Current and Delay Estimation in Deep Sub-micrometer CMOS Logic Circuits

By

Muaayad Al-Mosawy

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Master of Applied Science
in
Electrical Engineering

Department of Electronics
Carleton University
Ottawa, Ontario, Canada

January 2007

Your file Votre référence
ISBN: 978-0-494-26981-7
Our file Notre référence
ISBN: 978-0-494-26981-7

Abstract

The continual shrinking of CMOS device features raises some problems; one of the most relevant is signal propagation delay estimation. Using the circuit simulator is prohibitively time consuming. Faster, yet, accurate delay estimation methods are strongly needed. The propagation delay is affected by MOSFET current and capacitance, the applied signal slope and the number of serially connected MOSFETs.

This work proposes an extended MOSFET saturation current model for deep sub-micrometer technologies by adding the channel-length modulation effects. Moreover, the thesis proposes four classes of MOSFET capacitance models. These classes differ in the number of parameters they use to characterize MOSFET capacitances. In addition, the work adopts one more MOSFET capacitance class from the literature that recognizes the difference between the capacitances associated with a rising transition and those associated with falling transition.

Furthermore, the dissertation proposes four delay model levels that are related to the MOSFET capacitance classes. Also, two more delay levels are adopted from the literature: one that is based on logical effort technique and a more complicated one that is based on Shams model. Shams model is modified to give two more delay model levels. One of them uses a simpler way of accounting for the effect of serially-connected MOSFETs. The other one uses a more complicated expression for the input slope effect. An empirical technique to estimate the linear and quadratic effects of the input signal slope on the delay is also proposed.

The study is performed on 0.13 μm and 90 nm CMOS technologies to demonstrate how much complexity and details are required for reasonably accurate delay estimation.

In memory of

my father,

my mother,

and my brother Muhammad

Acknowledgements

First, foremost, and eternally, all thanks are due to Allah Almighty for innumerable blessings through out my life, one of which is this thesis.

Second, my great thanks are going to my thesis supervisor, Professor Maitham Shams. Absolutely, this thesis would not have been possible without him. His professional guidance, kind support, and remarkable patience are extremely appreciated.

I would like to thank my colleagues Raghid Shreih, Ziad El-Khatib, Reza Yousefi, and Syed Irfan Ahmed for their generous academic help whenever they could.

My thanks go to the faculty and staff of Department of Electronics at Carleton University, especially graduate study chair Professor Barry Syrett, graduate assistant Peggy Piccolo, department administrator Blazenka Power, computer systems manager Jacques Lemieux, and PC systems administrator Scott Bruce for their help during this thesis work. Also, I would like to thank interlibrary loan department staff in the university library for their help whenever needed. In addition, I would like to thank Dr. Marco S. Dragic, senior analog/RF IC engineer, from CMC Microsystems, Kingston, Ontario, Canada, for his professional and fast response to the threads I posted over CMC discussion forums.

Indeed, I owe my wife Suhaila special and deep thanks for her unlimited support to have me finish this degree. I will never forget her sacrifices in putting aside her plans and dreams whenever they would have conflicted with my study.

The valuable moral support and prayers of my brothers, sisters, and their families - especially my elder brother Sayed Dia - provided the restorative I needed at difficult times during the project.

Finally, I would like to thank everyone who helped in whatever way, but got forgotten when the above list was composed.

Contents

List of Tables

List of Figures

List of Abbreviations

BSIM	**Berkeley Short-Channel IGFET Model**
C_{db}	**Drain substrate capacitance**
$Cd\text{-}_{Bottom}$	**Bottom diffusion capacitance**
$C_{Diffusion}$	**Diffusion capacitance**
C_F	**Interconnect related fringing capacitance**
C_{gate}	**All gate related capacitance**
C_{gb}	**Gate- substrate capacitance**
C_{gCH}	**Gate-channel capacitance**
C_{gd}	**Gate-drain capacitance**
C_{gdOL}	**Gate-drain overlap capacitance**
C_{gOL}	**Overlapping related capacitance**
Cgs	**Gate-source capacitance**
$C_{dSide\text{-}Wall}$	**Side-wall diffusion capacitance**
C_{gsOL}	**Gate-source overlap capacitance**
C_H	**Interconnect related horizontal capacitance**
C_j	**Junction capacitance per unit area**
C_{jSW}	**Side-wall junction per meter capacitance**
C_L	**Load capacitance**
C_{ox}	**Per unit area gate capacitance**
C_{sb}	**Source-substrate capacitance**
C_V	**Interconnect related vertical capacitance**
C_{Wire}	**Interconnect capacitance per unit length**
D	**Delay**
D_G	**Gate delay**

DSMM	**Deep sub-micrometer**
E_x	**Horizontal electrical field**
Ex_c	**Horizontal critical electric field**
E_y	**Vertical electrical field**
g	**Gate logical effort**
h	**Electrical effort**
H	**Insulator thickness**
I_d	**MOSFET drain current**
L	**MOSFET device effective length**
LM	**MOSFET device mask length**
m	**Number of MOSFET connected in parallel to the same input**
M	**Mobility degradation index of this work**
MESFET	**Metal-Semiconductor Field Effect Transistor**
MOSFET	**Metal Oxide Semiconductor Field Effect Transistor**
n	**Number of MOSFET connected in series**
NMOS	**n-channel MOSFET**
p_{inv}	**NOT gate parasitic delay**
PMOS	**p-channel MOSFET**
q	**Number of MOSFET connected in parallel**
Q_n	**Electron charge density**
$S1$	**Input Signal First Order Slope Factor**
$S2$	**Input Signal Second Order Slope Factor**
S_f	**MOSFET dimensions scale-down factor**
SPICE	**Simulation Program with Integrated Circuit Emphasis**
S_x	**Slope factor**
T	**Interconnect height**
t_{ox}	**Oxide thickness**
TSMC	**Taiwan Semiconductor Manufacturing Company**
U_f	**MOSFET Supply and MOSFET terminals voltage scale-down factor**
V_{db}	**MOSFET drain to substrate voltage**
V_{dd}	**Supply voltage**

V_{dsat}	**Drain saturation voltage**
V_{ds}	**MOSFET source to drain voltage**
V_{gb}	**MOSFET gate to substrate voltage**
V_{gd}	**MOSFET gate to drain voltage**
V_{gs}	**MOSFET gate to source voltage**
V_{sb}	**MOSFET source to substrate voltage**
V_t	**Threshold voltage**
W	**MOSFET device width**
x	**Empirical parameter used to calculate Y**
X	**Delay degradation factor for load capacitance due to series connected MOSFETs**
X_j	**Junction depth**
y	**Empirical parameter used to calculate Y**
Y	**Delay degradation factor for internal capacitances due to series connected MOSFETs**
Z	**Velocity saturation index of this work**
β	**Transistor gain factor**
μ_{eff}	**Effective carrier mobility**
μ_n	**Electrons mobility**
μ_o	**Mobility in the substrate**
μ_p	**Holes mobility**
μ_x	**Horizontal mobility**
μ_y	**Vertical mobility**
v	**1 - Electron velocity** **2 – Mobility degradation index in Shams model**
v_{sat}	**Saturation velocity**
ν	**MOSFET resistance per unit width**
λ	**Channel-length modulation**
α	**1 - Velocity saturation index in α-power law** **2 - Empirical parameter used to calculate slope factor**
ζ	**Velocity saturation index in Shams model**
κ_t	**Empirical technology and device length dependent parameter of this work**

τ	**Input signal transition time**
τ_{inv}	**Inverter delay driving identical inverter with no parasitics**
δ	**MOSFET resistance in linear to its resistance in saturation**

Subscripts

C	**Related to cell gate**
D	**Related to drive gate**
equ	**Related to equivalent**
F	**Related to falling transition**
L	**Related to load gate**
N	**Related to NMOS**
P	**Related to PMOS**
R	**Related to rising transition**
S	**Related to Step**
T	**Related to total**
Wire	**Related to interconnect**

Chapter 1

Introduction

Over the past forty years, Complementary Metal Oxide Semiconductor (CMOS) technology has been introduced as a leading contender in VLSI systems. Today, CMOS is the dominant technology for microprocessors, memories, application specific integrated circuits (ASICs) and even analog circuits. CMOS mature fabrication facilities and its immunity against static power dissipation are two reasons among others that help in putting CMOS technology in such leading position. Current density increases as channel-length decreases and long-channel models no longer evaluate the device current accurately. Therefore, it is essential to introduce accurate methods to compute the current in deep sub-micrometer CMOS circuits. The delay has a strong relationship to the current.

The availability of reliable and precise computer aided design tools and process technologies has made possible a still-continuing evolution in circuit size and complexity, so that design of high speed digital circuits has become one of the concentrated areas of endeavour in the VLSI industry. The primary means of sustaining this situation has been device scaling. VLSI technology has pushed feature sizes down to the deep sub-micrometer level, with dimensions of 0.13 μm and 90 nm and smaller. Device dimensions down-scaling make some previously-ignorable physical effects to have great influence on the short-channel VLSI circuits. Effects like velocity saturation, mobility degradation, and channel-length modulation make it quite a challenge to predict CMOS circuit performance.

1.1 Thesis Motivation

In every technology generation, transistor per area unit is approximately doubled. Hence, VLSI chips grow in functionality and switching frequency. As a result, millions

of parasitic capacitances charge and discharge at aggressively increasing rates.

In CMOS digital circuits timing analysis, signal path delay is one of the most critical performance parameters. Estimating this delay is essential to evaluate the overall operation of these circuits, since it is an important figure of merit associated with a circuit's speed and capability.

MOSFET current and parasitic capacitances over the signal path determine how fast or slow the signal would pass through. Conventional models that are suitable for describing the long-channel technologies are no longer valid when applied to short-channel technologies.

The trade-off between accuracy and speed makes existing commercial circuit simulators models, such as SPICE, consume too much CPU time. Therefore, computer aided design (CAD) tools must include efficient techniques for rapid, yet accurate enough, simulation of digital ICs. In contrast to numerical approaches, analytical methods can offer faster and fairly accurate delay estimation.

1.2 Thesis Objective

In deep sub-micrometer CMOS digital design, propagation delay is of critical importance. Fast and accurate techniques to evaluate this delay are needed for VLSI digital designers to verify and optimize their designs. An analytical approach that does not need the full details of every single element in the circuits is a common technique that is used in this direction. This technique analytically derives timing models for the CMOS circuit from various MOSFET models that describe its I-V characteristics and parasitic capacitances.

The signal propagation delay is not only affected by how precise these models are, but also by the input signal transition time. This signal slope factor is of remarkable effect on the path propagation delay. The number of series connected MOSFETs (SCM) along the signal delay path is another factor that affects the propagation delay.

As the feature size decreases, gate delay decreases while interconnect delay increases. The overall logic path delay is the sum of path gates delay component and the path wiring delay component. In older circuits, the devices were large enough that the gate delay component was the dominant one and the interconnect component could be disregarded. Now, with the interconnect resistance and capacitance increasing due to smaller cross

sectional area and narrower minimum distance, in addition to larger interconnect trees, the interconnect delay component is becoming a significant portion of overall logic path delay.

The objective of this thesis is to investigate and to compare relatively simple CMOS delay models for deep sub-micrometer technologies and determine the level of detail necessary for reasonable accuracy. Model delay estimation with accuracy of less than 10% away from the simulator delay is achieved at some points.

This objective is reached through investigating, adopting, modifying and proposing aspects as summarized in Table 1.1 with related chapters and sections:

Table 1.1: Thesis work classification in four categories

Category	Work	Chapter	Section	Comment
Investigated	Channel-length modulation vs. channel-length	2	2.4.3	For NMOS and PMOS in five technologies
	Delay through serially connected MOSFETs vs. their number	3	3.5.1	For NMOS series in four technologies
	Input signal linear and quadratic slopes vs. technology	3	3.7	For NMOS and PMOS in four technologies
Adopted	MOSFET capacitance extraction Class 5	4 and Appendix B	4.3.5 and B5	Shams [24]
	Delay model Level 1	5	5.2	Shams [24]
	Delay model Level 7	5	5.3	Logical Effort [56]
Modified	Delay model Level 6	6	6.1.3	Based on delay model Level 7
	Delay model Level 8	6	6.1.5	Based on delay model Level 7
Proposed	Extended deep sub-micro meter drain saturation current model	3 and Appendix A	3.1, A1, and A2	Takes care of velocity saturation, mobility degradation, and channel-length modulation effects
	Empirical technique to extract input signal linear and quadratic slopes	3 and Appendix A	3.7, and A5	Based on a chain of three inverters
	MOSFET capacitance extraction Classes 1, 2, 3, and 4	4 and Appendix B	4.3, B2, B3, B4, and B5	The input signal slope is implemented within extracted value
	Delay model Levels 2, 3, 4, and 5	6	6.1.2	Based on MOSFET width, resistance, and capacitances (Class 1-4)

The required parameters and factors were extracted as required. The extraction procedures were fully explained in the appendices and the resulted values were tabulated in the related chapters.

Several applications were used to show how accurate the models calculate the delay comparing to the simulated delay using BSIM4 Cadence-Spectre simulator in TSMC 0.13 μm and TSMC 90 nm technologies.

1.3 Thesis Outline

This thesis is written in seven different chapters. This chapter, Chapter 1, gave motivation and objective. Chapter 2 is a literature-based review of MOSFET physical structure, basic operation, dimensional scaling, capacitance and delay.

Chapter 3 discussed several MOSFET saturation current models. Differences between those applying to long- and short-channel devices are noted. It also introduced an extended model that captures the three most important effects on channel current in short-channel technologies. An analytical delay model suited to deep sub-micrometer transistors is discussed. The factors used in current and delay models are described and extracted. The 2^{nd} order effects of the input signal are discussed and 1^{st} and 2^{nd} order slope factors are extracted.

In Chapter 4, the nature and origins of MOSFET capacitances are described in detail, with a mathematical model to estimate them in TSCM 0.13 μm and TSMC 90 nm technologies being presented. The way in which interconnect capacitances affect the delay is described, as well as how the worst-case interconnect capacitance might be estimated in both technologies. Four classes of MOSFET capacitance model are proposed, and a fifth is taken from [24]; Chapter 4 offers a detailed procedure to extract the parameters of these five capacitance models for use with 0.13 um and 90 nm technologies.

CMOS gate delay modeling is the subject of Chapter 5. The collapsing technique is discussed. The delay of a chain of MOSFETs as described in [24] is presented in a mathematical form to be later modified, giving two more levels at which to estimate path delay. A further four delay-model levels are proposed, ones based on the four MOSFET capacitance models of Chapter 4. Logical effort, which is the basis of the delay model Level 1, is also discussed in this chapter.

In Chapter 6, five different applications are designed. Their critical delay paths are identified and the corresponding delays are calculated. The custom layout of Application 1 is done in both 0.13 μm and 90 nm technologies. Simulation of this application circuit was performed according to schematic and extracted views. All other applications are simulated based on the schematic views only. This chapter offers an excellent opportunity to confirm the delay model accuracies with respect to the circuit simulator delay measurements.

The five applications delay estimations using the delay model eight levels are compared to the simulator in Chapter 7. This compression was done in two directions, speed and accuracy. Chapter 8 gave concluding remarks, thesis contribution and possible future work.

The references are followed by two appendices, A and B. These two appendices were assigned to give detailed procedures of extracting the current, delay and capacitances different parameters and values.

Chapter 2

The MOSFET in Review

The relatively simple manufacturing process, higher density of devices per chip, higher yield, and low cost make the Metal Oxide Semiconductor Field Effect Transistor (MOSFET) the heart of almost every modern electronic circuit.

With four terminals, the MOSFET device can be thought of as a perfect switch. The Applied voltage to the Gate (G) terminal controls if and how much current is to flow between the Source (S) and Drain (D) terminals. The fourth MOSFET terminal is the Body (B) or substrate itself. This terminal's purpose is auxiliary as it is only used to modulate the MOSFET parameters [1].

The MOSFET can be one of two types, the n-channel MOSFET (NMOS) in which the source and drain are an n+ type doped region implanted in a p-substrate. The other type is the p-channel MOSFET (PMOS). In PMOS, the source and the drain are of p+ type material implanted in an n-substrate. Usually, the PMOS is fabricated in an n-well on a p-substrate.

MOSFETs are said to operate in two regions, the "Linear" or "Triode" region and the "Saturation' or "Active" region. Basically, the current flows between the source and drain; can be written for these two regions as a function of the applied voltages to the device terminals, the device physical dimensions, the mobility of moving carriers in the channel and the gate to substrate capacitance value.

Figure 2.1 shows the physical structure of an n-channel MOSFET and the circuit symbols of NMOS and PMOS, devices. The gate (G) is usually implanted using polysilicon.

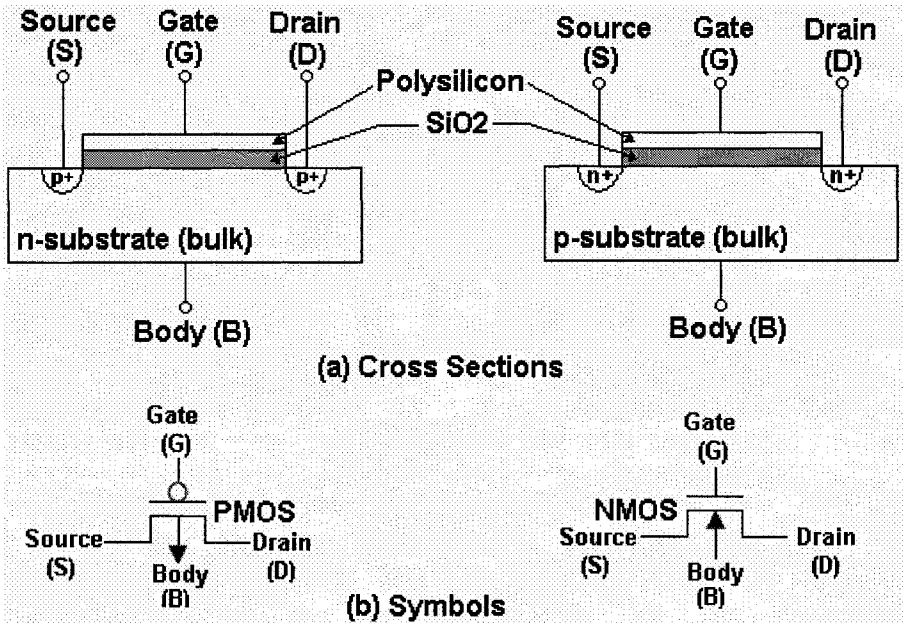


Figure 2.1: MOSFET, NMOS and PMOS

2.1 Brief History

Historically, the development of MOSFET can be divided into three phases: the discovery phase (1928-1958), the technology development and new device structure invention phase (1959-1968), and lastly the MOS transistor integration and integrated circuit manufacturing phase (1968 – today) [2].

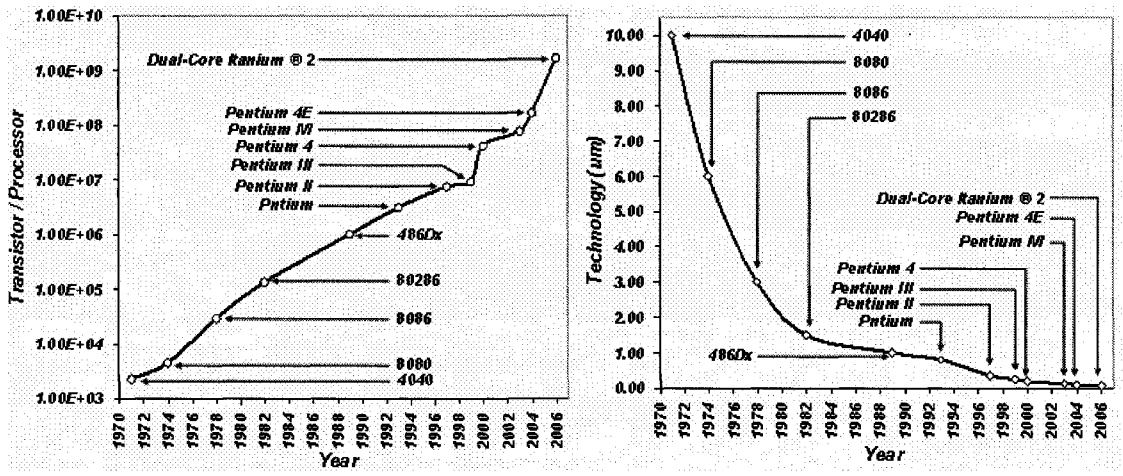
In the 1926, Edgar Lilienfeld filed a patent proposing the Metal-Semiconductor Field Effect Transistor (MESFET) [1]. Lilienfeld filed another patent after two years of promoting the idea of depletion mode MOSFET [3]. There is no evidence that he ever built and tested any prototype out of these patents. In 1935, Oskar Heil applied for a patent on "Improvement in, or relating to electric amplifiers and other control arrangements and devices" in which he proposed the "Inversion-Mode MOSFET" [4].

A research group headed by William Shockley and including Walter Brattain and John Bardeen discovered the "Point-Contact Transistor" in 1947. The "Bipolar Junction Transistor" was proposed by Shockley a year later and the first BJT was made in 1950 displacing the point-contact transistor and starting the transistor era. In 1956, the three Bell Telephone Laboratories researchers, Bardeen, Brattain and Shockley shared the Nobel Prize for their researches that led to discover the transistor effect [5].

Some surface problems like leakage current and device stability were common with the BJTs. In 1956, Bell Labs made Martin Atalla a head of a group to work on this issue. In 1960 and over their work, along with that of Dawon Kahng (a member of Atalla group) invented the metal oxide semiconductor field-effect transistor (MOSFET) [6]. The first two commercial MOSFETs saw the light in late 1964 [2] by Fairchild and RCA Companies and it was not before 1966 when Bell statrted using MOS technology in its projects [5].

The computer industry and the development of integrated circuits provided the MOSFET technology its current importance. The year 2000 Nobel Prize in Physics reciver, Jack Kilby succesed in 1958 to fabricate a simple flip-flop circuit on a single chip of germanium [7]. Another milestone came in 1971, when Intel began manufacture of its 2300-transistor 4004 microprocessor [3]. In the early 1980's, the domination of the integrated-circuit industry by MOSFET-based circuits had begun [3].

In 1982, Intel released its 80286 microprocessor. Based on CMOS technology, this procssor contained about 134 k transistor and operated from a 5V supply. The internal logic 'high' state was also 5 volts [8]. Using 0.8 μm technology, Intel released the Pentium® processor in 1993 with 3.1 M transistor [8]. Ten years later, Intel introduced its Pentium® M processor with 77 M transistor built in 0.13 μm technology [9]. In 2006, Intel released its Dual-Core®Itanium®2 processor with 1.7 G transistor built in 0.65 μm technology [10]. Figure 2.2 shows the number of transistors and the used technology in Intel processors over 35 year of developments (plotted depending on data gathered from [2] - [10]).

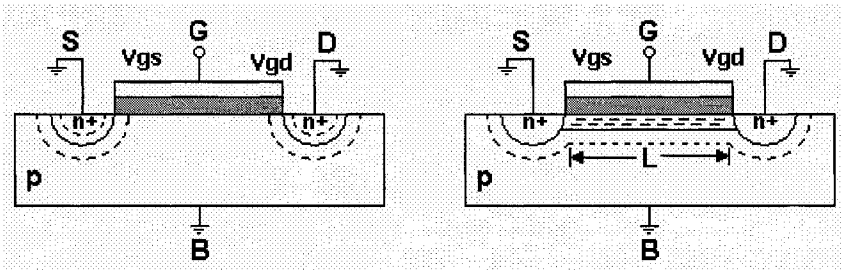


(a) Transistor/Processor **(b) Technology/Processor**

Figure 2.2: Intel processors, transistors and technology as a function of year

2.2 MOSFET Basic Operation

In NMOS transistor, when applying a positive voltage to the gate, this attracts the electrons to the region underneath the gate forming a channel. For the common operating conditions, usually the source and body are grounded. When the gate voltage is smaller than the device threshold voltage (V_t), Figure 2.3 (a), it is not strong enough to attract electrons to the area under the Si-SiO_2 junction. As the gate voltage increases to be greater than V_t, Figure 2.3 (b), enough electrons are attracted to the area under the Si-SiO_2 junction.



(a) $V_{gs} < V_t$ (b) $V_{gs} > V_t$

Figure 2.3: NMOS Device with V_{gs} less and greater than V_t

This will form a channel that provides an electrical path between the drain and source. As the potential applied to the gate, V_{gs}, increases, the channel electron density does so too. The electrons charge density, Q_n, is given as [11]:

$$Q_n = C_{ox} \cdot (V_{gs} - V_t) \qquad (2.1)$$

where V_t is the device threshold voltage and C_{ox} is the gate capacitance per unit area:

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \qquad (2.2)$$

where ε_{ox} is the relative permittivity of Silicon Oxide, and t_{ox} is the oxide thickness.

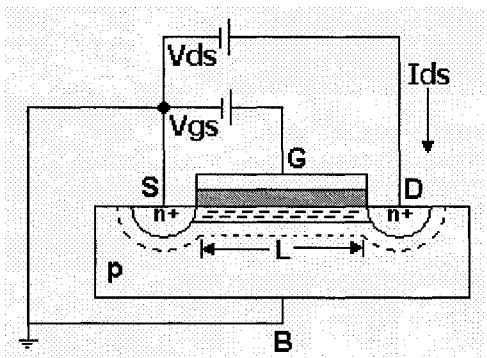


Figure 2.4: NMOS Device with $V_{gs} > V_t$ and $V_{ds} > V_t$ (small difference)

When a potential difference exists between the source and the drain as in Figure 2.4, a current will flow from the drain to the source. For small V_{ds}, I_{ds} is [12]:

$$I_d = \mu_n . Q_n . \frac{W}{L} . V_{ds} = \mu_n . C_{ox} . \frac{W}{L} . (V_{gs} - V_t) . V_{ds} \qquad (2.3)$$

Figure 2.5 shows I_{ds} vs. Vd_s plots, one as obtained from the circuit simulator (Cadence-Specter, in 0.13 μm technology for minimum length and W/L = 5) and the other as determined from Equation 2.3.

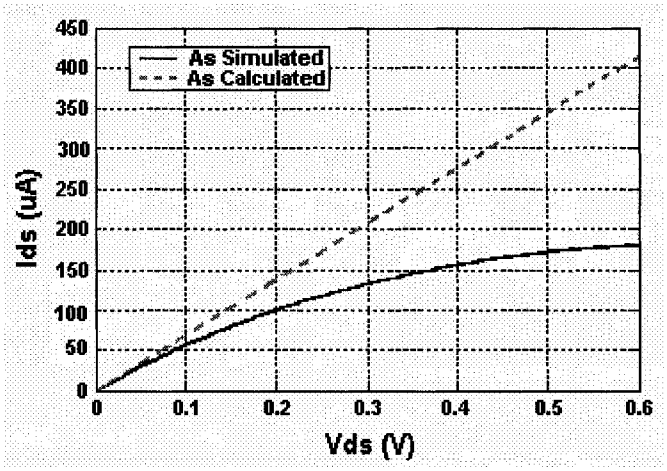


Figure 2.5: NMOS, Ids vs. Vds

For a very small V_{ds} values, the two curves match closely, but diverge as V_{ds} increases. The reason for this is that increasing V_{ds} is accompanied by decreasing V_{gd} and a resulting decrease in the channel charge density at the drain end. The simulator model takes account of this 'pinching off', whereas Equation 2.3 does not.

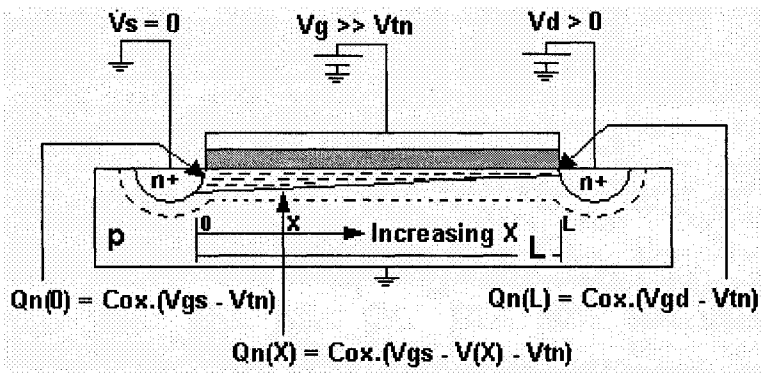


Figure 2.6: n-channel charge density

Now, let us consider Figure 2.6. The drop in the voltage with the direction of the channel length, dV_{ch}, at a certain distance in the channel, dx, can be written in term of the electron mobility, μ_n, and the channel charge at point x, channel current, I_d at x, and the channel width as [11]:

$$\mu_n.W.Q_n(x).dV_{ch} = I_d.dx \qquad (2.4)$$

At the distance x from the source in the channel, the charge density is [11]:

$$Q_n(x) = C_{ox}.(V_{gs} - V_{ch}(x) - V_t) \qquad (2.5)$$

Substituting $Q_n(x)$ from Equation 2.5 into Equation 2.4 gives:

$$\mu_n.W.C_{ox}.(V_{gs} - V_{ch}(x) - Vt).dv = I_d.dx \qquad (2.6)$$

The voltage across the channel is changing from 0 to V_{ds} and the distance is changing from 0 to L. Integrating both sides of Equation 2.6 between these limits gives:

$$\int_0^{V_{ds}} \mu_n.W.C_{ox}.(V_{gs} - V(x) - V_t).dv = \int_0^{L} I_d.dx \qquad (2.7)$$

$$\mu_n.W.C_{ox}[(V_{gs} - V_t).V_{ds} - \frac{V_{ds}^2}{2}] = I_d.L \qquad (2.8)$$

Solving Equation 2.8 for Id will give:

$$I_d = \mu_n.C_{ox}.\frac{W}{L}.[(V_{gs} - V_t).V_{ds} - \frac{V_{gs}^2}{2}] \qquad (2.9)$$

Equation 2.9 is showing that the drain current, I_d, is changing with the change of V_{ds}. This region of operation is called the "Triode Region' or the "Linear Region".

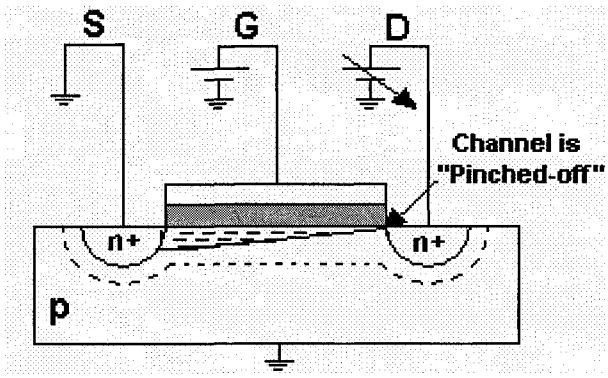


Figure 2.7: Channel pinched-off

As V_{ds} increases farther, V_{gd} decreases to a point where $V_{gd} = V_t$. At this value, the channel is no longer forming a complete path from the source to the drain. The channel is "pinched-off" as shown in Figure 2.7 and the transistor is operating now in the "Active Region" or "Saturation Region". The voltage across the channel then can be estimated by setting $V_{gd} = V_t$:

$$V_{gd} = V_t = V_{gs} - V_{ds} \qquad (2.10)$$

or:

$$V_{ds} = V_{g_s} - V_t \qquad (2.11)$$

By Substituting the V_{ds} value from Equation 2.11 into Equation 2.9, we can write the drain current as:

$$I_d = \frac{1}{2} \cdot \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot (V_{gs} - V_t)^2 \qquad (2.12)$$

Equation 1.12 is the well-known Shockley Square Law model [13].

2.3 MOSFET Down-Scaling

The industry's transition from bipolar to CMOS technology triggered a continuous and rapid development of MOSFET technology. The most important change in the devices has been decreasing channel length, expected to shrink as low as 7 nm by the year 2019[14].

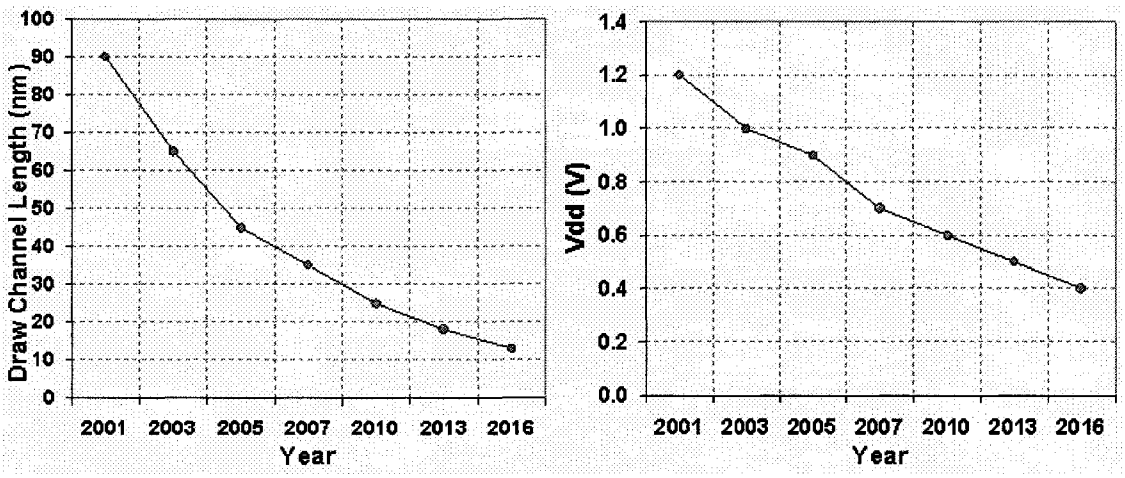


Figure 2.8: Device length and voltage supply expected scaling down

In parallel with channel length reduction, V_{dd} supply voltages are estimated to experience a reduction of 60% by the year 2016, to 0.4V from the present 1.0V [1]. Figure 2.8 illustrates recent behavior of these values and predictions for the future (data source: [1]).

The ultimately targeted goal of MOSFET down-sizing is one of smaller devices operating at higher speeds [15, 16] to be accompanied by lower power dissipation [17].

This scaling will continue until physical limitations prevent its continuance [18]. These limitations are ones imposed not only by Quantum Mechanics, but also by the changes in operational characteristics attending changes in device dimensions, and fabrication problems which may make further shrinkage impractical [18]. MOSFET scaling can be done in one of three ways, namely Full Scaling, Constant-Voltage Scaling, and General Scaling.

2.3.1 Full Scaling

In this approach, all device dimensions and potentials are reduced by the same factor. If $S_f > 1$ is defined as this factor, the dimensional changes shown in Figure 2.9 will occur.

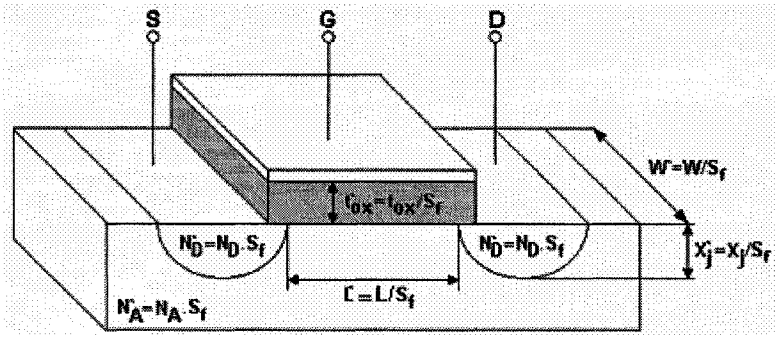


Figure 2.9: Scaling down the main dimensions of typical MOSFET by S_f

As the device length and width decrease by the factor S_f, the device area is reduced by S_f^2 resulting in a greater device per chip density. The total delay is the sum of the intrinsic delay and the load delay [19]. Scaling down the device sizes will reduce its intrinsic delay by $1/S_f$ and hence leading to a faster switching device.

There are disadvantages to this approach: the full scaling technique requires that the supply voltage and device terminal voltages be reduced, but for practical reasons this

reduction cannot be applied to all components in a system. For example, the interface circuitry may operate on different voltage levels and the multi-voltage supplies needed in such a situation would mean a more-complicated overall design. Furthermore, decreasing the device threshold voltage will increase the unwanted threshold-dependent leakage current by a factor of 12 for each 100mV decrease in V_t [20].

2.3.2 Constant-Voltage Scaling

The supply voltage and the device terminals voltages are kept un-changed while device dimensions are reduced by a factor of S_f, as in the full scaling technique.

Scaling down the oxide thickness will increase C_{ox} by the factor S_f. But because the terminal voltages remain unchanged, the saturation current after scaling will be increased by the scaling factor S_f. An increase in drain current may lead to other problems, such as electromigration and oxide breakdown. Also, it may result in severe heating problems and more-complicated heat dissipation procedures [21].

2.3.3 General Scaling

General Scaling attempts a compromise between the full and constant-voltage scaling techniques. Two different scaling factors are used in this method. The first one applies to all device features except the supply and the device terminal voltages, for which the other factor is used.

Table 2.1: Summary of scaling techniques

Feature	*Feature Scaled By*		
	Full Scaling	*Fixed-Voltage Scaling*	*General Scaling*
V_{dd}	$1/S_f$	1	$1/U_f$
V_t	$1/S_f$	1	$1/U_f$
Length	$1/S_f$	$1/S_f$	$1/S_f$
Width	$1/S_f$	$1/S_f$	$1/S_f$
Oxide Thickness	$1/S_f$	$1/S_f$	$1/S_f$
Area	$1/S_f^2$	$1/S_f^2$	$1/S_f^2$
Intrinsic Delay	$1/S_f$	$1/S_f$	$1/S_f$
Power	$1/S_f^2$	S_f	$1/U_f^2$
Power Density	1	S_f^3	S_f^2 / U_f^2

As in full scaling and constant-voltage scaling, S_f is the down-scaling parameter applicable to all of the device's physical features, and U_f is used for the supply and device terminal voltages. The relative magnitudes of the two parameters obeys $S_f > U_f > 1$.

Table 2.1 gives a summarization of the three scaling down techniques.

