

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MICROELETRÔNICA

ANDRÉS MAURICIO ASPRILLA VALDÉS

**Design and Evaluation of Logic Gates
Based on IG FinFET**

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Microelectronics

Advisor: Prof. Dr. Renato Perez Ribas

Porto Alegre
October 2016

CIP — CATALOGING-IN-PUBLICATION

Valdés, Andrés Mauricio Asprilla

Design and Evaluation of Logic Gates Based on IG Fin-FET / Andrés Mauricio Asprilla Valdés. – Porto Alegre: PGMI-CRO da UFRGS, 2016.

117 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Microeletrônica, Porto Alegre, BR-RS, 2016. Advisor: Renato Perez Ribas.

1. FinFETs. 2. Independent gates. 3. Logic gate design. 4. Delay analytical model. 5. VLSI. 6. Dual-gate transistors. 7. Low-power logic networks. 8. Logic network compaction. 9. Multi- V_{th} logic. I. Ribas, Renato Perez. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Vladimir Pinheiro do Nascimento

Diretor do Instituto de Informática: Prof. Luis da Cunha Lamb

Coordenadora do PGMICRO: Prof^a. Fernanda Gusmão de Lima Kastensmidt

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*Human behavior flows from three main sources:
desire, emotion, and knowledge. — PLATO*

ACKNOWLEDGMENTS

To complete this work some sacrifices and study hours were required. I want to express my gratitude to my friends and some special people who supported me during this year. Initially, I want to express my gratitude to my parents, Enrique and Luz Marina and my brother, Alberto. They are my inspiration, giving me the necessary courage to continue studying and researching.

Also, I want to express my gratitude to my advisor, prof. Renato Ribas, whose expertise, knowledge and useful comments, remarks and engagement helped me a lot with the development of this work and to be a better researcher. I appreciated all his knowledge and experience in digital design, and his support during the writing of this dissertation.

A very special thanks goes to the lab 214 colleagues: Vinicius Possani and Felipe Marranghello, who helped me with the development of this research work.

To the CEITEC's physical design (PRDIBA) team, specially to Gerson Scartezzini and Gabriel Tadeo, for our conversations, exchanges of knowledge and comprehension along this year while I was writing this work during lunch time and sometimes during work hours.

To my other CEITEC's colleagues: Thiago, Ussamah, Rafael and Guilherme who encouraged me a lot during this time.

I would also like to thank to my friends in Porto Alegre: Sebastian, Luis, David, Nathalia, Claudia, Eduardo and to rest. Specially to Patricia, who also encouraged me and made me laugh when I was worried and sad.

Finally, thanks to CNPq for the financial aid, to the IC-Brazil program and CEITEC for the invaluable experience during these two years. And to Brazil, a beautiful country that warmly welcomed me and gave me the opportunity of studying microelectronics.

ABSTRACT

The CMOS planar technology has been used in fabrication of integrated circuits in the last decades. However, short channel effects in the subthreshold operation region are becoming a critical restriction to the channel length reduction. With the use of FinFET devices, the scaling increases due to the reduction of short channel effects. The origin of the FinFET arises from the scaling limitations of planar devices, reducing the short-channel effects and continuing the scaling predicted by the Moore's Law. A variation of the standard FinFET device is the independent-gate FinFET device (IG FinFET), in which two independently connected gates control an unique channel. In this work, the independent-gate device was explored as a circuit element used for the implementation of different combinational logic networks. With independently connected gates, series/parallel arrays could be performed using a single device, but with direct impact in the electrical performance of logic gates. In this work, it is presented the electrical analysis in terms of signal delay propagation and energy consumption of compacted transistor networks. Different topologies derived from the independent-gate operation were tested through electrical simulations and the results demonstrate the existing trade-off between these two parameters. Also, an analytical delay expression was derived for logic networks which use IG FinFETs, deriving analytical expressions for the impact of reducing arrays of series transistors in logic networks. The analytical model for IG devices was tested in a logic data path and compared to SPICE simulation results, showing its utility for the timing analysis of digital circuits.

Keywords: FinFETs. Independent gates. Logic gate design. Delay analytical model. VLSI. Dual-gate transistors. Low-power logic networks. Logic network compaction. Multi- V_{th} logic.

Avaliação elétrica e modelo de atraso de redes lógicas combinacionais implementadas usando IG FinFETs

RESUMO

A tecnologia CMOS tem sido amplamente usada na fabricação de circuitos integrados durante as últimas décadas. Embora, os efeitos de canal curto na região sub-limiar restringem a diminuição do comprimento do canal. Com o uso de dispositivos FinFET, o escalamento continua devido à redução dos efeitos de canal curto, permitindo manter a tendência predecida pela lei de Moore. Um dispositivo derivado do FinFET, conhecido como IG FinFET, possui propriedades que são úteis no projeto de portas lógicas combinacionais. Com dispositivos de gates independentes (IG), arranjos de transistores série/paralelo podem ser realizadas utilizando um único transistor, porém, existe um impacto no atraso e no consumo das redes lógicas resultantes. Neste trabalho, é apresentada uma análise elétrica de atraso e consumo de redes lógicas compactadas usando dispositivos IG FinFET. Diferentes topologias de implementação derivadas da operação de gates independentes foram testadas por meio de simulações elétricas e os resultados mostram que existe um compromisso entre o consumo de potência e o atraso de propagação das redes resultantes. Também foi realizado um estudo do comportamento transiente, descrevendo analiticamente o impacto do atraso devido à redução do número de transistores. A análise realizada anteriormente, foi utilizada para calcular o atraso do caminho crítico de um circuito lógico, mostrando a sua utilidade na análise de atraso em circuitos digitais.

Palavras-chave: FinFETs, gates independentes, projeto de portas lógicas, modelo analítico de atraso, VLSI, transistores de duas portas, compactação de redes lógicas, redes lógicas de baixo consumo, lógica multi- V_{th} .

LIST OF ABBREVIATIONS AND ACRONYMS

ALU	Arithmetic-Logic Unit
CAD	Computer aided design
CMOS	Complementary metal-oxide-semiconductor
SCE	Short-channel effects
CMC	Compact model council
MG	Multi-gate
DELTA	Fully depleted lean-channel transistor
MFXMOS	Multi-fin XMOS
VeSFET	Vertical-slit field-effect transistor
FinFET	Fin field-effect transistor
MIGFET	Multiple independent gate field-effect transistor
MUGFET	Multiple gate field-effect transistor
MOSFET	Metal-oxide-semiconductor Field-effect transistor
FEOL	Front-end-of-line
IG	Independent gate mode/device
IG-LP	Independent gate-low power mode
IG-RED	Independent gate reduced mode
SG	Shorted-gate mode/device
HVT	High- V_{th} device
GAA	Gate-all-around
ITRS	International technology roadmap for semiconductors
PTM	Predictive technology model
TCAD	Technology CAD
VLSI	Very large scale integration

SOI	Silicon on Insulator Fin Field-effect transistor
PDSOI	Partially depleted silicon on insulator field-effect transistor
FDSOI	Partially depleted silicon on insulator field-effect transistor
UTBSOI	Ultra thin body silicon on insulator field-effect transistor
GIDL	Gate-induced drain lowering
DIBL	Drain-induced bias lowering
SiNW	Silicon-nano-wire
DG	Double-gate device
SDG	Symmetric double-gate device
ADG	Asymmetric double-gate device
DIBL	Drain-induced bias lowering
BSIM	Berkeley short-channel IGFET model
SPICE	Simulation program with integrated circuit emphasis
SCCT	Short-circuit charge transferred
PUN	Pull-up network
PDN	Pull-down network
FA	Full adder
PN	P-type, n-type junction
RCA	Ripple carry adder
PPA	Parallel-prefix adder
CLA	Carry look ahead

LIST OF FIGURES

Figure 2.1 Total power dissipation of an Integrated Circuit as a function of the technology node	19
Figure 2.2 Current behavior of long and short-channel devices I_D versus gate to source voltage V_{GS}	20
Figure 2.3 Energy bands for long and short-channel N-Type MOSFETs	22
Figure 2.4 Electric field distribution along the channel for a short-channel device for low and high V_{DS}	23
Figure 2.5 Drain current shift caused by DIBL effect	23
Figure 2.6 Band to band tunneling in a PN junction.....	24
Figure 2.7 V_{th} shift caused by GIDL effect	25
Figure 2.8 a) a PDSOI device and b) a FDSOI device.....	26
Figure 2.9 Double gate transistor structures built on silicon wafers. a) planar double-gate, b) fin structure, c) vertical double gate device, with perpendicular channel	27
Figure 2.10 DELTA MOSFET device	28
Figure 2.11 Multiple Gate device evolution and families	29
Figure 2.12 Double gate transistor types: symmetric and assymatric	29
Figure 2.13 Electron density for SDG transistors for different channel thicknesses	30
Figure 2.14 Mobility dependence on silicon thickness for a double-gate MOSFET	31
Figure 2.15 a) a triple-gate FinFET and b) a double-gate FinFET	32
Figure 2.16 a) a SOI FinFET and b) bulk FinFET	32
Figure 2.17 SOI vs. Bulk FinFET leakage current for different gate lengths.....	33
Figure 2.18 Body effect on SOI and Bulk FinFETs, compared with a planar device.....	34
Figure 2.19 Silicon Nanowire Gate-all-around MOSFET device.....	34
Figure 2.20 The first silicon Nanowire Gate-all-around MOSFET device	35
Figure 2.21 DIBL and Subthreshold slope of cylindrical GAA devices.....	36
Figure 2.22 Independent Gate FinFET fabrication process	37
Figure 2.23 a) Shorted Gate FinFET (all gates tied together), b) Independent Gate FinFET (gates connected separately).....	37
Figure 2.24 Independent Gate FinFET operation: when the channel is controlled by one of the gates and when the channel is controlled by the two gates connected simultaneously	38
Figure 2.25 Threshold voltage (V_{th}) variation with respect to second gate biasing (V_{G_2})	41
Figure 2.26 Drain current (I_D) variation with respect to second gate biasing (V_{G_2})	41
Figure 2.27 Independent Gate FinFET cross sections: a) low- V_{th} b) high- V_{th}	44
Figure 2.28 Devices supported by BSIM-CMG compact model	45
Figure 2.29 Devices supported by BSIM-IMG compact model	46
Figure 2.30 IG-FinFET model scheme for SG to IG FinFET conversion	47
Figure 2.31 UFDG equivalent circuit model.....	48
Figure 3.1 A standard CMOS logic function implementation	52
Figure 3.2 A standard CMOS logic function implementation for 2-input NAND gate..	53
Figure 3.3 A NAND2 logic gate implemented with IG FinFET in single-gate mode (SG).....	53
Figure 3.4 A NAND2 logic gate implemented with IG FinFET in LP mode (IG-LP) ...	53
Figure 3.5 A NAND2 logic gate implemented with IG-FinFET, merging a and b inputs into a single P-type device	54

Figure 3.6 A NAND2 logic gate implemented with IG-FinFET devices in reduced mode (IG-RED)	54
Figure 3.7 A NOR2 logic gate implemented with IG-FinFET devices in reduced mode (IG-RED)	55
Figure 3.8 Proposed test bench for electrical evaluation of transistor stacks	56
Figure 3.9 Results for delay of stack transistors for arrays of 1 to 4 series IG FinFETs	56
Figure 3.10 a) Standard CMOS implementation of a tri-state buffer b) reduced version using IG-FinFETs.....	57
Figure 3.11 a) Standard CMOS implementation of a two-input multiplexer b) its reduced version using IG-FinFETs.....	58
Figure 3.12 a) Standard CMOS implementation of a two-input XOR logic gate b) its reduced version using IG-FinFETs	59
Figure 3.13 a) Standard CMOS implementation of a full-adder logic gate b) its reduced version using IG-FinFETs.....	60
Figure 3.14 A logic circuit implemented with: a) FinFETs in SG mode, and, b) reduced network using FinFETs in IG mode and defactorization proposed by (ROSTAMI; MOHANRAM, 2011)	61
Figure 3.15 A series logic network with: a) FinFETs in SG mode, and, b) reduced network using FinFETs in IG mode and defactorization proposed by (ROSTAMI; MOHANRAM, 2011).....	61
Figure 3.16 A series logic network with: a) FinFETs in SG mode, and, b) reduced network using high- V_{th} FinFETs and merging series transistors	62
Figure 3.17 A series logic network with: a) FinFETs in SG mode, b) reduced network using defactorization (ROSTAMI; MOHANRAM, 2011), and c)Reduced version with low- V_{th} FinFETs and merging parallel transistors directly	62
Figure 3.18 Binary tree of function F	63
Figure 3.19 a) Function F optimized using the method proposed by (POSSANI et al., 2016) and b) its implementation using IG-FinFETs	63
Figure 3.20 Ring oscillator used to estimate signal delay propagation and power consumption of the logic gates, design under test (DUT)	65
Figure 3.21 Energy delay product of each function.....	66
Figure 4.1 Static transfer curve (V_{out} vs. V_{in} of a CMOS inverter)	72
Figure 4.2 Measured ramp response of a CMOS inverter ($V_o(t)$ vs. $V_{in}(t)$)	72
Figure 4.3 IG FinFET inverter modes for delay analysis.....	77
Figure 4.4 I_D vs. V_{DS} curves simulated and modeled using α -power for an N-type IG-FinFET device	85
Figure 4.5 IG FinFET inverter modes for delay analysis.....	86
Figure 4.6 Delay response (simulated and modeled) for inverters (minimum dimensions) in different modes (SG,IG and high- V_{th}), varying the input slope (τ_{in}). Thick lines represent the slow input response.....	88
Figure 4.7 Delay response (simulated and modeled) for inverters in IG mode, varying the input slope (τ_{in}) and the N_{FIN_p}/N_{FIN_n} ratio. Thick lines represent the slow input response.....	89
Figure 4.8 Delay response (simulated and modeled) for inverters in IG mode, varying the input slope (τ_{in}) and the load capacitance (C_L). Thick lines represent the slow input response.....	89
Figure 4.9 Delay arcs of a standard CMOS NAND3 logic gate	90
Figure 4.10 a) NAND3 logic gate using IG-LP mode, and b) the equivalent inverter formed by the equivalent transistor with $K_{n_{sat}}$	91
Figure 4.11 Tested topologies for delay estimation for a NAND4	92

Figure 4.12 a) NAND3 logic gate using IG-LP mode, and b) the equivalent inverter formed by the equivalent transistor with $K_{n_{sat}}$	95
Figure 4.13 Tested topologies for delay estimation for a NAND4	95
Figure 4.14 IG FinFET implementation for the function $F = a + \overline{bc}$	96
Figure 4.15 Equivalent inverters for each input	97
Figure 4.16 Rise time response for different inverters built using IG FinFETs, varying the input ramp slope (τ_{in}) and $N_{FINp}/N_{FINn} = 1$ ratio.	99
Figure 4.17 Fall time response for different inverters built using IG FinFETs, varying the input ramp slope (τ_{in}) and $N_{FINp}/N_{FINn} = 1$ ratio.	100
Figure 5.1 Block diagram of a standard Ripple Carry Adder architecture	102
Figure 5.2 Parallel Prefix Adder block diagram.....	103
Figure 5.3 Propagate and generate signal implementation using IG-FinFETs	104
Figure 5.4 Prefix operator used by PPA adder to calculate the group propagate and group generate signals.....	104
Figure 5.5 Prefix operator circuit implementation using IG FinFET: AND gate built using 4 devices, and AND-OR gate built using 6 devices	105
Figure 5.6 Kogge-Stone prefix tree implementation and the critical path indication. ..	105
Figure 5.7 8-bit Kogge-Stone adder critical path circuit.....	107
Figure 5.8 Proposed algorithm to calculate the data path delay	108
Figure 5.9 Equivalent inverters for the AND2 logic gate	109
Figure 5.10 Equivalent inverters for the AND-OR logic gate.....	109
Figure 5.11 Equivalent inverters for the XOR2 logic gate.....	109

LIST OF TABLES

Table 2.1 Process and wafer cost comparative between SOI and bulk FinFET	32
Table 2.2 Differences in physical and electrical dimensions between high- V_{th} and low- V_{th} IG-FinFETs	44
Table 2.3 Comparison between available electrical simulation models.....	49
Table 2.4 Relevant SPICE parameters used for electrical simulation of IG-FinFET devices.....	49
Table 3.1 Naming conventions for IG-FinFET devices	51
Table 3.2 Summary of possible IG-FinFET modes for the NAND2 logic gate	55
Table 3.3 Device count for F_1 to F_{10} logic functions implementation using T_1 to T_5 implementation modes	59
Table 3.4 Logic functions implementation results for power, energy and device count using the different implementation modes.....	67
Table 4.1 Alpha-power parameters for IG FinFETs operating in SG, IG and High- V_{th} devices	84
Table 4.2 Extracted capacitance values for IG FinFETs, for the different operating modes	87
Table 4.3 Comparison between simulation and delay model for series transistors	93
Table 4.4 Comparison between simulation and delay model for series transistors, using a 1 fF load and a 10 ps input ramp	96
Table 4.5 Comparison between simulation and delay model for a logic gate, using a 1 fF load and a 10 ps input ramp.....	97
Table 5.1 Comparison between calculations and SPICE simulations for delay the critical path of the 8-bit PPA circuit.....	110

CONTENTS

1 INTRODUCTION.....	15
1.1 Motivation.....	15
1.2 Objectives.....	16
1.3 Thesis organization	16
2 INDEPENDENT DUAL-GATE FINFET	18
2.1 Introduction.....	18
2.2 Short-Channel Effects in conventional MOSFET transistors	18
2.2.1 Subthreshold swing	20
2.2.2 Drain-Induced Barrier Lowering (DIBL)	22
2.2.3 Gate-Induced Drain Leakage (GIDL).....	24
2.3 Alternatives to reduce the Short Channel Effects.....	24
2.3.1 Silicon Over Insulator transistor (SOI)	26
2.3.2 Multiple gate devices	27
2.3.2.1 Double-gate MOSFET (DG).....	27
2.3.2.2 FinFET	31
2.3.2.3 Gate-all-around device (GAA).....	33
2.4 Independent Gate FinFET	36
2.4.1 Basic Independent-gate device physics.....	38
2.5 Simulation models.....	44
2.5.1 BSIM-CMG model	44
2.5.2 BSIM-IMG Model	45
2.5.3 IG-FinFET Model	46
2.5.4 UFDG model.....	47
2.6 Chosen model and SPICE parameters.....	48
2.7 Summary.....	50
3 IG FINFET BASED LOGIC GATES	51
3.1 Introduction.....	51
3.2 Naming conventions	51
3.3 Basic IG FinFET usage in pull-up/pull-down networks.....	52
3.4 Delay evaluation of stacked transistor networks	55
3.5 Transistor network compactations	57
3.5.1 Compacting standard logic gates	57
3.5.2 Compacting logic complex functions using de-factorization	58
3.6 Electrical evaluation of compacted networks.....	64
3.6.1 Methodology	65
3.6.2 Results.....	66
3.7 Summary.....	68
4 DELAY MODELING OF IG-FINFET LOGIC GATES	69
4.1 Introduction.....	69
4.2 IG FinFET inverter delay model	71
4.2.1 Static behavior of CMOS inverter.....	71
4.2.2 Dynamic behavior of CMOS inverter	71
4.2.2.1 Overshoot region.....	73
4.2.2.2 Short-circuit region	73
4.2.2.3 Discharge region	73
4.2.3 Delay definition.....	74
4.2.4 Transistor current equations.....	74
4.2.5 Inverter virtual step response	75

4.2.6 Inverter ramp response	77
4.2.6.1 Fast input ramps	78
4.2.6.2 Fast and slow input reference time (t_{ref})	80
4.2.6.3 Slow input ramps	81
4.3 Inverter delay analysis for IG-FinFETs.....	83
4.3.1 Transistor parameters	83
4.3.2 Physical parameters	84
4.3.3 Delay evaluation of inverter gates.....	88
4.4 Delay of series connected IG-FinFETs.....	90
4.4.1 Switching device order	91
4.4.1.1 Top switching transistor.....	91
4.4.1.2 Bottom switching transistor.....	93
4.4.1.3 Delay analysis of complex logic gates.....	96
4.5 Transition time modeling	98
4.5.1 Transition time for an inverter	98
4.6 Summary.....	99
5 DELAY ANALYSIS OF IG FINFETS IN LOGIC DATA PATHS	101
5.1 Introduction.....	101
5.2 Parallel Prefix Adder (PPA)	101
5.3 Critical path delay analysis	106
5.4 Results	109
5.5 Summary.....	110
6 CONCLUSIONS	111
6.1 Future work.....	111
REFERENCES.....	113

1 INTRODUCTION

1.1 Motivation

The scaling of MOS technology has been a major factor of contributing to the continuous improvement of circuit performance. However, aggressive shrinking of MOS transistor dimensions led to severe degradations of physical properties of these devices. In order to maintain such an electronic scaling, novel devices presenting improved electrical characteristics have been developed. In this way, double-gate devices, such as the FinFET one, has attained great success in replacing the standard planar MOSFET due to better electrostatic control that decreases leakage current and reduces short-channel effects. Consequently, it has been possible to scale these devices to dimensions that were unlikely to be achieved with conventional planar fabrication process.

The standard FinFET used in integrated circuits is a double-gate or triple-gate structure, built over SOI or bulk substrates. In such devices the gates are physically tied together, acting as a planar MOS device. Despite the differences in electrical performance, transistor arrangements using standard FinFETs are very similar to those ones built in planar technology.

Another FinFET structure is the independent-gate (IG) FinFET. In this case, each gate is controlled independently to the other. The utilization of IG device may open new possibilities of transistor network topologies, not suitable when using only planar or standard FinFET devices. These novel structures have motivated new investigations on how to exploit efficiently IG FinFET technology in VLSI design.

The utilization of IG FinFET may reduce the number of devices required to build logic gates, suggesting that the use of IG FinFET may lead to an interesting tradeoff between power consumption and delay when compared to planar transistors or standard FinFETs.

The reduction on the transistor count in a logic gate arises from the possibility to implement OR and AND operations using a single device. To perform the OR operation using a single device, a transistor with a low threshold voltage is applied. Thus, if at least one of the gates is active, a conducting channel is formed regardless of the other transistor gate state. Naturally, whether both gates are active, the inversion level at the channel is even higher. On the other hand, in order to obtain the AND behavior through a single device, a transistor with a higher threshold voltage is adopted. In this case, both gates

must be active for the creation of a conducting channel.

In this work, the impact of reducing the number of transistors of logical networks built using IG FinFETs is studied, by initially showing the basic operation principle of independent double-gate devices. Also, different logic functions are implemented using different topologies derived from the use of IG FinFET devices, showing the impact on the delay and power consumption of each evaluated topology through electrical simulations. Furthermore, the transient behavior of the output response of IG FinFET logic networks is modeled for a variety of input ramps, output capacitances and stacks of series transistors, showing its application for evaluating the delay of the critical path of a logic circuit.

1.2 Objectives

The main objective of this work is to explore the different implementation alternatives of logic networks using IG FinFETs. Showing different topologies derived from the use of these devices in logic networks and also, observe and quantify the performance effects the impact on the performance caused by the compaction of logic networks implemented with such kind of devices through simulations and analytical models. This could be achieved through the following steps:

- Set-up an electrical simulation environment, capable of showing the behavior of IG FinFET devices and the gate coupling.
- Electrical transient simulation of different logic networks built using IG FinFET devices, measuring their performance in terms of delay and power consumption.
- Determination of analytical equations for delay and transition times of different IG FinFET-based logic structures in order to quantify the delay impact of logic network compactions.

1.3 Thesis organization

This work is organized as follows:

- Chapter 2 gives an introduction to IG FinFET devices. Showing the origin of the multiple-gate devices and the basic physical principle of a thin channel device operated by two independent gates. Also, the different SPICE simulation models created

for simulating IG FinFET devices are shown, the chosen model and its simulation parameters.

- In chapter 3, the IG FinFET device is explored as a circuit element, creating different logic network topologies taking advantage of the device properties for reducing the number of required transistors for implementing logic networks and evaluating their impact in terms of power and delay of that reduction.
- Chapter 4 shows analytically shows the delay and transition time behavior of different logic structures built using IG FinFET devices using charge-based analytical models for a variety of input transitions and output capacitances.
- Chapter 5 applies the analytical model shown in the previous chapter for analyzing the delay response of a critical path, using a parallel-prefix adder (PPA) as an example.

2 INDEPENDENT DUAL-GATE FINFET

2.1 Introduction

MOS planar technology has been used in fabrication of integrated circuits (ICs) in the last decades. However, the short-channel effects in the subthreshold operation region are becoming a critical restriction to the channel length reduction. With the use of multiple-gate field-effect transistors (MUGFETs), the scaling increases due the reduction of short-channel effects (SCE). One of the most important type of MUGFET is the FinFET device which offers the possibility of independent gate controlling, giving more useful features for digital IC design. In this chapter, an introduction to double-gate devices and FinFETs is presented, showing the different electrical properties and behavior of those transistors. After that, the basic operation principle of independent gate transistors is presented, showing the different configurations, threshold voltage behavior and their utility in combinational logic gates. Then, some compact simulation models useful for simulating IG-FinFETs are shown and some conclusions are drawn.

2.2 Short-Channel Effects in conventional MOSFET transistors

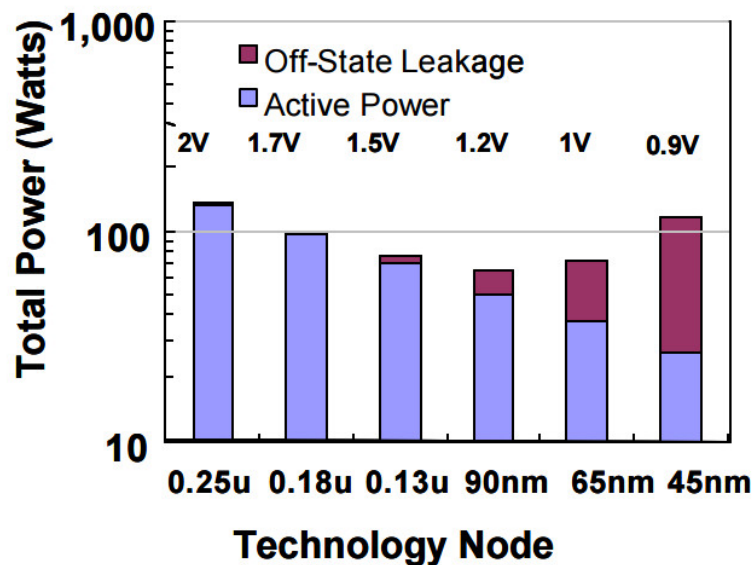
In digital circuits, the main application of Field-Effect Transistors (FET) is to act as a logic switch, in which the drain current (I_D) is controlled by the gate to source voltage (V_{GS}). To build faster switches, it is necessary to create smaller devices in order to improve the response time of the device. Trough the technology scaling, more transistors can be integrated and hence the circuit complexity increases. In 1965, Moore's law predicted that the transistor density in an IC will quadruple every four years (MOORE, 1965).

The Moore's law became difficult to be achieved as the feature size of the transistors started to be lesser than $1\mu\text{m}$, and some undesirable effects started to influence the transistor behavior because the source and drain terminals of the transistors become to be closer. As a consequence of that, the gate terminal started to lose the control of the channel and undesirable effects appeared and influenced negatively on the transistor behavior. Those undesirable effects were called "short-channel effects" (SCE) and clearly affect the transistor miniaturization. Basically, short-channel effects are seen when the controllability of the channel region by the gate is affected by electric field lines coming from source

and drain regions (COLINGE, 2007).

For an ideal switch, the "off" state current (or leakage current) is zero and the "on" state current is high, being its I_{ON}/I_{OFF} ratio an infinite value. Using a MOS transistor as a logic switch, the I_{ON}/I_{OFF} ratio is finite and considerably low for standard submicron devices. Fig. 2.1 shows the evolution of the power consumption for different feature sizes, showing a high leakage power consumption for smaller feature sizes. In order to reduce the leakage consumption in conventional CMOS devices, new fabrication techniques needed to be applied to decrease the "off" state current and to achieve a better I_{ON}/I_{OFF} ratio (CHUANG et al., 2004).

Figure 2.1: Total power dissipation of an Integrated Circuit as a function of the technology node



Source: Doyle et al. (2002)

As shown in Fig. 2.1 and mentioned previously, the total power consumption in new technology nodes is dominated by the leakage component, caused mainly by three different issues, called subthreshold swing, drain-induced barrier lowering (DIBL) and gate-induced drain leakage (GIDL), which will be explained later in this work. New fabrication efforts are focused on reducing the effect of these issues on the transistor behavior. Some research works show that it is possible by changing the device architecture, creating scalable devices with lower leakage current (NOWAK et al., 2004).

The most important short-channel effects which affect the deep submicron devices will be shown along this section, giving a brief description of each one, and how they affect the device performance in the subthreshold region. Then, an introduction of multi-

gate and thin silicon body devices is given, showing their scalability and how each one improve the SCEs present in planar devices.

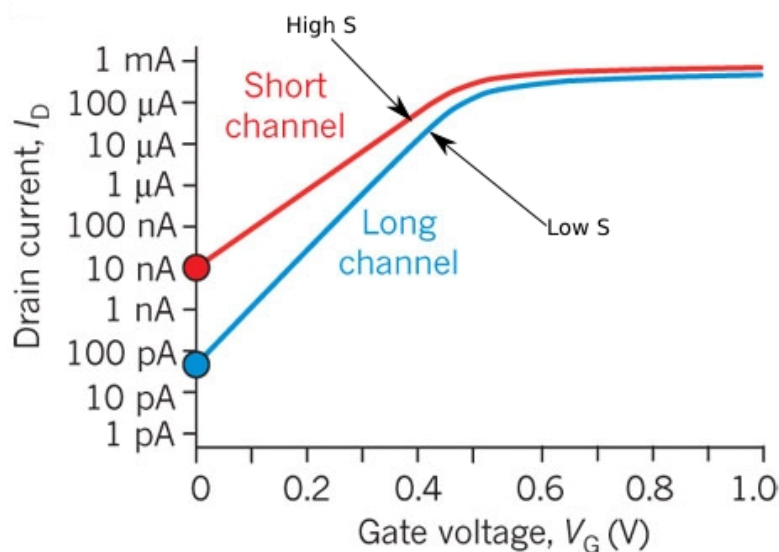
2.2.1 Subthreshold swing

Logic gates use a set of switching components which turn on and off through controlling input signals. An ideal switch has zero current when the controlling input is off, and a maximum current when the input is on. For a MOS device acting as a switch, the "off" state current is different from an ideal one. When the gate voltage is zero ($V_{GS} = 0$) or below the threshold voltage, there is a quiescent current flowing through the channel from drain to source caused by minority carriers, affecting the "off" state current. The subthreshold swing factor (S), is a figure-of-merit factor of the device performance in terms of power, and is defined as the required change in gate bias voltage (V_{GS}) to change the subthreshold drain current (I_D) by one decade.

$$S = \frac{\partial V_{GS}}{\partial \log I_D} \quad (2.1)$$

Fig. 2.2 shows a real response of typical short and long channel devices. Notice that in $V_{GS} = 0$ V, there is a little current component flowing through the drain. For a short-channel device, that current is usually higher than the long-channel one.

Figure 2.2: Current behavior of long and short-channel devices I_D versus gate to source voltage V_{GS}



Source: Ferain et al. (2011).

It is possible to see the difference between the current of a long-channel device and a short-channel device in Fig. 2.2. A long-channel device has a higher slope (slower S) compared to a short-channel one. Subthreshold slope (the reciprocal of S) is important because it defines clearly two different states of the device, a high slope (low S) is desirable in order to reduce leakage power consumption and hence, reduce the total power consumption. A device with high S factor is undesirable, because it could not have two clearly defined "on" and "off" states and could not be used as a logic switch.

To find a mathematical expression for the S factor, it could be defined as the gate voltage swing required to change the drain current by one decade. For a device operating in the subthreshold region, the drain current could be defined as:

$$I_D = \left(I_0 e^{\frac{V_G}{nkT}} \right) \left(1 - e^{-\frac{qV_D}{kT}} \right) \quad (2.2)$$

The right part of the equation 2.2 is almost 1 because $qV_D \gg kT$, so the drain current could be approximated to:

$$I_D \approx I_0 e^{\frac{V_G}{nkT}} \quad (2.3)$$

For two different gate voltages (V_{G1} and V_{G2}), one decade of change of the drain current could be calculated as:

$$I_D \approx I_0 e^{\frac{V_{G1}}{nkT}} \quad (2.4)$$

$$10I_D \approx I_0 e^{\frac{V_{G2}}{nkT}} \quad (2.5)$$

Taking the natural logarithm and subtracting the equations 2.4, 2.5, the voltage variation (ΔV_G) could be expressed as:

$$V_{G2} - V_{G1} = \ln(10) \left(\frac{nkT}{q} \right) = \Delta V_G = S \quad (2.6)$$

Where n is called the "interface trap density", defined as (for a standard MOS device):

$$n = 1 + \frac{C_B}{C_G} \quad (2.7)$$

Where C_B is the bulk equivalent capacitance which contains the interface capacitance

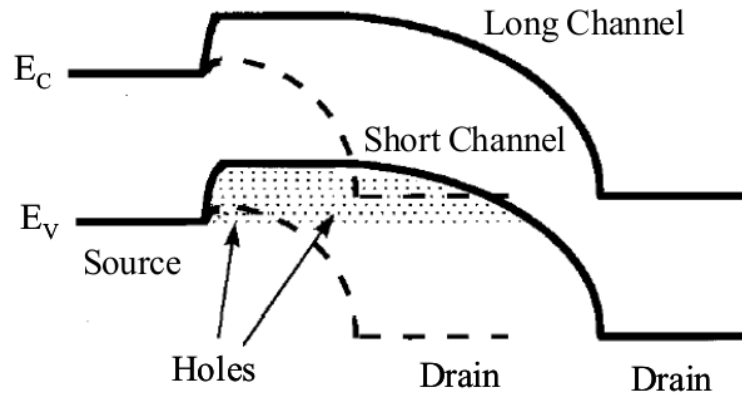
and the depletion layer capacitance, and C_G is the gate equivalent capacitance which is proportional to ϵ_{ox} , t_{ox} , W and L . The minimum value of the subthreshold swing (S) at $T = 300$ K is $S \geq \ln(10) \frac{kT}{q} \approx 60$ mV/dec. For a planar MOS device, the typical value for S is higher than 70 mV/dec.

2.2.2 Drain-Induced Barrier Lowering (DIBL)

Drain-induced barrier lowering is a short channel effect which affects the drain current behavior of submicron devices. Basically, the DIBL effect is due to a significant electric field penetration from drain to source, causing the drain depletion region moves closer to the source depletion region. Due to the depletion region shifting, the potential barrier at the source is lowered, causing a carrier mobility increment and thus, a threshold voltage diminution (QU et al., 2011).

Fig. 2.3 shows an energy band diagram for short and long channel N-type MOSFETs with $V_G = V_S = 0$ V and $V_D = V_{DD}$.

Figure 2.3: Energy bands for long and short-channel N-Type MOSFETs



Source: Chaudhry e Kumar (2004).

For long-channel devices, the drain current is mostly influenced by gate voltage (V_G), and a small component on the total depletion charge (Q_B) appears, being proportional to the drain to source voltage (V_{DS}). For short-channel devices, the depletion region charge component is higher, causing an increasement of the depletion charge and hence a threshold voltage (V_{th}) diminution. Figs. 2.4 and 2.5 show the depletion charge behavior and current curves for short-channel devices.

