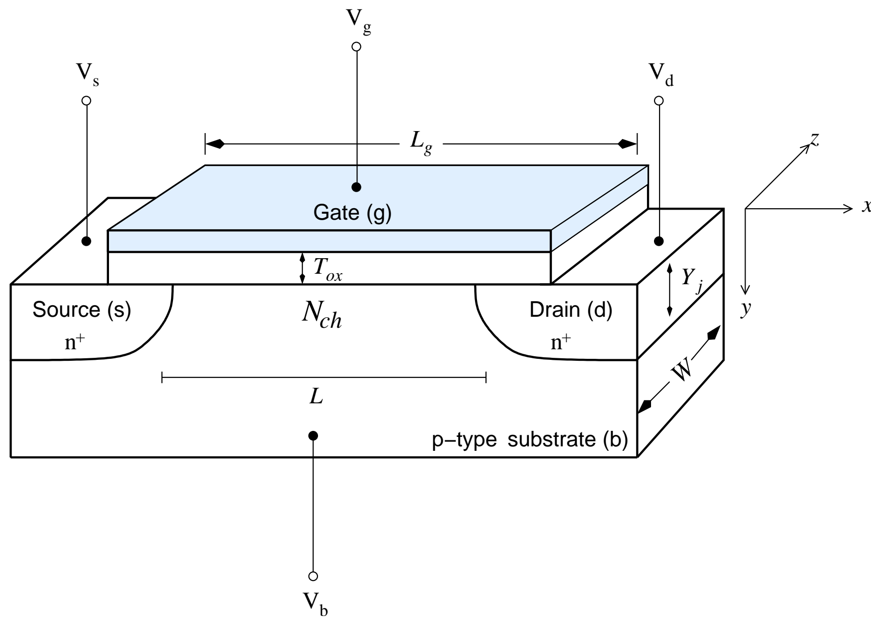

Figure 2.1: Architecture of an n-channel MOSFET

2.1 Power Consumption

The power consumption of a system sets up how much energy is consumed per operation, and how much heat is dissipated. The upper power limits determine the maximum number of transistors that are integrated on a single chip, heat removal system, chip package, and especially, the frequency at which the transistor switches [50]. In digital CMOS circuits, the power consumption is composed of two components: dynamic power due to the active switching activity, and static power due to the leakage current. These two sources of power consumption are represented by

$$P_{total} = P_{dynamic} + P_{static} = \alpha C_L V_{dd}^2 f_{clk} + V_{dd} I_{leak}. \tag{2.1}$$

To reduce the dynamic power component, the first target has been aggressive supply voltage (V_{dd}) scaling, since it leads to quadratic power reductions; the switching activity (α), operating frequency (f_{clk}), and load capacitance (C_L) reductions provide linear decreases in the dynamic power. For the static power component, besides the supply voltage scaling which provides linear power reductions, the objective is to keep the leakage current (I_{leak}) as low as possible . Leakage power exponentially increases for each technology node, and eventually becomes the dominant component of the total power (P_{total}) as the technology scales beyond $65nm$ [9, 29], as seen Figure 2.2 [7]. This is why low-power strategies, especially those for static power reduction, are necessary at almost any design level in recent digital circuits.

The power dissipation of high performance applications such as microprocessors, digital signal processors, and random access memories has increased along with the progress in CMOS technologies, where the design emphasis has been on maximizing the operational frequency (f_{clk} in (2.1)). The increased power consumption raises a chip temperature which leads to electromigration reliability problems, and degradation in the device perfor-

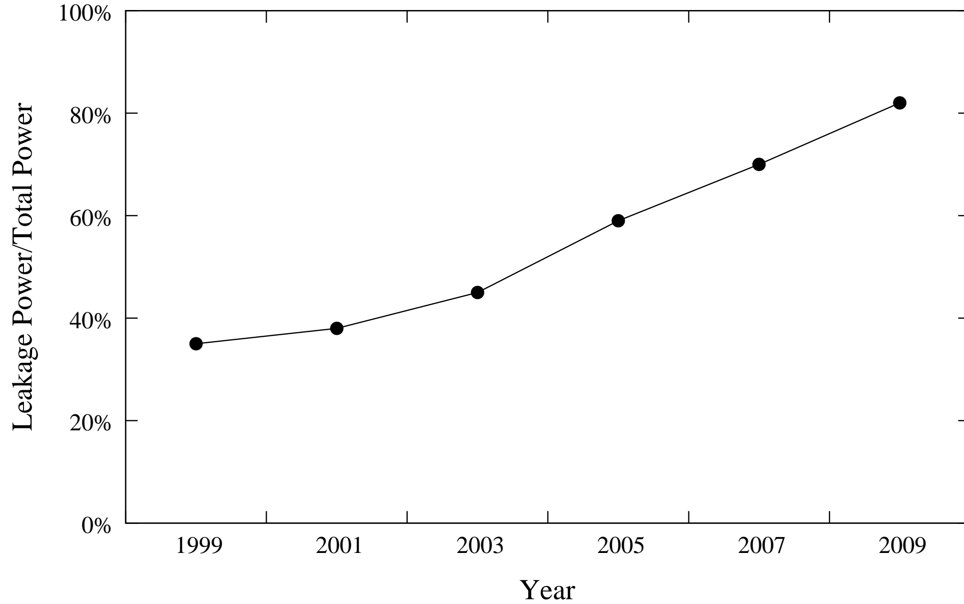

Figure 2.2: Projected leakage power as a fraction of the total power consumption according to the ITRS.

mance. Thus, lowering the power dissipation is crucial for high performance VLSI designs [34]. Also, applications are emerging for which the energy consumption is the key metric, and the speed of operation becomes less relevant. Generally, energy-constrained VLSI applications such as micro-sensor networks and nodes, radio frequency identification, and biomedical devices have low activity rates and low speed requirements (α, and f_{clk} in (2.1), respectively); but the concern is to lengthen battery life. Ideally, the power consumption of these systems should decrease to the extent that they can harvest energy from environmental resources such as solar power, thermal gradients, radio-frequency, and mechanical vibration [17], and theoretically have unlimited lifetimes. Such ultra-low-power applications have established a significant niche for subthreshold circuits [70].

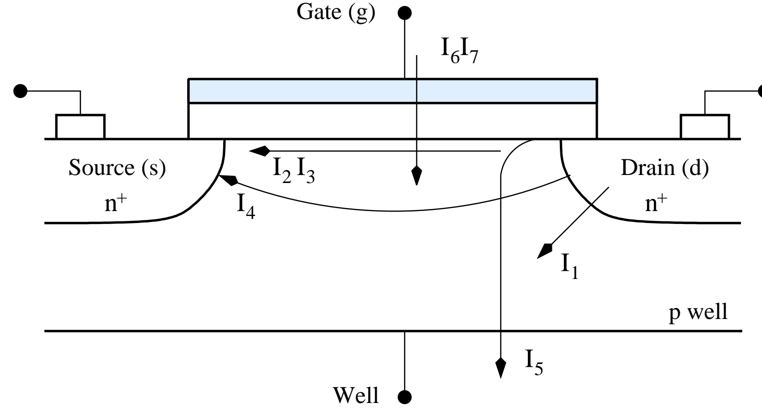

Figure 2.3: Short-channel transistor mechanisms: (I_1), reverse biased p-n junction; (I_2), subthreshold or weak inversion; (I_3), drain-induced barrier lowering; (I_4), punch-through; (I_5), gate-induced drain leakage; (I_6), gate oxide tunneling; and (I_7), hot-carrier injection.

2.1.1 Static Power: Leakage Mechanisms

Static power is dissipated during the idle time, that is, when no transition or switching activity occurs. As the transistor threshold voltage, channel length, and gate oxide thickness are reduced in DSM regimes, the static power dissipation becomes a challenging obstacle for the development of modern ICs. Consequently, the identification of the different leakage components is pivotal for the analysis and design of low-power applications. Figure 2.3 [57] denotes the seven transistor intrinsic leakage mechanisms in short channel devices.

- I_1 is the **reverse bias p-n junction leakage**. Drain-to-substrate and source-to-substrate junctions are normally reverse biased, occasioning a p-n junction leakage current. It has two components: i) the minority carrier diffusion/drift near the edge of the depletion region, and ii) the electron-hole pair generation in the depletion region of the reverse biased junction. If both the n- and p-regions are heavily doped, which is the case of an advanced MOS, to mitigate short-channel effects, Band-To-

Band Tunneling (BTBT) can also be present. The effect dominates the p-n junction leakage component. I_{BTBT} occurs when a high electric field ($> 10^6 V/cm$), across the reverse biased junction, leads to electrons from the valence band of the p-side to migrate to the conduction band of the n-side, as denoted in Figure 2.4 [65].

- I_2 is the **weak inversion or subthreshold conduction current** between the source and drain. It occurs when the gate voltage is below the threshold voltage ($V_g < V_{th}$). Recently, this current dominates device off-state leakage mechanisms due to the low V_{th} values of transistors [5]. This weak inversion current is the *drive current* in the subthreshold regime. Consequently, this leakage component is looked at the next section.

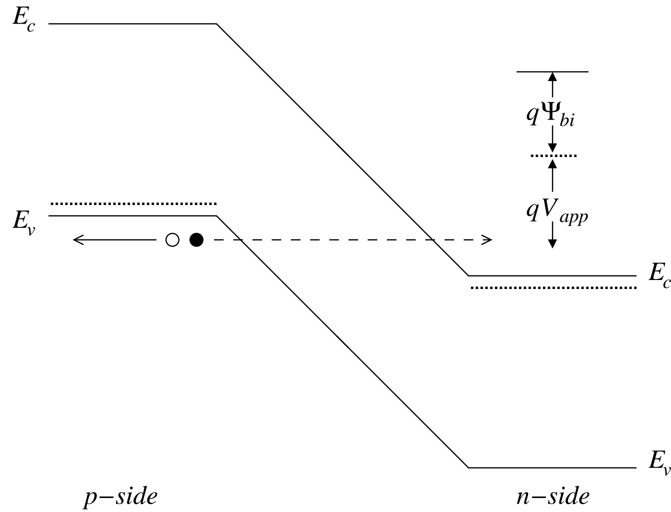

Figure 2.4: Band-to-band tunneling in an nMOS: valence band electron tunneling from the valence band of the p-side to the conduction band of the n-side; the total voltage drop across the junction, the reverse bias voltage and the built-in voltage is greater than the energy-band gap, ($V_{app} + \psi_{bi} > E_g$).

- I_3 is the **Drain-Induced Barrier Lowering** (DIBL). It occurs when a high drain voltage is applied to a short-channel device, and thus, the potential (voltage) barrier (to the electrons for an nMOS) at the surface between the source and drain is lowered. For example, consider the potential energy barrier at the surface between the drain and source, depicted in Figure 2.5 [65]. At the off-condition, this potential prevents the flow of electrons between the terminals. However, as the drain voltage is increased, the potential barrier is reduced in short-channel devices. In this way,

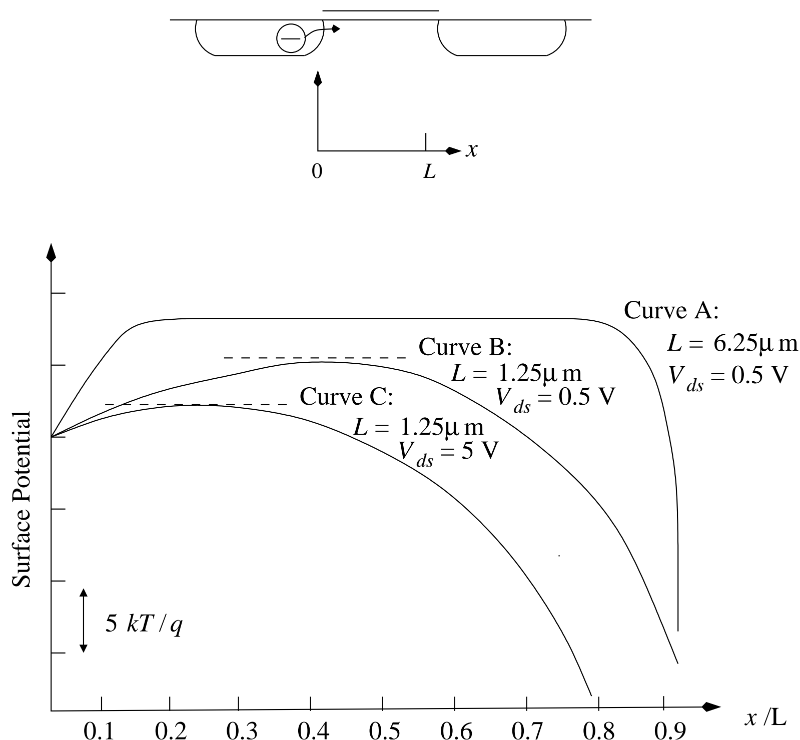

Figure 2.5: Normalized channel length from the source to the drain versus the surface potential: *Curve A*: a long-channel MOSFET, *Curve B*: a short-channel MOSFET at a low drain bias, and *Curve C*: a short-channel MOSFET at a high drain bias; the gate voltage is constant for the three cases.

the higher the drain voltage applied to a short channel device, the lower the barrier height is, and thus, the source injects carriers into the channel surface without the control-gate voltage playing a role [65].

- I_4 is the **channel punch-through**. At even higher drain voltages and channel length reductions, the drain and source depletion regions approach each other and eventually merge in the deep substrate. As a result, the gate totally loses control over the channel, and the flow of the drain current becomes independent of the control voltage [65].

- I_5 represents the **Gate-Induced Drain Leakage** (GIDL). It is the result of the influence of high electric fields on the gate-drain overlap region. Consequently, the depletion width of the drain to substrate p-n junction is thinned out [53]. Carriers are generated in the substrate and drain from the direct band-to-band tunneling, trap-assisted tunneling, or a combination of thermal emission and tunneling [5]. Oxide thickness (T_{ox}) reductions and higher supply voltages lead to a higher potential between the gate and the drain, which in turn, enhances the electric field dependent GIDL.

- I_6 refers to **oxide leakage tunneling**. The continuous reduction of the oxide thickness leads to an increase in the field across T_{ox}. The high electric field results in the tunneling of electrons from the inverted substrate-to-gate and also from the gate-to-substrate through T_{ox}. This current flow is known as oxide leakage tunneling. The direct tunneling of electrons is signified in Figure 2.6 [54, 56].

- I_7 is the gate current due to **hot carrier injection**. If a region with a high electric field is located near the Si-SiO_2 interface (as it occurs in the pinch-off condition),

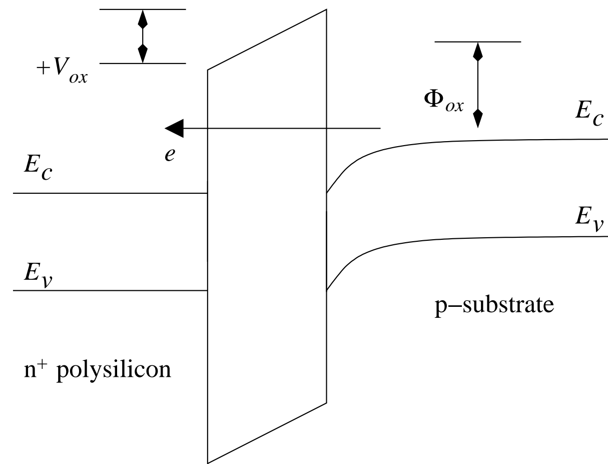

Figure 2.6: Tunneling of electrons: direct tunneling occurs when the potential drop across the gate oxide is lower than the barrier height of the tunneling electron ($V_{ox} < \phi_{ox}$).