2.4 Scaling Effects on the MOSFET Behavior

When the effective channel length of a MOSFET is comparable to its source or drain junction depth, the device is categorized a short channel device [18]. In long-channel devices, the current is assumed to flow on the Silicon surface with the electric field being distributed evenly along that plane [22]; this is not the case with short channel device. Short channel MOSFETs are subject to performance-influencing effects not seen in their long-channel counterparts, three of them being

1 – Velocity Saturation,

2 – Mobility Degradation, and

3 – Channel Length Modulation

These have dramatic effects on short channel device characteristics and will therefore be incorporated into the extended saturation drain current model proposed in Chapter 3 of this work. Items 1, 2 and 3 will be described in following sections.

The reader can refer to [1], [20] and [23] for information on other phenomena such as the Hot-Carrier, Punch-Through and Drain-Source Series Resistances effects.

2.4.1 Velocity Saturation

The precise scaling of the MOSFET requires scaling down both of device supply and terminal voltages. Such scaling cause a decrease in the device noise immunity and reduces the input-output compatibilities with the surrounding devices. In General Scaling, section 2.3.3, two scaling factors are used to scale down the supply and terminal voltages slower that the other device features. This attempt increases the electrical fields inside the MOSFET and it is no longer represented by the conventional theories [1].

Drain to source voltage, V_{ds}, is higher in short channel device comparing to long-channel device. As a result, the lateral electrical field, Ex, increases and the carriers in the

channel have an increased velocity.

Electrons velocity, v, is a function of electrons mobility, μ, and the electric field E. When the electrical field is small, μ has an independent relation with E and is constant. Moving electrons in the channel use to collide with Silicon atoms; this is known as "Scattering".

For weak electric field, carriers scattering is a linear process as illustrated in Figure 2.10 by straight line that has a constant slope. However, under a stronger field, the carriers fail to maintain this linear relation. The slope decreases to zero at the "saturation velocity", v_{sat}, where further increases in E have no effect on the velocity.

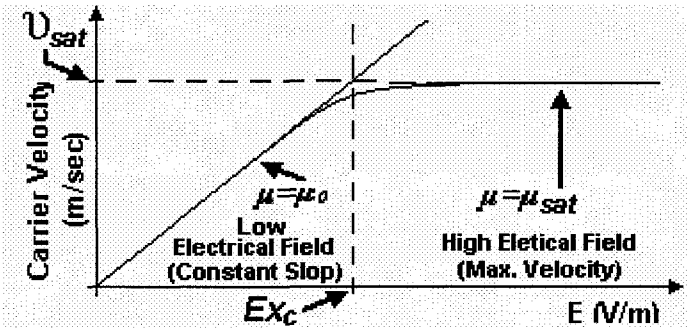


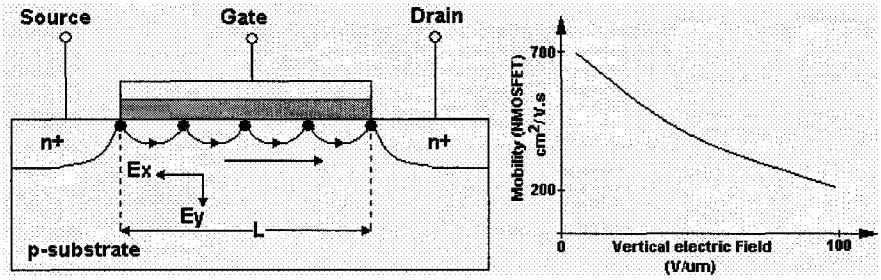
Figure 2.10: Carrier velocity vs. electrical field

The lateral electric field for which the velocity is saturated is called the critical electrical field, Ex_c. The velocity saturation causes the drain current to saturate at a lower voltage than $(V_{gs}-V_t)$. The magnitude of the saturation current deviates from the quadratic dependence of the Shockley model (Equation 2.12), which was developed for the long channel devices and neglects velocity saturation.

2.4.2 Mobility degradation

The carriers in the channel are subject to longitudinal (E_x) and transverse (E_y) electric fields, as shown in Figure 2.11. The corresponding electron mobilities are the horizontal and vertical mobilities, μ_x and μ_y.

The lateral or horizontal electric field, Ex, is related to the applied voltage to the device drain terminal. The horizontal mobility is a function of mobility in the substrate, μ_o, V_{ds}, effective channel length, and the horizontal critical electric field, Ex_c.



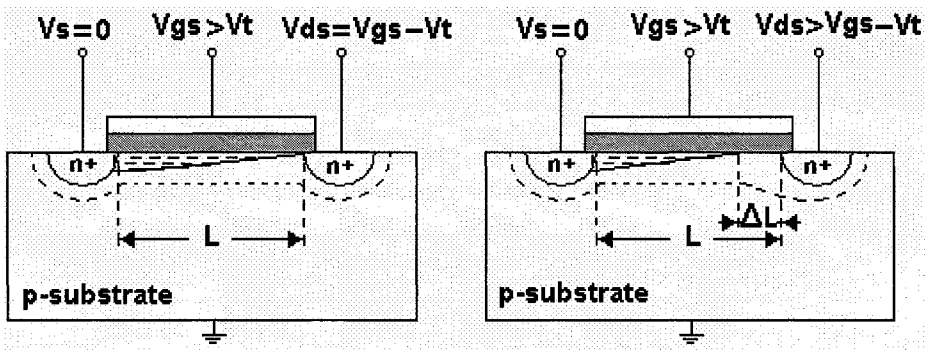
(a) Vertical and horizontal fields **(b) Mobility vs. Ey**

Figure 2.11: Electric fields and mobility

The transverse or vertical electric field, *Ey*, is related to the device gate terminal voltage. When the electrons move across the channel under *Ex* effects, the vertical electric field attracts these electrons to Si-SiO_2 interface causing them to "Rebound". The higher the field is the grater the rebounding. Because Si-SiO_2 interface is rough, the carriers will move with more difficulties. Electrons vertical mobility, μ_y, will include gate voltage rather than drain-source voltage.

2.4.3 Channel-Length Modulation

Shockley model in Equation 2.12 does not show the influence of V_{ds} on I_{ds} in the saturation region. Practically, increasing V_{ds} in the saturation region causes I_{ds} to increase as well.



(a) $V_{ds} = (V_{gs}-V_t)$ **(b) $V_{ds} > (V_{gs}-V_t)$**

Figure 2.12: Channel length modulation

At $V_{ds} = (V_{gs}-V_t)$, the channel is pinched-off as shown in Figure 2.12 (a). Increasing V_{ds} further, $V_{ds} > (V_{gs}-V_t)$, will increase the drain electric field and increase the size of the depletion region about the drain connection (Figure 2.12 (b)). The channel's effective length

is now less by ΔL because of the expanded depletion region. We replace L in the Shockley Equation with its reduced value ($L - \Delta L$) to get:

$$I_d = \frac{1}{2}.\mu_n C_{ox}.\frac{W}{L-\Delta L}.(V_{gs}-V_t)^2 \qquad (2.13)$$

It is clear now that as the channel length decreases due to increasing V_{ds}, the drain saturation current, I_{ds} will increase. By means of the binomial theorem, Equation 2.13 can be converted to the convenient form [1]:

$$I_d = I_{dsat}.(1+\lambda.V_{ds}) \qquad (2.14)$$

I_d has been written in terms of I_{dssat}, the drain saturation current of the Shockley model (Equation 2.12), and λ the "channel-length modulation parameter". λ is an empirical and technology-dependant parameter that is proportional to the inverse of device length.

The smaller MOSFET size becomes, the more important the channel-length modulation effect becomes [1]. Five different technologies were studied in the course of this work and values of λ for the NMOS and PMOS devices of each were extracted. We moved from extremes of 0.8 μm and 90 nm technologies and checked 0.35 μm, 0.18 μm and 0.13 μm technologies on the way: it was found that λ increases as the device length decreases. Figure 2.13 gives an illustrated of this finding. In each case, channel length was the minimum for that technology and the value of W/L was 5.

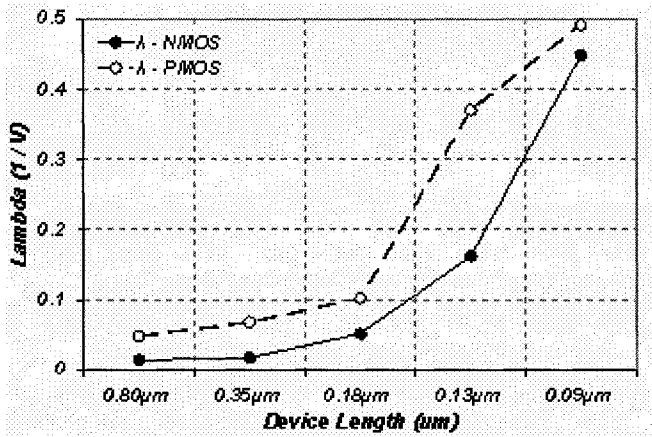


Figure 2.13: Lambda, λ, as a function of device length

2.5 MOSFET Capacitance Review

MOSFET operation is affected by two types of capacitances, external capacitances due to the interconnecting lines and the load, and ones that are device-intrinsic (parasitic capacitances). MOSFET parasitic capacitances originate with three sources: MOS structural capacitance, channel capacitance and junction capacitances [1].

These parasitic capacitances have differing characteristics; they might be distributed, voltage-dependent or nonlinear, and their exact modeling can be a very complicated process. Modeling is simplified by assigning constant values, dependant on their origins, to them

2.5.1 Gate Capacitance

Figure 2-14 shows the cross-sectional and top views of an n-channel MOSFET. LM is the mask gate length. Theoretically, the source and drain diffusion should end exactly at the mask borders, but in practice this is not the case.

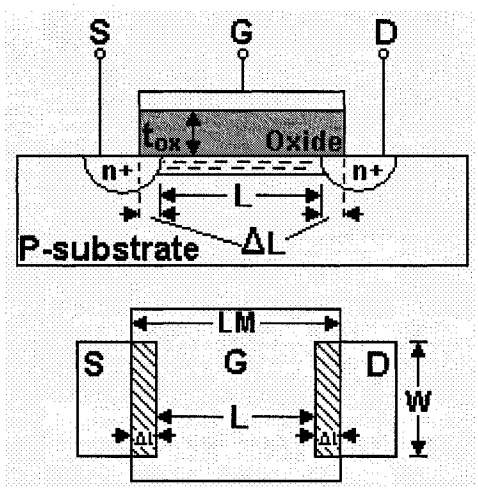


Figure 2.14: MOSDET cross-sectional and top views

A lateral diffusion occurs when the source and drain regions diffuse underneath the gate. This makes the channel length, L, less than the designed mask length:

$$L = LM - 2.\Delta L \qquad (2.15)$$

The lateral diffusion creates two parasitic capacitances, one between the gate and the source and the other between the gate and the drain. They are called "Overlap" capacitances and we will give them the symbols C_{gsOL} for gate-source overlap capacitance and C_{gdOL} for

gate-drain overlap capacitance. Each one can be represented by the gate oxide capacitance per unit area, C_{ox}, and the lateral diffusion area as:

$$C_{gsOL} = C_{gdOL} = C_{ox}.\Delta L.W \qquad (2.16)$$

C_{gsOL} and C_{gdOL} can be joined together in one overlapping related capacitance as:

$$C_{gOL} = C_{gsOL} + C_{gdOL} = 2.C_{ox}.\Delta L.W = C_{OL}.W \qquad (2.17)$$

The silicon oxide isolating the gate from the channel serves as a dielectric for three gate-channel related capacitances, involving the gate and the source (C_{gs}), the gate and the drain (C_{gd}), and finally the gate and the body (C_{gb}). The total gate to channel capacitance, C_{gCH} can then be written as:

$$C_{gCH} = C_{gs} + C_{gd} + C_{gb} \qquad (2.18)$$

For each of the three MOSFET operation regions, C_{gCH} can be described. When the MOSFET is off (cut-off region) there is no gate voltage applied to create the channel. Therefore, C_{gs} and C_{gd} are both equal to zero. The third capacitance, C_{gb}, is written as [1]:

$$C_{gb} = C_{ox}.LW \qquad (2.19)$$

In the linear region, there is a continuous channel between source and drain which lies between the substrate and the entire gate. In this case, C_{gb} is zero and there will be an almost equal sharing of gate-channel capacitance between the source and drain [1]:

$$C_{gs} = C_{gd} = \frac{1}{2}.C_{ox}.LW \qquad (2.20)$$

Finally, in the saturation region the MOSFET channel will be pinched-off. Electrically, the channel still connects the source and drain, and so the gate will again be isolated from the substrate and C_{gb} again be equal to zero. We will have [1]:

$$C_{gs} = \frac{2}{3}.C_{ox}.LW \qquad (2.21)$$

All gate related capacitances can be combined in one big gate capacitance, C_{gate}:

$$C_{gate} = C_{gOL} + C_{gCH} \qquad (2.22)$$

Table 2.2 gives a summary of the gate capacitance for the different MOSFET operation regions.

Table 2.2: Gate capacitance for different operation regions

Operation region	*Cut-off*	*Linear*	*Saturation*
C_{gb}	$C_{ox}.L.W$	0	0
C_{gs}	0	$\frac{1}{2}.C_{ox}.L.W$	$\frac{2}{3}.C_{ox}.L.W$
C_{gd}	0	$\frac{1}{2}.C_{ox}.L.W$	0
C_{gCH}	$C_{ox}.L.W$	$C_{ox}.L.W$	$\frac{2}{3}.C_{ox}.L.W$
C_{gOL}	$C_{OL}.W$	$C_{OL}.W$	$C_{OL}.W$
C_{gate}	$C_{ox}.L.W + C_{OL}.W$	$C_{ox}.L.W + C_{OL}.W$	$\frac{2}{3}.C_{ox}.L.W + C_{OL}.W$

2.5.2 Diffusion Capacitance

This capacitance is also called the "Junction Capacitance". Two diffusion capacitances are associated with the MOSFET, the source to substrate, C_{sb} and the drain to substrate, C_{db}. They are due to the reverse-biasing of the source-substrate and drain-substrate pn-junctions.

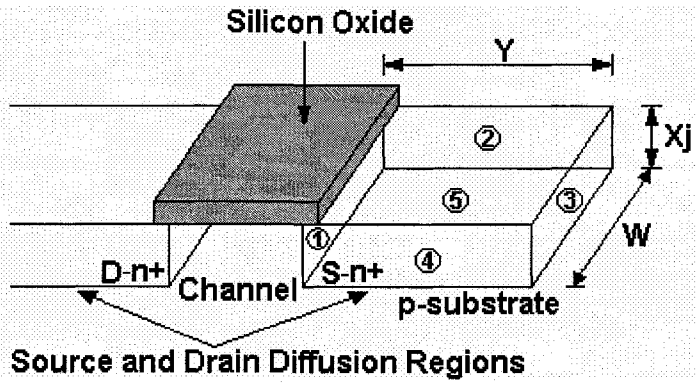


Figure 2.15: MOSFET cross-sectional and top views

Figure 2.15 (re-plotted from [18]) shows a 3D view of NMOS device. The drain n+ region, and the source n+ regions are involved in five planes of pn-junctions with the surrounding p-substrate of dimensions Y, W and X_j. Plane 1 is representing the conducting

channel and will not be considered in computing diffusion capacitance, $C_{Diffusion}$. This will leave us with four planers (planer 2 to planer 5). Bottom diffusion capacitance, $C_{d\text{-}bottom}$, is due to planer 5 and is given as:

$$C_{d-bottom} = C_j . Y . W \qquad (2.23)$$

where C_j is the junction capacitance per unit area, Y and W are the bottom dimensions. The planers 2, 3 and 4 are responsible for the side-wall diffusion capacitance, $C_{d\text{-}side\ wall}$, which can be written as:

$$C_{d-sidewall} = C_{SW} . X_j . (W + 2Y) = C_{j_{SW}} . (W + 2Y) \qquad (2.24)$$

X_j is the junction depth and it is a technology dependent parameter, it is added to the technology dependent C_{SW} perimeter capacitance to form a new technology dependent capacitance per unit perimeter, C_{jsw}. The diffusion capacitance now can be written as:

$$C_{Diffusion} = C_{d-bottom} + C_{d-sidewall} = C_j . Y . W + C_{j_{SW}} . (W + 2Y) \qquad (2.25)$$

2.6 MOSFET Delay

Figure 2.16 shows an NMOS device connected to a load capacitance, C_L. Let us assume that C_L is initially charged to V_{dd}. When the input changes from low (0) to High (V_{dd}), the NMOS device turns on and starts discharging the load capacitance.

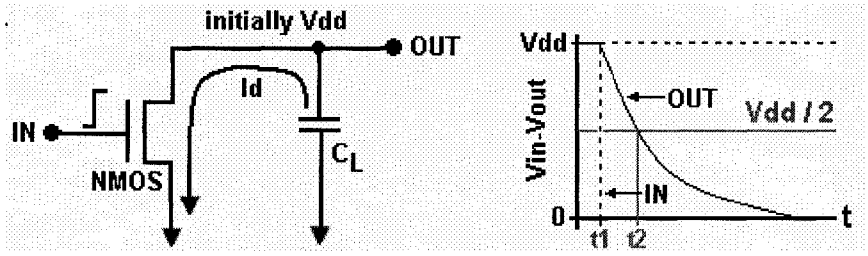


Figure 2.16: NMOS, discharging process

The capacitor current is proportional to the rate at which the voltage across the capacitor terminals changes with time, hence, we can write a mathematical relation between the applied voltage and the capacitor current as [18]:

$$i = C_L . \frac{dv}{dt} \qquad (2.26)$$

This is solved for dt and then integrated:

$$dt = C_L \cdot \frac{dv}{Id} \qquad (2.27)$$

$$\int_{t1}^{t2} dt = \frac{C_L}{I_d} \cdot \int_{V1}^{V2} dv \qquad (2.28)$$

$$(t2 - t1) = \frac{C_L}{I_d} \cdot (V2 - V1) \qquad (2.29)$$

For a step voltage input (infinite slope), (*t2* - *t1*) will be called the "delay time", the time required for C_L to lose half of its charge. If initially, the load capacitor was charged to V_{dd} and then this capacitor is fully discharged to 0 V, the "step delay" would be:

$$DS = \frac{C_L}{I_d} \cdot (V_{dd} - \frac{V_{dd}}{2}) \qquad (2.30)$$

The same conclusion can be reached when charging the load capacitor from an initial 0V to a final $V_{dd}/2$. Therefore, the step delay to charge or discharge the output capacitance, C_L, is [24]:

$$D_S = \frac{C_L . V_{dd} / 2}{I_d} \qquad (2.31)$$

Figure 2.17 shows where to measure the delay (*D*) for a ramp input signal (a) and for an ideal (step) input signal (b).

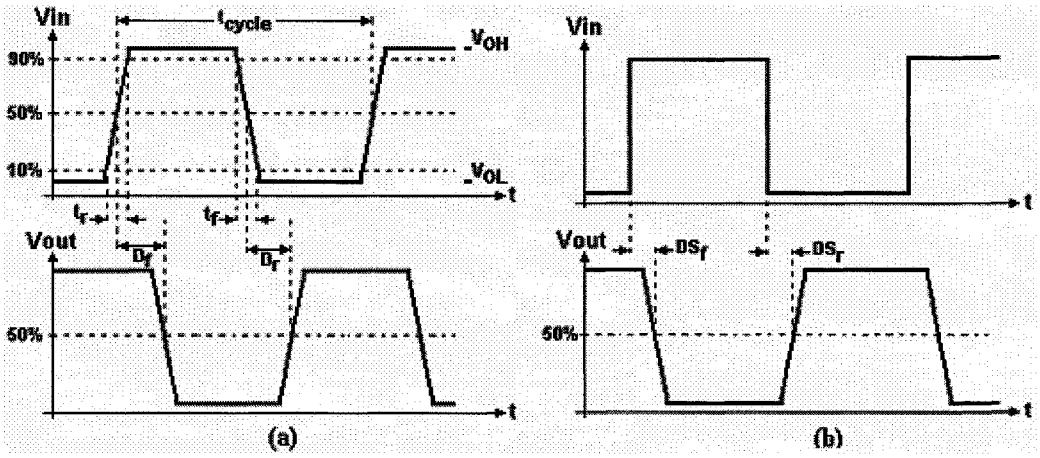


Figure 2.17: Propagation delay, ramp (a) and step (b)

In digital circuits, we either transfer logic 0 through the circuit or we transfer logic 1. Shams [24] stated that transferring logic 0 or logic 1 half-way through is taking place in saturation region of operation. The same conclusion was reached by [18] who also suggested adding a 10% error to the delay estimation.

2.7 MOSFET Saturation Drain Current

With five parameters, the Shockley model is widely used as drain current model of the conventional long-channel MOSFET when the following assumptions are true [25]:

- The vertical electric field due to the gate voltage is the dominant one and completely controls the channel charge density.
- Carrier mobility is constant.

This model does not consider any effects of dimensional or voltage scaling, and is not adequate for modern short-channel MOSFETs, with substantial deviations between predicted and simulated characteristics being observed (Figure 2.18).

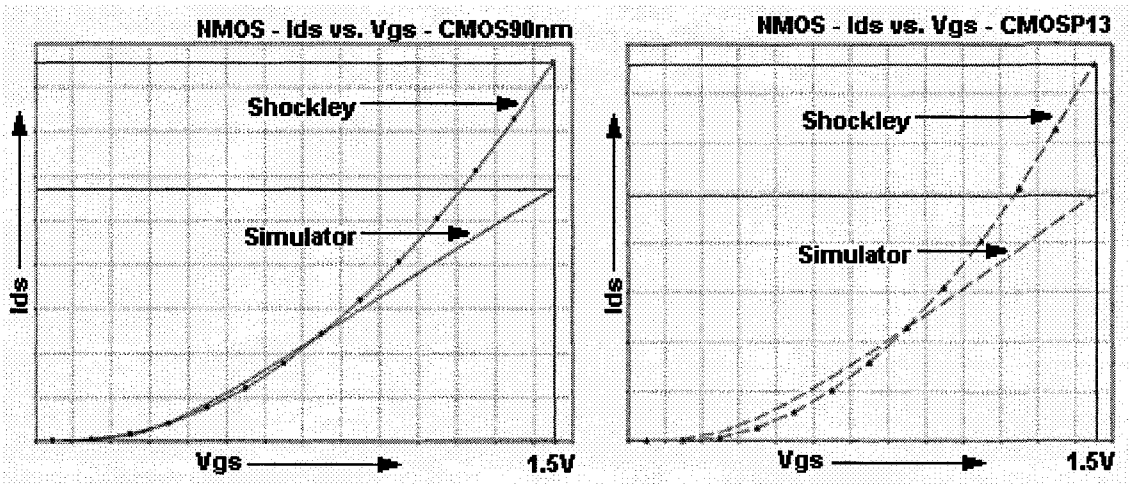


Figure 2.18: NMOS I_{ds} vs. V_{gs}, simulator and Shockley model

For the minimum device length in each technology with $W/L = 5$ and at $V_{dd} = 1.5V$, Figure 2.18 shows the I_{ds}-V_{gs} plot for an NMOS device in two deep sub-micrometer technologies, 90 nm and 0.13 μm. It is very obvious that the Shockley model is over estimating the drain current compared to the actual readings from the circuit simulator (Cadence-Spectre).

Started at the University of California in late 1960s and still under continuous development [26], SPICE is a simulation model that widely used for circuit simulation. This model, and so are the other professional commercial models like BSIM, employs a lot of parameters which complicates the computation and increases the simulation time. For example SPICE2 requires about 30 parameters [27] and BSIM3v3 requires over 200 [1].

This situation inspires researchers to devise efficient models with modest parameterization which will be acceptably accurate and sufficiently fast as device physical dimensions keep shrinking.

Working on NMOS devices of length from 1 μm to 3 μm and PMOS devices from 1.3 μm to 3.3 μm, K. Toh, P. Ko and R. Meyer [28] introduced an engineering model for short-channel MOS transistors, one including carrier velocity and mobility effects. Defining the drain saturated voltage, V_{dsat}, as the drain voltage at which the carrier velocity saturates, this model defines the linear region of operation as the region where $V_{ds} \leq V_{dsat}$ and the saturation operation region is where $V_{ds} \geq V_{dsat}$. These definitions are not quite accurate as they implement that for $V_{ds} = V_{dsat}$, the device operates in both regions, the linear and saturation. The drain saturation current is given as:

$$I_d = \upsilon_{sat} \cdot \frac{\varepsilon_{ox}}{t_{ox}} . W . (V_{gs} - V_t - V_{dsat}) \qquad (2.32)$$

$$V_{dsat} = (1 - \frac{1}{1 + \{(L - \Delta L) . \frac{2\upsilon_{sat}}{\mu_{eff}}\} / (V_{gs} - V_t)}) . (V_{gs} - V_t) \qquad (2.33)$$

where ΔL is the depletion width into the channel from the drain, υ_{sat} is the saturated velocity value and μ_{eff} is the effective carrier mobility. This model needs seven parameters to compute the drain current in saturation region of operation.

As an extension of Shockley's square-law model in the saturation region, T. Sakurai and R. Newton [29] introduced their α-power law model. This model focuses mainly on velocity saturation effects and gives the drain current, I_{ds} in the saturation region as:

$$I_d = \frac{W}{L} . Pc . (V_{gs} - V_t)^{\alpha} \qquad (2.34)$$

, where Pc is a technology dependent empirical parameter that includes the effect of the carrier mobility (μ) and the gate-oxide capacitance (C_{ox}). α is an index that varies from 2 (Square-law) to 1 as the carrier velocity saturation becomes severe; α is called velocity saturation index [29]. By taking two points on the I_d-V_{gs} plot as (V_{gs1}, I_{d1}) and (V_{gs2}, I_{d2}), α can be determined as:

$$\alpha = \frac{\log(I_{d1} / I_{d2})}{\log((V_{gs1} - v_t) / (V_{gs2} - v_t))} \qquad (2.35)$$

Pc then can be computed as:

$$Pc = \frac{I_{d2}.L}{W.(V_{gs2} - V_t)^{\alpha}} \qquad (2.36)$$

For 1 μm gate length device operated on 5 V supply, [29] claimed that α is 1.2 for NMOS device and 1.5 for PMOS device. With only four parameters, α-power law model is very handy for fast computation. It has been widely use in delay analysis work as a basic device current model For example, [30], [31] and [32] based their work on α-power law model.

Working on 0.5 μm technology, M. Shams [24] split the index α into two indexes. He wrote the drain saturation current as:

$$I_d = \kappa.W.(V_{gs} - V_t)^{(\xi + \frac{\upsilon}{V_{gs}})} \qquad (2.37)$$

ζ is an index sets to capture the velocity saturation while index υ is set to capture the mobility degradation. K is a technology and effective channel length dependent parameter.

With four parameters, this model needs three points on I_d-V_{gs} plot to find ζ, υ and K. For points (V_{gs1}, I_{ds1}), (V_{gs2}, I_{ds2}) and (V_{gs3}, I_{ds3}), ζ, υ and K can be computed as:

$$\xi = \frac{(X2.X6 - X1.X7)}{(X4.X6 + X4.X7 - X5.X6 - X3.X7)} \qquad (2.38)$$

$$\upsilon = \frac{(X1 - (\xi.(X3 - X4))}{X6} \qquad (2.39)$$

$$\kappa = \frac{I_{ds2}}{W.(V_{gs2} - V_t)^{(\xi + \frac{\upsilon}{V_{gs2}})}} \qquad (2.40)$$

where X1 to X7 are:

$$X1 = \log(\frac{I_{ds1}}{I_{ds2}}) \qquad (2.41)$$

$$X2 = \log(\frac{I_{ds2}}{I_{ds3}}) \qquad (2.42)$$

$$X3 = \log(V_{gs1} - V_t) \qquad (2.43)$$

$$X4 = \log(V_{gs2} - V_t) \qquad (2.44)$$

$$X5 = \log(V_{gs3} - V_t) \qquad (2.45)$$

$$X6 = \frac{X3}{V_{gs1}} - \frac{X4}{V_{gs2}} \qquad (2.46)$$

$$X6 = \frac{X4}{V_{gs2}} - \frac{X5}{V_{gs4}} \qquad (2.47)$$

Later, Shams modified this formula so that the mobility degradation index becomes V_{gs} and V_t dependent and the index would be unitless [33].

2.8 Summary

Three MOSFET development phases are summarized and brief historical review is given with numbers and charts illustrating the rapid evolution of VLSI. MOSFET basic operation is covered. Device features shrinking improves operation speed, reduces consumption power, and increase device per area ratio.

As the device shrinks, MOSFET traditional models are no longer giving accurate results. Physical effects that were ignored in long-channel technologies such as velocity saturation, mobility degradation and channel-length modulation become of great influence on short-channel technologies performance. These three effects are reviewed. We show that as the device channel-length keeps decreasing, the channel-length modulation keeps increasing.

The origin and description of MOSFET parasitic capacitances are overviewed. A brief explanation of their nature and effects with the formulas that can be used to express each one of these capacitances is given.

An intuitively-appealing introduction to MOSFET delay has been made. This will serve as a foundation when the matter of delay is discussed more thoroughly later in this work.

The understanding of MOSFET current modeling has evolved since Shockley proposed his "Square Power Law". Three notable improvements have been discussed briefly, those of Koh, To and Meyer, Sakurai and Newton, and Shams, in chronological order.

Chapter 3

MOSFET Current and Delay Modeling

Circuit speed is one of the most important performance parameters in VLSI design. The ability to accurately estimate this speed is an essential one when it comes time to compare the merits of one circuit alternative to those of another. However, estimation of critical path delay in large VLSI systems using commercial models such as BSIM or SPICE is prohibitively time consuming.

For accurate estimation, considering the current flows through the MOSFET is fatal. As the device features keep shrinking, using the square-law model for long-channel device is no longer giving a trustable results because it over estimating the current in short-channel technologies.

Alpha-power law [29] took care of the rising effects of velocity saturation in the short-channel devices. In this model, the drain current is proportional to $(V_{gs} - V_t)^{\alpha}$, where V_{gs} is the gate to source voltage, V_t is the threshold voltage, and alpha (α) is a carrier velocity saturation index. The short-channel device is also suffering from the mobility degradation due to the high transverse electric field. Shams included the mobility effects as V_{gs} dependent index in the formula he proposed in [24]. Later, he also modified the index to be threshold voltage dependent [33].

When choosing a delay estimation method, one must be mindful of an accuracy-speed tradeoff which gets worse as devices are scaled down [34].

Delay estimation is strongly affected by the input signal type. A ramped input signal will be subject to greater delay than that simulated for a step input, with the delay being proportional to the slope [35].

The importance of signal slope on the delay of a simple inverter of minimum length and W_P/W_N = $5L/3L$ driving a capacitance load of 100 fF as in the Figure 3.1 (a) is investigated. Figure 3.1 (b) Shows the transient response of the output according to three input signals, step, fast ramp and slow ramp signals. Input signal slope could be responsible for 30% of the gate delay [36].

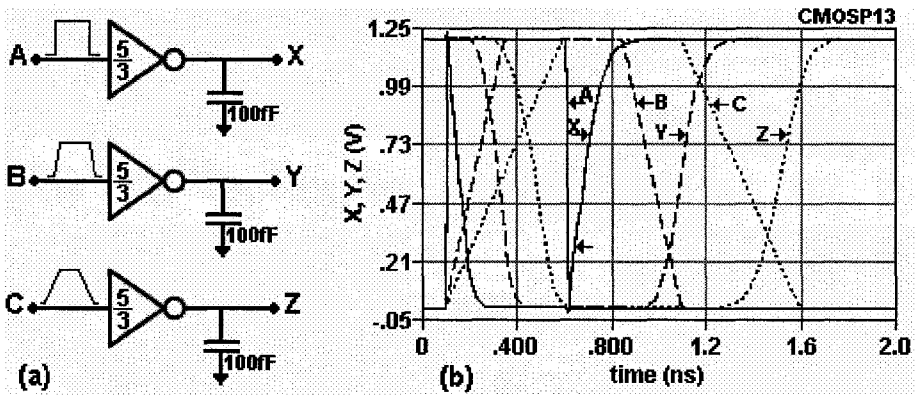


Figure 3.1: Effect of input signal slope on rise and fall delay

Table 3.1 tells that rise and fall delays are proportional to the input signal slope.

Table 3.1: Effect of input signal slope on inverter rise and fall delay

Input	*Output*	*Input Rise/Fall Time*	*Fall Delay*	*Rise Delay*
A	X	10.0 ps	52.85 ps	76.42 ps
B	Y	250 ps	98.94 ps	129.5 ps
C	Z	500 ps	127.4 ps	176.7 ps

The delay estimation through a chain of serially connected MOSFETs (SCM) is another factor that complicates the delay modeling. serially connected transistors can be found in gates and blocks like NAND, NOR, XOR, PLAs and arithmetic units. Using two degradation factors (X and Y) that count for the Load and internal capacitances due to the SCM, Shams [24] found that the Delay of transistors connected in series is proportional to a single transistor delay of the same load.

3.1 Short-Channel MOSFET Extended Saturation Current Model

Taking in to consideration the velocity saturation, mobility degradation, and channel-length modulation effects that discussed in chapter 2, we came to propose the following model to express the saturation drain current in deep sub-micrometer devices:

$$I_d = \kappa_t . W . (V_{gs} - V_t)^{(Z+\frac{M}{V_{gs}})} . (1 + \lambda . V_{ds}) \qquad (3.1)$$

κ_t is a parameter depending on the technology which is a function of the minimum channel length allowed by that technology. W is the device width, Z is a parameter which compensates for velocity saturation effects, and M one to compensate for mobility degradation effects (as a function of the vertical electric field, and in turn the gate voltage V_{gs}). Finally, λ is the channel-length modulation parameter. This formula contains five parameters compared to the three parameters Sakurai formula [29] and four parameters shams formula [24] as indicated in Table 3.2.

Table 3.2: Sakurai, Shams, and this work saturation current parameters

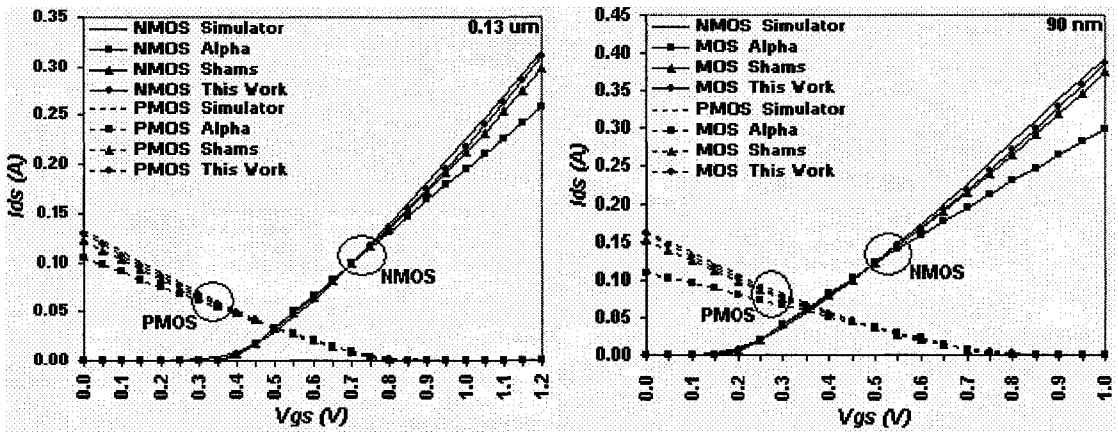
Feature	*Sakurai [29]*	*Shams [24]*	*This Work*
Technology Dependent Parameter (Including the device length at its minimum value for [24] and this work)	*Pc*	κ	κ_t
Device Length	-------	-----	-----
Device Width	*W*	*W*	*W*
Velocity Saturation	α	ξ	*Z*
Mobility degradation	-----	υ	*M*
Channel-length Modulation	-----	-----	λ
Total Number of Parameters	3	4	5

Table 3.3 summarizes Equation 3.1 parameters values (λ, Z, M and κ_t) for 0.13 μm and 90 nm technologies as extracted in Appendix A, Section A.1.

Table 3.3: Parameters for the proposed I_{ds} model

Technology	*NMOS*				*PMOS*			
	λ	*Z*	*M*	κ_t	λ	*Z*	*M*	κ_t
0.13um	0.163	1.46	- 0.15	515	0.370	1.83	- 0.27	262
90nm	0.448	1.56	- 0.11	735	0.491	1.97	0.30	541

Figure 3.2 shows plots of I_{ds} *vs.* V_{gs} as simulated by the Cadence simulator, the α-power law, the Shams law, and the model of this Thesis for NMOS and PMOS devices in 0.13 μm (a) and 90 nm (b) technologies.



(a) In 0.13 μm technology **(b) In 90 nm technology**

Figure 3.2: NMOS and PMOS I_{ds} vs. V_{gs}

For both of the 0.13 μm and 90 nm technologies, the predictions of the new I_{ds} model match closely those of the Cadence simulator. In each technology, the device length is set to the minimal length allowed by the technology and $W/L = 5$.

3.2 Delay Modeling Approaches

Several methods are used in the investigation of propagation delay. Their principal characteristics [37-40] are summarized here.

1. Numerical Approach: This approach provides an accurate simulation and verification in VLSI designing. An example of such approach is the use of BSIM model in Spice. In this approach, every circuit element is characterized by many parameters. Therefore, the analysis requires large hardware resources such as memory and it takes a lot of CPU computation time. This makes the numerical approach suitable to analyse smaller digital circuits.
2. Analytical Approach: The propagation delay is obtained from the transient response of the differential equations that describe the devices in the circuit. This approach is faster than the numerical approach as a detailed description of every single element in the circuit is not needed. It has the advantage that the gate delay can be calculated

by substitution of the used technology and the gate parameters. Several analytical models have been introduced, among them being the well-known and widely-used ones of Hendesterna and Jepson [41], and Sakurai and Newton [29, 42].

3. RC Delay: Linearization of non-linear circuits and approximation by means of RC networks are used to produce fair accuracy quickly. As the non-linear properties of the devices are neglected here, this approach only gives an average estimation of the circuit behavior.
4. Table Lookup Approach: This approach is based on tables of data that pre-simulated. The data in these tables include propagation delays as a function of technology parameters. Any change in the circuit devices parameters needs a re-calculated data. Preparing these tables is a memory and time consuming issue. This approach incorporates interpolation errors as well [43].

3.3 Effect of Input Signal Slope on Delay

Propagation delay is not only a technology and topology dependent matter, but it is also a function of the input signal slope.