Figure 2.4: Electric field distribution along the channel for a short-channel device for low and high V_{DS}

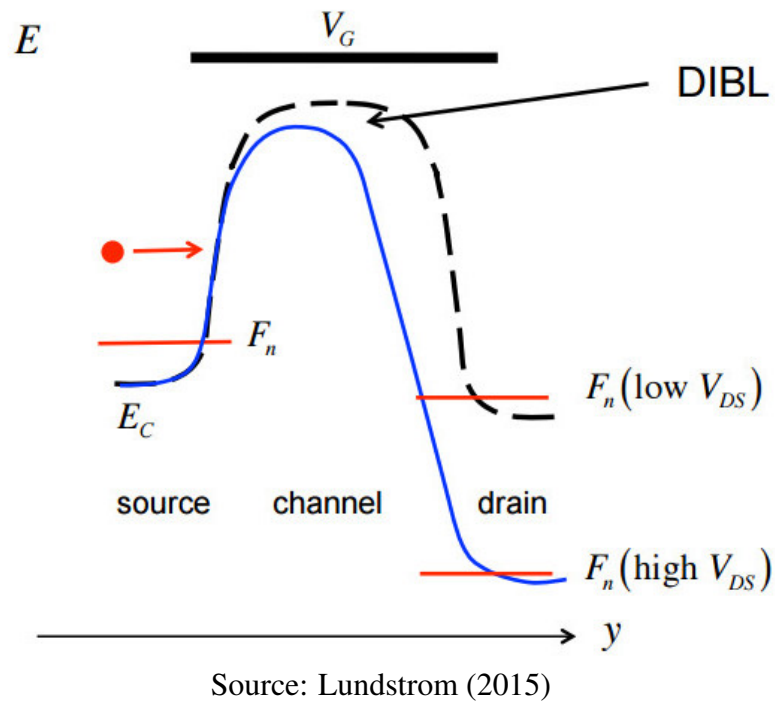
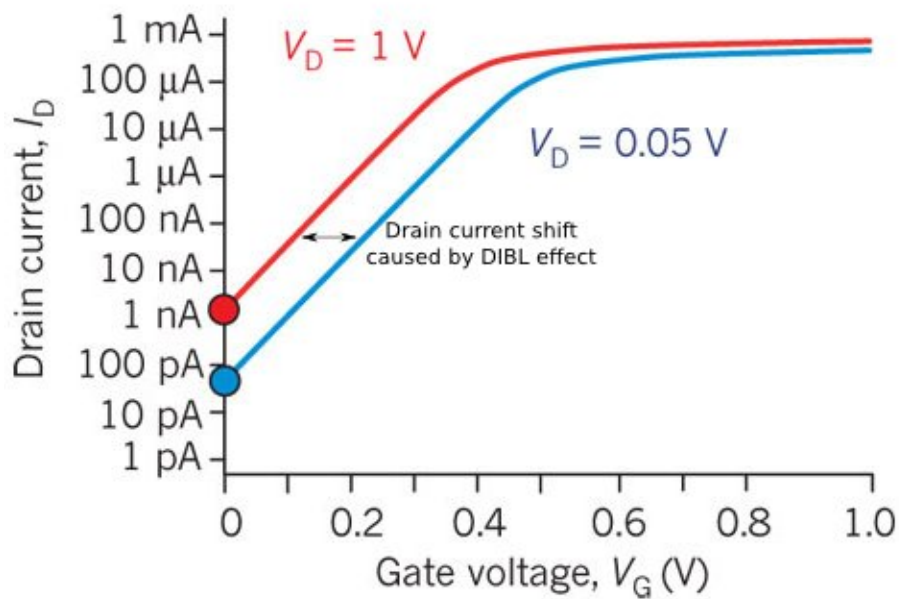


Figure 2.5: Drain current shift caused by DIBL effect

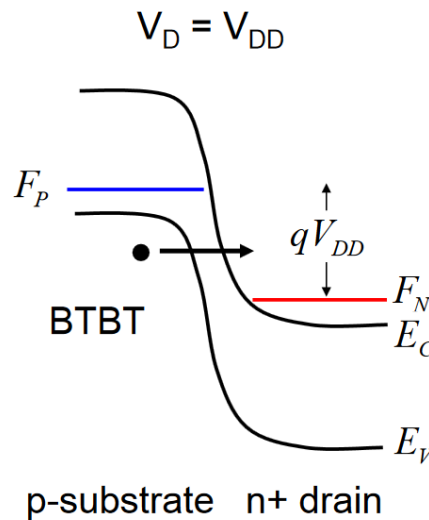


2.2.3 Gate-Induced Drain Leakage (GIDL)

To understand the GIDL effect, it is important to define the band-to-band tunneling effect (BTBT) concept. Basically, BTBT is caused by high reverse biasing of a PN junction, creating a high electric field and hence, a high band bending (higher than the silicon energy gap E_g). As a consequence, an electron in the valence band of the semiconductor tunnels across the band gap to the conduction band without the use of traps.

In MOS transistors, the drain to channel junction could be affected by tunneling effect when the gate voltage is low and the drain voltage is (V_{DD}), causing a reverse junction biasing, a high energy bending and so, an undesirable carrier mobility along the channel as seen in Fig. 2.6.

Figure 2.6: Band to band tunneling in a PN junction

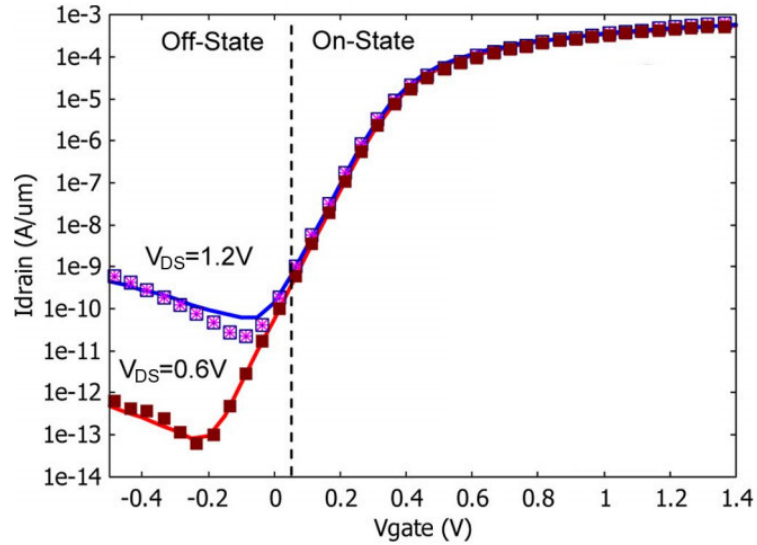


Source: Kerber et al. (2013)

When the channel is biased in accumulation region ($V_{GS} < V_T$) and the drain voltage equals to V_{DD} , carriers can flow from valence to the conduction band causing a new leakage current component when the gate voltage is low as seen in Fig. 2.7. The band-to-band tunneling current becomes relevant as the channel dopant charge increases.

2.3 Alternatives to reduce the Short Channel Effects

To reduce the effects explained above, geometrical changes to the standard planar MOSFET structure were adopted in order to increase the device efficiency (I_{ON}/I_{OFF}

Figure 2.7: V_{th} shift caused by GIDL effect

Source: Kerber et al. (2013)

ratio) and improve the channel controllability. Such changes are the reduction of the channel thickness and the adding of more gates in order to control the electron flow through the channel (ENEMAN, 2015). Those geometrical changes will be explained along this section.

An important concept to evaluate the short-channel effects in MOSFETs is the natural length (λ) (FRANK; TAUR; WONG, 1998). It could be derived from the Poisson equation and the set of boundary conditions which describe the potential distribution of the channel.

$$\frac{\partial^2 \Phi(x, y, z)}{\partial x^2} + \frac{\partial^2 \Phi(x, y, z)}{\partial y^2} + \frac{\partial^2 \Phi(x, y, z)}{\partial z^2} = q \frac{N_a}{\epsilon_{si}} \quad (2.8)$$

The λ parameter uses the potential distribution and its boundary conditions to represent the length of the region of the channel controlled by the drain (COLINGE, 2008). λ depends on physical parameters of the transistor (oxide permittivity ϵ_{ox} , channel permittivity ϵ_{si} , channel thickness t_{si} and oxide thickness t_{ox}). For a planar MOSFET, the natural length is:

$$\lambda = \sqrt{\frac{\epsilon_{si}}{\epsilon_{ox}} t_{ox} t_{si}} \quad (2.9)$$

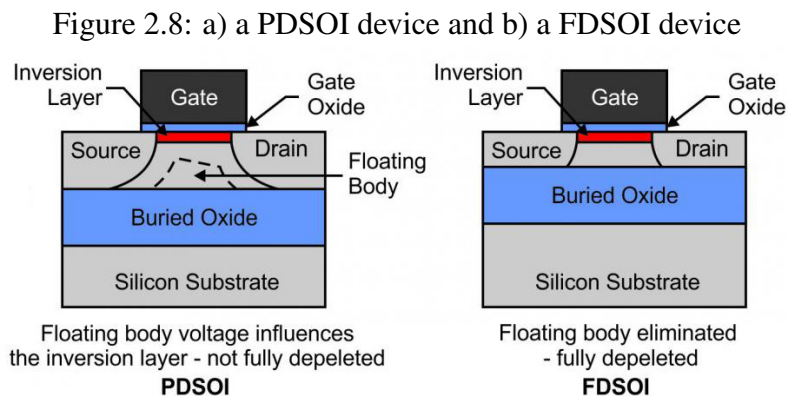
A device with low influence of short-channel effects, the gate length is at least 4 to 6 times longer than the natural length (FERAIN et al., 2011). If the equation 2.9 is

considered, to scale the gate length with low SCEs, some physical parameters such t_{ox} and t_{si} could be decreased, and high-k or different gate oxide materials could be considered in order to reduce the λ parameter.

2.3.1 Silicon Over Insulator transistor (SOI)

As seen in Equation 2.9, to decrease the natural length, one of the alternatives is the reduction of the channel thickness, thus, to increase the channel controllability and improve the I_{ON}/I_{OFF} , the channel thickness is reduced by changing its properties and materials. One of the most important devices which take advantage of this property is the silicon of insulator (SOI) MOSFET, which started to be considered to overcome the short-channel effects and mainstream bulk silicon to build low power and high performance circuits.

SOI device is not as much different from standard bulk MOSFET. The main difference is the insertion of a buried silicon oxide layer in order to limit the channel thickness (t_{Si}). The silicon channel thickness defines two types of SOI devices: the partially depleted SOI (PDSOI) (whose t_{Si} is higher than 20nm) and the fully depleted SOI (FDSOI) (whose t_{Si} is 5 to 20nm, typically 1/4 of the gate length). Fig. 2.8, shows the FDSOI and PDSOI structure.



SOI devices could have better response to short-channel effects, specially the fully depleted structure. It could be possible if the buried oxide is thin and the silicon substrate is connected to ground. With that, the electric field lines from drain or source terminate on the ground plane instead of the channel region (COLINGE, 2007).

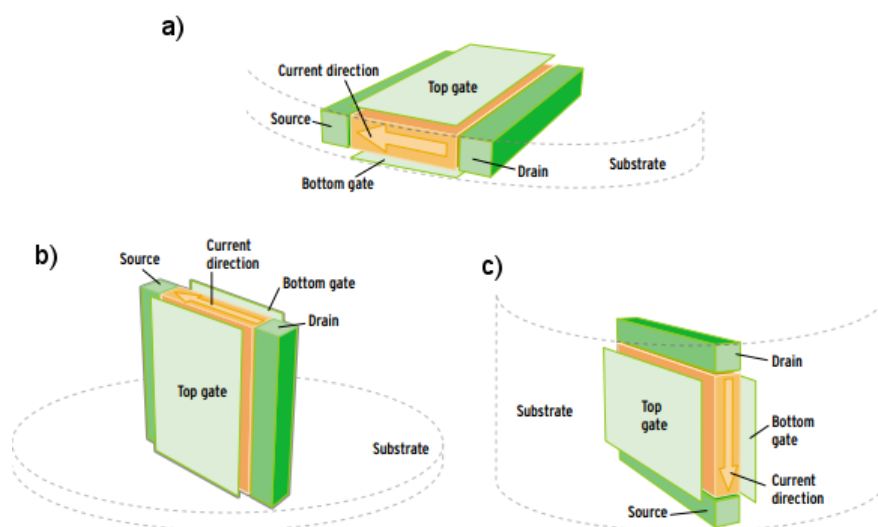
2.3.2 Multiple gate devices

Another important approach to improve the MOS electrostatic behavior is to add more gates to control the electron flow of the channel. With this, the electric field lines from source and drain terminate on the second gate electrode and cannot reach the channel region (COLINGE, 2007), improving the channel controllability by the gate. Different device geometries were developed in order to improve the electrostatic behavior of the channel and reduce the short-channel effects such as double-gate (DG), FinFETs and gate-all-around (GAA).

2.3.2.1 Double-gate MOSFET (DG)

The DG structure is comprised of a conductive undoped silicon film with two gate electrodes wrapping the conductive channel. Using the DG configuration, the electric field is better controlled, and the majority carriers are widely forced to flow in the middle of the channel (CRISTOLOVEANU et al., 2001). Double-gate transistors have excellent properties such as the reduction of short-channel effects, better mobility and transconductance. With double-gate transistors, the natural length (previously defined as λ) becomes smaller, getting easier to build devices with lower gate length. Double-gate devices could be planar or vertical. Fig. 2.9 shows different double-gate geometries.

Figure 2.9: Double gate transistor structures built on silicon wafers. a) planar double-gate, b) fin structure, c) vertical double gate device, with perpendicular channel



Source: Geppert (2002).

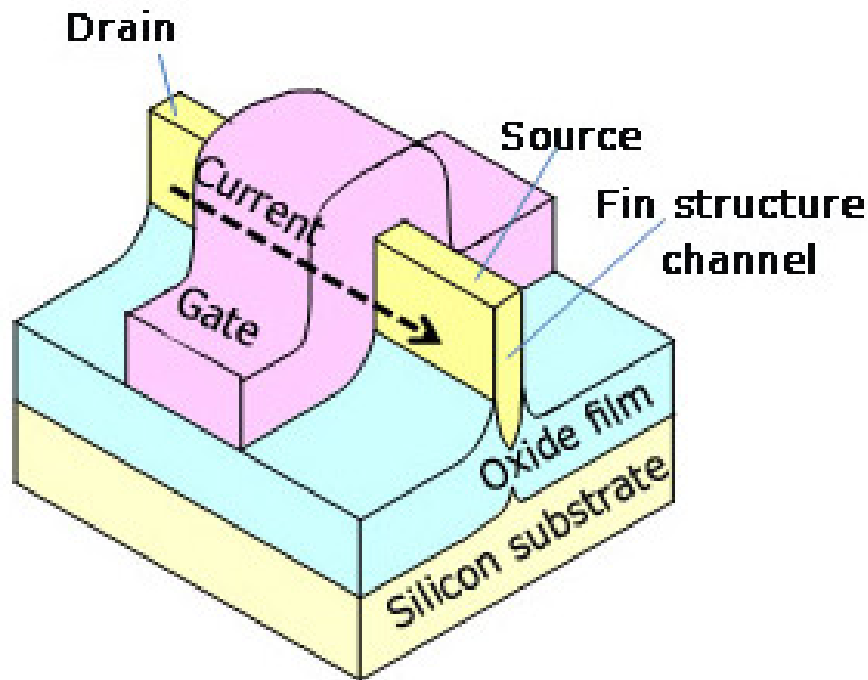
As mentioned before, the natural length of the double gate device becomes:

$$\lambda = \sqrt{\frac{\epsilon_{\text{si}}}{2\epsilon_{\text{ox}}} t_{\text{ox}} t_{\text{si}}} \quad (2.10)$$

Which is lower than the obtained in Eq. 2.9, that is, the double-gate device could be more scalable than planar structures, having a lower influence of short-channel effects.

The first transistor built using double-gate structures is the fully depleted lean-channel transistor (DELTA), created by Hitachi Labs (Japan) in 1989 (HISAMOTO et al., 1989). It was the first vertical (or 3D structure) built over a SOI substrate, offering some interesting properties such as the SCEs reduction and high scalability. Fig. 2.10 shows an illustration of the DELTA device. Based on DELTA, other double-gate devices were fabricated, such as the MFXMOS or the triangular wire SOI MOSFET (COLINGE, 2004). Fig. 2.11 shows the multiple-gate transistor evolution.

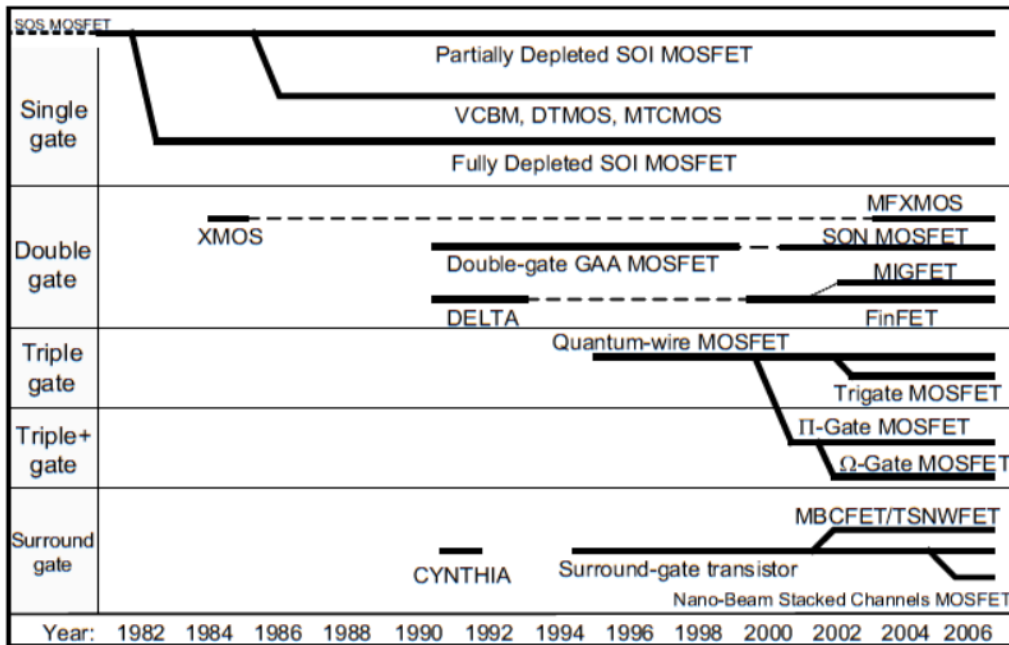
Figure 2.10: DELTA MOSFET device



Source: Hisamoto et al. (1989)

There are two different types of double-gate MOSFETS: symmetric and asymmetric devices (SDG and ADG respectively). In symmetric devices, both gates are biased at the same potential and have the same properties ($\phi_{G_1} = \phi_{G_2} = \phi_G$, $t_{\text{ox}_1} = t_{\text{ox}_2} = t_{\text{ox}}$ and $V_{G_1} = V_{G_2} = V_G$). In asymmetric devices, both gates have different properties and are fabricated using different materials, therefore, they have different work functions, oxide thicknesses and generally are biased to different voltages (GOEL; TRIPATHI, 2012)

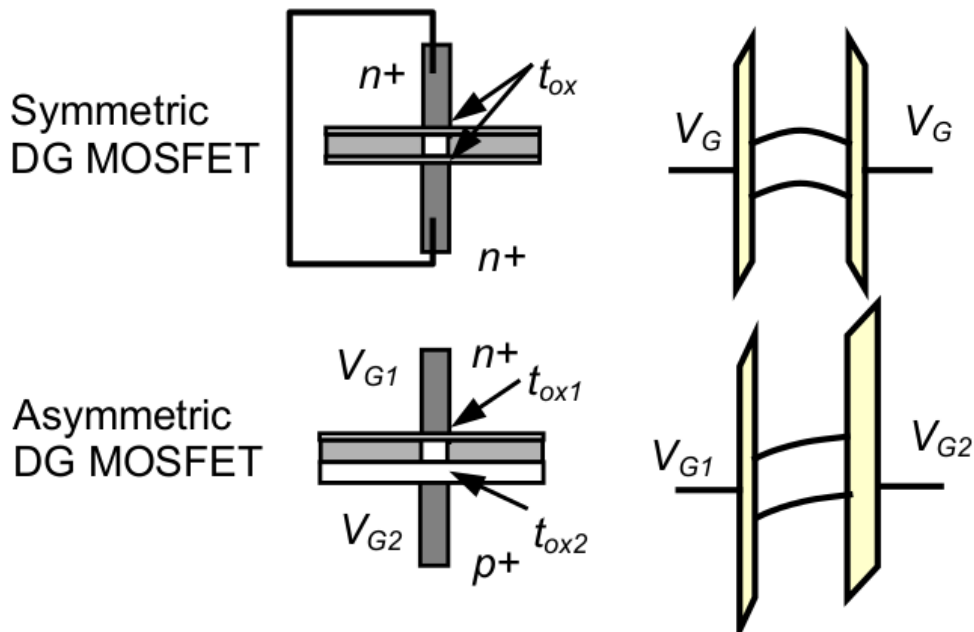
Figure 2.11: Multiple Gate device evolution and families



Source: Colinge (2004)

($\phi_{G_1} \neq \phi_{G_2}$, $t_{ox1} \neq t_{ox2}$ and $V_{G_1} \neq V_{G_2}$). Fig. 2.12 shows the different operation modes of a DG MOSFET and their energy band representations.

Figure 2.12: Double gate transistor types: symmetric and assymetric



Source: Chan et al. (2003)

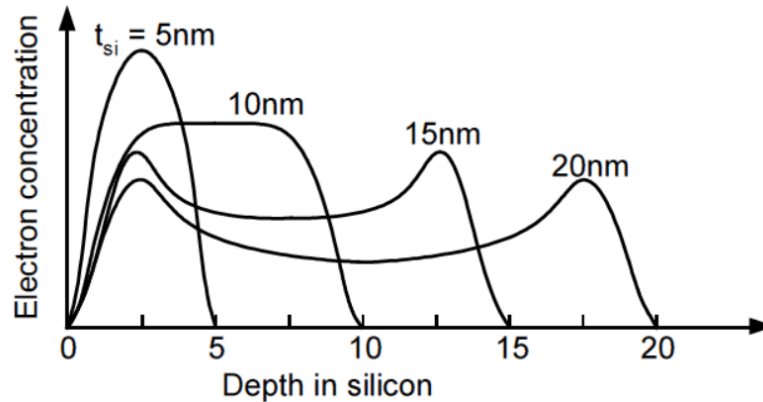
The silicon channel thickness (t_{Si}) is an important parameter to define the device behavior, if the device has a thick or highly doped channel, there is enough charge on

either side of the body in the form of depleted dopant atoms to mirror the gate charge to the other side of the channel (FRIED, 2004), that is, the electrical potential is the same from the first to the second gate, and the depletion region is small compared to the channel thickness, reaching the equilibrium at the center of the channel. At the middle of the channel, the electrical field is minimum. This case is similar to a PDSOI device with two gates.

For a thin channel device the depletion region width is higher than its physical thickness, giving interesting properties for electron mobility electric field along the channel. The first effect seen in such devices is the volume inversion (BALESTRA et al., 1987).

Volume inversion (or bulk inversion) is an important phenomenon of thin channel multi-gate devices. It appears due to the fact that inversion carriers are not confined near the gate interface as happens with planar devices, but rather at the middle of the conducting channel (COLINGE, 2007), (FOSSUM; TRIVEDI, 2013). Fig. 2.13 shows the electron density in a SDG device for different channel thicknesses.

Figure 2.13: Electron density for SDG transistors for different channel thicknesses



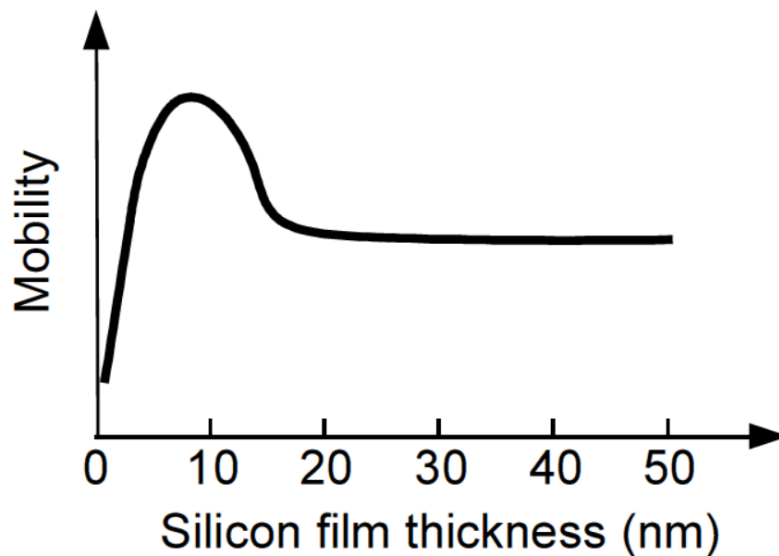
Source: Colinge (2007)

As seen in Fig. 2.13, for a thick channel double-gate device, the maximum electron concentration is near to the channel interfaces. For thin channels the maximum carrier density is near to the center. For ADG devices, the volume inversion phenomenon is similar to the SDG ones (GE; FOSSUM; GAMIZ, 2003).

The first advantage of thin channel DG MOSFET is that the carrier mobility is increased, because in volume inversion the interface scattering of the carriers is lower than in surface inversion. Resulting in a mobility increment as seen in Fig. 2.14, where clearly shows that for the case of thin devices, the mobility of thin channel devices is

nearly twice the mobility of thick channel ones (COLINGE, 2007).

Figure 2.14: Mobility dependence on silicon thickness for a double-gate MOSFET



Source: Colinge (2007)

2.3.2.2 FinFET

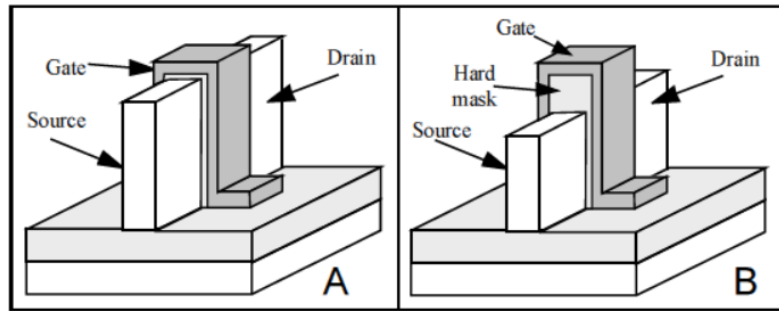
The FinFET is an special type of MOSFET transistor, in which the conducting channel is vertical, and the gate surrounds it. This device, together with FDSOI, extended the Moore's law life, due to their excellent mobility, transconductance, scalability and low SCEs. For modern feature sizes, the FinFET replaced the standard MOSFET transistors. Because it is, in essence, a vertically folded planar MOSFET, with the gate stack wrapped over the "fin" body. The physical dimensions are different from planar MOSFETs, because the device width is proportional to the silicon fin height (FOSSUM; TRIVEDI, 2013).

There are two different types of FinFETs: the double and triple gate devices. Double gate FinFET has two sidewall gates along the channel and the triple gate has the third gate on the top of the fin. Fig. 2.15 shows the double gate and triple gate FinFET structures.

Recent works on circuit design using FinFETs exemplify its outstanding performance for nanoscale CMOS, reflecting excellent SCEs (DIBL 50 mV/V and a subthreshold swing near to 70 mV/dec).(FOSSUM; TRIVEDI, 2013)

The FinFET device is usually fabricated using SOI substrates. But also can be fabricated using standard bulk substrates. Samsung in 2003 designed the first bulk FinFET

Figure 2.15: a) a triple-gate FinFET and b) a double-gate FinFET

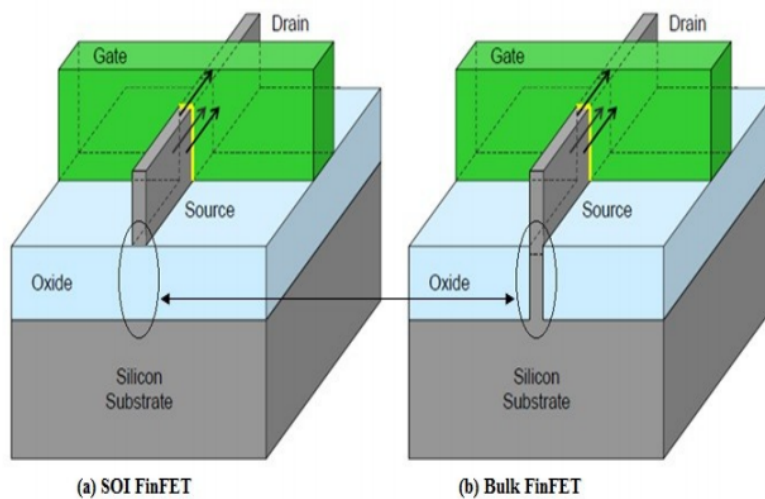


Source: Colinge (2007)

reducing the wafer costs but increasing the fabrication costs (DESHMUKH et al., 2015).

Fig 2.16 shows the difference between SOI vs. Bulk FinFET transistors.

Figure 2.16: a) a SOI FinFET and b) bulk FinFET



Source: Deshmukh et al. (2015)

Table 2.1 compares the fabrication costs between SOI and bulk FinFET structures.

Table 2.1: Process and wafer cost comparative between SOI and bulk FinFET

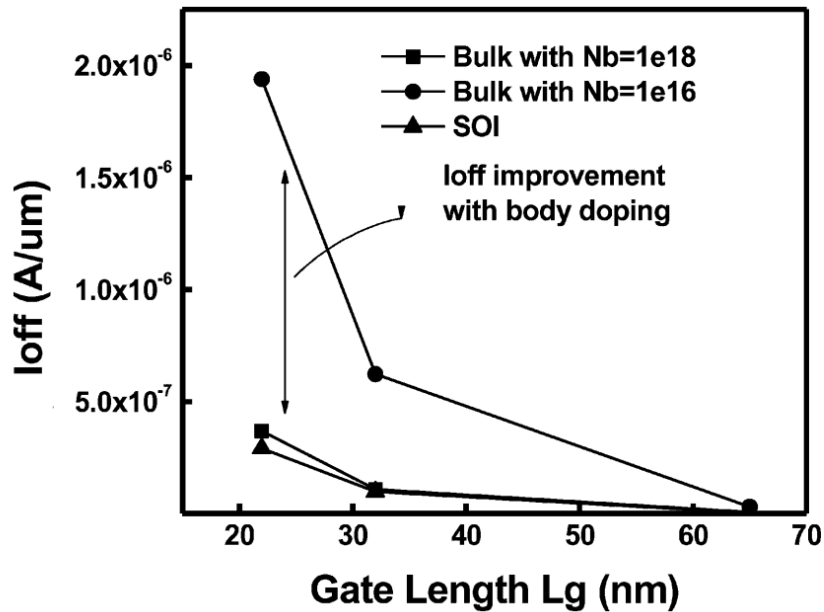
Technology	SOI FinFET			Bulk FinFET			Diff (\$)
	Litho Steps	Process steps	Cost (\$)	Litho Steps	Process Steps	Cost (\$)	
Substrate			500			120	-380
FEOL	7	56	561	9	91	805	244
Total difference							-136

Source: Deshmukh et al. (2015)

As seen in table 2.1, the bulk FinFET fabrication costs are lower. However, the process variability is higher, specially the fin width and height parameters (LATATRI-

PATHI et al., 2012). Fig. 2.17 shows the leakage current of SOI and bulk devices for different gate sizes.

Figure 2.17: SOI vs. Bulk FinFET leakage current for different gate lengths



Source: Manoj et al. (2008)

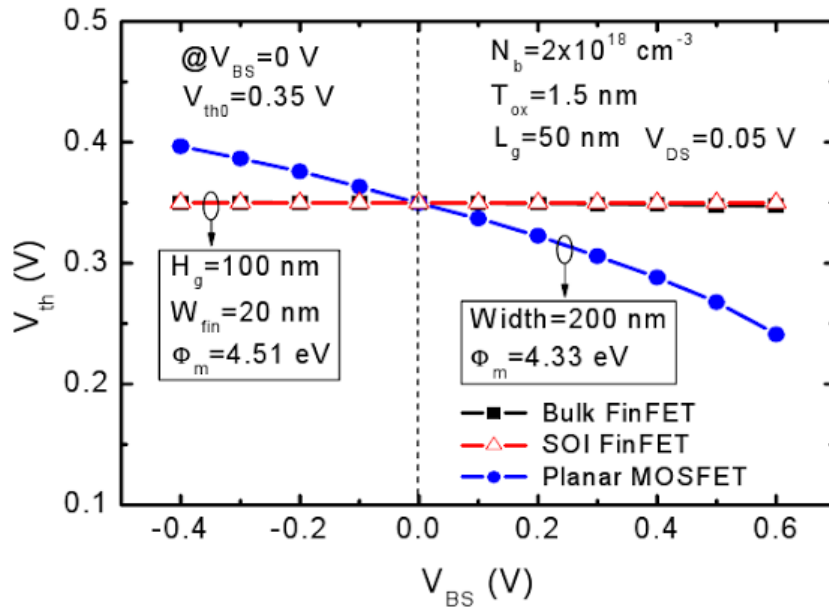
It is important to mention that the body effect impact on SOI and Bulk FinFETs is lower and could be negligible due to orientation of the thin channel with respect to the substrate (LEE et al., 2009). The presence of lateral gates reduces the electrostatic coupling between the fin and the substrate (FREI et al., 2004). The influence of bulk bias on the threshold voltage (V_{th}) is seen in Fig. 2.18. However, for gate lengths lesser than 15 nm and special body doping profiles, the bulk bias could to be considered for low-power circuit design (SACHID; KHANDELWAL; HU, 2014).

2.3.2.3 Gate-all-around device (GAA)

Another alternative of MUGFETs is the Gate-all-around device. The device channel is formed with an undoped silicon wrapped structure called "nanowire" as the conducting channel. The gate oxide and insulator surrounds the channel, creating a MOS structure. The device is also known as Silicon-Nano-Wire structure (SiNW), and it has some electrical advantages over the conventional FinFET transistor. Fig 2.19 shows an illustration of a standard GAA SiNW structure.

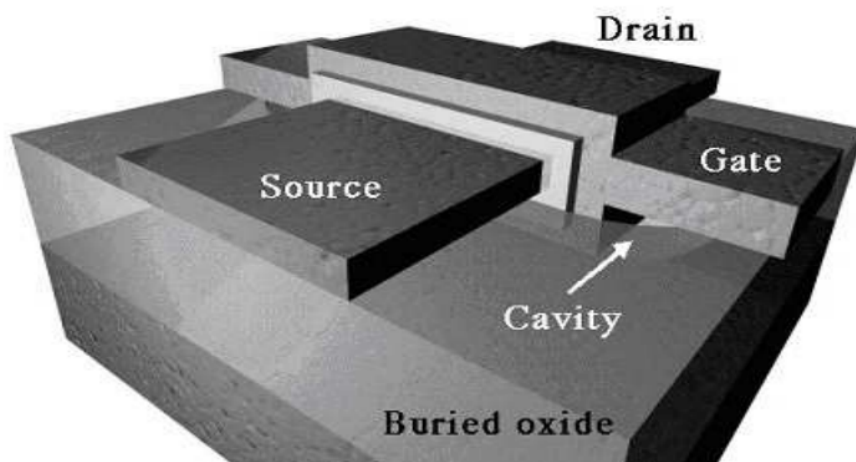
The first GAA device was created in 1990s, the device had a polysilicon gate, that wraps all the silicon channel. Due the width of the device was much larger than the silicon

Figure 2.18: Body effect on SOI and Bulk FinFETs, compared with a planar device



Source: Lee et al. (2009)

Figure 2.19: Silicon Nanowire Gate-all-around MOSFET device

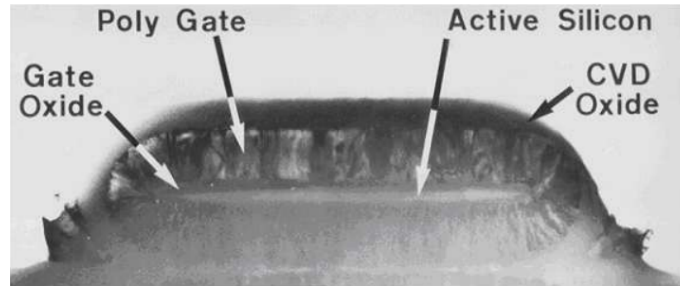


Source: Manoj et al. (2008)

channel thickness, the device was really a double-gate structure, and the contribution of the sidewall gates to electrostatic control of the channel was negligible (COLINGE, 2008).

Fig 2.20 shows the first SOI GAA device created by Colinge et al. (1990).

