A Technology Portable Analytical Model for DSM CMOS Inverter Short-circuit Power Estimation

Shihong Deng

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfilment of the Requirements
for the Degree of Master of Applied Science at
Concordia University
Montreal, Quebec, Canada

March 2004

©Shihong Deng, 2004

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisisitons et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-91018-0
Our file Notre référence
ISBN: 0-612-91018-0

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this dissertation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de ce manuscrit.

While these forms may be included in the document page count, their removal does not represent any loss of content from the dissertation.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canadä

ABSTRACT

A Technology Portable Analytical Model for DSM CMOS Inverter Short-circuit Power Estimation

Shihong Deng

In this thesis, an accurate analytical model for DSM (Deep Sub-Micron) CMOS inverter short-circuit power estimation is presented. Compared with previous works, which always require extracted or fitting parameters, the proposed model depends only on the inverter dimension and SPICE parameters, which are usually provided with the given technology, resulting in the technology portability of this model. To achieve accurate modeling, the effect of the gate-to-drain coupling capacitance and some main DSM effects such as velocity saturation and mobility degradation are taken into account.

The accuracy and portability validations of the proposed model have been performed for UMC 1.2V 0.13μm and TSMC 1.8V 0.18μm CMOS technologies, and for a wide range of input transition times, output loading capacitances and aspect ratios. The results produced by the proposed model show good agreement with Spectre simulation using BSIM3v3.2 model in TSMC 1.8V 0.18μm technologies, indicating its portability. Its accuracy is better than that of latest methods that require extraction. Based on a Maple implementation, the proposed model always offers much less average CPU time than Spectre simulator.

Acknowledgements

First of all, I would like to express my deepest appreciation and respect to my supervisor, Professor Asim J. Al-Khalili, who proposed the original idea of this thesis and offered all of his insightful guidance, valuable advice and continuing encouragement throughout this work. He always had time and patience to answer numerous questions but also gave me room to learn on my own.

The fellow graduate students that I worked with in the VLSI Lab are very helpful. The support given by Ted has been greatly appreciated.

The proposed model was tested in UMC 1.2V 0.13μm and TSMC 1.8V 0.18μm CMOS technologies, which are gracefully provided by the Canadian Microelectronics Corporation (CMC).

Finally, I am happy to dedicate this work to my beloved wife Yuelin and daughter Chuyu, who have so much confidence in me and continuously encourage me through all the hard times.

Table of Contents

List of Figures vii

List of Tables ix

List of Acronyms ix

Nomenclature ix

Chapter 1. Introduction 1

1.1 Motivation and Objective 1

1.2 Thesis Scope and Overview 4

Chapter 2. Methodology of Inverter Power Modeling 6

2.1 MOSFET Capacitance Modeling 6

2.1.1 MOSFET Parasitic Capacitances 7

2.1.2 Gate Capacitances 9

2.1.3 Junction Capacitances 11

2.2 The CMOS Inverter 13

2.3 Literature Review 14

2.4 Inverter Power Modeling 18

2.4.1 Contributions to the Power Dissipation of Digital CMOS ICs 18

2.4.2 Estimation Approach of Short-Circuit Power Dissipation 19

Chapter 3. Submicron MOSFET Models 26

3.1 Square-law long-channel MOSFET Model 27

3.2 Short-channel Effects 28

3.2.1 Velocity Saturation 29

3.2.2 Mobility Degradation 31

3.2.3 Threshold Voltage Variations 33

3.2.3.1 Non-Uniform Doping Effects 34

3.2.3.2 Short-Channel Effects....35
3.2.3.3 Narrow Width Effect....36
3.3 α-power Law MOSFET Model....36
3.3.1 Model Equations....37
3.3.2 Extraction of Model Parameters....37
3.3.3 Discussions on the Model....38
3.4 Toh-Ko-Meyer MOS Engineering Model....39
3.4.1 Toh-Ko-Meyer MOS Engineering Model....39
3.4.2 Normalized Inverter Current Model Equations....42

Chapter 4. Inverter Short-Circuit Power Modeling....44
4.1 Short-circuit Energy Dissipated in Rising Input Transition....44
4.2 Short-circuit Energy Dissipated in Falling Input Transition....54
4.3 Inverter Short-circuit Power Dissipation....55

Chapter 5. Model Validations and Comparisons....57
5.1 Illumination of Comparison Method....57
5.2 Model Accuracy Validation....59
5.2.1 Accuracy Validation Using a Randomly Sized Inverter....59
5.2.2 Accuracy Validation Using Large Size Inverters....62
5.2.3 Accuracy Validation Using Inverters with Different Ratios....64
5.2.4 Accuracy Comparisons with Latest Methods....66
5.3 Model Portability Validation....69
5.4 CPU Time....71

Chapter 6. Conclusion....73

References....74

List of Figures

Figure 2.1 An n-channel MOSFET with the intrinsic and extrinsic capacitances.........7
Figure 2.2 Lumped capacitances of the MOSFET...8
Figure 2.3 CMOS inverter circuit with parasitic capacitances modeled explicitly......13
Figure 2.4 CMOS inverter circuit with rising/falling input voltage............................20
Figure 2.5 Inverter response in the case of rising input..22
Figure 2.6 The short-circuit current and its piecewise linear approximation...............24
Figure 3.1 Effect of electrical field on electron velocity...29
Figure 3.2 Mobility degradation effect...31
Figure 3.3 Vertical substrate doping distribution and its approximation.....................34
Figure 3.4 Threshold variations...35
Figure 4.1 Rising input inverter with explicitly modeled parasitic capacitances.........45
Figure 4.2 Operation regions of a typical inverter...46
Figure 4.3 The first approximation of the PMOS current in region 2..........................50
Figure 4.4 The second approximation of the PMOS current in region 2.....................53
Figure 5.1 Inverter short-circuit energy dissipation in one switching cycle as a function of input transition time...60
Figure 5.2 Transient waveforms of the inverter in Figure 5.1 (rising input)................61
Figure 5.3 Inverter short-circuit energy dissipation in one switching cycle as a function of output loading capacitance...61
Figure 5.4 Transient waveforms of the inverter in Figure 5.3 (rising input)................62
Figure 5.5 Large size inverter (Wn=0.52μm, and Wp=1.04μm) short-circuit energy dissipation in one switching cycle as a function of loading capacitance 63
Figure 5.6 Large size inverter (Wn=1.04μm, and Wp=2.08μm) short-circuit energy dissipation in one switching cycle as a function of loading capacitance 63
Figure 5.7 Inverter short-circuit energy in one switching cycle as a function of input transition time for different aspect ratio (Wp/Wn=.5, 1, 2, and 4) 64
Figure 5.8 Inverter short-circuit energy in one switching cycle as a function of input

transition time for different aspect ratio (Wp/Wn=8, 12, and 16) 65

Figure 5.9 Comparison of inverter short-circuit energy in one switching cycle as a function of input transition time (Inverter size: Ln=Lp=0.8μm, Wn=4μm, Wp=6.55μm), using different method (Copied from [8])...........................69

Figure 5.10 Inverter short-circuit energy in one switching cycle as a function of input transition time using 0.18μm CMOS technology...70

Figure 5.11 Inverter short-circuit energy in one switching cycle as a function of loading capacitance using 0.18μm CMOS technology..71

List of Tables

Table 2.1 Approximation of intrinsic CMOS gate capacitances.................................10

Table 2.2 Definition of model parameters for the MOSFET junction capacitance......12

Table 3.1 Typical mobility values for electrons and holes..32

Table 4.1 The interchanges between rising input inverters and falling input inverters..55

Table 4.2 Summary of parameters used in the proposed model..................................56

Table 4.3 Summary of equations used in the proposed model....................................56

Table 5.1 Performance comparison of MOSFET models...58

Table 5.2 Short-circuit energy and total dynamic energy dissipation per switching event estimated using the method in [7] 66

Table 5.3 The relative errors of the inverter short-circuit energy estimation of the cases in Table 5.2 67

Table 5.4 Inverter short-circuit energy results obtained by using the proposed model, in the cases corresponding to Table5.3...68

List of Acronyms

CMOS	Complementary Metal Oxide Semiconductor
VLSI	Very Large Scale Integration
DRAM	Dynamic Random Access Memory
DSM	Deep Submicron meter
SPICE	Simulation Program Integrated Circuits Especially
MOSFET	Metallic Oxide Semiconductor Field Effect Transistor
PMOS	Positive-channel Metal Oxide Semiconductor
NMOS	Negative-channel Metal Oxide Semiconductor
IC	Integration Circuit
BSIM	Berkeley Short-channel IGFET Model
ODE	Ordinary Differential Equation

Nomenclature

This thesis uses the following notational conventions:

- The subscripts d, g, s and b denote, respectively, the drain, gate, source, and bulk terminals of the MOS transistors.
- The subscripts n and p are used to associate the values to the NMOS or PMOS device in the CMOS inverter circuit, respectively.
- The subscripts i and o are used to associate the values to the input or output of the CMOS inverter, respectively.

Chapter 1
Introduction

To set the scene for this thesis, this chapter begins by outlining the motivation and objective of the topic addressed in this work and the necessity of a CMOS inverter model. A description of how this thesis is organized then follows.

1.1 Motivation and Objective

Power dissipation is one of the most critical design parameters in VLSI circuits. In past, many circuit designers believed "CMOS circuits dissipates little power by nature". But in reality, CMOS power dissipation as a whole has increased by 4 times every 3 years, which is the same pace as increase in bit density of state-of-the-art DRAMs [1]. The power dissipation has increased by 1000 times over the past 15 years and already exceeded 50 watts. For example, the third-generation 64-bit microprocessor UltraSpectre dissipates 53 watts at 1.3v, 1.1GHz [2].

With the considerable increase in power consumption and heat dissipation resulting from the continuing growing in the clock frequency, density and size of VLSI chips, accurate and efficient evaluation of power dissipation during the design phase becomes more important in order to select appropriate packaging and adequate cooling. The power dissipation of CMOS logic circuits consists of three components: the static, the switching and the short-circuit power dissipation [3].

- The static power comes from subthreshold conduction and reverse bias leakage between diffusion regions and the substrate. The contribution of this part to the total power consumption was negligible in a well-designed CMOS logic circuit [4], however is increasing as devices become smaller.
- The switching power consumption, resulting from charging and discharging the load capacitance of a CMOS circuit, is well understood and easy to be estimated if the load capacitance is known.
- The short-circuit power is due to the direct-path current from supply to ground when both the pull-up and pull-down structures in static CMOS circuits are simultaneously active.

The estimation of short-circuit power is not easy, because the short-circuit current is affected by many factors such as the input transition time, the capacitive load, and the transistors' sizes, etc. [5]. Therefore, power estimation and optimization for deep submicron (DSM) CMOS logic circuits requires accurate and efficient modeling of short-circuit power dissipation. However, some common time-stepped or iteration required circuit simulators such as SPICE are very time consuming, although they play an important role in circuit design. Most of other existing techniques, which can be efficiently used with modern deep submicron CMOS technologies, are depending on either empirical parameters, or extracted or fitting parameters. As a result, there is a strong need for analytical expressions to accurately model the design performance and help designers quickly assess the impacts of different design parameters.

Recently, a novel kind of model -- technology portable model has been proposed to estimate the transition time of DSM CMOS logic circuits [6]. The technology portable

model is formulated depending only on SPICE parameters, which are usually provided with the given technology. Technology portability means that the model does not depend on any extracted or fitting parameters, and no extra effort is required for any pre-simulations or pre-measurements. Therefore, considering the fast shifting pace of nanometer processes and the reuse design methodology, it is highly desirable for the DSM models to be technology portable. In light of the same idea, a technology portable analytical model has been presented in this thesis to address the short-circuit power estimation for DSM CMOS inverters.

So far, many techniques for estimating power dissipation and/or delay in CMOS logic circuits have been presented. Most of them develop analytical models particularly for the inverter circuit. The emphasis on modeling the inverter comes from the following [7][8].

- In timing analysis, the worst-case delay of any complex CMOS gate can be simulated by replacing this gate with its worst-case equivalent inverter. The worst-case delay can help designers to determine the critical paths in VLSI systems.
- Many efficient transistor-level techniques for reducing CMOS logic gates to equivalent inverters are available [9][10][11][12]. Thus, with these techniques and by developing a delay/power model for the CMOS inverter, any CMOS logic circuit can be easily simplified and analyzed.
- In a digital VLSI chip, the clock distribution network and buses are based on inverters or inverter like circuits which must be carefully designed and modeled. These circuits account for a large fraction of the total power consumption[13]. In addition, the power dissipated in off-chip drivers (pads,

also based on inverters) could be a large percentage of the chip power, i.e. more than 15% of the total chip power[14].

1.2 Thesis Scope and Overview

In this thesis, an accurate technology portable analytical model for estimating short-circuit power dissipation of deep submicron CMOS inverter is presented. The Toh-Ko-Meyer engineering MOSFET model [15] is used to relate the terminal voltages to the drain current in submicron transistors. In the course of modeling, the effect of input-to-output coupling capacitance is considered. The accuracy and portability of the proposed model have been validated by comparing its results with BSIM3v3 MOSFET model simulation results for both 1.8V 0.18μm and 1.2V 0.13μm CMOS technologies, and for a wide range of inverter sizes, input transition times, and capacitive loads.

After giving the motivation of the topic and the thesis overview in Chapter 1, Chapter 2 begins by introducing the different MOSFET capacitances. To simplify the transient analysis of the CMOS inverter, the nonlinear voltage-dependent parasitic capacitances in the inverter circuit are modeled by equivalent constant capacitances. This is described in Section 2.1. The CMOS inverter circuit and the associated parasitic capacitances, which must be taken into account in current and power calculations, are analyzed in Section 2.2. Section 2.3 is dedicated to a detailed literature review of techniques for computing the current and/or power in CMOS inverters. The inverter power components and the method of estimating the inverter short-circuit power dissipation are presented in the last section.

Chapter 3 begins by briefly introducing the square-law MOSFET model and the reasons why it is no longer valid for short-channel devices. A description of some main

short-channel effects, which must be taken into account when analyzing submicron devices, then follows. In the third and fourth section, the α-power law MOSFET model and the Toh-Ko-Meyer MOS engineering model are introduced respectively. In the last section of this chapter, the model equations used in this thesis are normalized and listed for easy reference.

Following the methodology described in Chapter 2, the procedures and expressions to estimate the CMOS inverter short-circuit energy dissipation in rising input transition is given first in Chapter 4. Then, they are extended to the case of falling input transition. In the last section of this chapter, the expression of short-circuit power dissipation is given, and the model parameters and equations are summarized.

In the first section of Chapter 5, the method using the results of Spectre simulator with BSIM3v3.2 MOSFET model as "exact" values, instead of directly measured values, is illuminated. Then, the proposed inverter short-circuit power model, which is implemented in Maple, has been tested with a wide range of inverters in two different CMOS technologies. The model accuracy is verified by using the UMC 1.2V 0.13μm CMOS technology, and its portability is verified by using the TSMC 1.8V 0.18μm CMOS technology. Various switching conditions of input transition time, capacitive load, and inverter ratio have been considered. To show the rapidness of the proposed model, the CPU time is compared with SPECTRE simulator.

The last chapter is dedicated to conclusions.

Chapter 2
Methodology of Inverter Power Modeling

In this chapter, the methodology used in the thesis is presented. It begins by introducing the different MOSFET capacitances. To simplify the transient analysis of the CMOS inverter, the nonlinear voltage-dependent parasitic capacitances in the inverter circuit are modeled by equivalent constant capacitances, as described in Section 2.1. The CMOS inverter circuit and the associated parasitic capacitances, which must be taken into account in current and power calculations, are analyzed in Section 2.2. Section 2.3 is dedicated to a detailed literature review of techniques for computing the current and/or power in CMOS inverters. The inverter power components and the method of estimating the inverter short-circuit power dissipation are presented in the last section.