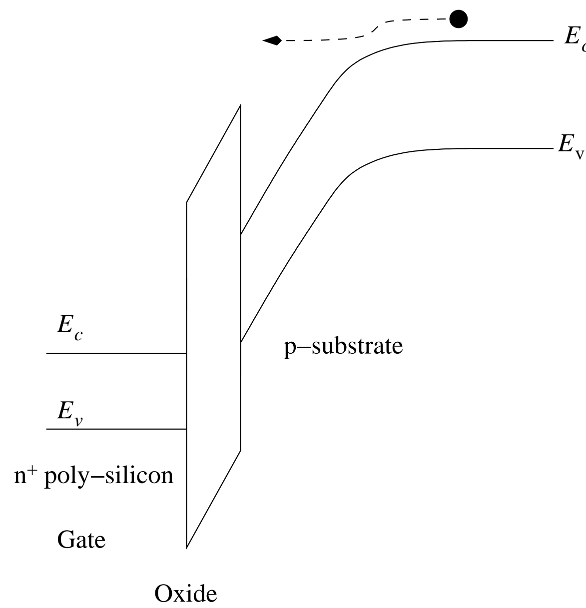

Figure 2.7: Injection of hot electrons from the substrate to the oxide.

some of the electrons or holes can gain sufficient energy from the field to cross the interface potential barrier and enter the oxide layer. This phenomenon, called a hot-carrier injection, is represented in Figure 2.7 [65].

Currents I_1 to I_5 are off-state leakage mechanisms; I_6 is present when the transistor is in the on-state. Finally, I_7 can occur in the off-state, but is more typical during the transitions of the transistor bias states [5]. Figure 2.8 [30] summarizes the relative contributions of the leakage components for $0.35\mu m$ CMOS technology.

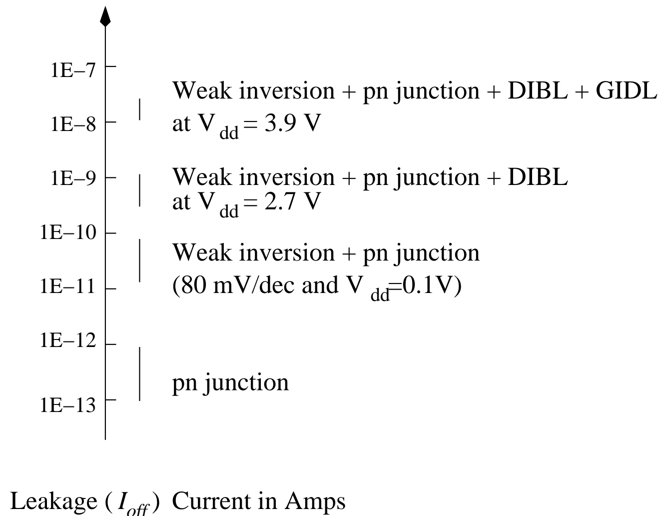

Figure 2.8: Leakage components of $0.35\mu m$ technology for a $20\mu m$ wide transistor; currents for various leakage mechanisms are accumulated for a given drain bias.

2.1.2 Technology Scaling and Leakage Power

In the last three decades, the CMOS technology evolution has resulted in substantial device scaling for achieving density, speed, and power improvements, as predicted by Moore's Law [40]. The direct result of device scaling is the reduced intrinsic capacitance, enabling a faster

switch. Simultaneously, the power supply voltage scaling has reduced the switching energy. To maintain the speed enhancement for each technology node, threshold voltage V_{th} must also scale down in order to retain enough gate overdrive V_{dd}/V_{th}. However, reducing the V_{th} results in an exponential increase in the subthreshold leakage current, as shown by (2.5) in the next section. The oxide thickness scaling, required to maintain reasonable short channel effects, results in a considerable amount of direct oxide tunneling leakage current. Finally, the higher substrate doping density and the application of the halo profiles to reduce short channel effects in scaled devices, cause substantially large junction band-to-band tunneling leakage. In this way, among the seven leakage mechanisms in scaled devices, the three more contributive components of leakage power are: subthreshold, oxide tunneling, and reverse bias p-n junction BTBT [52]. The magnitudes of each component depend strongly on the device constitution, that is, oxide thickness, channel length, and doping profile in Figure 2.9 [41].

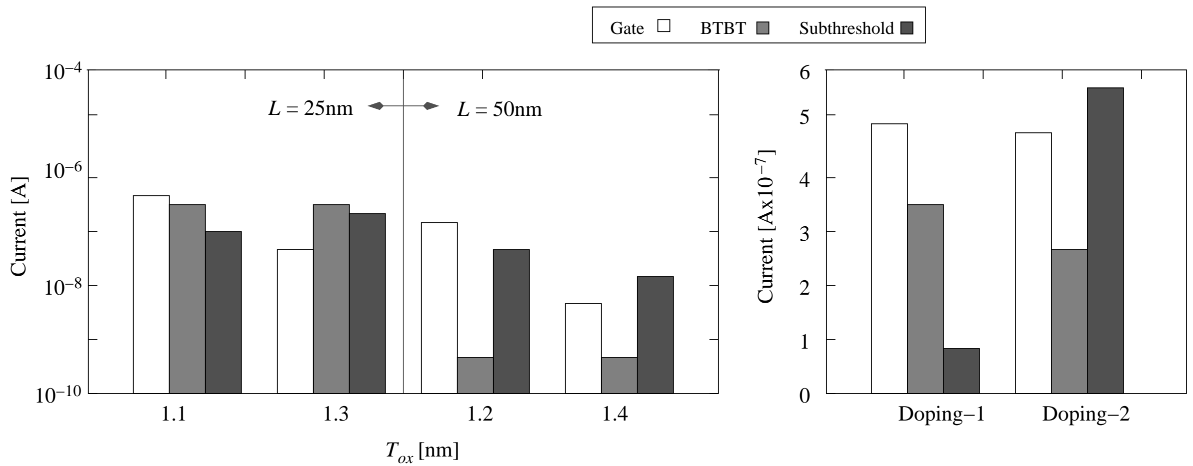

Figure 2.9: Variation of leakage components: (a) oxide thickness and channel length, and (b) doping profile in a 50nm device; “Doping-1” has a stronger halo profile than “Doping-2”.

There are two MOS scaling theories: Constant Electric (CE) field scaling [19] and Constant Voltage (CV) scaling [18]. In CE scaling, it has been proposed that for keep the short channel effect under control, the horizontal and vertical dimensions should be scaled down. In addition, the applied voltage should be decreased and the substrate doping concentration should be increased proportionally, as summarized in Figure 2.10.

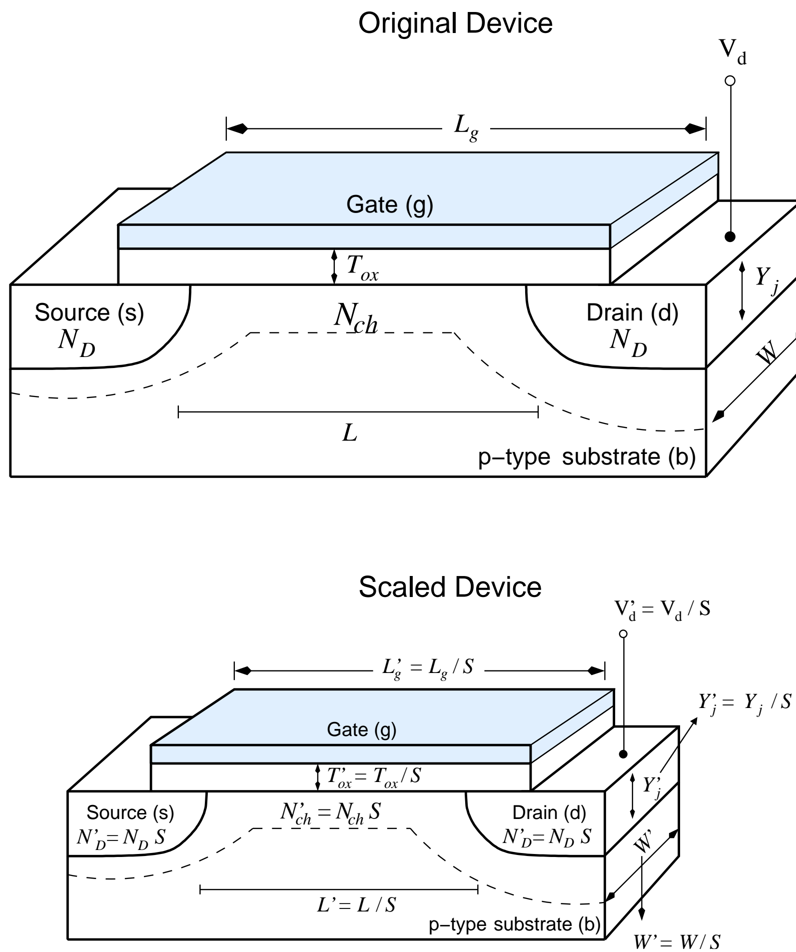

Figure 2.10: MOS constant electric field scaling by a factor of S.

The CV theory retains the same scaling down theory as the CE, but the power supply voltage remains unchanged. Therefore, to maintain the charge-field, the doping densities are increased by a factor of S^2 [18]. However, the CV scaling inherently leads to a continuous increase of the transistor's internal electric field, which can cause reliability problems such as electron migration, hot carrier degradation, and oxide breakdown in recent CMOS process generations. As a result, since the $0.5\mu m$ MOS technology, the industry's scaling methodology has been the CE [23].

Table 2.1: CE scaling of MOSFET device and circuit parameters.

Parameter	1/S Constant Field Scaling	30% Scaling Field Scaling
Physical Device Dimension	$1/S$	0.7
Supply and threshold voltage	$1/S$	0.7
$C_{ox} = (\epsilon Area)/T_{ox}$	$1/S$	0.7
Gate Capacitance $= WL/T_{ox}$	$1/S$	0.7
Current $= (W/L)(1/T_{ox}V_{dd}^2)$	$1/S$	0.7
Propagation Delay $= CV_{dd}/I$	$1/S$	0.7
Frequency	S	1.43
Dynamic Power $= CV_{dd}^2 f_{clk}$	$1/S^2$	0.5
Leakage Power	Exponential	Exponential
Energy	$1/S^3$	0.34

The principle of the CE theory is that the physical dimensions (gate length, transistor width, and oxide thickness), and voltages (power supply and threshold) are scaled by a factor of 1/S ($S > 1$). Consequently, the current, gate capacitance, and propagation delay

also scale by a factor of 1/S. Hence, with a 30% reduction of all the device parameters (1/S = 0.7), improvements, close to 50% in the operation frequency, are achieved for each generation. Table 2.1 lists the CE scaling rules for various device parameters and circuit performance factors. Evidenlty, the resulting switching energy scales by $1/S^3$, whereas the dynamic power scales by $1/S^2$. However, for a constant die size, the power dissipation due to the dynamic switching currents remains relatively constant with scaling, because the number of switching elements for the same die size increases by a factor of S^2. The negative effect of the CE is that the subthreshold current increases exponentially, as the threshold voltage scales. For example, consider a device with a V_{th} of 400mV that is to be scaled by 0.7. For a constant die size, scaling provides around a 43% ($S = 1.43$) frequency improvement, and doubles the number of devices. The dynamic power dissipation scales by unity, but the leakage current increases by a factor of $1.43 \times 10^{(V_{th}/S(1-0.7))} = 45$ [28]. Figure 2.11 [55] demonstrates the power breakdown composition for Intel's process technologies. Leakage power in the $0.25\mu m$ technology corresponds to 0.1% of the active power, but

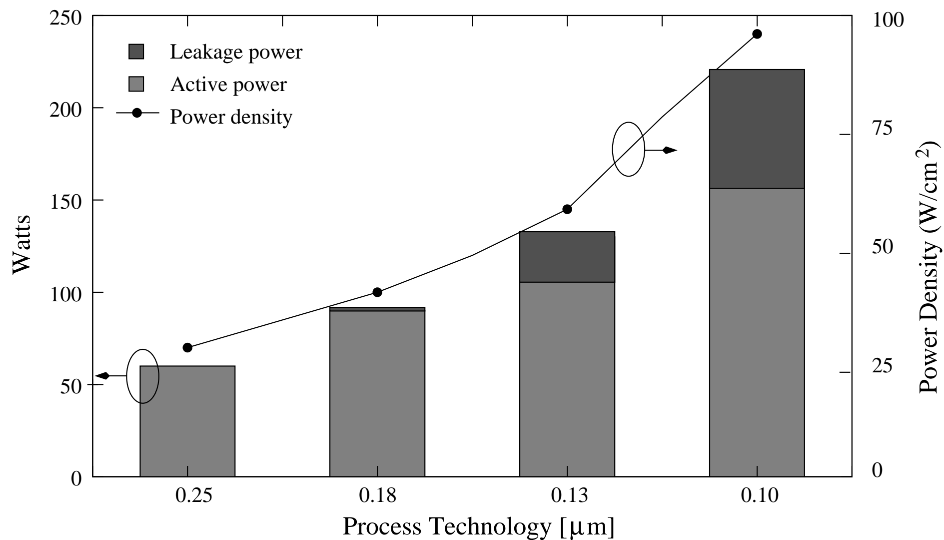

Figure 2.11: Power breakdown trend over technology generations.

dramatically increases to approximately 25% of the active power in $0.1\mu m$ technology.

Since the primary reasons for technology scaling are performance and device integration, the CE scaling theory serves as an essential blueprint. However, as the device's dimensions and voltages are shrunk, the leakage power becomes a significant barrier in present and future technologies. Therefore, the scaling of VLSI technology leads to multiple challenges, including power dissipation, leakage management, and short channel effects. This problem worsens for portable energy-constrained VLSI applications, where battery power is drained needlessly during long idle periods [3].

In Subsection 2.1.2, the three contributive leakage components in recent sub-micron technologies are the subthreshold, oxide tunneling, and reverse bias p-n junction BTBT, whereas subthreshold leakage is the *drive current* in the subthreshold regime. Thus, consideration of this component is developed in the next section.

2.2 MOS Transistor in the Subthreshold Region

In this section, the behavior of a MOS transistor in the subthreshold or weak inversion region is examined. As the gate voltage drops below V_{th}, the strong inversion model erroneously predicts a drain current; thus, a model for the subthreshold region is necessary. The vital parameters for the subthreshold regime include subthreshold swing coefficient and subthreshold slope.

A current flow in silicon is carried out by two mechanisms: the drift of carriers caused by the presence of an electric field, and diffusion of carriers which is caused by a gradient concentration of electrons or holes. In weak inversion, the channel inversion layer charge is much lower than that of the substrate or depletion charge ($Q_I << Q_B$). Since the substrate is weakly doped, Q_B is small, and the electric field along the channel direction is not enough to pull the electrons from the source to the drain; as a result, the drift current is

negligible [65]. Thus, unlike the strong inversion region where the drift current dominates, in the subthreshold region, the diffusion current is the major component, as portrayed in Figure 2.12.

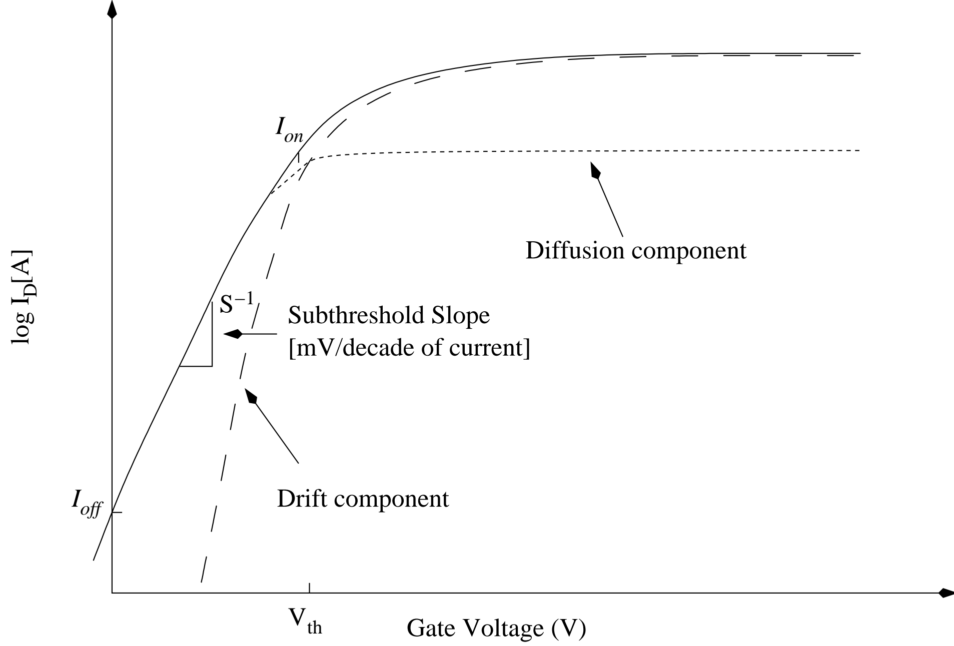

Figure 2.12: Drift and diffusion components of the total current, represented by the solid curve.