Figure 2.20: The first silicon Nanowire Gate-all-around MOSFET device



Source: Colinge (2008)

Gate-all-around devices could be rectangular (4-gate) or cylindrical, and the channel controllability is higher than double or triple-gate devices. Previous studies show that GAA devices have an excellent channel controllability and represent a natural evolution of FinFET structures, providing the best geometry for electrostatic control over the channel, and consequently, superior scalability properties (MARCHI et al., 2014). The natural length of the GAA device with rectangular channel (quadruple-gate device) is given by:

$$\lambda = \sqrt{\frac{\epsilon_{Si}}{4\epsilon_{ox}} t_{ox} t_{Si}} \quad (2.11)$$

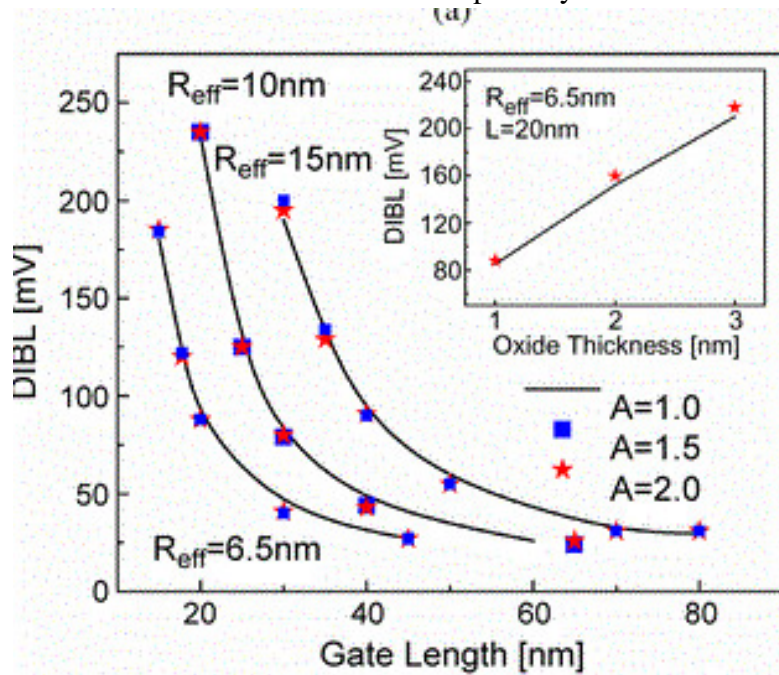
And for the cylindrical gate:

$$\lambda = \sqrt{\frac{2\epsilon_{Si} t_{Si}^2 \ln\left(1 + \frac{2t_{ox}}{t_{Si}}\right) + \epsilon_{ox} t_{Si}^2}{16\epsilon_{ox}}} \quad (2.12)$$

Which is lower than the obtained for double-gate or FinFET devices. As seen in all device geometries, the λ factor is a metric to measure the minimum gate length to avoid short-channel issues. For the case of cylindrical GAA devices, (ZHANG et al., 2011) proposed the R_{eff} concept to evaluate the SCEs present in those devices. Similarly to double-gate devices, the DIBL of GAA devices increases when the channel radius increases (higher R_{eff} value), as seen in Fig. 2.21.

As seen in Fig. 2.21, the short-channel effects for a GAA could be minimized depending on the silicon radius for cylindrical GAAs. For higher R_{eff} factors, the DIBL needs to be considered for low gate size devices. Also, similarly to other MOS devices, the gate oxide thickness is directly proportional to DIBL ratio.

Figure 2.21: DIBL and Subthreshold slope of cylindrical GAA devices



Source: Zhang et al. (2011)

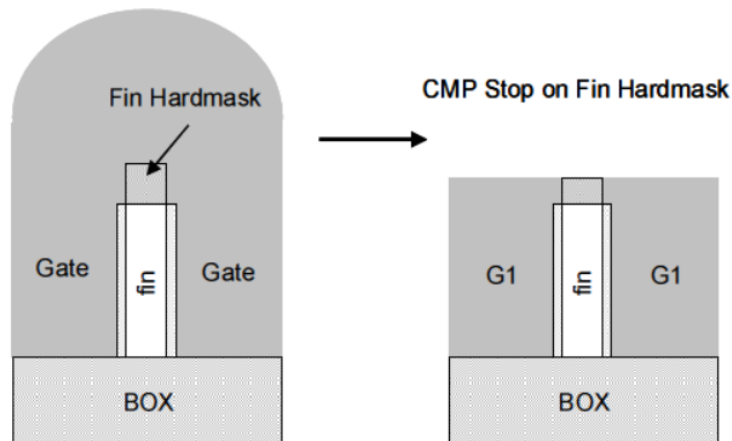
2.4 Independent Gate FinFET

FinFET devices are versatile and highly used in recent integrated circuit designs. As shown previously, they offer excellent short-channel properties and are highly scalable. Another design flexibility achieved by the FinFET is the independent-gate device which could be seen as a vertical version of the double gate MOSFET explained previously. The back gate can be used for threshold voltage (V_{th}) and SCE control as well as for interesting circuit applications (FOSSUM; TRIVEDI, 2013) which are not possible to achieve using planar devices. The study of this device in digital applications is the main goal of this work, showing the multiple possibilities which could be achieved by using two independent gates controlling a unique channel.

To fabricate an independent gate device, the gate material on top of the fin is removed in order to have two independently biased gates. To do this, etching processes such as the chemical mechanical polishing (CMP) is used. Some applications require selective etching because not all devices need to be independently controlled, to do that, etching masks are needed for selecting the devices whose gate will be etched (COLINGE, 2007). In Fig. 2.22, the IG FinFET fabrication process is shown.

The IG FinFET, called Four-Terminal FinFET (or 4T-FinFET), has two gates which are capable of controlling the channel independently, and could be connected to

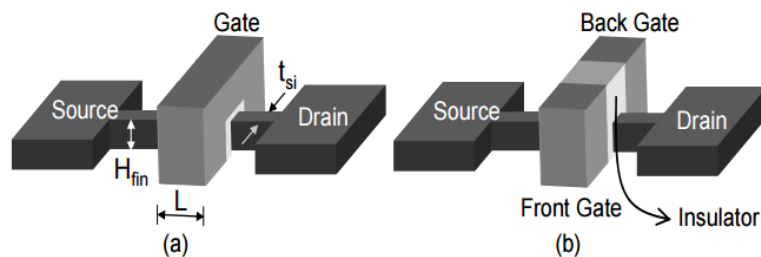
Figure 2.22: Independent Gate FinFET fabrication process



Source: Colinge (2007)

different input signals. It has the same physical behavior of the DG-MOSFET explained above, but shares the same features of a standard FinFET device. Figure 2.23 shows the structure of a Shorted-Gate (SG) and Independent-Gate (IG) modes respectively.

Figure 2.23: a) Shorted Gate FinFET (all gates tied together), b) Independent Gate FinFET (gates connected separately)



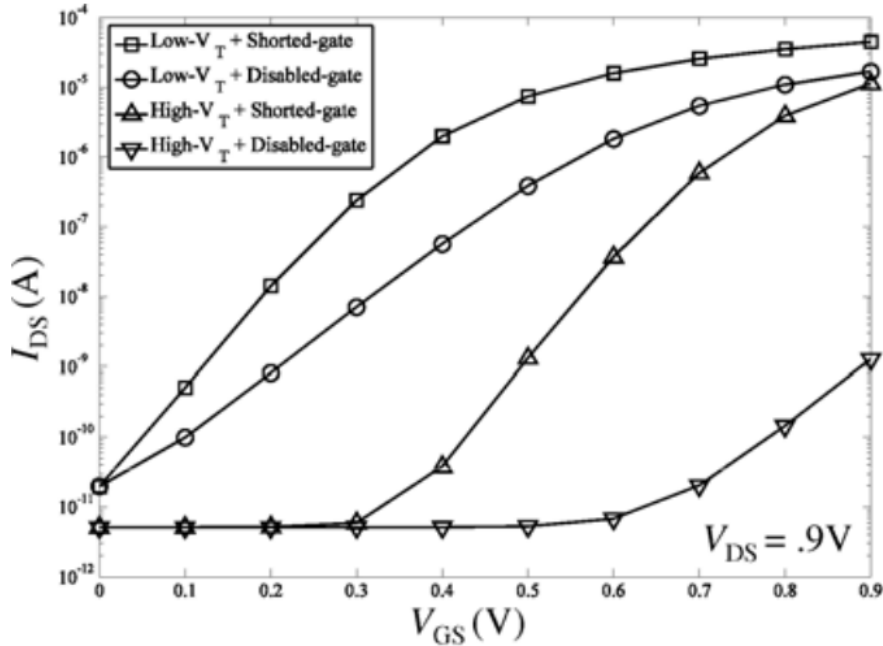
Source: Tawfik, Liu e Kursun (2007)

Independent Gate devices offer two different design possibilities: one of these is to connect the gates together to the same input signal, which gives a good channel controllability and low SCEs (the λ parameter is the same of double-gate device, this mode is called "shorted-gate" configuration (SG). The other possibility is to connect the channel gates to different signals, which is called "independent-gate" mode (IG).

Using the IG mode, the transistor current flow could be controlled by biasing one of the gates, or both simultaneously. Biasing only one gate, the device will have current flow, which is lower than the SG mode current but is enough to have a logic transition if the device is used as a logic switch. Fig. 2.24 shows the drain current when one gate is controlled independently and when both gates control the channel simultaneously.

It is possible to infer that if two different logic signals are connected to the gates of the transistor, it will implement the logic 'OR' operation. This feature is unique of the independent gate devices and will be explored and evaluated later in this work.

Figure 2.24: Independent Gate FinFET operation: when the channel is controlled by one of the gates and when the channel is controlled by the two gates connected simultaneously



Source: (ROSTAMI; MOHANRAM, 2011)

2.4.1 Basic Independent-gate device physics

IG FinFET shares the same properties of thin channel double-gate devices shown above, because they are built using a FinFET structure, which usually has a low fin thickness and both gates are equally aligned. FinFET devices also share the volume inversion and surface inversion properties shown previously for the double-gate devices. As explained above, one gate could be used to control the threshold voltage (V_{th}) and the other could be used as the signal input.

The first property of double-gate devices is the variation of the threshold voltage ΔV_{th} with respect to the biasing of the second gate V_{G2S} . This is an important parameter of IG FinFET devices and could be explored in order to design logic gates with IG FinFETs.

For symmetrical DG devices, with the same gate material, work functions and physical dimensions ($t_{ox1} = t_{ox2} = t_{ox}$, $\phi_{G1} = \phi_{G2} = \phi_G$) the front and back gate voltages

with respect to source (V_{G_1S} and V_{G_2S} respectively) could be defined as: (KIM; FOSSUM, 2001)

$$V_{G_1S} = V_{FB} + \left(1 + \frac{C_b}{C_{ox}}\right) \psi_{s_1} - \frac{C_b}{C_{ox}} \psi_{s_2} - \frac{0.5Q_b + Q_{c_1}}{C_{ox}} \quad (2.13)$$

$$V_{G_2S} = V_{FB} + \left(1 + \frac{C_b}{C_{ox}}\right) \psi_{s_2} - \frac{C_b}{C_{ox}} \psi_{s_1} - \frac{0.5Q_b + Q_{c_2}}{C_{ox}} \quad (2.14)$$

Where $C_{ox} = \epsilon_{ox}/t_{ox}$ is the gate capacitance, $Q_b = -qN_a t_{Si}$ is the depletion charge density defined for a doping level, $C_b = \epsilon_{Si}/t_{Si}$ is the body capacitance, $\psi_{s_{1,2}}$ is the surface potential created by front and back gates and $Q_{c_{1,2}}$ is the inversion charge density for front and back gates respectively.

If both gates are connected together $V_{G_1S} = V_{G_2S} = V_{GS}$, the inversion charge could be expressed as (CHIANG et al., 2006):

$$Q_c = Q_{c_1} + Q_{c_2} = -2C_{ox}(V_{GS} - V_{th(DG)}) \quad (2.15)$$

Using the equations 2.13, 2.14 and $\psi_{s_1} = \psi_{s_2} = \psi_s$, $Q_{c_1} = Q_{c_2} = Q_c$, $V_{G_1S} = V_{th}$ and equation 2.15, it is possible to deduce the threshold voltage for shorted-gate mode IG-FinFET ($V_{th(SG)}$) as:

$$V_{th(SG)} = V_{FB} + \psi_{S(SG)} - \frac{Q_b}{2C_{ox}} \quad (2.16)$$

When one of the gates is connected to ground ($V_{G_2S} = 0$), and assume that the inversion charge density of the channel is predominantly caused by the front gate ($Q_{c_2} \approx 0$), it is possible to define the front gate inversion charge density as:

$$Q_{c_1} = C_{ox_1}(V_{G_1S} - V_{th(IG)}) \quad (2.17)$$

Replacing $V_{G_2S} = 0$ and $Q_{c_2} = 0$ in Eq. 2.14, it is possible to calculate the threshold voltage for IG mode as:

$$V_{th(IG)} = (1 + r)V_{FB} + (1 + r)\psi_{S(IG)} - (1 + r)\frac{Q_b}{C_{ox}} \quad (2.18)$$

Where r is defined as the coupling factor. It could be expressed as:

$$r = \frac{c_b}{c_o + c_b} \quad (2.19)$$

And also, could be approximated to:

$$r = \frac{3t_{\text{ox}}}{t_{\text{Si}} + 3t_{\text{ox}}} \quad (2.20)$$

(FOSSUM; TRIVEDI, 2013), defines the r factor as a first order approximation for the variation of the threshold voltage with respect to the back-gate voltage. It neglects DIBL and other quantum mechanical effects.

If 2D quantum mechanical effects are considered, the previous equation does not show the real variation of the threshold voltage. Some electrical simulations show that V_{th} changes due back-gate biasing are not a constant value and depends on the volume inversion concept explained above. A better approximation proposed by (ZHANG et al., 2005) shows a dependency on the inversion layer centroid, which changes when front and back interfaces are biased, defining the effective oxide thickness ($t_{\text{ox,eff}}$) and the effective transistor width (W_{eff}) concepts, whose values depend on the charge centroid position. With those new values, the variation of the threshold voltage due to V_{G_2} voltage is shown in equation 2.21.

$$\frac{\Delta V_{\text{th}}}{\Delta V_{G_2}} = r_{\text{eff}} \approx -\frac{3t_{\text{ox1,eff}}}{3t_{\text{ox2}} + W_{\text{eff}}} = \frac{3(t_{\text{ox1}} + \frac{x_c}{3})}{3t_{\text{ox2}} + W_{\text{eff}} - x_c} \quad (2.21)$$

Where x_c is the position of the electron density centroid ($0 < x_c < W_{\text{eff}}$). Notice the negative sign in the middle expression of equation 2.21, which means the thicker the channel, less influence will have the second gate in the device threshold voltage. In Fig. 2.25, the behavior of $V_{\text{th}}(V_{G_2})$ is shown.

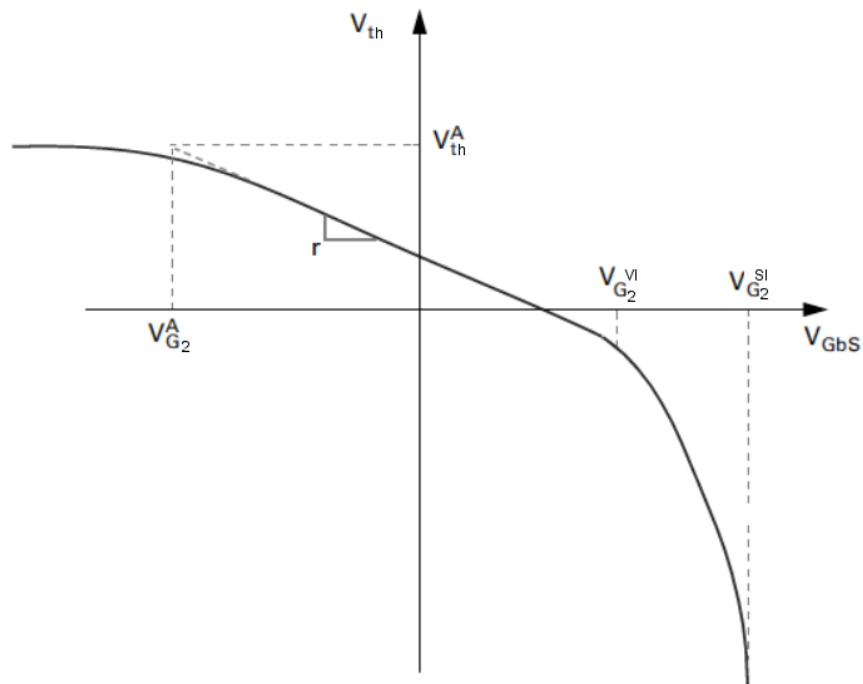
As seen in Fig. 2.25, there are three important regions for the V_{th} variation: when the back gate is accumulated, when the volume inversion (VI) predominates and when surface inversion (SI) predominates (FOSSUM; TRIVEDI, 2013), and the variation is linear when the back interface is accumulated and the channel is in volume inversion.

Also, the biasing of the gates causes different channel current characteristics. Due to gate coupling (r), the drain current also varies when back-gate voltages are applied. Fig. 2.26 shows the behavior of the drain current with respect to variations in V_{G_2} voltages.

As seen in Fig. 2.26, when the back-gate interface is biased, the variation of the drain current is minimum and vice versa, because the channel is active due the biasing of one of the interfaces. The behavior is symmetrical for the front-gate, showing that the S value is affected by the back-gate voltage.

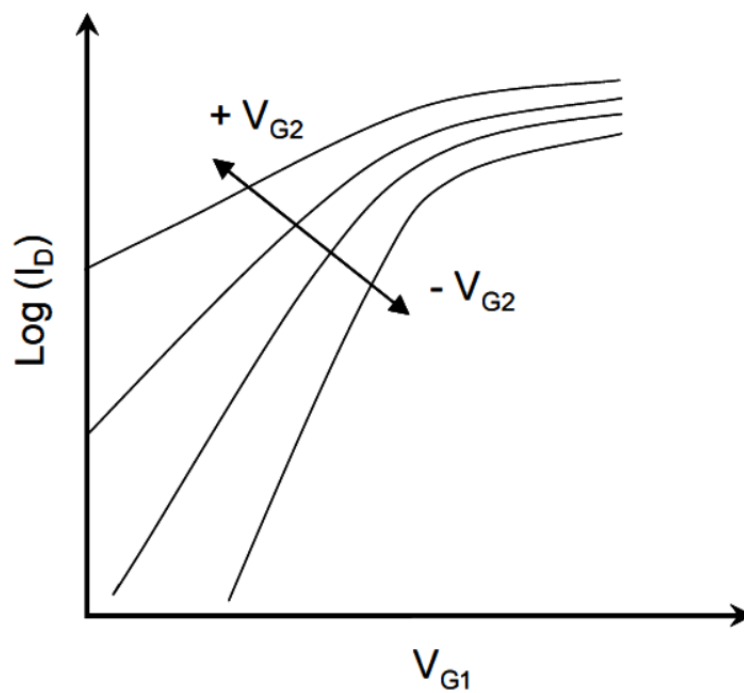
It is important to see the subthreshold swing behavior of two independent gate de-

Figure 2.25: Threshold voltage (V_{th}) variation with respect to second gate biasing (V_{G_2})



Source: Fossum e Trivedi (2013)

Figure 2.26: Drain current (I_D) variation with respect to second gate biasing (V_{G_2})



Source: Colinge (2007)

vices. For the case of IG mode could be evaluated as follows (KIM; FOSSUM, 2001):

$$S_{(IG)} = \left(\frac{kT}{q} \ln(10) \right) \frac{dV_{G_1S}}{\psi_{S_{(IG)}}} \quad (2.22)$$

Masahara et al. (2007), simplifies the calculation of the surface potential ($\psi_{S_{(IG)}}$) shown in equation 2.22 defining two IG operation modes: when the back-gate is depleted (mode 1) and when is inverted (mode 2).

For the mode 1, the $S_{(IG), \text{mode 1}}$ is equal to:

$$S_{(IG), \text{mode 1}} = \left(\frac{kT}{q} \ln(10) \right) \left(\frac{3t_{ox1} + 3t_{ox2} + t_{Si}}{3t_{ox2} + t_{Si}} \right) \quad (2.23)$$

And for the mode 2, the $S_{(IG), \text{mode 2}}$ is equal to:

$$S_{(IG), \text{mode 2}} = \left(\frac{kT}{q} \ln(10) \right) \left(\frac{3t_{ox1} + 3t_{ox2} + t_{Si}}{3t_{ox2}} \right) \quad (2.24)$$

For a IG FinFET, $t_{ox1} = t_{ox2} = t_{ox}$, so equations 2.23 and 2.24 could be simplified to:

$$S_{(IG), \text{mode 1}} = 0.06 \left(\frac{6t_{ox} + t_{Si}}{3t_{ox} + t_{Si}} \right) \quad (2.25)$$

$$S_{(IG), \text{mode 2}} = 0.06 \left(\frac{6t_{ox} + t_{Si}}{3t_{ox}} \right) \quad (2.26)$$

As seen in those equations, the S factor for the mode 2 (inversion) is higher than the value for the mode 1. Which is consistent with the behavior shown in Fig. 2.26. This also shows an interesting property of independent-gate devices. When the back-gate interface is depleted, a desirable S value could be adjusted by biasing the back-interface, which could be useful for low-power applications.

As mentioned before, an interesting property of IG FinFETs is the implementation of 'OR' logic function if both gates are used as signal inputs. If the threshold voltage of the device is increased, it could have another interesting property if Eq. 2.21 is considered. As seen previously, the threshold voltage could be decreased if the back-gate is biased. It could be useful for creating a device in which the current flows through the channel only if both gates are biased, being possible to merge two series standard transistors and to implement 'AND' operations between the inputs.

In order to implement the 'AND' logic series using IG FinFETs, higher values of V_{th} are needed. To do that, physical dimensions and properties are needed to be changed.

Their values could be chosen, being careful to not degrade the short-channel behavior of the resulting high- V_{th} device. The S parameter is a good metric criterion to determine the adequate I_{on}/I_{off} ratio needed to implement such type of device. When only one gate is biased to V_{DD} , $S_{(IG)}$ needs to be low (close to 60mV/dec) to have a minimum current flowing through the channel. As seen in equation 2.25, increasing t_{ox} helps to achieve the desired S . On the other hand, to have lower $S_{(IG)}$ when both gates get biased, looking the equation 2.26, it is easy to see that increasing t_{ox} , the S factor becomes close 60mV/dec, which is the desired value.

As seen in equation 2.18, the flat-band voltage (V_{FB}), which is proportional to the gate work function should be increased to achieve a desired value of V_{th} which permits the current flow only when both interfaces are biased. Choosing a gate material with a high work function is important to achieve a higher V_{th} value.

Another important physical parameter which needs to be changed to increase the threshold voltage is the silicon fin thickness t_{si} . It is an important value to define the device performance because a high- V_{th} FinFET creates the inversion region and conduct current only when both gates are biased above threshold voltage. Thus, the high- V_{th} FinFET needs to have a thin silicon channel.

(FOSSUM; TRIVEDI, 2013) shows the utility of the channel length underlap parameter (L_u), which is a variation of the nominal channel length value. It is important because in the weak inversion region, the underlap affects the effective gate length ($L_{eff} = L + L_u$), reducing the I_{off} current. In strong inversion, this length variation is negligible, resulting in a small reduction in the current flow, helping to improve I_{on}/I_{off} ratio (TRIVEDI; FOSSUM; CHOWDHURY, 2005).

Some authors (ROSTAMI; MOHANRAM, 2011), (FOSSUM; TRIVEDI, 2013) and (CHIANG et al., 2006) suggested the use of low and high- V_{th} devices in the same die, to build circuits using both types of switching devices. With this, some interesting applications could be explored. In this work, the use of multi- V_{th} devices is explored to create reduced logic gates and see the impact of those reduced implementations.

Table 2.2 compares the physical dimensions of high- V_{th} and low- V_{th} , and Fig. 2.27 shows the structure of high- V_{th} and low- V_{th} devices.

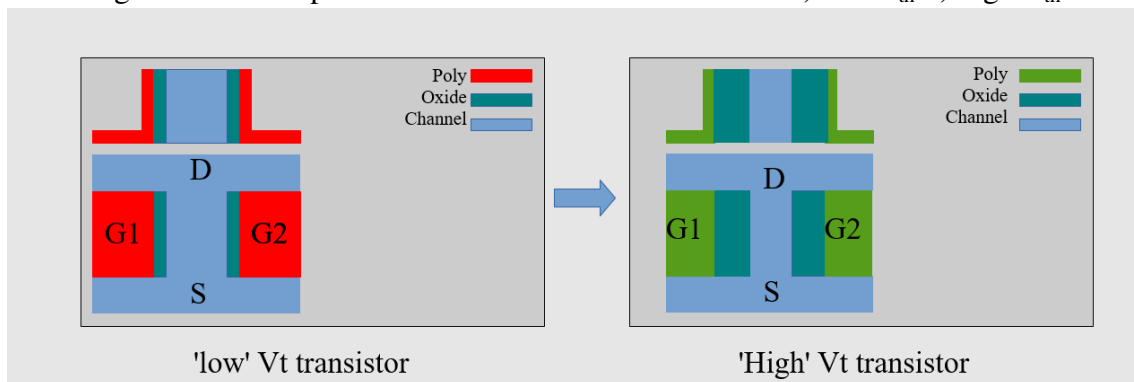
Later in this work, all the simulation models and parameters used for low- V_{th} and high- V_{th} IG FinFET devices will be shown.

Table 2.2: Differences in physical and electrical dimensions between high- V_{th} and low- V_{th} IG-FinFETs

Variable	High- V_{th} FinFET	Low- V_{th} FinFET
t_{ox}	Higher	Lower
t_{Si}	Lower	Higher
ϕ_{MS}	Higher	Lower
Gate Overlap	Higher	Lower

Source: The author

Figure 2.27: Independent Gate FinFET cross sections: a) low- V_{th} b) high- V_{th}



2.5 Simulation models

For IG-FinFET simulations, different compact models written in Verilog-A language are available to be used with SPICE simulators such as HSPICE or SPECTRE. The available simulation models will be explored and the selected model and its main SPICE parameters will be shown.

2.5.1 BSIM-CMG model

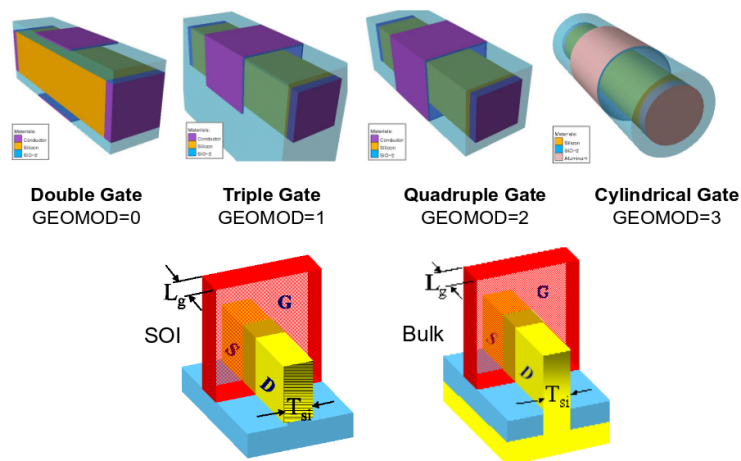
BSIM-CMG, which stands for BSIM "common multi-gate" is one of the most popular compact models for FinFET simulation. It was developed by the University of California, Berkeley. This model can simulate and extract electrical characteristics of dual-gate, triple-gate and gate-all-around devices using SOI or bulk substrate and is surface potential based. The compact model coalition (CMC) has chosen BSIM-CMG as the first and only industry-standard compact model for advanced circuit design (DUARTE et al., 2015). This model accurately simulates all the short-channel effects and quantum-

mechanical effects of multi-gate devices, however, this model only simulates devices in symmetric shorted-gate mode (SG), that is, all the gates are tied to the same potential (V_G).

To simulate the new trends of physical dimensions of multi-gate devices, the ITRS and ARM developed different device model-cards which describe the behavior of standard FinFET devices for new technology nodes. Those descriptions were developed based on BSIM-CMG model and simulated using TCAD. This set of specifications are called predictive technology model (PTM), and was made covering emerging physical effects and alternative structures. Based on physical models and early stage silicon data, PTM of bulk and double-gate devices are successfully generated from 130nm to 32nm technology nodes, with effective channel length down to 13nm. (ZHAO; CAO, 2007).

Figure 2.28 shows the different devices which BSIM-CMG model could be capable of performing electrical simulations.

Figure 2.28: Devices supported by BSIM-CMG compact model



Source: Silvaco (2014)

2.5.2 BSIM-IMG Model

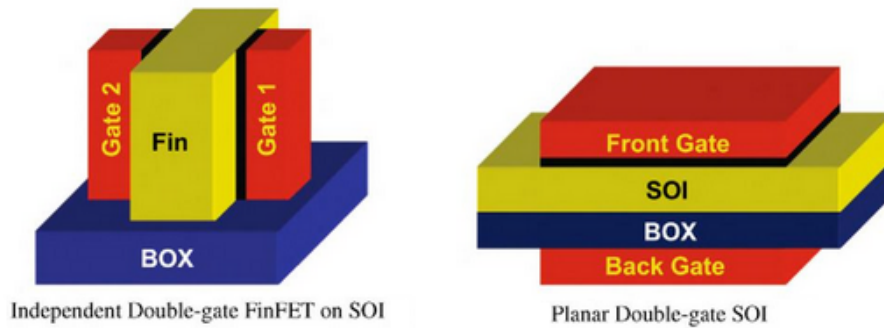
BSIM-IMG stands for BSIM "independent multi-gate". It was developed by the University of California, Berkeley and can simulate and extract electrical characteristics of asymmetric dual-gate devices connected to different potentials and fabricated over SOI substrate.

This model is a surface-potential based, and intended to be used mainly for ultra

thin body fully depleted SOI transistors (UTBSOI) and multiple gate structures, where the second gate acts as a V_{th} control. Due to surface potential equation solutions, this model considers the same analysis shown in previous sections to calculate the surface potential (based on equation 2.22, and assumptions made in equations 2.23 and 2.24) thus, this model does not assume back gate entering in strong inversion mode (KHANDELWAL et al., 2012).

The BSIM-IMG model was developed in Verilog-A and can be used by commercial EDA tools such as HSPICE and SPECTRE. Fig. 2.29 shows the double-gate devices supported by BSIM-IMG.

Figure 2.29: Devices supported by BSIM-IMG compact model



Source: Goldenblat (2010)

2.5.3 IG-FinFET Model

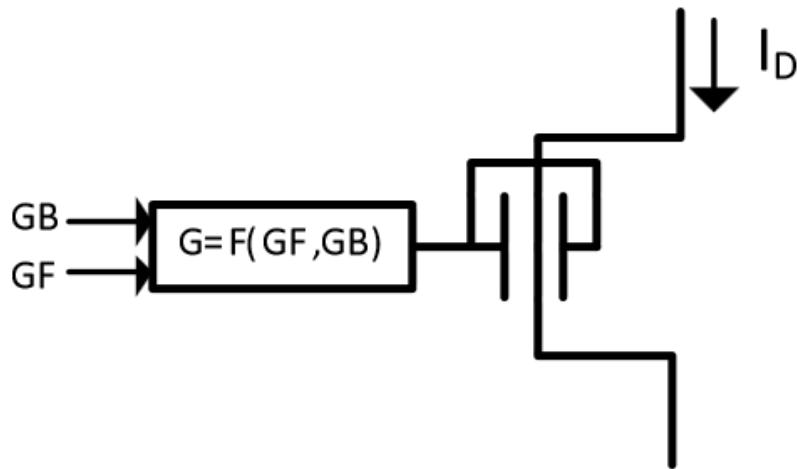
IG-FinFET model is intended for symmetrical IG-FinFET devices, with the same gate work functions ($\phi_{G_1} = \phi_{G_2} = \phi_G$) and oxide thicknesses ($t_{ox_1} = t_{ox_2} = t_{ox}$). It uses the PTM-MG model information and parameters, and convert them for IG FinFET.

Basically, the conversion technique is to use a voltage controlled voltage source (VCVS), which takes the gates voltage (V_{G_1} and V_{G_2}) and creates an output voltage which is the input for a standard FinFET (V_G) whose parameters are defined in PTM model card. Figure 2.30 shows the scheme of this implementation.

IG-FinFET model was validated using 2D TCAD simulations based on experimental published results. TCAD model was developed from PTM model parameters, comparing the I-V characteristics and tuning them to have the best I_{on}/I_{off} ratio matching in comparison to experimental results (ZAREI et al., 2013).

Based on the information given by the authors, the model converts a double-gate

Figure 2.30: IG-FinFET model scheme for SG to IG FinFET conversion



Source: Zarei et al. (2013)

FinFET into an independent-gate FinFET, acting as a SG to IG mode converter, but it does not model the surface potential between front and back gate, and does not provide real charge responses for capacitances and other values required for figure-of-merit estimations.

2.5.4 UFDG model

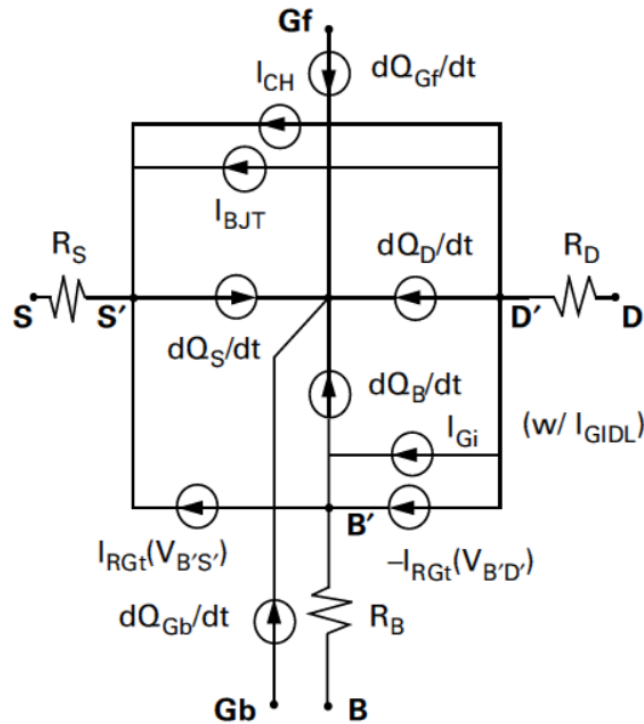
The UFDG model (University of Florida double-gate), is a compact model based on the standard DG MOSFET configuration. This model can be applied to symmetric or asymmetric DG MOSFETs, as FinFETs or Fully Depleted SOI MOSFETs (FOSSUM; TRIVEDI, 2013).

The UFDG model is physically based on the electric potential and the inversion charge in the channel body, in which the surface potential is solved using two different expressions for weak and strong inversion along the thin silicon body t_{Si} .

UFDG model needs some tuning to be used with different technology nodes. Also this model suffers from convergence problems for newer technology nodes (below 22nm) (ZAREI et al., 2013).

Fig. 2.31 shows the UFDG model equivalent circuit for SPICE simulations.

Figure 2.31: UFDG equivalent circuit model



Source: Fossum e Trivedi (2013)

2.6 Chosen model and SPICE parameters

For this work, delay and power of IG-FinFET circuits will be evaluated. The choose of an electrical simulation model is important because it will be used as a reference for logic gate implementation and evaluation. The most important aspects considered for choosing were: the model accessibility, documentation and characteristics.

BSIM-CMG is the most popular model for SG FinFETs, and is tested by CMC, but for this work is not usable because it only supports devices in SG mode. However, the PTM parameter information is written using this model as reference. For this work, physical device parameters are needed to simulate FinFET devices and PTM model gives a reference for model parameter extraction and coherent physical dimensions according to ITRS.

Table 2.3 shows a comparison between the BSIM-IMG, UFDG and IG-FinFET model, which shows the criteria parameters used by the author to chose the simulation model and 2.4 shows the most important SPICE model parameters using for IG-FinFET low and high- V_{th} .

Table 2.3: Comparison between available electrical simulation models

Item	BSIM-IMG	UFDG	IG-FinFET
Accessibility	Open Source	License Fee: \$2.500 USD	Open source
Real device Parameter modeling	Yes	Yes	No
SCE and quantum mechanical behavior	Yes	Yes	Data fitting
Compatibility	HSPICE/ SPECTRE	ngspice/ HSPICE	HSPICE

Source: The author

Table 2.4: Relevant SPICE parameters used for electrical simulation of IG-FinFET devices

Parameter	SPICE NAME	Value (SI)		Description
		Low V_{th}	High V_{th}	
$t_{ox1,2}$	EOT1,EOT2	1 nm	2 nm	Effective oxide thickness
t_{Si}	TSI	12 nm	8 nm	Silicon channel thickness
$\phi_{ms1,2}$	PHIG1,PHIG2	4.55 eV	4.85 eV	Gate work function
$L_{uD,S}$	LOVD	2 nm	5 nm	Gate/Drain Overlap
H_{FIN}	W	40 nm	40 nm	Fin Height
L	L	32 nm	32 nm	Gate Length

Source: The author

As shown in Table 2.3, the BSIM-IMG model has the necessary features for IG-FinFET simulations, so it will be used for the electrical evaluation of logic networks in the later chapters. This model could represent the logic functionality of transistor networks based on IG FinFET devices.