2.1 MOSFET Capacitance Modeling

Before analyzing the operation of the CMOS inverter, it is necessary to understand the capacitances in a MOSFET at first. The dynamic response of CMOS circuits is strongly dependent on the different capacitances associated with the CMOS devices and interconnection capacitances. Thus, the accuracy of the inverter model depends on those parasitic capacitances to a large extent. However, the MOSFET parasitic capacitances are distributed, voltage-dependent, and nonlinear, and their exact modeling is very complex. Therefore, in order to simplify the calculations in the proposed model, these capacitances are replaced by equivalent constant capacitances, as described later. This simplification is

widely accepted and used [3][4][7][8]. This comes of course at the expense of accuracy.

2.1.1 MOSFET Parasitic Capacitances

MOSFETs have a number of different capacitances which must be accounted for in AC and transient analysis. Figure 2.1 shows an n-channel MOSFET with its different capacitances. Generally, MOSFET capacitance can be divided into two groups, the intrinsic and the extrinsic capacitances [16].

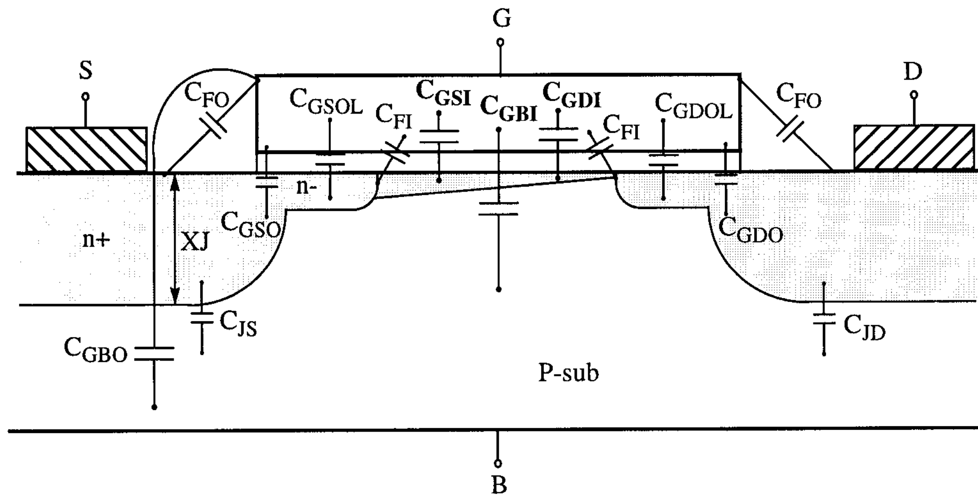


Figure 2.1 An n-channel MOSFET with the intrinsic (bold-faced) and extrinsic capacitances [16].

The extrinsic capacitance is further divided into five components:

1) the outer fringing capacitance between the polysilicon gate and S/D regions, C_{FO};

2) the inner fringing capacitance between the polysilicon gate and S/D regions, C_{FI};

3) the overlap capacitances between the gate and the heavily doped S/D regions (and the bulk region), C_{GSO} & C_{GDO} (C_{GBO}), which are relatively insensitive to terminal voltages;

4) the overlap capacitances between the gate and the lightly doped S/D regions, C_{GSOL} & C_{GDOL}, which changes with bias; and

5) the source/drain junction capacitances, C_{JD} & C_{JS}.

The intrinsic capacitance is related to the region between the metallurgical source and drain junctions. In Figure 2.1, C_{GSI}, C_{GDI} and C_{GBI} are respectively the gate-to-source intrinsic capacitance, the gate-to-drain intrinsic capacitance, and the gate-to-bulk intrinsic capacitance. This picture of the intrinsic capacitance is overly simplistic. In fact, the intrinsic capacitance is much more complex than the extrinsic components, and consists of up to 16 nonreciprocal capacitance components [16]. These MOSFET capacitances are nonlinear and voltage-dependent, and their exact modeling is very complex. However, some intrinsic capacitance models that simply treated the MOSFET capacitance as several separate lumped capacitances, have been used widely in simulators and continue to be used as optional models for their simplicity and efficiency.

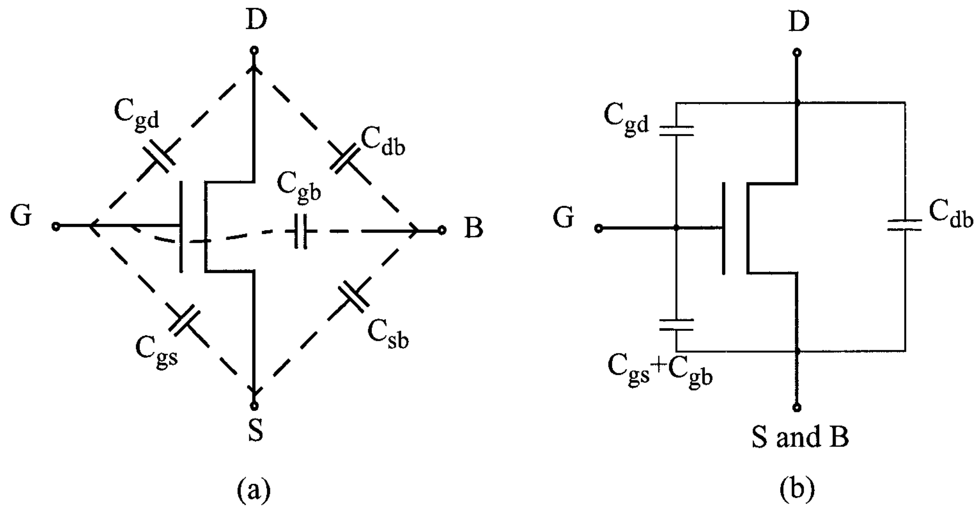


Figure 2.2 Lumped capacitances of the MOSFET. (a) The lumped capacitances of a MOS transistor. C_{gd}, C_{gs} and C_{gb} are gate capacitances. C_{db} and C_{sb} are junction capacitances. (b) Lumped capacitance model for the MOS transistor of an inverter (the source and bulk terminals are at the same potential: Vsb = 0).

Figure 2.2 (a) is a diagrammatic representation of the lumped capacitances of a CMOS transistor. For a CMOS inverter, the bulk is connected to the source. So the capacitances

of a transistor in an inverter can be lumped as shown in Figure 2.2 (b). C_{gs}, C_{gd} and C_{gb} are gate capacitances, where C_{gs} and C_{gd} are gate-to-channel capacitances including the gate-to-source and gate-to-drain overlap capacitances, and C_{gb} is the gate-to-bulk capacitance. C_{sb} and C_{db} are p-n junction capacitances describing the source-bulk and drain-bulk junction capacitances, respectively.

It should be noted that in this chapter, symbols with upper-case subscripts are for the capacitances per unit area or per unit length; Symbols with lower-case subscripts are for the total capacitances. C_{GS}, C_{GD}, C_{GB}, C_{gs}, C_{gd}, and C_{gb} here include both intrinsic and extrinsic components.

2.1.2 Gate Capacitances

The total gate capacitance C_G of a CMOS transistor is given by

$$C_G = C_{GB} + C_{GS} + C_{GD} \qquad (2.1)$$

These gate capacitances are expressed as the sum of their respective intrinsic and extrinsic (overlap) capacitances:

$$C_{GB} = C_{GBint} + C_{GBext} \qquad (2.2)$$

$$C_{GS} = C_{GSint} + C_{GSext} \qquad (2.3)$$

$$C_{GD} = C_{GDint} + C_{GDext} \qquad (2.4)$$

where the subscripts *int* and *ext* stand for intrinsic and extrinsic, respectively.

The intrinsic capacitance is a fraction of the effective gate-oxide capacitance and function of the MOSFET biasing voltages. A simple, yet accurate enough, approximation of the intrinsic capacitance is

$$C_{int} \approx \gamma W_g L_e C_{ox} \qquad (2.5)$$

where γ is a constant over each MOSFET mode of operation.

The gate-oxide capacitance per unit area is given by

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \quad (2.6)$$

where ε_{ox} and t_{ox} are the permittivity and thickness of the gate oxide respectively. L_e is the effective channel length. W_g is the effective gate width. L_e and W_g are defined as following.

$$L_e = L + XL - 2LD \quad (2.7)$$

$$W_g = W + XW \quad (2.8)$$

Where L and W are the drawn channel length and width, respectively. XL and XW are, respectively, the gate length and width correction to account for masking and etching effects. LD is the lateral diffusion into the channel from the source/drain diffusion region.

In terms of the accuracy of the transient analysis, the optimal set of values for γ depends on the MOS process technology used, and must therefore be extracted experimentally. However, to keep the inverter model simple and general, the conservative estimates, given in Table 2.1 [3], for the gate intrinsic capacitances are used.

Table 2.1 Approximation of intrinsic CMOS gate capacitances [3].

	MOSFET Mode of Operation		
Parameters	Off	Linear	Saturation
C_{GBint}	$W_g L_e C_{ox}$	0	0
C_{GSint}	0	$\frac{1}{2} W_g L_e C_{ox}$	$\frac{2}{3} W_g L_e C_{ox}$
C_{GDint}	0	$\frac{1}{2} W_g L_e C_{ox}$	0

The gate extrinsic capacitances, C_{GBext}, C_{GSext}, and C_{GDext}, account for the overlap

capacitances of the gate over the bulk, source, and drain regions, respectively. To estimate these gate extrinsic capacitances, three extrinsic values of C_{GS0}, C_{GD0} and C_{GB0}, which are specified in the SPICE MOSFET model by CGS0, CGD0, and CGB0 respectively, are used as follows [3][17]

$$C_{GBext} = L_e \cdot CGB0 \tag{2.9}$$

$$C_{GSext} = W_g \cdot CGS0 \tag{2.10}$$

$$C_{GDext} = W_g \cdot CGD0 \tag{2.11}$$

2.1.3 Junction Capacitances

The junction (depletion-layer) capacitance C_j of a p-n junction diode is a function of the reverse-bias voltage V_r across the junction [18]:

$$C_j(V_r) = \frac{C_{j0}}{\left(1 + \frac{V_r}{V_{bi}}\right)^m} \tag{2.12}$$

where C_{j0} is the junction capacitance at equilibrium (i.e. at zero bias). V_{bi} is the junction built-in potential and m is the junction grading coefficient (m is 1/2 for an abrupt p-n junction and 1/3 for a linearly-graded one).

An equivalent constant capacitance, which sees the same change in charge as the voltage-dependent junction capacitance for a change in the reverse-bias voltage from V_{r1} to V_{r2}, is then given by [18]

$$C_{eq} = \frac{C_{j0}}{(1-m)} \frac{V_{bi}}{(V_2 - V_1)} \left[\left(1 + \frac{V_{r2}}{V_{bi}}\right)^{1-m} - \left(1 + \frac{V_{r1}}{V_{bi}}\right)^{1-m}\right] \tag{2.13}$$

Similarly, in the CMOS inverter circuit, the drain-bulk junction capacitances can be modeled by constant capacitances that will force equivalent changes in charge of

transitions in the reverse-bias drain-bulk voltages between 0 and V_{DD}.

$$C_{db} = C_{ja} + C_{jf} + C_{jg} \quad (2.14)$$

where

$$C_{ja} = \frac{A_d \cdot CJ0}{(1-MJ)} \frac{V_{bi}}{V_{DD}}\left[\left(1+\frac{V_{DD}}{V_{bi}}\right)^{1-MJ} - 1\right] \quad (2.15)$$

$$C_{jf} = \frac{(P_d - W_g) \cdot CJF0}{(1-MJF)} \frac{V_{bi}}{V_{DD}}\left[\left(1+\frac{V_{DD}}{V_{bi}}\right)^{1-MJF} - 1\right] \quad (2.16)$$

$$C_{jg} = \frac{W_g \cdot CJG0}{(1-MJG)} \frac{V_{bi}}{V_{DD}}\left[\left(1+\frac{V_{DD}}{V_{bi}}\right)^{1-MJG} - 1\right] \quad (2.17)$$

In the above equations, the model parameters for the MOSFET drain-bulk junction capacitance, defined in Table 2.2, are characteristic of the MOS process technology used. A_d and P_d are the area and periphery of the drain region respectively. C_{ja}, C_{jf}, and C_{jg} are the equivalent junction capacitances over the area of the drain region (A_d), along the field-oxide periphery of the drain region (P_d-W_g), and along the gate-oxide periphery of the drain region (W_g), respectively.

Table 2.2 Definition of the model parameters for the MOSFET junction capacitance.

V_{bi}	Bulk-junction built-in potential
CJ0	Zero-bias bulk-junction capacitance, per unit area over the drain region
CJF0	Zero-bias bulk-junction capacitance, per unit length along the field-oxide periphery (sidewall) of the drain region
CJG0	Zero-bias bulk-junction capacitance, per unit length along the gate-oxide periphery (gate-edge sidewall) of the drain region
MJ	Bulk-junction grading coefficient for area component
MJF	Bulk-junction grading coefficient for the field-oxide periphery (side-wall) component

V_{bi}	Bulk-junction built-in potential
MJG	Bulk-junction grading coefficient for the gate-oxide periphery (gate-edge sidewall) component

2.2 The CMOS Inverter

In Figure 2.3(a), the CMOS inverter circuit is depicted along with the parasitic capacitances modeled explicitly. As mentioned in the last section, C_{gs} and C_{gd} are gate-to-channel capacitances, which are lumped at the source and the drain region of the channel, respectively. C_{gb} is the gate-to-bulk capacitance. They also include overlap capacitances respectively. C_{db} is a lumped drain-bulk junction capacitance.

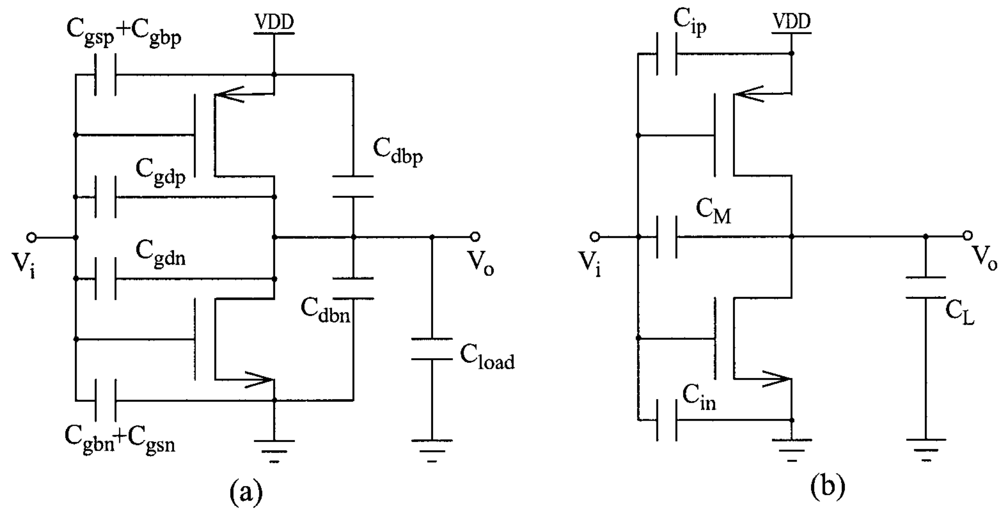


Figure 2.3 (a) CMOS inverter circuit with the parasitic capacitances modeled explicitly. (b) The equivalent circuit of (a).