To obtain an expression for the subthreshold diffusion current, the surface channel potential in weak inversion is explored. The inversion charge in weak inversion is negligible, such that the gate-to-bulk capacitance is principally influenced by the oxide capacitance (C_{ox}) and the depletion capacitance (C_{dep}), as signified in Figure 2.13.

Assuming that $V_b = 0$, the relationship between the surface potential (ψ_s) and the applied gate voltage (V_g) is the capacitive divider

$$\psi_s = \kappa V_g = \frac{C_{ox}}{C_{ox} + C_{dep}} V_g, \tag{2.2}$$

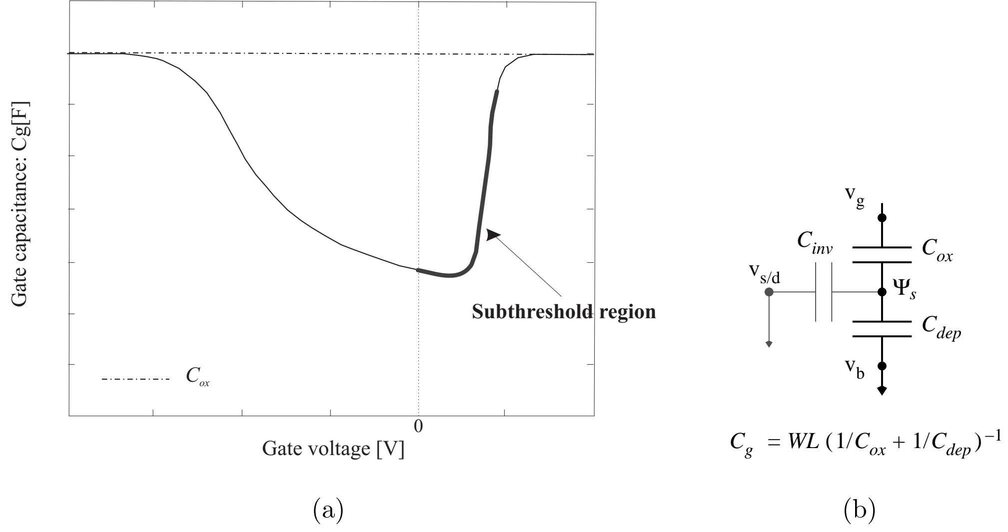

Figure 2.13: Capacitance in weak inversion.

where κ is the coupling coefficient of the gate voltage to the surface potential. At the surface level, the charge concentration in the source ($x = 0$) and the drain ($x = L$) are given by

$$|Q'_{I0}| \propto exp\left(\frac{V_s - \kappa V_g}{V_T}\right),$$
$$|Q'_{IL}| \propto exp\left(\frac{V_d - \kappa V_g}{V_T}\right), \tag{2.3}$$

$V_T = kT/q$ is the thermal voltage. The concentration gradient of electrons in the channel decreases linearly from $x = 0$ to $x = L$ (i.e., the concentration gradient is constant). Then, the diffusion current is written as

$$I_d = -WD_n\frac{Q'_{I0} - Q'_{IL}}{L} = -\frac{W}{L}\mu_0 V_T(Q'_{I0} - Q'_{IL}), \tag{2.4}$$

$\frac{W}{L}$ is the width over length ratio of the device, and μ_0 is the zero bias electron mobility in

the channel. This leads to a compact expression for the subthreshold current, $I_{sub} = I_{ds}$ as follows:

$$I_{sub} = I_o e^{(V_{gs}-V_{th})/nV_T} \left(1 - e^{-V_{ds}/V_T}\right), \tag{2.5}$$

where $I_o = \mu_0 C_{ox} \frac{W}{L} V_T^2$, and n is the subthreshold swing coefficient, defined as $n = 1/\kappa = 1 + C_{dep}/C_{ox}$. In subthreshold logic, the drive current (I_{on}) is precisely the subthreshold current modeled by (2.5) as $I_{on} = I_{sub}(V_{gs} = V_{ds} = V_{dd} < V_{th})$. Also, the transistor "off" state current (I_{off}) is the drain current, when the gate voltage is zero; that is $I_{off} = I_{sub}(V_{gs} = 0, V_{ds} = V_{dd} < V_{th})$. Throughout this thesis, I_{on} and I_{off} refer to the latter definitions, unless otherwise specified. In addition, note that for $V_{ds} > 4V_T \approx 100mV$ at room temperature, the term ($e^{-V_{ds}/V_T} << 1$) such that I_{sub} saturates and is independent of the V_{ds} as follows:

$$I_{sub} = I_o e^{(V_{gs}-V_{th})/nV_T} \quad \text{for} \quad V_{ds} > 4V_T. \tag{2.6}$$

It is observed that the drain current changes exponentially with the V_{gs}, whereas in strong inversion, I_D responds quadratically with V_{gs}. These relationships are illustrated in the logarithmic plot of I_D vs. V_{gs} in Figure 2.12. From this figure, the dependence of the gate voltage swing needed to change the drain current by one order of magnitude is defined as [65]

$$S = \frac{d(\log_{10} I_d)}{dV_{gs}}^{-1} = 2.3\frac{nkT}{q} = 2.3\frac{kT}{q}\left(1 + \frac{C_{dep}}{C_{ox}}\right). \tag{2.7}$$

The subthreshold slope, S, is an important device parameter in the subthreshold region. The smaller the S value is, the higher the drive I_{on} current is, and thus, the faster the device.

This section introduces the characteristics of a MOS transistor in the subthreshold region that differ from those of the strong inversion in the following ways.

- Current Flow Mechanism: In the subthreshold region, the drain current is governed by diffusion, whereas in strong inversion it is mainly by drift phenomenon.
- Intrinsic MOS Capacitances: Like that of the subthreshold region, the inversion charge is negligible, the gate-to-bulk capacitance is given by the serial combination of C_{ox} and C_{dep}.
- Drain Current. It is exponentially related to the gate and threshold voltages, as well as temperature.

The characteristics and state-of-the-art of the corresponding subthreshold logic are discussed in the next section.

2.3 Subthreshold Circuits

In this section, the paradigms of subthreshold logic are briefly outlined. Subsequently, the evolution and the state-of-the-art of subthreshold designs are described.

Subthreshold logic operates completely in the subthreshold region; that is, the drain on and off currents are composed entirely of subthreshold leakage. Therefore, the logic assumes a power supply voltage that is less than the threshold voltage, $V_{dd} < V_{th}$. Since the leakage current is orders of magnitude lower than the drain strong inversion current and since the power supply is reduced, subthreshold logic dissipates ultra-low-power. Due to the small drive leakage current, the subthreshold logic only fits in designs, where the performance is considerably poor, and not the main concern.

As mentioned in the former of this section, subthreshold logic shares important properties with traditional strong inversion CMOS logic.

- High Noise Margins: The output swing goes from V_{dd} to ground (GND).
- Low Output Impedance: In the steady state, a low impedance path to either V_{dd} or GND exists.

In addition, subthreshold logic has a number of advantages over its strong inversion counterpart.

- Lower Power Consumption: At the same frequency, subthreshold circuits consume orders of magnitude less power than strong inversion circuits [59].
- Higher Gain. The exponential relationship between I_{sub} and V_{gs} leads to a high transconductance, $g_m = \frac{\partial I_{sub}}{\partial V_{gs}} = \frac{I_{sub}}{nV_T}$ [68].
- Better Noise Margins. In (2.6), I_{sub} readily becomes independent of V_{ds}. This near-ideal current source characteristic improves the noise margin of the logic gates [60].

A notable difference between strong inversion CMOS logic and subthreshold circuits is their robustness. Strong inversion logic will always work, given that the appropriate complementary Pull-Up Network (PUN) and Pull-Down Network (PDN) are implemented, even if the transistors are erroneously sized. In subthreshold circuits, transistor sizing impacts the functionality of CMOS circuits due to low supply voltages [16]. For example, consider a simple inverter, operating in the subthreshold region. Subthreshold I_{off} leakage always flow through a large pMOS device (which forms the PUN) to a certain extent where, a smaller nMOS (which forms the PDN) cannot pull down the voltage at the output to a full logic 0 level, and viceversa. This problem is augmented by the effect of the process variations.

Besides the Subthreshold static logic (Sub-CMOS), other logic families such as Subthreshold pseudo-nMOS (Subpseudo nMOS), Variable Threshold voltage Subthreshold

CMOS (VT-Sub-CMOS), Subthreshold Dynamic Threshold voltage MOS (Sub-DTMOS), and subthreshold dynamic logic (Sub-Domino logic) have been proposed [60, 61, 62].

2.3.1 Origin and Evolution of Subthreshold Circuits

In the 1970s, subthreshold circuits were first considered for analog design in the development of micropower circuits [68]. Between the 1980s and the early 1990s, subthreshold circuits were proposed to implement analog VLSI designs that emulated functions of the brain [38, 69]. It was not until the late 1990s that subthreshold circuits were considered in the digital domain [59]. Subsequently, a growing number of successful implementations of subthreshold systems have occurred.

In 2001, Paul et al. [47] designed an 8x8 VT-Sub-CMOS array multiplier in $0.35\mu m$ technology. They demonstrated that the PDP of this multiplier is around 25 times lower than its strong inversion operation. In 2003, Kim et al. [31] reported an ultra-low-power Delayed Least Mean Square (DLMS) adaptive filter for hearing aid devices, operating in Sub-CMOS and Subpseudo nMOS in $0.35\mu m$ technology. This parallel DLMS adaptive filter achieves a 91% improvement in power compared with a nonparallel CMOS implementation. One year later, Wang and Chandrakasan [71, 72] fabricated a 180-mV subthreshold Fast Fourier Transform (FFT) processor in $0.18\mu m$ technology. At the optimum supply voltage, the FFT processor dissipates $155nJ$ for a 16 bits and 1024 point FFT. Later on, Calhoun and Chandrakasan [13, 14] proposed a 256k Sub-threshold SRAM in $65nm$ technology. They show that a traditional six transistor (6 T) SRAM cannot function in the subthreshold region, but solved the problem by introducing a 10 T bitcell. Finally, Zhai et al. [78] have fabricated a 2.60pJ/instruction subthreshold sensor processor in $0.13\mu m$ technology, and showed that the minimum energy consumption of the core is improved 10 times in that of previous sensor processors of the same MIPS.

Obviously, subthreshold circuits have been used in analog designs for a long time. Moreover, the authors have proven that it is possible to implement subthreshold logic circuits by using standard CMOS process technologies. However, it is prudent to analyze the suitability of these standard technologies in terms of power and performance in the subthreshold regime. This topic is addressed later after a look at the process variations in subthreshold circuits, one of the most challenging obstacles in recent DSM technologies and dramatically accentuated in subthreshold designs.

2.4 Process Variations

Besides the power consumption issue, the process variation issue in today's IC design is paramount [7]. This section briefly introduces this topic. Thereafter, a survey of the literature that address variability in subthreshold circuits is conducted. The identification of the dominant components of variations in the subthreshold regime is imperative to explore this issue.

Since fabrication process tolerances have not scaled proportionally with device dimensions, the relative impact of process variations (especially beyond $90nm$) on power and timing has become more significant with each technology generation [67]. For example, the magnitude of variations in a transistor's gate length is predicted to increase from 35% in 130nm technology to almost 60% in 70nm technology. Usually, these variations are expressed as a fraction, $3\sigma/\mu$, σ and μ are the standard deviation and the mean of a process parameter, respectively, thus, 3σ is considered the worst case shift in a parameter. Hence, a 60% variation in 70nm technology relates the σ of the distribution gate length across a large number of samples in 14nm. Consequently, to maintain the benefits of technology scaling, designers should treat these variations with alternative statistical strategies rather than traditional guard-band deterministic approaches [64].

Process variations are fluctuations around the desired value of design parameters introduced during chip device fabrication [4]. The process variations are, inter-die (die-to-die) and intra-die (within-die). Inter-die variations originate from factors such as the processing temperature, and equipment properties [11]; intra-die variations result from factors such as the random placement of dopant atoms in the channel region, and channel length variations across a single die [67]. Traditionally, inter-die variations have become the main concern in CMOS digital circuit design [20]. However, intra-die variations have become just as important, and their impact on frequency and power is becoming more and more pronounced [11]. Figure 2.14 reflects the trend in the ratio between the intra-die and total process variations for some key technology device parameters. The data is reported in [43], and represents projections of the ITRS [7] in conjunction with data from IBM processes. It can be seen that variations in the channel length are expected to increase significantly. In addition, variability in oxide thickness, as well as threshold and supply voltages also increase.

Parameter variability in present designs has intensified the need to consider the impact of statistical leakage current variations. Leakage variability is vital in subthreshold designs, since this current drives the circuits. For example, a 10% variation in the transistor effective channel length can lead to as much as a three-fold difference in the amount of subthreshold leakage current [51]. Gate leakage current exhibits an even greater sensitivity to variations; a 10% variation in the oxide thickness indicates a 15X difference in current (in a 100nm Berkeley Predictive Technology Model) [63]. In addition, random microscopic fluctuations in the number and location of dopant atoms in the channel region, directly produce variations in the threshold voltage, subthreshold swing, drain current, and subthreshold leakage current [74]. For example, consider a uniformly doped nMOS, where $L_g = 45nm$, $W = 3L_g$, impurity density $N_{ch} = 10^{18}cm^{-3}$, and $W_{dep} = 35nm$, the average number of acceptor $N = N_{ch}L_gWW_{dep}$ is approximately 200. As a result, any random

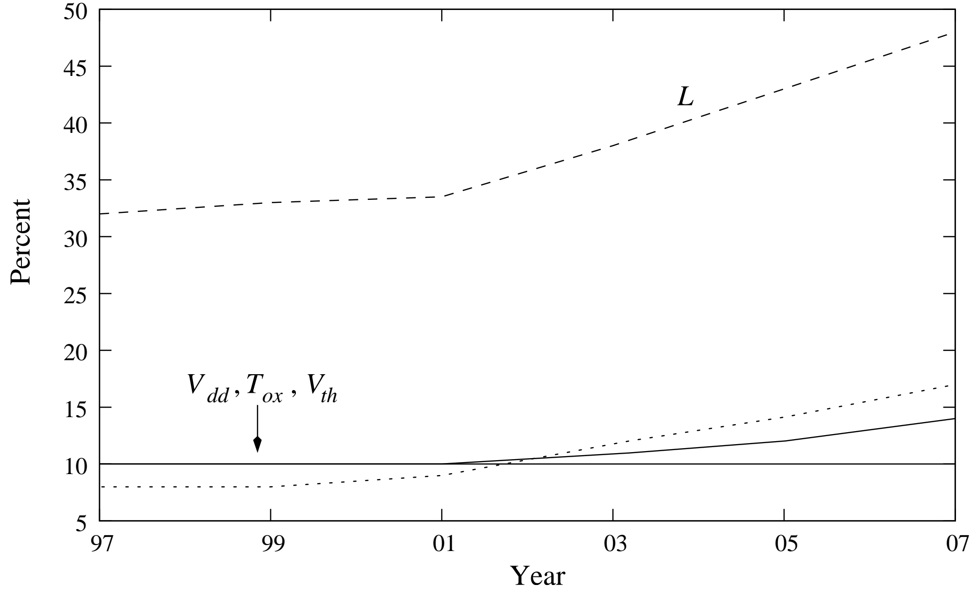

Figure 2.14: Percentage of the total variations accounted for intra-die variation for different technology generations.

variation in this small number of dopants is translated into a shift in the value of V_{th}. Some work in the literature [10] shows that for the chips that meet the required operating frequency, a large portion dissipates very large amount of leakage power and thus, are unsuitable for commercial use. In today's designs, the yield is determined according to the operating frequency and the leakage power [4].

Given the categories of variation and the impact of device parameter variability in circuit behavior, the three essential parameters to account for are oxide thickness, gate length, and channel doping concentration.

2.4.1 Process Variations in Subthreshold Designs

The exponential dependence of subthreshold leakage current (I_{on} and I_{off} in subthreshold circuits) on the threshold voltage, derived from (2.5), results in subthreshold circuits with

a marked sensitivity to V_{th} variation. The threshold voltage is strongly related to several device parameters, which undergo considerable variations in DSM technologies [64]. Therefore, it is not surprising that subthreshold designs are prone to process variations.