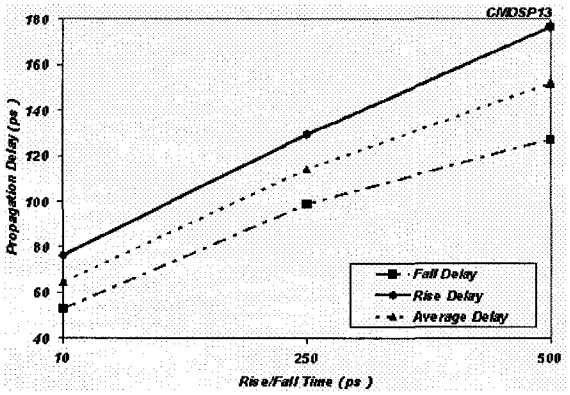


Figure 3.3: Simulated inverter delay as a function of input signal slope

Table 3.1 data are plotted as in Figure 3.3. When the slope is halved (from 250 ps to 500 ps), there is about 30% and 36% increase in the fall and rise delay respectively.

3.3.1 Hedenstierna and K. Jeppson Delay Model

Hedenstierna and Jeppson [41] proposed an analytical expression including the effect of input slope to compute gate delay. The propagation delay D, the time for the input level

of $V_{dd}/2$ to be transferred to an output level of $V_{dd}/2$, is approximated by:

$$D = D_S + S_\tau . \tau \quad (3.2)$$

where τ is the input signal transition time. D_S and S_x are the step delay and the slope factor respectively. They are given as:

$$D_S = \frac{C_L}{\beta . V_{dd} . (1 - V_t / V_{dd})} . [\frac{2 . V_t / V_{dd}}{1 - V_t / V_{dd}} + \ln\{4 . (1 - \frac{V_t}{V_{dd}}) - 1\}] \quad (3.3)$$

$$S_\tau = \frac{(1 + 2 V_t / V_{dd})}{V_{dd}} \quad (3.4)$$

C_L is the load capacitance and β is the transistor gain factor. Hedenstierna and K. Jeppson work is valid only when the input signal is three times faster than the output signal [41]. Their work is based on Shockley model, therefore the short-channel device effects are not taken in consideration.

3.3.2 Sakurai and Newton Delay Model

The α-power law model [29] proposed by Sakurai and Newton is one of the well-known MOSFET models. Based on this model, they extended the long-channel delay model to include some of the short-channel effects such as velocity saturation. Sakurai and Newton gave the rise or fall delay simulated at $V_{dd}/2$ of an inverter driven by a ramp input signal as:

$$D = \frac{C_L . V_{dd}}{2 . I_d} + (\frac{1}{2} - \frac{1 - V_t / V_{dd}}{1 + \alpha}) . \tau \quad (3.5)$$

where I_d is the drain current at $V_{gs} = V_{ds} = V_{dd}$. Comparison of Equation 3.5 and Equation 3.2 tells that S_x is:

$$S_\tau = \frac{1}{2} - \frac{1 - V_t / V_{dd}}{1 + \alpha} \quad (3.6)$$

α in Equation 3.6 can be driven from Equations 3.2, 3.5 and 3.6 as following:

$$S_\tau = \frac{1}{2} - \frac{1 - V_t / V_{dd}}{1 + \alpha} = \frac{D - D_S}{\tau} \quad (3.7)$$

$$\frac{1 - Vt / Vdd}{1 + \alpha} = \frac{1}{2} - \frac{D - D_S}{\tau} \quad (3.8)$$

Solving for (1+ α) gives:

$$(1 + \alpha) = (1 + \frac{V_t}{V_{dd}}) . (\frac{2\tau}{\tau - 2 . (D - D_S)}) \quad (3.9)$$

and then for α itself:

$$(1+\alpha) = (1+\frac{V_t}{V_{dd}}) \div (\frac{\tau - 2.(D - D_S)}{2.\tau}) \qquad (3.10)$$

$$\alpha = [(1+\frac{V_t}{V_{dd}}) \div (\frac{\tau - 2.(D - D_S)}{2.\tau})] - 1 \qquad (3.11)$$

After D_S and D are simulated results for both step and ramp inputs (transition time τ), α is calculated using Equation 3.11.

3.4 Extending α and the Slope Factor

Equation 3.11 is giving α as a general value that can be used for both of the devices, NMOS and PMOS. Depending on this α value, Equation 3.6 is estimating the slope that is required to calculate the delay according to Equation 3.5.

For the gates connected in series as in Figure 3.4, Shams [24] stated that the step delay is approximately half of the corresponding transition time τ ($\tau = 2.D_S$). Gate G_i delay can be expressed by its step delay and the driving gate G_{i-1} step delay as:

$$D_i = D_{Si} + S.D_{S(i-1)} \qquad (3.12)$$

S is the slope factor and equal to $2S_\tau$.

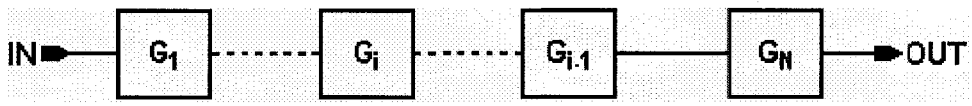


Figure 3.4: Cascaded gates

On examining Table 3.1, one sees that the rising delay is higher than the falling delay for the three input signals having different slopes. The use of this observation in calculations will be possible if two versions of the slope factor are defined, namely S_N and S_P. Doing so will require that two corresponding values for α are similarly defined – α_N and α_P and hence the slope factors S_N and S_P. α_N and S_N are to be used when calculating the fall delay through NMOS device and α_P and S_P are used to calculate the rising delay through PMOS device.

Furthermore, we can expand α_N, S_N, α_P and S_P to cover the rising and falling delay calculation for each device. The new four α and S values are α_{RP}, S_{RP}, α_{FP} and S_{FP} for

PMOS device and α_{RN}, S_{RN}, α_{FN} and S_{FN} for NMOS device [24]. The procedure for extracting these parameters from measurement data will be described in Appendix A, after the inclusion in the delay model of the chains of series-connected MOSFETs found in complex gates has been studied.

3.5 Series Connected Transistors

The series-connected MOSFET (SCM) structure appears frequently in VLSI designs. SCM complicates the delay estimation and needs a careful modeling to handle it. One of the methods used to work on the SCMS is to collapse the gate of SCMS into equivalent inverters [44, 45].

For N series-connected MOSFETs, Shams [24] found that the delay is proportional to the delay that a single transistor shows when loaded by the same load as the SCM. Using two degradation factors, X and Y, he gave the step delay, D_S, of SCM as:

$$D_S = \frac{v}{W}.(X.C_d.W + Y.C_L) \qquad (3.13)$$

where v is the MOSFET resistance times unit width, X is delay degradation factor for load capacitance due to series connected MOSFETs, Y is delay degradation factor for internal capacitances due to series connected MOSFETs, C_d is the diffusion capacitance and W is the device width. v, X, and Y are written as:

$$v = \frac{V_{dd}.W}{2.I_d} = \frac{V_{dd}}{2.\kappa_t.(V_{gs} - V_t)^{(Z+\frac{M}{V_{gs}})}.(1+\lambda.V_{ds})} \qquad (3.14)$$

$$X = \delta.(N^2 - 1) + 1 \qquad (3.15)$$

$$Y = (x - y.\frac{V_{dd}}{V_t}).(N - 1) + 1 \qquad (3.16)$$

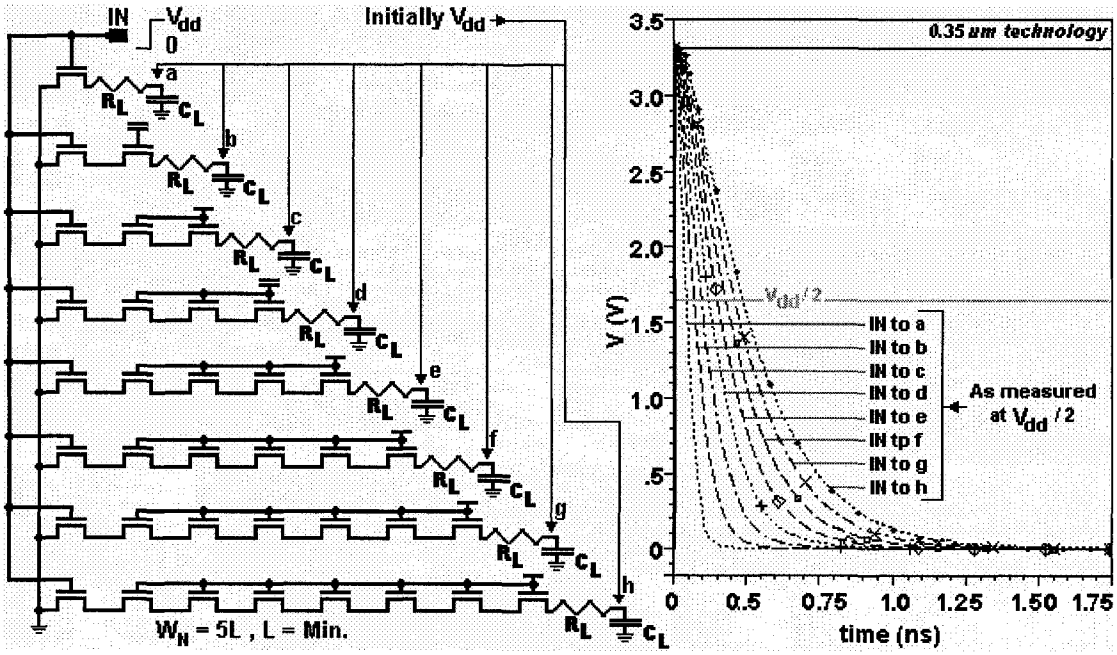
where δ is the MOSFET resistance in linear to its resistance in saturation, x and y are empirical values that we will extract in Appendix A. When the load capacitance is dominating, X can be set to zero [24] and Equation 3.13 is written as:

$$D_S \cong Y.\frac{v}{W}.C_L \qquad (3.17)$$

3.5.1 Channel-Length Effects on Delay in SCM Structure

Figure 3.5 (a) shows a test bench that used to study the effect of channel-length shrinking on the delay of serially connected MOSFET (NMOS with $W_N = 5L$ for minimum L) in four different technologies (0.35 μm, 0.18 μm, 0.13 μm, and 90 nm).

The worst-case condition is considered when all internal nodes in the chain were pre-charged to V_{dd}. Eight chains with 1, 2, 3, 4, 5, 6, 7, and 8 serially connected MOSFETs were investigated. The delay to discharge the load capacitance is simulated in each chain at $V_{dd}/2$ when a step input signal is applied. Figure 3.5 (b) shows the discharging time response of each serial connected MOSFET chain in 0.35 μm technology.



(a) (b)

Figure 3.5: Serially connected MOSFET test bench (a), and C_L discharging time response in 0.35 μm technology

The load in each technology has been chosen to represent 5 μm of an M1 metal interconnect that has a minimum width. The capacitance part of the load is obtained by extracting this piece of interconnect while the resistance part is obtained from the technology documentation per unit length for minimum width. Table 3.4 shows the resistance and the capacitance that used as a load in each technology.

Table 3.4: Resistance and capacitance loads values used in test bench (Figure 3.5 (a)) per μm length of minimum width interconnect in four technologies (M1 metal)

Technology	*Resistance*		*Capacitance*	
	(Ω/□)	*(Ω/5μm)*	*(fF/μm)*	*(fF/5μm)*
0.35 μm	0.0750	0.75	0.215	1.08
0.18 μm	0.0780	1.70	0.224	1.00
0.13 μm	0.0940	2.94	0.129	0.65
90 nm	0.0950	3.96	0.103	0.52

The delay measurements versus the number of MOSFETs in the SCM chain were plotted in Figure 3.6 (a). One can observes that for a longer channel technology (0.35 μm), there is a non-linear relation between the delay and the devices number in the SCM chain. This non-linearity decreases as we move to a shorter length technology as could be observed in 90 nm technology.

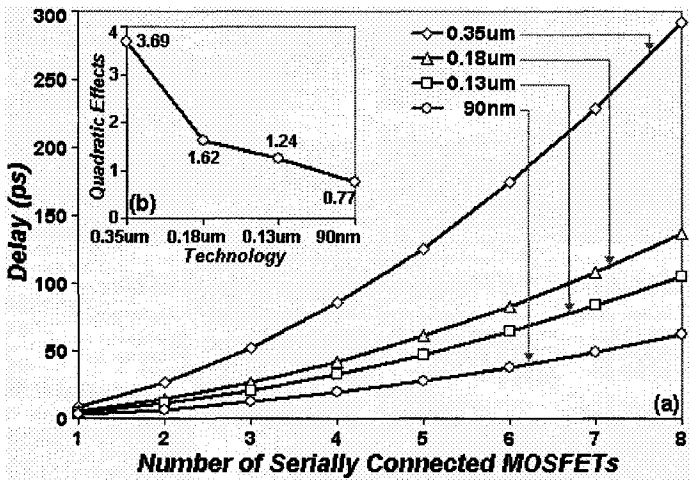


Figure 3.6: The delay through a serially connected transistors chain vs. the number of them in the chain in four technologies (a) and quadratic effects vs. technology (b)

To justify this observation, the series delay as a function of its devices number connected in series could be written as:

$$D = A.N^2 + B.N + C \qquad (3.18)$$

where A is the quadratic effects, B is linear effects, C is a constant and N is the number of serially connected MOSFETs. A, the quadratic effects, can be calculated as:

$$A = \frac{3.D8 - 7.D4 + 4.D1}{84} \qquad (3.19)$$

where *D8*, *D4*, and *D1* are the simulated delay for a chain of 8, 4, and 1 MOSFETs respectively. *A* is plotted vs. technology, Figure 3.6 (b) and we can see that as we move to deeper technology, *A* decreases.

3.6 Delay Parameters Extraction

In sections A.1.1 and A.1.2, the parameters of the extended saturation current model were extracted from the results of $I_d - V_{ds}$ and $I_d - V_{gs}$ simulations. Equation 2.31 gives the step delay as a function of I_d. In Appendix A, Section A.2 gives the procedures of extracting Equation 3.1 parameters based on the delay results from transient response simulation. The other delay parameters mentioned in sections 3.4 and 3.5 will be extracted in Appendix A, Sections A.3 and A.4.

In Figure 3.7 (a) and (b), NMOS and PMOS devices are connected together in two configurations to simulate for the rising and falling step delay. Four delays will be simulated, $D_{S\text{-}NF}$, $D_{S\text{-}NR}$, $D_{S\text{-}PF}$, and $D_{S\text{-}PF}$ as seen in Figure 3.7 (c) and (d) according to step input and according to ramp input.

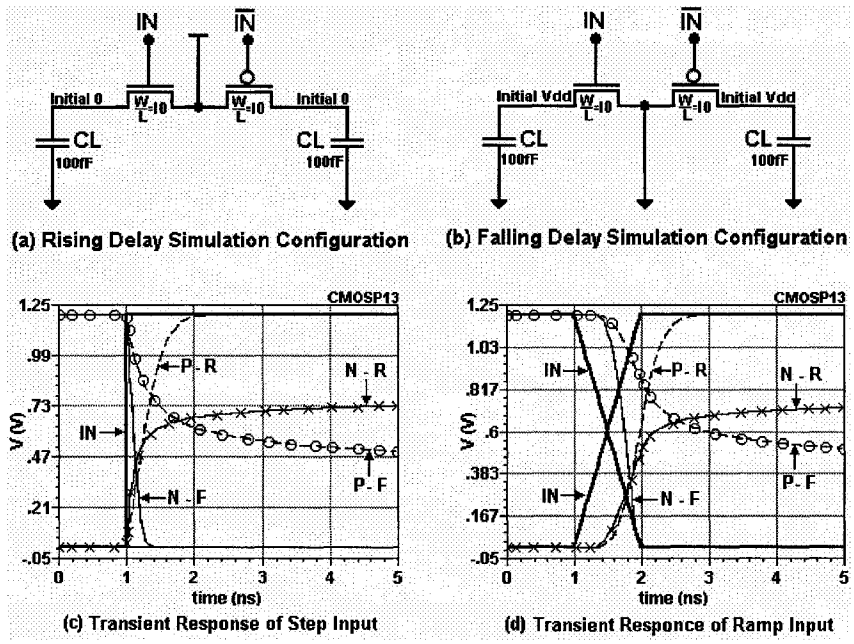


Figure 3.7: Delay parameters extraction test bench based on step and ramp signals

A complete summary of the current-model parameters computed based on delay in Section A.2 is presented in Table 3.5.

Table 3.5: Proposed current parameters as extracted depending on delay readings in 0.13 μm and 90 nm technologies

Technology and Transition		*NMOS*					*PMOS*				
		λ	Z	M	κ_t	v	λ	Z	M	κ_t	v
0.13 μm	*Fall*	0.171	1.073	0.154	506	0.0014	0.263	2.067	1.120	63	0.0153
	Rise	0.209	1.606	1.120	155	0.0064	0.381	1.258	-0.012	166	0.0034
90 nm	*Fall*	0.465	1.012	0.156	627	0.0010	0.517	1.564	0.649	118	0.0067
	Rise	0.477	1.526	0.476	275	0.0030	0.503	1.313	-0.028	241	0.0023

Note that negative M does not have physical interpretation. However, since the negative values are small, they are effectively zero. In these cases, the model is actually reduced to that of Sakurai's. Figure 3.8 shows the rise and fall step delay of NMOS and PMOS devices as function of the device width for the nominal V_{dd} and minimum length in 0.13 μm technology as simulated and according to the model. A close matching can be observed.

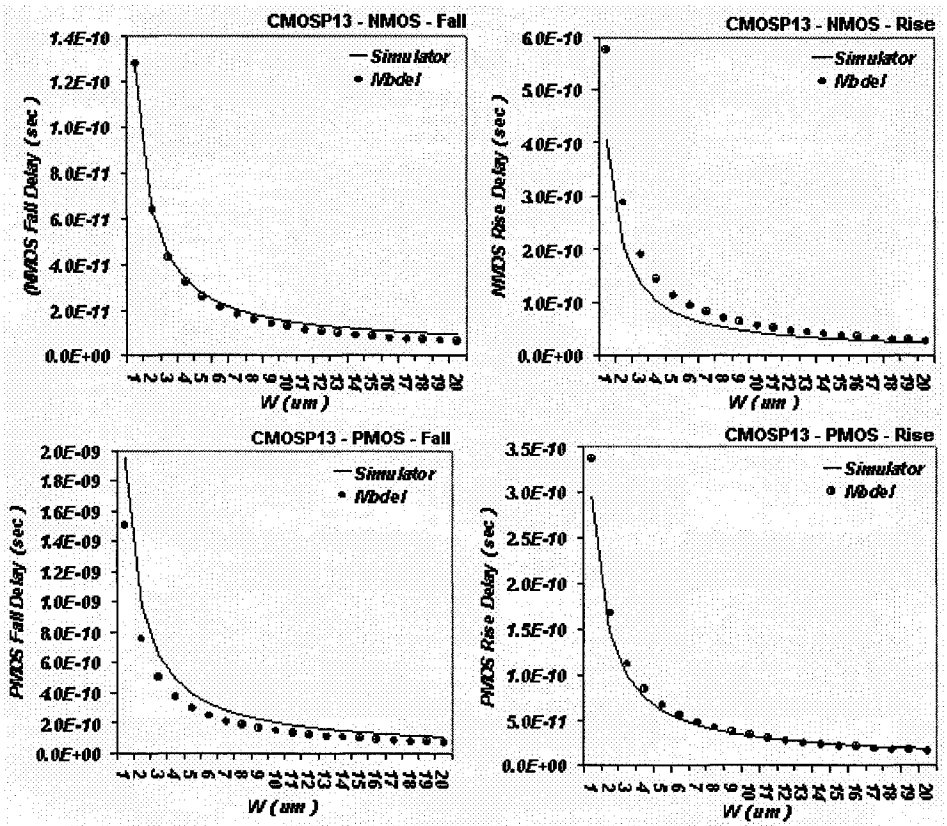


Figure 3.8: NMOS and PMOS step rising and falling delay vs. device width in 0.13 μm

Table 3.6 gives α, S, x, and y as extracted in Appendix A, section A3 and A.4 from delay for NMOS rise and fall, PMOS rise and fall in two technologies, 0.13 μm and 90 nm.

Table 3.6: Extracted slope factors and series-connection parameters based on delay

Technology and Transition		NMOS				PMOS			
		α	S	x	y	α	S	x	y
0.13 μm	Fall	0.834	0.316	1.879	0.359	1.941	0.558	1.477	0.010
	Rise	2.415	0.633	1.762	0.107	1.335	0.443	1.706	0.201
90 nm	Fall	0.787	0.309	2.039	0.502	1.504	0.507	1.560	0.063
	Rise	2.343	0.631	2.428	0.382	1.017	0.388	1.678	0.227

3.7 Input Signal Second Order Slope Effects

Gate delay is approximately linearly dependent on input transition time [1]. So far in this work, the input signal slope factor has been treated as a 1st order effect, which is a good approximation, but one that can be improved upon. In this section we attempt to do just that, and develop a 2nd order expression for the slope factor. The obvious starting point is the re-writing of Equation 3.12:

$$D_i = D_{Si} \cdot \{1 + S1 \cdot \frac{D_{i-1}}{D_{Si}} + S2 \cdot (\frac{D_{i-1}}{D_{Si}})^2\} \qquad (3.20)$$

where D_i, D_{Si}, D_{i-1}, $S1$, and $S2$ are Gate G_i propagation and step delay, Gate G_{i-1} propagation delay, input signal 1st and 2nd order slope effects. $S1$ and $S2$ are extracted as in Appendix A, Section A.5, The results for 90 nm and 0.13 μm technologies are listed in Table 3.7.

Table 3.7: Extracted 2nd order slope factors in 90 nm and 0.13 μm technologies

Technology and Transition	Rising		Falling	
	$S1$	$S2$	$S1$	$S2$
90 nm	0.318	0.0008	0.401	0.0020
0.13 μm	0.250	0.0010	0.368	0.0067

Figure 3.9 shows that as the device channel length shrinks, the first order slope effect increases and the second order one decreases.

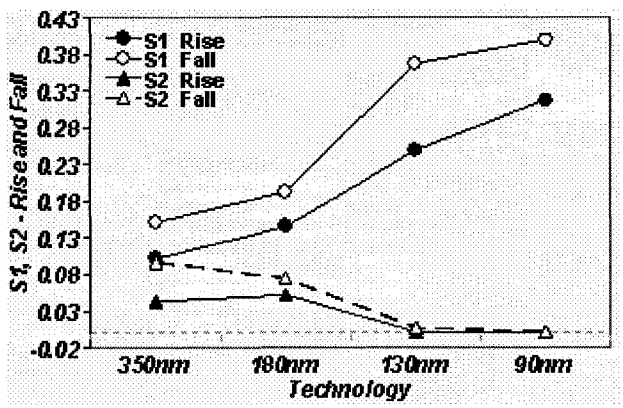


Figure 3.9: Input Signal 2nd order slope factors as a function of technology

3.8 Summary

An extended saturation current model for deep sub-micrometer technologies, which uses only five parameters, has been developed. The model parameters can be extracted in two ways, the first one depending on the I_{ds}-V_{ds} and $I_{ds} - V_{gs}$ simulation results. The second is based on the transient response of NMOS and PMOS delay simulation according to step and ramp input.

The effect of SCMS is taken in consideration when calculating the parameters based on delay. Delay through serially connected transistors does not only depend on the number of these transistors in the series, but also on the technology used. A linear dependency seems to be true as device size shrinks. Input slope factors S and the associated empirical factor α are also calculated from the delay simulation and so are the series-connection parameters x, and y.

The model shows very close results comparing to the actual circuit simulator results. The slope factor (S) is extended to include the 2nd order effects.

Chapter 4

MOSFET Capacitance Modeling

Accurate modeling of MOSFET capacitance is as important as the treatment of their DC characteristics. Charge in a MOSFET is stored on the gate electrode, the conducting channel and in the depletion layers. These structures can be regarded as capacitor electrodes.

The MOSFET has different parasitic capacitances that need to be taken into account when simulating a circuit, ones which can be classified into extrinsic and intrinsic capacitances. Extrinsic capacitance is the result of overlaps in the physical structure of the device and is due to the lateral diffusion which forms the gate-source and gate-drain overlap capacitances. Intrinsic capacitances are voltage dependent and are measured between the device terminals.

A MOSFET capacitance may be one of several types; it may be of a distributed structure, non-linear and may often be dependent on applied voltage. The precise modeling of capacitances such as these is a complicated matter, but usually they can be approximated by (constant) equivalent capacitances, an approach which is adequate for delay modeling and estimation purposes. Since the switching speed of a MOS digital circuit is determined by the time required to charge or discharge internal device capacitances, a delay model's accuracy will be dependent upon the accuracy of the MOSFET capacitance model used.

Accurate delay modeling does not depend solely on the accuracy of MOSFET capacitance approximations, but also on interconnect capacitances [18], the capacitances of the metal traces connecting the various devices, gates and blocks in an integrated circuit.

As the technology keeps shrinking, interconnect resistance and coupling capacitance increase due to smaller wire width and smaller wire spacing [40]. This leads to higher interconnect delay.

4.1 Toward Capacitance Modeling

MOS digital circuit speed is controlled by the time that this circuit needs to charge or discharge its internal nodes capacitances. MOSFET capacitances are computed as the product of a capacitance having units of capacitance per unit length (femto-Farad per micrometer), with the device's width (micrometer). It is difficult to measure these capacitances directly, and one must trust theory and calculation. In Section 2.5, seven important sources of MOS capacitance were described. Figure 4.1 is a detailed diagram of the capacitances and their locations in the device.

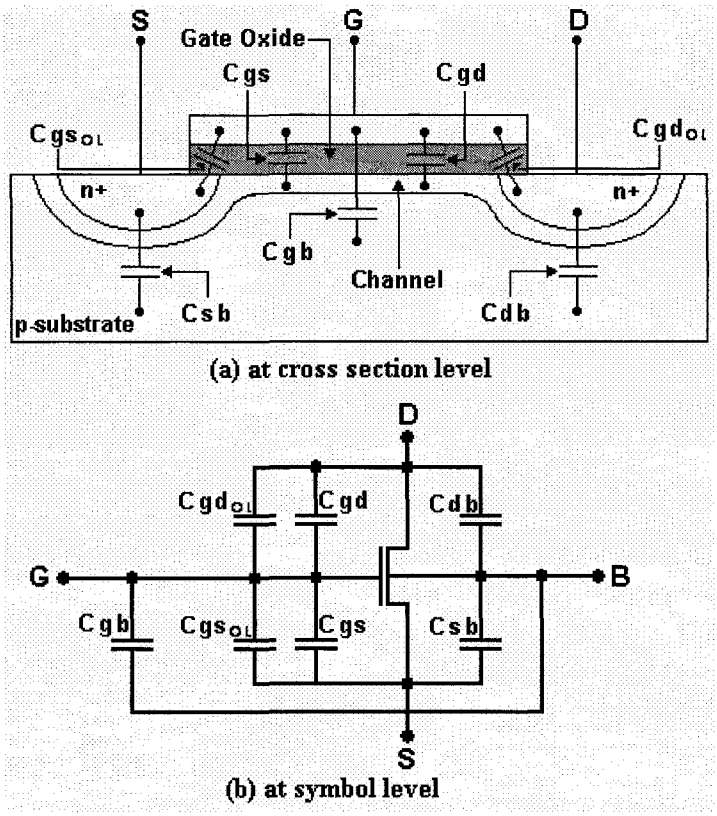


Figure 4.1: NFET capacitances; cross-sectional view and symbolic

These capacitances are non-linear and can be classified as following [20]:

1 - C_{db}, C_{sb}, and C_{gd}: Non-linear capacitances which are V_{db}, V_{sb}, and V_{gb} dependent respectively.

2 - C_{gd}, C_{gs}: Non-linear capacitances (through the channel) which are V_{gs}, V_{gd} and V_{gb} dependent respectively.

3 - C_{gdOL}, C_{gsOL}: Over-lap capacitances which are voltage-independent.

In Figure 4.2, two identical inverters are connected in series and the delay from IN to node X is to be estimated.

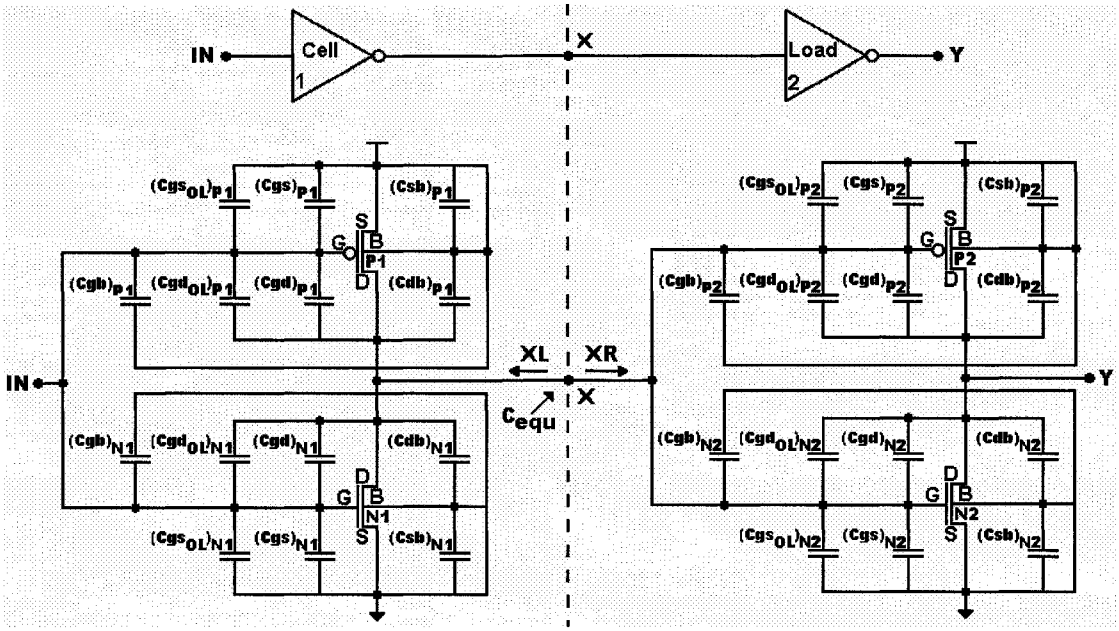


Figure 4.2: Two inverters in series

The propagation delay is simulated at $V_{dd}/2$. [24] and [18] stated that transferring logic 0 or logic 1 half-way through is taking place in saturation region of operation. Under this assumption, all C_{gb} and all C_{gd} capacitances can be set to zero as indicated in Table 2.2.

The capacitance seen at node X has contributions from the circuitry on either side. The capacitance contributions from the right of X (XR) are those associated with the load's gate:

$$C_{XR} = \sum C_{Gates} \qquad (4.1)$$

Each of the devices P2 and N2 makes three contributions to the sum, which can be expanded to:

$$C_{XR} = (C_{gsOL} + C_{gdOL})_{P2} + (C_{gsOL} + C_{gdOL})_{N2} + (C_{gs})_{P2} + (C_{gs})_{N2} \quad (4.2)$$

Each of the terms on Equation 4.2 is in turn expanded using the formulae summarized in Table 2.2:

$$C_{XR} = 2.C_{OL}.W_P + 2.C_{OL}.W_N + \frac{2}{3}.C_{ox}.L.W_P + \frac{2}{3}.C_{ox}.L.W_N$$

$$= (2.C_{OL} + \frac{2}{3}.C_{ox}.L).(W_N + W_P) \quad (4.3)$$

Turning to the XL direction, it is seen immediately that all source-related capacitances of both the P1 and N1 devices can be neglected as their terminals are either all connected to V_{dd} or to ground. Furthermore, P1 and N1 operate in either the saturation or cut-off regions and so all C_{gd} and C_{gb} capacitances of both devices equal to zero. Hence, to the left of node X there are but four capacitances that must be taken into account: $(C_{gdOL})_{P1}$, $(C_{gdOL})_{N1}$, $(C_{db})_{P1}$, and $(C_{db})_{N1}$. The part of the capacitance seen at Node X that is due to circuitry XL direction is:

$$C_{XL} = (C_{gdOL})_{P1} + (C_{gdOL})_{N1} + (C_{db})_{P1} + (C_{db})_{N1} \quad (4.4)$$

In what is called the Miller Effect, the two overlap capacitances $(C_{gdOL})_{P1}$ and $(C_{gdOL})_{N1}$ allow a current to flow directly from the input to the output during rising/falling transitions. Weste and Eshraghian [35] stated that this effect is hardly ever important in digital circuits but of major importance in analog circuits. We will follow Rabaey *et al* [1] in compensating for the Miller Effect by doubling the overlap capacitances between the gate and the drain (C_{gdOL}). Doing this, results in a slight modification to equation 4.4:

$$C_{XL} = 2.C_{OL}.W_P + 2C_{OL}.W_N + (C_{db})_{P1} + (C_{db})_{N1} \quad (4.5)$$

The diffusion capacitance, C_b, is given by Equation 2.25 which can in this case be simplified to:

$$C_d = C_j.Y.W + C_{jSW}.W + 2.C_{jSW}.Y = W.(C_j.Y + C_{jSw}) + 2.C_{jSW}.Y \quad (4.6)$$

C_d has been written as a sum of two terms. If the relatively small second one is neglected, what remains is a width-dependent formula, and if the further approximation that $C_{jSW} << Y.C_j$ is made, a formula simple enough for hand calculation results:

$$C_d = C_j.W \quad (4.7)$$

Substitution of this into Equation 2.5 gives

$$C_{XL} = 2C_{OL}.(W_P + W_N) + C_j(W_P + W_N) = (2C_{OL} + C_j).(W_P + W_N) \quad (4.8)$$

Let C_{equ} now denote the capacitance at node X:

$$C_{equ} = C_{XR} + C_{XL} = (2C_{OL} + \frac{2}{3}.C_{ox}.L).(W_P + W_N) + (2C_{OL} + C_j).(W_P + W_N)$$

$$= (W_P + W_N).(C_j + 4C_{OL} + \frac{2}{3}.C_{ox}.L) \quad (4.9)$$

After the replacement of all parasitic capacitances to the right and left of X with C_{equ}, a simplified equivalent inverter results (Figure 4.3).

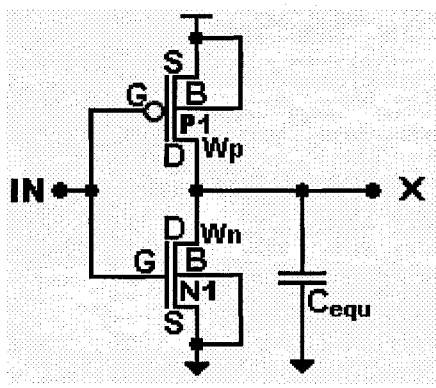


Figure 4.3: The cell inverter with the equivalent devices parasitic capacitances

A test of the validity of the approximations leading to Figure 4.3 can be made by comparing the rising and falling delays resulting from the C_{equ} model with the same ones computed by simulation. In the case of 0.13 μm technology, C_{OL} and C_j are estimated to be 0.25fF/μm and 0.80fF/μm respectively and t_{ox} is 2.7nm [47]. When $W_P = 2.5W_N = 5$ μm, Spectre simulation calculates a falling delay of 13.36 ps. Equation 2.31 is then used to compute the falling step delay. The load capacitance C_L appearing in this equation can be calculated with Equation 4.9 and I_d with Equation 3.1, for the parameters shown in Table 3.3. The final result is a falling step delay of $D_{Sf} = 12.18$ ps for an identical device. This calculated delay is less than the simulated one by 9.7%, mainly because the interconnect capacitances were not included. This is addressed in the next section.

4.2 Interconnect Capacitance

When long-channel devices are used, signal propagation delay is roughly proportional to device capacitances, and the interconnection capacitance (C_{wire}) can be disregarded. But

this approximation becomes less and less accurate as device sizes diminish. C_{wire} comes to have significant effects not only on delay, but on reliability and signal integrity as well [20].

The move to deep sub-micrometer technologies is also accompanied by increasing interconnection influence. With these processes, C_{wire} acquires a dependence on geometry in the form of interconnect shape, and the separation between signal paths and the substrate and neighboring interconnections [1].

Interconnect capacitance in general is three dimensional by nature and the expense of precise calculations makes these unsuitable for routine VLSI work [36]. Figure 4.4 is a simple 3D illustration of the capacitances that might be associated with one conductor in an array of interconnect. Multilevel interconnection arrays are normally used in VLSI design.

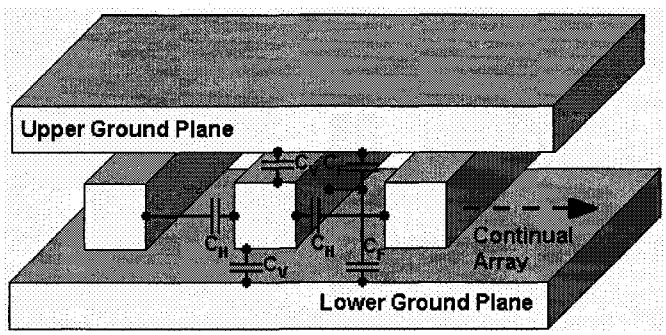


Figure 4.4: Interconnect capacitances in simple 3D view

Three capacitances are associated with an interconnection trace [18]. They are:

1 - C_H: Horizontal capacitance, existing between a conductor and its right and left neighbours.

2 - C_V: Vertical capacitances, existing between a conductor and each of the upper and lower ground planes.

3 - C_F: Fringing capacitance, existing between different edges and surfaces.