For the sake of simplification, an equivalent inverter circuit in Figure 2.3(b) can be used, where

$$C_i = C_{ip} + C_{in} = C_{gbp} + C_{gsp} + C_{gbn} + C_{gsn} \tag{2.18}$$

$$C_M = C_{gdn} + C_{gdp} \tag{2.19}$$

with C_i and C_M are input capacitance and the coupling capacitance between input node and output node (Miller capacitance), respectively. The property of C_M will be introduced in Section 2.4.2.

The effective load capacitance is defined as

$$C_L = C_{op} + C_{on} + C_{load} \tag{2.20}$$

where, the output capacitance $C_o = C_{db}$. Then, C_L includes the NMOS and PMOS drain-bulk junction capacitances, the load capacitance and the interconnect capacitance.

The values of the inverter parasitic capacitances may be calculated by using expressions given in Section 2.1.

2.3 Literature Review

Since the first closed-form expression for the evaluation of the short-circuit energy dissipation in a CMOS inverter was presented by Veendrick [5] in 1984, much research has addressed the development of this field. In this section, some of the previous works on computing the power dissipation in a CMOS inverter are reviewed.

Veendrick [5] observed that the short-circuit current is dependent on the input transition time, capacitive load, and transistor sizes of the logic gate, and he gave an expression of computing the short-circuit energy dissipation in a CMOS inverter, as following.

$$P_{SC} = \frac{\beta}{12}(V_{DD} - 2 \cdot V_t)^3 \cdot \frac{T_{rf}}{T_p} \tag{2.21}$$

where $V_t = V_{tn} = |V_{tp}|$ is the threshold voltage of NMOS and PMOS transistors;

$\beta = \beta_n = \beta_p$ is the gain factor ($\mu A/V^2$) of an MOS transistor;

$T_{rf} = T_r = T_f$ is the transition time of the rising/falling input signal, and

T_p is the period of the input signal.

Although he gave this first closed-form expression to compute the short-circuit power dissipation P_{SC}, he ignored the load capacitance, and assumed that the saturation short-circuit current flows from V_{DD} to GND during the entire transition period. Furthermore, this expression is based on the Shichman and Hodges [19] square-law MOS model that ignores the carriers' velocity saturation effects of submicronmeter devices. So, this expression gives an upper bound of the P_{SC} of a CMOS inverter. Consequently, it only provides a qualitative measure of the P_{SC} and is not sufficiently accurate for power estimation and optimization.

Veendrick's formula was extended to submicron inverters in [20], by replacing the square-law MOSFET equations with the corresponding α-power law equations, while all the assumptions of [5] are retained. The α-power law model (n-th power law model in [21]) is more accurate than the square-power law model because it considers carrier velocity saturation effect that becomes prominent in short-channel devices. However, this model requires the extraction of some parameters from the static device characteristics for each transistor width, such as the empirical velocity saturation index (α or n), the V_{D0} (drain saturation voltage ar $V_{GS} = V_{DD}$) and the I_{D0} (drain current at $V_{GS} = V_{DS} = V_{DD}$). For the derivation of the output expression in [20], both the short-circuit current and the coupling capacitance are neglected.

In [22] and [23], an expression for the short-circuit energy dissipation of the CMOS inverter without the simplifications of [5] was derived. However, the formula for P_{sc} is based on the long-channel square-law MOSFET model. In their analysis, the expression of the output waveform, was derived without consideration of the short-circuit current I_{sc} and

the gate-to-drain coupling capacitance C_M.

Following the same analysis as [22] and [23] which neglected the I_{SC} and C_M, Vemuru and Scheinberg [24][25] proposed a formula for P_{SC} in submicron inverters, based on α-power law MOSFET model. Furthermore, to get an analytical solution for the integral equation of P_{SC}, the time when the short-circuit transistor changes its mode of operation was assumed to occur at $t = t_0 - t_1$, where t_1 and t_0 are, respectively, the times when the charging/discharging transistor turns on and the short-circuit transistor turns off. Consequently, the error in both formulas is large for relatively slow input ramps and/or low fanouts.

A formulation of P_{SC} through an equivalent short-circuit capacitance is presented in [26], where a rough linear approximation of the output waveform is used. In [27], the waveform of I_{SC} was approximated with a piecewise linear function of time, in order to estimate the P_{SC}. However, the energy of the reverse current due to the gate-to-drain coupling capacitance is subtracted from the short-circuit energy dissipation resulting in an underestimation.

Wang and Vrudhula [28] proposed a semi-empirical short-circuit power model, which still used the Shichman and Hodges square-law MOS model [19]. For better accuracy, two fitting parameters are inserted to the expression of the output waveform derived in [22], in order to include the reaction of the short-circuit transistor to the output charge or discharge operation.

In [29] and [30], an analytical short-circuit power model was presented, the model which takes into account the influences of the short-circuit current and the gate-to-drain coupling capacitance. Analytical expressions of the inverter output ramp response for all

the case of input ramps have been derived. However, the derived model is based on a less accurate and simpler bulk-charge MOS model [31]. Its accuracy and efficiency are improved in [8], which is based on the α-power law MOS model [20] and considers the influence of the input-to-output coupling capacitance and the short-circuit transistor's gate-source capacitance. Recently, Hamoui and Rumin [7] proposed an analytical model for estimating short-circuit energy dissipation, which is based on a modified version of the nth power law MOSFET model that they proposed.

Above all, a number of methods for estimating the short-circuit power dissipation in CMOS inverters have been presented. These methods can be divided into two classes in terms of the current model they used.

The first class, such as [5][22][23][28][29][30], uses the very simple square-law MOS model or other simple models, which are appropriate for long-channel and uniform-doping devices used two decades ago and do not take into account the main short-channel effects of modern submicron devices.

Another class of methods, such as [7][8][20][24][25][27], uses the α-power (or nth power law) MOSFET model, which will be reviewed in more detail in Chapter 3. In the α-power law MOS model, four parameters require extraction from measured data [20]. Still, the α-power law MOS model is not sufficiently accurate in DSM regime. To improve its accuracy, more extracted parameters are introduced in some modified versions [7][32][33]. In a word, this kind of methods requires some extra efforts to extract and optimize empirical parameters to characterize a given MOS process technology.

2.4 Inverter Power Modeling

2.4.1 Contributions to the Power Dissipation of Digital CMOS ICs

There are three major sources of power dissipation in digital CMOS circuits which are summarized in the following equation [3]:

$$\begin{aligned} P_{total} &= P_{static} + P_{switching} + P_{short-circuit} \\ &= I_{leakage} \cdot V_{DD} + \alpha_{0-1} \cdot C_L \cdot V_{DD}^2 \cdot f_{clk} + I_{SC} \cdot V_{DD} \end{aligned} \tag{2.22}$$

The first term is due to the leakage current $I_{leakage}$, which arises from substrate injection and subthreshold effects, and is primarily determined by fabrication technology considerations. With the ongoing trend towards scaling down the minimum feature size and the supply voltage in CMOS ICs, the subthreshold currents are growing due to short-channel effects and scaled-down threshold voltages [34][35], thereby increasing the power dissipation. However, the contribution of the static power component to the total power consumption remains negligible in a well-designed CMOS logic circuit [4].

The second term represents the switching component of power, where C_L is the load capacitance, f_{clk} is the clock frequency and α_{0-1} is the node transition activity factor (the average number of times the node makes a power consuming transition in one clock period). For an inverter circuit, the energy drawn from the power supply for the positive going transition is $C_L V_{DD}^2$, half of which is stored in the output capacitor and half is dissipated in the PMOS transistor. On the V_{DD} to 0 transition at the output, no charge is drawn from the supply, however, the energy stored in the capacitor ($C_L V_{DD}^2/2$) is dissipated in the pull-down NMOS transistor.

The third term, short-circuit power dissipation is due to the direct path short-circuit

current, I_{SC} from power supply to ground when both the NMOS and PMOS devices are simultaneously active. The difference of $P_{switching}$ and $P_{short\text{-}circuit}$ lies in that the former is independent of the rise and fall time at the input of logic gates. However finite rise and fall times of the input waveform result in a direct path between V_{DD} and GND. Specifically when the condition $V_{tn} < V_{in} < V_{DD} - |V_{tp}|$ holds for the input voltage, there will be a conductive path open between V_{DD} and GND, because both the NMOS and PMOS transistors will be simultaneously on.

Consequently, the total power dissipation in the CMOS inverter circuit can be readily calculated, provided that the short-circuit power dissipation (or short-circuit energy dissipation per switching event, E_{SC}) can be evaluated. With the ongoing trend towards scaling down the supply voltage and the minimum feature size in CMOS ICs, the total power consumption is largely underestimated if its short-circuit component is neglected. As shown in [7], the short-circuit power dissipation can no longer be neglected in submicron CMOS circuits, even with 0.8-μm channel length.

A simple, yet accurate enough, approach for evaluating inverter short-circuit power dissipation in this thesis is presented below.

2.4.2 Estimation Approach of Short-Circuit Power Dissipation

To simplify the analysis, an input waveform approximation, which is widely accepted for its simplicity and effectiveness, is used in this thesis. Assume that the input voltage waveform is a rising/falling ramp with transition time τ_r/τ_f:

$$\text{For rising inputs:}\quad V_{in}(t) = \begin{cases} 0 & t \le 0 \\ V_{DD} \cdot (t/\tau_r) & 0 \le t \le \tau_r \\ V_{DD} & t \ge \tau_r \end{cases} \qquad (2.23)$$

$$\text{For falling inputs:} \quad V_{in}(t) = \begin{cases} V_{DD} & t \le 0 \\ V_{DD} \cdot (1 - t/\tau_f) & 0 \le t \le \tau_f \\ 0 & t \ge \tau_f \end{cases} \tag{2.24}$$

From above, the slope of the rising and falling input voltage ramps can be determined:

$$S_r = V_{DD} / \tau_r, \tag{2.25}$$

$$\text{and } S_f = V_{DD} / \tau_f \tag{2.26}$$

Corresponding to the different input ramps, there are two different transitions for an inverter in one switching cycle. One results from a rising input voltage, the other from a falling input voltage.

If the input voltage V_i is a rising signal as showing in Figure 2.4(a), when $V_i \ge V_{tn}$, the NMOS transistor will turn on, and discharge the load capacitance. However, before the PMOS transistor turns off when $V_i \ge V_{DD} - |V_{tp}|$, there is a direct conduct path between the power and ground rails. Through this direct conducting path, the corresponding current, i.e. i_{sp} in Figure 2.4(a), flows. This current is called the short-circuit current in the case of a rising input (discharging inverter).

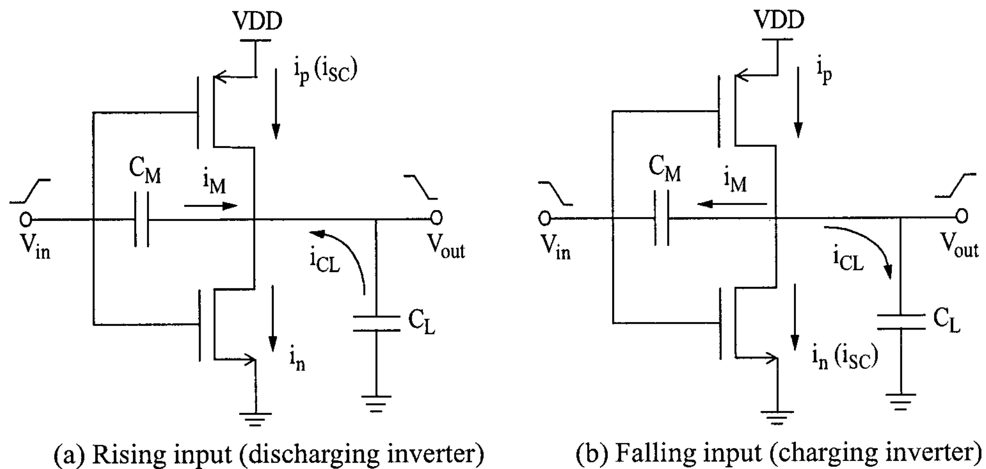


Figure 2.4 CMOS inverter circuit with rising/falling input voltage.

If the input voltage V_i is a falling signal as showing in Figure 2.4(b), when $V_{tn} \leq V_i \leq V_{DD} - |V_{tp}|$, both the PMOS and the NMOS transistor are on. Flowing through this direct conducting path, the corresponding current, i.e. i_{sn} in Figure 2.4(b), is the short-circuit current in the case of a falling input (charging inverter).

In the course of inverter transitions, the Miller capacitance C_M allows a current i_M to flow between the input and output nodes. It can be given by

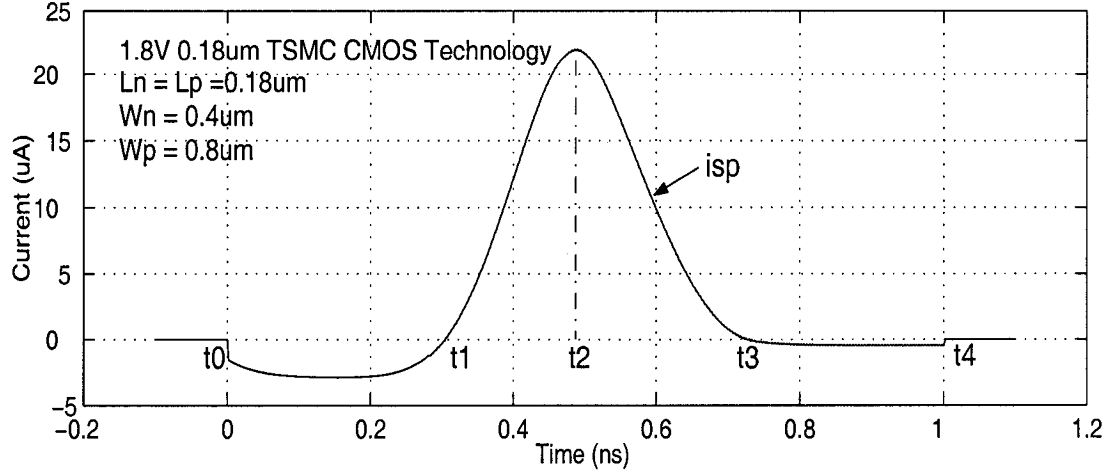
$$i_M = C_M \cdot \left(\frac{dVin}{dt} - \frac{dVout}{dt}\right) \tag{2.27}$$

As V_{in} starts rising/falling from ground/V_{DD}, C_L gets actually overcharged/discharged until the NMOS/PMOS transistor turns on. So, i_M results in an overshoot/undershoot at the early part of the output voltage waveform [36] and an initially negative current flowing from the output node to the power supply, as shown in Figure 2.5(a). Generally, because C_M comes from tiny parasitic capacitances, i_M is also very small. Hence, the overshoot/undershoot at the early part of the output voltage V_{out} is almost not noticeable. For example, as shown in Figure 2.5(b), for an inverter using 1.8V 0.18μm CMOS technology, the maximum voltage of the overshoot is only 1.806V. Its normalized value is 1.003.