As CMOS devices are further scaled in the nanometer regime, variations in the number and placement of dopant atoms in the channel region, called Random Dopant Fluctuation, (RDF), cause random variations in the threshold voltage, where the V_{th} standard deviation is roughly $\sigma_{V_{th}} \propto \sqrt{\frac{N_{ch}}{WL}}$. This source of variation is the most significant phenomenon in subthreshold operation, as shown by Zhai et al. [77]. They have reduced the impact of RDF through circuit sizing and the choice of circuit logic depth (i.e., averaging this random effect). In addition, the authors have derived statistical models for circuit delay, power, and energy efficiency as a function of the circuit parameters. Also, Kim et al. [32] have reduced the sensitivity to RDF by a device sizing optimization process which uses the Reverse Short Channel Effect (RSCE) present in standard CMOS non-uniform halo doping profile devices.

A major concern, related to process variations, is robustness or the correct functionality of subthreshold circuits. If, for example, the variations strengthen an nMOS, relative to a pMOS device, a PUN cannot drive the logic gate output fully to V_{dd} because of the increased idle leakage in the PUN, and viceversa. To cope with this robustness issue, Kwong and Chandrakasan [35] have introduced a criterion to determine how sizing affects the variability in the output logic swing and active I_{on} current in several topologies. With this consideration, they have proposed a design methodology for minimum energy subthreshold circuits by device sizing and optimal supply voltage. They have concluded that upsizing is necessary to achieve robustness at reduced voltages.

With small V_{th} variations, drive-I_{on} pMOS and nMOS currents can easily differ by an order of magnitude, or even more, in subthreshold circuits. For example, with a reasonable $6\sigma_{V_{th}} = 100mV$, the V_{th} variation disturbs the MOS current ratios by approximately 1.17

times in strong inversion operation, whereas a similar V_{th} variation upsets current matching by at least 10 times in subthreshold operation [25]. As a consequence, the rise and fall times differ, impacting the switching frequency, and thus, power consumption. So, Melek et al. [39] have equalized mismatched currents by body bias compensation circuits. With a similar body biasing approach, Jayakumar and Khatri [27] have targeted inter- and intra-die Process, supply Voltage, and Temperature (PVT) variations. These techniques inherently lead to area and power overheads due to the extra-circuitiry, but these penalties are not included.

Obviously, process variation plays a key role in energy efficiency and robustness in subthreshold designs. In addition, it should be noted that the previous research involve circuit and architecture level techniques, whereas guidelines for designing a transistor for subthreshold operation is not considered under process variations. The next section conveys some considerations for constructing a subthreshold optimized transistor.

2.5 Devices in Subthreshold Circuits

In Subsection 2.3.1 is discussed that subthreshold designs can be constructed with standard CMOS process technologies. In 2005, Bipul Paul, Arijit Raychowdhury, and Kashik Roy [46] have optimized a transistor structure for digital subthreshold operation. They have demonstrated that the optimized device improves the delay and PDP of an inverter chain by 44% and 51%, respectively, of standard Bulk-MOS transistors operated in the subthreshold region. This is a key result for ultra-low-power designs and one of the sources for inspiration of this thesis. This section describes the features and design considerations of the optimized transistor for subthreshold operation.

Throughout this dissertation, the term *subthreshold transistor* refers to the device that is optimized for operating in the subthreshold region, whereas *standard transistor* corre-

sponds to an ordinary transistor oriented to strong inversion operation.

Highly scaled CMOS technologies have satisfied the demand for higher levels of integration and performance in modern VLSI systems. Since subthreshold designs inherently do not operate at high frequencies, it appears that modern highly scaled devices are not necessary in this regime. However, oxide thickness scaling benefits the subthreshold slope, (2.7), and the device length scaling results in a reduction in the total capacitance, as portrayed in Fig. 2.13. Consequently, device intrinsic delay ($\tau \propto C_g$), and dynamic power are reduced such that, subthreshold designs also take advantage of technology scaling.

The behavior of scaled DSM devices differs considerably from that of the well known strong inversion (linear and saturation) region due to undesirable Short-Channel Effects (SCE). In order to mitigate the SCE such as V_{th} roll-off, drain induced barrier lowering, and body punchthrough, previously explained in Subsection 2.1.1, standard transistors incorporate nonuniform, halo, and retrograde doping profiles [66]. The DIBL is a function of a high drain voltage, applied to a short-channel device, resulting in a decreased V_{th} which substantially increases the subthreshold current. At even higher drain voltages, the device reaches the punch-through condition, and the gate loses control over the channel, developing high drain currents that are independent of the control gate voltage.

It is noted worthy that the described SCEs occur at high drain voltages. However, in the subthreshold regime, the power supply is small ($V_{dd} \approx [400, 100]mV$), and the DIBL and body punchthrough are very low. As a result, to simplify the fabrication process technology and to significantly lessen the junction capacitances (resulting in a faster device operation and lower power consumption), the subthreshold device is constructed without any halo and retrograde doping profiles [46]. For this reason, the subthreshold transistor is characterized with a simplified uniform high-to-low doping profile in conjunction with the symmetrical Bulk-nMOS structure introduced in Fig. 2.1.

Note that the previous authors have not followed nor proposed an automatic framework

to perform the device optimization process, and have not accounted for process variations. The result is the absence of research on device level techniques to mitigate process variations. This thesis offers an automatic device optimization technique which includes process variations in a subthreshold transistor design.

2.6 Summary

This chapter describes some issues in reducing power consumption in scaled devices. A MOS's intrinsic leakage mechanisms are explained. The three more contributive leakage components in DSM technologies are subthreshold, oxide tunneling, and reverse bias p-n junction BTBT. Subthreshold leakage is the drive current in the subthreshold regime. In addition, the characteristics of MOS transistors in the subthreshold region are examined with respect to the strong inversion regime, including current flow mechanism, intrinsic capacitances, and the exponential relation between the gate voltage and drain current. Then, the paradigms of subthreshold logic are outlined. Subthreshold logic and strong inversion CMOS logic share several properties such as high noise margins, and low output impedance with improved features in power consumption and gain. Beginning with the analog field, it was not until the late 1990s that subthreshold circuits attracted attention in the digital domain; since then, several subthreshold systems have been implemented with standard DSM technologies. Later, process variations and their impact on circuit behavior are presented. The three essential parameters to account for variations are: oxide thickness, channel length, and channel doping concentration. Variability is one of the most challenging obstacles in recent technologies, and is accentuated in subthreshold designs. Finally, an optimized transistor structure for subthreshold operation is discussed, as a promising construction block in terms of simplified fabrication processes, faster operation, and lower power consumption.

Chapter 3

Problem Definition and Yield Maximization Technique

This chapter extends the qualitative concept of process variations from the last chapter by examining analytically the variations of device parameters, and their impact on performance and power consumption in a subthreshold device. After, the yield maximization problem is defined according to selected device parameters, lastly, the yield maximization technique is developed.

3.1 Design Parameters and Variations

This section covers the proposed device structure, the constraints of interest, and device parameters which are the most susceptible to variability. It should be noted that the device parameter models provide an understanding of why certain parameters are or not included in the set of design variables for the device optimization problem. The actual technique is carried out by the MEDICI device simulator [1].

3.1.1 Device Structure and Constraints

For convenience, the characteristics of a subthreshold transistor, models and definitions outlined in Section 2, are re-stated.

It is an established fact that for scaled MOS standard transistors, it is essential to incorporate halo and retrograde doping profiles to mitigate SCE [66]. However, in the subthreshold regime, the power supply is small ($V_{dd} \approx [400, 100]mV < V_{th}$), and the SCE such as DIBL and body punchthrough are minimal. As a result, for a more simplified fabrication process technology and significantly fewer junction capacitances (resulting in a faster device operation and a lower power consumption), the subthreshold device is characterized without any halo and retrograde doping profiles [46]. For this reason, a simplified uniform doping profile, in conjunction with the symmetrical Bulk-nMOS structure are depicted in Figure 3.1.

Since, in subthreshold logic, circuits are driven entirely by the subthreshold leakage current, it is important to further examine the major transistor leakage components in order to form a realistic estimation scheme.

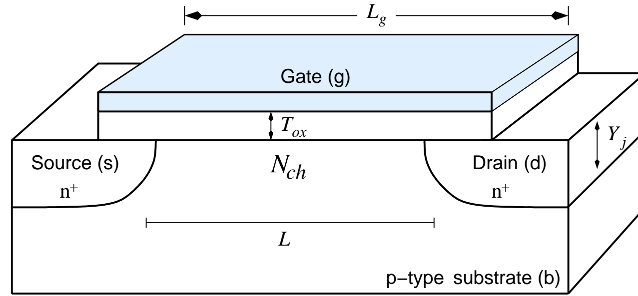

Figure 3.1: Architecture of the subthreshold device.

Transistor Leakage Currents and Intrinsic Delay in the Subthreshold Regime

The most contributive leakage components in nanometer technologies that are identified in Chapter 2 are subthreshold leakage, I_{sub}, gate leakage, I_{gate}, and reverse biased drain-substrate and source-substrate junction BTBT, I_{BTBT}.

I_{sub} is the *drive current* in the subthreshold regime (i.e., $I_{on} = I_{sub}(V_{gs} = V_{ds} = V_{dd} < V_{th})$, where V_{dd} is the power supply voltage). In the transistor off-state, I_{sub} is the drain current, when the gate voltage is zero (i.e., $I_{off_{sub}} = I_{sub}(V_{gs} = 0, V_{ds} = V_{dd} < V_{th})$). This off-state current is the dominant contributor to static power consumption. Therefore, for computing the total leakage current, $I_{off_{sub}}$ should be taken into account. As derived in the last chapter, the subthreshold current is exponentially dependent on the threshold voltage as follows:

$$I_{sub}(V_{gs}, V_{ds}) = I_o e^{(V_{gs}-V_{th})/nV_T}\left(1 - e^{-V_{ds}/V_T}\right), \tag{3.1}$$

where $I_o = \mu_0 C_{ox}\frac{W}{L}V_T^2$, $V_T = kT/q$ is the thermal voltage, C_{ox} is the oxide capacitance, μ_0 is the zero bias mobility, $\frac{W}{L}$ is the width over length ratio of the device, and $n = 1 + C_{dep}/C_{ox}$ is the subthreshold swing coefficient.

Direct gate leakage is a byproduct of the oxide thickness scaling, T_{ox}, required to overcome the V_{th} roll-off in scaled technologies. Such a current is due to the tunneling of an electron (or hole) from the Si-bulk through the gate-oxide potential barrier into the gate, and is an exponential function of T_{ox} [56], computed by

$$J_{DT} = A\left(\frac{V_{ox}}{T_{ox}}\right)\exp\left\{-\frac{BT_{ox}\left[1-\left(1-\frac{V_{ox}}{\phi_{ox}}\right)^{\frac{3}{2}}\right]}{V_{ox}}\right\}, \tag{3.2}$$

where V_{ox} is the drop across the thin oxide, and ϕ_{ox} is the barrier height for the tunneling particle (electron or hole). A and B are physical parameters dependent on the barrier height

and are given in the literature [56]. Because the tunneling current increases exponentially with a decrease in T_{ox}, the gate tunneling leakage cannot be neglected, when the oxide thickness is less than $3nm$ [53]. Therefore, $I_{gate} = J_{DT} \times W \times L$ should be considered in computing the total leakage current, first to have a realistic indication of the leakage mechanisms which contribute to the leakage power consumption, and secondly, to consider the tradeoff between the V_{th} roll-off improvement by means of T_{ox} reductions and I_{gate} exponential increases.

In scaled standard CMOS technology, the higher substrate doping density and the necessary incorporation of halo profiles generate a significantly large junction BTBT current to flow through the reverse biased drain-substrate and source-substrate junctions [52]. The BTBT current, I_{BTBT} is estimated as [41]

$$I_{BTBT} = \left(\frac{W X_{jSDE} \hat{A}}{E_g^{1/2}} \right) \xi V_{dd} \exp \left(- \frac{\hat{B} E_g^{3/2}}{\xi} \right), \tag{3.3}$$

where

$$\xi = \sqrt{\frac{2q N_{aside} N_{sdside}}{\epsilon_{Si} (N_{aside} + N_{sdside})} \left[V_{dd} + \frac{kT}{q} \ln \frac{N_{aside} N_{sdside}}{n_i^2} \right]}. \tag{3.4}$$

X_{jSDE} is the source/drain extension junction depth (for standard CMOS DSM devices), N_{aside} and N_{sdside} are the p-side and n-side junction doping. E_g is the band-gap of the silicon, q is the electronic charge, and lastly, $\hat{A}$ and $\hat{B}$ are the physical coefficients given by Taur and Ning [65]. However, the BTBT leakage is mainly present in the halo and retrograde doping profiles, which are not incorporated in the subthreshold structure. Besides, as V_{dd} is reduced, the BTBT leakage is negligible [15, 46], and is ignored.

Therefore, to gain a realistic indication of the total leakage, TL, the first two leakage components are added for the worst case; that is,

$$TL = I_{sub}(V_{gs} = 0, V_{ds} = V_{dd}) + I_{gate}(V_{gs} = V_{dd}, V_{ds} = 0). \tag{3.5}$$

$V_{dd} < V_{th}$ to ensure the subthreshold operation. Figure 3.2 portrays a typical scheme, where these two leakage components contribute to the total leakage power consumption. This unwanted leakage power consumption represents a concern in energy-constrained systems for extending battery and system lifetime [73]. Therefore, TL is chosen as the first constraint in the optimization design problem.

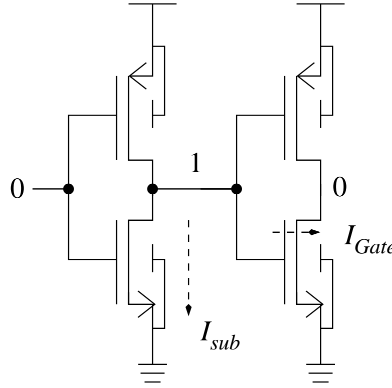

Figure 3.2: Total leakage (TL) estimation scheme.

The primary metric for transistor speed is the transistor intrinsic delay [76], τ, defined as

$$\tau = C_g \frac{V_{dd}}{I_{on}}, \tag{3.6}$$

where C_g is the gate capacitance per micron of the transistor width, and I_{on} is the subthreshold drive current/μm. τ is the time required for a MOSFET to charge or discharge the gate of another identical MOSFET with a potential difference of V_{dd}. In fact, τ is the metric for MOSFET performance suggested by the ITRS [7], and constitutes the performance constraint in the optimization design problem.

To incorporate the device design parameters in the design problem, it is prudent to look at their variability impact on the defined power and performance constraints.

3.1.2 Device Parameters and Variations

To establish a set of device design parameters for the optimization problem, device parameters such as physical gate length (L_g), oxide thickness (T_{ox}), junction depth (Y_j), transistor width (W), and channel doping concentration (N_{ch}) are considered in the following. A good starting point for the discussion is the exponential dependence between the subthreshold leakage I_{sub} and V_{th} in (3.1). Therefore, the variations in V_{th} impact both $I_{off_{sub}}$ and driving current I_{on}. In turn, any variation in I_{on} echoes for τ as well. Thus, the impact of device parameters on the V_{th} should be considered.