We will assume a unit (1 μm) length when computing C_V (the electrodes of which are M1 and the ground plane; the dielectric is the insulation.). The insulator is silicon oxide (dielectric constant ε_o = 3.9; permittivity ε_{ox}, = 88.5x10 $^{-4}$ fF/μm.). Using a standard equation of electrostatic theory:

$$C_V = \varepsilon_o . \varepsilon_{ox} . \frac{W}{H} \qquad (4.10)$$

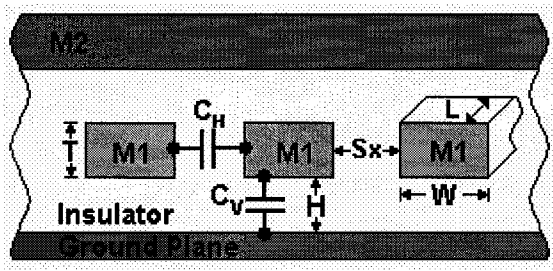


Figure 4.5: Vertical and horizontal capacitance interconnect component

The horizontal capacitance is that of two neighbouring conductors of height T (at the same level) with separation Sx:

$$C_H = \varepsilon_o.\varepsilon_{ox}.\frac{T}{Sx} \qquad (4.11)$$

The fringing capacitance is ([48]):

$$C_F = \frac{\varepsilon_o.\varepsilon_{ox}}{\pi}.ln(1+\frac{2.T}{H}+2.\sqrt{\frac{T}{H}+\frac{T^2}{H^2}}) \qquad (4.12)$$

Finally, the total interconnect capacitance is:

$$C_{Wire} = \varepsilon_o.\varepsilon_{ox}.[\frac{W}{H}+\frac{T}{Sx}+\frac{1}{\pi}.\ln(1+\frac{2.T}{H}+2.\sqrt{\frac{T}{H}+\frac{T^2}{H^2}})] \qquad (4.13)$$

The conductor and insulation thicknesses, T and H respectively, are fixed when the technology is chosen and cannot be altered by the designer. Also fixed by the choice of technology is the minimum conductor width, W, and the minimum spacing between them, S_x. Their values are designer-adjustable upward from these minimums

Table 4.1 gives the values of these parameters for 0.13 μm and 90 nm technologies [49, 50] for Metal 1 (M1) conductors running between ground planes, and Metal 2 (M2) as seen in Figure 4.4.

Table 4.1: W, Sx, T, and H parameters and the worst case interconnect capacitances in 0.13 μm and 90 nm technologies

Technology	*0.13μm*				*90nm*			
Parameter (μm)	W_{min}	Sx_{min}	T	H	W_{min}	Sx_{min}	T	H
	0.160	0.160	0.300	0.600	0.160	0.180	0.26	0.600
C_{Wire}	0.089 fF/μm				0.073 fF/μm			

If Figure 4.2 is re-drawn with C_{wire} included, one gets Figure 4.6 where C_T is the sum of the interconnect capacitance and the inverters total parasitic capacitances at node X.

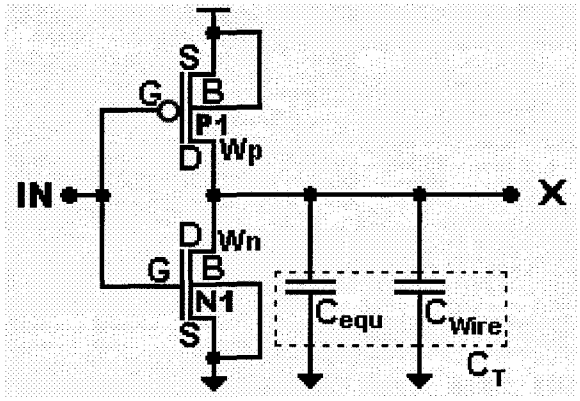


Figure 4.6: The cell inverter with total parasitic and interconnect capacitances

4.3 MOSFET Capacitance Extraction

In this section, we will extract MOSFET capacitances at five classes of detail. The simplest of these is that of Class 1 where a single MOSFET capacitance applies to either NMOS or PMOS devices. The most detailed MOSFET capacitance specification is that of Class 5. Here each device is separately characterised by four capacitance parameters, for a total of eight in all [24]. For each device type (NMOS or PMOS), Class five offers two gate and two diffusion capacitances, one for each of the rising and falling transition directions. The remaining intermediate classes of capacitance specification are, briefly:

- Class 2: proposes two capacitance parameters, one for each of the device types.
- Class 3: also proposes two capacitance parameters. Here, the subdivision is into gate and diffusion parameters, with no distinction between device types being made.
- Class 4: is like Class-3, except that device types are distinguished, doubling the number of capacitance parameters to four. Gate and diffusion capacitances are considered.

The full procedures of extracting the MOSFET capacitance in each class is fully described in Appendix B over five sections. Each section is assigned for specific MOSFET capacitance class extraction.

Table 4.2 gives the MOSFET capacitance values (in fF/μm) that were the results of the parameter extractions for the five classes of description. The full extraction procedures can be found in Appendix B, Sections B.1 to B.5.

Table 4.2: Class 1 to Class 5 extracted MOSFET capacitances

Class 1 MOSFET Capacitances					
Tech.	C_{mosfet} *(fF/μm)*				
90 nm	1.79				
0.13 μm	2.59				
Class 2 MOSFET Capacitances					
Tech.	C_P *(fF/μm)*		C_N *(fF/μm)*		
90 nm	1.22		2.94		
0.13 μm	2.20		3.36		
Class 3 MOSFET Capacitances					
Tech.	C_g *(fF/μm)*		C_d *(fF/μm)*		
90 nm	1.96		1.62		
0.13 μm	1.93		3.24		
Class 4 MOSFET Capacitances					
Tech.	*NMOS (fF/μm)*		*PMOS (fF/μm)*		
	C_g	C_d	C_g	C_d	
90 nm	2.07	1.68	1.84	0.60	
0.13 μm	2.08	2.90	2.08	2.31	
Class 5 MOSFET Capacitances					
Tech.	*Transition*	*NMOS (fF/μm)*		*PMOS (fF/μm)*	
		C_g	C_d	C_g	C_d
90 nm	*Fall*	1.01	2.86	0.68	1.73
	Rise	0.86	2.33	0.91	1.72
0.13 μm	*Fall*	1.31	2.78	0.87	2.89
	Rise	0.91	3.20	1.24	2.01

4.5 Summary

Accurate modelling of MOSFET capacitance is essential to delay calculations involving these devices. Determinations of the parameters needed by five increasingly sophisticated classes of capacitance descriptions have been done. The complexity varied

from a simple-minded single parameter model to one recognizing the independent nature of gate and diffusion capacitances and the effect of transition direction on capacitance.

A simple but acceptably-accurate mathematical model resulted from the estimations of the magnitudes of various MOSFET parasitic capacitances, and the judicious choice of approximations. It was found possible to 'lump sum' all of them into a single equivalent capacitance.

A general mathematical formula to calculate interconnection capacitance is written, where the capacitance is taken to be a sum of three types of capacitances that are associated with an interconnect. Using this formula, the worst case interconnect capacitance is calculated for each of the technologies being studied, 0.13 μm and 90 nm processes.

Four increasingly elaborate formulations for describing MOSFET capacitance have been proposed, leading up to Class 5. In this class, it is recognized that the voltage-dependent nature of device capacitances can mean that the direction in which a signal transition is occurring (rising, falling) has a strong effect on delay. Since the parameters for the first four of the models were extracted from measurements on closed-loop inverter chains, the effect of signal slope is implicit in their values. Slope effects must be explicitly dealt with in delay calculations involving the fifth class of description.

Chapter 5

CMOS Gate Delay Modeling

Effective modeling of gate delay in day-to-day circuit design has gone from being desirable to being vital as multi-million transistor ICs executed in deep submicron processes are becoming commonplace. By 'effective' is meant the combination of accuracy and speed which will make this possible – fast enough to enable the designers to identify and quantify critical paths, and accurate enough that modeling results are not differ from those of a full-scale simulation by more than a few percentage points.

Modeling any aspect of a large circuit at the individual transistor level is prohibitively time-consuming and so it will be agreed upon in advance that any new means of computing propagation times will be done at the gate level (macro-modeling), wherein the very large number of parameters necessary for transistor level modeling is significantly reduced. One of the techniques that used to estimate the delay at the gate level is the equivalent inverter technique. This technique seeks to replace a complex gate with an inverter that is 'equivalent' (for the purpose of delay determination) [51]. This is done by replacing each parallel and series multi-transistor structure with a single "equivalent" MOSFET of the appropriate type. A second method considers the effects on one another of nearest neighbors in a chain of gates [24]. Logical effort technique is another method that can be used to estimate the gate delay.

5.1 Equivalent inverter Technique

This technique is a widely used in practice and we are not aware who proposed it first.

In [51], Nabavi used this technique to collapse NOR and NAND gates to their equivalent inverter. The equivalent series connected transistor width is given as $(W_{N,(P)})_{equ} = (W_{N,(P)} / n)$, where $W_{N,(P)}$ is the width of NMOS (PMOS) transistor in the series connected transistors chain and n is the number of the transistors in this chain. Beside the condition of using a very fast input to validate this estimation, Nabavi [51] does not take in consideration the different operation regions that the series connected transistors may work in.

In Figure 5.1 (a), *n* NMOS transistors are connected in series. Let us assume that all transistors will get the same signal in the same time. At $t = 0$, $V_{out} = V_{dd}$, $V_0 = V_1 = V_2 = ... = V_{n-1} = 0$; these conditions are the worst case conditions [52].

Under such conditions and when applying an input signal of V_{dd}, transistor N_n starts working in saturation region as $V_{ds} > (V_{gs} - V_t)$ and then works in the linear region. V_{ds} across the transistors N_0 to N_{n-1} is zero and no current will flow; they are all working in the linear region with out leaving it.

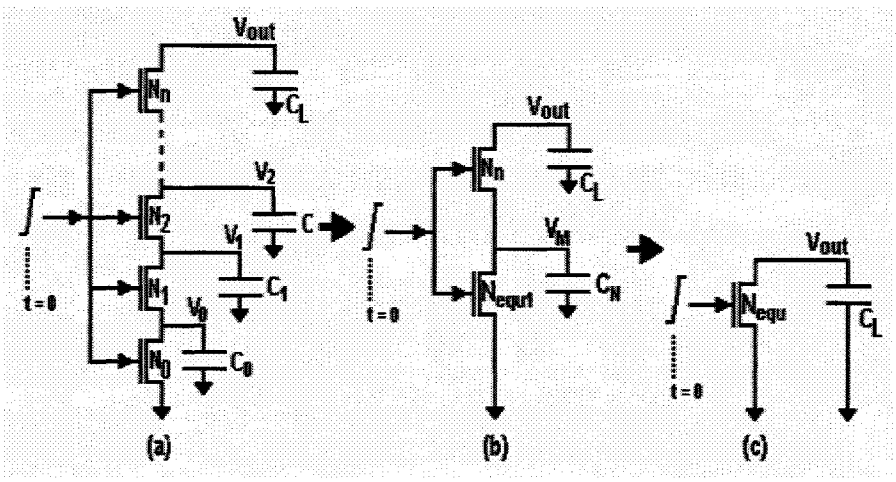


Figure 5.1: Series-connected n NMOS devices

The condition that describes a transistor working in saturation region is $v_{ds} > (V_{gs}-V_t)$. For $V_{gs} = V_{dd}$, to have transistors N_0 to N_{n-1} work in saturation V_{ds} across each one should be higher than $(V_{dd}-V_t)$. In a series connected transistors chain, the parasitic capacitance at each drain-source node (C_0 to C_{n-1}) is charged by the transistor above it. The maximum V_{ds} that can be achieved is $(V_{dd}-V_t)$ which is not enough to have the transistor working in saturation. One more thing that impedes the operation in saturation is the body effect. V_t increases as we move to higher transistors from the ground.

In [53] collapsing series-connected transistors is done in two steps. First, all transistors that work in the linear region (N_0 to N_{n-1}) are collapsed to one equivalent transistor, Figure 5.1 (b).Then this equivalent transistor is collapsed with the transistor that connected to the output node (N_n) to another one equivalent transistor, Figure 5.1 (c).

W_{Nequl} is given as in [51] while W_{Nequ} is given as the summation of W_N multiplied by saturation capacitance, C_{ast}, and W_{Nequl} multiplied by the linear capacitance, C_{lint}. C_{sat} is the division of (C_L . V_{dd}) by the integration of the current through W_N for the time interval during which W_N is working in saturation region. C_{lin} is ($1 - C_{sat}$) [54].

The parallel-connected transistors of the same width can be replaced by their equivalent single transistor of a width equal to the width of a transistor in the chain multiplied by the number of the transistors in the chain [54].

Using the collapsing technique, a complex CMOS circuit could be collapsed first to its equivalent NOR/NAND gates which then collapsed to their equivalent inverters [55].

5.2 Logical Effort Technique

In the Logical Effort technique, delay is expressed in terms of a basic delay unit (τ_{inv}), which is the delay of an inverter, driving another that is identical to itself, but having no parasitic capacitances. The absolute gate delay, D_G, is then simply defined as the product of the normalized delay of the gate (d) and τ_{inv} [56]:

$$D_G = \tau_{inv}.d = \tau_{inv}.(g.h + p) \qquad (5.1)$$

where p is the gate parasitic delay, g is the gate logical effort that depends on the gate topology. h is the electrical effort that equals to the ratio of the gate's output load capacitance to its input capacitance. The gate logical effort, g, is given as the gate input capacitance divided by the input capacitance of an inverter having $W_P/W_N = 2/1$. Table 5.1.

Table 5.1: Logical effort of NOT, NAND, NOR, and XOR gates [1, 56]

Gate	*Number of Inputs*									
	1		*2*		*3*		*4*		*n*	
	g	*p*	*g*	*p*	*g*	*p*	*g*	*p*	*g*	*p*
NOT	1	p_{inv}	-----	-----	-----	-----	-----	-----	-----	p_{inv}
NAND	---	---	4 / 3	$2.p_{inv}$	5 / 3	$3.p_{inv}$	6 / 3	$4.p_{inv}$	(n+2) / 3	$n.p_{inv}$
NOR	---	---	5 / 3	$2.p_{inv}$	7 / 3	$3.p_{inv}$	9 / 3	$4.p_{inv}$	(2n+1) / 3	$n.p_{inv}$
XOR	---	---	4	$4.p_{inv}$	12	$12.p_{inv}$	32	$64.p_{inv}$	------	------

gives the logical efforts and the parasitic delays of the four basic gates as a function of the number of inputs (n).

5.2.1 Extracting of τ_{inv} and P_{inv} in 0.13 µm and 90 nm Technologies

Referring to Equation 5.1, [56] stated that "The straight line that connects the points will have slope τ and will intercept the $h = 0$ axis at $d = \tau p_{inv}$". Table 5.2 summarizes these values for both of the technologies used in this work.

Therefore, to find τ_{inv} the appropriate plot is required: Let τ_{inv} be the delay of an inverter which is driving an identical copy of itself free of parasitic capacitances [56]. An inverter of minimum length and $W_P/W_N = 10L/5L$ is simulated six times, loaded by 1, 2, 3, 4, 5, and 6 identical inverter (s) successively, and its delay simulated. The delay measurements are plotted as a function of electrical effort, h, as in Figure 5.2.

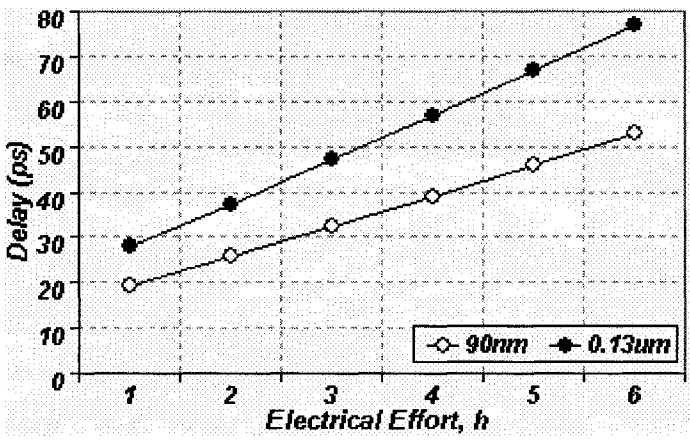


Figure 5.2: Inverter delay vs. electrical effort in 0.13 µm and 90 nm technologies

Table 5.2: τ_{inv} and P_{inv} as extracted in 0.13 µm and 90 nm technologies

Technology	τ_{inv} *(ps)*	p_{inv} *(ps)*	p_{inv}/τ_{inv}
90 nm	6.15	12.36	2.01
0.13 µm	9.17	17.70	1.93

p_{inv}/τ_{inv} has been plotted as a function of the technology in Figure 5.3. This figure shows that as feature-size decreases, this ratio increases.

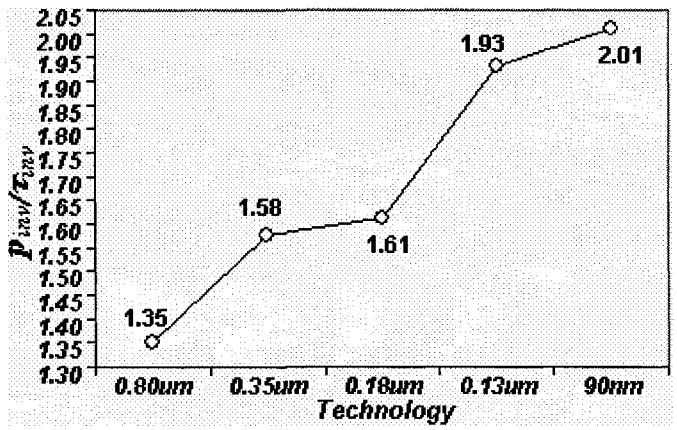


Figure 5.3: P_{inv}/τ_{inv} ratio vs. technology

5.3 Shams Delay Model

Figure 5.4 shows a chain of three CMOS gates.

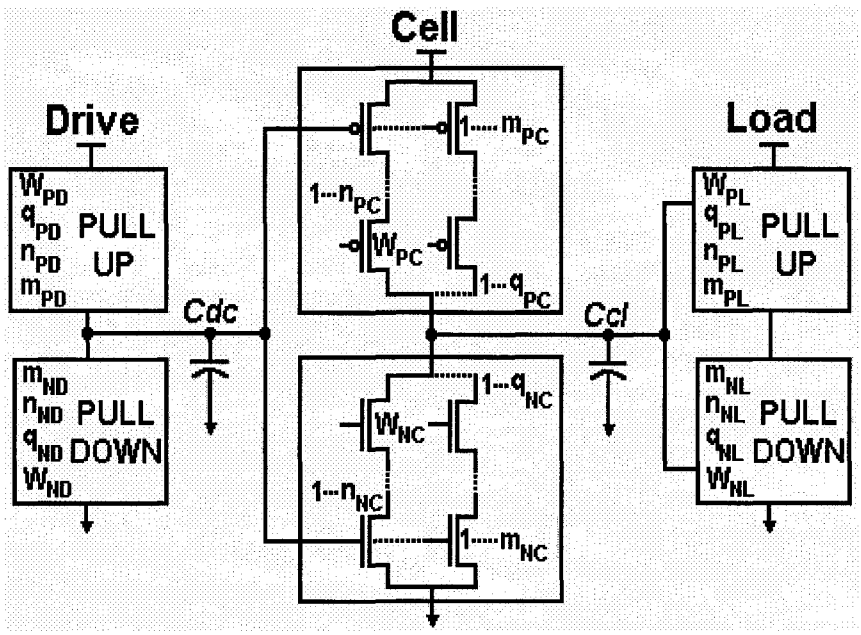


Figure 5.4: CMOS cell between drive gate and load gate [24]

W_{PD}, W_{ND}, W_{PC}, W_{NC}, W_{PL}, and W_{NL} are the pull up and pull down device width of the drive, cell, and load gates respectively. n_{PD}, n_{ND}, n_{PC}, n_{NC}, n_{PL}, and n_{NL} are the number of devices connected in series in pull up and pull down nets of the drive, cell, and load gates respectively. q_{PD}, q_{ND}, q_{PC}, q_{NC}, q_{PL}, and q_{NL} are the number of branches connected in

parallel in pull up and pull down nets of the drive, cell, and load gates respectively. m_{PD}, m_{ND}, m_{PC}, m_{NC}, m_{PL}, and m_{NL} are the number of devices connected in parallel to the same input in pull up and pull down nets of the drive, cell, and load gates respectively.

Finally, C_{dc} is gate, diffusion, and interconnect capacitances between drive gate and cell gate while C_{cl} is gate, diffusion, and interconnect capacitances between cell gate and load gate respectively.

All capacitances at drive gate output excluding the ones related to the cell gate can be expressed in two formulae [24]:

$$C_{DR} = C_{dcR} + q_{PD}.W_{PD}.C_{d\,PR} + q_{ND}.W_{ND}.C_{d\,NR} \quad (5.2)$$

$$C_{DF} = C_{dc\,F} + q_{PD}.W_{PD}.C_{d\,PF} + q_{ND}.W_{ND}.C_{d\,NF} \quad (5.3)$$

In a similar way, the load capacitances can be given as:

$$C_{LR} = C_{cl\,R} + m_{PL}.W_{PL}.C_{g\,PR} + m_{NL}.W_{NL}.C_{g\,NR} \quad (5.4)$$

$$C_{LF} = C_{cl\,F} + m_{PL}.W_{PL}.C_{g\,PF} + m_{NL}.W_{NL}.C_{g\,NF} \quad (5.5)$$

where C_{dcR}, C_{dcF}, C_{clR}, and C_{clF} are written as:

$$C_{dc\,R} = C_{Wire} + C_{d\,PR}.W_{PD} + C_{g\,PR}.W_{PC} \quad (5.6)$$

$$C_{dc\,F} = C_{Wire} + C_{d\,NF}.W_{ND} + C_{g\,NF}.W_{NC} \quad (5.7)$$

$$C_{cl\,R} = C_{Wire} + C_{d\,PR}.W_{PC} + C_{g\,PR}.W_{PL} \quad (5.8)$$

$$Ccl_F = C_{Wire} + Cd_{NF}.W_{NC} + C_{g\,NF}.W_{NL} \quad (5.9)$$

where C_{Wire} is interconnect capacitance as given in Equation 4.13.

The rise and fall propagation delay from the drive gate input to the cell gate output can be written as [24]:

$$D_R = D_{SR-Cell} + (1 + S_{PR}).D_{SF-Drive} \quad (5.10)$$

$$D_F = D_{SF-Cell} + (1 + S_{NF}).D_{SR-Drive} \quad (5.11)$$

where the cell rise and fall step delay are:

$$D_{SF-Cell} = \frac{v_{NF}.Y_{NCF}}{W_{NC}}.(q_{NC}.W_{NC}.C_{d\,NF} + q_{PC}.W_{PC}.C_{dPF} + C_{LF}) \quad (5.12)$$

$$D_{SR-Cell} = \frac{v_{PR}.Y_{PCR}}{W_{PC}}.(q_{NC}.W_{NC}.C_{d\,NR} + q_{PC}.W_{PC}.C_{d\,PR} + C_{LR}) \quad (5.13)$$

, and drive rise and fall step delay are:

$$D_{SF-Drive} = \frac{v_{NF}.Y_{NDF}}{W_{ND}}.(m_{NC}.W_{NC}.C_{g\,NF} + m_{PC}.W_{PC}.C_{g\,PF} + C_{DF}) \quad (5.14)$$

$$D_{SR-Drive} = \frac{v_{PR}.Y_{PDR}}{W_{PD}}.(m_{NC}.W_{NC}.C_{g\,NR} + m_{PC}.W_{PC}.C_{g\,PR} + C_{DR}) \quad (5.15)$$

The delay is given as:

$$D = (D_{SR-Cell} + D_{SF-Cell}) + \{(1 + S_{PR}).D_{SF-Drive} + (1 + S_{NF}).D_{SR-Drive}\} \quad (5.16)$$

5.3.1 Basic Gates Delay

In this section, we will describe modeled vs. simulated delay comparisons for the four basic gates, the NOT, the 2-input NOR, the 2-input NAND, and the 2-input XOR. The model used will be that described in section 5.2. These 'micro-application' calculations will have the same flavor as those for the full applications of Chapter-6 (delay model Level 7).

Assuming no off-path branches, there are 64 distinct ways of connecting the gates in chains of three (see Figure 5.4).This situation can be simplified if only four connections are used, where each of the gates will in turn take on the role of cell and will always be connected between two NOT gates, one acting as drive, the other acts as load. Therefore, the parameters *n*, *m* and *q* of the pull-up and pull-down nets of all drive and drive gates are equal to 1.

No branches will be considered at this point. (see Figures 5.5, 5.8, 5.10 and 5.12). Worst case paths only are to be considered (where the path involves connection between cell input and the SCM farthest from the output node). The delay is calculated assuming static conditions in the drive and load; that is they have been in their initial states for a long time before being presented with an input edge. Thus all nodes on the delay path will be either fully charged to V_{dd} or else discharged to ground.

The interconnect capacitance, C_{Wire}, is arbitrarily calculated for a 10 μm length.

5.3.1.1 NOT Gate

This simplest of the gate delay calculations is diagrammed in Figure 5.5. There is only one transistor in each pull-up and pull-down net in the chain and so the delay

degradation factor for internal capacitances due to SCMs, $Y = 1$ (see Equation 3.33).

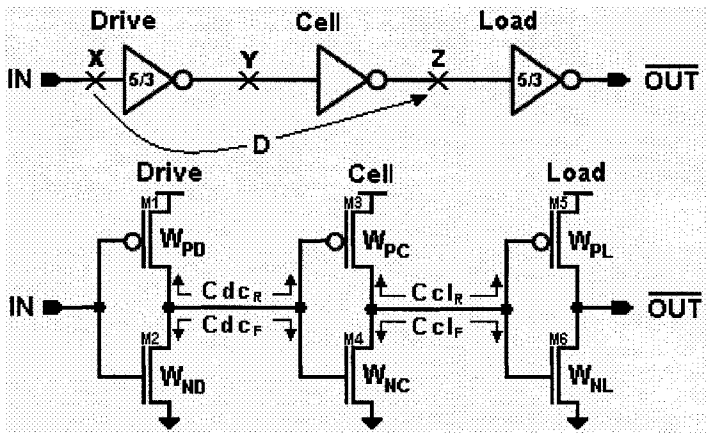


Figure 5.5: NOT-NOT-NOT chain

Figure 5.6 shows the chain delay (node X to node Z) as a function of the cell's gate width W_N in the two used technologies, 0.13 µm and 90 nm. A close match is achieved between model and simulator results in both technologies

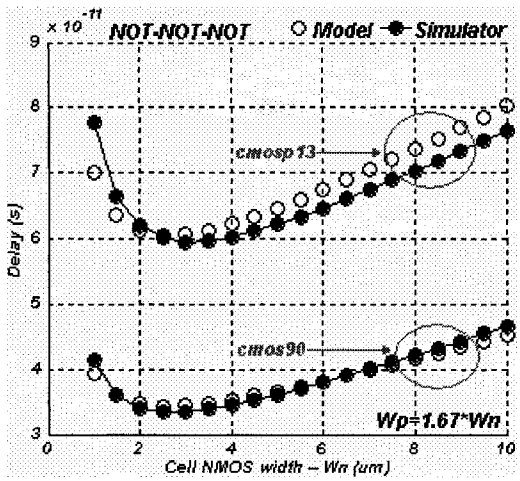


Figure 5.6: NOT-NOT-NOT chain model and simulator delay vs. W_{NC} (0.13 µm and 90 nm technologies)

A better feeling for this accuracy can be had by studying the plot of model *vs.* simulator delay difference percentage in Figure 5.7. This percentage is calculated as:

$$(Delay\ Difference)\% = \frac{Model\ Delay - SimulatorDealy}{Simulator\ Delay}.100 \qquad (5.17)$$

When this delay difference is positive, then the model is over-estimating the delay.

The negative percentage indicates that the model is under-estimating the delay. The average delay percentage is also calculated. It is the sum of absolute values of delay percentage percentages divided by their total number. The average percentage was 4.08% (0.13 μm) and 1.81% (90 nm).

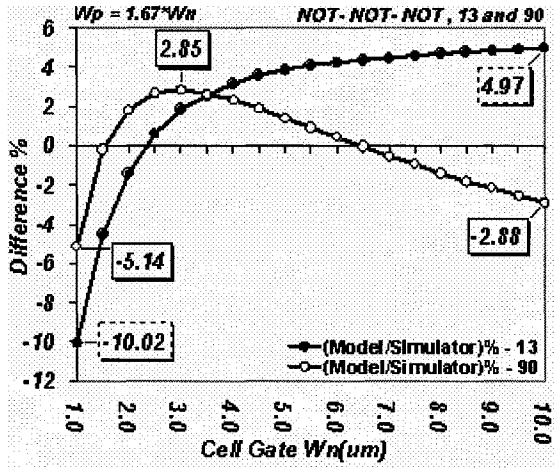


Figure 5.7: NOT-NOT-NOT chain delay percentage error in 0.13 μm and 90 nm technologies

5.3.1.2 2-Input NOR Gate

The cell gate is now a 2-input NOR gate (Figure 5.8). It can be seen that two drain diffusion capacitances (of M5 and M6) must be considered when computing C_{clF}.

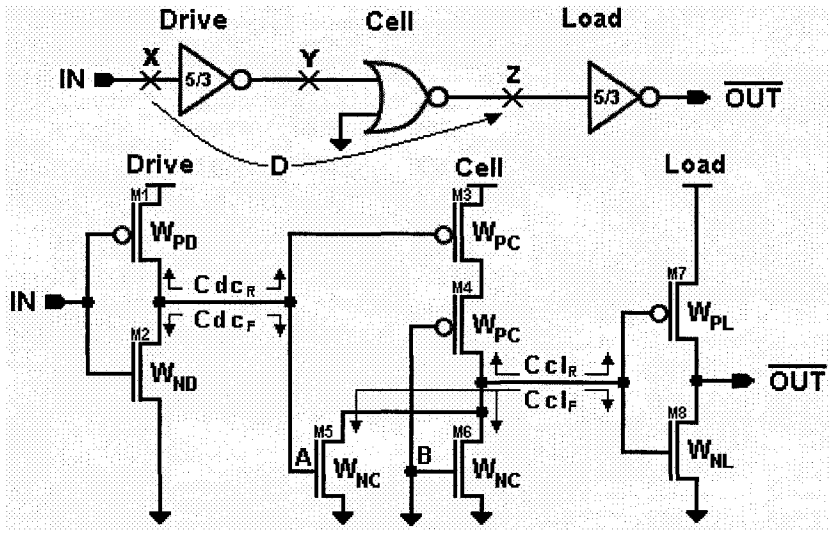


Figure 5.8: NOT-NOR-NOT chain

Transistors M4 and M6 are held at ground potential, and so the pull-up path is controlled by M3 and pull-down path by M5. The cell gate pull-up parameters are $n_{PC} = 2$, $m_{PC} = 1$ and $q_{PC} = 1$, the pull-down parameters $n_{NC} = 1$, $m_{NC} = 1$, and $q_{NC} = 2$.

The PMOS width was set at $W_{PC} = 4W_{NC}$, and W_{NC} scanned from 1 μm to 10 μm in steps of 0.5 μm. The model and simulator delays (X to Z) are plotted against W_n in Figure 5.9 (a), and the differences percentages between them in Figure 5.9 (b).

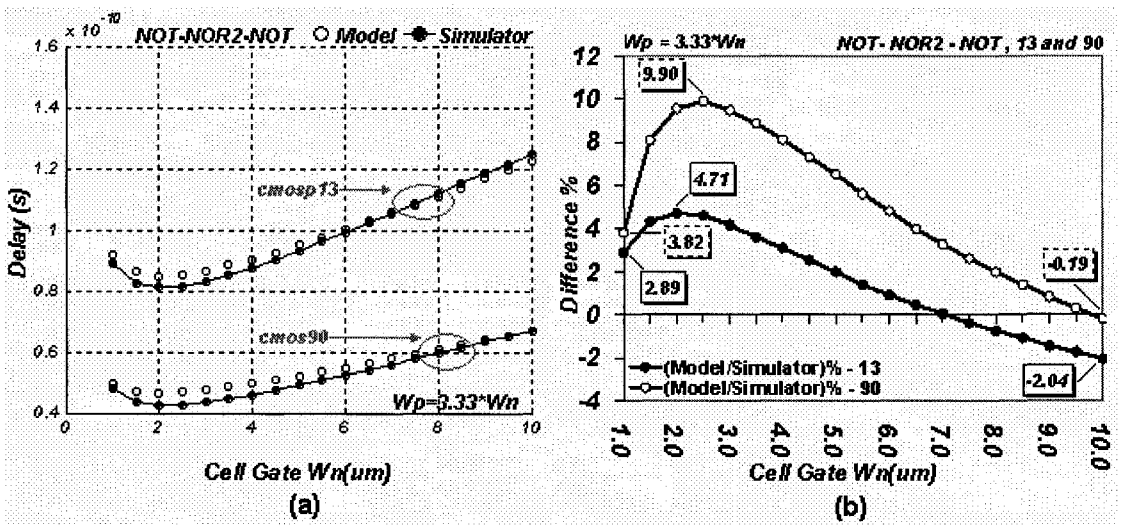


Figure 5.9: NOT- 2-Input NOR -NOT chain model and simulator delay (a) and delay difference percentage (b) vs. W_{NC}

Figure 5.9 (b) shows the delay difference as a function of W_{NC}. The average delay difference is 2.22% and 5.09% in 0.13 μm and 90 nm technologies respectively.

5.3.1.3 2-Input NAND Gate

This example is complementary to the previous one in that the SCM is on the cell's pull-up net (Figure 5.10), making $n_{NC} = 2$. The two branches in the pull-up net makes $q_{PC} = 2$ as well.

One gate and one diffusion capacitance, plus C_{Wire}, are to be considered when computing C_{dcR}, C_{dcF}, and C_{clF}. In the case of C_{clR}, one must take into account two drain diffusion capacitances, one load gate capacitance, and C_{Wire}. Note that transistors M4 and M5 are held to V_{dd}.

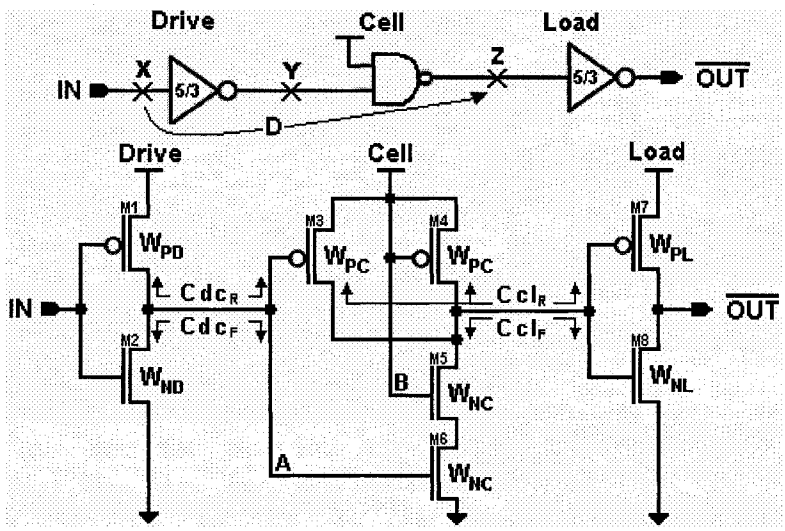


Figure 5.10: NOT-NAND-NOT chain

Figure 5.11 (a) shows the simulator and the model delays as a function of W_{NC} while Figure 5.11 (b) shows the delay difference vs. W_{NC}. The maximum delay difference is 7.68% and 11.66, and minimum delay difference is 0.09% and 1.20% with average delay difference of 5.72% and 9.08% in 0.13 μm and 90 nm technologies respectively.

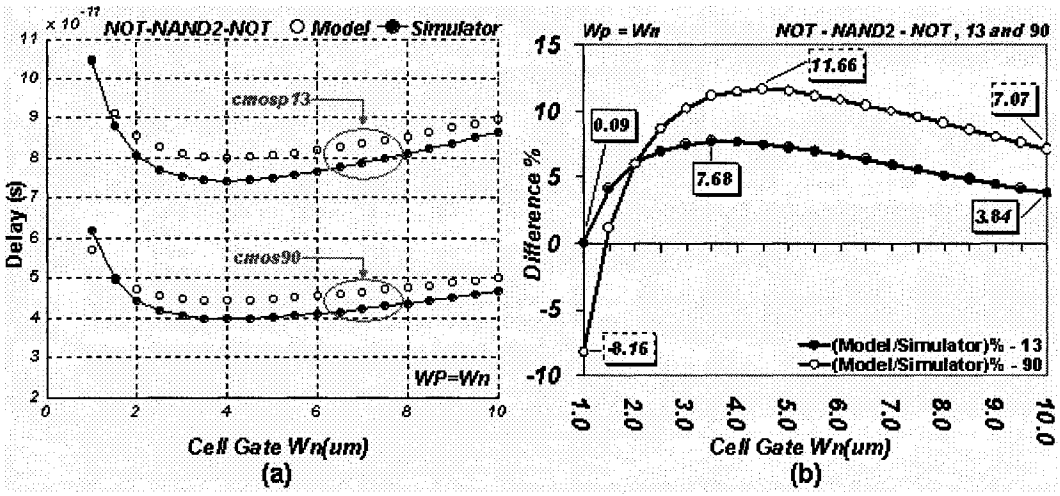


Figure 5.11: NOT- 2-Input NAND -NOT chain model and simulator delay (a) and delay difference percentage (b) vs. W_{NC}

5.3.1.4 2-Input XOR Gate

This gate has two branches in each of the pull-up and pull-down nets, both having two devices in series as indicated in Figure 5.12.