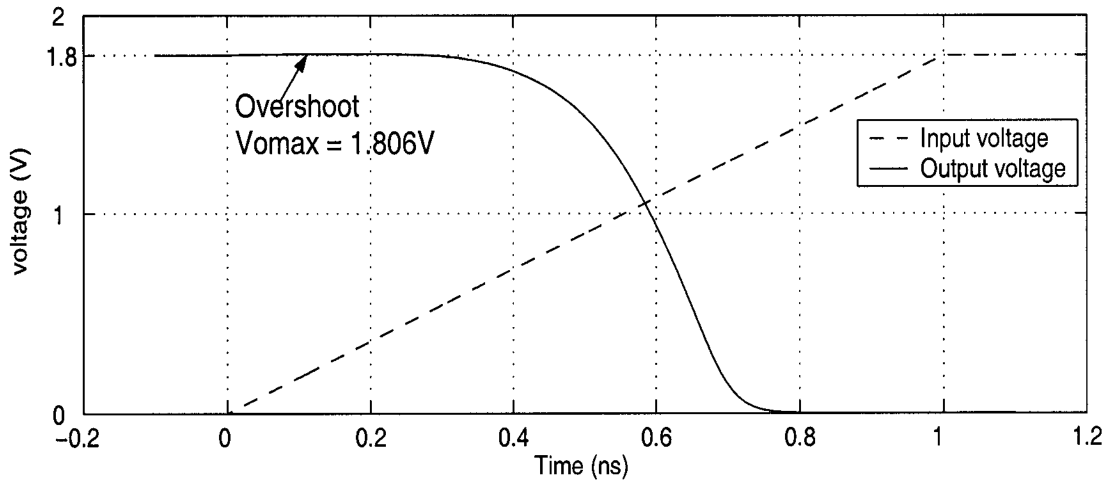
The analysis approach for the rising input transition is symmetrical to that of the falling input transition. The following introduction of the short-circuit energy estimation approach is based on the rising input transition.

In case of rising inputs as shown in Figure 2.4(a), by applying Kirchoff's Current Law at the output node, the current equation taking into account the gate-drain coupling capacitance may be written as:

$$i_n = i_p + i_M + i_{CL} \tag{2.28}$$



(a) Inverter short-circuit current waveform (rising input transition).



(b) Input and output voltage waveforms

Figure 2.5 Inverter response in the case of rising input (1.8V 0.18μm TSMC technology, Ln=Lp=0.18μm, Wn=0.4μm, Wp=0.8μm, and τ_r=1ns). (a) The inverter short-circuit current waveform. (b) The inverter input and output voltage waveforms.

By using Equation 2.27, and $i_{CL} = C_L \frac{dV_{out}}{dt}$, a differential equation describing the inverter behavior of the rising input transition is derived as following:

$$\frac{dV_{out}}{dt} = \frac{i_p - i_n}{(C_M + C_L)} + \frac{C_M \cdot S_r}{C_M + C_L} \tag{2.29}$$

The Toh-Ko-Meyer MOS engineering model [15] can be modified and used to express the drain currents of the NMOS and PMOS transistors in terms of their respective terminal voltages. Along with some appropriate approximations depending on the operation region of the inverter, the short-circuit current in case of rising input can be determined. Among these approximations, the most important approach is linearization.

To illustrate the linearization method more clearly, a typical short-circuit current waveform is shown in Figure 2.6, where the rising input transition begins at time t_0 and ends at time t_4. The short-circuit current is equal to zero at time t_0, t_1, t_3, and t_4 respectively, and reaches its maximum value, i_{spmax}, at time t_2. On the short-circuit current curve, A, C, D, E and F are the points on the curve corresponding to the time t_0, t_1, t_2, t_3, and t_4. While at the point B, the short-circuit current i_{sp} reaches its minimum value.

The short-circuit energy dissipation for this rising input transition, E_{SCr}, is obtained by multiplying V_{DD} with the area between the short-circuit current curve and the zero-current axis, i.e.,

$$E_{SCr} = V_{DD} \int_{t1}^{t3} i_{SC}(t) dt \tag{2.30}$$

Note that the area ABC should not be included in the computation of the short-circuit energy dissipation. This part of energy is provided from the input or the power supply of the previous gate. Only the part under the curve segment CDE, i.e. in the interval t_1 to t_3 is virtually the short-circuit energy dissipation. To compute this area, it is reasonable to approximate the curve by the piecewise linear approach, which is widely accepted

[7][8][27][30]. Let t_2 be the time when the short-circuiting PMOS transistor leaves the linear region and enters saturation. Simulation results show that, for the purpose of computing the short-circuit energy, it is valid to assume that the short-circuit current reaches its maximum value at $t=t_2$. Therefore, the short-circuit energy dissipation during rising input transition consists of two parts, i.e.,

$$E_{SCr} = V_{DD}\left(\int_{t1}^{t2} i_{SC}(t)dt + \int_{t2}^{t3} i_{SC}(t)dt\right) \tag{2.31}$$

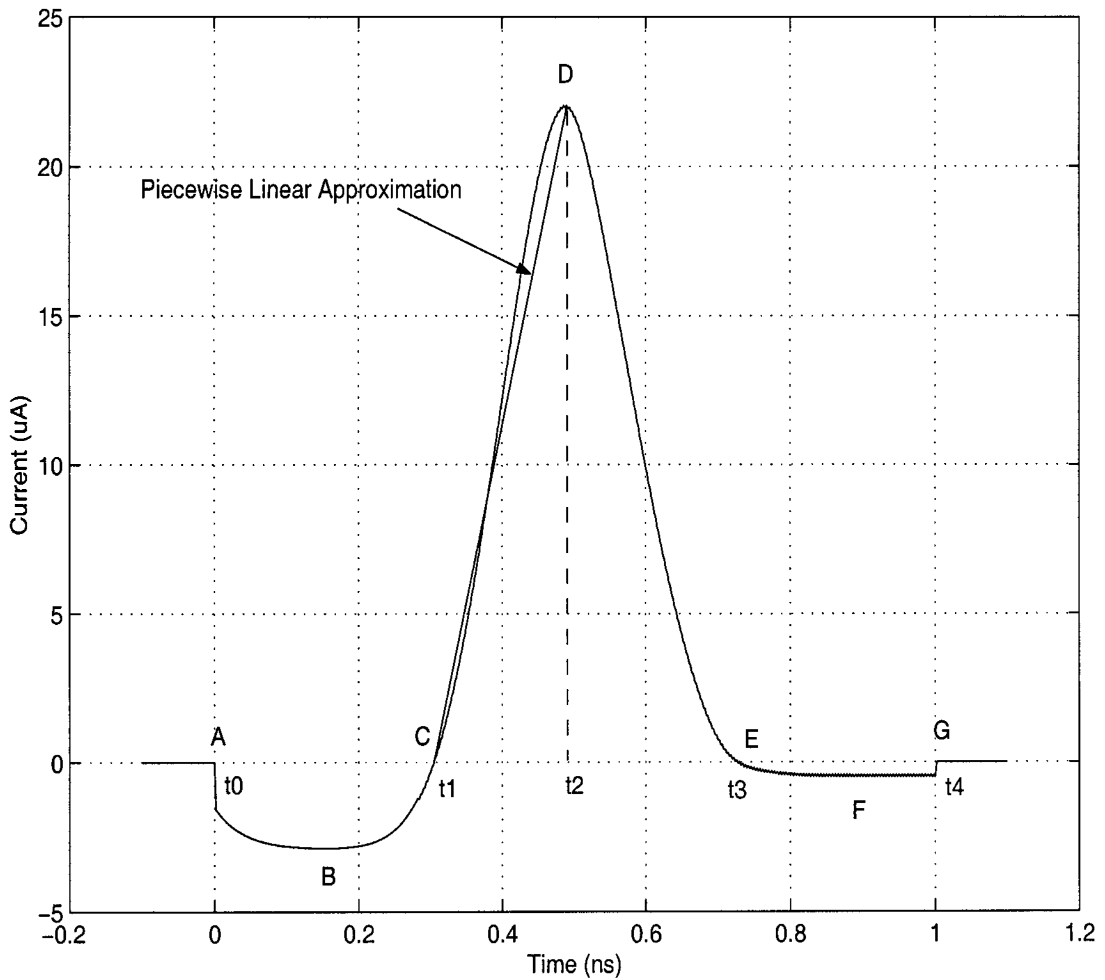


Figure 2.6 The short-circuit current and its piecewise linear approximation.

The expressions for calculating the short-circuit current in the interval t_1-t_2 and t_2-t_3

will be derived in detail in Chapter 4.

The short-circuit energy dissipation for the falling input transition, E_{SCf}, may be derived in a similar way. Then, the inverter short-circuit power dissipation in one switching cycle is obtained by using the follow equation:

$$P_{SC} = f \cdot (E_{SCr} + E_{SCf}) \tag{2.32}$$

where f is the inverter switching frequency.

Chapter 3
Submicron MOSFET Models

To describe the device behavior and to relate the terminal voltages to the drain current in CMOS devices, a MOSFET model should be used. Because of the complexity of the behavior of the MOS transistor and its many parasitic effects, a number of models have been developed for varying degree of accuracy and computing efficiency. Generally, more accuracy means more complexity and, hence, an increased run time. Although the traditional square-law MOSFET model is very simple, it is no longer valid for modern submicron devices, because it neglects the short-channel effects. On the other hand, some accurate submicron MOSFET models involve a large number of parameters. For example, the well-known BSIM3v3 (BSIM for Berkeley Short-channel IGFET Model) is a physically-based, accurate, continuous and scalable MOSFET model. However, it includes almost 200 parameters and a lot of equations [37]. Therefore, these models can not give circuit designers intuitive insight on what parameters are dominant under certain operating conditions, and they are not suitable for quick analysis. Hence, some 'engineering' models for short-channel MOSFETs have been proposed[15][21].

This chapter begins by briefly introducing the square-law MOSFET model and the reasons why it is no longer valid for short-channel devices. A description of the short-channel effects, which must be taken into account when analyzing submicron devices, then follows. In the third and fourth section, the α-power law MOSFET model and the Toh-Ko-Meyer MOS engineering model are introduced respectively. In the last section of

this chapter, the model equations used in this thesis are normalized and listed for easy reference.

3.1 Square-Law Long Channel MOSFET Model

The square-law MOSFET model is basically the model proposed by Shichman and Hodges [19], and it is also referred as LEVEL1 model in SPICE [17]. In this model, the transistor drain current is expressed as following [3][38].

For $0 < V_{ds} < V_{DSsat}$ (Saturation region)

$$I_D = I_{Dsat} \cdot (1 + \lambda \cdot V_{ds}) \tag{3.1}$$

For $0 < V_{DSsat} < V_{ds}$ (Linear region)

$$I_D = \mu_0 \cdot C_{ox} \cdot \frac{W}{L_{eff}} \cdot \left[(V_{gs} - V_t) \cdot V_{ds} - \frac{1}{2} \cdot V_{ds}^2 \right] \tag{3.2}$$

where

$$I_{Dsat} = \frac{1}{2} \cdot \mu_0 \cdot C_{ox} \cdot \frac{W}{L_{eff}} \cdot (V_{gs} - V_t)^2 \tag{3.3}$$

and $$V_{DSsat} = V_{GS} - V_t \tag{3.4}$$

In the above equations, μ_0 is the effective surface mobility of the carriers in the channel; C_{ox} is the gate oxide capacitance per unit area; ε_{ox} and t_{ox} are the permittivity and thickness of the gate dielectric; λ is the channel-length modulation factor, which has a value in the range 0.02 to 0.005 V^{-1} and represents the small influence of V_{ds} on I_D in saturation. To avoid discontinuities in the I_D-V_{ds} characteristic, the empirical correction term $(1+\lambda \cdot V_{ds})$ may be included for both saturated and linear regions [39][40].

The square-law model equations were derived for long-channel MOSFETs where the following two assumptions are valid [41]:

- All current flows on the surface of the silicon and the electrical fields are oriented along the plane. In other words, the current flowing in the conducting channel is taken to be exclusively in the source-drain direction. Then a long-channel device is adequately described by a one-dimensional model.
- The carrier mobility is assumed to be constant, hence the velocity of the carriers is thought to be proportional to the electrical field, independent of the value of that field.

As the dimensions of the device reach the μm-range or below, the channel length becomes comparable to other device parameters such as the depth of drain and source junctions and the width of their depletion regions. In this case, the behavior of a submicron device can deviate substantially from that of a long-channel device, resulting in a breakdown of those assumptions. So, the main purpose of the square-law model, which does not handle short-channel effects, is to verify a manual analysis. To describe the submicron device behavior more accurately, some important short-channel effects must be taken into account.

3.2 Short-Channel Effects

As the device dimensions shrink into the submicron range, the voltage-current relations of a short-channel CMOS device deviate considerably from the expressions of square-law model [20]. The most important reasons for this difference are the velocity saturation and the mobility degradation effects. The short-channel effects also bring some variations with

threshold voltage. Those effects and variations are discussed in this section.

3.2.1 Velocity Saturation

In the analysis of the long-channel devices, it is assumed that the carrier velocity ν is proportional to the electrical field E along the channel, independent of the value of that field:

$$\nu = \mu \cdot E \tag{3.5}$$

where the carrier mobility μ is taken to be a constant. However, when the electrical field along the channel reaches a critical value E_{sat}, the velocity of the carriers tends to saturate as illustrated in Figure 3.1.

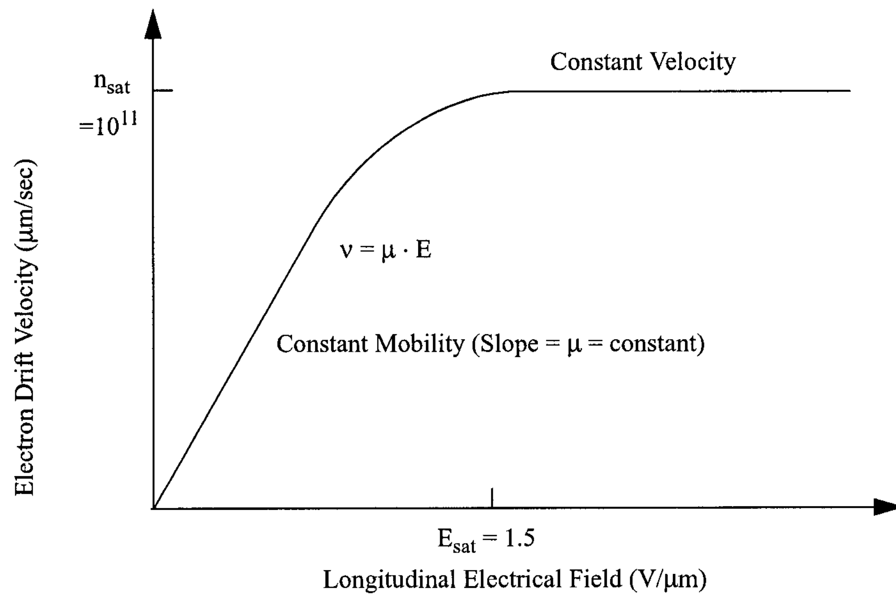


Figure 3.1 Effect of electrical field on electron velocity [34].

At 300K, the average carrier drift velocity saturates at $\nu_{sat} = 10^{11}$ μm/sec when the electrical field reaches its critical value (1.5 V/μm for p-type silicon, and 10V/μm for n-type silicon). Thus in an NMOS device with a channel length of 1 μm, only a couple of

volts between drain an source are enough for electron drift velocity to reach the saturation point [38]. The velocity saturation effect has a profound impact on the operation of the transistors and it can no longer be ignored in the analysis of submicron devices.

The velocity saturation effect (and also the mobility degradation effect) can be combined into an approximate but manageable model for the short-channel MOSFET transistor [38].