Physical Gate Length (L_g)

The threshold voltage of a short-channel device decreases as L_g is reduced. This is due to the closer proximity of the source and drain areas, whose surrounding depletion regions penetrate into a considerable portion of the channel, as signified in Figure 3.3 [75]. W_{dep} is the depletion-layer width, L is the channel length, and L' is the reduced channel region. Consequently, the total charge in the channel, Q'_B, is reduced to trapezoidal region $Q'_B \propto W_{dep} \times (L + L')/2$, in contrast to the long-channel device case, where $Q_B \propto W_{dep} \times L$ [75]. Therefore, less charge (Q'_B) must be inverted by the gate voltage to reach the V_{th} which is defined as follows [65]:

$$V_{th} = V_{fb} + \psi_s + \frac{Q'_B}{C_{ox}}. \tag{3.7}$$

V_{fb} is the flat-band voltage, ψ_s is the surface potential, and Q'_B is the total gate depletion (trapezoidal) charge. The shift in the V_{th}, originated by channel length scaling ΔV_{th}, is

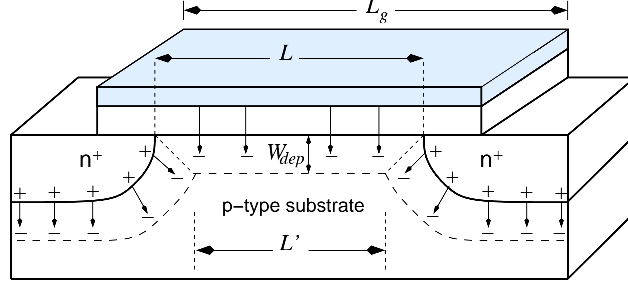

Figure 3.3: Charge sharing model.

approximated as [36]

$$\Delta V_{th} = [2(V_{bi} - \psi_s) + V_{ds}](e^{-L/2l} + 2e^{-L/l}), \tag{3.8}$$

where V_{bi} is the built-in potential, and l is the characteristic length defined as

$$l = \sqrt{\frac{\epsilon_{Si} T_{ox} W_{dep}}{\epsilon_{ox} \eta}}. \tag{3.9}$$

ϵ_{Si}, ϵ_{ox} are the silicon and oxide permittivity, respectively, and W_{dep}/η is the average depletion layer width along the channel. This analytical approximation defines a short channel effect known as the V_{th} roll-off. As a result, any reduction in L_g increases the subthreshold leakage current, and hence, the power consumption [65]. This does lead to a faster device, since the reduction not only boosts the driving current, but also reduces the gate capacitance. Therefore, in finding an appropriate gate length, this tradeoff should be considered in the design problem.

Oxide Thickness (T_{ox})

The oxide thickness has a considerable effect on the V_{th} since any variation in T_{ox} impacts the oxide capacitance, C_{ox} per unit area, given by

$$C_{ox} = \frac{\epsilon_{ox}}{T_{ox}}. \tag{3.10}$$

Any variation in C_{ox} affects V_{th} in (3.7) and I_{sub} in (3.1). Moreover, the SCE is affected by T_{ox}, as given in (3.9); thus, a thinner oxide is needed to overcome the V_{th} roll-off in scaled devices. However, this oxide reduction increases the gate leakage current exponentially in (3.2), and hence, the power consumption. Careful engineering of the oxide thickness dimension is crucial to meet the desired device delay and total leakage constraints. As a consequence, T_{ox} should be considered in the design problem.

Channel Doping (N_{ch})

As CMOS devices are scaled further into the nanometer regime, variations in the number, and the placement of dopant atoms in the channel region cause random variations in V_{th} (as discussed in Section 2). This effect, called Random Dopant Fluctuation (RDF), is accentuated with technology scaling, since the average number of dopant atoms in the channel is rather reduced [37]. A simple first order model of the V_{th} standard deviation ($\sigma_{V_{th}}$), due to RDF is given by [65]

$$\sigma_{V_{th}} = \frac{q}{\epsilon_{ox}}\sqrt{\frac{N_{ch}W_{dm}}{3L_gW}}. \tag{3.11}$$

Reinforcing this discussion, Zhai et al. [77] have concluded that RDF is the dominant source of variations in the subthreshold operation. It is observed that by reducing N_{ch}, the threshold voltage variation is also reduced. However, since in (3.7), $Q'_B = \sqrt{4\varepsilon_{si}qN_{ch}\psi_B}$, (where ψ_B is the difference between the Fermi potential and the intrinsic potential, $|\psi_f - \psi_i|$), the threshold voltage also drops when the channel doping is reduced, increasing the subthreshold off-state leakage current. For these reasons, the channel doping concentration should also be included in the subthreshold device design.

Junction Depth (Y_j)

Aggressive device scaling has necessitated shallow junction depths to reduce the SCE and suppress the depletion layer penetration into the channel. The result is an increase in parasitic device resistance and involves a complex fabrication process [45, 42]. Consequently, any treatment of Y_j is limited by the sheet resistance, R_{sh}, of the source/drain region which is defined as [65]

$$R_{sh} = \rho_{sd} \frac{S_{sd}}{W}, \tag{3.12}$$

where S_{sd} is the spacing between the gate edge and the source/drain contact edge in Fig. 3.4, and the sheet resistivity, ρ_{sd}, is given by

$$\rho_{sd} = \frac{\rho}{Y_j}, \tag{3.13}$$

where ρ is the resistivity of the diffusion region. Therefore, it is desirable to keep the parasitic resistance low to achieve sufficient current-drivability. As a result, the design parameter, Y_j, should not be included as part of the optimization problem and follows the S/D profile from [6].

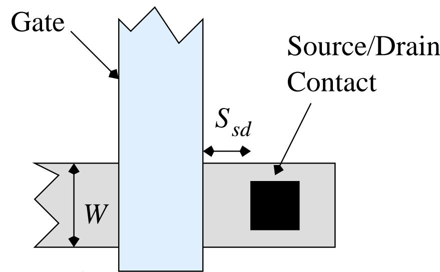

Figure 3.4: Top view of the transistor.

Transistor Width (W)

The transistor width is the principal parameter that circuit designers can change to meet the required specifications of circuits and systems. In section 2 all the state-of-the-art research addressing variability involves circuit and architecture level techniques which require to resize W. In addition to the device level approach presented in this thesis, designers can still use such circuit and architecture level techniques to further mitigate the effect of variations. For these reasons, the transistor width should not be included as part of the device optimization problem.

It is evident from the modeling of the relationship between the parameters and their variations, that I_{gate}, and I_{sub} are exponentially dependent on T_{ox}, and V_{th}, respectively; whereas V_{th} is a function of all the selected device parameters, that is,

$$\begin{aligned} &\Delta_{V_{th}} = f(T_{ox}, L_g, N_{ch}), \\ &\Delta_{TL} \propto I_{gate_{nominal}} e^{f(\Delta_{T_{ox}})} + I_{sub_{nominal}} e^{f(\Delta_{V_{th}})}, \\ &\Delta_{\tau} \propto \frac{f(\Delta_{T_{ox}})}{I_{sub_{nominal}} e^{f(\Delta_{V_{th}})}}. \end{aligned} \tag{3.14}$$

Thus, the problem is composed by the following three device parameters since their variations directly impact the design constraints, TL and τ.

- Oxide thickness (T_{ox})
- Gate length (L_g)
- Channel doping concentration (N_{ch})

The next section incorporates the three design parameters and the two constraints in the optimization technique.

3.2 Problem Statement and Yield Maximization Technique

This section begins with a general formulation of the device optimization problem that is composed of the selected device design parameters and constraints. After the idea behind the yield maximization process is established, the problem is formalized by the yield maximization technique.

3.2.1 General Problem Statement

The novel approach [26] adapted in this investigation to a subthreshold transistor design, consists of exploiting a 3-D parameter design space, constructed by T_{ox}, L_g, and N_{ch}. Hence, the yield optimization problem is declared as follows:

$$\begin{aligned} &\text{Given:} \quad \sigma_{T_{ox}}, \sigma_{L_g}, \sigma_{N_{ch}}, \\ &\max_{x \in \Re^3} \quad \text{Yield} = P\{C(x) = 1\}, \end{aligned} \tag{3.15}$$

where $x = [T_{ox}, L_g, N_{ch}]$ is the set of design variables, σ_i is the standard deviation of the i-th design parameter, and $C(x)$ is a Boolean random variable function, defined by the bounds of the critical delay (τ_{max}), and the maximum total leakage current (TL_{max}). $C(x)$ is formulated as

$$C(x) = (\tau(x) \leq \tau_{max}) \text{ AND } (TL(x) \leq TL_{max}). \tag{3.16}$$

Therefore, $P\{C(x) = 1\}$ is the probability that a device $x = (T_{ox}, L_g, N_{ch})$ meets the performance and power constraints in the presence of variations in the design parameters.

3.2.2 Qualitative Approach

To solve the optimization problem, (3.15), the first step is to find a 3-D space, generated by the three device design parameters, bound by the power and performance constraints. This space is called the feasible region, F_c. In addition, an estimate of the probability of placing a device in F_c should be calculated; that is, the probability that a device $x_i = (T_{ox_i}, L_{g_i}, N_{ch_i})$ can satisfy the desired constraints, τ and TL. To estimate such a probability, $P\{C(x_i) = 1\}$, a cube is formed in the 3-D parameter design space, where all points within the cube satisfy the constraints.

For clarification, a similar problem with *two* design variables T_{ox} and L_g is denoted in Figure 3.5. Note that any point inside this plane represents the construction device dimensions, corresponding to the respective ordered pair (T_{ox}, L_g). A feasible region is defined in terms of the problem constraints. Any device x_i above the TL curve in Figure 3.5 satisfies the power constraint, and any device below the τ_{max} curve meets the performance constraint. Therefore, all the devices lying in the intersection of the defined zones can satisfy both constraints, as depicted by the shaded region in Figure 3.5. For the last constraint, the yield maximization problem is reduced to an inscribed rectangle that is formed by four corner devices: (T^l_{ox}, L^l_g), (T^l_{ox}, L^u_g), (T^u_{ox}, L^l_g), and (T^u_{ox}, L^u_g) in the 2-D feasible region. The center of the maximum yield rectangle $x^c = (T^c_{ox}, L^c_g)$ represents a device with the set of design values most immune against the variations. Finally, Monte Carlo simulations are carried out to verify the optimal design (x^c) yield, which is defined as the percentage of the total devices (scattered points) whose τ and TL values fall within feasible region F_c.

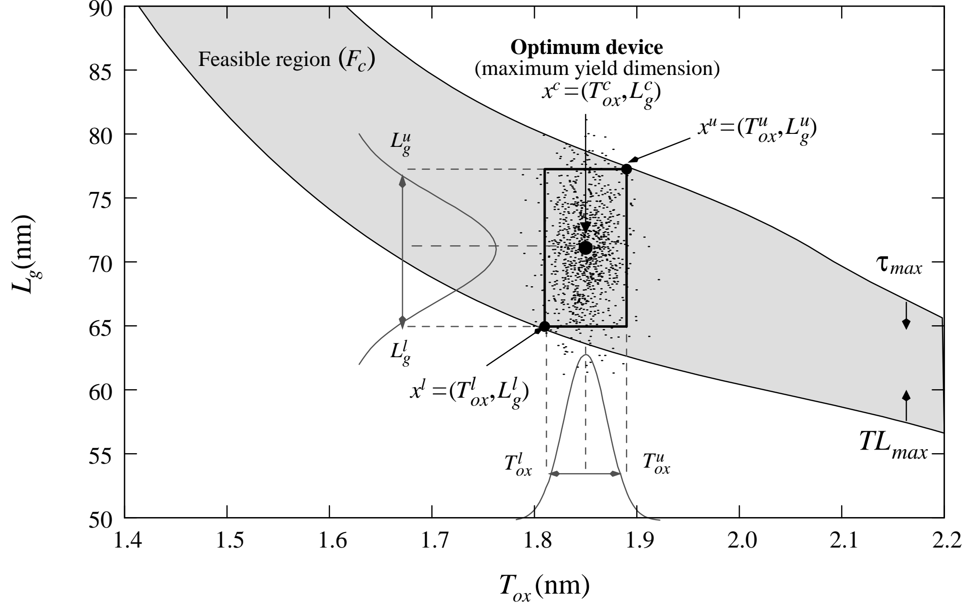

Figure 3.5: Simplified problem in 2-D.

3.2.3 Formal Problem Statement

To translate the idea explained in the last section to the 3-D space, the feasible region of the new design space should be identified from (3.17), a 3-D space, where any device $x_i = (L_{g_i}, T_{ox_i}, N_{ch_i})$ satisfies the condition in (3.16). As mentioned earlier, all device constraints are verified directly by the MEDICI device simulator throughout this investigation. The feasible region, F_c, is formulated by

$$F_c = \{x \in \Re^3 | C(x) = 1\}. \tag{3.17}$$

In addition to the feasible region, for the three device parameters, a cube should be formed and inscribed in (3.17). Therefore, all the devices, lying inside this cube, also satisfy (3.16). The cube is defined as follows:

$$cube(x^l, x^u) = \{x \in \Re^3 | x^l \leq x \leq x^u\}, \tag{3.18}$$

and

$$cube(x^l, x^u) \subseteq F_c,$$

where x^l and x^u are the coordinates of the extreme corners.

Before solving the yield optimization problem in (3.15) the Probability Distribution Function (PDF) of each design variable are modeled. In this thesis, the variation of each design parameter, T_{ox}, L_g, and N_{ch} is considered to be independent and the distribution is assumed to be Gaussian [63]. Since this distribution function does not have a closed form integral (Cumulative Distribution Function CDF), which represents the yield evaluation, Kumaraswamy's double-bounded density function [33] is adopted. This model has a simple closed form for both PDF ($f(z)$) and CDF ($F(z)$) as follows:

$$f(z) = abz^{a-1}(1 - z^a)^{b-1}, \tag{3.19}$$

$$z = \frac{x - x^{lb}}{x^{ub} - x^{lb}}, \quad x^{lb} \leq x \leq x^{ub}, \tag{3.20}$$

and

$$F(z) = 1 - (1 - z^a)^b, \tag{3.21}$$

where x^{ub} and x^{lb} are the upper and lower bounds of a random variable x, respectively. Depending on the a and b values, $f(z)$ takes different shapes. A truncated Gaussian shape with range $x^{ub} - x^{lb} = 6\sigma_x$ is used by setting a and b to 3.6 and 8, respectively. Finally, $x^{ub} = x^c + 3\sigma$ and $x^{lb} = x^c - 3\sigma$, where x^c is found by (3.23).

Supported by the last definitions, the yield problem (3.15) is extended to the following:

$$\text{Given}\begin{cases} x = [L_g, T_{ox}, N_{ch}] \\ x^l = [L_g^l, T_{ox}^l, N_{ch}^l] \\ x^u = [L_g^u, T_{ox}^u, N_{ch}^u], \end{cases}$$

$$\begin{aligned}
&\text{Yield}(x^l, x^u) \\
&= P\{C(x) = 1\}, \\
&= \textstyle\prod_{i=1}^{3} P\{x_i^l \le x_i \le x_i^u\}, \\
&= \textstyle\prod_{i=1}^{3} \left(F\left(\frac{x_i^u - x_i^{lb}}{x_i^{ub} - x_i^{lb}}\right) - F\left(\frac{x_i^l - x_i^{lb}}{x_i^{ub} - x_i^{lb}}\right)\right), \\
&= \textstyle\prod_{i=1}^{3} \left(F\left(\frac{x_i^u - (x_i^c - 3\sigma_{x_i})}{6\sigma_{x_i}}\right) - F\left(\frac{x_i^l - (x_i^c - 3\sigma_{x_i})}{6\sigma_{x_i}}\right)\right), \\
&= \textstyle\prod_{i=1}^{3} \left(F\left(\frac{x_i^u - x_i^l + 6\sigma_{x_i}}{12\sigma_{x_i}}\right) - F\left(\frac{x_i^l - x_i^u + 6\sigma_{x_i}}{12\sigma_{x_i}}\right)\right).
\end{aligned} \tag{3.22}$$

Note that the symmetrical assumption of the design variable distribution leads to easily locating the final device, lying at the center of the cube by computing

$$x^c = \frac{x^l + x^u}{2}. \tag{3.23}$$

The optimization problem is finally expressed as

$$\text{Given}\begin{cases} \text{Technology - Specific Variances:} \\ \quad \sigma_x = [\sigma_{L_g}, \sigma_{T_{ox}}, \sigma_{N_{ch}}] \\ \text{Constraints:} \\ \quad \tau_{max} \;\; \text{and} \;\; TL_{max} \end{cases}$$

$$\begin{aligned} \max_{x^l, x^u} \quad & \text{Yield}(x^l, x^u), \\ \text{Subject to:} \quad & cube(x^l, x^u) \subseteq F_c \\ & x^l \leq x^u. \end{aligned} \tag{3.24}$$

To effectively solve the constrained non-linear optimization problem, the Sequential Quadratic-Programming (SQP) algorithm of $Matlab^{®}$ is used [49, 2]. The variances of the design parameters (σ_{L_g}, $\sigma_{T_{ox}}$ and $\sigma_{N_{ch}}$) and the two bounds (TL and τ) are given to the optimization engine. Subsequently, the engine attempts to find a cube in the feasible region, and maximize *Yield*. Finally, the optimum device design parameters are the center point, coordinate (T^c_{ox},L^c_g,N^c_{sub}) of the cube which represents the set of the most immune values against variations. This framework appears in Figure 3.6.