2.7 Summary

In this chapter, the different issues regarding to scaling issues for planar devices are shown. The most important effects that affect the leakage current related to scaling in planar devices were explained, and the new strategies to reduce the short-channel effects (SCEs) were shown. Devices with thin channels and multiple gates have increased the scaling of MOSFET transistors and extended the Moore's law life. In addition to the scaling, IG FinFETs have new features as shorted and independent gate operation modes, offering new interesting properties for logic gate design. The following chapters will explore the IG FinFET transistor as a circuit element that could reduce area and power consumption of digital circuits, and some different digital gate topologies will be evaluated through electrical simulations using the BSIM-IMG model to see its utility for digital design.

3 IG FINFET BASED LOGIC GATES

3.1 Introduction

The previous chapter shows a brief introduction about dual independent gate (IG) FinFETs, showing its particular features for circuit design such as threshold voltage control using the second gate of the device. In this chapter, the usage and utility of IG-FinFETs in digital circuits will be explored, showing the different alternatives which this type of transistors bring for reducing area in logic gates, showing that it is possible to expand a single topology created using CMOS implementation into a variety of circuits with different delays and power consumptions. The impact of using the shorted gate (SG) and independent gate (IG) modes for the different logic networks in terms of delay and power consumption will be shown, comparing the results using the figure-of-merit concept for the different explored topologies.

3.2 Naming conventions

In this work, some different logic functions will be implemented using IG FinFETs and different configurations will be explored and evaluated. To facilitate the reader comprehension, some conventions will be adopted in order to show the different topologies that could be created using those devices. Table 3.1 shows the conventions for the different transistor types and threshold voltages.

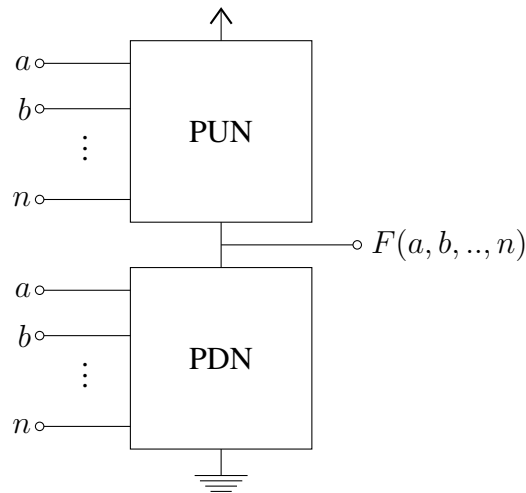
Table 3.1: Naming conventions for IG-FinFET devices

Mode	N-type	P-type
Low- V_{th} IG		
Low- V_{th} SG		
High- V_{th} IG		

3.3 Basic IG FinFET usage in pull-up/pull-down networks

CMOS digital circuits use two complementary networks for logic function implementations. A logic function $F(x_1, x_2, \dots, x_n)$ uses two different logic circuits implemented using N-type MOSFETs for the pull-down network (PDN) and P-type MOSFETs for the pull-up network (PUN). Each network determines the binary value of the output depending on the logic value of the inputs for the case of combinational circuits. The PDN determines whether the output is low and PUN determines whether the output is high. Fig. 3.1 shows a conventional CMOS logic function implementation using a PUN and a PDN.

Figure 3.1: A standard CMOS logic function implementation



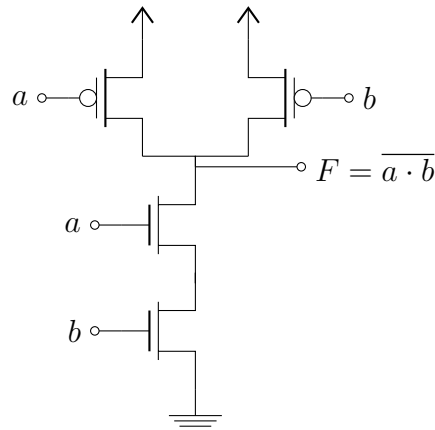
Each logic network (PDN and PUN) uses a single device type. PDN always uses only NMOS transistors and PUN uses only PMOS transistors, those transistors are connected in series/parallel arrangements in order to open or close the current flow from source/ground to the load.

For the case of a NAND logic gate, its logic function is $F(a, b) = \overline{a \cdot b}$ and needs two series NMOS transistors, and two parallel PMOS transistors, each one connected to a and b inputs respectively. Fig. 3.2 shows a standard CMOS implementation of a two-input NAND gate.

As seen in Fig. 3.2, a 2-input NAND gate needs 4 transistors for its implementation. In general terms, a N -variable logic function needs $2 \cdot N$ transistors to be implemented using standard CMOS logic. 2 N-type and 2 P-type MOSFETs for PDN and PUN respectively.

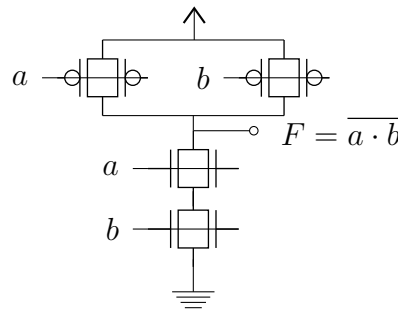
With an independent-gate FinFET device, there are three different implementation

Figure 3.2: A standard CMOS logic function implementation for 2-input NAND gate



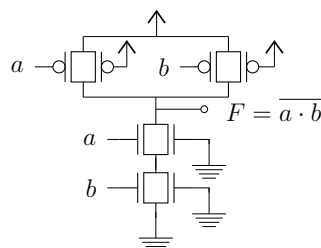
topologies derived from standard CMOS implementation. The first one is shorting the two gates and connecting both gates to the same input as shown in Fig. 3.3. This model will be called as shorted-gate mode (SG) later in this work.

Figure 3.3: A NAND2 logic gate implemented with IG FinFET in single-gate mode (SG)



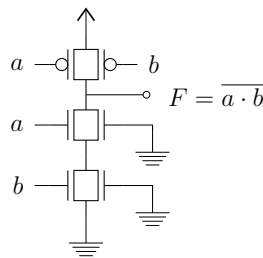
As shown in the previous chapter, an IG FinFET device could control its V_{th} of any of the gates by biasing the other ($\Delta V_{th}/\Delta V_{G_2}$ relation shown in equation 2.21). If the second gate is tied to ground ($V_{G_2} = 0$), the threshold voltage of the first gate is high (as seen in Fig. 2.25). In IG mode, it is possible to decrease the power consumption of the gate. Fig. 3.4 shows an implementation of this mode, which will be called IG-LP mode later in this work.

Figure 3.4: A NAND2 logic gate implemented with IG FinFET in LP mode (IG-LP)



One of the features of IG transistors shown in the previous chapter, is the channel controllability by any of the gates. The channel could be activated by biasing one of the two gates, implementing an 'OR' function between the inputs for the case of low- V_{th} devices. Using the NAND gate shown in Fig. 3.2, the PUN could be replaced by an unique P-type IG device, achieving the same functionality with less area. Fig. 3.5 shows the NAND gate of Fig. 3.4 with the inputs a and b connected to the same IG device.

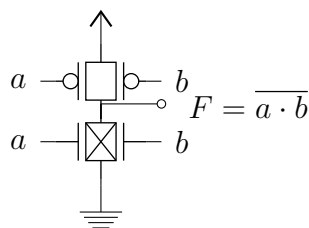
Figure 3.5: A NAND2 logic gate implemented with IG-FinFET, merging a and b inputs into a single P-type device



Now, the circuit of Fig. 3.5 has three transistors, and the same functionality of the circuit shown in Fig. 3.2. This could be useful for area minimization in circuit design. Now, as shown in the previous chapter, a low- V_{th} transistor could be used for implementing 'OR' operations with the inputs, and using IG-FinFETs with high- V_{th} , it is possible to perform 'AND' operations with the inputs, because the transistor drives current only when the two inputs are on, and could be useful for reducing the PDN network of Fig. 3.5.

This functionality is possible to be implemented only with IG transistors and is a new alternative for circuit design, and will be explored in terms of delay and power in this work. 3.6 will show a reduced NAND gate implementation with two FinFET devices. This mode will be called IG-RED later in this work.

Figure 3.6: A NAND2 logic gate implemented with IG-FinFET devices in reduced mode (IG-RED)



The same process could be applied for a NOR logic function, which uses series transistors on its PUN and parallel on its PDN. 3.7 shows the reduced IG-FinFET version

of a two-input NOR gate.

Figure 3.7: A NOR2 logic gate implemented with IG-FinFET devices in reduced mode (IG-RED)

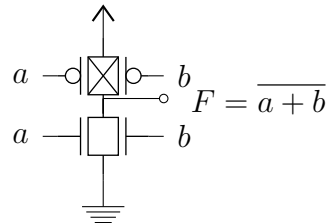


Table 3.2 shows a summary of the different topologies illustrated previously. Those configurations will be explored in later in this work, showing their delay/power impact of their usage in digital circuits.

Table 3.2: Summary of possible IG-FinFET modes for the NAND2 logic gate

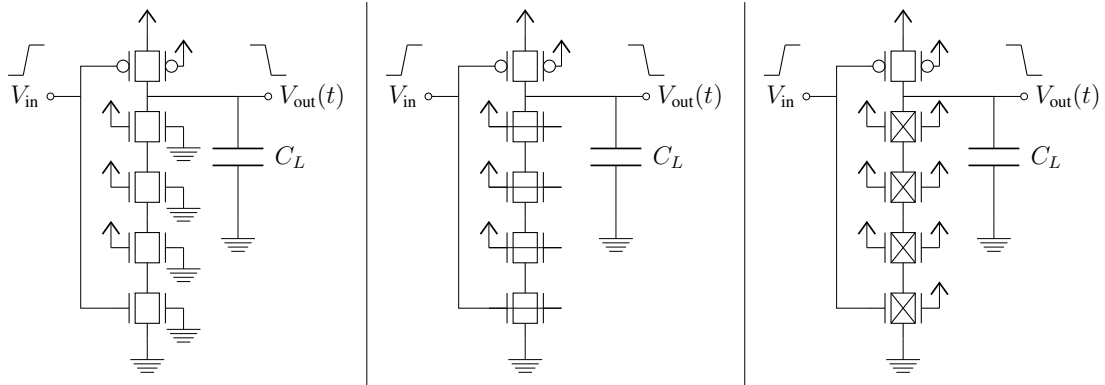
SG Mode	IG-LP Mode	IG-RED Mode

3.4 Delay evaluation of stacked transistor networks

Logic gates use arrays of series and parallel devices to implement logic functions. Serial arrays of transistors cause an impact on the logic gate performance. For example, to implement a 4-input NAND function, a series array of 4 N-type devices is required, causing an impact on the gate performance. In order to evaluate the performance of series arrays built using IG FinFETs, electrical simulations of different serial arrays will be performed to show the performance of the different transistor arrays derived from the different IG FinFET implementation modes.

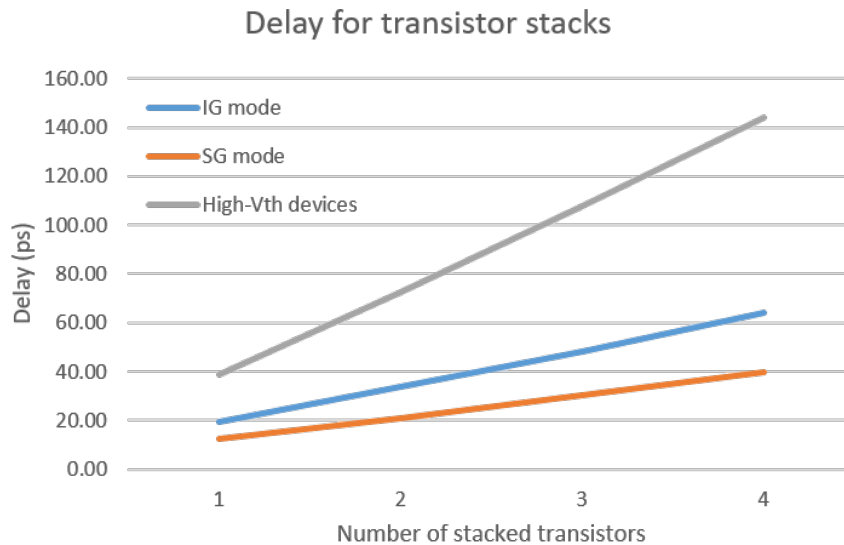
Figure 3.8 shows an example of the simulated test bench for 4-transistor stacks. Stacks of 1, 2, 3 and 4 series transistors were simulated using IG FinFETs in SG, IG modes and high- V_{th} devices using 1 fF load capacitance and 10 ps input transition time.

Figure 3.8: Proposed test bench for electrical evaluation of transistor stacks



Electrical simulations were performed using the BSIM-IMG compact model and HSPICE. The high-to-low delay was measured as the time when the output crosses 50% of its value starting to count after 50% of the input transition time. Simulations results are shown in Fig 3.9.

Figure 3.9: Results for delay of stack transistors for arrays of 1 to 4 series IG FinFETs



As seen in Fig. 3.9, there is a high penalty in terms of delay when stacks of high- V_{th} devices are used. That is because they have lower current capability compared to low- V_{th} devices, caused by their higher threshold voltage. When the stack size is increased, the current capability of the array is reduced causing an increment in their delay response. For the case of low- V_{th} arrays, there is a difference between their delay response caused mainly by their current capabilities, which is better for the devices connected in SG mode (as shown in in Eq. 2.16).

3.5 Transistor network compactions

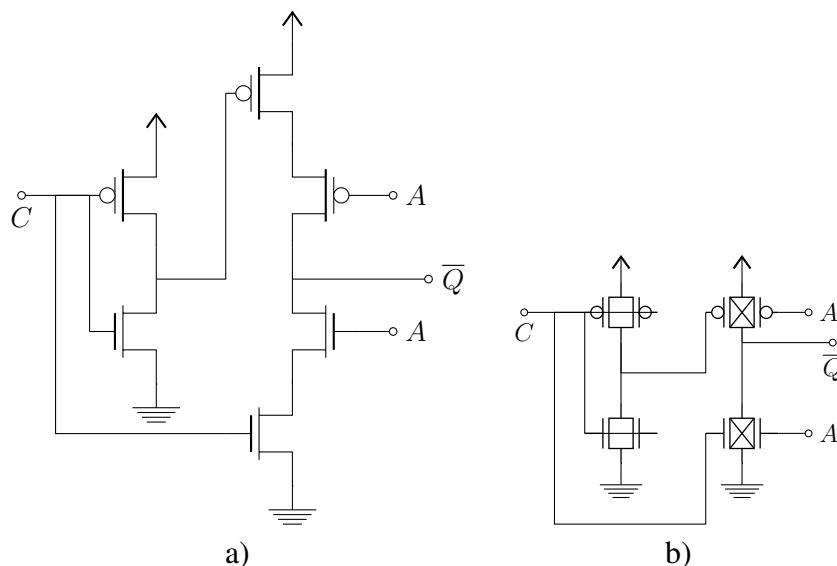
In this section, logic functions with more than two variables are used as an example of logic network implementation using IG-FinFET devices. As mentioned above, standard CMOS technology limits the alternatives of implementation of a certain logic function. With IG FinFET devices, those possibilities are increased and now, more topologies could be chosen depending of the application: high-performance, low-area and low-power. Previously was shown that IG-FinFET devices are useful for merging two parallel MOS transistors (using low- V_{th} devices) and two series transistors (using the high- V_{th} ones), reducing the number of required devices for implementing a two-input NAND/NOR gate by a half.

3.5.1 Compacting standard logic gates

Standard logic gates could be compacted using FinFETs in IG mode. Previously, compaction of NAND/NOR gates was shown to illustrate the utility of IG FinFET in transistor reductions. Now, more examples of complex logic structures are shown in order to see more complex implementations.

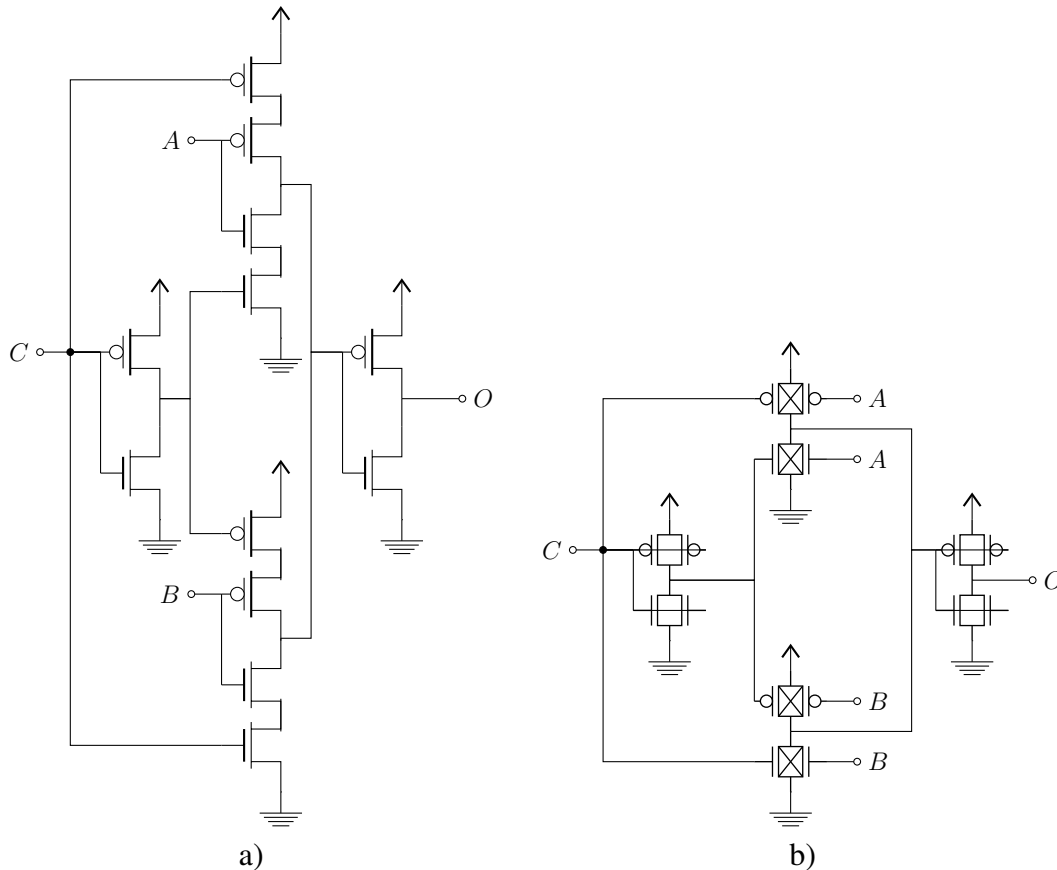
A standard tri-state buffer could be reduced using High- V_{th} IG-FinFET devices as shown in Fig. 3.10. The High- V_{th} devices help to reduce arrays from 4 to 2 series transistors required to implement the tri-state function.

Figure 3.10: a) Standard CMOS implementation of a tri-state buffer b) reduced version using IG-FinFETs



The same methodology could be applied to a 2-input multiplexer, which could be reduced using FinFETs in IG mode as shown in Fig. 3.11. Note that in this case, only 8 IG transistors were needed to implement the MUX function.

Figure 3.11: a) Standard CMOS implementation of a two-input multiplexer b) its reduced version using IG-FinFETs



For a XOR2 logic gate, only 4 devices are needed for implement the logic function, as shown in Fig. 3.12b) instead of 8 needed by standard CMOS technology.

Finally, a full-adder circuit is shown in Fig. 3.13a). Taking advantage of its symmetry, a 28-transistor standard CMOS version could be reduced to 20 IG-FinFET devices as shown in Fig. 3.13b).

Table 3.3 shows a summary of the device count used to implement each combinational logic gate.

3.5.2 Compacting logic complex functions using de-factorization

Several methods have been proposed to optimize transistor networks in conventional single-gate MOS technology (KAGARIS, 2016), (POSSANI et al., 2016). How-

Figure 3.12: a) Standard CMOS implementation of a two-input XOR logic gate b) its reduced version using IG-FinFETs

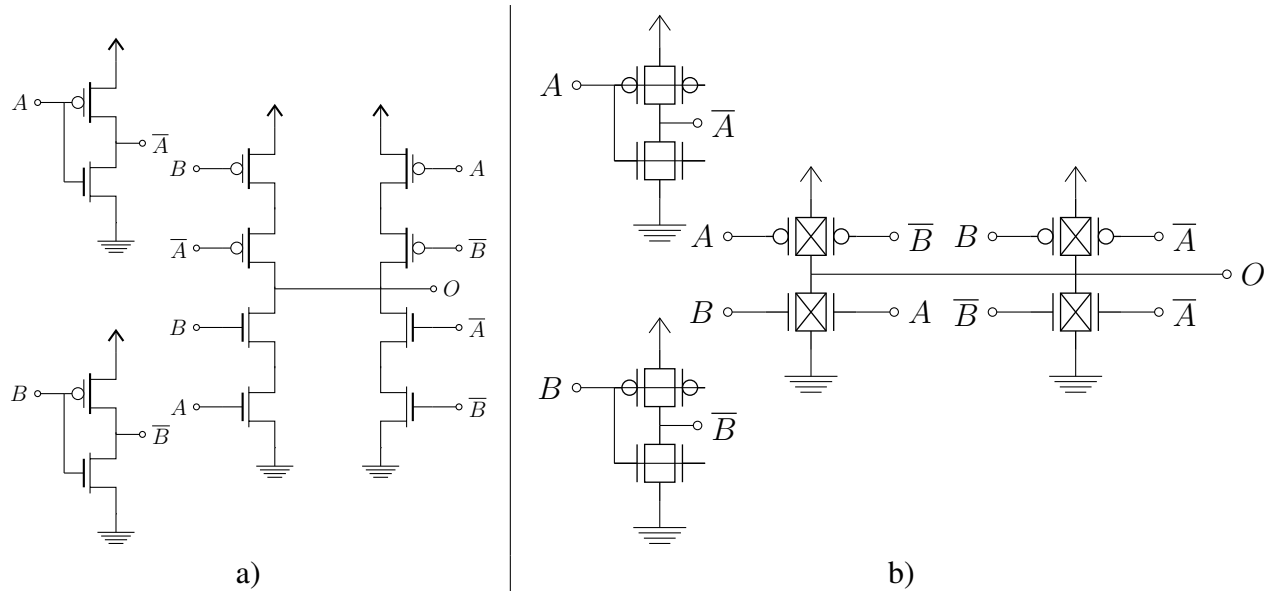
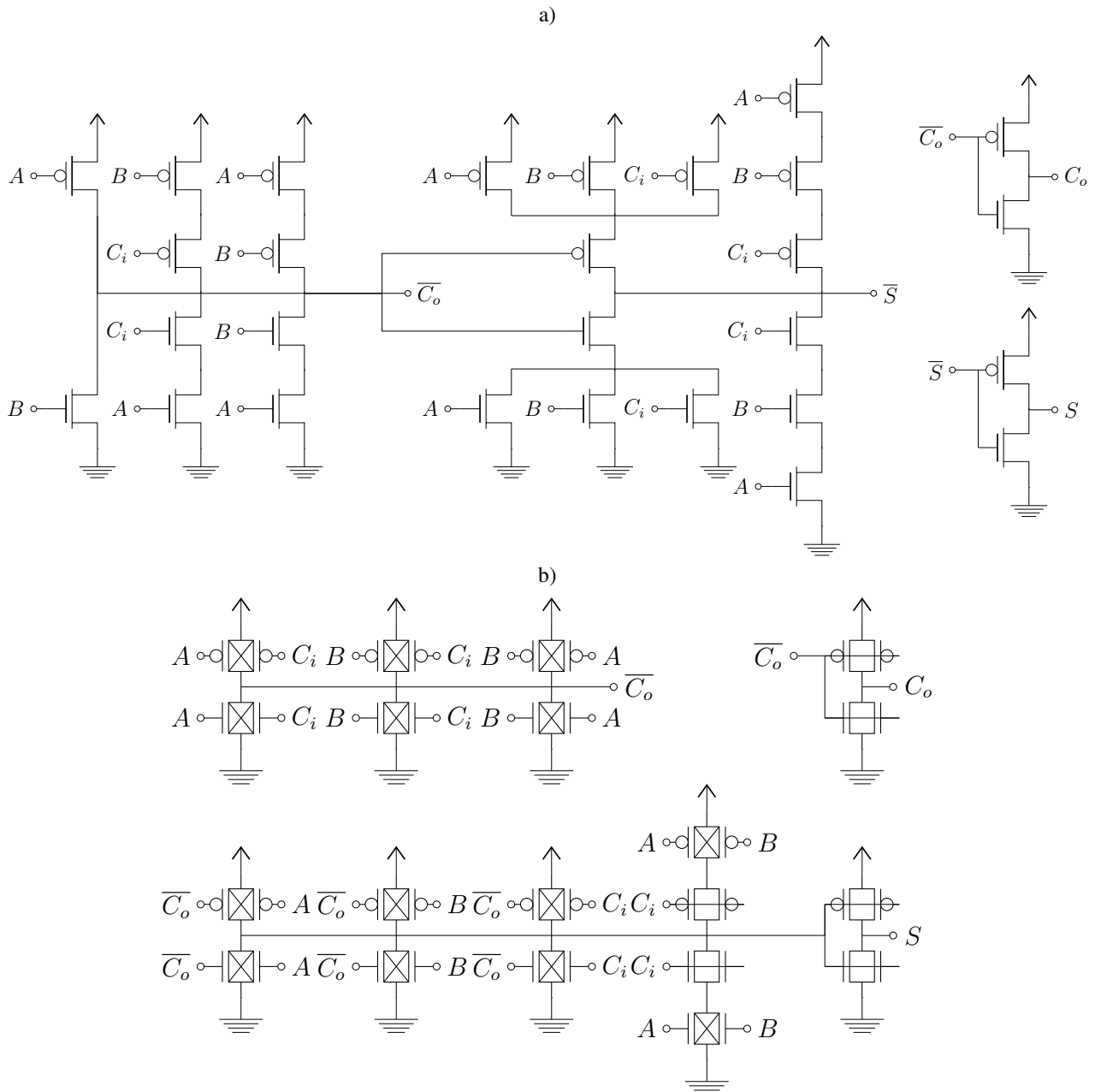


Table 3.3: Device count for F_1 to F_{10} logic functions implementation using T_1 to T_5 implementation modes

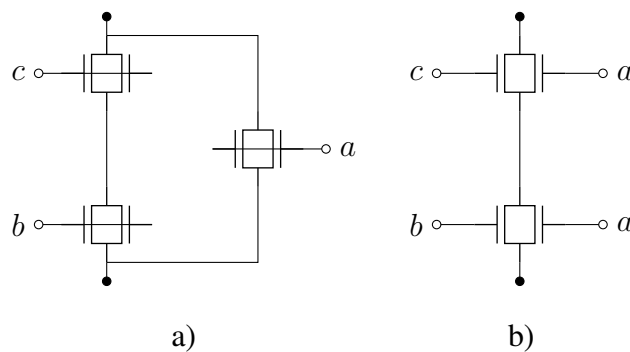
Logic Gate	Device count		% Red
	Std. CMOS	IG FinFET	
INV-TRI	6	4	33.3
MUX21	12	8	33.3
XOR2	12	8	33.3
Full-Adder	28	20	28.5

Figure 3.13: a) Standard CMOS implementation of a full-adder logic gate b) its reduced version using IG-FinFETs



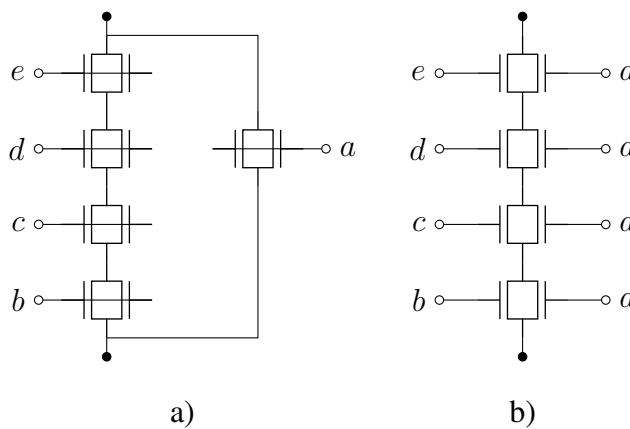
ever, when considering the design of logic gates using IG FinFETs, the best transistor arrangement may differ from ones the obtained through traditional methods. In this sense, an alternative is to apply a defactorization technique over the factored forms produced by those methods. In (ROSTAMI; MOHANRAM, 2011), the authors presented two ways for defactoring Boolean expressions. The first defactorization technique starts from a conventional single-gate transistor arrangement, as shown in Fig. 3.14 a), and replicates literals that enable the merging of transistor into single IG FinFET, as depicted in Fig. 3.14 b).

Figure 3.14: A logic circuit implemented with: a) FinFETs in SG mode, and, b) reduced network using FinFETs in IG mode and defactorization proposed by (ROSTAMI; MOHANRAM, 2011)



As seen in Fig. 3.14, defactoring option could be useful for reducing transistors in a logic network. However, could not be the best option when the logic network has series transistors. Fig. 3.15 shows an example of a series transistor network.

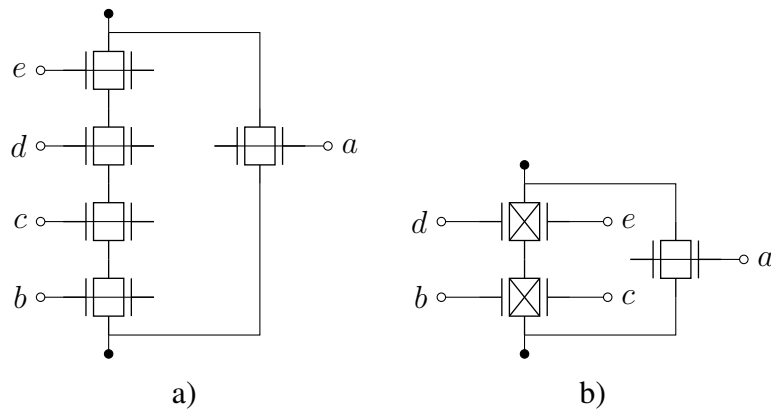
Figure 3.15: A series logic network with: a) FinFETs in SG mode, and, b) reduced network using FinFETs in IG mode and defactorization proposed by (ROSTAMI; MOHANRAM, 2011)



The example shown in Fig. 3.15 using defactorization shows that distributing the a literal, reduces only one transistor. In terms of transistor count, series transistors could be

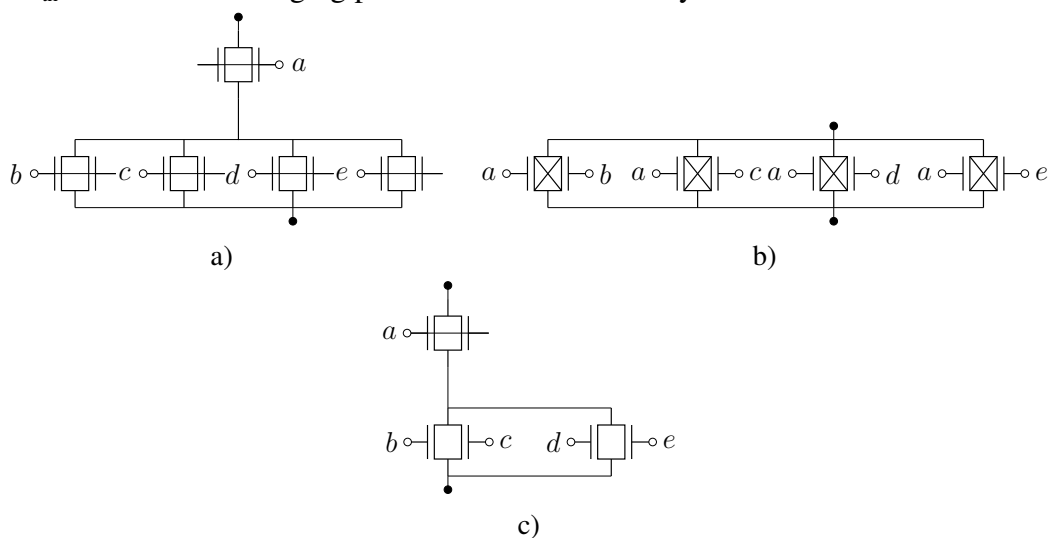
merged, obtaining better results in terms of transistor count. Fig. 3.16 shows the resulting circuit if the literals a, b, c and d are merged using high- V_{th} transistors.

Figure 3.16: A series logic network with: a) FinFETs in SG mode, and, b) reduced network using high- V_{th} FinFETs and merging series transistors



For the case of Fig. 3.16, two transistors could be reduced using High- V_{th} IG FinFETs instead defactorization and the transistor stack could be reduced. Analogously, if defactorization method proposed by (ROSTAMI; MOHANRAM, 2011) is applied to the circuit of Fig. 3.17 a), the circuit of Fig. 3.17 b) is obtained. Note that could be possible to merge parallel transistors instead of using defactorization, and better results could be achieved, as shown in Fig 3.17 c).

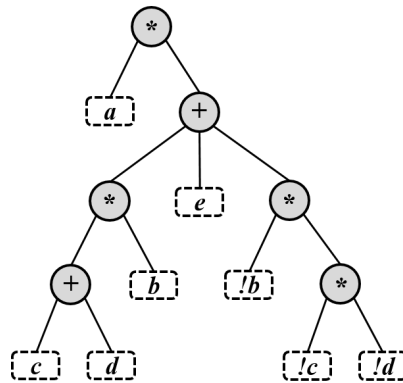
Figure 3.17: A series logic network with: a) FinFETs in SG mode, b) reduced network using defactorization (ROSTAMI; MOHANRAM, 2011), and c) Reduced version with low- V_{th} FinFETs and merging parallel transistors directly



As seen in the previous examples, using directly the defactorization method as proposed by (ROSTAMI; MOHANRAM, 2011) is not the optimal solution. A new method

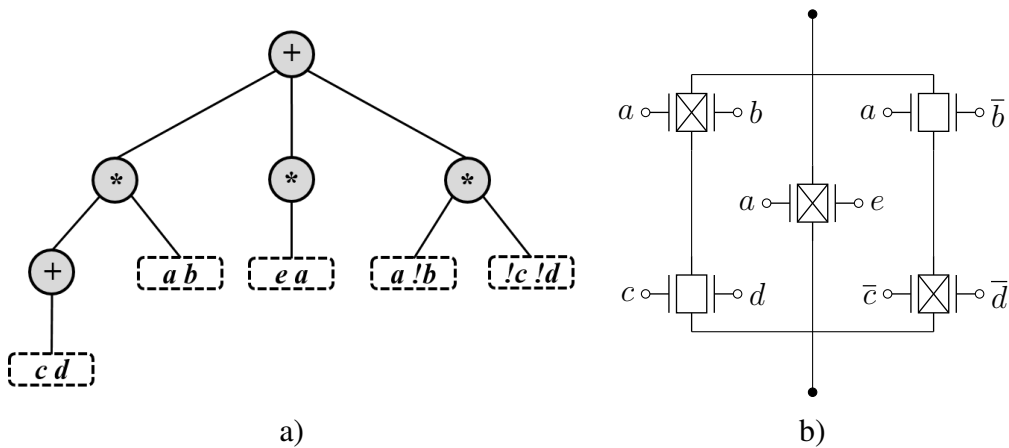
proposed by (POSSANI et al., 2014) and presented in (VALDES et al., 2016), tries to improve the transistor reduction, and looks to the factored branches before applying defactorization, and those procedure could be done by looking to the binary tree of the function. As an example, consider the PDN of the logic function: $F = \overline{a \cdot (b \cdot (d + c) + e + \bar{b} \cdot (\bar{c} \cdot \bar{d}))}$ and its binary tree shown in Fig. 3.18.

Figure 3.18: Binary tree of function F



Using the information of 3.18, low level literals could be merged using IG FinFETs, for example the c and d literals could be merged using a low- V_{th} device, and \bar{c} with \bar{d} are merged using a high- V_{th} one, and would have three single inputs (b, \bar{b} and e). Those signals could be merged with a , getting a reduced implementation, whose tree shown in Fig. 3.19 a) and its circuit using IG FinFETs in Fig. 3.19 b).