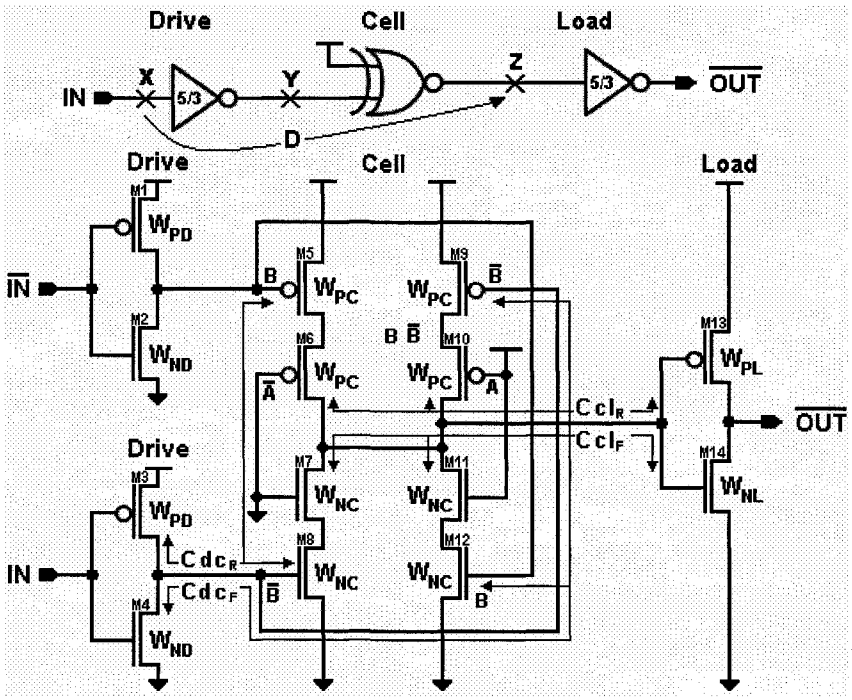


Figure 5.12: NOT- 2-Input XOR -NOT chain

The pull-up parameters are $n_{PC} = 2$, $m_{PC} = 1$ and $q_{PC} = 2$, and those for the pull-down net are $n_{NC} = 2$, $m_{NC} = 1$ and $q_{NC} = 2$.

Besides C_{Wire}, two gate and two diffusion capacitances must be considered when computing C_{cdR} and C_{cdF} at node Z, and C_{clR} and C_{clF} at node Y. Because input A is connected to V_{dd}, M6 and M11 are held ON so that the delay path trough these two branches is controlled by M5 and M12.

When IN is high, input B is high too. This switches ON M12, and pulls node Z to ground. The pull-down worst-case delay path is then from IN through M12 and M11, to node Z. Changing IN to low will cause input B to be taken low also, resulting in M5's switching ON and connecting node Z to V_{dd}. The worst-case pull-up delay path is that from IN through to M5 and M6, to node Z.

The simulated and calculated delays are plotted against W_{NC} in Figure 5.13 (a). The 90 nm technology result shows nearly perfect matching between the two. For the larger 0.13 μm technology, the matching of the curve derivatives is near-perfect, but the model results are displaced downward by a small constant amount.

Figure 5.13 (b) shows the delay difference percentages, plotted against W_{NC}. In both

used technologies (0.13 μm and 90 nm), the maximum delay differences are -10.24% and -10.36%; the minimum ones -4.71% and -0.59%. The averaged delay differences are respectively 6.62% and 2.12%.

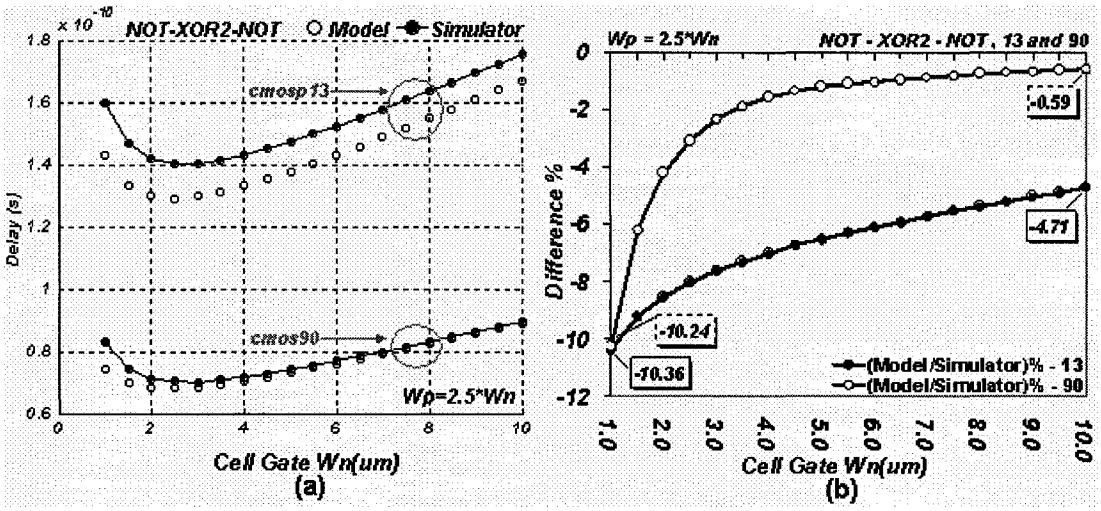


Figure 5.13: NOT-2- Input XOR -NOT chain model and simulator delays (a) and delay difference percentage (b) vs. W_{NC}

5.4 Summary

Delay estimation at transistor level has the advantage of high accuracy. Such accuracy is not obtained without price. A tradeoff between the accuracy and speed seems a must. Estimation at gate level is accelerating the modeling process and maintaining a good accuracy.

The equivalent inverter technique is systematically reducing the complexity of a gate to obtain a much simpler 'equivalent' structure; an equivalent inverter.

The Logical Effort technique is another micro-modeling technique that can be used to estimate the delay of a gate or of an entire logical path.

The much more elaborate and general method of dividing a gates chain into 3 gates 'sliding windows' involves estimation of the delay from the input of a drive gate to the output of a cell gate in terms of the size and topology of the drive, cell and load components of the gate triple. The delays of the four basic gates, NOT, 2-input NOR, 2-input NAND, and 2-input XOR gates were investigated using this method.

Chapter 6

Applications

Addition is one of the basic arithmetic operations. It is not only used to add two binary numbers, but it is also the base of some other operations like subtraction, multiplication, and division. The Adder can be found in many VLSI circuits such as DSP blocks and microprocessors. In these circuits, the adder lies on the critical path that evaluates the circuit operation.

Converting an n-bit input to 2^n-bit output and for each input combination, the decoder offers a single output line that has 1 while all other outputs lines are zero. The decoder is the nucleus of applications like memory addressing and decoding CPU instructions. In such applications, the decoder has great influence on the signal delay.

The binary comparator is a circuit that compares two binary numbers of n bit and provides 1 or 0 at the output depending on whether these numbers are equal or not. Comparators are used in a wide range of applications such as communication systems, encryption devices and microprocessors.

Two adders, 4-bit and 16-bit, are used to investigate the delay from the carry input, Cin, to each output of the adders. In the two other applications, 6x64 decoder and 8x256 decoder, the delay along the longest path is investigated. The fifth application is 8-bit comparator. The longest path from input to the output is investigated. The five different applications schematics are simulated in two technologies, 90 nm and 0.13 μm. The layout of the 16-bit adder is done. For this application, the delay simulation is performed according to the schematic and to the extracted views.

6.1 Delay Calculation Procedures

With each test application, the delay is calculated in eight levels. The first delay calculation is done based on logical effort technique that discussed in Section 5.2. Levels 2 to 5 calculations based on the four MOSFET capacitances levels discussed in Section 4.3 and extracted in Sections B1 to B4. Levels 6 to 8 are performed according to the MOSFET capacitance extracted in Section B5. Level 7 is based on the set of equations from Equation 5.2 to 5.16 and Equation 3.33. In Level 6, the delay is calculated the same way as in Level 7 but without including the empirical factors x and y in Equation 3.33. Level 8 employs the same equations that are used in Level 7 with one modification. Equation 5.16 is modified to include the second order effects of the input signal as discussed in Section 3.7.

The results of computation and measurement were compared in all cases and interconnect capacitance (C_{Wire}) of 10 μm length was assumed when calculating the delay.

The descriptions of delay calculation in the following five subsections will be made in terms of the arrangement of Figure 6.1. There may be branches at any node on the path, nodes k and n being chosen only for illustration. Gate 5 is working as a load and it is always a NOT gate.

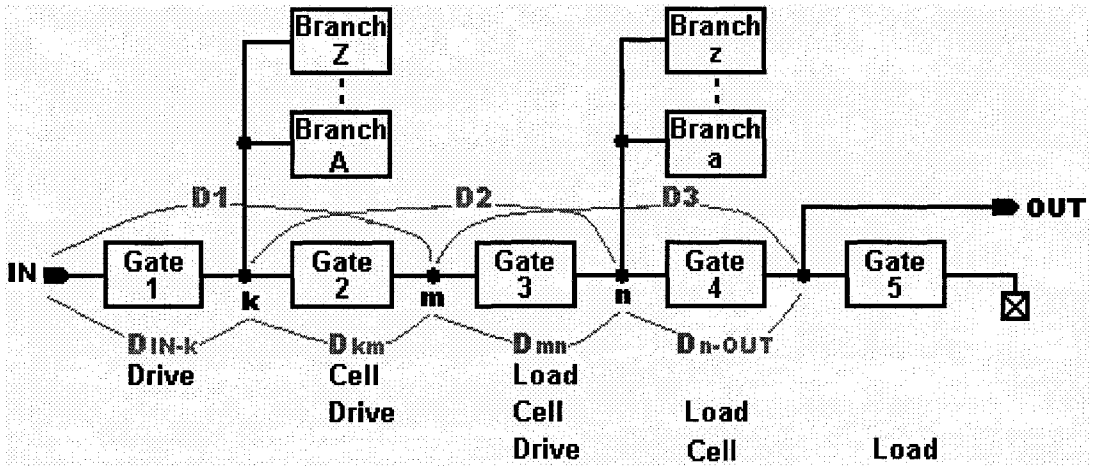


Figure 6.1: Path delay estimation

6.1.1 Delay Calculation: Level 1

The total path delay is the sum of the individual delays of Gates 1 to 4, as computed

with the on Logical Effort technique:

$$D_{LE(IN-OUT)} = D_{LE(IN-k)} + D_{LE(km)} + D_{LE(mn)} + D_{LE(n-OUT)}$$
$$= \{(g_1.h_1 + p_1) + (g_2.h_2 + p_2) + (g_3.h_3 + p_3) + (g_4.h_4 + p_4)\}\tau_{inv} \qquad (6.1)$$

where g_1, g_2, g_3, g_4, p_1, p_2, p_3, and p_4 values for a particular gate type are found in Table 5.1. The symbols h_1, h_2, h_3, and h_4, are the electrical efforts of Gates 1 to 4, computed as the quotient obtained when the gate's output capacitance is divided by its input capacitance.

6.1.2 Delay Calculation: Levels 2 to 5

For each gate (1 to 4), the delay is:

$$D_T = C.\left(\frac{v_P.W_{ND} + v_N.W_{PD}}{W_{PD}.W_{ND}}\right) \qquad (6.2)$$

v_P and v_N are calculated as in Equation 3.31. Their values can be found in Table 3.4. C is the capacitance at each node that works as a load of the gate. When calculating the node capacitance, cell gate, drive gate, and gate (s) on any branch connected to the node, and interconnect on which this node is located are considered. The total delay for the path is the sum of the individual gate delays:

$$D_T = D_{IN-k} + D_{km} + D_{mn} + D_{n-OUT}$$
$$= C_k.\left(\frac{v_P.W_{N1} + v_N.W_{P1}}{W_{P1}.W_{N1}}\right) + C_m.\left(\frac{v_P.W_{N2} + v_N.W_{P2}}{W_{P2}.W_{N2}}\right) +$$
$$C_n.\left(\frac{v_P.W_{N3} + v_N.W_{P3}}{W_{P3}.W_{N3}}\right) + C_{OUT}.\left(\frac{v_P.W_{N4} + v_N.W_{P4}}{W_{P4}.W_{N4}}\right) \qquad (6.3)$$

6.1.3 Delay Calculation: Level 6

At this level, the gate delay model includes the consequences of cascading gates. Cascades of three gates, a drive gate followed in by a cell gate and then a load gate, are considered at each step in the calculation. Based on the central gate, the delay is calculated, taking into consideration the gates on either side, and then the cascade is moved forward one gate position, the roles of the gates reassigned, and the calculation repeated, and so on, until the end of the chain has been reached. The delay for the entire path is the sum of the individual delays calculated in this manner.

In Figure 6.1, there are three cascades to be considered. They are (Drive, Cell, Load) = (1 2 3), (2 3 4) and (3 4 5). Three delays are computed, one for each cascade:

► D_1:

$$D1 = (D1_{SR-Cell} + D1_{SF-Cell}) + \{(1+S_{PR}).D1_{SF-Drive} + (1+S_{NF}).D1_{SR-Drive}\} \quad (6.4)$$

► D_2:

$$D2 = (D2_{SR-Cell} + D2_{SF-Cell}) + \{(1+S_{PR}).D2_{SF-Drive} + (1+S_{NF}).D2_{SR-Drive}\} \quad (6.5)$$

► D_3:

$$D3 = (D3_{SR-Cell} + D3_{SF-Cell}) + \{(1+S_{PR}).D3_{SF-Drive} + (1+S_{NF}).D3_{SR-Drive}\} \quad (6.6)$$

$D_{SF\text{-}Cell}$, $D_{SR\text{-}Cell}$, $D_{SF\text{-}Drive}$, and $D_{SR\text{-}Drive}$ are calculated as in Equations 5.12, 5.13, 5.14 and 5.15. The factor 'Y' in these equations is found with Equation 3.33, but with the part containing the empirical parameters x and y omitted (Y=N). The total delay is:

$$D_{T(IN-OUT)} = D1 + D2 + D3 \quad (6.7)$$

6.1.4 Delay Calculation: Level 7

Level 7 delay calculations are done in the same way as for Level 6, except that the quantities x and y are this time included in the calculation of Y.

6.1.5 Delay Calculation: Level 8

In section 3.7, the quadratic effects of the input signal were explored. Delay model Level 8 extends the slope-effect calculation to second order. Apart from this, the calculation proceeds in the same way as for Levels 6 and 7.

► D_1:

$$D1 = (D1_{SR-Cell} + D1_{SF-Cell}) + \{(1+S1_R + S2_R.D1_{SF-Drive}).D1_{SF-Drive} + (1+S1_F + S2_F.D1_{SR-Drive}).D1_{SR-Drive}\} \quad (6.8)$$

► D_2:

$$D2 = (D2_{SR-Cell} + D2_{SF-Cell}) + \{(1+S1_R + S2_R.D2_{SF-Drive}).D2_{SF-Drive} + (1+S1_F + S2_F.D2_{SR-Drive}).D2_{SR-Drive}\} \quad (6.9)$$

► D_3:

$$D3 = (D3_{SR-Cell} + D3_{SF-Cell}) + \{(1+S1_R + S2_R.D3_{SF-Drive}).D3_{SF-Drive} + (1+S1_F + S2_F.D3_{SR-Drive}).D3_{SR-Drive}\} \quad (6.10)$$

6.2 16-Bit Ripple Carry Adder

Figure 6.2 shows how to build a 16-Bit Ripple Carry Adder (RCA) from a 1-Bit RCA. Input A is connected high while input B is connected low. Now, as input Cin changes state, outputs S0 to S15 alternate the high and low status with output Cout.

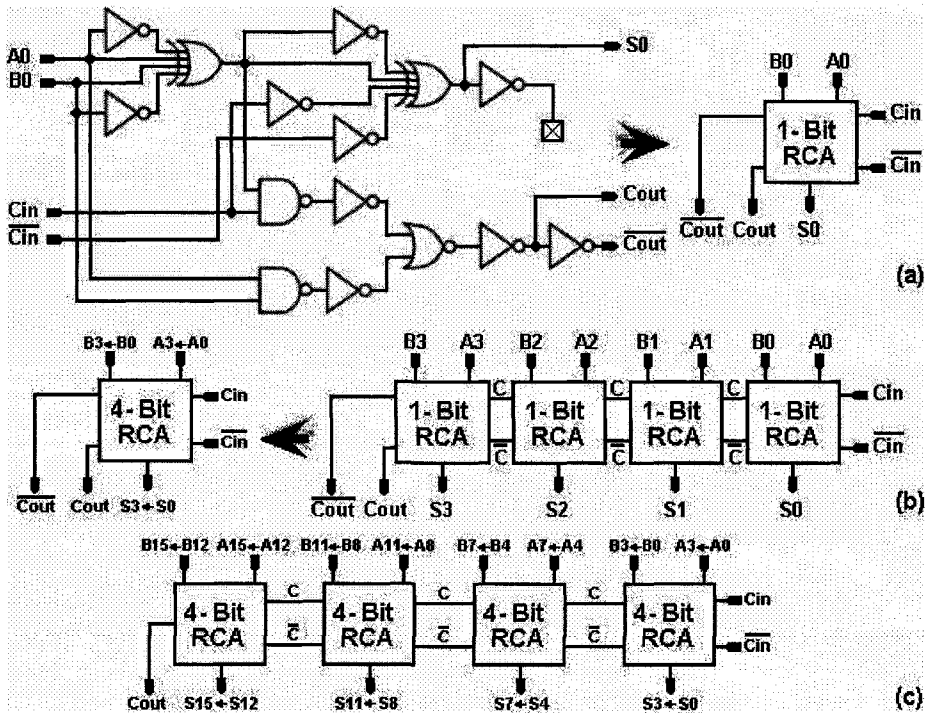


Figure 6.2: Ripple Carry Adder, 1-Bit schematic and symbol (a), 4-Bit block diagram and symbol (b), and 16-Bit block diagram(c)

Delays of signals on paths from the carry-in port (Cin) to the outputs (S0 to S15 and Cout) are considered. Figure 6.3 is showing the delay paths Cin-S0 and Cin-S1.

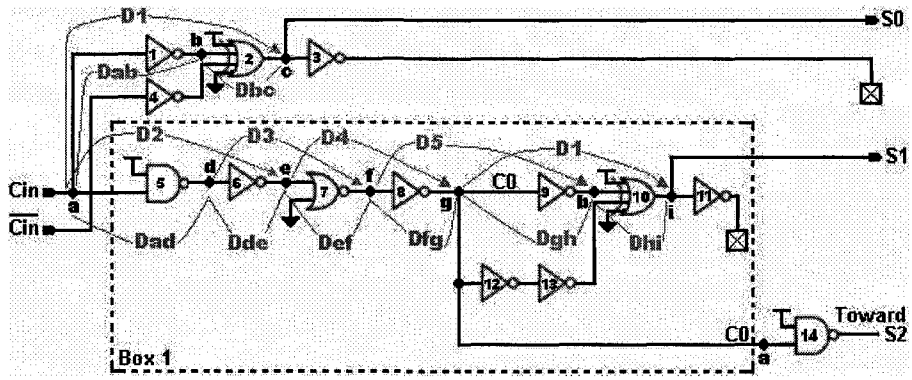


Figure 6.3: Delay Paths from Cin to S0 and Cin to S1

Box 1 is repeated for paths leading to outputs S_2 to S_{15} as shown in Figure 6.4. The final contribution to the delay of C_{out} occurs after box-15, where the buffer made of the sequence of Gates 5, 6, 7, and 8 (with 9 acting as a load) has the same general structure as is illustrated in Figure-6.3. The delays $D_{Cin\text{-}Si}$ (i = 2, 3... 15) are calculated as the delay over the entire preceding part of the path added to the delay from C_{in} to C_0 ($D_{Cin\text{-}C0}$). For example, $D_{Cin\text{-}S3} = (D_{Cin\text{-}S2} + D_{Cin\text{-}C0})$. The delay over the path from the carry-in to the carry-out ports (Cin-Cout) is the sum of the carry delay at each stage (box), added to the carry delay going from box-15 to C_{out} ($D_{C14\text{-}Cout}$).

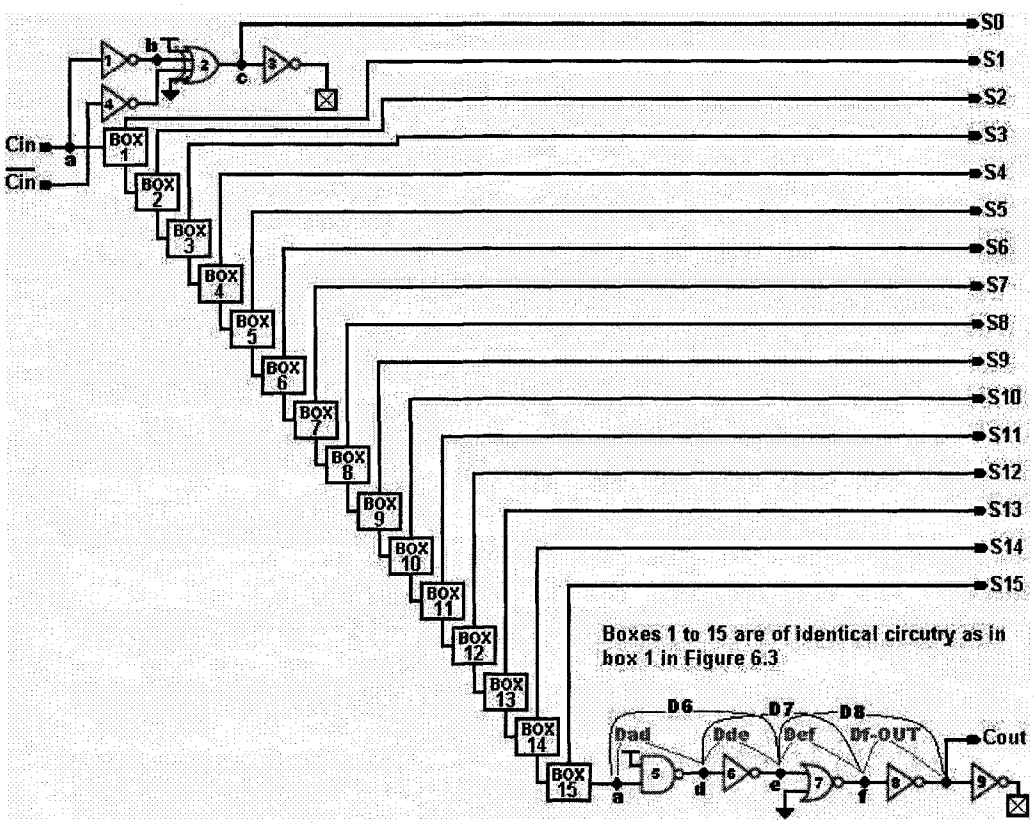


Figure 6.4: 16-Bit RCA full delay paths

The propagation delays of the 16-bit RCA, computed with the delay model Level 1 to Level 8 delay models are, in summary:

► Levels 1:

$$D_{LE\,(Cin-S0)} = D_{LE\,(ab)} + D_{LE\,(bc)} \qquad (6.11)$$

$$D_{LE(Cin-Co)} = D_{LE(ad)} + D_{LE(de)} + D_{LE(ef)} + D_{LE(fg)} \quad (6.12)$$

$$D_{LE(Cin-S1)} = D_{LE(Cin-C0)} + D_{LE(gh)} + D_{LE(hi)} \quad (6.13)$$

$$D_{LE(Cin-S2)} = D_{LE(Cin-C0)} + D_{LE(Cin-S1)} \quad (6.14)$$

. . . . .

$$D_{LE(Cin-S15)} = D_{LE(Cin-C0)} + D_{LE(Cin-S14)} \quad (6.15)$$

$$D_{LE(Cin-C14)} = D_{LE(ad)} + D_{LE(de)} + D_{LE(ef)} + D_{LE(f-Cout)} \quad (6.16)$$

$$D_{LE(Cin-Cout)} = 15.D_{LE(Cin-C0)} + D_{LE(Cin-C14)} \quad (6.17)$$

► Levels 2-5:

$$D_{Cin-S0} = D_{ab} + D_{bc} \quad (6.18)$$

$$D_{Cin-Co} = D_{ad} + D_{de} + D_{ef} + D_{fg} \quad (6.19)$$

$$D_{Cin-S1} = D_{Cin-C0} + D_{gh} + D_{hi} \quad (6.20)$$

$$D_{Cin-S2} = D_{Cin-C0} + D_{Cin-S1} \quad (6.21)$$

. . . . .

$$D_{Cin-S15} = D_{Cin-C0} + D_{Cin-S14} \quad (6.22)$$

$$D_{Cin-C14} = D_{ad} + D_{de} + D_{ef} + D_{f-Cout} \quad (6.23)$$

$$D_{Cin-Cout} = 15.D_{Cin-C0} + D_{Cin-C14} \quad (6.24)$$

► Levels 6-8:

$$D_{Cin-S0} = D1 \quad (6.25)$$

$$D_{Cin-Co} = D2 + D3 + D4 \quad (6.26)$$

$$D_{Cin-S1} = D_{Cin-C0} + D5 + D_{Cin-S0} \quad (6.27)$$

$$D_{Cin-S2} = D_{Cin-C0} + D_{Cin-S1} \quad (6.28)$$

. . . . .

$$D_{Cin-S15} = D_{Cin-C0} + D_{Cin-S14} \quad (6.29)$$

$$D_{Cin-C14} = D6 + D7 + D8 \quad (6.30)$$

$$D_{Cin-Cout} = 15.D_{Cin-C0} + D_{Cin-C14} \quad (6.31)$$

6.2.1 16-Bit RCA Layout

Figure 6.5 shows the fully costumed layout of the 16-Bit RCA based on 0.13 μm technology. The 90 nm technology version is similar.

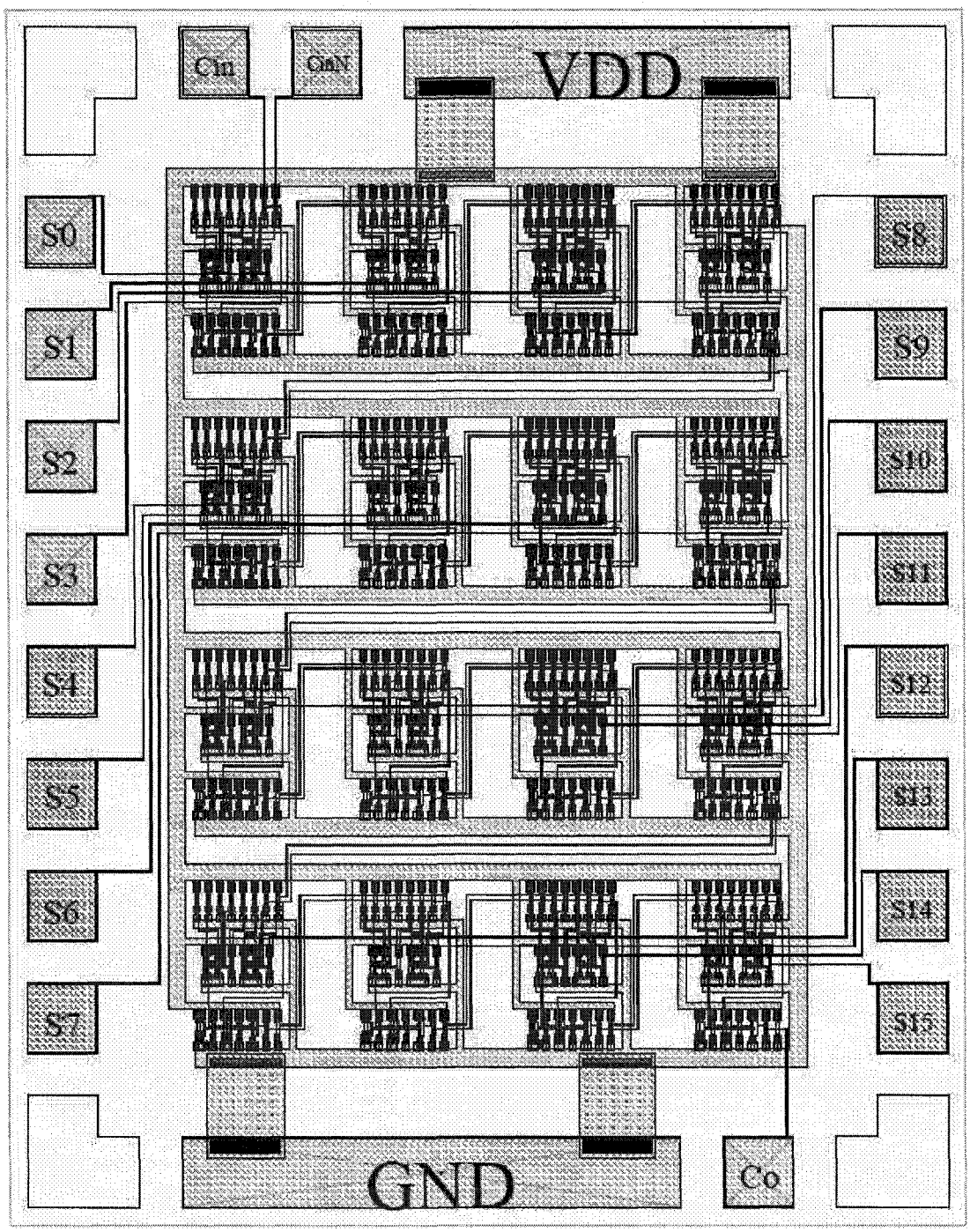


Figure 6.5: 16-Bit Ripple carry adder (RCA) layout in 0.13 μm technology

6.2.2 Model versus Simulation Comparison: 16-Bit RCA

Table 6.1 shows delay values for the paths Cin-S0 to Cin-S15 and Cin-Cout resulting from schematic- and extracted-view simulation measurements. The same delays, this time obtained with models, are also tabulated. These data are relevant to the 90nm implementation of the RCA.

Table 6.1: 16-Bit RCA delay simulator (from schematic, Sch., and from extracted, Ext.) and model readings in 90 nm technology

From To	*Simulator Delay(ps)*		*Model Delay (ps)*							
	Sch.	*Ext.*	*Level 1*	*Level 2*	*Level 3*	*Level 4*	*Level 5*	*Level 6*	*Level 7*	*Level 8*
Cn-S0	75.76	95.53	126.1	81.77	81.92	82.49	67.76	74.46	70.48	69.96
Cn-S1	246.5	329.9	271.4	296.2	304.5	301.6	256.1	275.3	268.3	264.4
Cn-S2	398.4	530.9	416.8	444.7	457.8	452.9	385.3	437.3	427.4	420.7
Cn-S3	550.3	732.7	562.1	593.2	611.2	604.3	514.4	599.4	586.4	577.1
Cn-S4	702.1	943.3	707.5	741.7	764.6	755.7	643.6	761.4	745.5	733.4
CnS5	854.0	1140	852.8	890.3	917.9	907.0	772.8	923.4	904.5	889.7
Cn-S6	1006	1342	998.2	1039	1071	1058	902.0	1085	1064	1046
Cn-S7	1158	1544	1144	1187	1225	1210	1031	1248	1223	1202
Cn-S8	1310	1755	1289	1336	1378	1361	1160	1410	1382	1359
Cn-S9	1462	1953	1434	1484	1531	1513	1290	1572	1541	1515
Cn-S10	1613	2155	1580	1633	1685	1664	1419	1734	1700	1671
Cn-S11	1765	2358	1725	1781	1838	1815	1548	1896	1859	1828
Cn-S12	1917	2569	1870	1930	1992	1967	1677	2058	2018	1984
Cn-S13	2069	2768	2016	2078	2145	2118	1806	2220	2177	2140
CnS14	2221	2970	2161	2227	2298	2269	1935	2382	2336	2297
Cn-S15	2373	3174	2306	2376	2452	2421	2065	2544	2495	2453
Cn-Cout	2401	3201	2313	2376	2454	2422	2067	2589	2542	2498

Based on Equation 5.17 and Table 6.1, the delay differences are plotted in figure 6.6. Levels 6, 7, and 8 are giving the most stable differences between all the eight levels with minimum delay difference of -1.72% and maximum delay difference of 11.68% with an average delay difference of 6.01%.

Levels 2 to 5 overestimate the delay of very short paths, but tend to become more accurate as the path length increases and actually begin to give superior results by the time the paths have become moderately-long. With all paths being considered, Levels 2 – 5

exhibit relative deviations of 23.51% to -13.92% with an average of 6.92%.

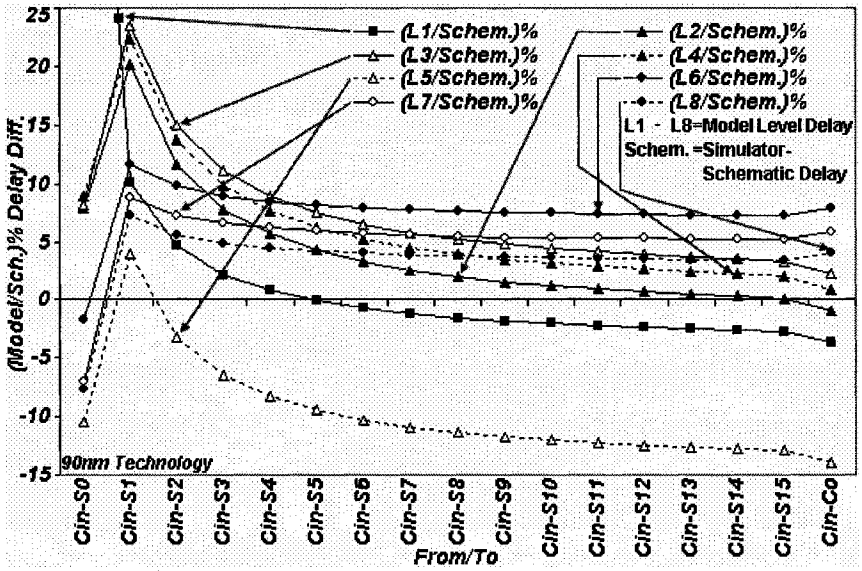


Figure 6.6: Delay vs. the 16-Bit RCA paths in 90 nm technology

Level 1, built around the Logical Effort technique, highly overestimates the very shortest path delay (Cin-S0). The 66.45% relative deviation cannot be seen on the plot – but the situation rapidly improves, the model producing competitive and superior results for all but the two shortest paths. Level 1 gives an average difference over all the paths at 6.37%. Averages are taken over absolute (unsigned) values.

Table 6.2 and Figure 6.7 convey information analogous to that of Table 6.1 and Figure 6.6, the difference being that they apply to the 0.13 μm process.

There are some interesting differences between Figures 6.6 and 6.7. A summary of approximate changes in accuracy on going from 90 nm to 0.13 μm can be made on studying the Figures:

- Level 1: Much worsened overall
- Level 2, Level-3, Level-4: Little change
- Level 5: Small improvement
- Level 6, Level-7, Level-8: Much improved overall

The qualitative features of the 90 nm plots are still evident in the 0.13 μm ones (for example, Level-8 is better than Level-7 in both). The severe worsening of performance at Level 1 and the substantial improvements at levels 6, 7 and 8 should be noted.

Table 6.2: 16-Bit RCA delay simulator (from schematic, Sch., and from extracted, Ext.) and model readings in 0.13 μm technologies

From To	Simulator Delay(ps)		Model Delay (ps)							
	Sch.	Ext.	Level1	Level2	Level3	Level4	Level5	Level6	Level7	Level8
Cn-S0	147.3	192.0	189.0	159.0	158.8	154.9	139.5	134.4	133.4	133.3
Cn-S1	470.0	566.6	405.9	574.4	580.7	545.3	501.2	492.6	489.2	483.8
Cn-S2	755.4	884.5	622.8	862.2	872.3	817.8	752.3	781.8	776.0	766.7
Cn-S3	1041	1221	839.7	1150	1164	1090	1003	1071	1063	1050
Cn-S4	1326	1530	1057	1438	1456	1363	1255	1360	1350	1333
CnS5	1612	1844	1274	1726	1747	1635	1506	1650	1636	1616
Cn-S6	1897	2162	1491	2014	2039	1908	1757	1939	1923	1898
Cn-S7	2183	2499	1708	2301	2330	2180	2008	2228	2210	2181
Cn-S8	2468	2835	1924	2589	2622	2453	2259	2517	2497	2464
Cn-S9	2754	3126	2141	2877	2914	2725	2510	2807	2783	2747
Cn-S10	3039	3429	2358	3165	3205	2998	2761	3096	3070	3030
Cn-S11	3324	3733	2575	3453	3497	3270	3012	3385	3357	3313
Cn-S12	3610	4094	2792	3740	3789	3543	3264	3674	3644	3596
Cn-S13	3895	4385	3009	4028	4080	3815	3515	3963	3931	3879
CnS14	4180	4688	3226	4316	4372	4088	3766	4253	4217	4162
Cn-S15	4466	4994	3443	4604	4663	4360	4017	4542	4504	4445
Cn-Cout	4517	5011	3453	4605	4666	4360	4018	4622	4583	4521

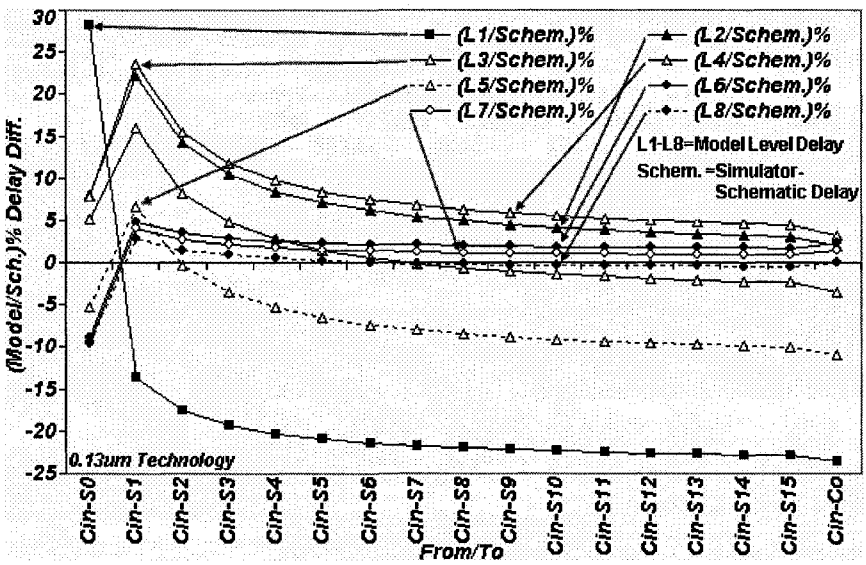


Figure 6.7: Delay Differences vs. the 16-Bit RCA paths in 0.13 μm

As one might expect, the predicted delays resulting from extracted view simulations are larger than those coming from schematic view simulations and higher still than those produced by all eight models (true in both 90 nm and 0.13 μm cases). In Figure 6.8 are

plotted the delay differences between Level 7 calculations and simulator (extracted) measurements for all paths. The figure shows that the differences increase on going from 0.13 μm to 90 nm technology. The model-to-extracted-differences percent relative deviations lie between -26.22% and -18.67% with an average of -21.08% (90 nm), to be compared with minimum=-30.53%, maximum= -8.55%, average= -12.25% (0.13 μm).