For $V_{ds} \geq V_{DSsat}$ (Saturation Region)

$$I_d = K \cdot v_{sat} \cdot C_{ox} \cdot W \cdot (V_{gs} - V_t) \cdot (1 + \lambda V_{ds}) = I_{Dsat} \cdot (1 + \lambda V_{ds}) \tag{3.6}$$

For $V_{ds} \leq V_{DSsat}$ (Linear Region)

$$I_d = \mu_n \cdot C_{ox} \cdot \frac{W}{L} \cdot \left[(V_{gs} - V_t) \cdot V_{ds} - \frac{V_{ds}^2}{2}\right] \tag{3.7}$$

where V_{DSsat} is the drain-source voltage at which the carrier drift velocity saturates:

$$V_{DSsat} = (1 - K) \cdot (V_{gs} - V_t) \tag{3.8}$$

with K a measure of the velocity-saturation degree, given by

$$K = \frac{1}{1 + E_{sat}/E} = \frac{1}{1 + E_{sat} \cdot L/(V_{gs} - V_t)} \tag{3.9}$$

where the factor K, which models the velocity saturation effect, depends on the manufacture technology and increases with (V_{gs}-V_t), though it may be regarded as constant under certain conditions [15]. For devices with finite channel length and operating under practical bias conditions, it is found that $0<K<1$ in saturation region.

In Eq.(3.7), the mobility μ_n is not a constant, but a function of the applied electrical field. And the variation of μ_n is also due to the mobility degradation effect, which will be

discussed in more detail in Section 3.3.2.

By comparing Eq.(3.1) to (3.4) with Eq.(3.6) to (3.9), it can be observed that the velocity saturation has three main impacts on device characteristics:

- In contrast with the squared dependence for long-channel devices, I_{Dsat} exhibits an almost linear dependence on (V_{gs}-V_t). Consequently, the current drive capability of submicron devices is less sensitive to changes in the operating voltage.
- I_d is independent of L in velocity-saturated devices if ignoring the influence of L on V_{DSsat}, suggesting that current cannot be further improved by decreasing the channel length as in long-channel transistors.
- V_{DSsat} is reduced, thereby extending the saturation region of a short-channel MOSFET as compared to a long-channel device.

3.2.2 Mobility Degradation

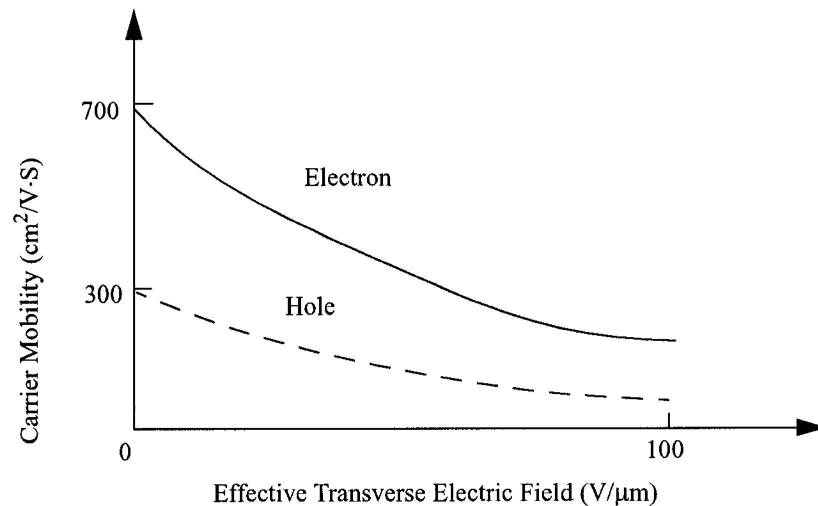


Figure 3.2 Mobility degradation effect

In a long-channel transistor, the carrier mobility is assumed to be constant. However,

after reducing the channel length even at normal electric field levels, the mobility degradation effect, which is illustrated in Figure 3.2[15][38], can be observed. The reduction in carrier mobility can be attributed to the vertical component of the electrical field, which is no longer ignorable in these small devices.

The value of mobility is critical to the accuracy of a MOSFET model [15]. In general, the mobility depends on many process parameters and bias conditions, such as the gate oxide thickness, substrate doping concentration, threshold voltage, gate and substrate voltage, etc. [37]. A lot of mobility models have been developed. One of them is a simple first-order empirical equation used for LEVEL3 SPICE model [39]:

$$\mu = \frac{\mu_0}{1 + \theta \cdot (V_{gs} - V_t)} \tag{3.10}$$

where μ_0 is the low-field mobility; θ (named THETA in SPICE) is an empirical parameter that takes into account the mobility modulation.

For better accuracy, an unified mobility formulation can be used [37]:

$$\mu = \frac{\mu_0}{1 + (E_{eff}/E_0)^{\nu}} \tag{3.11}$$

where μ_0, E_0, and ν are empirical constants, and their values are shown in Table 3.1.

Table 3.1 Typical mobility values for electrons and holes [37].

Parameter	Electron (surface)	Hole (surface)
μ_0 (cm^2/Vsec)	670	160
E_0 (mV/cm)	0.67	0.7
ν	1.6	1.0

This equation fits experimental data very well [42], but it involves a very time consuming power function in numerical simulation. Thus, the Taylor expansion of

Eq.(3.11), shown as following, is used in BSIM3V3 model [37].

$$\mu = \frac{\mu_0}{1 + (U_a + U_c \cdot V_{bs}) \cdot \left(\frac{V_{gs} + V_t}{t_{ox}}\right) + U_b \cdot \left(\frac{V_{gs} + V_t}{t_{ox}}\right)^2} \tag{3.12}$$

where U_a, U_b, and U_c are the first-order carrier mobility degradation factor, the second-order carrier mobility degradation factor, and the body bias sensitivity coefficient of mobility, respectively. V_{bs} is the bulk-to-source voltage.

3.2.3 Threshold Voltage Variations

Accurate modeling of threshold voltage is one of the most important requirements for precise description of device electrical characteristics. To take into account their impacts on the operation of MOSFETs, the short-channel effects may be introduced in the calculation of threshold voltage [39].

In a long-channel MOSFET with uniform substrate doping concentration, the threshold voltage V_t is assumed to be a function of only the device technology and the applied body bias V_{sb} [37][38]:

$$V_t = V_{t0} + \gamma \cdot (\sqrt{\phi_s - V_{bs}} - \sqrt{\phi_s}) \tag{3.13}$$

where V_{th0} is the threshold voltage of the long-channel device at zero substrate bias, γ is the body bias coefficient, and ϕ_s is the surface potential.

Equation (3.13) is valid only when the substrate doping concentration is constant and the channel length is long. Under these conditions, the potential is uniform along the channel [37]. As the device dimensions shrink into μm range, modifications have to be made on the threshold voltage equation because of some physical phenomena such as non-

uniform doping effect, short and narrow channel effects.

3.2.3.1 Non-Uniform Doping Effects

There are two kinds of non-uniform doping effect — vertical non-uniform doping effect and lateral non-uniform doping effect. The former is due to the non-uniform substrate doping profile in the vertical direction as shown in Figure 3.3 [37].

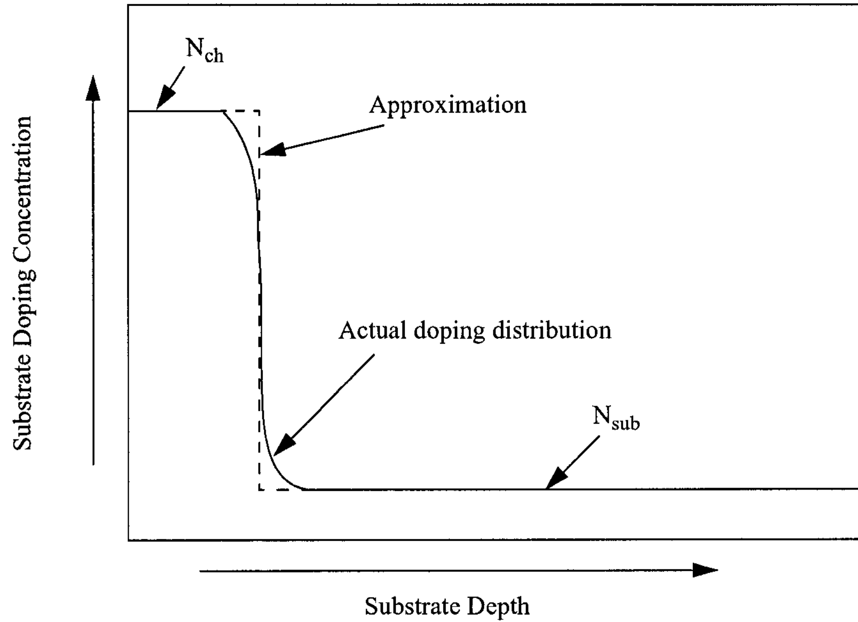


Figure 3.3 Vertical substrate doping distribution and its approximation.

As shown above, the substrate doping concentration is usually higher near the Si/SiO_2 interface than deep into the substrate. The distribution curve of impurity atoms inside the substrate looks like a half gaussian distribution. This non-uniformity makes γ in Eq.(3.14) a function of the substrate bias.

The lateral non-uniform doping effect results from the higher doping concentration near the source/drain than that in the middle of the channel. As the channel length becomes shorter, this effect will cause V_t to increase in magnitude because the average doping concentration in the channel is larger.

3.2.3.2 Short Channel Effects

As the device dimensions are reduced, the threshold voltage becomes a function of L and V_{ds}, and decreases with shrinking channel length L, as illustrated in Figure 3.4 (a) [38][41].

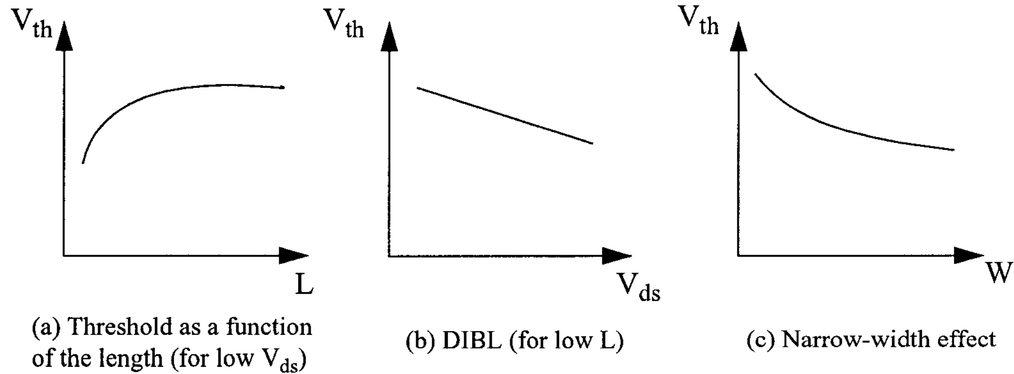


Figure 3.4 Threshold variations

Before an inversion layer or channel forms in a transistor, the substrate region under the gate must be first depleted from the majority carrier. For long-channel devices, the value of V_t is predicted using the gradual-channel approximation, which is based on a one-dimensional theory. It is assumed that all the depletion region charge beneath the gate originates from the MOS field effects. This ignores the depletion regions of the source and reverse-biased drain junctions.

With device dimensions shrinking into the nm range, the channel length becomes comparable to the depletion region width. Consequently, less gate charge is required to invert the channel, because a part of the region below the gate is already depleted by the source and drain fields. This effect is also referred to as drain/source charge sharing [37]. Hence, a smaller V_t suffices to cause strong inversion. In other words, V_t decreases with L

for short-channel devices.

A similar effect can be observed by raising the drain-source (bulk) voltage, as this increases the drain-junction depletion region. Therefore, V_t decreases with increasing V_{ds}. This effect, called the drain-induced barrier lowering (DIBL), causes the threshold voltage to be a function of the transistor operating voltage, as illustrated in Figure 3.4(b) [38].

3.2.3.3 Narrow-Width Effect

In narrow-channel MOSFETs, V_t is found to monotonically increase with decreasing channel width W, as illustrated in Figure 3.4(c) [41].

The actual depletion region in the channel is always larger than what is usually assumed under the one-dimensional analysis due to the existence of fringing fields [43]. As the channel width decreases, the depletion region underneath the fringing field becomes comparable to the "classical" depletion layer formed from the vertical field. Then, the gate-controlled charge on either side of the gate width is no longer negligible as compared to the charge directly beneath the W-width of the gate. Consequently, as W decreases, this effect becomes more substantial, and more gate charge is required to invert the channel because a large percentage of the gate-induced space charge is lost in fringing fields. Therefore, the result is an increase in V_t [37].

3.3 α-power law MOSFET model

As mentioned in Chapter 2, since many short-circuit power estimation methods use the α-power law MOSFET model [20] (extended as the nth power law model in [21]), it is necessary to review this model in more detail. The α-power law MOSFET model, first proposed by Sakurai and Newton in [20], offers a simple, yet accurate enough, empirical

model for the MOSFET drain current.

3.3.1 Model Equations

The equations of the four-parameter α-power law MOSFET model is as following:

$$I_D = \begin{cases} 0 & (V_{gs} \le V_{th},\ cutoff) \\ \frac{I_{D0}}{V_{D0}} \cdot \left(\frac{V_{gs} - V_{th}}{V_{DD} - V_{th}}\right)^{\frac{\alpha}{2}} \cdot V_{ds} & (V_{ds} < V_{DSsat},\ triode) \\ I_{D0} \cdot \left(\frac{V_{gs} - V_{th}}{V_{DD} - V_{th}}\right)^{\alpha} & (V_{ds} \ge V_{DSsat},\ saturation) \end{cases} \tag{3.14}$$

where

$$V_{DSsat} = V_{D0} \cdot \left(\frac{V_{gs} - V_{th}}{V_{DD} - V_{th}}\right)^{\frac{\alpha}{2}} \tag{3.15}$$

As shown above, the α-power law MOSFET model is based on four parameters: V_{th} (threshold voltage), α(velocity saturation index), V_{D0}(drain saturation voltage at V_{gs}=V_{DD}), and I_{D0}(drain current at V_{gs}=V_{ds}=V_{DD}). These parameters require to be extracted from measured data.

3.3.2 Extraction of Model Parameters

Among the four parameters of α-power law MOSFET model, V_{D0}(the drain saturation voltage at V_{gs}=V_{DD}), and I_{D0}(the drain current at V_{gs}=V_{ds}=V_{DD}) may be easily obtained from the measured data.

As for V_{th} and α, there are two extraction methods described in [20]. One uses brute force, and the other involves equation solving. The procedure of the second method is:

1. From the measured I-V characteristics, pick three points, namely (V_{gs1}, I_{D1}), (V_{gs2}, I_{D2}), and (V_{gs3}, I_{D3}).

2. Then, V_{th} can be obtained by solving the following equation:

$$f(V_{th}) = \log\left(\frac{I_{D1}}{I_{D2}}\right) \cdot \log\left(\frac{V_{gs2} - V_{th}}{V_{gs3} - V_{th}}\right) - \log\left(\frac{I_{D2}}{I_{D3}}\right) \cdot \log\left(\frac{V_{gs1} - V_{th}}{V_{gs2} - V_{th}}\right) = 0 \quad (3.16)$$

3. α can be obtained from the following expression:

$$\alpha = \frac{\log(I_{D1}/I_{D2})}{\log((V_{gs1} - V_{th})/(V_{gs2} - V_{th}))} \quad (3.17)$$

3.3.3 Discussions on the Model

From Section 3.3.1 and 3.3.2, the four parameters of α-power law MOS model require extraction from measured I-V characteristics for each transistor width [36]. When the transistor width is changed in a wide range, the effort to extract those parameters will significantly increase.

Still, the α-power law MOS model provides enough accurate estimates of the drain current only when used for wide transistors, for example, the transistors with width at least 14X the minimum width for 0.25μm technology [33], since those transistors do not suffer from narrow-width effects. Hence, in DSM regime the α-power law MOS model is not sufficiently accurate for digital circuits, where the transistor width ranges from minimum size up to 8X the minimum width.