Figure 3.19: a) Function F optimized using the method proposed by (POSSANI et al., 2016) and b) its implementation using IG-FinFETs



3.6 Electrical evaluation of compacted networks

In order to evaluate the electrical behavior of network compactions using IG-mode devices, some complex functions (F_1 to F_{10}) are chosen in order to see the impact of optimizing logic networks using defactorization. Five different topologies (named T_1 to T_5) to build transistor networks using IG FinFETs were considered. The first two topologies can be seen as single-gate device implementations because merging of series and parallel transistors is not used. The other three topologies compact the structure by merging transistors in different manners. For all cases, merging of series (parallel) transistors is done by using a high- V_{th} (low- V_{th}) devices. The evaluated topologies are described as follows:

- Topology T_1 : The first topology T_1 is similar to a single-gate device implementation. All transistors are low- V_{th} IG FinFET. One of the gates is driven by an input and the other tied to VDD if the transistor is P-type or GND if the transistor is N-type (IG-LP).
- Topology T_2 : similar to a single-gate device implementation. The difference to T_1 is that the two gates of the transistor are connected to the same input signal (SG mode).
- Topology T_3 : obtained by performing basic reduction of series and parallel transistors in T_1 are merged without logic path defactorization (IG-RED mode).
- Topology T_4 : obtained by performing defactorizations according to (ROSTAMI; MOHANRAM, 2011).
- Topology T_5 : obtained by defactoring logic paths as explained previously.

It is well know that the electrical characteristics of logic gates are greatly influenced by the transistor arrangements. Thus, to carry out the experiments proposed in this work, a set of logic functions were chosen to obtain distinct transistor arrangements when considering different optimization techniques. The evaluated functions are presented as follows:

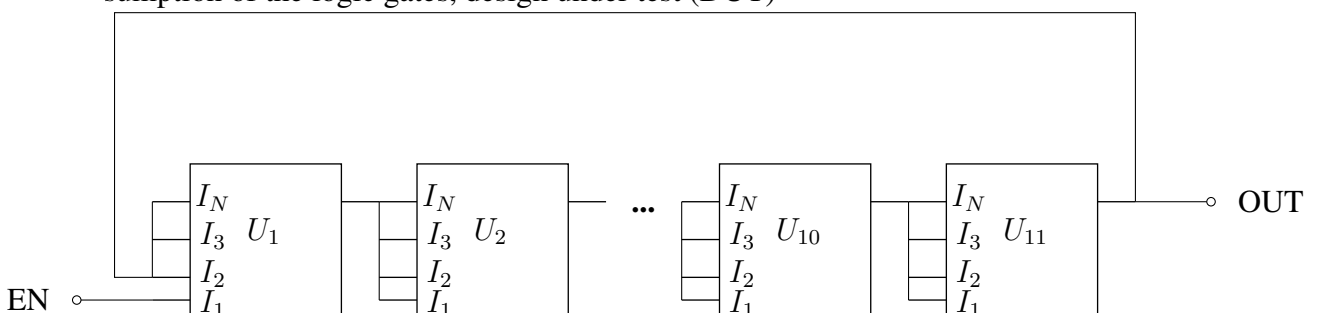
$$\begin{aligned}
 F_1 &= \overline{(a \cdot (b + (c \cdot d \cdot e) + (f \cdot d \cdot h)))} & F_6 &= \overline{(a \cdot ((f + (g \cdot h \cdot i)) \cdot (b + c + d + e)))} \\
 F_2 &= \overline{(a \cdot ((b \cdot c \cdot d) + (e \cdot (f + g)) + h))} & F_7 &= \overline{(a \cdot ((b \cdot (c + d + e + f)) + (g \cdot h \cdot i) + j))} \\
 F_3 &= \overline{(a \cdot ((b \cdot c \cdot d) + e) + f + g)} & F_8 &= \overline{(a \cdot (b \cdot (c + d) + e \cdot ((f \cdot g) + h)))} \\
 F_4 &= \overline{(a \cdot ((b \cdot (c + d + e + f)) + (g \cdot h \cdot i) + j))} & F_9 &= \overline{(a \cdot (((b + c) \cdot d) + e) \cdot (f + g + h))} \\
 F_5 &= \overline{(a \cdot (b \cdot (c + d) + (e \cdot f \cdot g)))} & F_{10} &= \overline{(a \cdot (b \cdot ((c \cdot d) + e + f) + g))}
 \end{aligned}$$

The size of the generated FinFET networks, according to each method discussed above and corresponding to the topologies T_1 to T_5 , is summarized in Table 3.4. It is clear the compaction obtained when applying the defactoring processes presented in (ROSTAMI; MOHANRAM, 2011) (topology T_4) and the one seen previously (topology T_5), resulting in significant circuit area saving. As seen in table 3.4, the proposed implementation method (T_5) explained above, reduces the number of required transistors for implementing the PDN of the evaluated logic functions, because it first search the stand-alone literals before defactoring, obtaining a maximum reduction of 40%.

3.6.1 Methodology

The functions F_1 to F_{10} were implemented using the modes T_1 to T_5 explained above. To measure the performance of each implementation in terms of area and power consumption, SPICE simulations were performed in which each instance was replicated 11 times and connected in cascade, creating a 11-stage ring oscillator as shown in Fig. 3.20. Tying all inputs together has already been adopted as a metric to evaluate the electric behavior of standard cells in the literature (ABOUZEID et al., 2011). It is important to notice that the gate output when all inputs are in high logic level (1) may be different to the case when all inputs are in the low logic level (0). All functions evaluated herein respect this property. In the first stage, some inputs act as enable signal to control the beginning of the oscillation. The oscillation period is considered as gate delay metric whereas the integral of the supply voltage current has been adopted for measuring the power dissipation. The gate dissipation corresponds to the consumption of one period of oscillation, and this period was measured on a node of the circuit between two instants when this node is not transitioning.

Figure 3.20: Ring oscillator used to estimate signal delay propagation and power consumption of the logic gates, design under test (DUT)

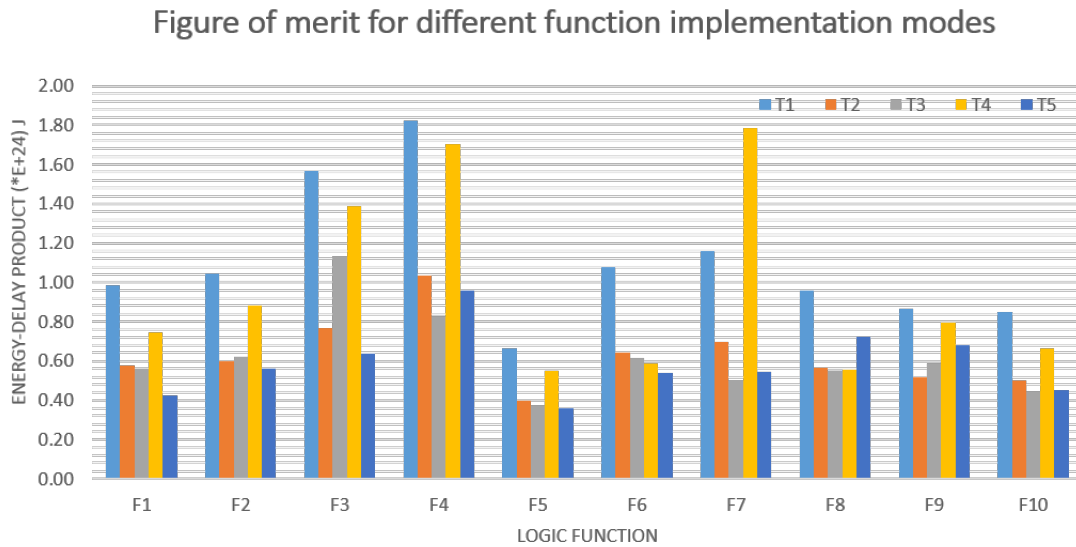


The gate delay and power consumption of the tested topologies were extracted from the transient simulation and shown in Table 3.4. Notice that T_2 implementation mode (SG), provides the lowest delay, but the power consumption tends to be higher. It is expected because the Single-Gate transistor mode (SG) has a better channel control and its threshold voltages is lower.

3.6.2 Results

The energy-delay product of logic gates is given in Fig. 3.21, for all five topologies and their benchmarking functions. It is interesting to notice that the implementation mode T_2 has a good trade-off between area and power compared to the other functions, but the best results are achieved with the proposed defactorization methodology in almost all functions.

Figure 3.21: Energy delay product of each function



Some research works as (CHIANG et al., 2006) study the effects of circuits with high number of stacked transistors and the advantages of reducing stacked transistors with two-gate high- V_{th} devices.

For the defactorization proposed by (ROSTAMI; MOHANRAM, 2011), there are some functions like F_4 and F_7 where the energy-delay product is critically high. That is because the defactorization methods do not consider a maximum number of high- V_{th} series transistors, causing a high delay penalty compared with the methods that use less

Table 3.4: Logic functions implementation results for power, energy and device count using the different implementation modes

Logic Function	Device count					Delay (ns)					Energy (pJ)				
	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
F_1	8	8	6	6	5	2.38	1.34	1.84	2.42	2.82	4.14	4.32	3.06	3.08	1.50
F_2	8	8	6	6	5	2.48	1.37	1.95	2.76	3.35	4.21	4.38	3.17	3.19	1.68
F_3	7	7	5	5	4	4.60	2.02	4.64	5.31	3.77	3.41	3.81	2.44	2.62	1.69
F_4	10	10	7	8	6	3.28	1.80	2.24	3.94	4.26	5.55	5.76	3.69	4.33	2.25
F_5	7	7	5	5	4	1.92	1.11	1.54	2.33	2.24	3.45	3.61	2.44	2.35	1.61
F_6	8	8	6	6	5	2.45	1.40	1.86	1.84	2.86	4.41	4.61	3.30	3.19	1.88
F_7	9	9	6	7	5	2.52	1.45	1.75	4.99	2.59	4.60	4.82	2.86	3.58	2.10
F_8	8	8	6	6	5	2.31	1.31	1.81	1.96	3.12	4.15	4.32	3.05	2.83	2.32
F_9	8	8	6	6	5	2.18	1.25	1.88	2.99	2.85	3.97	4.15	3.11	2.66	2.39
F_{10}	7	7	5	6	4	2.22	1.25	1.66	1.99	2.77	3.83	4.00	2.67	3.33	1.63

series transistors.

In some cases as F_7 and F_{10} , a simple series and parallel transistor merging offers a good figure of merit, but in the other cases is possible to see that the proposed defactorization (POSSANI et al., 2016) method has an excellent power-delay product compared with the other topologies.

3.7 Summary

An interesting functionality of the dual-gate transistor is presented in this chapter, showing an interesting feature of IG FinFET devices. The independent dual gate mode (IG) for channel controllability was used for merging series and parallel devices using high and low- V_{th} devices respectively. Also, some proposed complex logic functions were simulated using different topologies and showing that is possible to reduce the number of transistors using FinFET transistors with dual independent gates. Finally, power and delay simulations of those functions were performed to show that IG FinFET devices could be useful in low leakage and low area designs. With the compaction through logic defactorization (POSSANI et al., 2016), the power consumption is reduced in approximately 54% with regard to single-gate logic gate, and approximately 61% of gate delay increasing.

4 DELAY MODELING OF IG-FINFET LOGIC GATES

4.1 Introduction

Previous chapters show an introduction to FinFETs, their advantages as logic switches, the IG operation mode and its utility for CMOS combinational logic gates, showing that it is possible to create more than one logic topology for the same logic function. IG FinFET devices are useful for reducing the required transistors to implement a logic function due their capability of merging series or parallel pairs of transistors using multi- V_{th} devices, having some consequences in terms of delay and power consumption as seen previously, having a consequence for the system performance.

Transistor sizing is one of the most important steps for logic gate design, because it defines the transient characteristics and output response for different input stimulus. The determination of the appropriate device dimensions is an important step of digital circuit synthesis. In small circuits, those timing constraints could be checked using first-order expressions or electrical simulations, however, in bigger and complex logic gates, simulation time could be impractical.

A fast and reliable method for estimating the delay of a logic gate is achieved by using analytical delay models. Such models can be used in the analysis and optimization of logic circuits as well as in the characterization and sizing of standard cell libraries (MARRANGHELLO; REIS; RIBAS, 2012).

Analytical delay modeling is widely used in digital VLSI design to determine the delay of logic gates using some input parameters such as the input transition time, output load capacitance, threshold voltage, supply voltage and physical parameters of MOS transistors. Usually, to model a delay of a logic gate, first-order approximations are used. One of the most common used is the Elmore delay model, which considers the transistors as linear RC networks switching and charging a load capacitance. This model could be efficient for fast estimations, but could be inaccurate with the presence of short-channel effects which affects the linearity of the switching elements (DUTTA; SHETTI; LUSKY, 1995).

(SAKURAI; NEWTON, 1990), proposed the α -power model to represent the behavior of a short-channel MOS device current operating in linear and saturation regions, introducing the "velocity saturation" and α parameters, which are useful for short-channel devices, also, they proposed an analytical delay model which considers a fast ramp input

and neglects the input to output capacitance (C_M) and short-circuit current.

Some other delay models were developed using the α -power transistor model such (NABAVI-LISHI; RUMIN, 1994) and (DAGA; AUVERGNE, 1999). These models are charge-based, which means that the delay is calculated using the charge flowing through the circuit elements. These models could consider fast or slow input slopes, and are capable of determine the delay of series transistors depending of the position, but they use semi-empirical parameters to calculate the output ramp response (MARRANGHELLO; REIS; RIBAS, 2015). The model presented by (BISDOUNIS et al., 1996) considers only physical parameters, dividing the inverter transient response into operation regions, solving the differential equations for those intervals and approximating the delay using Taylor series approximations, however the model does not consider the transistor SCEs and it is difficult to predict the operating regions of the inverter before calculating its delay.

A practical and comprehensive delay model for IG FinFET was proposed by (DATTA et al., 2007). This model predicts the delay for an inverter using FinFETs in IG mode and extends the delay calculation for series transistors as equivalent RC networks, however, this model neglects the short-channel effects and does not consider logic gates with multi- V_{th} devices.

In this chapter, a delay model for FinFET logic gates in IG mode will be presented. The objective is to analyze the delay of logic gates implemented using multi- V_{th} IG FinFETs. The development of this model is based on (DAGA; AUVERGNE, 1999), (ROSSELLÓ; SEGURA, 2004) and (MARRANGHELLO; REIS; RIBAS, 2015) delay models and extending the analysis for IG devices using the V_{th} expressions shown previously. Also the calculation of the main device parameters for the different modes (SG, IG and LP) is shown, and comparisons with SPICE simulations are performed in order to show the model accuracy.

The rest of the chapter is divided as follows: initially, the static and transient characteristics of IG-FinFET based inverters is shown, analyzing the different operation regions and their influence on the output response. Then, delay of stacks of FinFETs is analyzed, considering the switching device position. Then, the accuracy of the model is shown through a comparison with HSPICE simulations and a summary of the chapter is given.

4.2 IG FinFET inverter delay model

The CMOS inverter is an important logic structure which defines an important set of properties that could be applied to any static CMOS logic gate. It is a simple gate whose properties could be used as a base to model more complex CMOS structures. This chapter will describe the static and dynamic characteristics of those logic gates, and extend the calculations for different IG FinFET devices. Then, the analysis of the inverter analytical model will be extended to series-connected and more complex networks, also, an empirical expression for the calculation of the transition time will be presented.

4.2.1 Static behavior of CMOS inverter

A standard CMOS inverter, is composed by a pair of P-type and a N-type MOS-FET transistors, which set the output to different logic levels depending on the input voltage. Since the switching elements are not ideal, the output signal will have different values depending on the operating region of the transistors, creating a transfer function between the input and output signal in the static domain. Fig. 4.1 shows a CMOS standard transfer curve, in which the output response is divided into sections depending on the operating region of the P and N devices (linear, saturation, cut-off).

As seen in Fig. 4.1, the static response of a CMOS inverter depends on the correspondent MOS device operating region. When $V_{out} = V_{in}$, the current through N and P transistors will be the maximum, this point is called the inverter threshold voltage. The threshold voltage could be moved from left to right of the figure depending on the P/N device ratio (W_p/W_n). Using the static transfer curve, important parameters such as the noise margin and nominal voltage levels are defined.

4.2.2 Dynamic behavior of CMOS inverter

To model the delay of a CMOS inverter, the transient response curve is used. For a standard CMOS logic structure, three operation regions could be defined: overshoot, short-circuit and discharge as seen in Fig. 4.2.

Applying the Kirchoff's current law in V_o , the input to output voltage will be

Figure 4.1: Static transfer curve (V_{out} vs. V_{in} of a CMOS inverter)

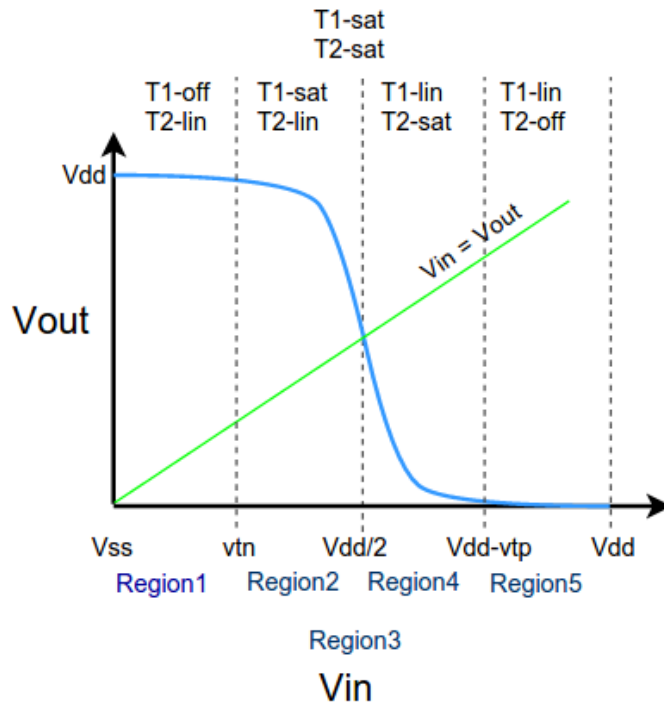
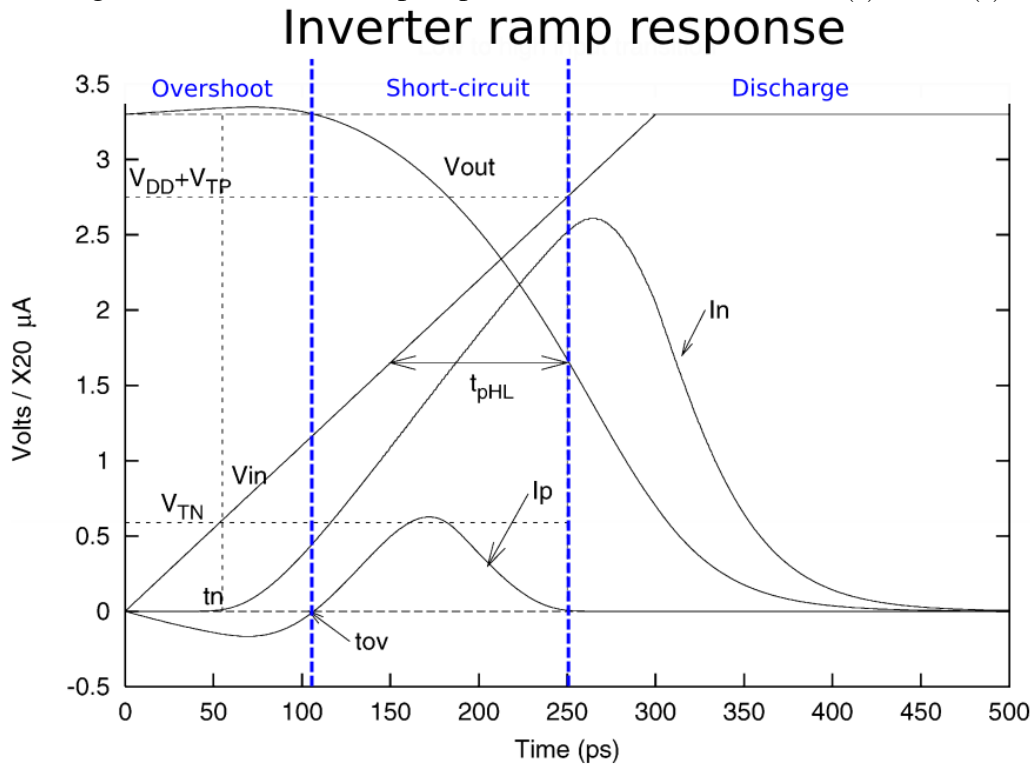


Figure-2

Figure 4.2: Measured ramp response of a CMOS inverter ($V_o(t)$ vs. $V_{in}(t)$)



Source: (ROSSELLÓ; SEGURA, 2004)

expressed as:

$$\frac{dV_o(t)}{dt}(C_M + C_L) - C_M \frac{dV_{in}(t)}{dt} = I_p - I_n \quad (4.1)$$

4.2.2.1 Overshoot region

As seen in Fig. 4.2, in $t = 0$ the output stays in high logic level $V_o(t) = V_{DD}$. when $V_{in}(t)$ starts to increase, the output voltage starts to increase to a value higher than the supply voltage, mainly caused by two factors: the input slope and the I/O coupling capacitance (called C_M later in this work). The overshoot range ends when the accumulated charge in C_M is evacuated through P and N transistors and the load capacitance (C_L). As seen in Fig. 4.2, the current through P transistor is negative along the overshoot range.

It is observed that the output overshoot magnitude is higher when the input slope gets faster or the P/N ratio is high. However, the overshoot amplitude is limited to a few percent of the supply-voltage value. This overshoot value is important, because it happens when the current flow through the devices is low (the P transistor is in linear region and the N device is off or saturated with low V_{GS}), being the recovery time after the overshoot an important component of the total delay (DAGA; AUVERGNE, 1999).

4.2.2.2 Short-circuit region

In this region, the P transistor moves from linear region to saturation and the N transistor moves from cut-off to saturation as seen in Fig. 4.1. Basically the load capacitance C_L , and C_M starts to discharge under the total $I_p - I_n$ current.

As seen in Eq. 4.1, this region also depends on the input slope (dV_{in}/dt), the faster the input ramp, its derivative will be higher and the current difference through N and P transistors $I_p - I_n$ gets lower, until the load will be discharged using the N-type device only, becoming to enter to the discharge region.

The short-circuit zone, which depends on input slope, load capacitance and P/N ratio of the transistors, is the most difficult region to be modeled (DAGA; AUVERGNE, 1999).

4.2.2.3 Discharge region

In the discharge region, the N-type transistor is active, discharging the load capacitance directly (the current through P-type transistor is negligible). This region is also

dependent on the input slope and load capacitance dependent. For fast inputs, the P-type transistor is in cut-off region and the N-type transistor is in linear region, acting as a low impedance path for load discharging. For slow input ramps, the input is already lower than V_{DD} , letting the current flow through N and P transistors.

4.2.3 Delay definition

For a logic gate, the delay from the input to the output is defined as the elapsed time between the input and output when their value cross the half part of the supply voltage. For an input i of a logic network, the $t_{dLH}(i)$ defines the time between the falling input signal of the input i reaches $V_{DD}/2$ and the rising output crosses $V_{DD}/2$, and $t_{dHL}(i)$ defines the time between the half of the rising signal of the input i and the half of the falling output signal.

4.2.4 Transistor current equations

To model the delay of logic gates built using MOS transistors, the α -power current model is used to model the transistor current characteristics (SAKURAI; NEWTON, 1990). This model is preferred because it considers devices affected by SCEs, being accurate for modern feature sizes. It uses the velocity saturation index to give a simple expressions for the drain current considering the different operating regions. The α -power current model considers the mobility degradation of short-channel devices caused by electron scattering. The velocity saturation index (called α) is an empirical parameter which is related to the saturation velocity of the carriers, and usually varies between 1 and 2. Being 1 for devices totally affected by mobility degradation and 2 for long-channel devices. For an independent-gate FinFET, the α -power model is still applicable and could be expressed as:

$$I_D = \begin{cases} 0 & \text{Cut-off} \\ K_{lin} N_{FIN} H_{FIN} (V_{G1S} - V_{th})^{\frac{\alpha}{2}} V_{DS} & \text{Linear} \\ K_{sat} N_{FIN} H_{FIN} (V_{G1S} - V_{th})^{\alpha} (1 + \lambda V_{DS}) & \text{Saturation} \end{cases} \quad (4.2)$$

Eq. 4.2 shows a drain current dependence on the number of fins and the fin height (N_{FIN} and H_{FIN} respectively). It clearly shows that the current values are discrete, and

depend on the number of fins of the device. As seen previously, the threshold voltage depends on the back-gate voltage, and the dependence is linear and proportional to the physical dimensions of the device. Using the information of Eq. 2.21, the threshold voltage for a FinFET could be expressed as:

$$V_{\text{th}}(V_{\text{G}_2\text{S}}, V_{\text{DS}}) = \begin{cases} V_{\text{th0(IG)}} - rV_{\text{G}_2\text{S}} - \eta V_{\text{DS}} & \text{IG-Mode} \\ V_{\text{th0(SG)}} - \eta V_{\text{DS}} & \text{SG-Mode} \end{cases} \quad (4.3)$$

Where r is the coupling factor seen previously and η is the DIBL coefficient. Notice that those equations neglect the body effect due the FinFET properties, as shown in chapter 2. All values are extracted from HSPICE simulations using the chosen compact model and parameters shown previously.

4.2.5 Inverter virtual step response

Daga and Auvergne defined the step response of an inverter as an important metric to measure the performance of an inverter structure. It could be a first order approximation to calculate the total delay response of an inverter. For this case, the input signal is an unitary step $V_{\text{in}}(t) = V_{\text{DD}}\mu(t)$ for the case of a rising signal. To evaluate the output response, the charge conservation law is applied in the output node, being possible to define the output voltage evolution (ΔV_o) as the average charge transferred from the load (C_L) through the switching element, neglecting the charge of the equivalent I/O coupling capacitance and the short-circuit current, the average charge transferred to the N-type device is defined as $Q_{N_{\text{avg}}}$.

$$Q_{N_{\text{avg}}} = C_L \Delta V_o \quad (4.4)$$

For an ideal step input voltage $V_{\text{in}}(t)$, the charge $Q_{N_{\text{avg}}}$ could be defined as the integral of the average current flowing through the N transistor (I_{avg}). For the case of a short-channel device, the saturation current dominates the process of charging and discharging the load capacitances, and the modeling process becomes simpler. To know the time delay between the starting of the transition and the half of the output voltage excursion, the time delay derived from a step response is defined ($t_{d_{\text{HLS}}}$) as the time interval at

which the output goes from V_{DD} to $V_{DD}/2$, thus Eq. 4.4 becomes:

$$\int_0^{t_{d_{HLs}}} I_n(t) dt = C_L \frac{V_{DD}}{2} \quad (4.5)$$

For short-channel devices, the saturation voltage is below $V_{DD}/2$ for $V_{GS} = V_{DD}$ and it could be seen also in FinFETs operating in SG and IG modes. However, for devices in SG mode the saturation voltage is higher (close to $V_{DD}/2$ as seen in Table 4.1), the devices will be considered in saturation region during the output range. Using the α -power expressions of Eq. 4.2, an expression for the maximum average current flowing through the NMOS transistor is:

$$I_{max} = K_N H_{FINn} N_{FINn} (V_{DD} - V_{th})^\alpha (1 + \lambda V_{DS_{avg}}) \quad (4.6)$$

From Eq. 4.6, α is the velocity saturation index (SAKURAI; NEWTON, 1990), and λ represents the channel modulation effect in saturation region. (MARRANGHELLO; REIS; RIBAS, 2015) proposed an expression for V_{DS} . Neglecting the I/O coupling capacitance, V_{DS} could be approximated as the average of the maximum and minimum value of the output excursion, so $V_{DS_{avg}} = 0.5(V_{DD} + V_{DD}/2) = 0.75V_{DD}$.

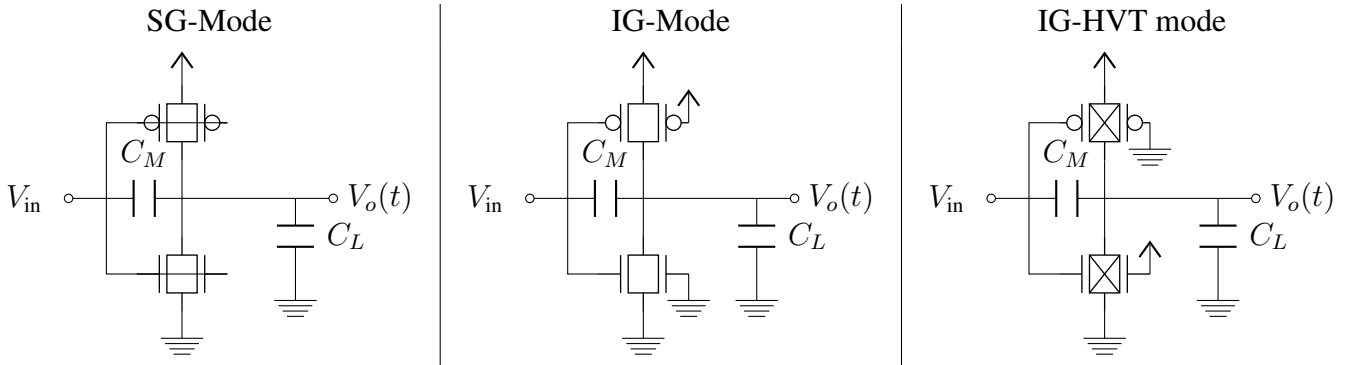
Using the information of Eqs. 4.4 and 4.6, the delay response from a step input is defined as:

$$t_{d_{HLs}} = \frac{C_L V_{DD}}{2K_N H_{FINn} N_{FINn} (V_{DD} - V_{thn})^\alpha (1 + 0.75\lambda V_{DD})} \quad (4.7)$$

Using IG FinFET devices, three different modes will be defined: when the second gate is zero biased $V_{G_{2S}} = 0$ (IG mode), when both gates are biased to the same potential $V_{G_{1S}} = V_{G_{2S}}$ (SG mode) and when a high- V_{th} device is used as a switching element (IG-red mode), which is useful to reduce two-device stacks as seen in the previous chapter. Fig. 4.3 shows the different evaluated configurations for IG FinFET inverters.

Notice that for high- V_{th} devices the second gate interface needs to be biased ($V_{G_{2S}} = V_{DD}$). With this, the transistor is activated and could generate a valid output transition from high to low (for the case of N-type devices).

Figure 4.3: IG FinFET inverter modes for delay analysis



4.2.6 Inverter ramp response

To model the real behavior of the inverter, the information of Fig. 4.2 and Eq. 4.1 are useful. The input ramp could be defined as:

$$V_{in}(t) = \begin{cases} 0 & \text{if } t < 0 \\ V_{DD} \frac{t}{\tau_{in}} & \text{if } 0 \leq t \leq \tau_{in} \\ V_{DD} & \text{if } t > \tau_{in} \end{cases} \quad (4.8)$$

Where τ_{in} is the input ramp duration. As shown in Fig. 4.2 and mentioned previously, there are three important operation regions for a CMOS inverter: overshoot, short-circuit and discharge. The output response for fast and slow inputs will be modeled.

Using the information of Eq. 4.1, the output voltage variation $dV_o(t)/dt$ depends on the P and N current difference ($I_p - I_n$), the input variation ($dV_{in}(t)/dt$) and the capacitances (C_M and C_L).

To get an accurate response, some authors use different approaches to solve the differential equation by neglecting the PMOS current for fast inputs (as seen in (DAGA; AUVERGNE, 1999) or (ROSSELLÓ; SEGURA, 2004)) using empirical parameters to fit the delay response, or dividing the response in 6 different regions which depend on the transistor operating regions (as seen in (BISDOUNIS; NIKOLAIDIS; KOUFOPAVLOU, 1998)).

For fast inputs, the delay response could be obtained using charge conservation laws, similarly to (MARRANGHELLO; REIS; RIBAS, 2015) and (ROSSELLÓ; SEGURA, 2004) and used previously for the case of the virtual step response. The amount of charge delivered to the NMOS transistor will depend on the charge delivered by the I/O coupling capacitance (C_M) and the charge delivered by the output capacitance (C_L^*),

which contains the load capacitance and the diffusion capacitance of the P and the N transistors respectively.

$$C_L^* = C_L + C_d^p + C_d^n \quad (4.9)$$

As defined by (ROSSELLÓ; SEGURA, 2004), the charge evacuated through NMOS/PMOS device (for rising or falling inputs respectively) $Q_{n,p}$ could be generically expressed as the sum of four different charge components:

$$Q_{n,p} = Q_{n,p_{lin}^{rise}} + Q_{n,p_{sat}^{rise}} + Q_{n,p_{lin}^{high}} + Q_{n,p_{sat}^{high}} \quad (4.10)$$

Where $Q_{n,p_{lin}^{rise}}$ and $Q_{n,p_{sat}^{rise}}$ define the charge evacuated by the NMOS/PMOS transistor when the input is rising/falling and the transistor is in linear and saturation regions respectively, $Q_{n,p_{lin}^{high}}$ and $Q_{n,p_{sat}^{high}}$ are the charge components evacuated by the NMOS/PMOS device when the input is high and the transistor is in linear and saturation regions, respectively.

Notice that Eq. 4.10 also could be used to define the current flowing through PMOS transistors for the case of a rising output. The charge expression shown in Eq. 4.10 is fundamental to describe the delay response for fast and slow input ramps. For each case, the charge components could be neglected and an expression for the delay could be obtained.

4.2.6.1 Fast input ramps

For fast input ramps, the output voltage goes from approximately V_{DD} to $V_{DD}/2$, range in which the switching device (NMOS for falling outputs and PMOS for rising ones) could be considered in saturation region, as seen previously for the virtual step response. Using this assumption, for falling output transitions, the charge evacuated by the NMOS device could be expressed as:

$$Q_n = Q_{n_{sat}^{rise}} + Q_{n_{sat}^{high}} \quad (4.11)$$

$Q_{n_{sat}^{rise}}$ could be calculated using the integral of the drain current for the rising input ramp of Eq. 4.54:

$$Q_{n_{sat}^{rise}} = \int_{t_n}^{\tau_{in}} I_n^{rise}(t) dt \quad (4.12)$$

Where t_n represents the time in which the input voltage reaches the NMOS threshold voltage for $V_{in}(t) = V_{th}$ and τ_{in} represents the input ramp duration. For a PMOS transistor the analysis is similar. I_n^{rise} could be expressed as:

$$I_n^{rise}(t) = K_n H_{FINn} N_{FINn} \left(\frac{V_{DD}}{\tau_{in}} t - V_{th} \right)^\alpha (1 + \lambda V_{DSavg}) \quad (4.13)$$

V_{DSavg} was defined by (MARRANGHELLO; REIS; RIBAS, 2015) to consider the channel modulation parameter (λ). It could be defined as the average between the maximum voltage (generated by C_M capacitance in the overshoot region) and $V_{DD}/2$:

$$V_{DSavg} = \frac{1}{2} \left(\frac{V_{DD}}{2} + V_{max} \right) \quad (4.14)$$

And V_{max} could be calculated as:

$$V_{max} = V_{DD} \left(1 + \frac{C_M}{C_M + C_L} \right) \quad (4.15)$$

Solving the Eq. 4.12 using Eqs. 4.17 to 4.15 the rising saturation charge $Q_{n_{sat}}^{rise}$ is expressed as:

$$Q_{n_{sat}}^{rise} = K_n H_{FINn} N_{FINn} (1 + \lambda V_{DSavg}) \int_{t_n}^{\tau_{in}} \left(\frac{V_{DD}}{\tau_{in}} t - V_{thn} \right)^\alpha dt \quad (4.16)$$

$$Q_{n_{sat}}^{rise} = \frac{K_n H_{FINn} N_{FINn} \tau_{in} (1 + \lambda V_{DSavg})}{V_{DD} (\alpha + 1) (V_{DD} - V_{thn})^\alpha} \left(\frac{V_{DD}}{\tau_{in}} t - V_{thn} \right)^{\alpha+1} \Bigg|_{t_n}^{\tau_{in}}$$

$$Q_{n_{sat}}^{rise} = \frac{K_n H_{FINn} N_{FINn} \tau_{in} (1 + \lambda V_{DSavg})}{V_{DD} (\alpha + 1) (V_{DD} - V_{thn})^\alpha} (V_{DD} - V_{th})^{\alpha+1}$$

And could be simplified to:

$$Q_{n_{sat}}^{rise} = \frac{I_{max} \tau_{in} (V_{DD} - V_{thn})}{(1 + \alpha) V_{DD}} \quad (4.17)$$

Where I_{max} is the maximum saturation current defined in Eq. 4.6 with V_{DSavg} calculated using Eq. 4.15.