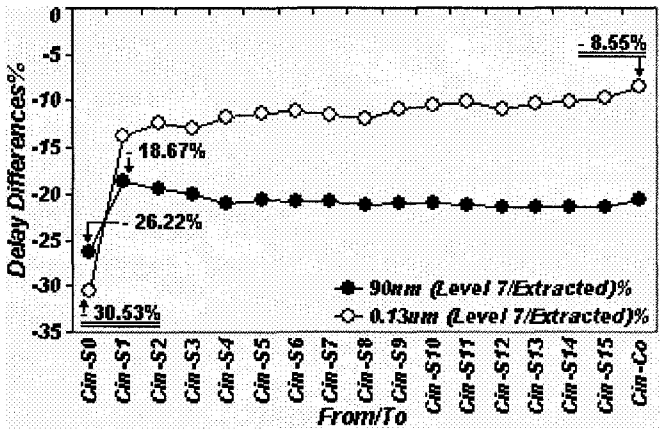


Figure 6.8: Level 7 model to extracted delay differences vs. the 16-Bit RCA paths in 90 nm and 0.13 μm Technologies

6.3 6x64 NAND Decoder

Figure 6.9 shows the evolution of a 2x4 NAND decoder design into a 6x64 device. After correct operation had been verified, the delay along the path from A0 to F0 was simulated and estimated with the delay models. This delay is calculated as:

► Level 1:

$$D_{LE(A0-F0)} = D_{LE(ab)} + D_{LE(bc)} + D_{LE(cd)} + D_{LE(de)} + D_{LE(ef)} + D_{LE(fg)} + D_{LE(gh)} \qquad (6.32)$$

► Level 2, Level 3, Level 4, and Level 5:

$$D_{A0-F0} = D_{ab} + D_{bc} + D_{cd} + D_{de} + D_{ef} + D_{fg} + D_{gh} \qquad (6.33)$$

► Level 6, Level 7, and Level 8:

$$D_{A0-F0} = D1 + D2 + D3 + D4 + D5 + D6 \qquad (6.34)$$

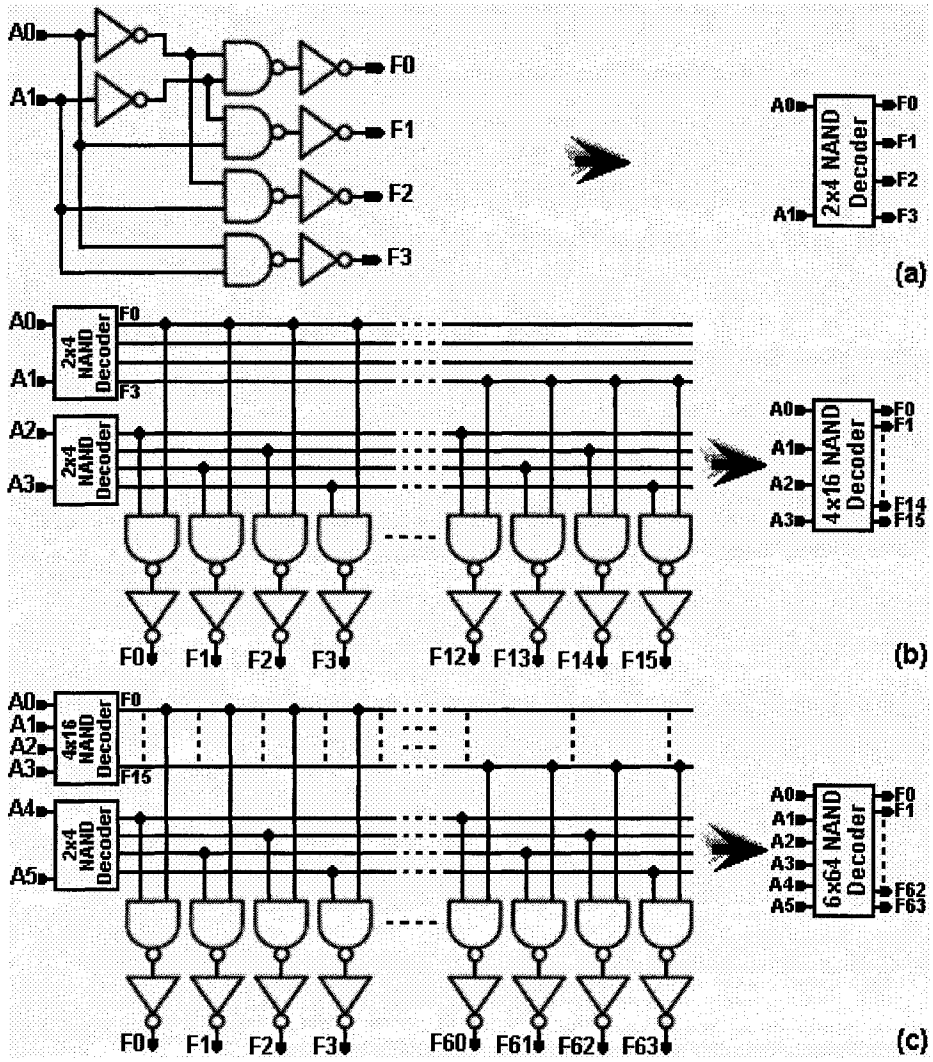


Figure 6.9: 6x64 decoder (c) as built from 4x16 decoder (b), and 2x4 decoder (a)

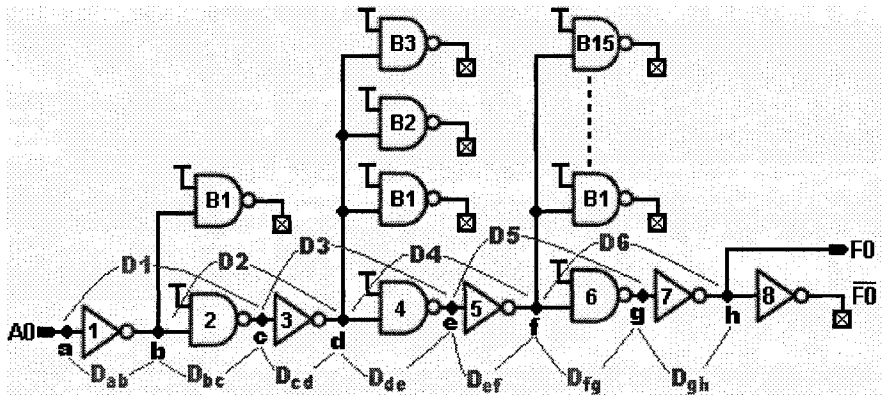


Figure 6.10: 6x64 decoder, A0 to F0 delay path

The results obtained by simulation and modeling are presented in Table 6.3

Table 6.3: 6x64 NAND decoder: delay along the A0-F0 Path by various

Technology	*Simulator Delay(ps)*	*Model Delay (ps)*							
	Schematic	*Level1*	*Level2*	*Level3*	*Level4*	*Level5*	*Level6*	*Level7*	*Level8*
90nm	349.90	328.52	404.15	456.44	424.65	394.87	362.22	361.87	355.85
0.13μm	598.20	489.88	784.67	830.44	676.97	665.34	635.66	645.93	622.04

The usual models vs. simulation relative errors are plotted in Figure 6.11 as a function of model level. Its is very obvious that levels 6, 7 and 8 provide the best match to simulator results in both technologies, with average differences of 2.88% (90 nm) and 6.08% (0.13 μm). Level 1 shows a close matching to the simulator in 90 nm technology with -6.11% delay difference percentage. Levels 2, 3, 4 and 5 formulae over-estimate the delay (the average relative deviation, taken over these four levels is 20.04% (90 nm) and 23.60% (0.13 μm).)

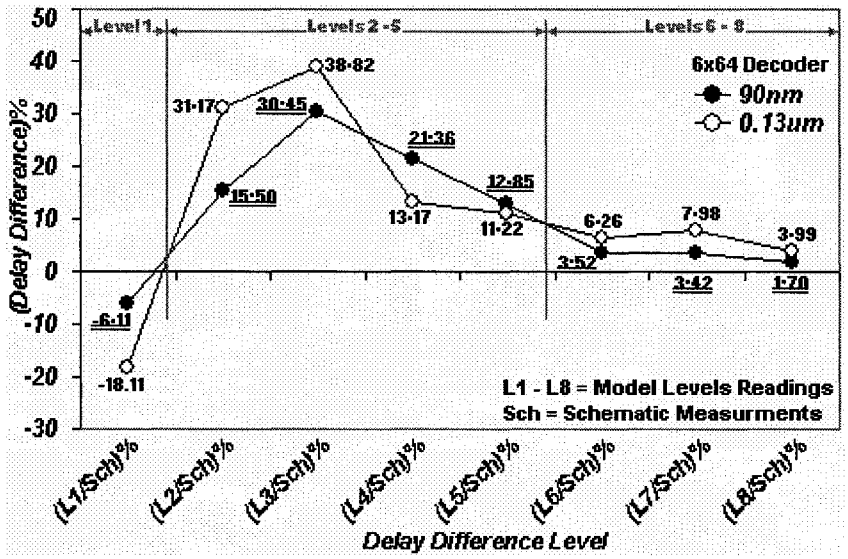


Figure 6.11: 6x64 NAND decoder delay differences vs. model eight

6.4 8x256 NAND Decoder

As with the 6x64 decoder, an 8x256 one can be built up from simpler ones (Figure 6.11 (a)). The delay along the path from A_0 to F_0 (Figure 6.12 (b)) will be studied.

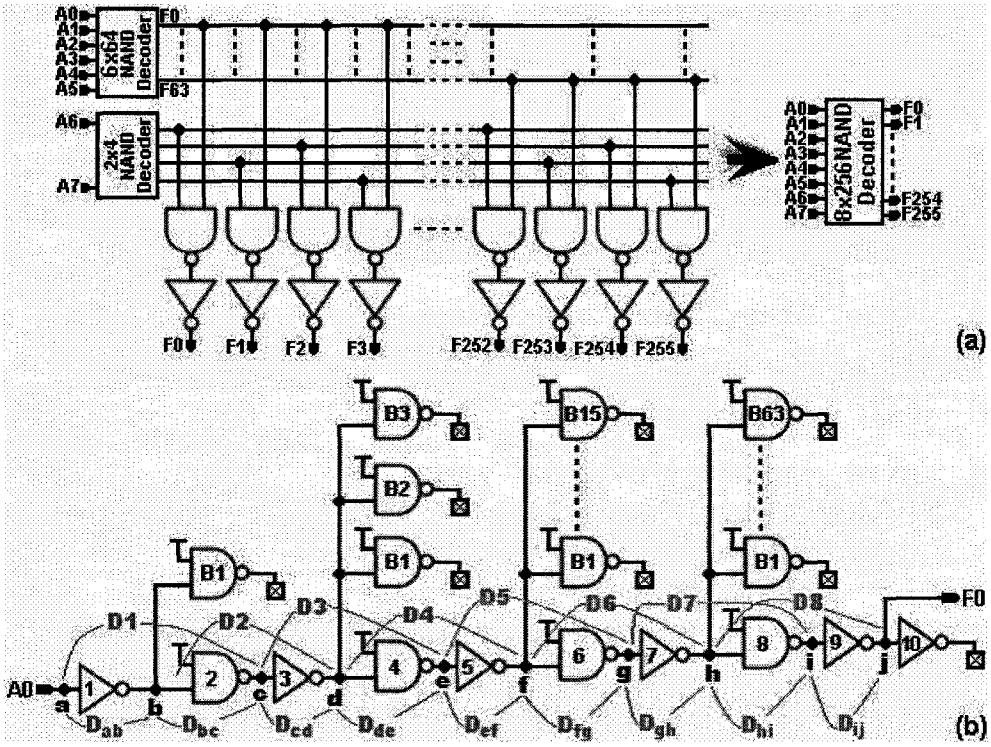


Figure 6.12: 8x256 decoder (a) and the delay path from A0 to F0 (b)

The signal path of interest is longer in this decoder than in the previous one, and a larger number of off-path (branching) gates are present, in particular there is an additional branch having 63 gates, which should make the path delay larger than in the case of the 6x64 decoder. Simulation and model results are in Table 6.4.

Table 6.4: 8x256 NAND decoder delay simulator measurements and model readings in 90 nm and 0.13 μm technologies

Technology	*Simulator Delay(ps)*	*Model Delay (ps)*							
	Schematic	*Level1*	*Level2*	*Level3*	*Level4*	*Level5*	*Level6*	*Level7*	*Level8*
90nm	900.10	778.73	1184.4	1359.1	1272.9	1231.4	813.61	810.86	799.33
0.13μm	1556.0	1161.1	2300.5	2453.9	1837.8	1894.8	1461.0	1477.1	1430.0

The delay along the path from A_0 to F_0 was simulated and estimated with the delay models:

► Level 1:

$$D_{LE(A0-F0)} = D_{LE(ab)} + D_{LE(bc)} + D_{LE(cd)} + D_{LE(de)} + D_{LE(ef)} + D_{LE(fg)} + D_{LE(gh)} + D_{LE(hi)} + D_{LE(ij)} \quad (6.35)$$

► Level 2, Level 3, Level 4, and Level 5:

$$D_{A0-F0} = D_{ab} + D_{bc} + D_{cd} + D_{de} + D_{ef} + D_{fg} + D_{gh} + D_{hi} + D_{ij} \quad (6.36)$$

► Level 6, Level 7, and Level 8:

$$D_{A0-F0} = D1 + D2 + D3 + D4 + D5 + D6 + D7 + D8 \quad (6.37)$$

Levels 6, 7 and 8 provide the best match to simulator results in both technologies, with average differences of 10.24% (90 nm) and 6.43% (0.13 μm). The Level 1 formula underestimates the delay by 13.48% (90 nm) and 2.54% in (0.13 μm). The Level 2, Level 3, Level 4 and Level 5 formulae over-estimate it, the differences ranging between highs of 50.99% (90 nm) and 57.71% (0.13 μm), and lows of 18.11% (90 nm) and 31.59% (0.13 μm).

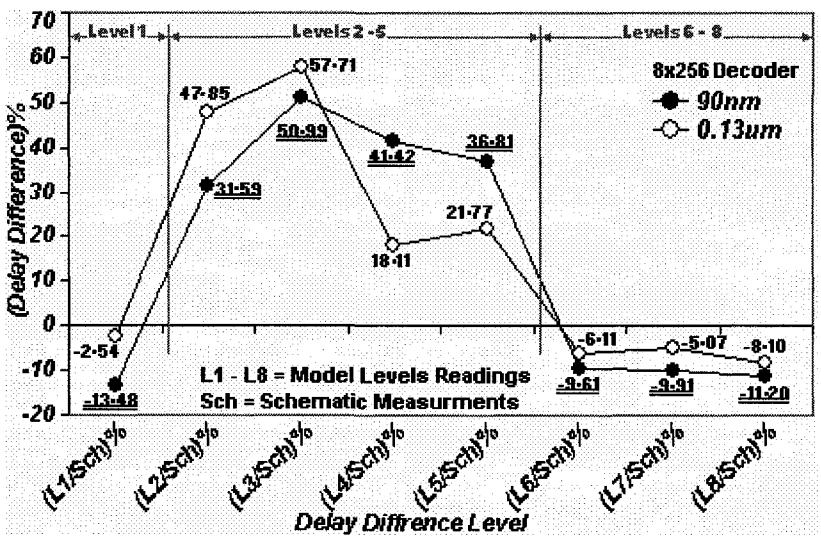


Figure 6.13: 8x256 NAND decoder: model *vs.* simulator delay differences

6.5 8-Bit Comparator

The general method of the adder and decoder designs is applied to the 8-bit comparator design. An evolutionary process proceeding from a 2-bit starting point is used (Figure 6.14 (b)). The design method is implicit in its architecture (Figure 6.14 (a)). Two

inverters having a *Wn/Wp* ratio of 5 μm/3 μm were connected to the comparator outputs (X and Y) to act as a load.

The comparator outputs encode the relationship between the magnitudes of the 8-bit inputs:

$$(A = B) \Rightarrow (X=0, Y=0) \quad (A > B) \Rightarrow (1, 0) \quad (A < B) \Rightarrow (0, 1)$$

Figure 6.14 (c) shows the longest path in each stage of (a). It is seen that the longest path through the whole comparator will be three times this.

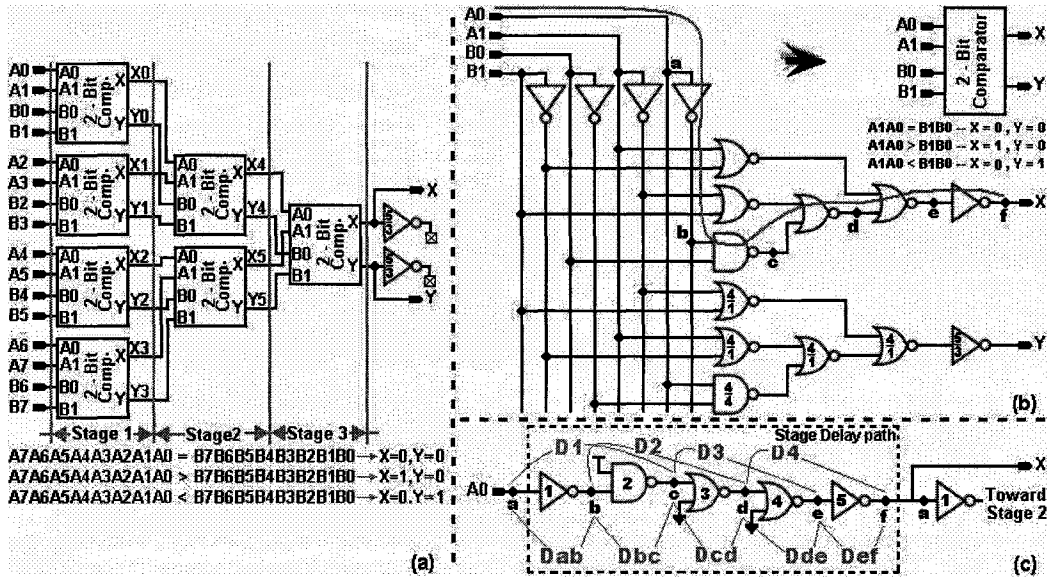


Figure 6.14: 8-Bit Comparator (a) , 2-Bit Comparator Circuitry and Symbol (b), and the Single-Stage Longest Path (c)

Simulation and model results are in Table 6.5. Relative prediction errors are plotted in Figure 6.15.

Table 6.5: 8-Bit comparator: delay along longest signal path by various means

Technology	*Simulator Delay(ps)*	*Model Delay (ps)*							
	Schematic	*Level1*	*Level2*	*Level3*	*Level4*	*Level5*	*Level6*	*Level7*	*Level8*
90 nm	670.00	526.04	493.42	489.07	495.06	401.34	698.43	694.16	685.39
0.13 μm	1151.0	785.18	954.99	949.00	944.65	843.53	1229.2	1223.3	1200.9

Delay estimates computed by model:

► Level 1:

$$D_{LE(A0-F0)} = 3.[D_{LE(ab)} + D_{LE(bc)} + D_{LE(cd)} + D_{LE(de)} + D_{LE(ef)}] \quad (6.38)$$

► Level 2, Level 3, Level 4, and Level 5:

$$D_{A0-F0} = 3.[D_{ab} + D_{bc} + D_{cd} + D_{de} + D_{ef}] \quad (6.39)$$

► Level 6, Level 7, and Level 8:

$$D_{A0-F0} = 3.[D1 + D2 + D3 + D4] \quad (6.40)$$

Investigation of Figure 6.15 tells that levels 6 to 8 are still giving a stable and close match to the simulator measurements. For the 90 nm process, the relative differences come between a high of 4.24% and a low of 2.30%, with an average of 3.38%. On going to the 0.13μm technology, these values become 6.79%, 4.34% and 5.80% respectively. Model levels 2 to 5 underestimate the delay by an average of 29.89% (90 nm) and 19.81% (0.13 μm). Model Level-1 also underestimates by 21.49% (90 nm) and 31.78% (0.13μm).

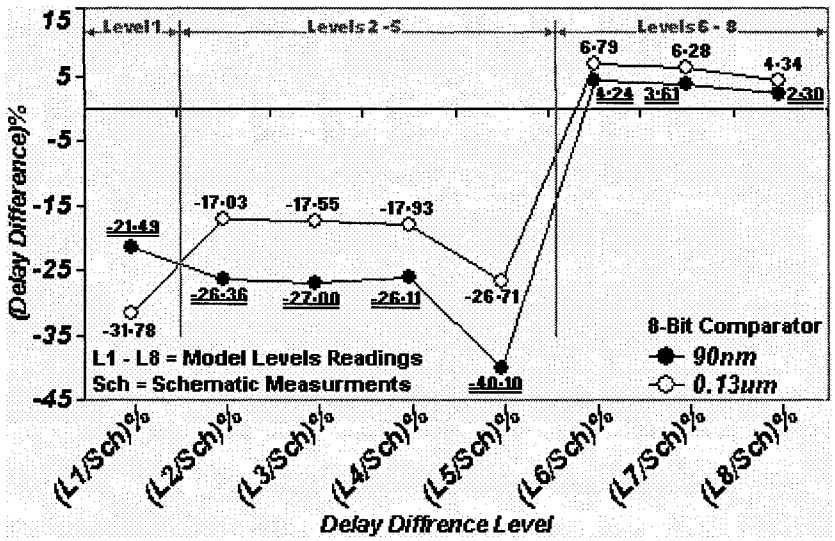


Figure 6.15: 8-Bit comparator: model *vs.* simulator delay differences

6.6 4-Bit NAND Adder

In the previous four applications, the highest number of series-connected transistors in any delay path is two. In this fifth and last application, a 4-Bit adder built from 2-, 3-, and 4-input NAND gates, which respectively contain two, three and four series-connected

transistors. This adder will thus provide a 'test bed' which will be used to study how the behaviours of our delay models are affected by changes in this parameter. It will be found that the delay models can be quite sensitive in this respect.

Figure 6.16 (a) shows, for reference, a 1-Bit adder constructed with multi-input NAND gates. To appreciate the nature of the tests, one need only know about the paths over which the carry-in signal reaches the four output bits. These are shown in Figure 6.2 (b).

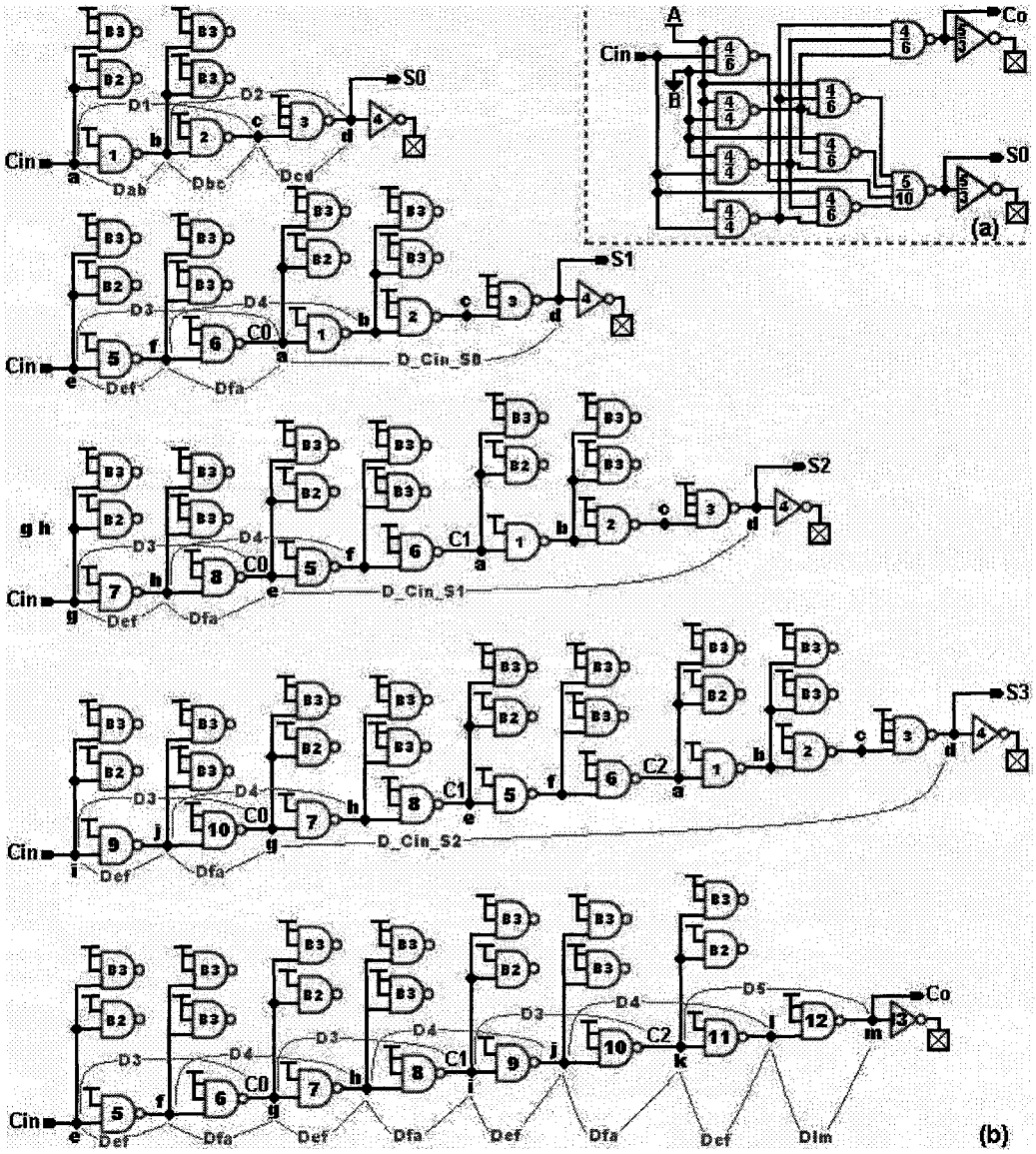


Figure 6.16: 1-Bit NAND adder built from 2, 3 and 4 Input Gates (a), Cin-Si and Cin-Cout Paths of the 4-Bit NAND Adder (b)

The delays along the Cin-Si and Cin-Co paths are calculated as follows:

► Levels 1:

$$D_{LE(Cin-S0)} = D_{LE(ab)} + D_{LE(bc)} + D_{LE(cd)} \quad (6.41)$$

$$D_{LE(Cin-S1)} = D_{LE(Cin-S0)} + D_{LE(ef)} + D_{LE(fa)} \quad (6.42)$$

$$D_{LE(Cin-S2)} = D_{LE(Cin-S1)} + D_{LE(ef)} + D_{LE(fa)} \quad (6.43)$$

$$D_{LE(Cin-S3)} = D_{LE(Cin-S2)} + D_{LE(ef)} + D_{LE(fa)} \quad (6.44)$$

$$D_{LE(Cin-Co)} = 4.D_{LE(ef)} + 3.D_{LE(fa)} + D_{LE(lm)} \quad (6.45)$$

► Level 2, Level 3, Level 4, and Level 5:

$$D_{Cin-S0} = D_{ab} + D_{bc} + D_{cd} \quad (6.46)$$

$$D_{Cin-S1} = D_{Cin-S0} + D_{ef} + D_{fa} \quad (6.47)$$

$$D_{Cin-S2} = D_{Cin-S1} + D_{Lf} + D_{fa} \quad (6.48)$$

$$D_{LCin-S3} = D_{Cin-S2} + D_{LEef} + D_{fa} \quad (6.49)$$

$$D_{Cin-Oo} = 4.D_{ef} + 3.D_{fa} + D_{lm} \quad (6.50)$$

► Level 6, Level 7, and Level 8:

$$D_{Cin-S0} = D1 + D2 \quad (6.51)$$

$$D_{Cin-S1} = D_{Cin-S0} + D3 + D4 \quad (6.52)$$

$$D_{Cin-S2} = D_{Cin-S1} + D3 + D4 \quad (6.53)$$

$$D_{Cin-S3} = D_{Cin-S2} + D3 + D4 \quad (6.54)$$

$$D_{Cin-Co} = 3.D3 + 3.D4 + D5 \quad (6.55)$$

The simulations were done twice, once for the carry-in signal being connected to the transistor closest to the output node in the SCM, and then for the carry-in connected to the farthest transistor (worst-case delay). These two connections are shown in Figure 6.17 for 4-input NAND gate.

As in Application 1 (16-Bit RCA), Input A is connected high while input B is connected low and the outputs will completely depend on the status of input Cin.

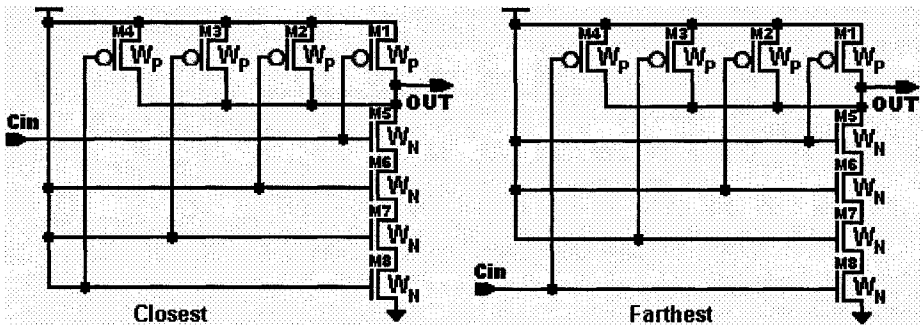


Figure 6.17: Cin as connected to closest and farthest transistors in the SCM.

The results of measurements and calculations are listed in table 6.6.

Table 6.6: 4-Bit NAND Adder: Delay from Cin to Nearest and Farthest Devices in SCM Chain.

Technology	From To	Schematic Delay(ps)		Model Delay (ps)							
		Closest	Farthest	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8
90 nm	Cn-S0	169.60	176.20	222.92	155.16	198.63	162.92	151.81	184.69	157.25	154.66
	Cn-S1	298.30	314.70	356.70	295.41	375.03	313.96	296.25	356.28	306.30	301.20
	Cn-S2	427.30	453.30	490.48	435.65	551.44	465.00	440.69	527.86	455.35	447.74
	Cn-S3	556.70	592.20	624.25	575.90	727.85	616.04	585.12	699.45	604.40	594.26
	Cn-Co	482.10	507.50	516.65	526.34	661.68	566.22	538.59	600.00	521.33	512.56
0.13 µm	Cn-S0	296.80	333.50	329.50	301.80	344.69	288.40	274.52	322.83	303.83	259.90
	Cn-S1	542.10	608.00	528.07	574.78	653.97	532.26	512.46	626.24	591.40	578.82
	Cn-S2	786.90	882.30	726.64	847.75	963.26	776.12	750.41	929.66	878.96	855.74
	Cn-S3	1032.0	1157.0	925.21	1120.7	1272.5	1020.0	988.35	1233.1	1166.5	1135.7
	Cn-Co	905.40	1005.0	766.77	1024.3	1161.2	925.07	899.87	1061.0	1005.7	979.19

Delay differences figures were calculated for the two sets of measurements, and were used to make the bar plots Figure 6.18. Discussion starts with the 90 nm data (Figure 6.18a and 6.18c):

Level 1 consistently overestimates the delay, often severely, the results for case closest being the worst. In both closest and farthest cases, the estimation accuracy improves with increasing path length, to the point of being negligible for the Cin-Cout path (farthest). The simple Level 2 gives reasonably good performance, tending to underestimate delay for short paths and overestimate them for longer paths (closest and farthest). Level 3 makes poor predictions overall (closest and farthest).Level 4 shows strong path-length dependence, underestimating at first, and then overestimating with errors getting worse as the path length

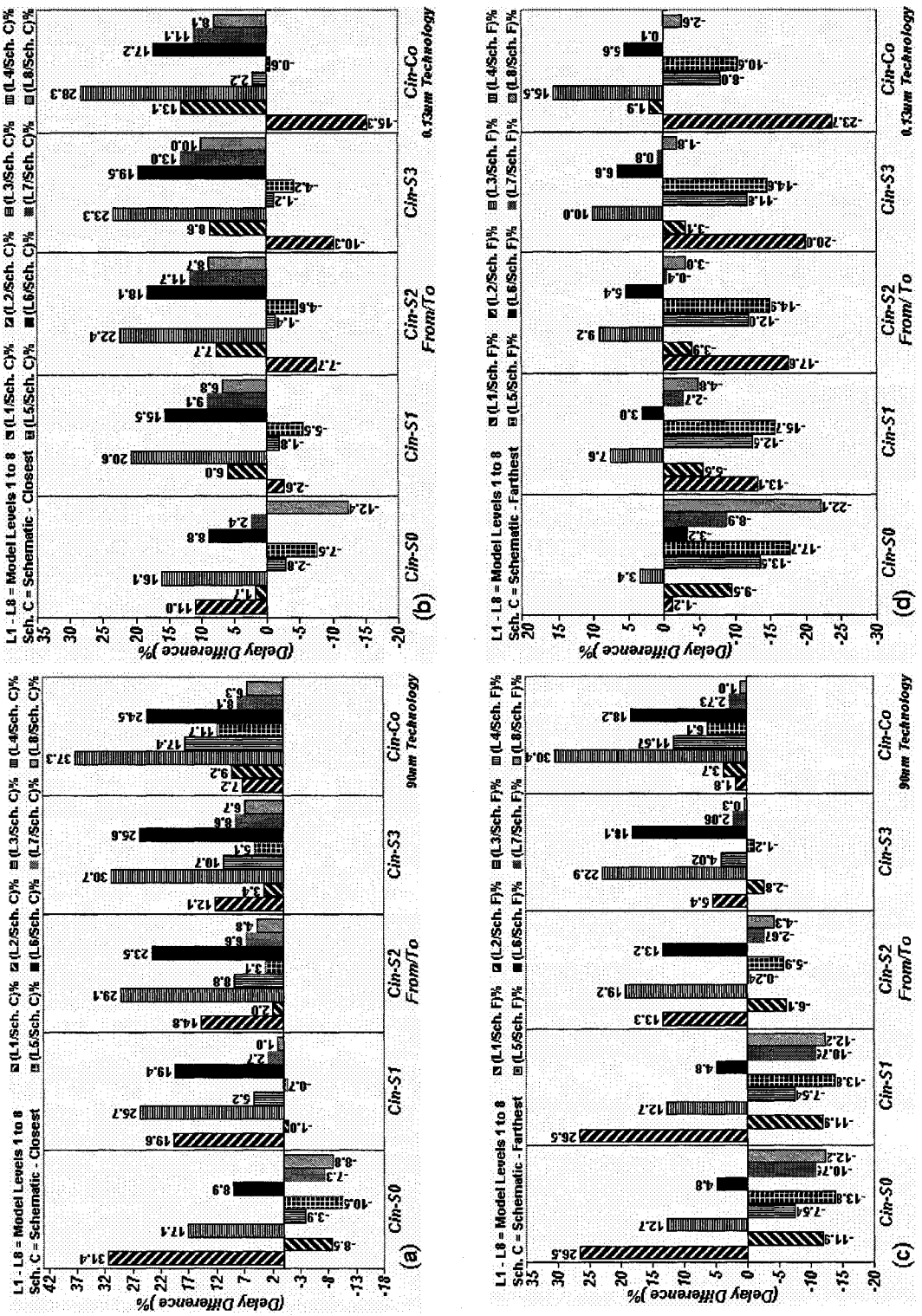


Figure 6.18: 4-Bit adder delay differences percentages according to closest and farthest connections in 90 nm (a) and (c) and 0.13 µm (b) and (d) respectively

increases (closest and farthest).Level 6 model consistently overestimates path delay, but with the degree of error seeming to stabilize as path length increases, settling down to about 25% (closest) and 18% (farthest). Levels 7 and 8 underestimate the delay of short (farthest) and very short (closest) path lengths, and then overestimate it somewhat, with the error settling down to about 8% and 7% respectively (closest) and 3% and 1% (farthest).

In case of 0.13 μm technology (Figure 6.18b and 6.18d), The Logical Effort-based method of Level 1 seriously underestimates propagation delay except for the shortest paths, with the situation getting steadily worse as path length increases. The farthest situation is worse than the closest one, but both are bad enough to make the method unusable. Level 2 model performs poorly in the closest case, and very well in the farthest case. Level 3 model continues to perform poorly overall. The path-length dependence of Level 4 it produces very good results in the closest case. In the farthest case, delays are moderately underestimated, but consistently so. Delay differences (underestimation) arising through the use of the Level 5 model would probably make it unusable in farthest situations, but they are small and decrease with path length in the closest case. The Level 6 model has the same difficulties noted in the 90 nm plots, but to a somewhat lesser extent. The difference attending the use of Levels 7 and 8 are consistent and small (closest), and very small (farthest).

The dependence of relative error on path length for models Levels 6, 7 and 8, for closest and farthest connections, is summarized in Figures 6.19 (90 nm) and 6.20 (0.13 μm).

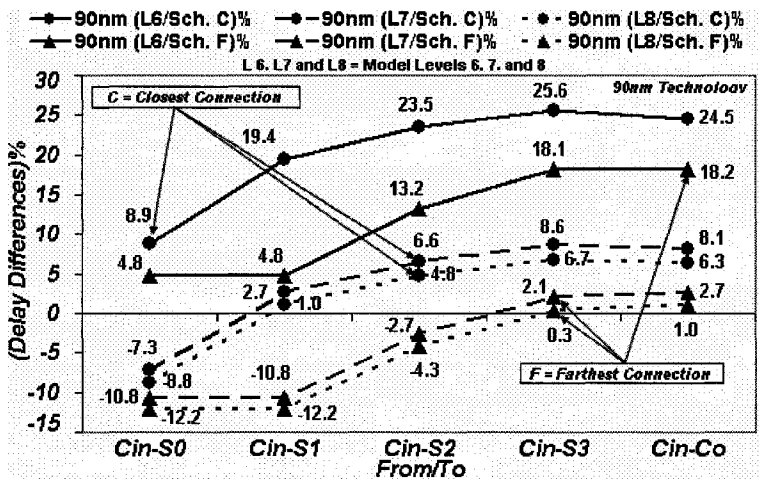


Figure 6.19: 4-Bit Adder: Delay differences using model Levels 6, 7, 8 in closest and farthest connection cases. (90 nm technology)

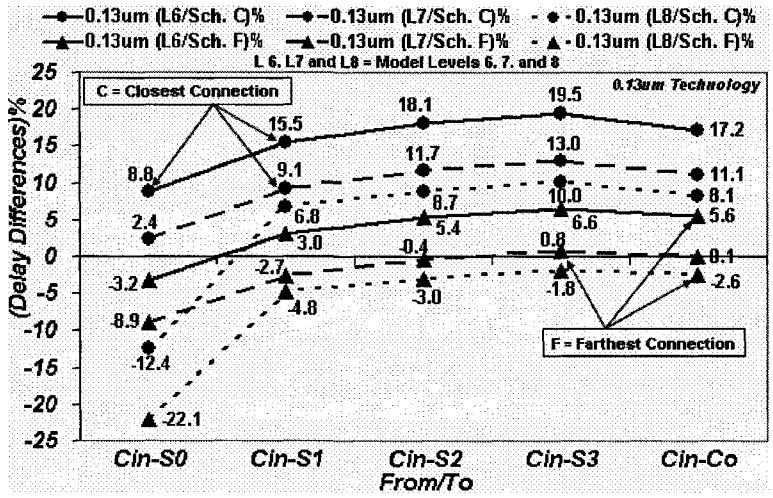


Figure 6.20: 4-Bit Adder: Delay differences using model Levels 6, 7, 8 in closest and farthest connection cases. (0.13 μm technology)

Of the three, Level 7 is the one giving the minimum average delay difference of 5.82% (90 nm process) and 2.58% (0.13 μm).