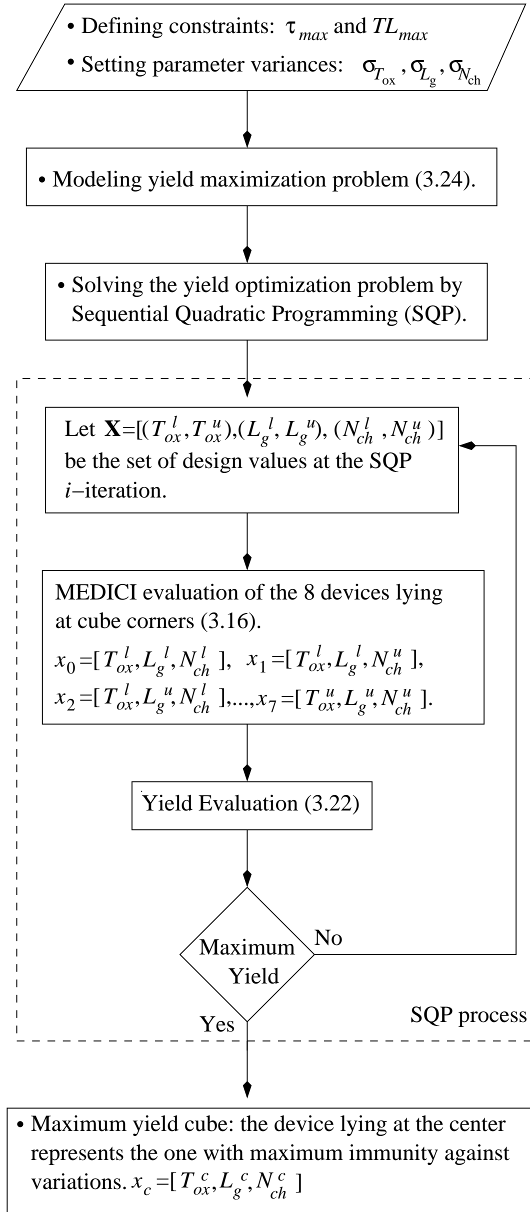

Figure 3.6: Flowchart of the yield maximization technique. For each iteration, a set of design values are determined by the optimization engine, and directly evaluated by MEDICI

3.3 Summary

In this chapter, the principal device parameters and their variations impact on transistor performance, and power dissipation is discussed. V_{th} is a function of T_{ox}, L_g, and N_{ch}, and I_{gate} and I_{sub} are exponentially dependent on T_{ox}, and V_{th}, respectively. Thus, the V_{th} variations should be considered in the design, since both the $I_{off_{sub}}$ current and the driving current I_{on} are impacted, and in turn any variation in I_{on} occurs in τ. The second part of this chapter describes a simple technique to find the maximum yield design which meets the power and performance constraints, where the selected device parameters, T_{ox}, L_g, and N_{ch} experiment variations. Moreover, the technique is also technology scalable, and can be easily adapted to any number of design variables, technology process variances, statistical distributions, and design constraints.

Chapter 4

Results and Discussion

This chapter describes the implementation and evaluation of sample devices for $90nm$ and $65nm$ technologies. The method to optimize a device necessitates selecting suitable constraints for the subthreshold regime. After the parameters and specifications of the optimum designs are discussed, a brief guideline for optimizing subthreshold transistors is outlined.

4.1 Implementation

To carry out the optimization methodology, MEDICI template files are developed to simulate the Bulk-Si nMOS transistors. The SQP numerical optimization engine is an iterative-based algorithm to find the optimum design point inside the feasible space, F_c. Consequently, for each iteration, a set of design values are determined by the optimization engine, and directly evaluated by MEDICI. As seen in (3.24), the optimum 3-D cube should be inscribed in the feasible region. The containment of the cube in this region ($\text{cube}(x^l, x^u) \subseteq F_c$) is verified by checking the worst cases, where each x element attains

its extreme values. In this way, these cases are formed by the 2^3 combinations of x^l and x^u extreme parameter values, that is, $x_0 = (T^l_{ox}, L^l_g, N^l_{ch}), x_1 = (T^l_{ox}, L^l_g, N^u_{ch}), x_2 = (T^l_{ox}, L^u_g, N^l_{ch}), ..., x_7 = (T^u_{ox}, L^u_g, N^u_{ch})$. This fact is observed by returning to the simple 2-D problem in Figure 3.5. Once the optimum $2^2 = 4$ corners of the rectangle $(T^l_{ox}, L^l_g), (T^l_{ox}, L^u_g), (T^u_{ox}, L^l_g)$, and (T^u_{ox}, L^u_g) are located in the feasible region, any x-device that is lying inside this rectangle also satisfies the constraints. Therefore, the containment verification process is reduced to a cube corner case. Finally, once the SPQ engine finds the maximum yield cube inside the feasible region, Monte Carlo simulations are carried out to verify the yield of the device located at the center subject to technology-specific variances of the device parameters.

4.2 Results and Discussion

Sample devices for $90nm$ technology are designed to balance speed and power. $3\sigma_{T_{ox}}$, and $3\sigma_{L_g}$ are chosen as $4\% \times 1.5nm$ and $12\% \times 90nm$, respectively, (according to $90nm$ technology specific variances [7]). In addition, a $65nm$ transistor is designed to see how the design parameters scale, when a shrunk (faster but leakier) technology is applied. In this case, the 3σ values are $4\% \times 1.2nm$ and $12\% \times 65nm$ for T_{ox} and L_g, respectively. $3\sigma_{N_{ch}}$ is equal to 10% of its center value at each iteration for both technologies. The lower limits of the device design parameters are set as follows: L_g^{min} is $33nm$ and $28nm$ for $90nm$ and $65nm$ technologies, respectively; T_{ox}^{min} is kept at $1nm$ for both cases [7]. The supply voltage is chosen as $250mV$ for both the $90nm$ and $65nm$ technologies. Table 4.1 lists the defined bounds on τ and TL of the proposed devices. The TL values are selected to be close to the *low operating power (LOP) devices* for $90nm$ technology, defined by the ITRS [7]. A realistic ratio, I_{on}/I_{off} for subthreshold devices is approximately $1,000$ [48], and for superthreshold transistors this ratio is approximately $100,000$ with a τ value of

Table 4.1: Desired device constraints and their technology node for optimized transistors.

Device	Technology	$\tau_{max}(ps)$	$TL_{max}(nA/\mu m)$
$90_{300,5.0}$	90nm	≤ 300	≤ 5.0
$90_{300,2.5}$	90nm	≤ 300	≤ 2.5
$90_{200,5.0}$	90nm	≤ 200	≤ 5.0
$65_{150,12.5}$	65nm	≤ 150	≤ 12.5

1.5ps or so (again, for LOP devices and 90nm technology [7]). In this way, as the intrinsic delay is a function of I_{on} (3.6), the τ constraint values are proposed in accordance to this approximately *one hundred* ratio difference between superthreshold and subthreshold devices; that is, the intrinsic delay for subthreshold devices is expected to be around one hundred times the intrinsic delay of LOP 90nm superthreshold transistors. For the 65nm technology device, the constraints are estimated according to expected five-fold I_{off} increase for each generation and to achieve at least a 30% delay improvement [8].

Table 4.2 summarizes the optimum device parameters (x^c) and specifications, obtained from the technique. Unfortunately, there are no industrial subthreshold devices to compare the results with, but the standard CMOS 90nm and 65nm technologies in the literature and guidelines, provided by the ITRS are adopted. The various physical limits and variances of both technologies are reflected in the T_{ox} and L_g values. Thin gate oxide thicknesses lead to an exponential increase of I_{gate} and leakage power, as mentioned in Subsection 3.1.1. Therefore, to maintain an adequate control over gate leakage current, the T_{ox} values exhibit a constant trend and converge to 2.3nm for 90nm technology devices. This value is similar to that of the oxide thickness offered for low standby power 90nm technology devices, proposed in the literature [21] (T_{ox} = 2.2nm). The oxide thickness value of the $65_{150,12.5}$ device corresponds to that of the general purpose 65nm technology device

Table 4.2: Design parameters and specifications of optimum devices.

Device	$90_{300,5.0}$	$90_{300,2.5}$	$90_{200,5.0}$	$65_{150,12.5}$
Yield (%)	98.9	84.0	92.3	99.0
T_{ox} (nm)	2.35	2.23	2.36	1.50
L_g (nm)	84.7	87.0	86.4	62.8
N_{ch} ($\times 10^{18} cm^{-3}$)	1.0	1.1	0.9	1.9
I_{off} ($nA/\mu m$)	2.2	1.6	2.9	5.6
I_{on} ($\mu A/\mu m$)	2.4	2.0	3.0	5.0
I_{on}/I_{off}	1054	1222	1026	896
τ (ps)	161	210	129	86
V_{th} (mV)	328	336	319	229
S (mV/dec)	76	75	76	75
L_{chan} (nm)	74	78	77	57

suggested by Fung et al. [22], ($T_{ox} = 1.4$). Subthreshold leakage I_{off} current is reduced as the threshold voltage V_{th} is increased, as given by (2.5). It is evident that the highest V_{th} value corresponds to that of the $90_{300,2.5}$ device; the $90_{200,5.0}$ device presents the lowest threshold voltage, and thus, the leakiest transistor. The I_{off} values are in good agreement with those proposed for *LOP devices* by the ITRS ($3nA/\mu m$ and $5nA/\mu m$ for $90nm$ and $65nm$ technologies, respectively) [7]. In the former, it is argued that a realistic I_{on}/I_{off} for subthreshold devices is approximately 1000, whereas for superthreshold *LOP* transistors in $90nm$ technology the ratio is one hundred times greater, $100,000$ with a corresponding τ value about $1.5ps$ according to the ITRS. Therefore, since I_{off} values of the proposed $90nm$ subthreshold devices are similar to *LOP* superthreshold devices, the τ values are about one hundred times the intrinsic delay of the LOP superthreshold transistors, as expected.

It is noteworthy that process precision does not allow the T_{ox} and L_g parameters to be optimized continuously. To cope with this issue, when the optimum device parameters are obtained, they are rounded according to the achievable process values, and then the remaining parameter (N_{ch} here) is re-optimized, considering T_{ox} and L_g as fixed values. This re-optimization process can be considerably faster, since the number of design variables is reduced. Moreover, the resultant optimized values do not change significantly, because the fixed values are rounded to the nearest process achievable value. For example, consider the $90_{300,5.0}$ designed device, assuming a precision level of $0.1nm$ for T_{ox}, and $1nm$ for L_g. Thus, T_{ox} remains at $2.4nm$, and L_g at $85nm$ in the re-optimization process. The new N_{ch} optimized parameter changes from $1.0 \times 10^{18}cm^{-3}$ to $0.90 \times 10^{18}cm^{-3}$. Leading to a negligible yield reduction from 98.9% to 98.1%.

The yield values in Table 4.2 correspond to the percentage of devices that satisfy both the performance and power constraints for variability cases. To get a clearer picture of the yield behavior, additional $90nm$ technology devices are designed to the reduce step-differences among the constraints: $(\tau_{max}[ps], TL_{max}[nA/\mu m]) = (300, 10)$, $(300, 7.5)$, $(200, 10)$, $(200, 7.5)$, and $(200, 2.5)$. Figure 4.1 records this particular result. It is evident that reducing the τ_{max} constraint from $300ps$ to $200ps$ leads to more pronounced yield degradation. Therefore, in applications that require medium to low frequency device performance, the design optimization method is expected to provide design yield values close to 100% with low off-state leakage current.

Table 4.3 reflects the three leakage components from Section 3.1.1: subthreshold off-state leakage, gate leakage, and BTBT leakage. To achieve a realistic indication, estimations of the worst cases are considered, that is, for I_{sub} and I_{BTBT}, $V_{gs} = 0$ and $V_{ds} = V_{dd}$, whereas for I_{gate}, $V_{gs} = V_{dd}$ and $V_{ds} = 0$. Obviously, $I_{off_{sub}}$ is orders of magnitude greater than I_{gate} and I_{BTBT} for the $90nm$ devices, and this tendency still holds but with an increasing contribution of gate leakage for the $65nm$ technology device. Thus, $I_{off_{sub}}$ con-

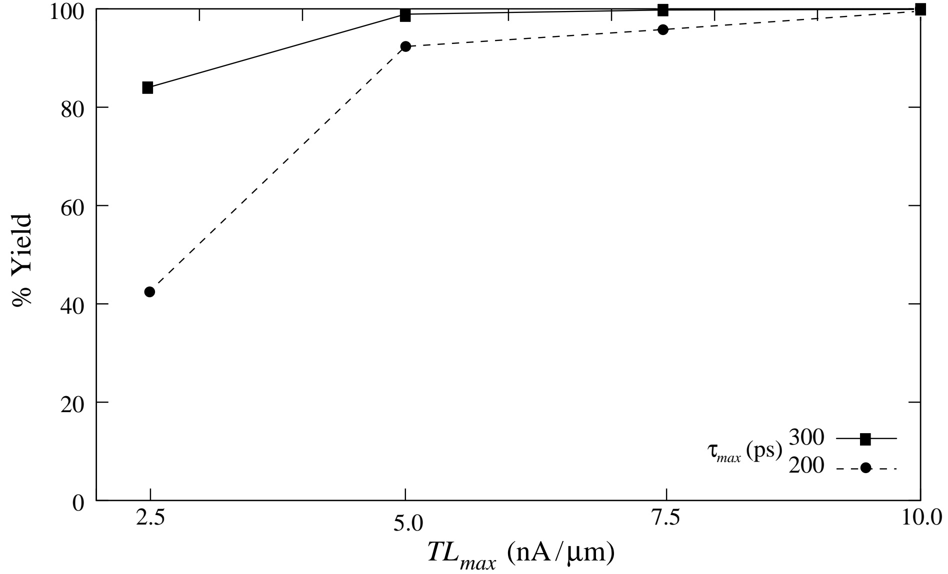

Figure 4.1: %Yield obtained from experimental $90nm$ devices.

Table 4.3: Various worst case leakage components of optimum devices.

Device	$I_{off_{sub}}(\times 10^{-9}A/\mu m)$	$I_{gate}(\times 10^{-12}A/\mu m)$	$I_{BTBT}(\times 10^{-16}A/\mu m)$
$90_{300,5.0}$	2.28	0.0373	4.63
$90_{300,2.5}$	1.65	0.1438	5.15
$90_{200,5.0}$	2.92	0.0337	4.37
$65_{150,12.5}$	5.62	384.14	8.61

tributes the most current to the static power consumption, and the most dominant leakage mechanism in the optimized subthreshold devices.

To verify the variability robustness and constraint satisfaction of the newly designed subthreshold transistors, Monte Carlo simulations (5,000 points) are performed based on the delay and current constraints in Table 4.1. Figures 4.2 and 4.3 signify the mean and standard deviation of the total leakage and delay of the optimum devices, respectively.

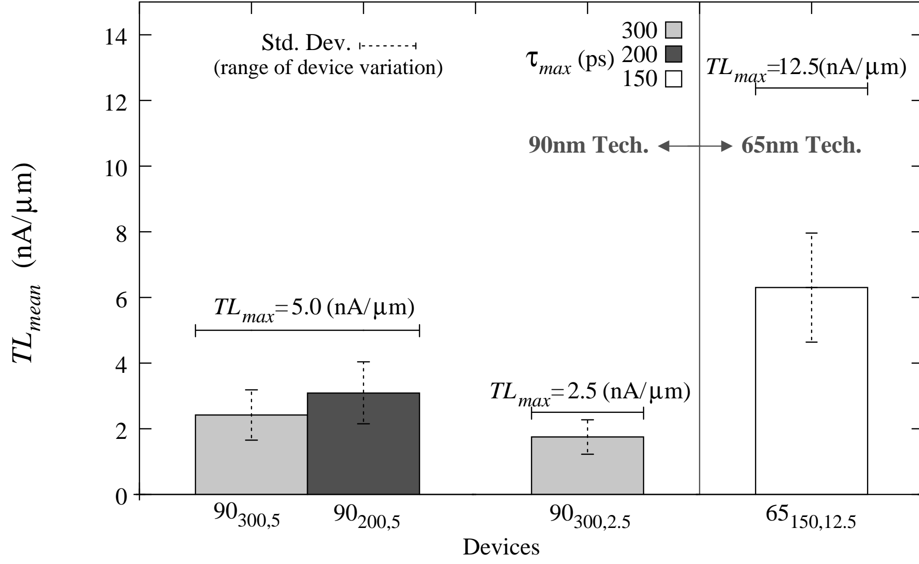

Figure 4.2: Mean and standard deviation of total leakage. The bars correspond to mean values whereas the dotted vertical lines are the standard deviations (range of device variation).