Similarly, the value of $Q_{n_{sat}}^{high}$ is calculated using the following expression:

$$Q_{n_{sat}}^{high} = \int_{\tau_{in}}^{t_{0.5}} I_{max} dt \quad (4.18)$$

Where I_{max} is the maximum current which is considered constant during the tran-

sition. $t_{0.5}$ is the time when the output is $V_{DD}/2$, where the delay time $t_{d_{HL,LH}}$ is calculated as:

$$t_{d_{HL,LH}} = t_{0.5} - \tau_{in}/2 \quad (4.19)$$

$$Q_{n_{sat}}^{high} = I_{max} \int_{\tau_{in}}^{t_{0.5}} dt = I_{max}[t_{0.5} - \tau_{in}] \quad (4.20)$$

The total charge evacuated by the transistor Q_n is the difference between the initial charge in the output node $V_{DD}(C_M + C_L)$ and the charge when $t = t_{0.5}$, which for fast input ramps could be defined as $(C_L + C_M)V_{DD}/2 - C_M V_{DD}$, thus:

$$Q_{total} = V_{DD}(C_M + C_L) - \left(\frac{V_{DD}}{2}(C_L + C_M) - C_M V_{DD} \right)$$

$$Q_{total} = \frac{V_{DD}(C_M + C_L)}{2} + C_M V_{DD} \quad (4.21)$$

Using Eqs. 4.11 and 4.21 and using the charge conservation principle, the charge evacuated through the switching element must be equal to the charge stored in the capacitances (Q_{total}). Thus, if 4.11 is equal to Q_{total} , and applying some algebra, $t_{0.5}$ could be expressed as (MARRANGHELLO; REIS; RIBAS, 2015):

$$t_{0.5fast} = \frac{Q_{total}}{I_{max}} + \frac{\tau_{in}(\alpha V_{DD} + V_{th})}{V_{DD}(\alpha + 1)} \quad (4.22)$$

Eq. 4.22 could be used for rising or falling output transitions. The analysis is similar to the case of PMOS devices.

4.2.6.2 Fast and slow input reference time (t_{ref})

(MARRANGHELLO; REIS; RIBAS, 2015) defined the particular case when $\tau_{in} = t_{0.5}$ to define whether an input is fast or slow. It could be found using the Eq. 4.21 as:

$$t_{ref} = \frac{Q_{total}(\alpha + 1)}{I_{max} \left(1 - \frac{V_{th}}{V_{DD}} \right)} \quad (4.23)$$

if $\tau_{in} < t_{ref}$, the input is considered as a fast stimulus in which the previous analysis could be applied. When $\tau_{in} \geq t_{ref}$, the input ramp does not reach its final value (V_{DD}) when the output crosses $V_{DD}/2$, which its analysis has different properties in terms of charge and needs to be considered in order to have a complete model of the inverter

structure. Also, t_{ref} is useful to calculate more parameters such as the transition time.

4.2.6.3 Slow input ramps

For slow inputs, the delay analysis is different from the case of fast ramps and its analysis is difficult to be modeled. In this case, the current through both transistors needs to be considered (called short-circuit current). (DAGA; AUVERGNE, 1999) defined an empirical expression for the delay of slow input ramps, which considers the P/N ratio of the transistors, the threshold voltage of the switching device (NMOS for falling outputs) and the fast input delay response using empirical constants (ϕ, β, γ) which are obtained through SPICE simulations of the different inverter configurations.

$$t_{d_{\text{HL}}} = t_{d_{\text{HL,fast}}} \left[1 - \frac{\phi(1 - V_{\text{th}n}/V_{\text{DD}})}{1 + \beta \frac{N_{\text{FIN}p}H_{\text{FIN}P}}{N_{\text{FIN}n}H_{\text{FIN}N}}} \left(\frac{\tau_{\text{in}}}{t_{d_{\text{HL}s}} \right)^\gamma \right] \quad (4.24)$$

Some regressions were performed using simulation data and multiple input transition times in order to have values for these parameters. However, when multiple- V_{th} networks are used, the model loses accuracy and gives wrong values for the delay, needing some recalibration to get new values of the constants.

Another approach given by (ROSSELLÓ; SEGURA, 2004) proposed an empirical expression for the calculation of the short-circuit current, which depends on the overshoot time. The model of short-circuit current also needs empirical constants, which needs calibration as the previous model. (MARRANGHELLO; REIS; RIBAS, 2015) proposed an expression for the SCCT which approximates the short-circuit current and calculates the short-circuit duration (using the t_{ref}) and the threshold voltage of P and N devices.

In this case, the output reaches $V_{\text{DD}}/2$ before the input slope reaches its maximum value ($V_{\text{in}}(t_{0.5}) < V_{\text{DD}}$), and similarly from the previous case, the main switching device (PMOS for falling inputs and NMOS for the rising ones) is considered in saturation region. For rising inputs, Q_n (defined in Eq. 4.10) becomes:

$$Q_n = Q_{n_{\text{sat}}}^{\text{rise}} \quad (4.25)$$

And could be expressed as:

$$Q_n = \int_{t_n}^{t_{0.5}} K_n N_{\text{FIN}n} H_{\text{FIN}n} \left(\frac{V_{\text{DD}}}{\tau_{\text{in}}} t - V_{\text{th}n} \right)^\alpha (1 + \lambda V_{\text{DSavg}}) dt \quad (4.26)$$

$$Q_n = \frac{I_{\max}}{(V_{DD} - V_{thn})^\alpha} \int_{t_n}^{t_{0.5}} \left(\frac{V_{DD}}{\tau_{in}} t - V_{thn} \right)^\alpha dt$$

The resolution of Eq. 4.26 is similar from the case of $Q_{n_{\text{sat}}}$ for fast input ramps. The main difference from 4.26 is that the input does not reach its maximum value when the output reaches $V_{DD}/2$. the charge evacuated by the NMOS transistor could be expressed as:

$$Q_n = \left(\frac{I_{\max}}{(V_{DD} - V_{thn})^\alpha} \right) \left(\frac{\tau_{in} \left(\frac{V_{DD}}{\tau_{in}} t - V_{thn} \right)^{\alpha+1}}{(1 + \alpha)V_{DD}} \right) \Bigg|_{t_n}^{t_{0.5}} \quad (4.27)$$

Evaluating Q_n in the timing interval, $(t_n, t_{0.5})$, and applying the conservation charge law, the total charge evacuated by the transistor Q_n is equal to the charge delivered by C_M and C_L (Q_{total}). The value of $t_{0.5}$ could be calculated as (MARRANGHELLO; REIS; RIBAS, 2015):

$$t_{0.5} = \left(Q_{\text{total}} \frac{\tau_{in}^\alpha (\alpha + 1) (V_{DD} - V_{thn})^\alpha}{I_{\max} V_{DD}^\alpha} \right)^{1/(\alpha+1)} + t_n \quad (4.28)$$

For slow input ramps, the total charge Q_{total} delivered by C_L and C_M is different from the case of fast input ramps, because it needs to include also the short-circuit component. As previously mentioned, the short-circuit charge is difficult to be modeled and some authors give empirical expressions which depend on the P/N ratio and empirical constants calibrated from simulations (DAGA; AUVERGNE, 1999), and other uses linear approximations of the saturation current of the complementary switching device. Both approaches were tested, and the best results were obtained using linear expressions to calculate the short-circuit charge component (as seen in (MARRANGHELLO; REIS; RIBAS, 2015)). For fast inputs the short-circuit current is neglected, so the t_{ref} factor could be considered as the start time of the short-circuit current. The short circuit current τ_{sc} duration could be approximated as (MARRANGHELLO; REIS; RIBAS, 2015):

$$\tau_{sc} = (\tau_{in} - t_{\text{ref}}) \left(\frac{V_{DD} - V_{thn} - |V_{thp}|}{V_{DD}} \right) \quad (4.29)$$

Where τ_{sc} represents the duration of the short-circuit current. Notice that the short-circuit duration is zero when the input slope is equal to the reference time, and increases when the input slope decreases. (MARRANGHELLO; REIS; RIBAS, 2015) defines an average current in the short-circuit regime, considering an average gate-to-source voltage

which produces an equivalent current for the short-circuit regime, called V_{ov} and defined as:

$$V_{ovp} = \left(1 - \frac{t_{ref}}{\tau_{in}}\right) \left(\frac{V_{DD} - V_{thn} - |V_{thp}|}{2}\right) \quad (4.30)$$

And the average SCCT current I_{scp} of the PMOS is obtained using the α -power law, as:

$$I_{scp} = K_p N_{FINp} H_{FINp} V_{ovp}^\alpha (1 + \lambda_p V_{DSavgp}) \left(1 - \frac{t_{ref}}{\tau_{in}}\right) \quad (4.31)$$

The SCCT component could be approximated as: $Q_{sc} = I_{scp} \tau_{sc}$

And the total charge delivered to the transistor, is, as in the previous case, the difference between the initial and final charges (when $t = t_{0.5}$) delivered by the PMOS transistor (for rising inputs) (Q_{sc}), C_M and C_L :

$$Q_{total} = Q_{sc} + C_M V_{DD} \frac{t_{ref}^{\frac{1}{1+\alpha}}}{\tau_{in}} + \frac{V_{DD}}{2} (C_M + C_L) \quad (4.32)$$

Using the information of Eqs. 4.28 and 4.32 it is possible to calculate the delay of slow inputs using IG FinFET devices, as seen in the next sections. Next, the parameters for the α -power model for IG FinFETs are extracted from electrical simulations and analysis of the different modes will be performed.

4.3 Inverter delay analysis for IG-FinFETs

To test the delay model accuracy, some tests were performed using inverters with FinFETs in IG mode and the configurations shown in Fig. 4.3. In order to apply the delay equations shown previously, the device parameters for α -power model were calculated. Also, the physical parameters such as input-to-output capacitance (C_M) and input capacitance (C_{in}) were estimated.

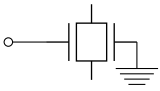
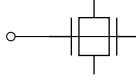
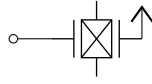
4.3.1 Transistor parameters

In order to extract the important transistor parameters mentioned above, the information of (SAKURAI; NEWTON, 1991) and operation curves of the FinFET in IG Mode were used. The operation curves were extracted using HSPICE simulations and the model

card parameters shown in Table 2.4.

The α -power model extracted parameters are shown in Table 4.1 for N and P-type IG FinFETs, and the I_D - V_{DS} curves are shown in Fig. 4.4 for N-type transistors operating in shorted-gate (SG) mode, independent gate (IG) mode ($V_{G_{2S}} = 0$) and for a high- V_{th} operating in IG mode ($V_{G_{2S}} = V_{DD}$).

Table 4.1: Alpha-power parameters for IG FinFETs operating in SG, IG and High- V_{th} devices

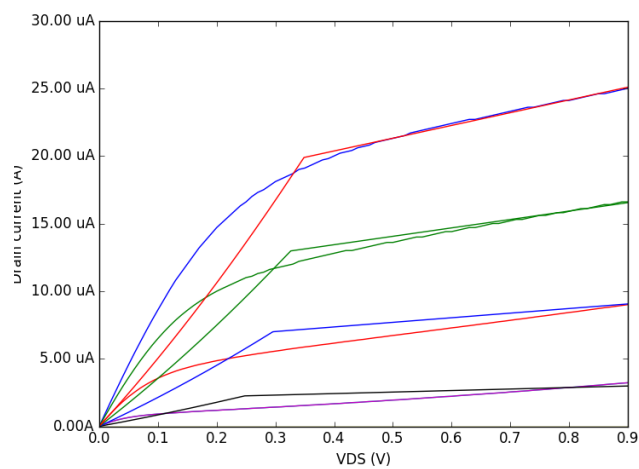
Mode	Parameter	Value	
		N-type	P-type
	K_{sat}	35.74μ	13.53μ
	K_{lin}	90.69μ	41.46μ
	α	1.359	1.617
	λ	0.568	0.895
	$ V_{D_{sat}} $	0.348	0.324
	$ V_{th0} $	0.331	0.327
	K_{sat}	51.56μ	23.5μ
	K_{lin}	83.35μ	38.40μ
	α	1.099	1.308
	λ	0.500	0.810
	$ V_{D_{sat}} $	0.464	0.449
	K_{sat}	23.53μ	10.38μ
	K_{lin}	80.17μ	27.94μ
	α	1.732	1.6359
	λ	0.362	0.503
	$ V_{D_{sat}} $	0.278	0.321
	$ V_{th0} $	0.276	0.263

4.3.2 Physical parameters

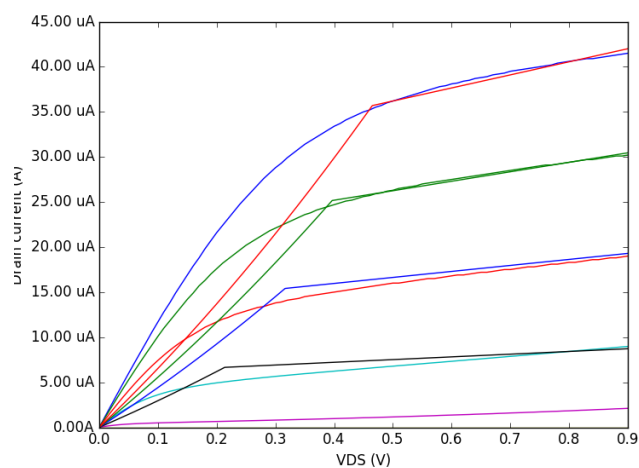
Besides the transistor voltage and current characteristics, C-V parameters are needed to model the transient characteristics of the logic gates. Those parameters are responsible for the delay in a CMOS circuit.

The intrinsic and extrinsic capacitance components play an important role in the circuit performance. To calculate the propagation delay of a CMOS circuit, important parameters such input to output and drain capacitances (C_M and C_p respectively) are needed to estimate the transient response of a standard CMOS inverter structure. Those equivalent capacitance values depend on physical parameters of the MOS structure (such $t_{ox1,2}$,

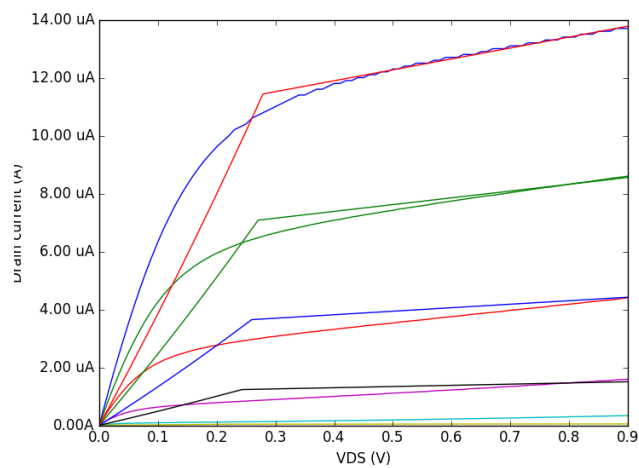
Figure 4.4: I_D vs. V_{DS} curves simulated and modeled using α -power for an N-type IG-FinFET device



(a) IG Mode



(b) SG Mode



(c) IG Mode for a High- V_{th} device

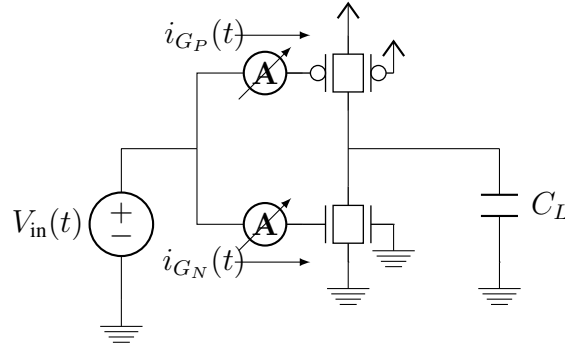
fin height, length and overlap/underlap dimensions) and the inversion charge distribution, which is voltage dependent (V_{G1S} and V_{G2S}).

To calculate the gate source capacitance and gate to drain capacitance, the device operation region needs to be considered. (FOSSUM; TRIVEDI, 2013) proposed an expression for an asymmetrical double gate device which is dependent on the coupling factor (r) and the charge centroid position, seen previously.

$$C_{G_{1,2eq}} \cong \frac{C_{ox1}}{1 - \frac{r}{3} \frac{x_c}{t_{ox1}}} \quad (4.33)$$

Equation 4.33 shows that for an independent double-gate device, the total gate capacitance could be higher than $C_{ox_{1,2}}$, and is dependent on the gate coupling factor (r). This is quite difficult to model, because depends on the charge characteristics along the channel and the volume inversion which are difficult to be estimated in thin-film devices. To calculate the different capacitance values, transient simulations were run and the charge through the gate was measured for a minimum-ratio inverter, the test bench was performed as shown in Fig. 4.5.

Figure 4.5: IG FinFET inverter modes for delay analysis



To measure the gate capacitance, the charge flowing through gate of N and P transistors, the front gate charge was measured by the following equation:

$$Q_{G_P} = \int_{t_0}^{t_0 + \tau_{in}} i_{G_P}(t) dt \quad (4.34)$$

$$Q_{G_N} = \int_{t_0}^{t_0 + \tau_{in}} i_{G_N}(t) dt \quad (4.35)$$

And the gate capacitance is calculated by the following equations:

$$C_{G_N} = \frac{Q_{G_N}}{\Delta V_{in}(t)} \quad (4.36)$$

$$C_{G_P} = \frac{Q_{G_P}}{\Delta V_{in}(t)} \quad (4.37)$$

Where $\Delta V_{in}(t) = V_{DD}$. The input capacitance of an inverter structure is the sum of the gate capacitances of N and P-type transistors.

$$C_{ineq} = C_{G_P} + C_{G_N} \quad (4.38)$$

For a n-input logic gate, the input capacitance for the input (i) depends on the capacitance of the i -th switching N-type transistor and its complementary P-type transistor, thus:

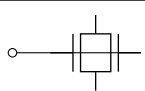
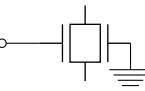
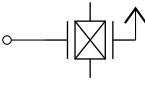
$$C_{ineq}(i) = C_{G_P}(i) + C_{G_N}(i) \quad (4.39)$$

And $C_{M_{avg}}$ is the average Miller equivalent capacitance, defined as (MARRANGHELLO; REIS; RIBAS, 2015):

$$C_{M_{avg}} = \frac{0.5C_{ox1p}N_{FINp}H_{FINp}L(V_{DD} - |V_{thp}|) + C_{ovp}|V_{thp}|}{V_{DD}} \quad (4.40)$$

The equivalent average input to output coupling capacitance is calculated using Eq. 4.40, which considers the overlap/underlap capacitance and the $C_{ox1,2}$ per unit area capacitance. Table 4.2 shows the measured equivalent gate capacitance ($C_{G_{N,P}}$), equivalent gate oxide capacitance ($C_{ox1,2}$) and input to output coupling capacitance ($C_{M_{avg}}$).

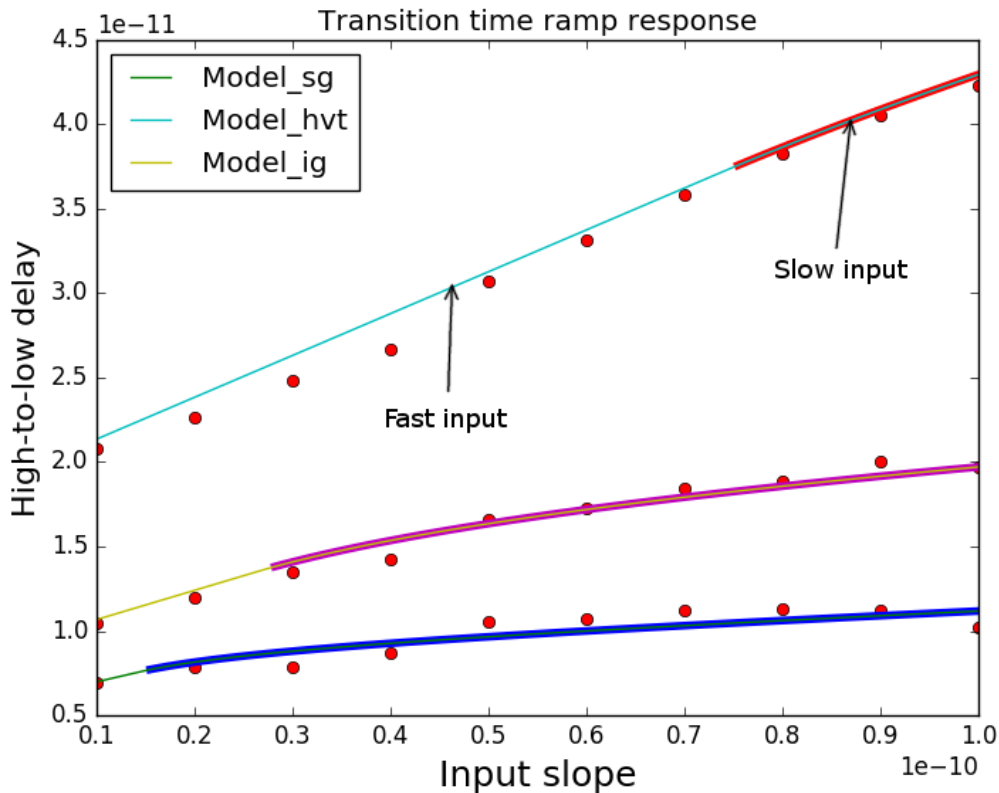
Table 4.2: Extracted capacitance values for IG FinFETs, for the different operating modes

Mode	Parameter	Value	
		N-type	P-type
	$C_{ox1,2}$ [fF/nm ²]	0.056	0.055
	$C_{M_{avg}}$ [fF]	0.075	0.073
	$C_{G_{N,P}}$ [fF]	0.153	0.149
	$C_{ox1,2}$ [fF/nm ²]	0.043	0.051
	$C_{M_{avg}}$ [fF]	0.588	0.068
	$C_{G_{N,P}}$ [fF]	0.118	0.139
	$C_{ox1,2}$ [fF/nm ²]	0.010	0.011
	$C_{M_{avg}}$ [fF]	0.014	0.015
	$C_{G_{N,P}}$ [fF]	0.028	0.031

4.3.3 Delay evaluation of inverter gates

To test the model accuracy, some simulations were performed using the BSIM-IMG model, HSPICE and the model parameters shown in Table 2.4. The simulation results were compared using the delay model proposed equations and parameters. The delay model algorithm was written in `python`. Fig. 4.6 shows the propagation delay $t_{d_{HL}}$ for minimum size inverters in SG, IG mode and using high- V_{th} devices, varying the input slope from 10 ps to 100 ps and $C_L = 0.5$ fF. Fig. 4.7 shows the delay response ($t_{d_{HL}}$ and $t_{d_{LH}}$) of inverters in IG mode varying the N_{FIN_p}/N_{FIN_n} ratio from 0.5 to 8. Also, Fig. 4.8 shows the delay response ($t_{d_{HL}}$ and $t_{d_{LH}}$) for inverters in IG mode with different fanout loads (0.5 fF, 1 fF, 2 fF, 5 fF and 10 fF).

Figure 4.6: Delay response (simulated and modeled) for inverters (minimum dimensions) in different modes (SG,IG and high- V_{th}), varying the input slope (τ_{in}). Thick lines represent the slow input response.



As shown in Fig. 4.6, the model shows a high accuracy, and a maximum error percentage of 7% for fast and slow input ramps. Also, it is possible to show that IG FinFET inverters implemented with high- V_{th} devices have the worst delay times compared to SG and IG implementation modes.

Figure 4.7: Delay response (simulated and modeled) for inverters in IG mode, varying the input slope (τ_{in}) and the N_{FIN_p}/N_{FIN_n} ratio. Thick lines represent the slow input response.

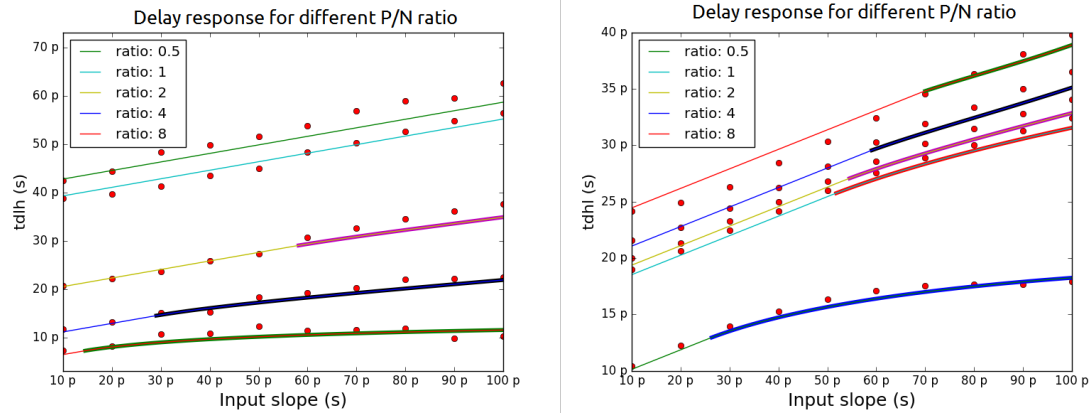
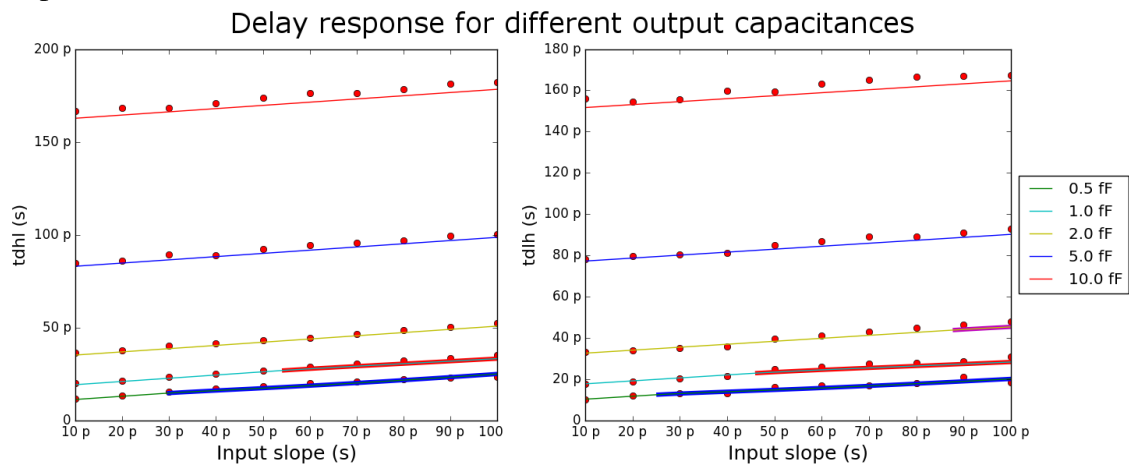


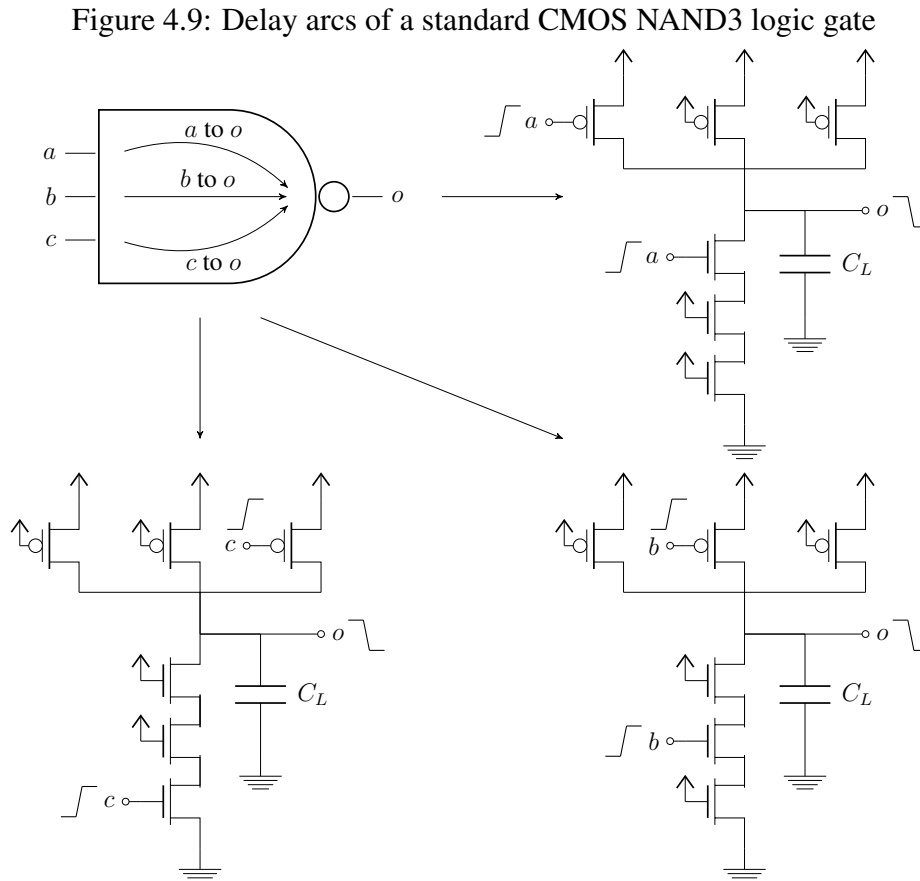
Figure 4.8: Delay response (simulated and modeled) for inverters in IG mode, varying the input slope (τ_{in}) and the load capacitance (C_L). Thick lines represent the slow input response.



4.4 Delay of series connected IG-FinFETs

A standard CMOS logic gate is composed by series and parallel arrangements of devices in the pull-up and down networks. As shown previously, a N -input CMOS logic gate could be implemented using $2N$ devices. With IG FinFETs, the required devices could be reduced by merging pairs of series/parallel transistors using low or multi- V_{th} devices. In this section, the delay analysis of series arrangements of IG FinFETs will be shown, considering the switching device position into the series arrangements.

To analyze the delay of a logic gate, each delay arc is analyzed separately. A logic arc is a path between each input and the output. Fig. 4.9 shows the different delay arcs of a three-input NAND gate.



As shown in Fig. 4.9, the gate contains three arcs, each one with a different delay ($t_{d_{LH}}(i)$, $t_{d_{HL}}(i)$) depending on the position of the i -th switching transistor. Each arc could be represented as an equivalent inverter driving a load. The order of the switching element in a series array of transistors is important, because each device contains a parasitic capacitance and a channel equivalent resistance that will be charged and discharged whether the input is switching. All the charging and discharging cases would be evaluated as follows:

first, the top switching transistor (close to the load) will be analyzed and then the bottom transistor (close to the source) for fast and slow input slopes. Then, the analysis will be extended to series and parallel IG FinFETs.

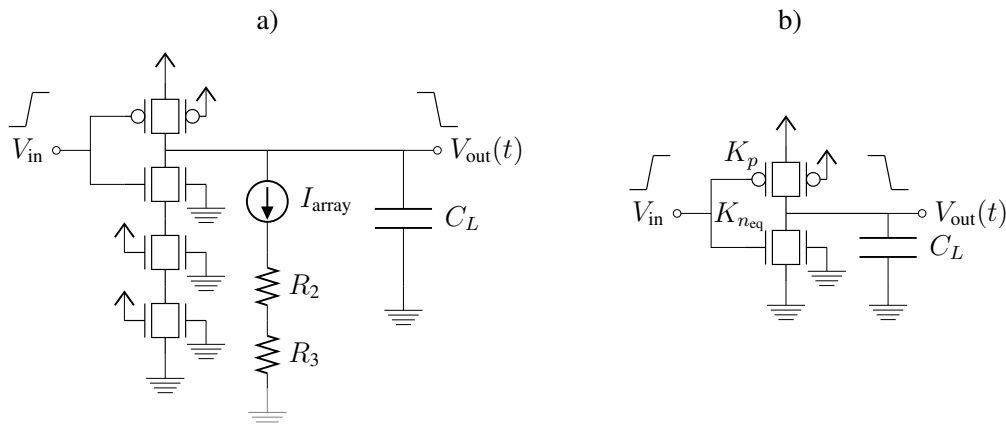
The switching device order analysis will be performed based on the information given by (DAGA; AUVERGNE, 1999) delay model and is extended for IG FinFET devices operating in different modes (SG-IG using multi- V_{th} devices). Then, HSPICE simulations will be performed in order to test the accuracy of the developed model.

4.4.1 Switching device order

4.4.1.1 Top switching transistor

As seen previously, the output response depends on the switching device position. When the controlling input is at top position, the drain voltage of the top transistor at the start of the transition is V_{DD} . When the input voltage is higher than V_{th} , the top transistor enters to saturation region and the other transistors work in linear region with an effective resistance (R_{eff_i}), as seen in Fig. 4.10.

Figure 4.10: a) NAND3 logic gate using IG-LP mode, and b) the equivalent inverter formed by the equivalent transistor with $K_{n_{sat}}$



If n transistors are connected in series and are connected in IG mode, the array current is given by the following expression:

$$I_{array} = K_{sat} H_{FIN} N_{FINn} \left(\frac{V_{DD}}{\tau_{in}} t - I_{array} \sum_{i=2}^n R_{eff_i} - V_{th}(V_{G2S}) \right)^\alpha (1 + \lambda_0 V_{DS_{avg}}) \quad (4.41)$$

For SG mode, the r factor is 0, because both gates are connected to the same poten-

tial, as explained previously. The equation above could be solved numerically, however, if $\alpha = 1$ for simplicity, the current could be expressed as: (DAGA; AUVERGNE, 1999)

$$I_{\text{array}} = \frac{\left(\frac{V_{\text{DD}}}{\tau_{\text{in}}}t - V_{\text{th0}}\right)}{\frac{1}{K_{\text{max}}} + \sum_{i=2}^n R_{\text{eff}i}} = K_{\text{eq}} \left(\frac{V_{\text{DD}}}{\tau_{\text{in}}}t - V_{\text{th}}(V_{\text{G2S}})\right) \quad (4.42)$$

Where $K_{\text{max}} = K_{\text{sat}}H_{\text{FIN}}N_{\text{FIN}n}(1 + \lambda_0V_{\text{DSavg}})$ and K_{eq} is the new saturation current coefficient, which is inversely proportional to the subsequent number of series transistors. If they are equal, the saturation current coefficient will be expressed as follows:

$$K_{n_{\text{eq}}} = \frac{K_{\text{max}}}{1 + K_{\text{max}}R_{\text{eff}}(n - 1)} \quad (4.43)$$

With the factor $K_{n_{\text{eq}}}$, the delay analysis could be performed for series IG-FinFET devices, using the parameters from Table 4.1.

The equivalent resistance for the $n - 1$ subsequent series transistors (R_{eff}) could be obtained using Eq. 4.2 for the linear region:

$$R_{\text{eff}} = \frac{1}{K_{\text{lin}}N_{\text{FIN}}H_{\text{FIN}}(V_{\text{DD}} - V_{\text{th0}})^{\frac{\alpha}{2}}} \quad (4.44)$$

Finally, to calculate the delay response, the $K_{n_{\text{eq}}}$ factor is used instead K_{sat} to calculate the maximum current I_{max} , seen previously. With the maximum current calculated Eq. 4.41, the calculation of the delay time response is performed as shown previously.

Using the top transistor as the switching device, some tests were performed for 2, 4 and 6-input NAND gates. Fig. 4.11 shows an example of the tested topologies for a four-input NAND gate. Results are shown in Table 4.3 for stacks.