6.7 Summary

The performances of eight delay models when computing delays along twenty five different signal paths in five test applications were evaluated. In each case, the performance was expressed in terms of a normalized delay difference to a circuit simulator's result for the same delay ((Model delay – Simulator delay)/Simulator delay); the simulator result was accepted as absolutely accurate.

Test applications were designed to investigate model performance in terms of path length, architecture, process technology and SCM length, as independently as possible.

Application-1 was a 16-bit ripple-carry adder, built from NOT, NAND, NOR, and XOR gates and having critical path SCM lengths no more than 2. Paths starting at Cin and ending with S0, S1… S15 and Cout were studied (17 paths in all).

Applications 2 and 3 were a 6x64 and a 8x256 NAND decoders, respectively. In each application, the same signal path was studied in all tests. Off-path structures (branching paths) having 1, 3, 15, and 63 gates were attached to the path under study. Acceptably accurate results could be obtained for both decoders. Two longest-delay paths were investigated, one for each decoder. SCM length = 2.

Application-4 was an 8-bit comparator whose paths of study were made of both NOT, NAND and NOR gates. SCM length = 2.

Application-5 differed from the first four in that the paths incorporated devices having SCM lengths of 2, 3 and 4. This one allowed the study of the effects of having the path pass through the SCMs at different places. In particular, calculations were done for two extreme cases, one in which Cin was always connected to the SCM transistor nearest the gate output node, and the other in which it was connected to the farthest one.

Chapter 7

Results and Discussion

Compared to the simulator, the different delay model levels that applied on the applications in Chapter 6 show a range of matching from relatively far to very close matching of less than 10% differences.

In this Chapter, these results will be compared to the corresponding simulator readings in two directions, speed and accuracy.

7.1 Model versus Simulation Speed Comparison

The elapsed time for simulating and calculating the delay for each application are listed in Table 7.1. The elapsed time is obtained from simulating and calculating in a UNIX environment on sparc Sun-Blade-155 machine.

In both used technologies, the investigation of simulator elapsed time tells that Application-1, the 16-bit ripple-carry adder, took the most simulation time, to be expected since the delay time on all 17 paths from the carry-in to each of the outputs was computed. (The length of the paths increased from the Cin to S0 minimum to a maximum with the Cin-S15 and Cin-Co paths.)

Applications 3 and 5 took the next largest simulation time. In Application-3, there were 82 branching gate, while Application-5 involved propagation delays along a total of five paths of differing length. Applications 2 and 4 required the least time – these were relatively simple cases with 8 gates with 19 branching gates (Application 2), and a simple path of 16 gates (Application 4).

While the times associated with model-based delay calculation are much smaller than the ones obtained by simulation, their magnitudes not unexpectedly follow the same sequence. The important thing is that the models are faster than the simulator by orders of magnitude. In terms of delay estimation speed, model delay Level 1 which employs five parameters only (see Table 7.1) is the recommended model.

Table 7.1: Elapsed times for five application circuits according to the simulator and delay models for 90 nm and 0.13 μm Technologies

Application			*Elapsed Time (s)*				
			16-Bit RCA	*6x64 Decoder*	*8x256 Decoder*	*8-Bit Comparator*	*4-Bit NAND Adder*
90 nm Technology	*Simulator*		*74.2400*	*4.0300*	*13.5100*	*4.6600*	*10.2900*
	Model	*Level 1*	0.1650	0.0516	0.0787	0.0362	0.0685
		Level 2	0.3848	0.1468	0.1978	0.1155	0.1793
		Level 3	0.2834	0.1996	0.1660	0.1576	0.1663
		Level 4	0.2875	0.1111	0.1882	0.0944	0.1167
		Level 5	0.2764	0.1187	0.2217	0.1175	0.1679
		Level 6	0.3075	0.1957	0.2033	0.1882	0.1531
		Level 7	0.3447	0.1384	0.2091	0.2082	0.1895
		Level 8	0.3730	0.1639	0.2273	0.1672	0.1991
		Average	*0.3028*	*0.1282*	*0.1865*	*0.1356*	*0.1551*
0.13 μm Technology	*Simulator*		*128.3100*	*4.9100*	*15.5700*	*5.9500*	*16.7600*
	Model	*Level 1*	0.1844	0.0469	0.0869	0.0616	0.1086
		Level 2	0.4093	0.1602	0.2149	0.1249	0.1678
		Level 3	0.2901	0.1653	0.1845	0.1779	0.2264
		Level 4	0.3041	0.1732	0.1632	0.1054	0.1565
		Level 5	0.2872	0.1746	0.2216	0.1258	0.1807
		Level 6	0.3128	0.1551	0.1621	0.1736	0.1615
		Level 7	0.3474	0.1708	0.1758	0.2038	0.1822
		Level 8	0.3893	0.1543	0.1743	0.1868	0.1944
		Average	*0.3156*	*0.1501*	*0.1729*	*0.1450*	*0.1723*

7.2 Model versus Simulation Accuracy Comparison

Based on Tables 6.1 to 6.6, the delay differences with these of simulations in each application is calculated and plotted in Figure 7.1 for 90 nm technology and Figure 7.2 for 0.13 μm. The delay differences in 16-Bit RCA and 4-Bit NAND adders are calculated from Carry-in, Cin, input signal to Carry-out, Co. This path is the longest in each adder. For the 4-Bit NAND adder, the delay difference is plotted according to the closest and farthest

connection of the Cin signal to the output node in series-connected transistors along the Cin-Co path.

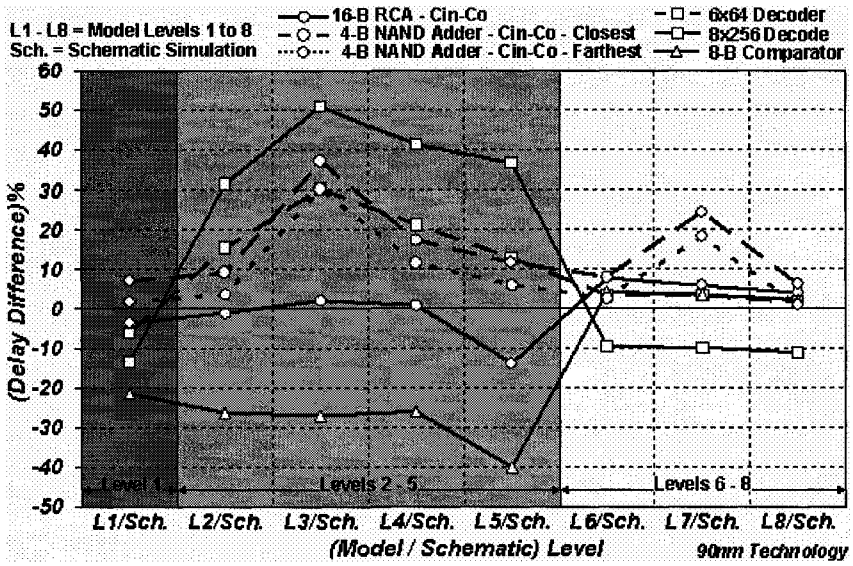


Figure 7.1: Model to simulator delay error of the five applications in 90 nm technology considering the longest path in adders' applications

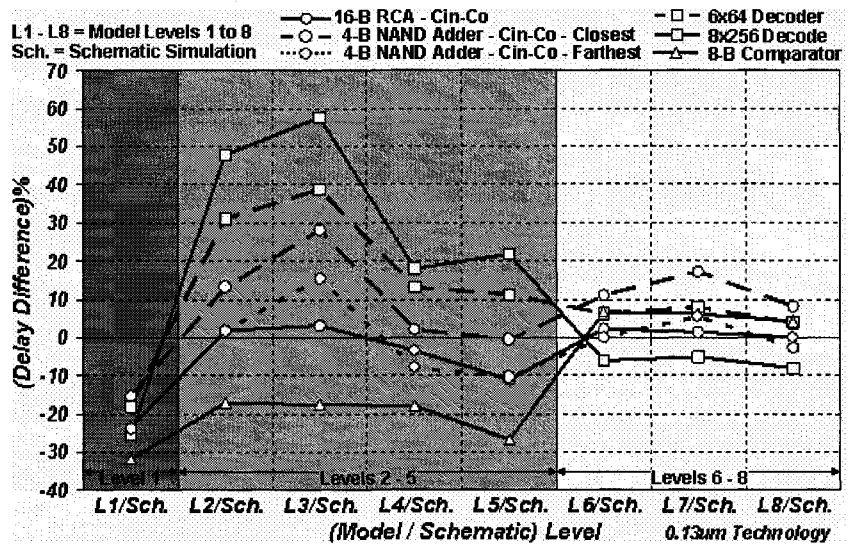


Figure 7.2: Model to simulator delay error of the five applications in 0.13 µm technology considering the longest path in adders' applications

Each plot is divided into three groups. The first for model Level-1, based on the logical effort technique, the second is assigned to model Levels 2 to 5, and the third to model Levels 6, 7 and 8.

It is seen that the most favourable results for all five applications implemented in both technologies appears in Group 3. Technology-influenced differences are seen in Groups 1 and 2: 90 nm technology gives better results in Group 1 than in Group 2, a situation that is reversed for the 0.13 μm process.

A numerical view of the plots illustrated by Figures 7.3 and 7.4 is given in Table 7.2 where the minimum, maximum, and average delay differences (model *vs.* simulation) are listed by group and technology.

Table 7.2: Maximum, minimum, and average delay difference in the three groups in 0.13 μm technology considering the longest path in adders' applications

Tech-nology	*Delay Difference Group 1*			*Delay Difference Group 2*			*Delay Difference Group 3*		
	Max.	*Min.*	*Ave.*	*Max.*	*Min.*	*Ave.*	*Max.*	*Min.*	*Ave.*
90 nm	-21.5%	1.8%	9.0%	51.0%	0.9%	21.1%	24.5%	1.0%	7.1%
0.13 μm	-31.8%	-15.3%	23.0%	57.7%	-0.6%	17.5%	17.2%	0.1%	5.7%

As mentioned, figure 7.1 and 7.2 consider the longest paths in the adders' applications. Figures 7.3 (90 nm technology) and 7.4 (0.13 μm technology) consider the shortest paths in the adders' applications. In each adder, the shortest path is from Cin to S0.

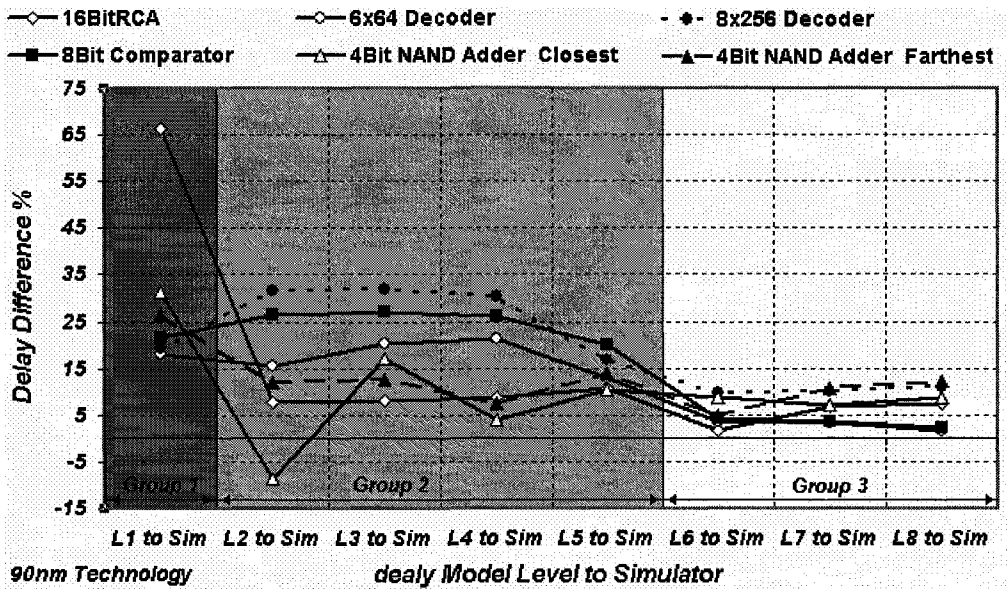


Figure 7.3: Model to simulator delay error of the five applications in 90 nm technology considering the shortest path in adders' applications

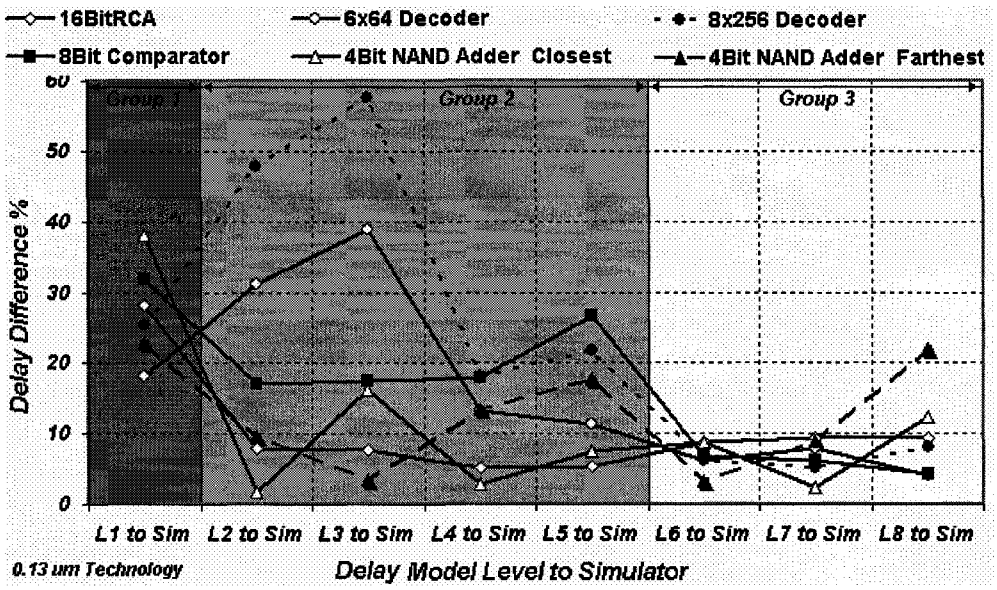


Figure 7.4: Model to simulator delay error of the five applications in 0.13 μm technology considering the shortest path in adders' applications

In both used technologies, considering the shortest path in adders' applications still keeps Group 3 as the group of the lowest delay error (the closest to the simulator). This leading is followed by Group 2 while Group 1 gives the largest delay error.

Again, a numerical view of the plots illustrated by Figures 7.3 and 7.4 is given in Table 7.3 where the minimum, maximum, and average delay differences (model *vs.* simulation) are listed by group and technology.

Table 7.3: Maximum, minimum, and average delay difference in the three groups in 0.13 μm technology considering the shortest path in adders' applications

Tech-nology	*Delay Difference Group 1*			*Delay Difference Group 2*			*Delay Difference Group 3*		
	Max.	*Min.*	*Ave.*	*Max.*	*Min.*	*Ave.*	*Max.*	*Min.*	*Ave.*
90 nm	66.5%	-6.1%	36.6%	51.0%	-3.9%	16.1%	-12.2%	1.7%	6.6%
0.13 μm	47.9%	1.7%	27.4%	57.7%	-2.8%	17.4%	-22.1%	2.4%	7.8%

7.3 Discussion

For both technologies, the average errors of each group that shown in Table 7.3 and 7.4 are plotted in Figure 7.5. Figure 7.5 (a) is considering the longest delay path in the

adders' applications while Figure 7.5 (b) is considering the shortest delay path in the adders' applications.

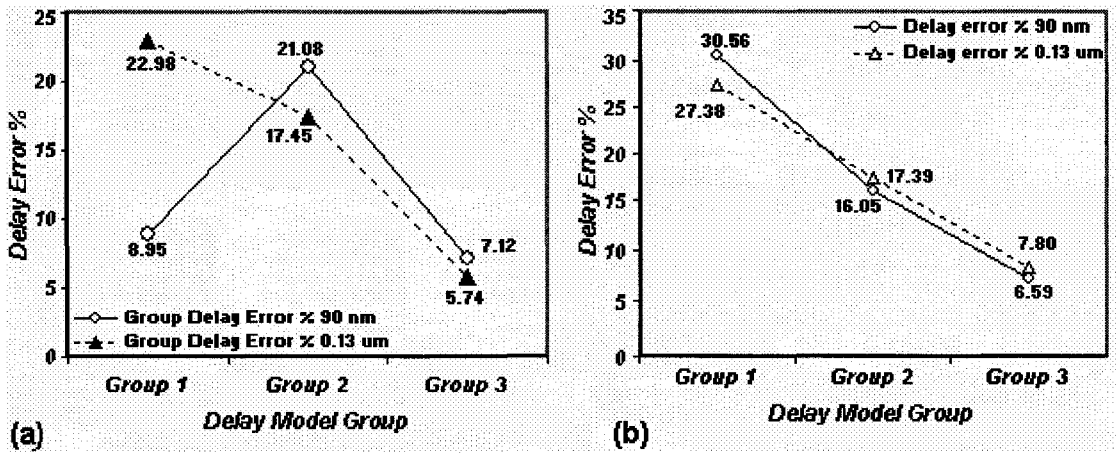


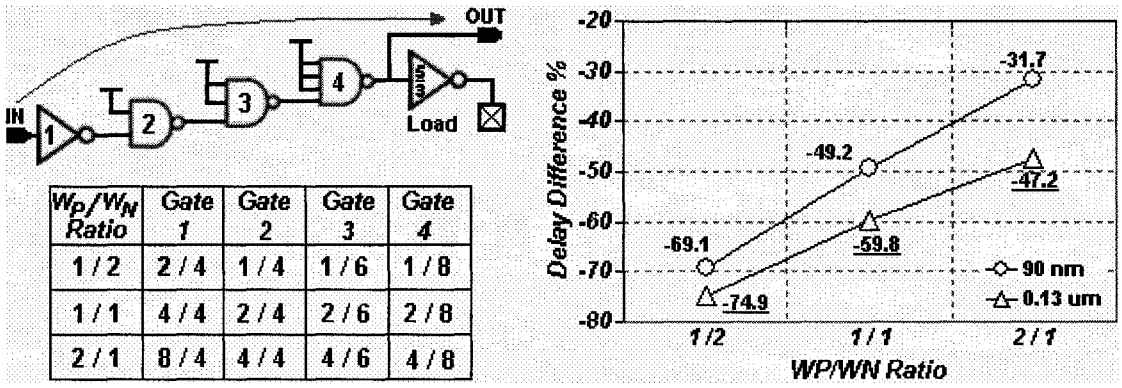
Figure 7.5: Delay error vs. Group considering the longest delay path in the adders' applications (a) and the shortest delay path in the adders' applications (b)

It can be seen that 0.13 μm technology succeeded to maintain the same sequence of delay error from the lowest (strongly matches the simulator) to the highest (weekly matches the simulator) for the longest and shortest delay path in the adders' applications. The delay error sequence from highest to lowest is Group 1 followed by Group 2 and lastly Group 3.

90 nm technology keeps Group 3 as the group of the lowest delay error and switches Group 2 and 3. For the longest delay path in the adders' applications, Group 1 is the second lowest delay error and Group 2 is giving the highest error. For the shortest delay path in the adders' applications, Group 2 is giving the second lowest error followed by Group 1.

One of the reasons that delay model Level 1 (in few occasions) is giving relatively close estimation compared to the simulator is that in all applications, the PMOS to NMOS size ratio is based on their 2/1 equivalent inverter. This centers the gates to give approximately equal rise and fall delay.

In Figure 7.6 (a), we showed a delay path that the gates on are sized based on three W_P/W_N ratio, 1/2, 1/1, and 2/1. Figure 7.7 (b) showed that logical effort technique is sensitive to this ratio and as we go from 1/2 ratio to 2/1 ratio, the difference between the simulator measurements and model calculations is improved.



W_P/W_N Ratio	Gate 1	Gate 2	Gate 3	Gate 4
1 / 2	2 / 4	1 / 4	1 / 6	1 / 8
1 / 1	4 / 4	2 / 4	2 / 6	2 / 8
2 / 1	8 / 4	4 / 4	4 / 6	4 / 8

(a) Delay path **(b) Delay difference vs. W_P/W_N ratio**

Figure 7.6: A delay path with gates W_P/W_N values in each ratio level (a) and delay difference vs. W_P/W_N ratio in both used technologies

It seems that increasing the level of input signal slope complexity from delay model Level 7 (linear) to Level 8 (quadratic and quadratic) does not provide a significant delay error improvement. This can be observed in Figure 7.7. This observation strengthens the result that we got in Section 3.7, Figure 3.9; there, we stated that as the device channel length shrinks, the 1st order slope increases and the 2nd order input slope decreases.

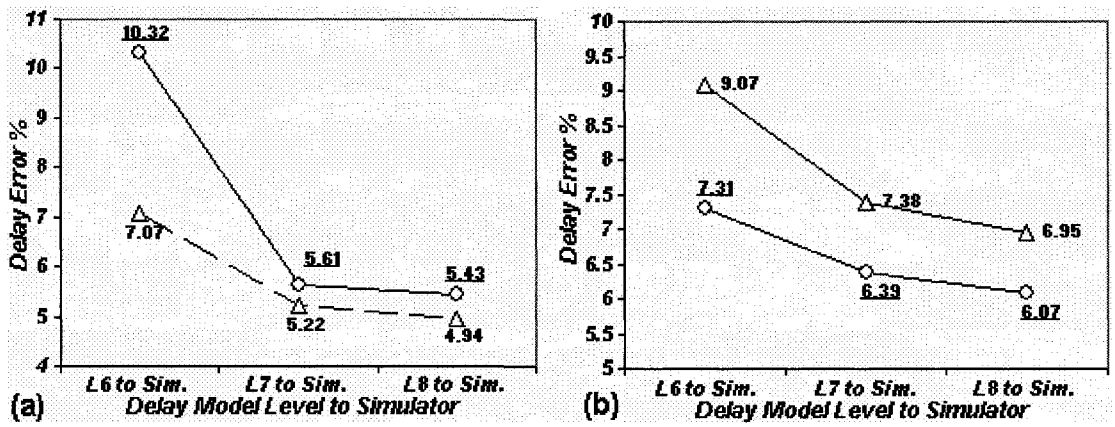


Figure 7.7: Delay error vs. delay model Levels 6 to 8 with respect to simulator considering the longest delay path in the adders' applications (a) and the shortest delay path in the adders' applications (b)

Chapter 8

Conclusions

Operation speed is one of the most important goals when designing VLSI. The issue of speed is to a very large extent one of delay along signal paths, the accurate modeling is therefore a crucially important part of the design process. Although commercial simulators like SPICE produce accurate and detailed information, they do so very slowly and a need for fast and acceptably accurate methods of path delay estimation is becoming ever more pressing as circuit sophistication increases

In this dissertation, seven delay models were investigated and compared with increasing level of details. These, and the already established method of Logical Effort were applied to five illustrative applications; the outcomes as to reliability and accuracy always compared with ones obtained by detailed simulations involving two important deep sub-micrometer technologies – 90 nm and 0.13 μm.

The extent of agreement varied from one level to another, and showed some dependence upon the nature of the application. The closest correspondences to simulator results were observed with the Level 6, 7, and 8 models

Beside schematic simulations, for the first application, a 16-Bit Ripple Carry adder, the simulation was performed on the fully custom layouts in 90 nm and 0.13 μm technologies.

8.1 Overview

In this section, the thesis is reviewed in four subsections on MOSFET current, MOSFET and wire capacitances, delay modeling, and applications.

8.1.1 MOSFET Saturation Current

Based on [24] and [29], This work proposed an extended MOSFET saturation current model which takes care of three effects that strongly affect the operation of deep sub-micrometer MOSFETs.

$$I_d = \kappa_t . W . (V_{gs} - V_t)^{(Z + \frac{M}{V_{gs}})} . (1 + \lambda . V_{ds})$$

where W is the effective width of the device, κ_t is a process and device length dependent parameter. Z is the velocity saturation effect parameter, and M is the mobility degradation effect parameter. Finally, λ is the channel-length modulation parameter.

8.1.2 MOSFET and Wire Capacitances

MOSFET capacitances are extracted in five different classes. The first four are ones proposed in this work, and the fifth is extracted based on [24]. At Class 1, a single general capacitance is used to calculate all MOSFET capacitances, PMOS and NMOS. At Class 2, the single capacitance of Class 1 is replaced by two, one for n-channel transistors and the other for p-channel devices. Further subdivision is done at Class 3. This time, types of capacitances are distinguished, gate capacitance and diffusion capacitance. Continuing in the same vein, Class 4 extraction is based on the recognition of four MOSFET capacitance types, a gate capacitance and a diffusion capacitance for each of the two transistor types. The fifth class, developed earlier by Shams [24], doubles the number of capacitance types in Class 4 (to a total of eight), this time according to transition direction, rising or falling.

Interconnect capacitances must also be dealt with, and in this work is proposed a simple model to compute them. The worst-case interconnect capacitances per unit length in 90 nm and 0.13 μm technologies are calculated to be 0.073 fF and 0.089 fF/μm respectively. In all applications, an interconnect capacitance of 10 μm length metal 1 of minimum width wire is assumed.

8.1.3 Delay Modeling

Eight different levels of delay model are used to estimate the signal path delay. Level 1 delay model, the first of eight, is based on the method of Logical Effort [56].

Table 8.1 give a summary of these models and the captured effects with each model.

Level 2 to Level 5 constructs developed in this work are based on MOSFET width, resistance and capacitance.The delay in these four levels is estimated based on single logic gate. The level of detail in the capacitance models used in developing them increases with the level of the delay calculation (see Table 8.1).

Delays in Level 6, 7, and 8 are caculated based on three cascaded gates technique [24]. Level 6 version of the three models uses only the simplest approximation for the effect of series-connected transistors in the path. It consideres the actual number of transistors in

Table 8.1: The eight delay models summary with the captured effects for each.

Delay Model	*Capacitance*	*Input Slope*	*Serial MOS Connection*	*# of Parameters*	*Comments*
Level 1	None	None	Linear	4	Based on Logical Effort Technique [56]
Level 2	One general MOSFET capacitance	None	Linear	5	The input Slope is implemented within the capacitance value
Level 3	One PMOS and one NMOS capacitance	None	Linear	6	The input Slope is implemented within the capacitance value
Level 4	One gate and one diffusion capacitance for both NMOS and PMOS	None	Linear	6	The input Slope is implemented within the capacitance value
Level 5	One gate and one diffusion capacitance for each of NMOS and PMOS	None	Linear	8	The input Slope is implemented within the capacitance value
Level 6	One gate and one diffusion capacitance for each of NMOS and PMOS in each input signal transion	Linear	Linear	24	Based on Shams Model [24]
Level 7	One gate and one diffusion capacitance for each of NMOS and PMOS in each input signal transion	Linear	Empirical Linear	32	Shams Model [24]
Level 8	One gate and one diffusion capacitance for each of NMOS and PMOS in each input signal transion	Linear and Quad-ratic	Empirical Linear	34	Modified from Shams Model [24]

the series and multiplies that by the channel resistance of a single MOSFET to obtain the overall effective resistance. Level 7 does consider some empirical factors when calculating the delay in a path containing serially connected transistors. The model is based on Elmore's approximation [24]. Level 8 estimates the delay the same way as in Level 7. The difference is when handling the input signal slope effect. Level 8 assumes a second-order quadratic expression while Level 7 assumes a first-order linear model for the delay including input slope effect. Level 8 is an extention of level 7 [24] that is proposed in this work.

8.1.4 Applications

A comparison with reality of the methods described herein, the verification of the original proposals and of the adaptations and modifications to existing concepts and techniques, was done by studying five mainstream applications. The application circuits were chosen to determine the predictive accuracy to be expected when using the new method in everyday work.

Four of these applications have a maximum of two series connected transistors in their signal paths. The fifth was designed to have a variable number of SCMs (two, three and four) in the path under investigation.

It is a rare signal path that does not have other, non-path, circuitry branching off from it, and two of the applications were designed with the study of the effects of branching circuitry in mind.

While three of these applications involve delays along the longest single path only, the other two compute the delays along a number of paths preparatory to comparison of path-to-path delay differences between the simulator and different delay model levels.

In addition to allowing a general appreciation of model and simulation differences, the applications provided an environment for evaluating the different delay models against one another.

8.2 Thesis Contributions

1 - An extended MOSFET saturation current model has been proposed and investigated.

- It has been approved that as the technology shrinks, the value of channel-length modulation factor, λ, increases either for PMOS or NMOS.

2 - Several CMOS delay models have been presented and studied. The models, labeled Level 1 to 8, gradually use more capacitances.

- Some delay models are from the literature and some are introduced in this work as modified or diluted versions of literature models. Modifications include second-order input slope effect.
- The delay models such as Shams's were applied to the older 0.5 μm technology before. We have made sure that the models are still valid and applicable to the more advanced technologies through our work here.
- The models are different in including the details regarding MOSFET capacitances, effect of input slope and effect of serially connected MOSFETs.

2 - For each delay model, the corresponding parameters are extracted for a couple of deep sub-micrometer CMOS technologies, 90 and 130 nm.

3 - Five different application circuits are designed to compare the accuracy of the models with each other and with a circuit simulator.

4 - Based on the work presented in this thesis, we can conclude the following important points:

- Including the channel-length modulation effect in estimating the saturation current for MOSFETs improves the accuracy only marginally.
- The accuracy of the model is generally improved by considering more capacitance parameters. Gradual improvement is illustrated between levels 2, 3, 4, 5, and 6.
- Group 3 delay model levels (Level 6, 7, and 8) seem to be giving fairly accurate results compared to simulation, while they do not impose the computational complexity of the simulators.
- As we move deeper into sub-micrometer CMOS region, the delay of serially connected MOSFETs seems to show more linear (rather than quadratic) dependence to the number of devices.
- As the CMOS device features keep shrinking, the dependence of the delay on input slope demonstrates a more linear relation. This is the difference between Level 7 and Level 8.

8.3 Future Work

This work may be extended in many ways, some of which are listed:

- Considering an application that includes extreme and exaggerated cases concerning the input slope and the number of serially connected MOSFETs.
- Extending the study to more advanced CMOS technologies to confirm or perhaps disapprove the calculations made here.
- Investigating the intuitive and physical reasons behind our conclusions, e.g., regarding the effect of serially connected MOSFETs and the input slope effect.
- Building a CAD tool based on the results of this work.
- Incorporating some non-ideal behaviors that are currently ignored in the presented models. These include the internal nodes capacitances and body effect.

We are not yet completely sure why the more complicated models sometimes give less accurate results compared to the simpler models. This could be because of the internal nodes capacitance. Perhaps the advantages of the more complicated models become more evident with more examples and considering some extreme scenarios.

References

[1] J. M. Rabaey, *Digital Integrated Circuits*. Prentice Hall, 2003.

[2] S. Chin-Tang, "Evolution of the MOS Transistor-from Conception to VLSI," *Proceeding of IEEE*, vol. 76, pp. 1280-1326, 1988.

[3] J. J. Lious and F. Schwierz, "RF MOSFET: Recent Advances and Future Trends", *IEEE Conference on Electron Devices and Solid-State Circuits*, pp. 185-192, 2003.

[4] M. Thumm, "Historical German Contributions to Physics and Applications of Electromagnetic Oscillations and Waves", ", *Int. Conference on Progress in Nonlinear Science*, Nizhny Novgorod, Russia, 2001.

[5] R.G. Arns, "The other transistor: Early History of the Metal-Oxide Semiconductor Field-Effect Transistor," *Engineering Science and Education Journal*, vol. 7, pp. 233-240, 1998

[6] W. F. Brinkman, E. Haggan and W. troutman, "A History of the Invention of the Transistor and Where It Will Lead Us," *IEEE Journal of Solid-State Circuits*, vol. 32, pp. 1858-1865, 1997.

[7] "In Memoriam – Jack Kilby (1923 – 2005) Inventor of the Integrated Circuits," *IEEE Signal Processing Magazine*, vol. 22, pp. 89-92, 2005.

[8] A. Yu, "The Future of Microprocessors," *Digital Object Identifier*, vol. 16, pp. 45-53, 1995.

[9] S. Gochman, "The Intel® Pentium®_Processor: Microarchitecture and Performance," *The Intel Technology Journal*, vol. 7, pp.21-36, 2003.

[10] Itanium Solutions Alliance, "Why Intel® Itanium® 2 Microprocessor Is Ideal for Multi-Core Performance Scaling," www.itaniumsolurionsalliance.org, 2006

[11] D. A. Johns and K. Martin, *Analog Integrated Circuit Design*. John Wiley & Sons Inc., 1997.

[12] S. M. Sze, *Physics of Semiconductor devices*. John Wiley & Sons Inc., 1981.

[13] W. Shockley, "A Unipolar Field Effect Transistor," *Proc. IRE*, vol. 40, pp.1365-1376, November 1952.

[14] J. A. Hutchby, V. V. Zhirnov, R. K. Cavin and G. I. Bourianoff, "Functional Scaling Beyond Ultimate CMOS," *IEEE 7th International Conference on Solid-State and Integrated Circuits Technology,* vol. 1, pp. 234-239, 2004.

[15] K.K Ng and W.T Lynch, "The Impact of Intrinsic Series Resistance on MOSFET Scaling," *IEEE Transactions on Electronic Devices*, vol. 34, pp.503-511, 1987.

[16] R. H Dennard and F. H. Gaensslen, "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions," *Proceeding of IEEE*, vol. 87, pp. 668-678, 1999.

[17] Yuan Taur and D. A. Buchanan, "CMOS Scaling into the Nanometer Regime;" *Proceedings of the IEEE*, vol. 85, pp. 486-504, 1997.

[18] Sung-Mo Kang and Yusuf Leblebici, *CMOS Digital Integrated Circuits.* McGraw Hill, 2003.

[19] Seongmoo Heo and K. Asanovic, "Load-Sensitiv Flip-Flop Characterizations," *IEEE Computer Society Workshop on VLSI*, pp. 87-92, 2001.

[20] Harry Vendrick, *Deep-Submicron CMOS ICs from Basics to ASICs.* Kluwer BedrijsfInformatie, 1998.

[21] E. Pop, S. Sinha and K. E. Goodson, "Heat Generation and Transport in Nanometer-Scale Transistors," *Proceedings of the IEEE*, vol. 94, pp. 1587-1601, 2006.

[22] K. Ko. Ping, "Approaches to Scaling," *Advanced MOS Device Physics*, vol. 18, pp.1-37, 1989.

[23] Y. Tsividis, *Operation and Modeling of the MOS Transistor.* McGraw Hill, 1999.

[24] M. Shams, "Modeling and Optimization of CMOS Logic Circuits with Application to Asynchronous Design," *PhD thesis*, University of Waterloo, 1999.

[25] K. Lee, M. Shur, T. A.Fjeldly and T. Ytterdal, *Semiconductor Device Modeling for VLSI*, Prentice-Hall, 1993.

[26] Y. Cheng and C. Hu, *MOSFET Modeling & BSIMS3 User's Guide*, Kluwer Academic Publishers, 1999.

[27] B. Sheu, D. Scharfetter, P. Ko, and M. Jeng, "BSIM: Berkely Short-Channel IGFET Model for MOS Transistors," *IEEE Journal of Solid-State Circuits*, vol. 22, pp. 558-556, 1987.

[28] K. Toh, P. Ko and R. G. Meyer, "An Engineering Model for Short-Channel MOS Devices', *IEEE Journal of Solid-State Circuits*, vol. 23, pp. 950-958, 1988.

[29] T. Sakurai and A. R. Newton, "Alpha-power Law MOSFET Model and its Applications to CMOS Inverter Delay and Other Formulas," *IEEE Journal of Solid-State Circuits*, vol. 25, pp. 584-594, 1990.

[30] E. Damodaran, "Delay Models for CMOS, BiMOS, and BiNMOS Circuits and Their Applications for Timing Simulations," *IEEE Transactions on Computer-Aided Design.* Vol. 13, pp. 1132-1142, 1994.

[31] L. Bisdounis, S. Nikolaidis, and O. Koufopavlou, " Analysis Transient response and Propagation Delay Evaluation of the CMOS Inverter for Short-Channel Devices," *IEEE Journal of Solid-State Circuits*, vol. 33, pp. 302-306, 1996.

[32] C. Forzan, B. Franzini, and C. Guardiani, "Accurate and Efficient Macromodel of Digital Standard Cells," Proceedings of the 34th Design Automation Conference, pp. 633 – 637, 1997.

[33] M. Shams, "A Unified Delay Model for CMOS Logic Styles," IEEE International Conference on Electronics, Circuits and Systems, vol. 2, pp. 874-877, 2003.

[34] M. Shams and M. Elmasry, "Delay Optimization of CMOS Logic Circuits using Closed-Form Expressions," *IEEE International Conference on Computer Design*, pp. 563-568, 1999.

[35] N. H. E. Weste and K. Eshraghian, *Principles of CMOS Design, A System Perspective*. Addison-Wesley Publication Co., 1993.