In Figure 4.2, it is noted that the TL_{mean} of the $90nm$ design with the lowest delay bound ($90_{200,5.0}$) is greater than the mean of $90nm$ designs with constraint $\tau_{max} = 300ps$. This is evident since the fulfillment of the lower delay bounds requires more drive current, and is thus, intrinsically, an increment of I_{off}. For the $65nm$ device, there is an approximate five-fold increment of leakage current. A reduction of τ_{mean} is expected, since the devices are bounded with higher TL limits as depicted in Figure 4.3. For the $65nm$ design, performance is improved in scaling technologies.

For a closer picture of the effect of the constraints, devices $90_{300,5.0}$ and $90_{200,2.5}$ are selected. Such devices can be manipulated to compare the delay and off-state leakage current dispersion, where the bounds are more relaxed ($300ps$ and $5nA/\mu m$), and others, where the one bound tightens ($300ps$ and $2.5nA/\mu m$). The spread of TL vs. τ is depicted

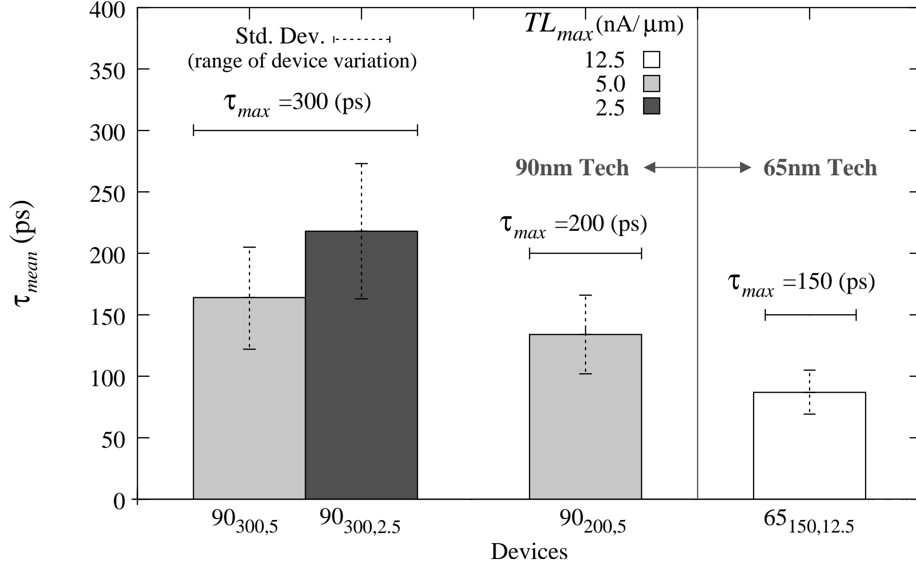

Figure 4.3: Mean and standard deviation of intrinsic delay. The bars correspond to mean values whereas the dotted vertical lines are the standard deviations (range of device variation).

in Figures 4.4 and 4.5, respectively.

In these figures, all the scattered devices that are inside the quadrant τ_{max} and TL_{max} represent the success of the transistors in meeting both constraints. The optimum devices are highlighted to observe their symmetric location with respect to the bounds. It can be seen in Figure 4.5 that as the TL_{max} constraint is reduced, many devices violate this power constraint as expected. Also there is an increase of the devices which violate τ_{max}, even though the performance constraint is constant for both newly designed devices ($300ps$). This occurs since the optimization process finds a center device, x^c, to meet a lower value of TL_{max}. Any device that has a reduction in its I_{off} current leads to a reduction in its I_{on} value, and therefore, an increase in its intrinsic delay τ. Since the optimum device has a greater value for τ, when the variations are incorporated, it is expected that the devices

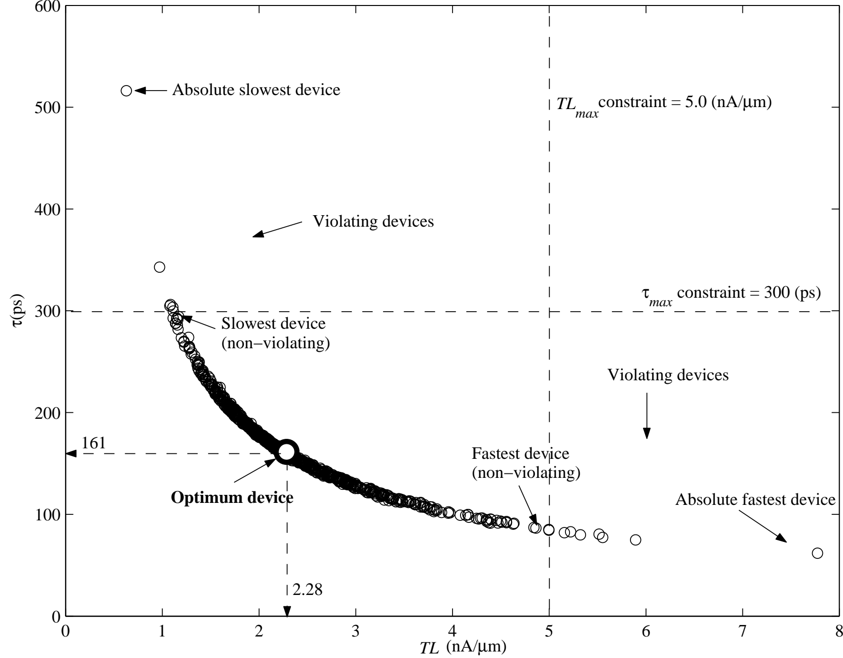

Figure 4.4: Monte Carlo simulation of the $90_{300,5.0}$ device, 98.9% of devices satisfy both constraints (relaxed TL bound).

also start to violate the τ_{max} constraint.

The extreme-violating devices in Figures 4.4 and 4.5, as well as the devices that are close to τ_{max} or TL_{max} but still meet both constraints (slowest/fastest devices, respectively) are captured, along with the optimum device (x^c), to depict their I-V curves in Figures 4.6 and 4.7. The *absolute fastest device (violating)* represents the most variation-affected device, and thus, violates the TL constraint, whereas the *absolute slowest device (violating)* is the most variation-affected transistor with the highest intrinsic delay, and thus, violates the τ constraint. The gray strip shows the variation of the output characteristics of the devices that meet both constraints; the *fastest device (non-violating)* and the *slowest device (non-violating)*, delimit this zone. In this way, the comparisons and effects of more relaxed

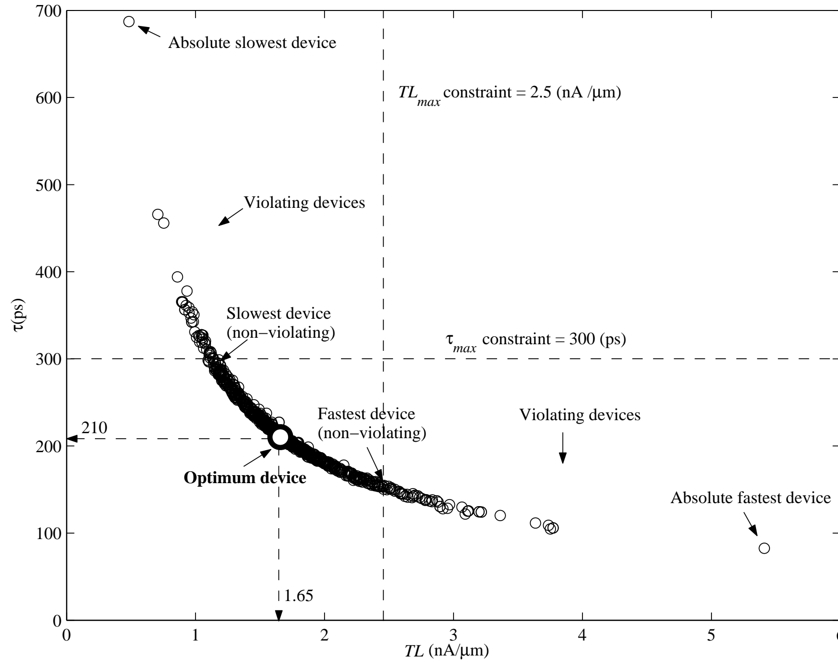

Figure 4.5: Monte Carlo simulation of the $90_{300,2.5}$ device, 84.0% of devices satisfy both constraints (tight TL bound).

or tighter constraints are observed for the two proposed designed devices. In fact, because the $90_{300,5.0}$ design has a yield value of 98.9% (more relaxed constraints), it is evident that the gray zone in Figure 4.6 is effectively wider than the corresponding zone in Figure 4.7 of the designed $90_{300,2.5}$ device which has a yield value of 84.0% (tighter constraints).

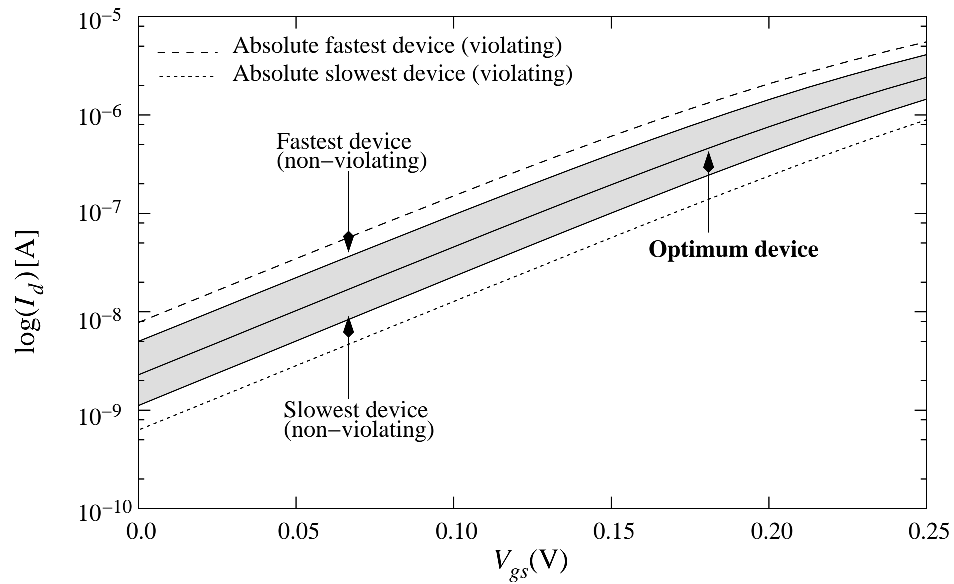

Figure 4.6: I-V characteristics of the $90_{300,5.0}$ device (more relaxed constraints, yield = 98.9%).

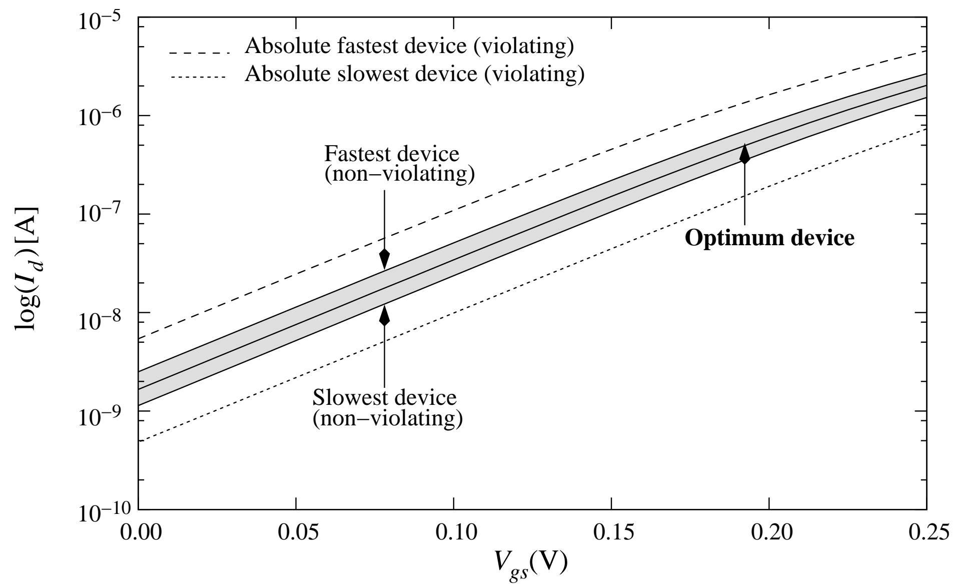

Figure 4.7: I-V characteristics of the $90_{300,2.5}$ device (tighter constraints, yield = 84.0%).

4.3 Optimizing a Subthreshold Transistor

Designing a subthreshold transistor requires the constraint values and the technology-specific variances of the device parameters. The new technique proposed to find the optimum device design parameters by a simple but efficient automatic framework. Figure 3.6 in Section 4 illustrates the steps in the framework. The previous discussion indicates that it is possible to obtain devices whose yield values are greater the 91% average ($90nm$ technology devices) by setting up an appropriate balance between the constraint values the total leakage and intrinsic delay. As an example, devices for $90nm$ are optimized for specific applications; that is, the $90_{300,5.0}$ device is appropriate for *general applications* to construct subthreshold circuits with a balance between the total leakage power and delay, $I_{off} = 2.2nA/\mu m, \tau = 161ps$. The device $90_{300,2.5}$ is good fit for subthreshold designs for *low power*, that is, $I_{off} = 1.6nA/\mu m$ with a sacrifice in speed, $\tau = 210ps$. Finally, for the *high speed* cases, the $90_{200,5.0}$ device provides an intrinsic delay $\tau = 129ps$ with an increase in I_{off}, equal to $2.9nA/\mu m$.

4.4 Summary

In this chapter the optimized transistors for digital subthreshold operation in $90nm$ and $65nm$ technologies are proposed. By finding the appropriate values for the oxide thickness, gate length, and channel doping concentration, sample devices satisfied the total leakage and intrinsic delay constraints, where technology-specific variances of the device parameters are considered. The resultant optimum device parameters for $90nm$ demonstrated that the oxide thickness values are close ($\approx 2.3nm$) to maintain a low gate current ($< 0.15pA/\mu m$). In addition, the leakage components show that $I_{off_{sub}}$ is the dominant current in the optimized devices, whereas I_{BTBT} can be completely ignored due to the sim-

plified uniform doping profile of the devices. The tradeoff between the performance and leakage requirements, I_{on}/I_{off} is satisfied by achieving the original projections (around 1,000). However, for the $65nm$ technology device is not obtainable, since $I_{off_{sub}}$ increases in a greater proportion than the I_{on} growth.

Chapter 5

Conclusion and Future Work

In this thesis, the focus is on process variations and ultra-low-power device design, challenging obstacles in modern IC development. On the one hand, subthreshold systems are a compelling approach to applications where, power reduction is the primary goal. On the other hand, the process variations are further accentuated in the regime. Therefore, variability-aware design strategies at all levels of abstraction device, circuit, and architecture, are imperative to ensure the success and functionality of power-efficient designs.