Figure 4.11: Tested topologies for delay estimation for a NAND4

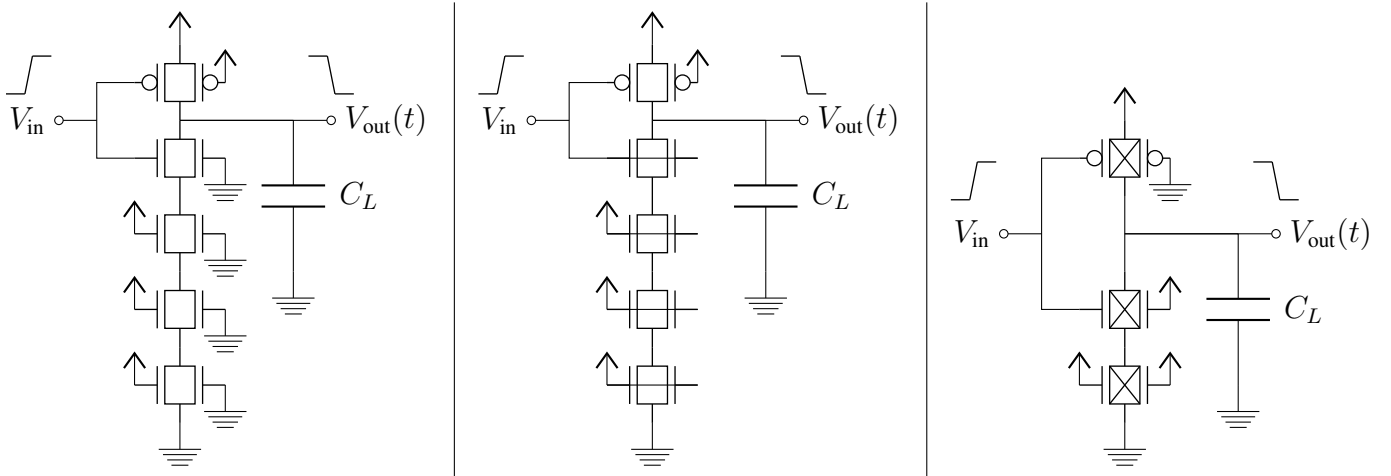


Table 4.3: Comparison between simulation and delay model for series transistors
Top switching transistor

# Series	Model	Simulation	Error	Model	Simulation	Error	Model	Simulation	Error
	(ps)	(ps)	(%)	(ps)	(ps)	(%)	(ps)	(ps)	(%)
	IG Mode			SG Mode			IG-HVT mode		
2	31.14	31.36	0.7	21.52	21.42	0.5	89.14	88.29	1.0
3	44.15	44.32	0.4	31.29	30.27	3.4	132.9	130.7	1.7
4	57	56.97	0.1	41.12	38.53	6.7	176.73	173.3	2.0

The inputs for an independent-gate device could be interchanged, thus, the delay response obtained for an input connected to the first gate is the same delay obtained for an input connected to the second gate because they have the same parasitic capacitances and gate materials, also the V_{th} from the back interface is the same. In this work, the delay response obtained by an input connected to the front-gate is considered as the same as the obtained by an input connected to the back-gate.

As seen in Table. 4.3, the response of high- V_{th} devices is slow due to their low current capability. Logic gates built using those devices should be used outside the critical path.

4.4.1.2 Bottom switching transistor

When the switching input is connected to the last transistor of the series array (transistor n), the output response is slower from the obtained from the previous case. This is due the parasitic capacitance of of the further transistors are previously charged and will be discharged together with the load. Here, the main assumption is that the load equivalent capacitance is higher than internal node parasitic capacitances of the switching transistors (C_p). For a fast input slope, the bottom transistor is initially saturated (because the drain voltage is $V_{DD} - V_{th0}$ and the gate voltage is above V_{th}), setting the current to the other transistors which are in the linear region. When the input reaches V_{DD} , the top transistor becomes saturated and the bottom transistor becomes to be in the linear region and the current will be imposed by the top transistor. The delay response could be approximated to the following expression (DAGA; AUVERGNE, 1999):

$$t_{dHL}(n) = t_{dHL}(1) + \Delta t_{\text{discharge}} \quad (4.45)$$

Where $\Delta t_{\text{discharge}}$ is the additional time interval caused by the discharge of the parasitic resistances of the previous transistors (C_{p_i}). (DAGA; AUVERGNE, 1999), defined the $\Delta t_{\text{discharge}}$ as the time required to evacuate the parasitic charge (Q_{par}) using approximately the maximum current available (I_{array}) imposed by the bottom transistor when it is switching. Thus:

$$\Delta t_{\text{discharge}} = \frac{Q_{\text{par}}}{I_{\text{rising}}} \quad (4.46)$$

For three devices connected in series, Q_{par} could be expressed as:

$$Q_{\text{par}} = C_{p_2}(V_{\text{DD}} - V_{\text{th}} - I_{\text{array}}(R_2 + R_3)) + C_{p_3}(V_{\text{DD}} - V_{\text{th}} - I_{\text{array}}(R_3)) \quad (4.47)$$

For the case of N-series connected transistors, the charge due parasitic capacitance could be expressed as:

$$Q_{\text{par}} = \sum_{i=2}^n C_{p_i} \left(V_{\text{DD}} - V_{\text{th}} - I_{\text{array}} \left(\sum_{j=i}^n R_{\text{eq}_j} \right) \right) \quad (4.48)$$

And I_{array} could be calculated using the maximum current through the device, with $V_{\text{GS}} = V_{\text{DD}}$ as:

$$I_{\text{rising}} = K_n N_{\text{FIN}} H_{\text{FIN}} (V_{\text{DD}} - V_{\text{th0}})^\alpha (1 + \lambda_0 V_{\text{DSavg}}) \quad (4.49)$$

With K_{eq} calculated as shown in Eq. 4.43. The value of C_{p_i} could be approximated as the half of the channel capacitance (C_b).

For a slow rising input slopes, the bottom transistor will be in saturation, imposing the current through the array. (DAGA; AUVERGNE, 1999) work, proposed a simpler problem to be solved: an RC equivalent circuit connected as a load of an equivalent current source formed by the bottom N-type device and its corresponding P transistor, loaded with a RC network as shown in Fig. 4.12. The delay response (for a rising input) could be expressed as:

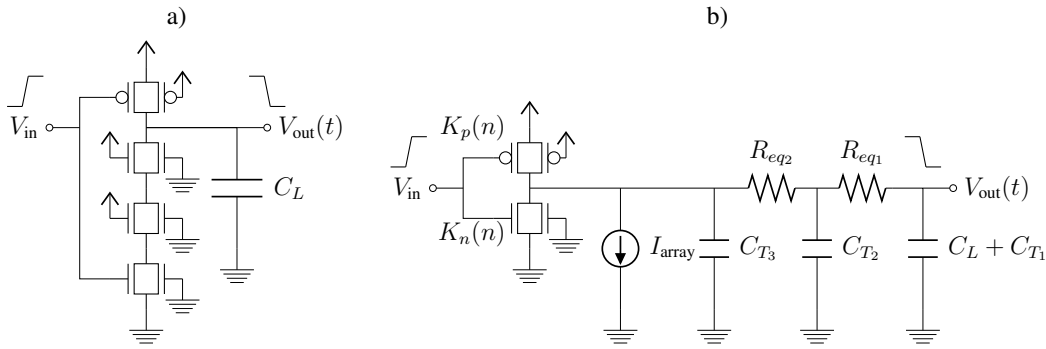
$$t_{d_{\text{HLslow}}}(n) = t_{d_{\text{HLslow}}}(1) \left(C_L + \sum_{i=1}^n C_{T(i)} \right) + \Delta t_{\text{propagation}} \quad (4.50)$$

Where $C_{T(i)} = C_{g(i)} + C_{p(i)}$ and $\Delta t_{\text{propagation}}$ is defined as (for equally sized tran-

sistors) (DAGA; AUVERGNE, 1999):

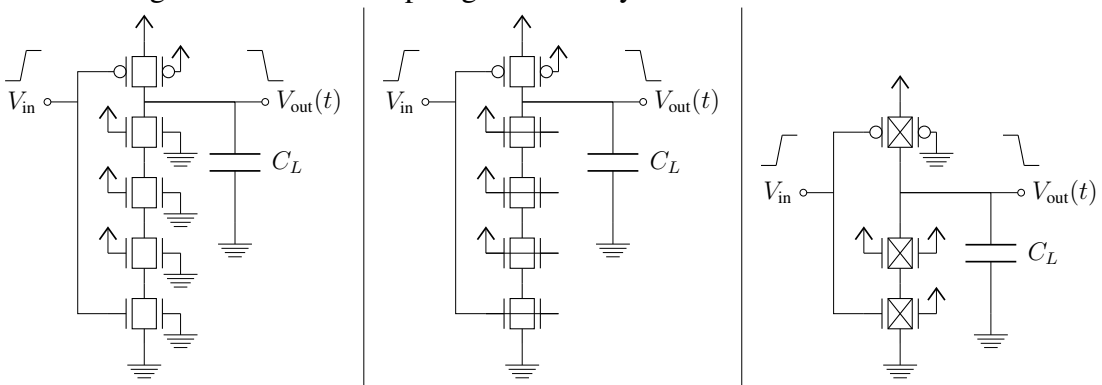
$$\Delta t_{\text{propagation}} = \frac{RC_p}{nC_p + C_L} \left(n(2n^2 + 1) \frac{C_p}{12} + n^2 \frac{C_L}{2} \right) \quad (4.51)$$

Figure 4.12: a) NAND3 logic gate using IG-LP mode, and b) the equivalent inverter formed by the equivalent transistor with $K_{n_{\text{sat}}}$.



Using the information of Eqs. 4.45 to 4.50, it is possible to evaluate the delay for a series array in which the bottom series transistor is switching. A python script was developed to calculate the delay response of n series IG FinFETs, the physical information of the devices and the parasitic capacitances (C_{in} and C_p). It was tested with fast input transitions and 1 fF of load capacitance. Fig. 4.13 shows an example of the tested topologies using a NAND 4 logic gate and Table 4.4 shows the results for the delay calculation using stacks of 2, 3 and 4 series transistors and the accuracy of the delay calculation using HSPICE simulation data.

Figure 4.13: Tested topologies for delay estimation for a NAND4



Using the data of Table 4.4, some information could be extracted to analyze the delay response of the different CMOS topologies implemented with IG FinFETs. For the case of four series transistors, using FinFETs in IG mode (Fig. 4.13 a)), the response is 17.39 ps slower compared with a stack of two devices in SG mode (Fig. 4.13 b)). If High- V_{th} IG-FinFET devices are used, only two series transistors are required (as seen in

Table 4.4: Comparison between simulation and delay model for series transistors, using a 1 fF load and a 10 ps input ramp

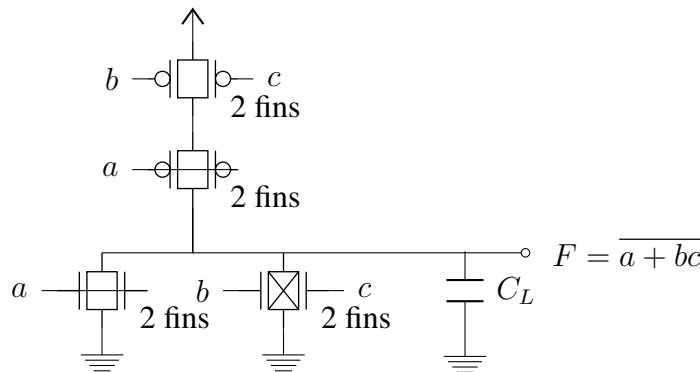
# Series	Bottom switching transistor								
	Model (ps)	Simulation (ps)	Error (%)	Model (ps)	Simulation (ps)	Error (%)	Model (ps)	Simulation (ps)	Error (%)
	IG Mode			SG Mode			IG-HVT Mode		
2	31.62	32.15	1.6	21.86	21.78	0.4	89.49	89.1	0.4
3	46.42	46.29	0.3	32.89	31.53	4.3	134.53	133	1.2
4	62.21	61.32	1.5	44.82	41.63	7.7	180.35	177.9	1.4

Fig. 4.13 c)) and the delay increases approximately four times compared to FinFETs in SG mode (a 135.53 ps of penalty).

4.4.1.3 Delay analysis of complex logic gates

Using the information given above, the delay of a N input logic gate with different series/parallel arrays of transistors could be analyzed. As an example, a logic gate which implements the function $F = \overline{a + bc}$ will be analyzed. Its schematic is shown in Fig. 4.14.

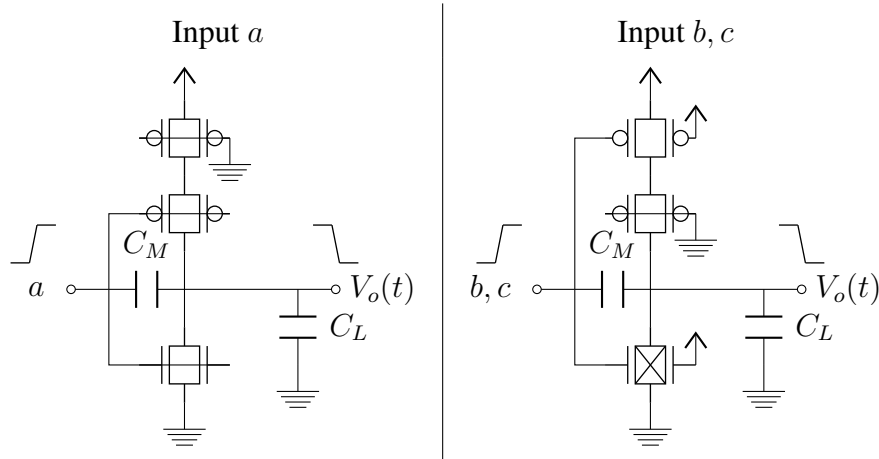
Figure 4.14: IG FinFET implementation for the function $F = \overline{a + bc}$



The analyzed logic gate contains four transistors, whose gates are connected to inputs a, b, c . The input a is connected as SG, while the inputs b and c are connected to a high- V_{th} device.

To analyze the delay response of each input, equivalent inverters will be created and the analysis shown previously for series transistors will be used in this example. Fig. 4.15 shows the different inverter configurations derived from the logic gate shown above.

Figure 4.15: Equivalent inverters for each input



The delay response ($t_{d_{HL}}$ and $t_{d_{LH}}$) for the inputs a , b and c was calculated using a 1 fF load. Table 4.5 shows the delay response of the equivalent inverters.

Table 4.5: Comparison between simulation and delay model for a logic gate, using a 1 fF load and a 10 ps input ramp

Input	High-to-low delay ($t_{d_{HL}}$)			Low-to-high delay ($t_{d_{LH}}$)		
	Sim. (ps)	Calc. (ps)	Error (%)	Sim. (ps)	Calc. (ps)	Error (%)
a	7.60	7.32	-3.73	19.80	20.25	2.27
b	24.75	25.59	3.79	27.14	26.65	-1.81
c	22.78	25.59	12.34	42.87	26.65	-37.84

As seen in Table 4.5, the simulation results for $t_{d_{LH}}$ and $t_{d_{HL}}$ for the inputs b and c are different, but the model considers the same response. This is due to lack of accuracy of the model to simulate the potential of the back-gate interface. The model can emulate the behavior of a dual-gate device, however, for simplicity, the back-gate potential analysis used to solve the Poisson's equation loses accuracy for the back-gate interface, causing loss of accuracy for the charge behavior of the back interface. As future work, some other compact models which are capable to simulate the back-interface for all the operating regions should be used. To model the behavior of IG FinFETs, and adaptation of the compact model Leti-UTSOI2.1 could be proposed as an alternative to correctly model the back-gate potential (POIROUX et al., 2015).

As seen previously, the current for a FinFET is proportional to the number of fins of the device (N_{FIN}). This is a problem for device sizing because the current values are discrete, and it is difficult to set discrete N_{FIN} values to get the desired rise and fall transition values. The use of IG-FinFET devices also could be useful to set a specific delay by setting a voltage in the back gate. Using the approximation of Eq. 4.3, the threshold voltage could be adjusted and hence, the saturation current (Eq. 4.2) could be increased

in order to decrease the delay response of a logic gate. The mentioned operation mode was used by (BHOJ; JHA, 2014) and (ALIOTO, 2011) to develop different combinational and sequential logic gates using back-biased devices.

4.5 Transition time modeling

To analyze the delay response of a logic data path, the delay information is not enough. The output transition time is also required because the output of a logic gate will be connected to other logic gates. To perform a complete modeling of a logic gate behavior it is necessary to evaluate the output transition time. It could be defined as a function of the output capacitance and the input transition time (MAURINE et al., 2002).

4.5.1 Transition time for an inverter

As a first order approximation, the virtual step response of a CMOS inverter could be used to define the transition time. When the input is a virtual step power source. the transition time could be approximated as the amount of charge which will be transferred to the load capacitance (Q_{C_L}) divided by the maximum current which will pass through the NMOS device (MAURINE et al., 2002), thus:

$$\tau_{\text{out}} = \frac{Q_{C_L}}{I_{\text{max}}} = \frac{C_L V_{\text{DD}}}{I_{\text{max}}} \quad (4.52)$$

For fast input transitions, the approximation of Eq. 4.52 could be used. However, the value of the maximum current diminishes when the input ramp duration is higher, causing an increasement of the output transition time. (ROSSELLÓ; SEGURA, 2004) and (MARRANGHELLO; REIS; RIBAS, 2015) proposed empirical equations which considered a percentage of the output voltage derivative at $V_{\text{DD}}/2$:

$$\tau_{\text{out}} = \frac{V_{\text{DD}}}{\alpha \left. \frac{dV_o(t)}{dt} \right|_{V_{\text{DD}}/2}} \quad (4.53)$$

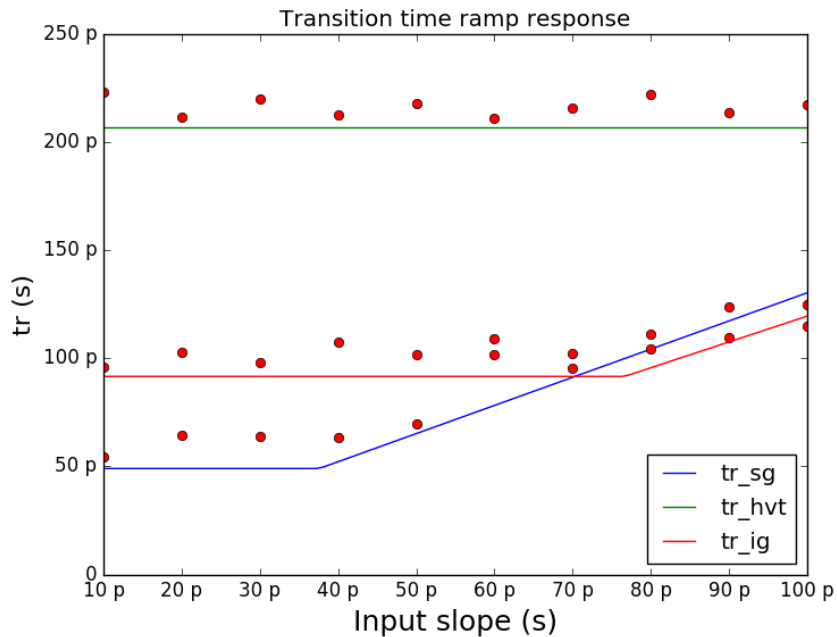
An empirical expression which uses the two approaches mentioned above will be used to find the output transition time of CMOS logic gates. As seen in Eq. 4.52, it depends on the maximum current flowing through the corresponding switching element I_{max} , the load capacitance and empirical constants which defines the maximum percentage

of the output voltage derivative at $V_{DD}/2$. As expressed by (MAURINE et al., 2002), the maximum output current decreases as the input slope increases, increasing the output transition time. Using HSPICE simulations, an empirical expression defined based on (MAURINE et al., 2002) and (ROSSELLÓ; SEGURA, 2004) approaches were used to define the transition time, and is defined as:

$$\max \left\{ \frac{V_{DD}C_L}{\alpha I_{\max}}, \beta \frac{V_{DD}C_L}{\alpha I_{\max}} \left(\frac{\tau_{in}}{t_{ref}} \right) \right\} \quad (4.54)$$

Where α and β are calibrated using HSPICE simulations ($\alpha = 0.65$ and $\beta = 1.2$). Figs. 4.16 and 4.17 show the results of transition times of different inverter configuration using IG FinFETs (SG, IG and inverters with high- V_{th}) varying the input ramp from 10 to 100ps and $C_L = 1$ fF. For simulations, the transition time was taken as the 70% of the derivative of the voltage at $V_{DD}/2$.

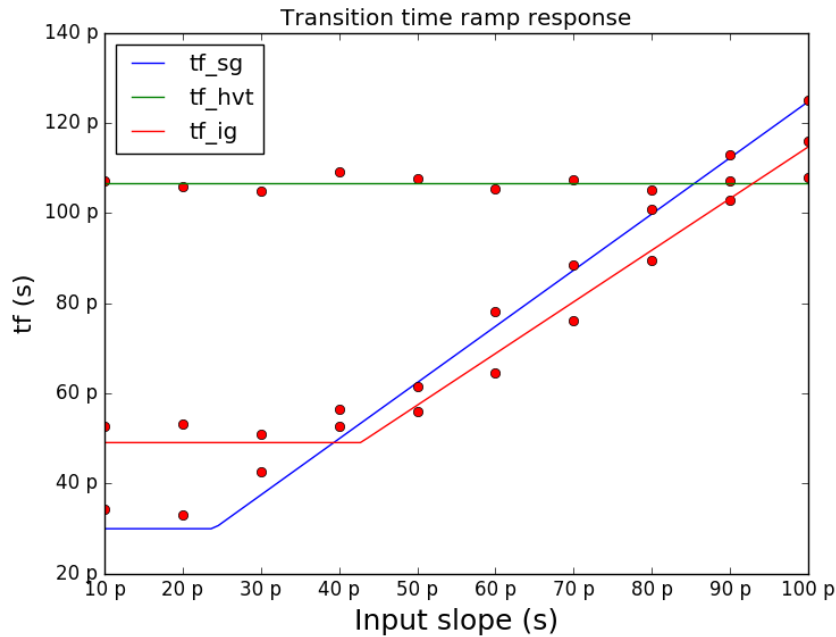
Figure 4.16: Rise time response for different inverters built using IG FinFETs, varying the input ramp slope (τ_{in}) and $N_{FINp}/N_{FINn} = 1$ ratio.



4.6 Summary

In this chapter, the static and dynamic behavior of CMOS inverters was introduced, showing the different stages of the transient output response of standard inverter structures. Then, the physical characteristics and features of the IG FinFET devices shown

Figure 4.17: Fall time response for different inverters built using IG FinFETs, varying the input ramp slope (τ_{in}) and $N_{FIN_p}/N_{FIN_n} = 1$ ratio.



in the previous chapters were used to analyze the delay of static CMOS inverters implemented using multi- V_{th} IG FinFETs, extending the analysis of the charge-based delay models proposed by (ROSSELLÓ; SEGURA, 2004) and (MARRANGHELLO; REIS; RIBAS, 2015) for those devices and showing the accuracy of the delay model. Then, the delay analysis used for inverters was extended for transistor stacks based on the analytical model proposed by (DAGA; AUVERGNE, 1999), and more complex topologies derived from the use of independent gate multi- V_{th} devices was shown. The delay model shows a good accuracy (a maximum error of 8%) and shows the delay penalty derived from the use of high- V_{th} devices to perform compactations of serial arrays of transistors. Also, an empirical expression for the transition time of inverters was proposed, showing its accuracy with HSPICE simulations. The analytical model shown in this chapter could be useful to design logic gates using IG FinFET devices. The next chapter will apply the delay analysis shown in this chapter to analyze the critical path delay response of a parallel-prefix adder circuit in order to see the utility of the delay model in the analysis of data paths.

5 DELAY ANALYSIS OF IG FINFETS IN LOGIC DATA PATHS

5.1 Introduction

Timing analysis in logic data paths has been an useful tool for evaluating the performance of a digital circuit and to check if timing requirements are met. In VLSI design, Static Timing Analysis (STA) and Statistical Static Timing Analysis (SSTA) tools play an important role in determining if a circuit meets the required timing constraints for the different operating points (Process variations, voltages and temperatures).

The information used by timing analysis tools is obtained through electrical simulations of each gate of the standard cell library, which is characterized using different input ramps and fan-out loads. The results of those set of simulations are indexed and saved in data files containing all the tested transitions, delays and power consumption. Restricting the analysis to a set of operating conditions given in the results files (FORZAN; PANDINI, 2007). When input transition or load capacitance values are outside the range of the tested values, the timing analysis loses accuracy, causing some errors in the timing analysis process.

The proposed approach in this chapter is to use the delay model presented in the previous chapter and to perform the delay analysis of a logic data path, in order to show an utility of the delay analysis. The proposed logic to be tested is an 8-bit Parallel Prefix Adder (PPA) which each building block is implemented and simulated using IG FinFETs and compacted using multi- V_{th} devices.

The rest of this chapter is divided as follows: initially, the Parallel Prefix Adder (PPA) architecture is explained, also the delay model is extended to perform calculations of transition times. Then, the analysis of the critical path is performed, extracting the critical path logic circuit and describing the used methodology for calculating its delay, showing the results and their accuracy. Finally, some concluding remarks are given.

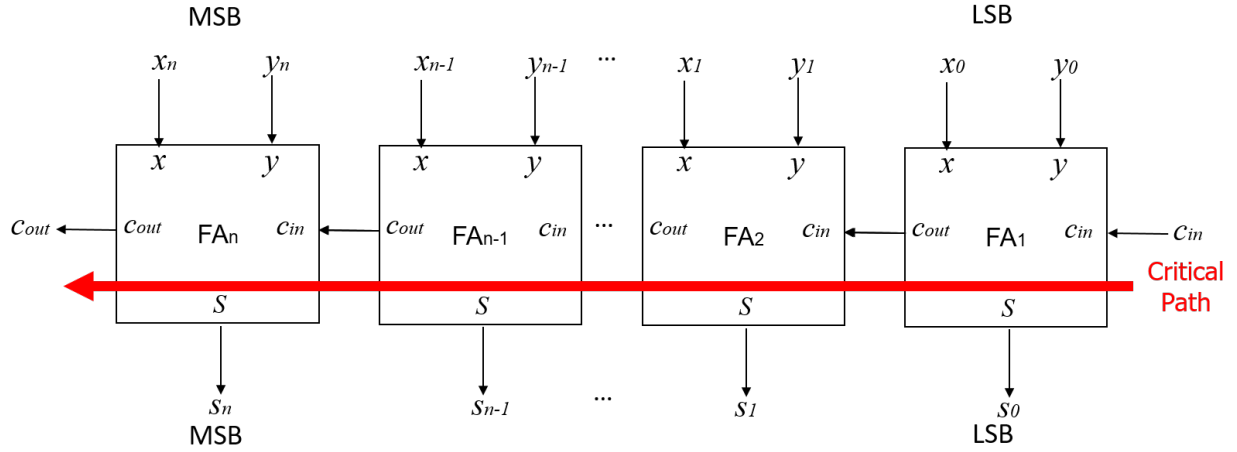
5.2 Parallel Prefix Adder (PPA)

Adder circuits are an important part of the modern digital systems, as processor data path. They compute two binary numbers and determine if a carry bit of that sum is generated.

One of the most adopted adder circuits is the ripple carry adder mainly due to be

the simplest one. It is composed by N full-adders connected in cascade (where N is the number of bits of each operand). Fig. 5.1 shows a block diagram of this topology.

Figure 5.1: Block diagram of a standard Ripple Carry Adder architecture



The full-adder (FA) block receives three different signals (x , y and C_{in}) and performs two basic operations: the first operation is the sum of the three bits (represented by the expression $S = x \oplus y \oplus C_{in}$), and the second is the carry out calculation (represented by the expression: $C_{out} = x \cdot y + x \cdot C_{in} + y \cdot C_{in}$).

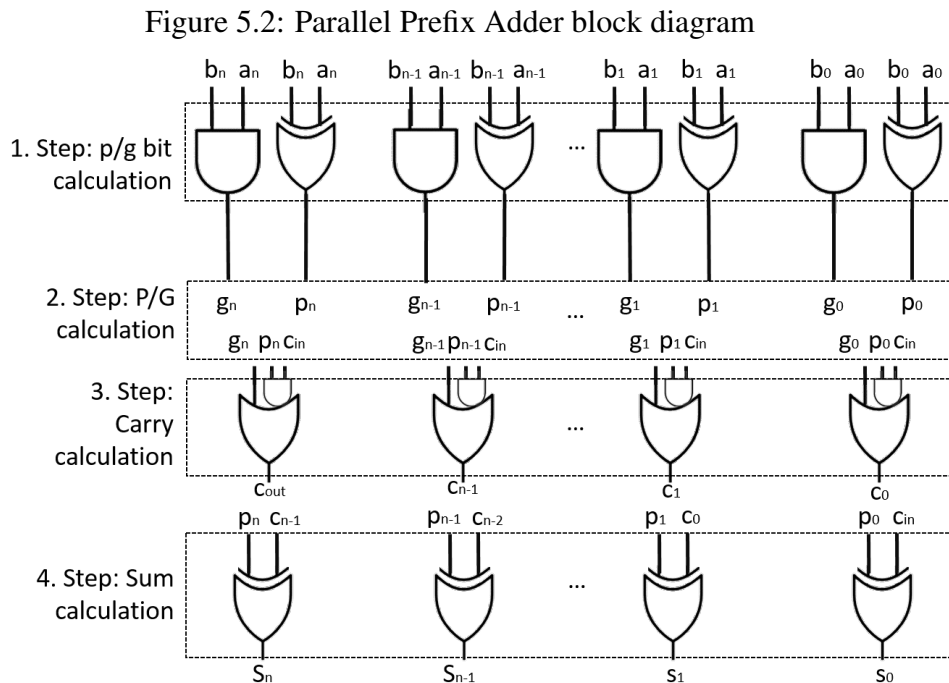
The FA circuit is implemented using 28 transistors in standard CMOS technology, as seen in Fig. 3.13 a). It is possible to reduce the number of transistors by using IG transistors and merging series and parallel switches as seen previously. Fig. 3.13 b) shows a reduced version of the FA circuit built using IG FinFET devices.

To compute the carry out signal, the circuit needs to perform all the calculations of the carry bit beginning from the least significant bit (LSB). If the adder has a high number of stages, the sum and carry out calculations become slower. That is one of the disadvantages of this topology.

Parallel-Prefix Adders are digital circuits that implement the binary addition using a more complex structure to perform the carry computation. This type of adders is more efficient than the ripple-carry adder approach in terms of performance, decreasing considerably the latency due to the parallelism of the associated algorithm.

This adder architecture calculates the sum using four steps. The first step is the calculation of propagate and generate signals of each bit of the operands. The second step is the grouping of that signals, which involves the calculation of group propagate and group generate signals. The third step calculates the internal carry bits. The last step calculates the result, using the carry bits and the propagated signals computed in the first

step. Fig. 5.2 shows a block diagram of standard implementation of the parallel-prefix adder algorithm.



The first step of the parallel adders is to compute the individual propagate and generate signals, performed by the following Boolean expressions:

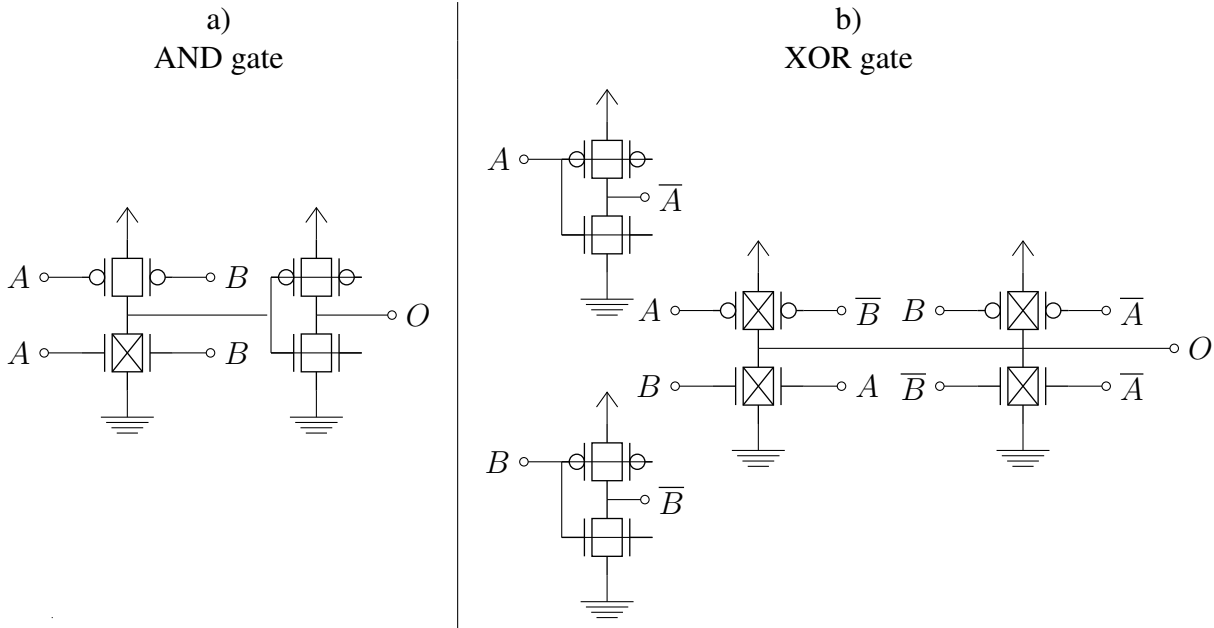
$$\begin{aligned} p_i &= a_i \oplus b_i \\ g_i &= a_i \cdot b_i \end{aligned} \quad (5.1)$$

Fig. 5.3 shows the standard CMOS topologies of the AND and exclusive-OR (XOR) gates used to implement those logic expressions and the reduced version using IG FinFETs.

The second step is the propagate/generate (P/G) association operations, which are performed using the prefix operators which uses the individual propagate and generate signals to calculate new pairs of P/G as well as the final group propagate and group generate signals. The most known association algorithms are Ladner-Fischer, Kogge-Stone and Brent-Kung topologies which has different delay, fanout and area characteristics.

The prefix operator takes two pairs of propagate and generate signals (p_i, g_i) , (p_j, g_j) calculated using Eq. 5.1 and compute a new pair of propagate and generate bits

Figure 5.3: Propagate and generate signal implementation using IG-FinFETs



by performing the following operations:

$$\begin{aligned}
 p_o &= p_i \oplus p_j \\
 g_o &= g_i + g_j \cdot p_i
 \end{aligned}
 \tag{5.2}$$

Fig. 5.4 shows a prefix operator block diagram. The operations performed by the prefix operator (p_o, g_o) , can be implemented using IG FinFETs, so reducing the number of transistors used in the conventional static CMOS design. Fig. 5.5 shows and circuits implemented using IG devices.