To improve the accuracy of this model, some modified versions of α-power law MOS model are proposed [7][32][33]. However, more extracted/fitting parameters are introduced in these modified versions. In a word, to utilize this kind of model, some extra

efforts are required to extract and optimize empirical parameters to characterize a given MOS process technology.

3.4 Toh-Ko-Meyer MOS Engineering Model

In order to give circuit designers an intuitive understanding of the MOS device electrical behavior from a relatively small number of parameters, an engineering model for short-channel MOS devices was proposed in [15]. This original model, which includes the effect of carrier drift velocity saturation, can relate the terminal voltages to the drain current of MOS devices with effective channel lengths down to 1 μm. To take into account the effect of carrier mobility degradation, this model is modified.

3.4.1 Toh-Ko-Meyer MOS Engineering Model

The simple Toh-Ko-Meyer MOS engineering model for short-channel devices gives the closed-form I-V relationship between the terminal voltages and drain current I_d, and much likes the well-known square-law I-V relationship for the long-channel devices. In Section 3.2.1, it is briefly introduced as an approximate but manageable model, because of the ease of combining the velocity saturation effect and the mobility degradation effect into this model. Here, its original formula is reviewed. The current is expressed as following [15].

$$I_{d(lin)} = \frac{\mu_{eff} \cdot C_{ox} \cdot W}{L_{eff}} \cdot \frac{1}{1 + \dfrac{V_{ds}}{E_c \cdot L_{eff}}} \cdot \left[(V_{gs} - V_t) \cdot V_{ds} - \frac{V_{ds}^2}{2} \right] \quad \text{For } V_{ds} \le V_{DSsat} \tag{3.18}$$

$$I_{d(sat)} = K \cdot v_{sat} \cdot C_{ox} \cdot W \cdot (V_{gs} - V_t) \quad \text{For } V_{ds} \ge V_{DSsat} \tag{3.19}$$

where C_{ox} gate capacitance per unit area,

L_{eff} effective channel length between the source and the drain,

v_{sat} saturated carrier drift velocity,

E_c critical value of the electrical field along the channel at which the carrier velocity tends to saturate, and it is given as

$$E_c \equiv \frac{2v_{sat}}{\mu_{eff}} \tag{3.20}$$

where, μ_{eff} is the effective carrier mobility, whose value is crucial to the accuracy of the device model. In [15], μ_{eff} is estimated as follows:

$$\frac{1}{\mu_{eff}} = \frac{1}{\mu_1} + \frac{1}{\mu_2} \tag{3.21}$$

where $$\mu_2 = \mu_1 \cdot (E_1/E_{t,\,eff})^n \tag{3.22}$$

μ_1, E_1, and n are empirical constants. $E_{t,eff}$ is the effective transverse field at the surface and is approximated as follows:

$$E_{t,\,eff} \approx \frac{V_{gs} - V_t}{6t_{ox}} + \frac{V_t + V_a}{3t_{ox}} \tag{3.23}$$

where, semi-empirically, $V_a \approx 0.5V$ for typical n^+ polysilicon gate devices.

In this thesis, the μ_{eff} expression of BSIM3v3 model, i.e. Eq.(3.12) is used to calculate the value of μ_{eff}. The reason will be given later.

In Eq.(3.18) and (3.19), V_{DSsat} is defined as the drain to source voltage, at which the carrier drift velocity v saturates, and is given as

$$V_{DSsat} = (1 - K) \cdot (V_{gs} - V_t) \tag{3.24}$$

where
$$K = \frac{1}{1 + E_c \cdot L_e / (V_{gs} - V_t)} \tag{3.25}$$

and
$$L_e = L_{eff} - X_d \tag{3.26}$$

In above equations, L_e is the device electrical channel length and X_d is the depletion width into the channel from the drain when $V_{ds} > V_{DSsat}$. X_d may be calculated as following:

$$X_d \approx \frac{1}{A} \cdot \ln\left[\frac{A \cdot (V_{ds} - V_{DSsat}) + E_d}{E_c}\right] \tag{3.27}$$

where
$$\frac{E_d}{E_c} = \sqrt{1 + \left[\frac{A \cdot (V_{ds} - V_{DSsat})}{E_c}\right]^2} \tag{3.28}$$

and
$$A^2 = \frac{\varepsilon_{ox}}{\varepsilon_{si}} \cdot \frac{1.5}{x_j \cdot t_{ox}} \tag{3.29}$$

in which 1.5 is a semi-empirical constant, and x_j is the source/drain junction depth.

To obtain the value of V_{DSsat} in Eq. (3.24), Eq. (3.25) to (3.29) should be used. However, in Eq. (3.27) and (3.28), V_{DSsat} is a parameter whose value is supposed to be known. Obviously, a process of iteration is required. To overcome this disadvantage, an approximation to the factor K [44] was proposed. By applying Taylor series expansion and eliminating terms containing the second and higher order derivatives, the expression of K can be found. However, in order to reduce the complexity of the computation, L_{eff} may be used instead of L_e in Eq.(3.25) [6]. For the same reason, Eq.(3.12) is used to calculate the

effective mobility μ_{eff}. Using Eq.(3.12) not only avoids the introduction of several empirical and semi-empirical parameters in Eq. (3.21) to (3.23), but also makes the model technology portable.

3.4.2 Normalized Inverter Current Model Equations

In this thesis, normalized voltages with respect to the supply voltage (V_{DD}), i.e., $u_{in}=V_{in}/V_{DD}$, $u_{out}=V_{out}/V_{DD}$, $n=V_{thn}/V_{DD}$, $p=|V_{thp}|/V_{DD}$, $u_{DSsatn}=V_{DSsatn}/V_{DD}$, $u_{DSsatp}=V_{DSsatp}/V_{DD}$, and the variable $x=t/\tau$ where τ is the input transition time, are used. Consequently, the current equations of the NMOS and the PMOS devices in a CMOS inverter are given as following.

The current equations for the NMOS device of an inverter are as follows.

$$i_{dn(lin)} = \frac{\mu_{effn} \cdot K_{l1n}}{1 + \mu_{effn} \cdot K_{l2n} \cdot u_{out}} \cdot \left(x - n - \frac{u_{out}}{2}\right) \cdot u_{out} \quad \text{, for } u_{out} \le u_{DSsatn} \tag{3.30}$$

$$i_{dn(sat)} = K_n \cdot K_{vn} \cdot (x - n) \quad \text{, for } u_{out} \ge u_{DSsatn} \tag{3.31}$$

where
$$K_n = \frac{\mu_{effn} \cdot K_{l2n} \cdot (x - n)}{1 + \mu_{effn} \cdot K_{l2n} \cdot (x - n)} \tag{3.32}$$

$$K_{vn} = v_{satn} \cdot C_{oxn} \cdot W_{effn} \cdot V_{DD} \tag{3.33}$$

$$K_{l1n} = \frac{C_{oxn} \cdot W_{effn} \cdot V^2_{DD}}{L_{effn}} \tag{3.34}$$

$$K_{l2n} = \frac{V_{DD}}{2 \cdot v_{satn} \cdot L_{effn}} \tag{3.35}$$

$$\mu_{effn} = \frac{\mu_{0n}}{1 + U_{an} \cdot V_{DD} \cdot \left(\frac{x+n}{t_{oxn}}\right) + U_{bn} \cdot V_{DD}^2 \cdot \left(\frac{x+n}{t_{oxn}}\right)^2} \tag{3.36}$$

$$u_{DSsatn} = \frac{V_{DSsatn}}{V_{DD}} = (1 - K_n)(x - n) \tag{3.37}$$

The current equations for the PMOS device of an inverter are as follows.

$$i_{dp(lin)} = \frac{\mu_{effp} \cdot K_{l1p} \cdot (1 - u_{out})}{1 + \mu_{effp} K_{l2p}(1 - u_{out})}\left(1 - x - p - \frac{1 - u_{out}}{2}\right) \quad \text{, for } 1\text{-}u_{out} \le u_{DSsatp} \tag{3.38}$$

$$i_{dp(sat)} = K_p \cdot K_{vp} \cdot (1 - x - p) \quad \text{, for } 1\text{-}u_{out} \ge u_{DSsatp} \tag{3.39}$$

$$\text{where} \quad K_p = \frac{\mu_{effp} \cdot K_{l2p} \cdot (1 - x - p)}{1 + \mu_{effp} \cdot K_{l2p} \cdot (1 - x - p)} \tag{3.40}$$

$$K_{vp} = \nu_{satp} \cdot C_{oxp} \cdot W_{effp} \cdot V_{DD} \tag{3.41}$$

$$K_{l1p} = \frac{C_{oxp} \cdot W_{effp} \cdot V^2_{DD}}{L_{effp}} \tag{3.42}$$

$$K_{l2p} = \frac{V_{DD}}{2 \cdot \nu_{satp} \cdot L_{effp}} \tag{3.43}$$

$$\mu_{effp} = \frac{\mu_{0p}}{1 + U_{ap} \cdot V_{DD} \cdot \left(\frac{1 - x + p}{t_{oxp}}\right) + U_{bp} \cdot V_{DD}^2 \cdot \left(\frac{1 - x + p}{t_{oxp}}\right)^2} \tag{3.44}$$

$$u_{DSsatp} = \frac{V_{DSsatp}}{V_{DD}} = (1 - K_p)(1 - x - p) \tag{3.45}$$

Chapter 4
Inverter Short-Circuit Power Modeling

Following the methodology described in Chapter 2, the procedures and expressions to estimate the CMOS inverter short-circuit energy dissipation in rising input transition is given first. Then, they are extended to the case of falling input transition. In the last section of this chapter, the expression of short-circuit power dissipation is given, and the model parameters and equations are summarized.

4.1 Short-circuit Energy Dissipated in Rising Input Transition

As mentioned in Section 2.4.2, it is widely accepted to assume that the input voltage waveform is a rising ramp with transition time τ_r:

$$V_{in}(t) = \begin{cases} 0 & t \le 0 \\ V_{DD} \cdot (t/\tau_r) & 0 \le t \le \tau_r \\ V_{DD} & t \ge \tau_r \end{cases} \tag{4.1}$$

Then, the slope of the rising input voltage ramp can be determined:

$$S_r = V_{DD} / \tau_r, \tag{4.2}$$

By applying Kirchoff's current law at the output node of the inverter shown in Figure 4.1, the current equation taking into account the gate-drain coupling capacitance may be written as:

$$i_n = i_p + i_M + i_{CL} \tag{4.3}$$

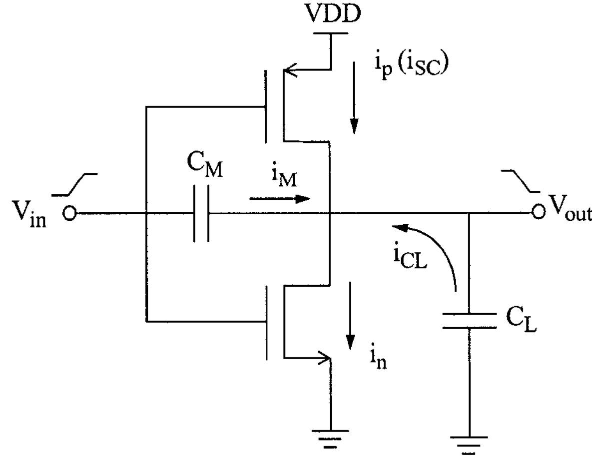


Figure 4.1 Rising input CMOS inverter with explicitly modeled parasitic capacitances

By using the normalized parameters described in Section 3.3.2, a differential equation describing the device behavior in the case of rising input transition is derived as following:

$$\frac{du_{out}}{dx} = \frac{i_p - i_n}{S_r \cdot (C_M + C_L)} + \frac{C_M}{C_M + C_L} \tag{4.4}$$

Consider application of such an input ramp that the NMOS transistor is still in saturation when the input voltage ramp reaches the value V_{DD}-$|V_{thp}|$, which is true for the most practical cases in VLSI circuits. The short-circuit current, input and output voltage waveforms are shown in Figure 4.2, where the operation of the inverter is divided into three regions.

Region 1 ($0 \leq x \leq n$): The PMOS transistor is in the linear region while the NMOS transistor is off, i.e. $i_n = 0$. Thus, there is no short-circuit dissipation. However, the expression of the output voltage is required, in order to prepare the initial condition for the computation of the next region.

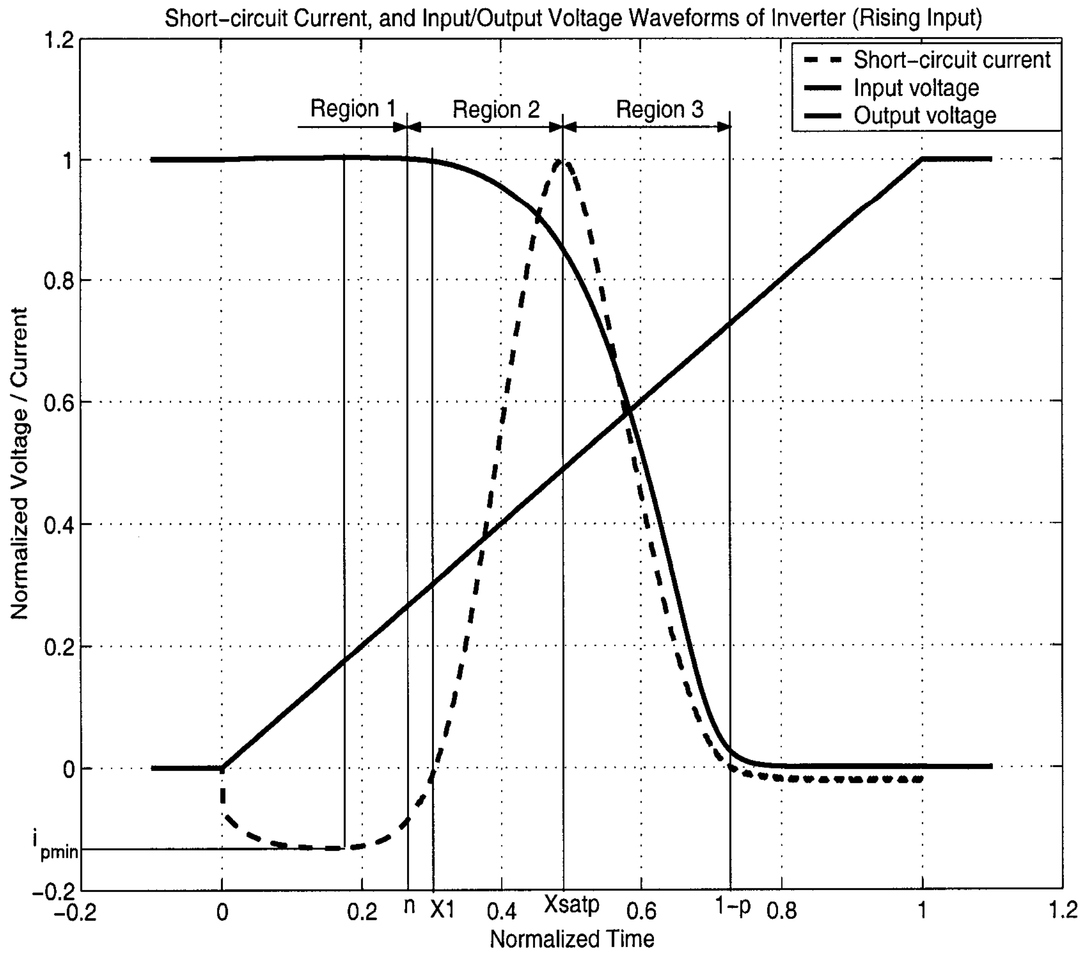


Figure 4.2 Operation regions of a typical inverter. The voltage, current and time are respectively normalized with respect to V_{DD}, i_{SCmax} and input transition time τ_r.