This thesis describes a novel device level technique to optimize transistors exclusively for digital subthreshold operation. By finding the appropriate values of the oxide thickness, gate length, and channel doping concentration, a MOS device is optimized for the subthreshold regime in terms of the desired total leakage and intrinsic delay constraints, where the device design-parameter variations are considered. The technique is technology scalable and can be adapted to any number of design parameters, technology process variances, statistical distributions, and design constraints. Moreover, since the approach is at the *device level*, circuit and architecture techniques should be applied to further mitigate process variations in subthreshold designs. It is worthy of noting that as the optimization

process occurs at device level, designers have just to focus on circuit and architecture issues, in addition, as the structure of the optimum devices is considered for subthreshold activity it is expected a poor operation in the strong inversion regime. It appears that the technique, is the first variability-aware design approach at device level for subthreshold devices. Sample optimized devices for $90nm$ and $65nm$ technologies are tested, and Monte Carlo simulations are employed to verify the process variations robustness and constraint satisfaction of the optimized transistors. The resultant optimum design parameters, oxide thickness, gate length, and channel doping concentration are comparable to actual parameters of standard devices. This indicates that it should be easy to tape-out the optimized subthreshold devices by modern lithography technologies with the advantage of a more simplified fabrication process.

Since there are no industrial subthreshold devices to compare the results with those of the proposed devices, the next step should be to build basic circuits with the proposed subthreshold devices. The objective is to compare issues such as variation immunity, power, and performance with respect to constructed subthreshold circuits with standard devices. In addition, a complete co-design, at all levels of hierarchy (device, circuit, and architecture) should further suppress the process variation effects, reduce the power consumption, and improve the performance.

Bibliography

[1] Medici user's manual (Synopsis, Inc) version X-2005.10, October 2005.

[2] T.F. Colemen and Y. Zhang, Optimization Toolbox for use with Matlab. The Mathworks, inc. 2005.

[3] M. Anis. Subthreshold leakage current: challenges and solutions. In *Proceedings of the International Conference on Microelectronics*, pages 77–80, December 2003.

[4] M. Anis and M. H. Aburahma. Leakage current variability in nanometer technologies. In *International Workshop on System-on-Chip for Real-Time Applications*, pages 60–63, July 2005.

[5] M. Anis and M. Elmasry. *Multi-Threshold CMOS Digital Circuits: Managing Leakage Power*. Kluwer, Norwell, MA, USA, 2003.

[6] D. A. Antoniadis, I. J. Djomehri, K. M. Jackson, and S. Miller. "Well-tempered" Bulk-Si NMOSFET device home page: http://www-mtl.mit.edu/researchgroups/Well/.

[7] Semiconductor Industry Association. International technology roadmap for semiconductors ITRS, http://www.itrs.net.

[8] S. Borkar. Design challenges of technology scaling. *IEEE Micro*, 19:23–29, July 1999.

[9] S. Borkar. Low power design challenges for the decade. In *Proceedings of the ASP-Design Automation Conference*, pages 293–296, January 2001.

[10] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De. Parameter variations and impact on circuits and microarchitecture. In *Proceedings Design Automation Conference*, pages 338–342, June 2003.

[11] K. Bowman, S. Duvall, and J. Meindl. Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration. *IEEE Journal of Solid State Circuits*, 37:183–190, February 2002.

[12] J. B. Burr and A. M. Peterson. Ultra low power CMOS technology. In *NASA VLSI Design Symposium*, pages 4.2.1–4.2.13, October 1991.

[13] B. H. Calhoun and A. P. Chandrakasan. A 256kb sub-threshold sram in 65nm CMOS. In *IEEE International Solid-State Circuits Conference*, pages 628–629, February 2006.

[14] B. H. Calhoun and A. P. Chandrakasan. A 256-kb 65-nm sub-threshold sram design for ultra-low-voltage operation. *IEEE Journal of Solid-State Circuits*, 42:680–688, March 2007.

[15] B. H. Calhoun, A. Wang, and A. Chandrakasan. Modeling and sizing for minimum energy operation in subthreshold circuits. *IEEE Journal of Solid-State Circuits*, 40:1778–1786, September 2005.

[16] B. H. Calhoun, A. Wang, and A. P. Chandrakasan. Device sizing for minimum energy operation in subthreshold circuits. In *IEEE Custom Integrated Circuits Conference*, pages 95–98, October 2004.

[17] B.H. Calhoun, D.C. Daly, Naveen Verma, D.F. Finchelstein, D.D. Wentzloff, A. Wang, Seong-Hwan Cho, and A. Chandrakasan. Design considerations for ultra-low energy wireless microsensor nodes. *IEEE Transactions on Computers*, 54:727–740, June 2005.

[18] P.K. Chatterjee, W.R. Hunter, T.C. Holloway, and Y.T. Lin. The impact of scaling laws on the choice of n-channel or p-channel for mos VLSI. *IEEE Electron Device Letters*, 1:220–223, October 1980.

[19] R.H. Dennard, F.H. Gaensslen, V.L. Rideout, E. Bassous, and A.R. LeBlanc. Design of ion-implanted mosfet's with very small physical dimensions. *IEEE Journal of Solid-State Circuits*, 9:256–268, October 1974.

[20] S.G. Duvall. Statistical circuit modeling and optimization. In *International Workshop on Statistical Metrology*, pages 56–63, June 2000.

[21] C.C. Wu et. al. A 90-*nm* CMOS device technology with high-speed, general-purpose, and low-leakage transistors for system on chip applications. In *IEEE International Electron Devices Meeting*, pages 65–68, December 2002.

[22] K. H. Fung et. al. 65*nm* CMOS high speed, general purpose and low power transistor technology for high volume foundry application. In *IEEE Symposium on VLSI Technology*, pages 92–93, December 2004.

[23] D. Foty. Perspectives on scaling theory and CMOS technology - understanding the past, present, and future. In *International Conference on Electronics, Circuits and Systems*, pages 631–637, December 2004.

[24] J. Frenkil. A multi-level approach to low-power IC design. *IEEE Spectrum*, 35:54–60, February 1998.

[25] S. Hanson, B. Zhai, K. Bernstein, D. Blaauw, A. Bryant, L. Chang, K. K. Das, W. Haensch, E. J. Nowak, and D. M. Sylvester. Ultralow-voltage, minimum-energy CMOS. *IBM Journal of Research and Development*, 50:469–490, July 2006.

[26] J. Jaffari and M. Anis. Variability-aware device optimization under i_{ON} and leakage current constraints. In *International Symposium on Low Power Electronics and Design (ISLPED 06)*, Tegernsee, Germany, October 2006.

[27] N. Jayakumar and S. P. Khatri. A variation-tolerant sub-threshold design approach. In *Design Automation Conference (DAC 05)*, pages 716–719, Anaheim, California, USA, June 2005.

[28] J. Kao. *Subthreshold Leakage Control Techniques for Low Power Digital Circuits*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, May 2001.

[29] J. Kao, S. Narendra, and A. Chandrakasan. Subthreshold leakage modeling and reduction techniques. In *IEEE/ACM International Conference on Computer Aided Design*, pages 141–148, November 2002.

[30] A. Keshavarzi, K. Roy, and C. F. Hawkins. Intrinsic leakage in deep submicron CMOS ICs-measurement-based test solutions. *IEEE Transactions on Very Large Scale Integration Systems*, 8:717–723, December 2000.

[31] C. H. Kim, H. Soeleman, and K. Roy. Ultra-low-power DLMS adaptive filter for hearing aid applications. *IEEE Transactions on Very Large Scale Integration Systems*, 11:1058–1067, December 2003.

[32] T. Kim, H. Eom, J. Keane, and C. Kim. Utilizing reverse short channel effect for optimal subthreshold circuit design. In *International Symposium on Low Power Electronics and Design 2006*, pages 127–130, October 2006.

[33] P. Kumaraswamy. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46:79–88, March 1980.

[34] J. B. Kuo and J. H. Lou. *Low-Voltage CMOS VLSI Circuits*. Wiley, New York, NY, USA, 1999.

[35] J. Kwong and A. P. Chandrakasan. Variation-driven device sizing for minimum energy sub-threshold circuits. In *International Symposium on Low Power Electronics and Design*, pages 8–13, October 2006.

[36] Z. H. Liu, C. Hu, J. H. Huang, T. Y. Chan, M. C. Jeng, P. K. Ko, and Y. C. Cheng. Threshold voltage model for deep-submicrometer MOSFETs. *IEEE Transactions on Electron Devices*, 40:86–95, January 1993.

[37] H. Mahmoodi, S. Mukhopadhyay, and K. Roy. Estimation of delay variations due to random-dopant fluctuations in nanoscale CMOS circuits. *IEEE Journal of Solid-State Circuits*, 40:1787–1786, September 2005.

[38] C. Mead. *Analog VLSI and neural systems*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

[39] L. A. P. Melek, M. C. Schneider, and C. Galup-Montoro. Body-bias compensation technique for subthreshold CMOS static logic gates. In *Symposium on Integrated Circuits and Systems Design*, pages 267–272, September 2004.

[40] G. Moore. Progress in digital integrated electronics. In *International Electron Devices Meeting*, pages 11–13, December 1975.

[41] S. Mukhopadhyay, A. Raychowdhury, and K. Roy. Accurate estimation of total leakage in nanometer-scale bulk CMOS circuits based on device geometry and doping profile.

Transactions on Computer-Aided Design of Integrated Circuits and Systems, 24:363–381, March 2005.

[42] Y. Nakahar, K. Takeuchi, T. Tatsumi, Y Ochiai, S. Manako, S. Samukawa, and A. Furukawa. Ultra-shallow in-situ-doped raised source/drain structure for sub-tenth micron CMOS. In *Symposium on VLSI Technology, 1996. Digest of Technical Papers. 1996*, pages 174–175, June 1996.

[43] S.R. Nassif. Delay variability: sources, impacts and trends. In *IEEE International Solid-State Circuits Conference*, pages 368–369, February 2000.

[44] L. Nazhandali, B. Zhai, J. Olson, A. Reeves, M. Minuth, R. Helfand, S. Pant, T. Austin, and D. Blaauw. Energy optimization of subthreshold-voltage sensor network processors. In *Proceedings of the 32nd Annual International Symposium on Computer Architecture*, pages 197–207, May 2005.

[45] M. Ono, M. Saito, T. Yoshitomi, C. Fiegna, T. Ohguro, and H. Iwai. Sub-50 nm gate length n-MOSFETs with 10 nm phosphorus source and drain junctions. In *International Electron Devices Meeting, 1993 Technical Digest*, pages 119–122, December 1993.

[46] B. C. Paul, A. Raychowdhury, and K. Roy. Device optimization for digital subthreshold logic operation. *IEEE Transactions on Electron Devices*, 52:237–247, February 2005.

[47] B. C. Paul, H. Soeleman, and K. Roy. An 8x8 subthreshold digital CMOS carry save array multiplier. In *Proceedings of the 27th European Solid-State Circuits Conference*, pages 377–380, September 2001.

[48] B.C. Paul and K. Roy. Device optimization for digital sub-threshold operation. In *Device Research Conference, 2004. 62nd DRC. Conference Digest [Late News Papers volume included]*, pages 113–114, June 2004.

[49] K. Ponnambalam, Abbas Seifi, and Jiri Vlach. Probabilistic design of systems with general distributions parameters. *International Journal of Circuit Theory and Applications*, 29.

[50] J. M. Rabaey, A. Chandrakasan, and B. Nikolic. *Digital Integrated Circuits A Design Perspective.* Prentice Hall, Upper Saddle River, NJ, USA, 2003.

[51] R. Rao, A. Srivastava, D. Blaauw, and D. Sylvester. Statistical analysis of subthreshold leakage current for VLSI circuits. *IEEE Transactions on Very Large Scale Integration Systems*, 12:131–139, February 2004.

[52] A. Raychowdhury, S. Mukhopadhyay, and K. Roy. Modeling and estimation of leakage in sub-90nm devices. In *International Conference on VLSI Design*, pages 65–70, January 2004.

[53] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. *Proceedings of the IEEE*, 91:305–327, February 2003.

[54] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. In *Proceedings of the IEEE*, pages 305–327, February 2003.

[55] S. Rusu. Trends and challenges in VLSI technology scaling towards 100nm. In *Proceedings of the Solid-State Circuits Conference*, pages 194–196, September 2001.

[56] K.F. Schuegraf and H. Chenming. Hole injection s_io_2 breakdown model for very low voltage lifetime extrapolation. *IEEE Electron Devices*, 5:761–767, May 1994.

[57] J.M. Soden, C.F. Hawkins, and A.C. Miller. Identifying defects in deep-submicron CMOS ICs. *IEEE Spectrum*, 33:66–71, September 1996.

[58] H. Soeleman. *Ultra-Low Power Digital Sub-Threshold Logic Design.* PhD thesis, Purdue University, West Lafayette, IN, USA, December 2000.

[59] H. Soeleman and K. Roy. Ultra-low power digital subthreshold logic circuits. In *International Symposium on Low Power Electronics and Design (ISLPED 99)*, pages 94–96, San Diego, California, USA, July 1999.

[60] H. Soeleman and K. Roy. Digital CMOS logic operation in the sub-threshold region. In *Tenth Great Lakes Symposium on VLSI*, pages 107–112, Chicago, Illinois, USA, March 2000.

[61] H. Soeleman, K. Roy, and B. Paul. Sub-domino logic: ultra-low power dynamic sub-threshold digital logic. In *International Conference on VLSI Design*, pages 3–7, January 2001.

[62] H. Soeleman, K. Roy, and B. C. Paul. Roboust subthreshold logic for ultra-low power operation. *IEEE Transactions on Very Large Scale Integration Systems*, 9:90–99, February 2001.

[63] A. Srivastava, R. Bai, D. Blaauw, and D. Sylvester. Modeling and analysis of leakage power considering within-die process variations. In *International Symposium on Low Power Electronics and Design 2002*, pages 64–67, August 2002.

[64] A. Srivastava, D. Sylvester, and D. Blaauw. *Statistical Analysis and Optimization for VLSI: Timing and Power.* Springer, New York, NY, USA, 2005.

[65] Y. Taur and T. H. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, New York, NY, USA, 1998.

[66] S. Thompson, P. Packan, and M. Bohr. MOS scaling: Transistor challenges for the 21st Century. *Intel Technology Journal*, Q3:1–19, September 1998.

[67] O. S. Unsal, J. W. Tschanz, K. Bowman, V. De, X. Vera, A. Gonzlez, and O. Ergin. Impact of parameter variations on circuits and microarchitecture. *IEEE Micro*, 26:30–39, November 2006.

[68] E. Vittoz and J. Fellrath. CMOS analog integrated circuits based on weak inversion operation. *IEEE Journal of Solid-State Circuits*, 12:224–231, June 1977.

[69] E. A. Vittoz. Analog VLSI implementation of neural networks. In *IEEE International Symposium on Circuits and Systems*, pages 2524–2527, May 1990.

[70] A. Wang, B. Calhoun, and A. Chandrakasan. *Sub-Threshold Design for Ultra Low-Power Systems*. Springler, New York, NY, USA, 2006.

[71] A. Wang and A. Chandrakasan. A 180mv fft processor using subthreshold circuit techniques. In *IEEE International Solid-State Circuits Conference*, pages 292–529, February 2004.

[72] A. Wang and A. Chandrakasan. A 180-mv subthreshold FFT processor using a minimum energy design methodology. *IEEE Journal of Solid-State Circuits*, 40:310–319, January 2005.

[73] A. Wang, Anantha P. Chandrakasan, and S. V. Kosonocky. Optimal supply and threshold scaling for subthreshold CMOS circuits. In *Proceedings of the IEEE Computer Society Annual Symposium on VLSI*, pages 5–9, April 2002.

[74] T. Xinghai, V.K. De, and J.D. Meindl. Intrinsic mosfet parameter fluctuations due to random dopantplacement. *IEEE Transactions on Very Large Scale Integration Systems*, 5:369–376, December 1997.

[75] L. D. Yau. A simple theory to predict the threshold voltage of short-channel IGFET's. *Solid State Electronics*, 17:1059–1063, October 1974.

[76] P. M. Zeitzoff and J. E. Chung. A perspective from the 2003 ITRS: MOSFET scaling trends, challenges and potential solutions. *IEEE Circuit & Devices*, 21:4–15, January 2005.

[77] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester. Analysis and mitigation of variability in subthreshold design. In *International Symposium on Low Power Electronics and Design (ISLPED 05)*, pages 20–25, San Diego, California, USA, August 2005.

[78] B. Zhai, L. Nazhandali, J. Olson, A. Reeves, M. Minuth, R. Helfand, S. Pant, D. Blaauw, and T. Austin. A 2.60pj/inst subthreshold sensor processor for optimal energy efficiency. In *Symposium on VLSI Circuits Digest of Technical Papers*, pages 154–155, June 2006.