Figure 5.4: Prefix operator used by PPA adder to calculate the group propagate and group generate signals

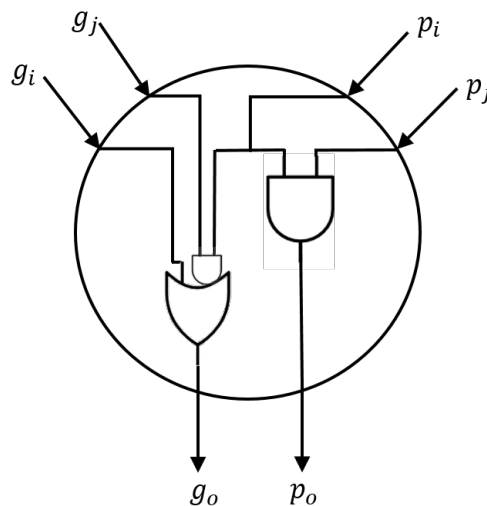
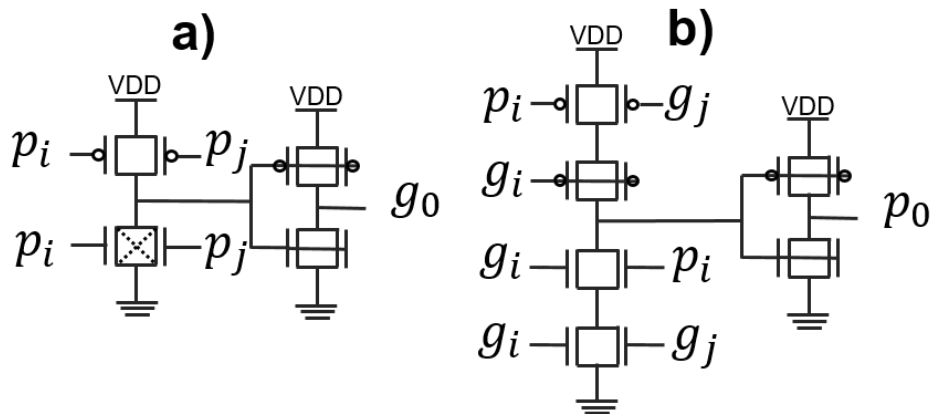
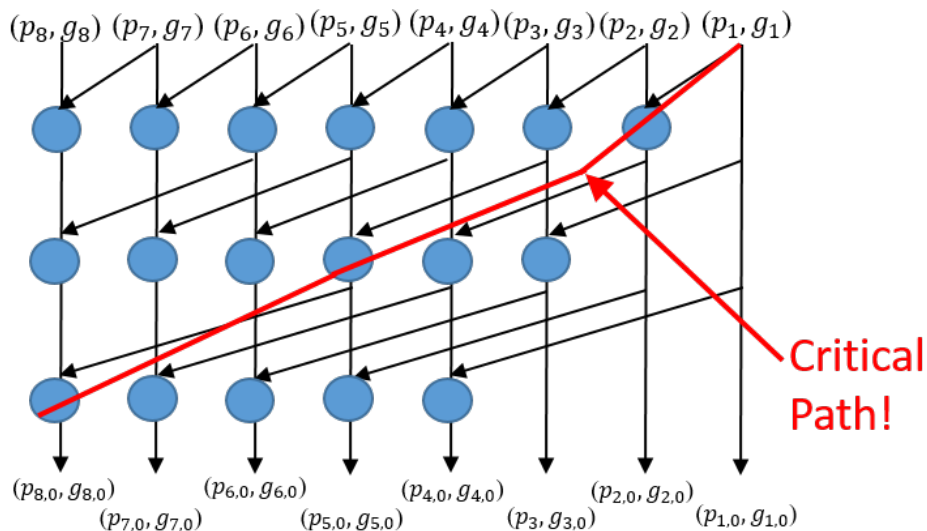


Figure 5.5: Prefix operator circuit implementation using IG FinFET: AND gate built using 4 devices, and AND-OR gate built using 6 devices



One of the most common topologies for a prefix tree is the one proposed by Kogge-Stone (NEHRU; SHANMUGAM; VADIVEL, 2012), which uses a larger quantity of operators (compared with other topologies). This topology is faster than the ripple-carry adder, but the power consumption is higher because it uses more devices. Fig. 5.6 shows an example of this prefix tree for 8-bit adder. In that, for example, the group generated and group propagated signals $p_{2,0}$, $g_{2,0}$ are calculated using the (p_2, g_2) and the (p_0, g_0) groups. The other propagate and generate signal groups are calculated in a similar way. The main idea is to compute $p_{j,0}, g_{j,0}$ where $1 \leq j \leq n$, being n the bit width of the operands.

Figure 5.6: Kogge-Stone prefix tree implementation and the critical path indication.



The next step is to calculate the carry bit for each of the bits of the sum (or C_i). The logic expression that represents the C_i value is $C_i = g_{i,0} + p_{i,0} \cdot C_{i-1}$ and can be implemented using an AND-OR gate. Fig. 5.5 shows an AND-OR gate implementation

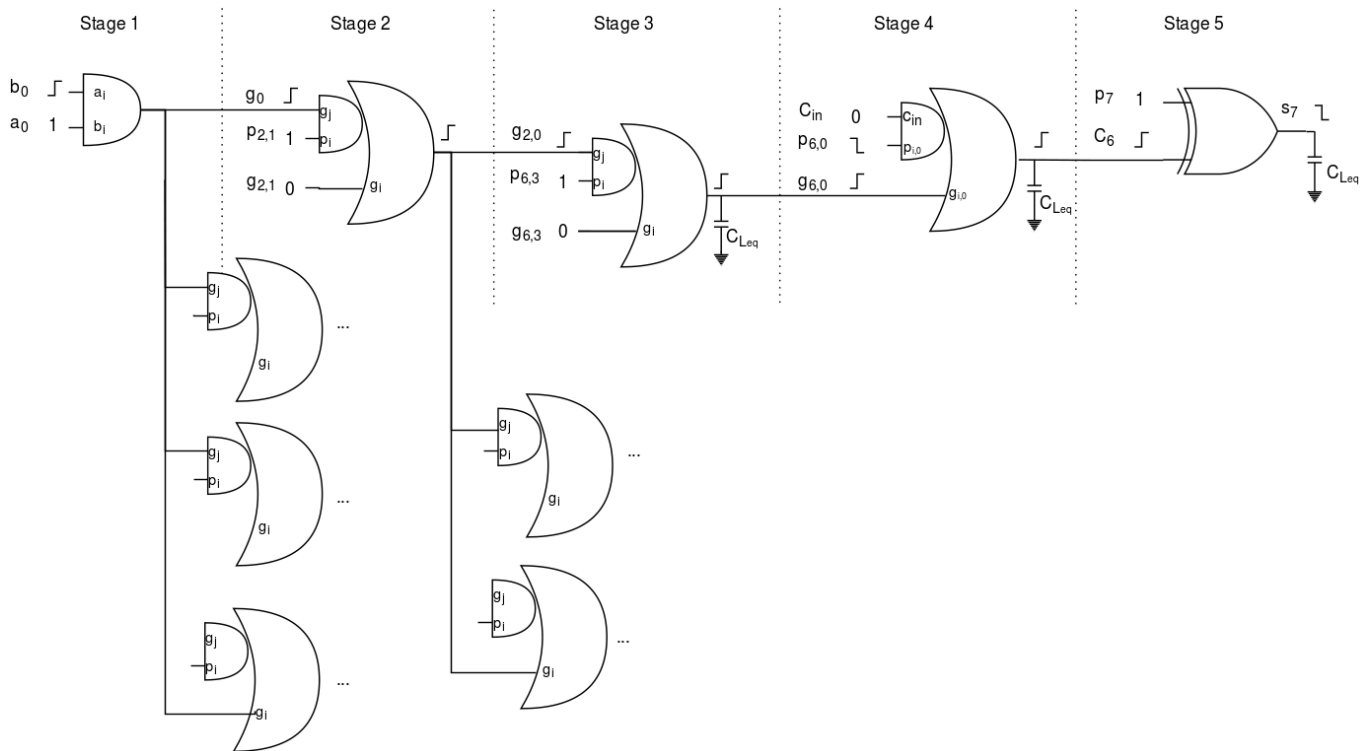
using IG FinFET. The last step of the parallel-prefix adder is the sum calculation. It is a simple XOR operation which uses the propagate signal of each bit (or $a_i \oplus b_i$), and the carry of the bit $i - 1$, so the i^{th} bit of the sum operation is $S_i = p_i \oplus C_{i-1}$. The XOR operation can be built using only 4 devices and 2 inverters, as shown in Fig. 5.3 b).

Notice that the critical path is different from the RCA, in where the critical path is from the carry in to the carry out signals, whereas in PPA the carry out calculation has shorter logic depth. In PPAs, the critical path goes from the least to the most significant sum bits, as illustrated in Fig. 5.6.

5.3 Critical path delay analysis

To estimate the critical path delay, the worst delay case is extracted from the previous information, in which is shown that the worst case delay is caused mainly by the prefix tree depth and the fan-out of each prefix operator. For the case of the Kogge-Stone adder shown in Fig. 5.6, each prefix operator nodes have a low fan-out and is lower than other prefix tree architectures such as Ladner-Fischer (which has less nodes with higher fan-out). For this chapter, an 8-bit Kogge-Stone architecture will be analyzed. The tested operation will be FF_{16} for the first operator (*a*) and a rising transition of the LSB of the second operator (*b*). The carry-in bit stays low during the operation. Fig. 5.7 shows the critical path which its delay analysis will be performed.

Figure 5.7: 8-bit Kogge-Stone adder critical path circuit



The circuit shown in Fig. 5.7 will be analyzed in different stages, in which each one the delay and transition time is calculated. Also, a comparison with simulation results will be performed in order to show the accuracy of the analyzed model.

One important variable which needs to be considered is the transition time of each stage ($t_t(i)$). Using the information of the previous chapter, the transition time of the stage i is a function of the equivalent load capacitance of each logic gate ($C_{Leq}(i)$), the input transition time (τ_{in}), which is the output transition time of the previous logic gate and the maximum current flowing through the transistor array which is being analyzed (I_{max}).

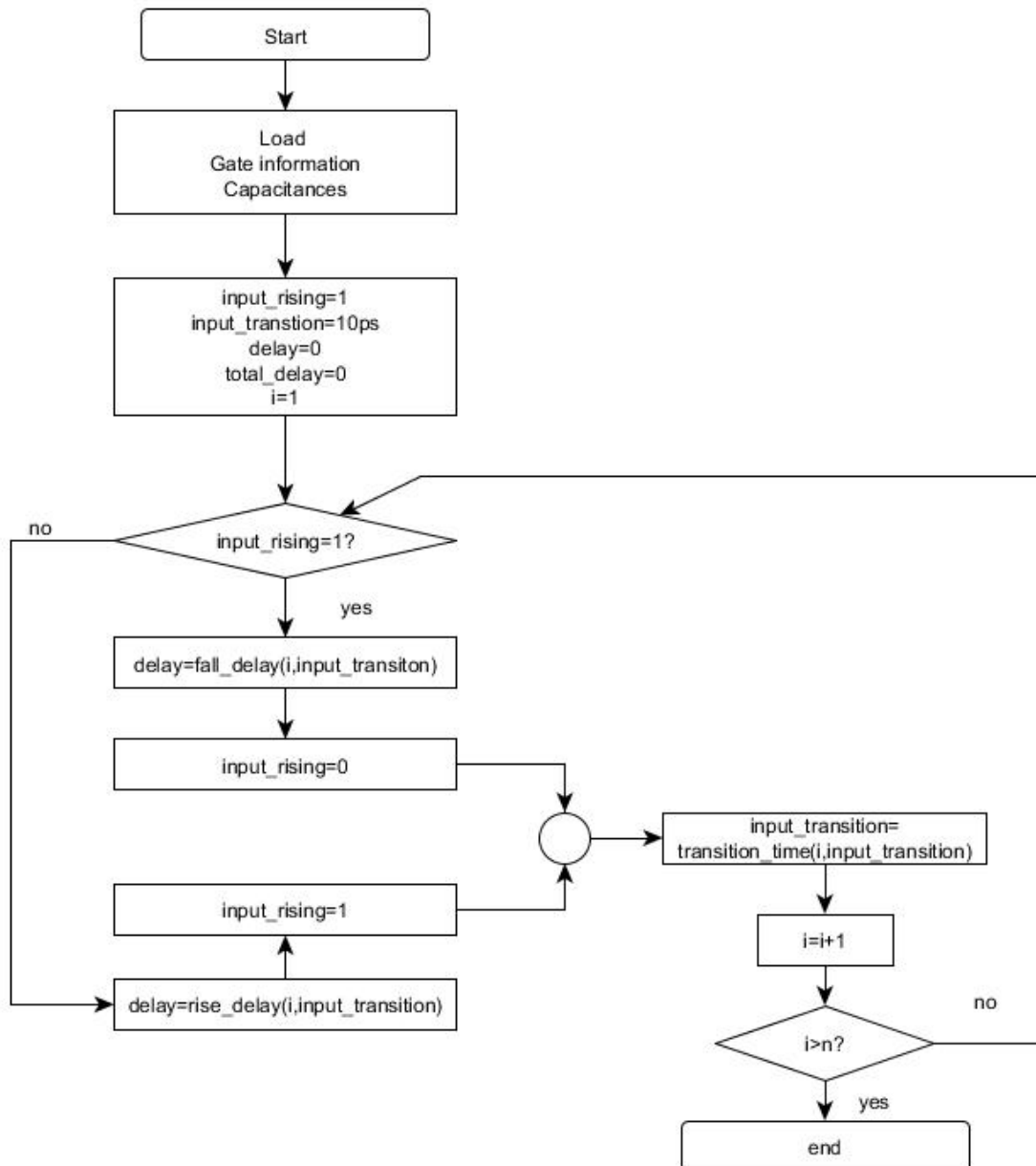
To analyze the delay of the circuit shown in 5.7, it will be divided in five stages, in which each one will be analyzed independently. Then, the total delay will be the sum of each delay response. Thus, the total delay time will be expressed as:

$$t_{d_{total}} = \sum_{i=1}^n t_{d_j}(i) \quad (5.3)$$

Where i represents the stage and j the corresponding input of the gate i . The input signal of the first stage is a 10 ps ramp. Notice that all the logic gates of the data path are positive unate. For the case of the AND gate (see Fig. 5.5 a)) and the AND-OR gate (see Fig. 5.5 b)), the delay of the series inverter, will affect the delay results, so they need to

be considered. For the case of the XOR gate (Fig. 5.3 b)), one of the inputs is high and the other is rising from low to high, so the delay of the input inverters could be negligible. The total delay response was calculated using the algorithm shown in Fig. 5.8.

Figure 5.8: Proposed algorithm to calculate the data path delay



The output capacitance of each stage was calculated using the information of Table 4.2 and the fan-out of each output node (based on the information of the Figs. 5.6 and 5.7. The equivalent inverters of each logic gate is shown in Figs. 5.9, 5.10 and 5.11.

Figure 5.9: Equivalent inverters for the AND2 logic gate

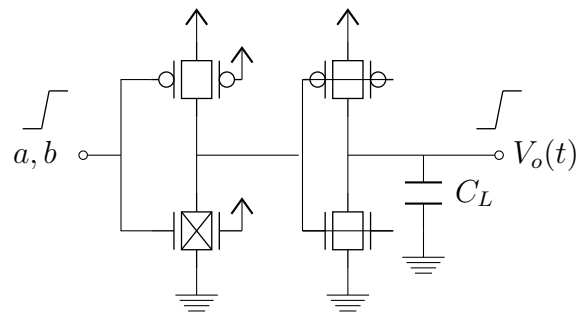


Figure 5.10: Equivalent inverters for the AND-OR logic gate

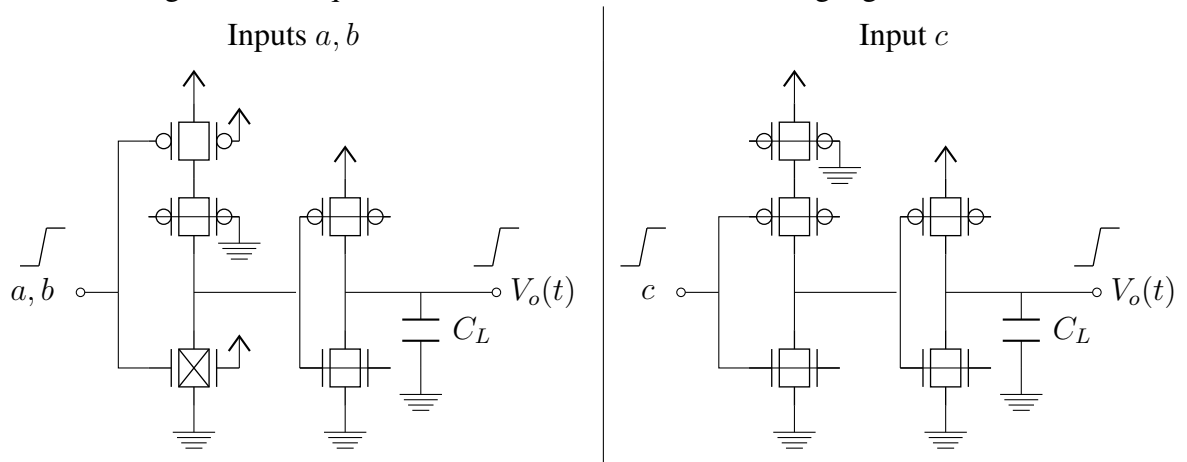
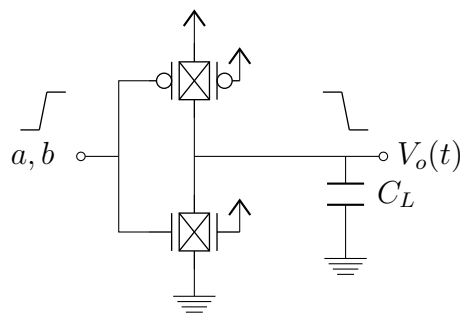


Figure 5.11: Equivalent inverters for the XOR2 logic gate



5.4 Results

Using the algorithm shown in Fig. 5.8, an input transition of 10 ps and 1fF of load capacitance, an 8-bit KSA adder was simulated using HSPICE and its critical path delay was measured in order to compare the delay model calculations. Table 5.1 shows the results of the delay and transition times (measured as 70% of the output response derivative at $V_{DD}/2$) for each stage.

As seen in Table 5.1, the total delay response of the critical path has a total error

Table 5.1: Comparison between calculations and SPICE simulations for delay the critical path of the 8-bit PPA circuit

stage	transition				delay		
	sim	calc	%err	type	sim	calc	%err
1	15.99	14.58	-8.82	fall	8.30	8.52	2.64
	24.65	24.73	0.32	rise	9.45	10.23	8.29
2	21.61	18.06	-16.43	fall	13.62	12.10	-11.16
	16.81	19.92	18.50	rise	6.97	6.32	-9.35
3	19.46	14.58	-25.08	fall	11.59	10.98	-5.26
	12.39	14.53	17.27	rise	5.26	5.09	-3.16
4	10.15	13.90	36.95	fall	3.83	3.50	-8.62
	17.99	15.15	-15.79	rise	5.81	5.76	-0.91
5	110.20	119.84	8.75	fall	40.36	40.12	-0.59
TOTAL					105.19	102.62	-2.44

of (-2.44%), and an average error of (-3.13%), which shows that the delay calculation using the delay analysis shown in the previous chapter has a high accuracy for the delay calculation. For the case of transition times, some cases shown a high difference between the simulation and the estimated delay times, however, in average, the average magnitude of the error is 16.43%, considering that the input stimulus is different from a ramp and the measuring of the output transition time response was set to a fixed percentage of the output derivative at $V_{DD}/2$.

5.5 Summary

In this chapter, the delay model explored in the previous chapter was used to analyze the timing delay of a logic circuit. As an example, an 8-bit parallel-prefix adder (PPA) was analyzed, in which each its critical path was extracted and the delay of each stage was analyzed. The architecture of each logic gate was implemented using IG Fin-FET devices and showing that the number of needed devices decreases by about 30%. The delay and transition time of each logic gate was analyzed and compared with SPICE simulation results, achieving about a mean percentage error of 3% for delay and 14.55% for transition times for standard operating conditions. With the development of this chapter, an interesting utility of the analytical model explored previously for the delay calculation of complex logic circuits and could be extended for timing analysis of logic circuits connected to sequential circuits in order to perform setup/hold slack calculations without the necessity of using electrical characterization of logic gates using SPICE simulations.

6 CONCLUSIONS

In this work, four-terminal device called independent-gate FinFET was explored, showing its basic operation principle and the variation of its threshold voltage in function of the back-gate bias, which gives interesting properties for logic gate design.

Taking advantage of the IG FinFET physical behavior, the reduction of logic networks using such kind of devices was studied, showing a method for the efficient reduction the number of required transistors to implement complex logic functions called de-factorization. With the de-factorization method, The energy consumption was reduced in approximately 54% whereas the delay increased in 61%, compared to single-gate implementations, giving a good delay-power product suitable for low-power designs.

After studying the electrical behavior of IG FinFET based networks, the analytical behavior for IG FinFET logic networks was studied, using analytical models to describe the delay response of the different logic arrangements derived from the use of independent gate devices. Also, an empirical expression for the transition time response was given, showing its accuracy for different configurations of IG FinFET devices.

After the study of the electrical behavior of IG FinFET logic networks through electrical simulations, the transient behavior was also analyzed based on charge-based delay models for describing the ramp response of different IG FinFET based logic gates, achieving an improvement of the actual IG FinFET delay models (DATTA et al., 2007), (LIN; WANG; PEDRAM, 2013) and considering the transistor stacks using multi- V_{th} devices. Also, the transition time analysis was also studied and an analytical expression was proposed for the different configurations of IG-FinFET logic gates.

Finally, the analytical model developed previously was applied to a combinational logic data path, using an 8-bit PPA adder as an example, analyzing its critical path and achieving a total error of 2.44% compared to SPICE simulations.

6.1 Future work

- An improvement on the electrostatic potential behavior is required to improve the response of the I-V characteristics of the back-gate interface. For that, a study of the last version (v2.1) of the Leti-UTSOI2 core compact model is proposed, and an adaption for IG FinFET devices is suggested. The Leti-UTSOI2 model is the first compact model able to describe FDSOI transistor behavior in all bias

configurations, including strong forward back bias (POIROUX et al., 2015).

- Delay model extension for other independent dual gate devices, such as the vertical slit field-effect transistor (VesFET), which is a dual dual-gate devices with present excellent properties and could be used for logic gate device compactions (YANG et al., 2016), (WEIS; EMLING; SCHMITT-LANDSIEDEL, 2009).
- Extension of the analytical model to support lower voltages and temperature variations. All the delay analysis presented here uses nominal operating conditions ($V_{DD} = 0.9$ V and $T = 25$ °C). The analysis could be extended to different variations of the operating conditions.
- The analytical model could be extended to model the characteristics of sequential logic gates, for creating equations capable of modeling the required setup/hold times.

REFERENCES

- ABOUZEID, F. et al. 40nm cmos 0.35V-optimized standard cell libraries for ultra-low power applications. **ACM Trans. Des. Autom. Electron. Syst.**, ACM, New York, NY, USA, v. 16, n. 3, p. 35:1–35:17, jun. 2011. ISSN 1084-4309. Disponível em: <<http://doi.acm.org/10.1145/1970353.1970369>>.
- ALIOTO, M. Comparative evaluation of layout density in 3t, 4t, and MT FinFET standard cells. **Very Large Scale Integration (VLSI) Systems, IEEE Transactions on**, IEEE, v. 19, n. 5, p. 751–762, 2011.
- BALESTRA, F. et al. Double-gate silicon-on-insulator transistor with volume inversion: A new device with greatly enhanced performance. **IEEE Electron Device Lett.**, Institute of Electrical & Electronics Engineers (IEEE), v. 8, n. 9, p. 410–412, sep 1987.
- BHOJ, A. N.; JHA, N. K. Parasitics-aware design of symmetric and asymmetric gate-workfunction FinFET SRAMs. **Very Large Scale Integration (VLSI) Systems, IEEE Transactions on**, IEEE, v. 22, n. 3, p. 548–561, 2014.
- BISDOUNIS, L.; NIKOLAIDIS, S.; KOUFOPAVLOU, O. Analytical transient response and propagation delay evaluation of the cmos inverter for short-channel devices. **IEEE Journal of Solid-State Circuits**, v. 33, n. 2, p. 302–306, Feb 1998. ISSN 0018-9200.
- BISDOUNIS, L. et al. Accurate timing model for the CMOS inverter. In: **IEEE. Electronics, Circuits, and Systems, 1996. ICECS'96., Proceedings of the Third IEEE International Conference on**. Rodos, Greece, 1996. v. 1, p. 89–92.
- CHAN, M. et al. A framework for generic physics based double-gate MOSFET modeling. In: **Proc. 2003 Nanotechnology Conf.** San Francisco, CA: [s.n.], 2003. p. 270–273.
- CHAUDHRY, A.; KUMAR, M. J. Controlling short-channel effects in deep-submicron SOI MOSFETs for improved reliability: a review. **IEEE Transactions on Device and Materials Reliability**, IEEE, v. 4, n. 1, p. 99–109, 2004.
- CHIANG, M.-H. et al. High-density reduced-stack logic circuit techniques using independent-gate controlled double-gate devices. **IEEE Transactions on Electron Devices**, IEEE, v. 53, n. 9, p. 2370–2377, 2006.
- CHUANG, C.-T. et al. Scaling planar silicon devices. **IEEE Circuits and Devices Magazine**, IEEE, v. 20, n. 1, p. 6–19, 2004.
- COLINGE, J. The new generation of SOI MOSFETs. **Romanian Journal of information science and technology**, v. 11, p. 3–15, 2008.
- COLINGE, J.-P. Multiple-gate SOI MOSFETs. **Solid-State Electronics**, Elsevier BV, v. 48, n. 6, p. 897–905, jun 2004.
- COLINGE, J.-P. **FinFETs and Other Multi-Gate Transistors**. 1st. ed. [S.l.]: Springer Publishing Company, Incorporated, 2007. ISBN 038771751X, 9780387717517.
- COLINGE, J. P. et al. Silicon-on-insulator ‘gate-all-around’ MOS device. In: **SOS/SOI Technology Conference, 1990., 1990 IEEE**. Key West, FL: [s.n.], 1990. p. 137–138.

CRISTOLOVEANU, S. et al. Double-gate MOSFETs: performance and technology options. In: IEEE. **Semiconductor Device Research Symposium, 2001 International**. [S.l.], 2001. p. 459–460.

DAGA, J. M.; AUVERGNE, D. A comprehensive delay macro modeling for submicrometer CMOS logics. **IEEE Journal of Solid-State Circuits**, IEEE, v. 34, n. 1, p. 42–55, 1999.

DATTA, A. et al. Modeling and circuit synthesis for independently controlled double gate FinFET devices. **IEEE transactions on computer-aided design of integrated circuits and systems**, IEEE, v. 26, n. 11, p. 1957–1966, 2007.

DESHMUKH, R. et al. Comparing finfets: Soi vs bulk: Process variability, process cost, and device performance. In: IEEE. **Computer, Communication and Control (IC4), 2015 International Conference on**. [S.l.], 2015. p. 1–4.

DOYLE, B. et al. Transistor elements for 30nm physical gate lengths and beyond. **Intel Technology Journal**, v. 6, n. 2, 2002.

DUARTE, J. P. et al. BSIM-CMG: Standard FinFET compact model for advanced circuit design. In: IEEE. **European Solid-State Circuits Conference (ESSCIRC), ESSCIRC 2015-41st**. [S.l.], 2015. p. 196–201.

DUTTA, S.; SHETTI, S. M.; LUSKY, S. L. A comprehensive delay model for CMOS inverters. **IEEE Journal of solid-state circuits**, IEEE, v. 30, n. 8, p. 864–871, 1995.

ENEMAN, G. ESSDERC 2015 tutorial: Thin channel silicon transistors. In: **45th European Solid-State Device Research Conference**. Graz, Austria: [s.n.], 2015. (ESSDERC).

FERAIN, I. et al. Multigate transistors as the future of classical metal oxide semiconductor field-effect transistors. **Nature**, Nature Publishing Group, v. 479, n. 7373, p. 310–316, 2011.

FORZAN, C.; PANDINI, D. Why we need statistical static timing analysis. In: **Computer Design, 2007. ICCD 2007. 25th International Conference on**. [S.l.: s.n.], 2007. p. 91–96. ISSN 1063-6404.

FOSSUM, J. G.; TRIVEDI, V. P. **Fundamentals of Ultra-Thin-Body MOSFETs and FinFETs**. [S.l.]: Cambridge University Press, 2013. ISBN 9781139343466.

FRANK, D. J.; TAUR, Y.; WONG, H.-S. Generalized scale length for two-dimensional effects in MOSFETs. **IEEE Electron Device Letters**, IEEE, v. 19, n. 10, p. 385–387, 1998.

FREI, J. et al. Body effect in tri-and pi-gate soi mosfets. **IEEE electron device letters**, IEEE, v. 25, n. 12, p. 813–815, 2004.

FRIED, D. **The design, fabrication and characterization of independent-gate FinFETs**. Tese (Doutorado) — Cornell University, May, maio 2004.

GE, L.; FOSSUM; GAMIZ. Mobility enhancement via volume inversion in double-gate MOSFETs. In: **2003 IEEE International Conference on Robotics and Automation SOI-03**. [S.l.]: Institute of Electrical & Electronics Engineers (IEEE), 2003.

- GEPPERT, L. The amazing vanishing transistor act. **IEEE Spectrum**, v. 39, n. 10, p. 28–33, Oct 2002. ISSN 0018-9235.
- GILDENBLAT, G. **Compact modeling: principles, techniques and applications**. [S.l.]: Springer Science & Business Media, 2010.
- GOEL, N.; TRIPATHI, A. Performance of double gate SOI MOSFET. **International Journal of Electronics Engineering**, Serials Publications, v. 4, n. 1, p. 57–59, 2012.
- HISAMOTO, D. et al. A fully depleted lean-channel transistor (DELTA)-a novel vertical ultra thin SOI MOSFET. In: IEEE. **Electron Devices Meeting, 1989. IEDM'89. Technical Digest., International**. [S.l.], 1989. p. 833–836.
- KAGARIS, D. MOTO-X: A multiple-output transistor-level synthesis cad tool. **Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on**, IEEE, v. 35, n. 1, p. 114–127, 2016.
- KERBER, P. et al. GIDL in doped and undoped FinFET devices for low-leakage applications. **IEEE Electron Device Letters**, IEEE, v. 34, n. 1, p. 6–8, 2013.
- KHANDELWAL, S. et al. Bsim-img: A compact model for ultrathin-body soi MOSFETs with back-gate control. **IEEE Transactions on Electron Devices**, IEEE, v. 59, n. 8, p. 2019–2026, 2012.
- KIM, K.; FOSSUM, J. G. Double-gate CMOS: Symmetrical-versus asymmetrical-gate devices. **IEEE Transactions on Electron Devices**, IEEE, v. 48, n. 2, p. 294–299, 2001.
- LATATRIPATHI, S. et al. Performance comparison of bulk FinFET with SOI FinFET in nanoscale regime. **International Journal of Electronics Signals and Systems (IJESS) ISSN**, v. 2, p. 2231–5969, 2012.
- LEE, J.-H. et al. Comparison of soi finfets and bulk finfets. **ECS Transactions**, The Electrochemical Society, v. 19, n. 4, p. 101–112, 2009.
- LIN, X.; WANG, Y.; PEDRAM, M. Joint sizing and adaptive independent gate control for finfet circuits operating in multiple voltage regimes using the logical effort method. In: IEEE PRESS. **Proceedings of the International Conference on Computer-Aided Design**. [S.l.], 2013. p. 444–449.
- LUNDSTROM, M. **MOSFETs: An Energy Band Treatment**. 2015. University Lecture. Disponível em: <https://nanohub.org/groups/ece305lundstrom/File:Eband_MOSFETs_S15.pdf>.
- MANOJ, C. et al. Device design and optimization considerations for bulk FinFETs. **IEEE transactions on electron devices**, IEEE, v. 55, n. 2, p. 609–615, 2008.
- MARCHI, M. D. et al. Top-down fabrication of gate-all-around vertically stacked silicon nanowire FETs with controllable polarity. **IEEE Transactions on Nanotechnology**, v. 13, n. 6, p. 1029–1038, Nov 2014. ISSN 1536-125X.
- MARRANGHELLO, F. S.; REIS, A. I.; RIBAS, R. P. Design-oriented delay model for CMOS inverter. In: **Integrated Circuits and Systems Design (SBCCI), 2012 25th Symposium on**. [S.l.: s.n.], 2012. p. 1–6.

MARRANGHELLO, F. S.; REIS, A. I.; RIBAS, R. P. Improving analytical delay modeling for CMOS inverters. **Journal of Integrated Circuits and Systems**, v. 10, n. 2, p. 123–134, 2015.

MASAHARA, M. et al. Demonstration of asymmetric gate-oxide thickness four-terminal FinFETs having flexible threshold voltage and good subthreshold slope. **Electron Device Letters, IEEE**, IEEE, v. 28, n. 3, p. 217–219, 2007.

MAURINE, P. et al. Transition time modeling in deep submicron cmos. **IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems**, IEEE, v. 21, n. 11, p. 1352–1363, 2002.

MOORE, G. E. Cramming more components onto integrated circuits. **Electronics**, v. 38, n. 8, April 1965.

NABAVI-LISHI, A.; RUMIN, N. C. Inverter models of cmos gates for supply current and delay evaluation. **IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems**, v. 13, n. 10, p. 1271–1279, Oct 1994. ISSN 0278-0070.

NEHRU, K.; SHANMUGAM, A.; VADIVEL, S. Design of 64-bit low power parallel prefix VLSI adder for high speed arithmetic circuits. In: **IEEE. Computing, Communication and Applications (ICCCA), 2012 International Conference on**. [S.l.], 2012. p. 1–4.

NOWAK, E. J. et al. Turning silicon on its edge (double gate CMOS/FinFET technology). **IEEE Circuits and Devices Magazine**, IEEE, v. 20, n. 1, p. 20–31, 2004.

POIROUX, T. et al. Leti-utsoi2. 1: A compact model for utbb-fdsoi technologies—part i: Interface potentials analytical model. **IEEE Transactions on Electron Devices**, IEEE, v. 62, n. 9, p. 2751–2759, 2015.

POSSANI, V. N. et al. Graph-based transistor network generation method for supergate design. **IEEE Transactions on Very Large Scale Integration (VLSI) Systems**, v. 24, n. 2, p. 692–705, Feb 2016. ISSN 1063-8210.

POSSANI, V. N. et al. Exploring independent gates in FinFET-based transistor network generation. In: **Proceedings of the 27th Symposium on Integrated Circuits and Systems Design**. Aracaju, BR: ACM, 2014. (SBCCI '14), p. 41:1–41:6. ISBN 978-1-4503-3156-2. Disponível em: <<http://doi.acm.org/10.1145/2660540.2661009>>.

QU, J. et al. Study of drain induced barrier lowering (dibl) effect for strained si nmosfet. **Procedia Engineering**, Elsevier, v. 16, p. 298–305, 2011.

ROSSELLÓ, J. L.; SEGURA, J. An analytical charge-based compact delay model for submicrometer cmos inverters. **IEEE Transactions on Circuits and Systems I: Regular Papers**, IEEE, v. 51, n. 7, p. 1301–1311, 2004.

ROSTAMI, M.; MOHANRAM, K. Dual-independent-gate FinFETs for low power logic circuits. **Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on**, IEEE, v. 30, n. 3, p. 337–349, 2011.

- SACHID, A. B.; KHANDELWAL, S.; HU, C. Body-bias effect in soi finfet for low-power applications: Gate length dependence. In: **Proceedings of Technical Program-2014 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA)**. [S.l.: s.n.], 2014.
- SAKURAI, T.; NEWTON, A. R. Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas. **IEEE Journal of solid-state circuits**, IEEE, v. 25, n. 2, p. 584–594, 1990.
- SAKURAI, T.; NEWTON, A. R. A simple mosfet model for circuit analysis. **IEEE transactions on Electron Devices**, IEEE, v. 38, n. 4, p. 887–894, 1991.
- SILVACO. **Berkeley Common-Gate Multi-Gate MOSFET Model**. 2014. Disponível em: <<http://www.silvaco.com/content/kbase/BSIM-CMG.pdf?20140403>>.
- TAWFIK, S. A.; LIU, Z.; KURSUN, V. Independent-gate and tied-gate FinFET SRAM circuits: design guidelines for reduced area and enhanced stability. In: IEEE. **2007 Internatonal Conference on Microelectronics**. Cairo, Egypt, 2007. p. 171–174.
- TRIVEDI, V.; FOSSUM, J. G.; CHOWDHURY, M. M. Nanoscale FinFETs with gate-source/drain underlap. **IEEE Transactions on Electron Devices**, IEEE, v. 52, n. 1, p. 56–62, 2005.
- VALDES, A. M. A. et al. Performance evaluation of optimized transistor networks built using independent-gate FinFET. In: **2016 IEEE 7th Latin American Symposium on Circuits & Systems (LASCAS)**. Florianópolis, BR: Institute of Electrical & Electronics Engineers (IEEE), 2016.
- WEIS, M.; EMLING, R.; SCHMITT-LANDSIEDEL, D. Circuit design with independent double gate transistors. **Advances in Radio Science**, Copernicus GmbH, v. 7, n. D. 5, p. 231–236, 2009.
- YANG, P.-L. et al. Vertical slit fet at 7-nm node and beyond. **IEEE Transactions on Electron Devices**, IEEE, v. 63, n. 8, p. 3327–3334, 2016.
- ZAREI, M. Y. et al. Modeling symmetrical independent gate FinFET using Predictive Technology Model. In: **Proceedings of the 23rd ACM International Conference on Great Lakes Symposium on VLSI**. New York, NY, USA: ACM, 2013. (GLSVLSI '13), p. 299–304. ISBN 978-1-4503-2032-0.
- ZHANG, L. et al. Modeling short-channel effect of elliptical gate-all-around mosfet by effective radius. **IEEE Electron Device Letters**, v. 32, n. 9, p. 1188–1190, Sept 2011. ISSN 0741-3106.
- ZHANG, W. et al. Physical insights regarding design and performance of independent-gate FinFETs. **IEEE Transactions on Electron Devices**, v. 52, n. 10, p. 2198–2206, Oct 2005. ISSN 0018-9383.
- ZHAO, W.; CAO, Y. Predictive technology model for nano-CMOS design exploration. **ACM Journal on Emerging Technologies in Computing Systems (JETC)**, ACM, v. 3, n. 1, p. 1, 2007.