With input voltage rising, the gate-drain coupling capacitance, C_M, allows a current i_M flow from the input node to the output node [36], causing an overshoot in the early part of the output voltage waveform and an initially negative current flowing from the output node to the power supply.

Substituting $i_n = 0$ and Eq.(3.38) in Eq.(4.6) yields

$$\frac{du_{out}}{dx} = \frac{C_M}{C_L + C_M} + \frac{\mu_{effp} C_{oxp} W_{effp} V_{DD} \tau_r}{L_{effp} \cdot (C_M + C_L)} \cdot \frac{(1-x-p)(1-u_{out}) - \frac{1}{2}(1-u_{out})^2}{1 + \frac{\mu_{effp} \cdot V_{DD}}{2 \cdot \nu_{satp} \cdot L_{effp}} \cdot (1-u_{out})} \tag{4.5}$$

From Eq.(3.44), it is known that μ_{effp} varies with the value of input voltage. However, in this region ($0 \le V_{in} \le n$), using the average value of μ_{effp}, i.e. $\mu_{effp,n/2}$, instead of μ_{effp} in Eq.(4.5) is reasonable. Thus, when $V_{in} = V_{thn}/2$, $x = n/2$, then,

$$\mu_{effp,\frac{n}{2}} = \frac{\mu_{0p}}{1 + U_{ap} \cdot V_{DD} \cdot \left(\frac{1-n/2+p}{t_{oxp}}\right) + U_{bp} \cdot V_{DD}^2 \cdot \left(\frac{1-n/2+p}{t_{oxp}}\right)^2} \tag{4.6}$$

With the above approximation, equation (4.5) can be written as:

$$\frac{du_{out}}{dx} = C_{MLM} + A_p \cdot \frac{(1-x-p)(1-u_{out}) - \frac{1}{2}(1-u_{out})^2}{1 + B_p \cdot (1-u_{out})} \tag{4.7}$$

where
$$C_{MLM} = \frac{C_M}{C_L + C_M} \tag{4.8}$$

$$A_p = \frac{\mu_{effp,\frac{n}{2}} C_{oxp} W_{effp} V_{DD} \tau_r}{L_{effp} \cdot (C_M + C_L)} \tag{4.9}$$

$$B_p = \frac{\mu_{effp,\frac{n}{2}} \cdot V_{DD}}{2 \cdot \nu_{satp} \cdot L_{effp}} \tag{4.10}$$

Equation (4.7) is an Abel's ODE (Ordinary Differential Equation) of the second kind, class A (Abel2a for short), for which there is no general solution [45]. Thus, by using a power-series expansion method, the solution of equation (4.7) may be given as following:

$$u_{out1} = 1 - \sum_{j=1}^{\infty} f_j \cdot x^j \tag{4.11}$$

where,

$$f_1 = -C_{MLM} \tag{4.12}$$

$$f_2 = \frac{A_p}{2}(p-1) \cdot f_1 \tag{4.13}$$

and for j > 2

$$f_j = \frac{A_p}{j}\left[f_{j-2} + (p-1)f_{j-1} + \frac{1}{2}\sum_{i=1}^{j-2}(f_i f_{j-i-1})\right] - \frac{B_p}{j}\sum_{i=1}^{j-2}[(j-i)f_i f_{j-i}] \tag{4.14}$$

The main purpose of Eq.(4.11) is to find the normalized output voltage value at x=n, i.e. $u_{out1}\big|_{x=n}$, which is the initial condition of the next region.

In region 1, another useful parameter that should be figured out is the minimum value of the PMOS current, i_{pmin}, as shown in Figure 4.2. The determination of its value is straightforward. Assume that the short-circuit current reaches its minimum value, i_{pmin}, when the output voltage reaches its maximum value. Since in this region, the NMOS device is off (i.e. i_n=0), equation (4.4) becomes

$$\frac{du_{out}}{dx} = \frac{i_{pmin}}{S_r \cdot (C_M + C_L)} + \frac{C_M}{C_M + C_L} = 0 \tag{4.15}$$

Then, the value of i_{pmin} may be easily obtained using following equation.

$$i_{pmin} = -C_M \cdot S_r \tag{4.16}$$

Simulation shows that the results calculated using the above equation are very close to Spectre simulation results for inverters with different sizes.

Region 2 ($n \leq x \leq x_{satp}$): x_{satp} is the normalized time value when the PMOS device is entering the saturation region. In this operation region, the NMOS transistor is saturated, while the PMOS transistor is still in the linear region. Both of the two transistors are on, resulting in a direct path from power supply to ground. However, as shown in Figure 4.2, the short-circuit current does not start from the point $x = n$, but from the point $x = x_1$, which is the normalized time value when the output voltage finishes its overshoot. To evaluate the short-circuit current in this region, a four-step approach is used, which includes two times approximation, resulting in better accuracy.

Step 1:

As shown in Figure 4.3, a rough approximation of the PMOS current, i_p should be derived first. Assuming that the minimum of the PMOS current appears when the input voltage arrives at the NMOS threshold voltage, i.e. $x = n$, then i_p can be approximated by a linear function of the normalized time [8]:

$$i_p = i_{pmin} + S_{1p} \cdot (x - n) \tag{4.17}$$

where the minimum PMOS current, i_{pmin} is given by Eq.(4.16). The PMOS current slope S_{1p} is calculated by equating the exact PMOS current in the linear region with the approximated one, at the point $x_c = (1-p)/2$.

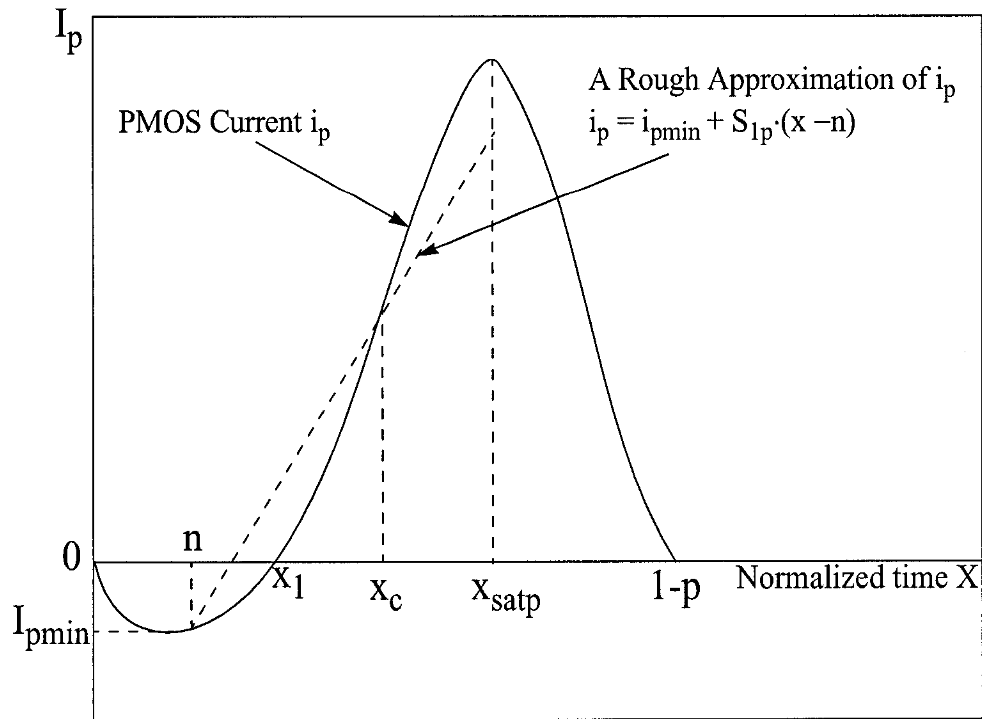


Figure 4.3 The first approximation of the PMOS current in region 2.

Step 2:

The analytical expression of the output voltage in region 2 is required for the evaluation of the short-circuit energy dissipation. To start with, however, the calculation of the NMOS current is simplified.

In region 2, because the NMOS transistor is saturated, Eq.(3.31) to (3.33) can be used to calculate the NMOS current $i_{dn(sat)}$. However, in Eq.(3.32), μ_{effn} is a function of x, as shown in Eq.(3.36). Thus, K_n is a very complex function of x, and that makes it very difficult to solve differential equation (4.4). In order to reduce the computation complexity, the factor K (K_n and K_p) can be approximated by a linear function of x for both NMOS and PMOS transistors when they are in saturation [6].

$$K_n \approx \frac{K_n^{dd}}{V_{DD} - V_{tn}}(V_{gsn} - V_{tn}) = S_{kn}(x - n) \tag{4.18}$$

$$K_p \approx \frac{K_p^{dd}}{V_{DD} - |V_{tp}|}(V_{sgp} - |V_{tp}|) = S_{kp}(1 - x - p) \tag{4.19}$$

where,
$$S_{kn} = \frac{K_n^{dd}}{1 - n} \tag{4.20}$$

$$S_{kp} = \frac{K_p^{dd}}{1 - p} \tag{4.21}$$

$$K_n^{dd} = K_n\big|_{x=1} \tag{4.22}$$

$$K_p^{dd} = K_p\big|_{x=0} \tag{4.23}$$

With the aid of Eq.(4.18) and (4.20), Eq.(3.31) becomes

$$i_{dn(sat)} = S_{kn} \cdot K_{vn} \cdot (x - n)^2 \tag{4.24}$$

With the above simplifications, the analytical expression of the output voltage in region 2 can be easily obtained by integrating Eq.(4.4). Substituting Eq.(4.17) and Eq.(4.24) for i_p and i_n in Eq.(4.4), and then solving it for u_{out} yields:

$$u_{out2} = \frac{S_{kn}K_{vn}(x - n)^3/3 - S_{1p}(x - n)^2/2 - x(i_{pmin} + C_M S_r)}{S_r \cdot (C_M + C_L)} + Const1 \tag{4.25}$$

where the integration constant, Const1, is inserted to ensure continuity with respect to region 1, and it can be given by $u_{out1}\big|_{x=n} = u_{out2}\big|_{x=n}$. Thus, with the aid of Eq.(4.11), the expression of Const1 is given as following.

$$Const1 = 1 - \sum_{i=1}^{\infty} f_j \cdot x^j - \frac{(i_{pmin} + C_M \cdot S_r) \cdot n}{S_r \cdot (C_M + C_L)} \tag{4.26}$$

Step 3:

In this step, the values of x_1 and x_{satp} are figured out.

At the time x_1, the output voltage finishes its overshoot. So x_1 may be calculated by the following equation

$$u_{out2}\big|_{x = x1} = 1 \tag{4.27}$$

x_{satp} is the normalized time value when the PMOS transistor is entering the saturation region. Thus, at the time x_{satp}, the saturation condition of Eq.(3.39) is satisfied.

$$1 - u_{out2}\big|_{x = xsatp} = u_{DSsatp}\big|_{x = xsatp} \tag{4.28}$$

x_{satp} can be solved by substituting Eq.(4.25) and Eq.(3.45) into the above equation.

It should be noted that both Eq.(4.27) and Eq.(4.28) are polynomial equations of degree 3. That means, each of the equations has three solutions, and only the solution between n and 1-p is reasonable. Although it is possible to find the analytical expressions of those solutions by using some symbolic computation tools such as Maple, it is not practical to give such kind of expressions, because those expressions are very complex and verbose, and even worse, it is impossible to discern which solution is reasonable just from those solution expressions.

Step 4:

As shown in Figure 4.4, in [x_1, x_{satp}], the PMOS current, actually the short-circuit current can be more accurately approximated by a linear equation [8]:

$$i_{SC2} = S_{2p} \cdot (x - x_1) \tag{4.29}$$

where S_{2p} is the slope of i_{SC} in the interval of $[x_1, x_{satp}]$, and is calculated by equating the exact PMOS current in the linear region (described by Eq.(3.38)) with the approximated one, at the middle of this interval, i.e. $x = (x_1+x_{satp})/2$. Then, the slope S_{2p} can be calculated from the following equation.

$$i_{dp(lin)}\Big|_{x = \frac{x1 + xsatp}{2}} = S_{2p}(x - x_1)\Big|_{x = \frac{x1 + xsatp}{2}} \tag{4.30}$$

Through the above four steps, the short-circuit current in region 2 (actually in the interval of $[x_1, x_{satp}]$), i.e. i_{SC2}, can be accurately described.

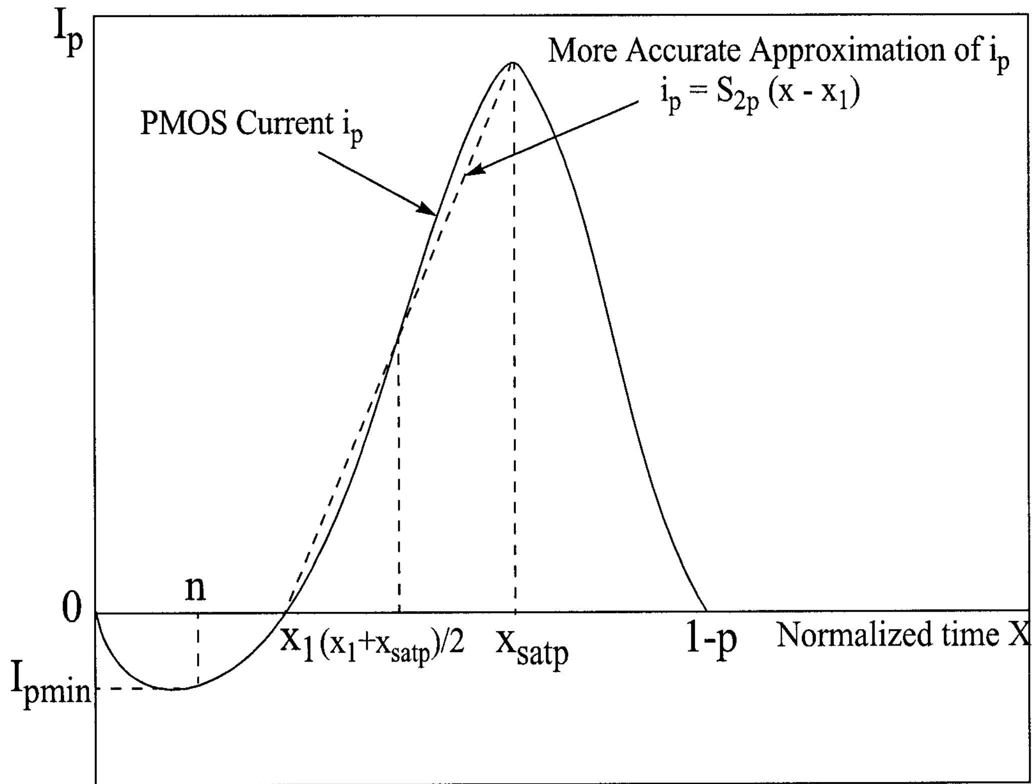


Figure 4.4 The second approximation of the PMOS current in region 2.