[36] E. Barker, "Line-to-Ground Capacitance Calculation for VLSI: A Comparison," *IEEE Transactions on Computer-Aided Design*, vol. 7, pp. 295-298, 1988.

[37] R. Gu and M Elmasry, "Power Dissipation and Optimization of Deep Submicron CMOS Digital Circuits," *IEEE Journal of Solid-State Circuits*, vol. 31, pp. 707-713, 1996.

[38] S. Embabi and R. Damodaran, "Delay Models for CMOS, BiCMOS and BiNMOS Circuits and Their Applications for Timing Simulation," IEEE Transaction on Computer-Aided Design of Integrated Circuits and Systems, vol. 13, pp. 1132-1142.

[39] W. Elmore, "The Transient Response of Damped Linear Networks with Particular Region to Wideband Amplifiers," *Journal of Applied Physics*, vol. 19, pp. 55-63, 1948.

[40] P. Cocchini, G, Piccinini and M. Zamboni, "A Comprehensive Submicrometer MOST Delay Model and its Application To CMOS Buffers," *IEEE Journal of Solid-State Circuits*, vol. 32, pp. 1254,1262, 1997.

[41] N. Hedenstierna and K. Jeppson, "CMOS Circuit Speed and buffer Optimization," IEEE Transaction on Computer-Aided Design, vol. 6, pp. 270-281, 1987.

[42] T. Sakurai and A. R. Newton, "Delay Analysis of Series-Connected MOSFET Circuits," *IEEE Journal of Solid-State Circuits*, vol. 26, pp. 122-130, 1991.

[43] J. Kong and D. Overhauser, *Digital Timing Macromodeling for VLSI Design Verification*. Kluwer Academic Publishers, 1995.

[44] H. Chen and S. Dutta, "A timing Model for Static CMOS Gates," *IEEE International Conference on Computer-Aided Design*, pp. 72-75, 1989.

[45] A. Nabavi-Lishi and N. Rumin, "Inverter Models of CMOS Gates for Supply Current and Delay Evaluation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 13, pp. 1271-1279, 1994.

[46] O. Xiaoning, A. Gyure, L. Yansheng, S. C Lo, M. Shahram, and K. Singhal, "Simulation of interconnect inductive impact in the presence of process variations in 90 nm and beyond," IEEE electron device letters, vol. 27, pp. 696-698, 2006.

[47] IBM Corporation, *CMRF8SF Model Reference Guide*. IBM, 2006.

[48] P.Schmidt, "A Nanoelectrode Lined Nanochannel for Single-Molecule DNA Sequencing," *M Sc. thesis*, Carnegie Mellon University, 2004.

[49] IBM Corporation, *CMOS8FR (CMRF8SF) Design Manual*. IBM, 2006.

[50] STMicroelectronics, PLS Flow and Solutions Training. STMicroelectronics, 2005.

[51] A. Nabavi-Lishi and N. Rumin, "Delay and Bus Current Evaluation in CMOS Logic Circuits," *IEEE Computer-Aided Design Digest and Technical Papers*, pp. 198-203, 1992.

[52] B. Cherkauer, E. Friedman, "Channel Width Tapering of Serially Connected MOSFET'swith Emphasis on Power Dissipation," *IEEE Transaction on VLSI Systems*, vol. 2, pp. 100-114, 1994.

[53] A. Chatzigeorgiou and S. Nikolaidis, "Collapsing the transistor chain to an effective single equivalent transistor," *IEEE Proceedings on Design, Automation and Test*, pp. 23-26, 1998.

[54] S. Taherzada, E. Iman, B. Amelfried, M. Farazan, A. Afzali and M. Nourian, "A Simple yet Accurate Method for Reducing CMOS Gate to Equivalent Inverters," *IEEE Symposium on Mixed-Signal Design*, pp. 112-115, 2003.

[55] A. Chatzigeorgiou, S. Spiridon, I. Tsoukalas, "A Modeling Technique for CMOS Gates," *IEEE Transaction on Computer-aided Design of Integrated Circuits and Systems*, vol. 18, pp. 557-576, 1999.

[56] I. Sutherland, *Logical Effort.* Morgan Kaufmann Publishers, Inc., 1999.

Appendix A

Current and Delay Parameters Extraction

This appendix shows how to extract the extended drain saturation current model (Equation 3.1) parameters λ, Z, M, and κ_t based on the DC analysis simulation in Section A.1 and based on AC analysis simulation in Section A.2. Section A.3 shows how the input signal slope and related parameters can be extracted. Series-connected parameters were extracted in Section A.4. Finally, input signal quadratic and linear slopes were extracted in Section A.5.

A.1 Based on DC Analysis Simulation

Here, the extraction is based on I_{ds} vs. V_{gs} simulation.

A.1.1 λ Parameter Extraction

This parameter may be extracted from measurements as following:

► Step 1:

- At $V_{gs} = V_{dd}$ (NMOS) or $V_{gs} = 0$ (PMOS) and on $I_{ds} - V_{ds}$ plot, take two points as (V_{ds1}, $Ids1$) at $V_{ds1} = V_{dd}$ and (V_{ds2}, I_{ds2}) at 90% of V_{dd}
- Write Equation 1.3 for these two points as:

$$I_{ds}1 = \kappa_t.W.(V_{gs} - V_t)^{(Z+\frac{M}{V_{gs}})}.(1+\lambda V_{ds\,1}) \qquad (A.1)$$

$$I_{ds}2 = \kappa_t.W.(V_{gs} - Vt)^{(Z+\frac{M}{V_{gs}})}.(1+\lambda.V_{ds\,2}) \qquad (A.2)$$

► Step 2:

- Division yields:

$$\frac{I_{ds1}}{I_{ds2}} = \frac{1+\lambda.V_{ds1}}{1+\lambda V_{ds2}} \qquad (A.3)$$

► **Step 3:**

- Then solve for λ:

$$\lambda = \frac{I_{ds2} - I_{ds1}}{V_{ds}2.I_{ds1} - V_{ds}1.I_{ds2}} \qquad (A.4)$$

A.1.2 *Z*, *M*, and κ_t Parameters Extraction

By setting $V_{ds} = V_{dd}$, step 4 to step 7 are used to extract Z (velocity saturation index), M (Mobility degradation index), and technology dependent parameter (κ_t):

► **Step 4:**

- Select three points on an $I_{ds} - V_{gs}$ plot: (V_{gs1}, I_{ds1}), (V_{gs2}, I_{ds2}), and (V_{gs3}, I_{ds3})
- Write Equation 3.1 for each of them:

$$I_{ds1} = \kappa_t.W.(V_{gs1} - V_t)^{(Z+\frac{M}{V_{gs1}})}.(1+\lambda.V_{dd}) \qquad (A.5)$$

$$I_{ds2} = \kappa_t.W.(V_{gs2} - V_t)^{(Z+\frac{M}{V_{gs2}})}.(1+\lambda.V_{dd}) \qquad (A.6)$$

$$I_{ds3} = \kappa_t.W.(V_{gs3} - V_t)^{(Z+\frac{M}{V_{gs3}})}.(1+\lambda.V_{dd}) \qquad (A.7)$$

► **Step 5:**

- Divide Equation 3.6 by Equation 3.7 and Equation 3.7 by Equation 3.8 and take logarithms:

$$\log\left(\frac{I_{ds1}}{I_{ds2}}\right) = \log\left(\frac{(V_{gs1}-V_t)^{Z+\frac{M}{V_{gs1}}}}{(V_{gs2}-V_t)^{Z+\frac{M}{V_{gs2}}}}\right) \qquad (A.8)$$

$$\log\left(\frac{I_{ds\,2}}{I_{ds\,3}}\right)=\log\left(\frac{(V_{gs2}-V_t)^{Z+\frac{M}{V_{gs_2}}}}{(V_{gs\,3}-V_t)^{Z+\frac{M}{V_{gs_3}}}}\right) \qquad (A.9)$$

► **Step 6:**

- Solving for Z and M will give:

$$Z=\frac{Y1.Y4-Y4.Y2}{Y3.Y5-Y6.Y2} \qquad (A.10)$$

$$M=\frac{Y1.Y6-Y4.Y3}{Y2.Y6-Y5.Y3} \qquad (A.11)$$

where:

$$Y1=\log\left(\frac{I_{ds1}}{I_{ds\,2}}\right) \qquad (A.12)$$

$$Y2=\frac{\log(V_{gs\,1}-V_t)}{V_{gs\,1}}-\frac{\log(V_{gs\,2}-V_t)}{V_{gs\,2}} \qquad (A.13)$$

$$Y3=\log(V_{gs\,1}-V_t)-\log(V_{gs\,2}-V_t) \qquad (A.14)$$

$$Y4=\frac{I_{ds2}}{I_{ds3}} \qquad (A.15)$$

$$Y6=\log(V_{gs\,2}-V_t)-\log(V_{gs_3}-V_t) \qquad (A.16)$$

► **Step 7:**

- κ_t can be found by:

$$\kappa_t=\frac{I_{ds\,2}}{W.(V_{gs}\,2-V_t)^{Z+\frac{M}{V_{gs\,2}}}.(1+\lambda V_{dd})} \qquad (A.17)$$

A.2 Based on AC Analysis Simulation

Here, the extraction is based on delay simulation according to step and ramp input signal (Figure 3.7).

A.2.1 λ, Z, M and κ_t Parameters Extraction

► **Step 1:**

- Measure sets of step delays ($D_{S\text{-}NF}$, $D_{S\text{-}NR}$, $D_{S\text{-}PF}$, $D_{S\text{-}PF}$) as D_{Si} for $i = 1 - 4$, at four different supply voltage V_1, V_2, V_3, and V_4. Sets numbered 3 and 4 are simulated with the MOSFETs fully ON, and their values used to compute the 4 values of λ:

$$\lambda = \frac{D_{S4}.V_3 - D_{S3}.V_4}{V_3.V_4.D_{S3} - V_3.V_4.D_{S4}} \qquad (A.18)$$

► **Step 2:**

- Calculate M and Z as:

$$M = \frac{b1.a2 - a1.b2}{b1.c1 - c.b2} \qquad (A.19)$$

$$Z = \frac{a2 - M.c2}{b2} \qquad (A.20)$$

where:

$$a1 = \log\left(\frac{D_{S1}.V_2.(1+\lambda.V_1)}{D_{S2}.V_1.(1+\lambda V_2)}\right) \qquad (A.21)$$

$$b1 = [\log(V2 - V_t) - \log(V1 - V_t)] \qquad (A.22)$$

$$c1 = \frac{\log(V_2 - V_t)}{V_2} - \frac{\log(V_1 - V_t)}{V_1} \qquad (A.23)$$

$$a2 = \log\left(\frac{D_{S2}.V_3.(1+\lambda.V_2)}{D_{S3}.V_2.(1+\lambda.V_3)}\right) \qquad (A.24)$$

$$b2 = \log(V_3 - V_t) - \log(V_2 - V_t) \qquad (A.25)$$

$$c2 = \frac{\log(V_3 - V_t)}{V_3} - \frac{\log(V_2 - V_t)}{V_2} \qquad (A.26)$$

► **Step 3:**

- Calculate κ_t as:

$$\kappa_t = \frac{C_L.V_2}{2.D_{S2}.(V_2 - V_t)^{\left(Z + \frac{M}{V2}\right)}.(1+\lambda.V_2)} \qquad (A.27)$$

► **Step 4:**

- Calculate v according to Equation 3.31.

A.3 Input Slope Factor and Related Parameters Extraction

► **Step 1:**

- At V_2, use a ramp input signal of τ transition time (1 ns in this work) and measure the delay (D). Calculate α as:

$$\alpha = [(1 + \frac{V_t}{V_{dd}}) \div (\frac{\tau - 2.(D - D_{S2})}{2.\tau})] - 1 \qquad (A.28)$$

► **Step 2:**

- Calculate the slope factor S as:

$$S = 2.[\frac{1}{2} - \frac{1 - V_t / V_{dd}}{1 + \alpha}] \qquad (A.29)$$

A.4 Series-Connection Parameters Extraction

► **Step 3:**

- Double the number of transistors connected in Figure 3.7, replacing each NMOS and PMOS with two in series.

► **Step 4:**

- Measure the step delay $D_{S1Series}$ and $D_{S2Series}$ at V_1 and V_2 respectively. Calculate x and y as:

$$y = \frac{c - a}{b - d} \qquad (A.30)$$

$$x = a - 1 + y.d \qquad (A.31)$$

where:

$$a = \frac{D_{S1\,Series}}{D_{S1}} \qquad (A.32)$$

$$b = \frac{V_1}{V_t} \qquad (A.33)$$

$$c = \frac{D_{S2\,Series}}{D_{S2}} \qquad (A.34)$$

$$b = \frac{V_2}{V_t} \qquad (A.35)$$

A.5 Input Signal Linear and Quadratic Slopes (S1 and S2) Extraction

- Connect two identical inverters of minimum L and $W_P/W_N = 6/3$ as in Figure A.1
- Apply a step input signal and measure the step rising and falling delay across G_i as D_{SRi} and D_{SFi}

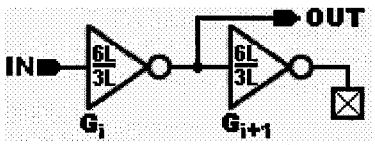


Figure A.1: Two cascaded identical inverters of minimal size

- Modify Figure A.1 as in Figure A.2

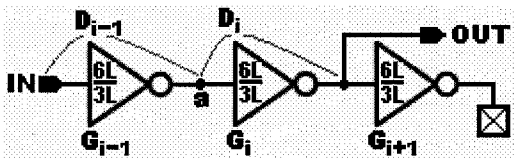


Figure A.2: Three cascaded identical inverters of minimal size

- As changing the input signal transition (for both rising and falling) from ≈ 0 s to 3 ns, measure D_{Ri}, D_{Ri-1}, D_{Fi} and D_{Fi-1}
- Plot D_{Ri} as a function of D_{Ri-1} and D_{Fi} as a function of D_{Fi-1}
- Take two points on each plot, (D_{1Ri}, D_{1Ri-1}) at small transition time and (D_{2Ri}, D_{2Ri-1}) at large transition time. Similarly, assign (D_{1Fi}, D_{1Fi-1}) and (D_{2Fi}, D_{2Fi-1})
- $S1$, $S2$ are calculated according to:

$$S2 = \frac{a - b}{c} \qquad (A.36)$$

$$S1 = \frac{a_2 - 1 - S2 \cdot b_2^2}{b_2} \qquad (A.37)$$

where:

$$a = \frac{a_2 - 1}{b_2} \qquad (A.38)$$

$$b = \frac{a_1 - 1}{b_1} \qquad (A.39)$$

$$c = b_2 - b_1 \qquad (A.40)$$

$$a_1 = \frac{D_{1i}}{D_{Si}} \qquad (A.41)$$

$$a_2 = \frac{D_{2i}}{D_{Si}} \qquad (A.42)$$

$$b_1 = \frac{D_{1i-1}}{D_{Si}} \qquad (A.43)$$

$$b_1 = \frac{D_{2i-1}}{D_{Si}} \qquad (A.44)$$

- Using (D_{1Ri}, D_{1Ri-1}) and (D_{2Ri}, D_{2Ri-1}) in Equations 3.54 to 3.61 give $S1_R$ and $S2_R$
- Using (D_{1Fi}, D_{1Fi-1}) and (D_{2Fi}, D_{2Fi-1}) in Equations 3.54 to 3.61 give $S1_F$ and $S2_F$

Appendix B

MOSFET Capacitance Extraction

This appendix is assigned to give the full procedures of extracting the MOSFET capacitances over five different sections. Each section is assigned for specific class of extraction.

B.1 MOSFET Capacitance Extraction: Class 1

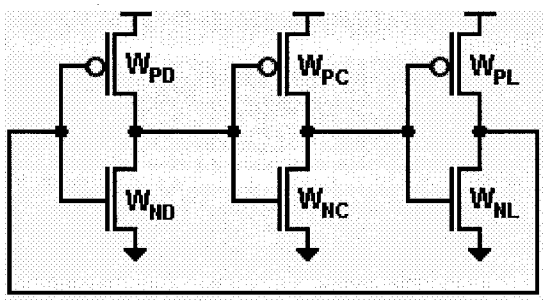


Figure B.1: Closed-loop chain of three inverters

A closed-loop chain of three inverters (Figure B.1) is used to extract the single capacitance parameters of Class 1. The extraction of C_{mosfet} takes place in the following steps:

► **Step 1:**

- Set all transistors length to L = minimum length set by used technology
- Set $W_{PD} = W_P = W_{PL} = W_{1\text{-}P} = 10L$
- Set $W_{ND} = W_N = W_{NL} = W_{1\text{-}N} = 5L$
- Measure the loop frequency as F_1

► **Step 2:**

- Calculate $I_{1\text{-}P}$ and $I_{1\text{-}N}$ and $I_{1\text{-}t}$ as:

$$I_{d1-P} = \kappa_{t-P} . W_{1-P} . (V_{gs} - V_{t-P})^{(Z_P + \frac{M_P}{V_{gs}})} . (1 + \lambda_P V_{ds}) \qquad (B.1)$$

$$I_{d1-N} = \kappa_{t-N} . W_{1-N} . (V_{gs} - V_{t-N})^{(Z_N + \frac{M_N}{V_{gs}})} . (1 + \lambda_N V_{ds}) \qquad (B.2)$$

$$I_{d1-T} = I_{d1-P} + I_{d1-N} \qquad (B.3)$$

where $\kappa_{t\text{-}P}$, Z_P, M_P, λ_P, $\kappa_{t\text{-}N}$, Z_N, M_N, and λ_N are as indicated in Table 3.3 for PMOS and NMOS respectively in both of used technologies, 0.13 μm and 90 nm.

► **Step 3:**

- Compute C_{mosfet} as:

$$C_{mosfet} = \frac{(2.I_{d1-P}.I_{d1-N})}{3.V_{dd}.F_1.I_{d1-T}.30.L} \qquad (B.4)$$

B.2 MOSFET Capacitance Extraction: Class 2

In this Class, one parameter each is used to describe the capacitances of NMOS and PMOS transistors, C_p for PMOS, and C_N for NMOS. The following steps are taken:

► **Step 1:**

- Measure F_1 and calculate $I_{d1\text{-}P}$, $I_{d1\text{-}N}$, and $I_{d1\text{-}T}$ as in step 1 and 2 in section B.1.

► **Step 2:**

- Keep all transistors length to L = minimum length set by used technology
- Set $W_{PD} = W_P = W_{PL} = W_{2\text{-}P}$ 15L
- Set $W_{ND} = W_N = W_{NL} = W_{2\text{-}N} = 5L$
- Measure the loop frequency as F_2

► **Step 3:**

- Calculate $I_{2\text{-}P}$ and $I_{2\text{-}N}$ and $I_{2\text{-}T}$ as:

$$I_{d2-P} = \kappa_{t-P}.W_{2-P}.(V_{gs} - V_{t-P})^{(Z_P + \frac{M_P}{V_{gs}})}.(1 + \lambda_P.V_{ds}) \quad (B.5)$$

$$I_{d2-N} = \kappa_{t-N}.W_{2-N}.(V_{gs} - V_{t-N})^{(Z_N + \frac{M_N}{V_{gs}})}.(1 + \lambda_N.V_{ds}) \quad (B.6)$$

$$I_{d2-T} = I_{d2-P} + I_{d2-N} \quad (B.7)$$

► **Step 4:**

- Compute C_P and C_N as:

$$C_P = X2 - X1 \quad (B.8)$$

$$C_N = X1 - 2.C_P \quad (B.9)$$

where $X1$ and $X2$ are given as:

$$X1 = \frac{(2.I_{d1-P}.I_{d1-N})}{3.V_{dd}.F_1.I_{d1-T}.10.L} \quad (B.10)$$

$$X2 = \frac{(2.I_{d2-P}.I_{d2-N})}{3V_{dd}.F_2.I_{d2-T}.10.L} \quad (B.11)$$

B.3 MOSFET Capacitance Extraction: Class 3

This class, Class 3, proposes two capacitances; a gate capacitance, C_g, and diffusion capacitance, C_d, are set for either NMOS or PMOS.

► **Step 1:**

- Measure F_1 and calculate $Id_{1\text{-}P}$, $Id_{1\text{-}N}$, and $Id_{1\text{-}T}$ as in step 1 and 2 in section B.1.

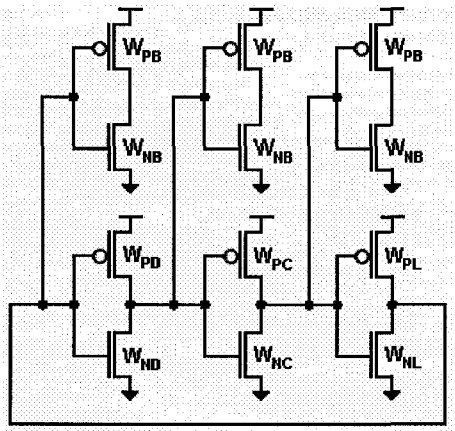


Figure B.2: Closed-loop chain of three inverters with branches

► **Step 2:**

- Load each inverter in Figure B.1 by one inverter as in Figure B.2.
- Keep all transistors length to L = minimum length set by used technology
- Set $W_{PD} = W_P = W_{PL} = W_{PB} = W_{3\text{-}P}$ $10L$
- Set $W_{ND} = W_N = W_{NL} = W_{NB} = W_{3\text{-}N} = 5L$
- Measure F_3

► **Step 3:**

- Calculate $I_{3\text{-}P}$ and $I_{3\text{-}N}$ and $I_{3\text{-}t}$ as:

$$I_{d3-P} = \kappa_{t-P}.W_{3-P}.(V_{gs} - V_{t-P})^{(Z_P + \frac{M_P}{V_{gs}})}.(1 + \lambda_P.V_{ds}) \quad (B.12)$$

$$I_{d3-N} = \kappa_{t-N}.W_{3-N}.(V_{gs} - V_{t-N})^{(Z_N + \frac{M_N}{V_{gs}})}.(1 + \lambda_N.V_{ds}) \quad (B.13)$$

$$I_{d3-T} = I_{d3-P} + I_{d3-N} \quad (B.14)$$

► **Step 4:**

- Compute C_g and C_d as:

$$C_g = X4 - X3 \quad (B.15)$$

$$C_d = X3 - C_g \quad (B.16)$$

where $X3$ and $X4$ are given as:

$$X3 = \frac{(2.I_{d1-P}.I_{d1-N})}{3V_{dd}.F_1.I_{d1-T}.15.L} \quad (B.17)$$

$$X4 = \frac{(2.I_{d3-P}.I_{d3-N})}{3V_{dd}.F_3.I_{d3-T}.15.L} \quad (B.18)$$

B.4 MOSFET Capacitance Extraction: Class 4

At this class, four MOSFET capacitance parameters are introduced, one set of two for each device type. These are denoted C_{Pg}, C_{Pf}, C_{Ng}, and C_{Nd}, their extraction is done in the following steps:

► **Step 1:**

- Measure F_1 and calculate $Id_{1\text{-}P}$, $Id_{1\text{-}N}$, and $Id_{1\text{-}T}$ as in step 1 and 2 in section B1.

► **Step 2:**

- Repeat step 2 in section B.2.

► **Step 3:**

- Calculate $I_{2\text{-}P}$ and $I_{2\text{-}N}$ and $I_{2\text{-}T}$ as in step 3, section B.2.

► **Step 4:**

- Keep all transistors length to L = minimum length set by used technology
- Set $W_{PD} = W_P = W_{PL} = W_{4\text{-}P} = 10L$
- Set $W_{ND} = W_N = W_{NL} = W_{4\text{-}N} = 10L$
- Measure the loop frequency as F_4 according to Figure B.1.

► **Step 5:**

- Calculate $I_{4\text{-}P}$ and $I_{4\text{-}N}$ and $I_{4\text{-}T}$ as:

$$I_{d4-P} = \kappa_{t-P}.W_{4-P}.(V_{gs} - V_{t-P})^{(Z_P + \frac{M_P}{V_{gs}})}.(1 + \lambda_P.V_{ds}) \qquad (B.19)$$

$$I_{d4-N} = \kappa_{t-N}.W_{4-N}.(V_{gs} - V_{t-N})^{(Z_N + \frac{M_N}{V_{gs}})}.(1 + \lambda_N.V_{ds}) \qquad (B.20)$$

$$I_{d4-t} = I_{d4-P} + I_{d4-N} \qquad (B.21)$$

► **Step 6:**

- Measure F_3 and calculate $I_{d3\text{-}P}$, $I_{d3\text{-}N}$, and $I_{d3\text{-}T}$ as in step 2 in section B.3.

► **Step 7:**

- Keep all transistors length to L = minimum length set by used technology
- Set $W_{PD} = W_P = W_{PL} = W_{5\text{-}P} = 10L$, and $W_{PB} = 15L$
- Set $W_{ND} = W_N = W_{NL} = W_{5\text{-}N} = 5L$, and $W_{NB} = 5L$
- Measure F_5.

► **Step 8:**

- Calculate $I_{5\text{-}P}$ and $I_{5\text{-}N}$ and $I_{5\text{-}T}$ as:

$$I_{d\,5-P} = \kappa_{t-P}.W_{5-P}.(V_{gs} - V_{t-P})^{(Z_P + \frac{M_P}{V_{gs}})}.(1 + \lambda_P.V_{ds}) \quad (B.22)$$

$$I_{d\,5-N} = \kappa_{t-N}.W_{5-N}.(V_{gs} - V_{t-N})^{(Z_N + \frac{M_N}{V_{gs}})}.(1 + \lambda_N.V_{ds}) \quad (B.23)$$

$$I_{d\,5-T} = I_{d\,5-P} + I_{d\,5-N} \quad (B.24)$$

► **Step 9:**

- Keep all transistors length to L = minimum length set by used technology
- Set $W_{PD} = W_P = W_{PL} = W_{5\text{-}P} = 10L$, and $W_{PB} = 10L$
- Set $W_{ND} = W_N = W_{NL} = W_{5\text{-}N} = 5L$, and $W_{NB} = 10L$
- Measure *F6*.

► **Step 10:**

- Calculate $I_{6\text{-}P}$ and $I_{6\text{-}N}$ and $I_{6\text{-}T}$ as:

$$I_{d\,6-P} = \kappa_{t-P}.W_{6-P}.(V_{gs} - V_{t-P})^{(Z_P + \frac{M_P}{V_{gs}})}.(1 + \lambda_P.V_{ds}) \quad (B.25)$$

$$I_{d\,6-N} = \kappa_{t-N}.W_{6-N}.(V_{gs} - V_{t-N})^{(Z_N + \frac{M_N}{V_{gs}})}.(1 + \lambda_N.V_{ds}) \quad (B.26)$$

$$I_{d\,6-T} = I_{d\,6-P} + I_{d\,6-N} \quad (B.27)$$

► **Step 11:**

- Compute C_{Pg} and C_{Ng} as:

$$C_{Pg} = X9 - X8 \quad (B.28)$$

$$C_{Ng} = X10 - X8 \quad (B.29)$$

- Compute C_{Pd} and C_{Nd} as:

$$C_{Pd} = (X6 - X5) - C_{Pg} \quad (B.30)$$

$$C_{Nd} = (X7 - X5) - C_{Ng} \quad (B.31)$$

where *X5, X, X7 X8, X9,* and *X10* as:

$$X5 = \frac{(2.I_{d\,1-P}.I_{d\,1-N})}{3.V_{dd}.F_1.I_{d\,1-T}.5.L} \quad (B.32)$$

$$X6 = \frac{(2.I_{d2-P}.I_{d2-N})}{3.V_{dd}.F_2.I_{d2-T}.5.L} \quad (B.33)$$

$$X7 = \frac{(2.I_{d3-P}.I_{d3-N})}{3.V_{dd}.F_3.I_{d3-T}.5.L} \quad (B.34)$$

$$X8 = \frac{(2.I_{d4-P}.I_{d4-N})}{3.V_{dd}.F_4.I_{d4-T}.5.L} \quad (B.35)$$

$$X9 = \frac{(2.I_{d5-P}.I_{d5-N})}{3.V_{dd}.F_5.I_{d5-T}.5.L} \quad (B.36)$$

$$X10 = \frac{(2.I_{d6-P}.I_{d6-N})}{3.V_{dd}.F_6.I_{d6-T}.5.L} \quad (B.37)$$

B.5 MOSFET Capacitance Extraction: Class 5

This is the most detailed of the MOSFET capacitance parameter subdivisions being considered. A chain of inverters (Figure B.3 [24]) provides the circuit model used to extract gate and diffusion parameters of PFETs and NFETs specialized to rising and falling transition directions.

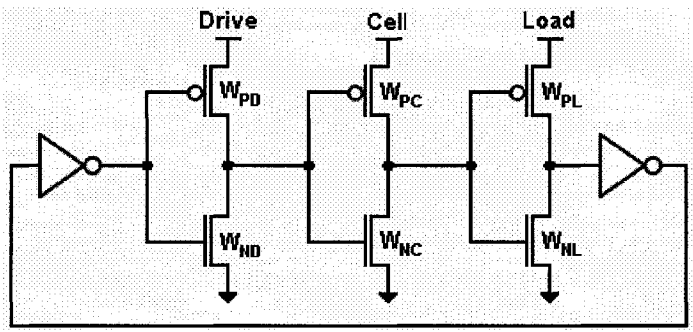


Figure B.3: CMOS inverters chain [24]

The falling propagation delay, $D_{F\text{-}Cell}$, and rising propagation delay, $D_{R\text{-}Cell}$, of the cell gate shown in Figure 4.9 are [24]:

$$D_{F-Cell} = D_{SF-Cell} + S_{FN}.D_{SR-Drive} \quad (B.38)$$

$$D_{R-Cell} = D_{SR-Cell} + S_{PR}.D_{SF-Drive} \quad (B.39)$$

where:

$$D_{SF-Cell} = \frac{v_{NF}}{W_{NC}}.(W_{PL}.C_{g\,PF} + W_{NL}.C_{g\,NF} + W_{PC}.C_{d\,PF} + W_{NC}.C_{d\,NF}) \quad (B.40)$$

$$D_{SR-Cell} = \frac{v_{PR}}{W_{NC}}.(W_{PL}.C_{g\,PR} + W_{NL}.C_{g\,NR} + W_{PC}.C_{d\,PR} + W_{NC}.C_{d\,NR}) \quad (B.41)$$

$$D_{SF-Drive} = \frac{v_{NF}}{W_{ND}}.(W_{PC}.C_{g\,PF} + W_{NC}.C_{g\,NF} + W_{PD}.C_{d\,PF} + W_{ND}.C_{d\,NF}) \quad (B.42)$$

$$D_{SR-Drive} = \frac{v_{PR}}{W_{ND}}.(W_{PC}.C_{g\,PR} + W_{NC}.C_{g\,NR} + W_{PD}.C_{d\,PR} + W_{ND}.C_{d\,NR}) \quad (B.43)$$

The subscripts D, C, and L distinguish the drive, cell, and load gates in the chain. The subscripts P and N indicate the device type. A subscript of R indicates relevance to rising transitions, and one of F relevance to falling transitions.

The capacitance parameters, named C_{gPR}, C_{gNR}, C_{gPF}, *and* C_{gNF} (gate parameters) and C_{dPR}, C_{dNR} C_{dPF}, and C_{dNF} (diffusion parameters), are extracted using the following steps:

► **Step 1:**

- Set $W_{PD} = W_{PC} = W_{PL} = W_{P1}$ and $W_{ND} = W_{NC} = W_{NL} = W_{N1}$
- Measure the cell gate rising and falling delay as $D_{R1} = D_{R\text{-}Cell}$ and $D_{F1} = D_{F\text{-}Cell}$

► **Step 2:**

- Set W_{PL}=W_{P2} and measure the cell gate rising and falling delay as $D_{R2} = D_{R\text{-}Cell}$ and $D_{F2} = D_{F\text{-}Cell}$
- Calculate:

$$C_{g\,PR} = \frac{W_{P1}}{v_{PR}}.\frac{D_{R1} - D_{R2}}{W_{P1} - W_{P2}} \quad (B.44)$$

$$C_{g\,NF} = \frac{W_{N1}}{v_{NF}}.\frac{D_{F1} - D_{F2}}{W_{P1} - W_{P2}} \quad (B.55)$$

► **Step 3:**

- Reset the widths to step 1.
- Set $W_{NL} = W_{N3}$
- Measure the cell gate rising and falling delay as $D_{R3} = D_{R\text{-}Cell}$ and $D_{F3} = D_{F\text{-}Cell}$.
- Calculate:

$$C_{g\,NR} = \frac{W_{P1}}{v_{PR}}.\frac{D_{R1} - D_{R3}}{W_{N1} - W_{N3}} \quad (B.56)$$

$$C_{g\,NF} = \frac{W_{N1}}{v_{NF}} \cdot \frac{D_{F1} - D_{F3}}{W_{N1} - W_{N3}} \qquad (B.57)$$

► **Step 4:**

- Reset the widths to step 1.
- Set $W_{PC} = W_{P4}$
- Measure the cell gate falling delay *as* $D_{F4} = D_{F\text{-}Cell}$
- Calculate:

$$C_{d\,PF} = \frac{W_{N1}}{v_{NF}} \cdot \left(\frac{D_{F1} - D_{F4}}{W_{P1} - W_{P4}} - S_{NF} \cdot \frac{v_{PR}}{W_{P1}} \cdot C_{g\,PR} \right) \qquad (B.58)$$

► **Step 5:**

- Reset the widths to step 1
- Set $W_{NC} = W_{N5}$
- Measure the cell gate falling delay as $D_{R5} = D_{R\text{-}Cell}$
- Calculate:

$$C_{d\,NR} = \frac{W_{P1}}{v_{PR}} \cdot \left(\frac{D_{R1} - D_{R5}}{W_{N1} - W_{N5}} - S_{PR} \cdot \frac{v_{NF}}{W_{N1}} \cdot C_{g\,NF} \right) \qquad (B.59)$$

► **Step 6:**

- Reset the widths to step 1.
- Set $W_{PD} = W_{PC} = W_{PL} = W_{P6}$
- Measure the cell gate rising and falling delay as $D_{R6} = D_{R\text{-}Cell}$ and $D_{F6} = D_{F\text{-}Cell}$.
- Calculate:

$$C_{d\,NF} = \frac{a - b}{c - d} \qquad (B.60)$$

$$C_{d\,PR} = b + c.C_{d\,NF} \qquad (B.61)$$

where:

$$a = \frac{D_{R1}.y6 - D_{F6}.x1}{v_{PR}.(D_{F6} - S_{NF}.D_{R1.})} \qquad (B.62)$$

$$b = \frac{D_{R6}.y1 - D_{F1}.x6}{v_{PR}.(D_{F1} - S_{NF}.D_{R6.})} \qquad (B.63)$$

$$c = \frac{v_{NF}.(D_{R6} - S_{PR}D_{F1.})}{v_{PR}.(D_{F1} - S_{NF}.D_{R6.})} \qquad (B.64)$$

$$d = \frac{v_{NF}.(D_{R1} - S_{PR}D_{F6.})}{v_{PR}.(D_{F6} - S_{NF}.D_{R1.})} \quad (B.65)$$

$$x1 = \frac{v_{PR}}{W_{P1}}.(W_{P1}.C_{g_{PR}} + W_{N1}.C_{g_{NR}} + W_{N1}.C_{d_{NR}}) + S_{PR}.\frac{v_{NF}}{W_{N1}}.(W_{P1}.C_{g_{PF}} + W_{N1}.C_{g_{NF}} + W_{P1}.C_{d_{PF}}) \quad (B.66)$$

$$x6 = \frac{v_{PR}}{W_{P6}}.(W_{P6}.C_{g_{PR}} + W_{N6}.C_{g_{NR}} + W_{N6}.C_{d_{NR}}) + S_{PR}.\frac{v_{NF}}{W_{N6}}.(W_{P6}.C_{g_{PF}} + W_{N6}.C_{g_{NF}} + W_{P6}.C_{d_{PF}}) \quad (B.67)$$

$$y1 = \frac{v_{NF}}{W_{N1}}.(W_{P1}.C_{g_{NF}} + W_{N1}.C_{g_{NF}} + W_{P1}.C_{d_{PF}}) + S_{NF}.\frac{v_{PR}}{W_{P1}}.(W_{P1}.C_{g_{PR}} + W_{N1}.C_{g_{NR}} + W_{N1}.C_{d_{NR}}) \quad (B.68)$$

$$y6 = \frac{v_{NF}}{W_{N6}}.(W_{P6}.C_{g_{NF}} + W_{N6}.C_{g_{NF}} + W_{P6}.C_{d_{PF}}) + S_{NF}.\frac{v_{PR}}{W_{P6}}.(W_{P6}.C_{g_{PR}} + W_{N6}.C_{g_{NR}} + W_{N6}.C_{d_{NR}}) \quad (B.69)$$

Table B.1 gives the widths to be used in steps 1 to 6 in the case of 0.13 μm technology, while Table B.2 gives them for the 90 nm case.

Table B.1: NMOS and PMOS widths to C_g and C_d in 0.13 μm technology

Step	W_{PD} (μm)	W_{PC} (μm)	W_{PL} (μm)	W_{ND} (μm)	W_{NC} (μm)	W_{NL} (μm)
1	9.00	9.00	9.00	3.00	3.00	3.00
2	9.00	9.00	12.00	3.00	3.00	3.00
3	9.00	9.00	9.00	3.00	3.00	6.33
4	9.00	12.00	9.00	3.00	3.00	3.00
5	9.00	9.00	9.00	3.00	6.00	3.00
6	6.00	6.00	6.00	3.00	3.00	3.00

Table B.2: NMOS and PMOS widths to C_g and C_d in 90 nm technology

Step	W_{PD} (μm)	W_{PC} (μm)	W_{PL} (μm)	W_{ND} (μm)	W_{NC} (μm)	W_{NL} (μm)
1	2.70	2.70	2.70	0.90	0.90	0.90
2	2.70	2.70	3.60	0.90	0.90	0.90
3	2.70	2.70	2.70	0.90	0.90	1.80
4	2.70	3.60	2.70	0.90	0.90	0.90
5	2.70	2.70	2.70	0.90	1.80	0.90
6	1.80	1.80	1.80	0.90	0.90	0.90