Region 3 ($x_{satp} \le x \le 1-p$): Both the NMOS transistor and the PMOS transistor are in saturation. The PMOS current, i.e. the short-circuit current can be easily calculated by using Eq.(3.39) and Eq.(4.19).

$$i_{SC3} = S_{kp} \cdot K_{vp} \cdot (1 - x - p)^2 \tag{4.31}$$

The short-circuit energy dissipation in the case of rising input is defined as

$$E_{SCr} = V_{DD}\int_{x1}^{1-p} i_{SC}\tau_r dx = V_{DD}\tau_r \cdot \left(\int_{x1}^{xsatp} i_{SC2}dx + \int_{xsatp}^{1-p} i_{SC3}dx\right) \tag{4.32}$$

$$= V_{DD}\tau_r \cdot \left(\int_{x1}^{xsatp} S_{2p}(x - x_1)dx + \int_{xsatp}^{1-p} S_{kp} \cdot K_{vp} \cdot (1 - x - p)^2 dx\right)$$

$$= V_{DD}\tau_r \cdot \left(\frac{1}{2}S_{2p}(x_{satp} - x_1)^2 + \frac{1}{3}S_{kp} \cdot K_{vp} \cdot (1 - x_{satp} - p)^3\right)$$

4.2 Short-circuit Energy Dissipated in Falling Input Transition

In the case of falling input transition, assume that the input voltage waveform is a falling ramp with transition time τ_f:

$$V_{in}(t) = \begin{cases} V_{DD} & t \le 0 \\ V_{DD} \cdot (1 - t/\tau_f) & 0 \le t \le \tau_f \\ 0 & t \ge \tau_f \end{cases} \tag{4.33}$$

From above, the slope of the falling input voltage ramp can be determined:

$$S_f = V_{DD} / \tau_f \tag{4.34}$$

Then, when $0 \le t \le \tau_f$ (i.e., $0 \le x \le 1$), the normalized format of input voltage is give as:

$$u_{in} = 1 - x \tag{4.35}$$

where $x=t/\tau_f$.

The method to evaluate the short-circuit energy dissipation in the falling input transition is same as that in the rising input transition, and the analysis is symmetrical. The method and equations presented in Section 2.4.2, 3.3.2 and Section 4.2 for the case of a rising input inverter will correspond to a falling input inverter if the following interchanges are made:

Table 4.1 The interchanges between rising input inverters and falling input inverters.

Type	Interchanges		
TERMS	NMOS	⇔	PMOS
	rising	⇔	falling
	discharging	⇔	charging
	overshoot	⇔	undershoot
SUBSCRIPTS	n	⇔	p
	r	⇔	f
	DS	⇔	SD
	GS	⇔	SG
VARIABLES	x	⇔	1-x

4.3 Inverter Short-circuit Power Dissipation

After deriving the short-circuit energy dissipated in rising and falling input transitions, the average short-circuit power dissipation P_{SC} can be easily obtained, i.e.

$$P_{SC} = f \cdot (E_{SCr} + E_{SCf}) \quad (4.36)$$

where f is the inverter switching frequency. Obviously, the inverter short-circuit power dissipation is proportional to the switching frequency of inverter input signal.

Though the above procedure described in this chapter, the inverter short-circuit power dissipation may be estimated by using those inverter topological parameters and

technology parameters summarized in Table 4.2, where no extracted or fitting parameters are included. Table 4.3 contains a summary of developed equations that are needed to calculate the inverter short-circuit power dissipation.

Table 4.2 Summary of parameters used in the proposed model

Types	Parameters
Topological Parameters	C_{load}, L_{eff}, W_{eff} L_e, W_g, A_d, and P_d
Technology Parameters	V_{DD}, V_{th}, U_0, U_a, U_b, v_{sat}, C_{ox}, C_{gd}, and C_j

Table 4.3 Summary of equations used in the proposed model

Types	Equations
Short-circuit energy in one switching cycle (rising input)	Eq.(4.4), (4.7), (4.11)~(4.14), (4.16), (4.17), (4.25)~(4.32)
Short-circuit energy in one switching cycle (falling input)	Eq.(4.34), (4.35), and transform the above equations using Table 4.1.
Short-circuit power	Eq.(4.36)

Chapter 5
Model Validations and Comparisons

In the first section of this chapter, the method using the results of Spectre simulator with BSIM3v3.2 MOSFET model as "exact" values, instead of directly measured values, is illuminated. Then, the proposed inverter short-circuit power model, which is implemented in Maple, has been tested with a wide range of inverters in two different CMOS technologies. In the second section, the model accuracy is verified by using the UMC 1.2V 0.13μm CMOS technology. Various switching conditions of input transition time, capacitive load, and inverter ratio have been considered. The portability of the proposed model is verified by using the TSMC 1.8V 0.18μm CMOS technology in Section 5.3. To show the rapidness of the proposed model, the average CPU time is compared with Spectre simulator in the last section.

5.1 Illumination of Comparison Method

In this section, all validations and simulation comparisons are carried out by comparing the proposed model computed results with the "exact" values obtained from Spectre simulation using the BSIM3v3.2 MOSFET model. It should be noted that in this thesis, measured I-V characteristics are not used, because such data are not available. Instead, the data are obtained by simulating the analyzed circuit with the BSIM3V3.2 model. To illustrate the effectivity of this method, it is necessary to briefly review this MOSFET model.

Many commercial SPICE simulators such as HSPICE (Avant!), Spectre (Cadence), ELDO (Mentor Graphics), PSPICE (MicroSim) are widely used for circuit simulation. In order to use these simulators, device models are needed to describe the device behaviors in the circuits. In other words, device models are the link between the physical world (technology, manufacturing,...) and the design world (device level simulation, timing simulator model, macro model, synthesis,...) of the semiconductor industry. So far, over a hundred MOSFET models have been developed.

Because of its outstanding performance, BSIM3v3 was selected as the first MOSFET model for standardization by an independent Compact Model Council, consisting of many leading companies in the semiconductor industry [16]. It has been verified extensively by both model developers [46] and model users from many different companies. The performance comparison of some common models is shown in Table 5.1. In current commonly used simulators with modern 0.13μm and 0.18μm CMOS technologies, the BSIM3v3.2 is always used as the default model. So, it is reasonable to regard the simulation results using BSIM3v3.2 as "exact" data. Thus, the discrepancy between the BSIM3V3.2 model and the actual MOSFET characteristic is not included when assessing the performance of the proposed model.

In this chapter, the results of the proposed model are compared with those of two latest short-circuit power estimation methods [7][8], which were published in IEEE journals.

Table 5.1 Performance comparison of MOSFET models [16]

Model	Minimum L(um)	Minimum Tox(nm)	Model Continuity	Id Accuracy in Strong Inversion	Id Accuracy in Subthreshold	Small Signal parameter	Scalability
MOS1	5	50	POOR	POOR	N.A.	POOR	POOR
MOS2	2	25	POOR	POOR	POOR	POOR	FAIR
MOS3	1	20	POOR	FAIR	POOR	POOR	POOR
BSIM1	0.8	15	FAIR	GOOD	FAIR	POOR	FAIR
BSIM2	0.35	7.5	FAIR	GOOD	GOOD	FAIR	FAIR
BSIM3v2	0.25	5	FAIR	GOOD	GOOD	GOOD	GOOD
BSIM3v3	0.15	4	GOOD	GOOD	GOOD	GOOD	GOOD

5.2 Model Accuracy Validation

The accuracy of the proposed model is verified by using UMC 1.2V 0.13μm CMOS technology. At first, the verification is conducted using a randomly sized inverter for different input transition times and output loads. Then, inverters with large widths and a very large range of aspect ratios are verified as well. At last, the accuracy of the proposed model is compared with that of two latest short-circuit power estimate methods.

5.2.1 Accuracy Validation Using a Randomly Sized Inverter

In digital circuits, the length of the devices is always the minimum feature size of the technology, i.e. 0.13μm for the UMC 1.2V 0.13μm CMOS technology. To validate the accuracy of the proposed model, an inverter with Wn=0.26μm and Wp=0.63μm is randomly selected, and tested for different input transition times and output capacitive loads, as shown in Figure 5.1 and 5.3, respectively. What are shown in Figure 5.1 and 5.3 are inverter short-circuit energy dissipation in one switching cycle, i.e. $E_{SCr} + E_{SCf}$. In

Figure 5.2 and 5.4, the transient waveforms of input/output voltage and short-circuit current responding to Figure 5.1 and 5.3 are shown.

The maximum relative errors (Relative error = (simulation-model)/simulation) for Figure 5.1 and 5.3 are 11% and 15% respectively, showing that the proposed model is very accurate for different input transition times and output loading capacitances.

In Figure 5.3, the maximum relative error occurs when C_{load}=80fF. At this point, the transition time of output voltage is very long, as shown in Figure 5.4, indicating the load of the inverter is too large. In digital IC design, a bigger inverter must be used to drive larger load capacitance. In the next part, the large size inverters are tested to verify the proposed model.

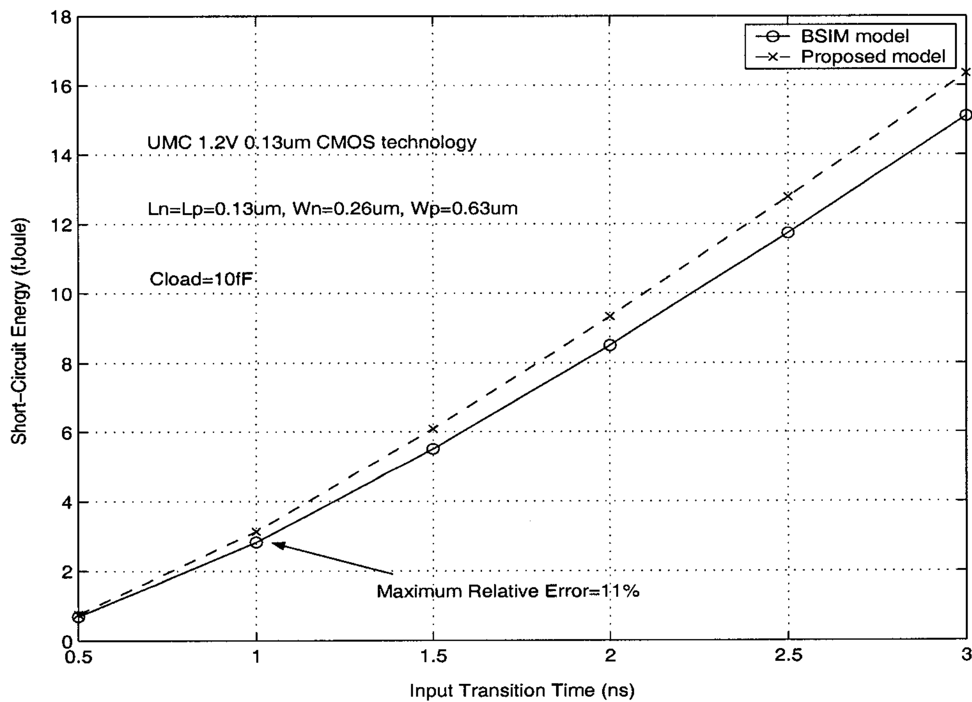
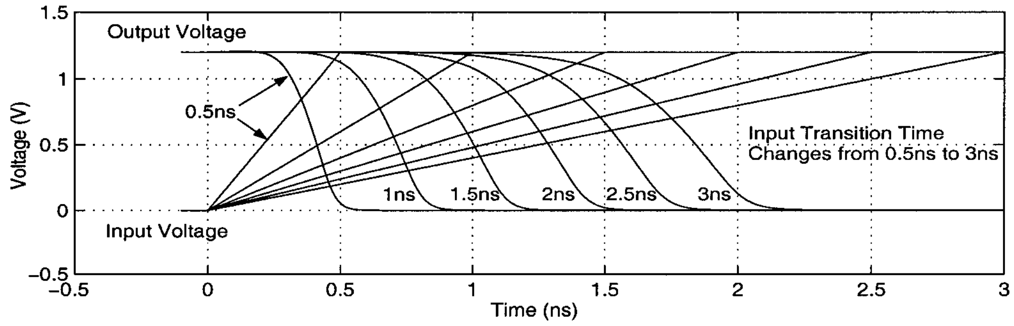
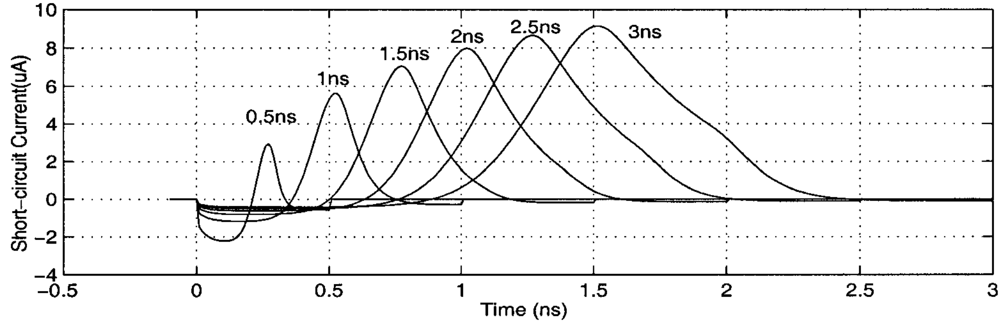


Figure 5.1 Inverter short-circuit energy dissipation in one switching cycle as a function of input transition time.



(a) Inverter transient input/output voltage waveforms in the cases of different input rising time.



(b) Inverter transient short-circuit current waveforms in the cases of different input rising time.

Figure 5.2 Transient waveforms of the inverter in Figure 5.1 in the case of rising input. (a) Inverter transient output/input voltage waveforms in the cases of different rising times; (b) Inverter transient short-circuit current waveforms in the cases of different rising time

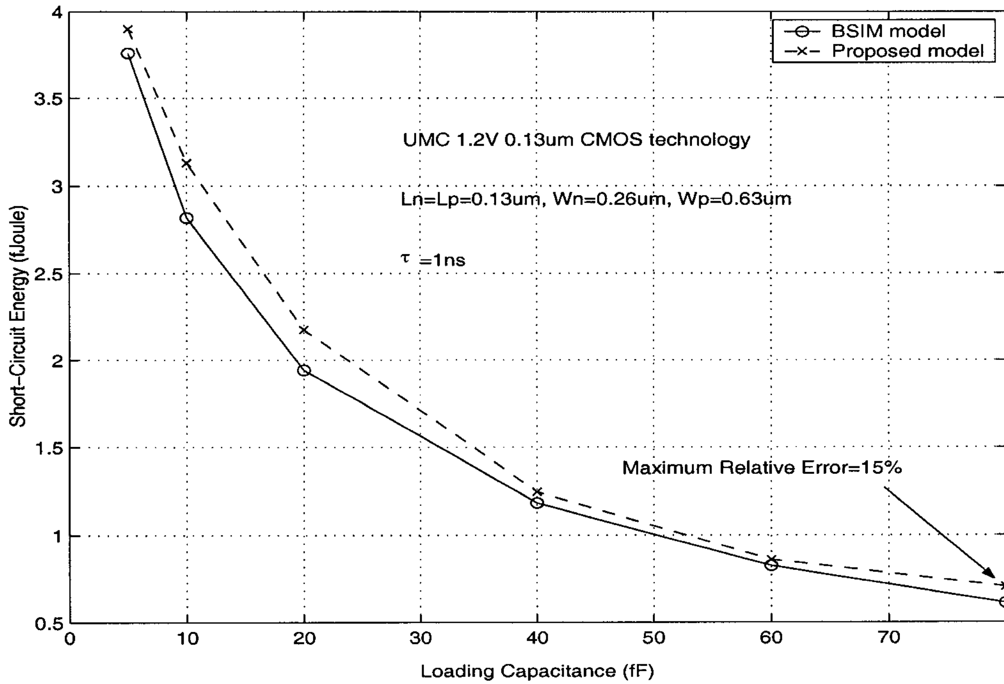
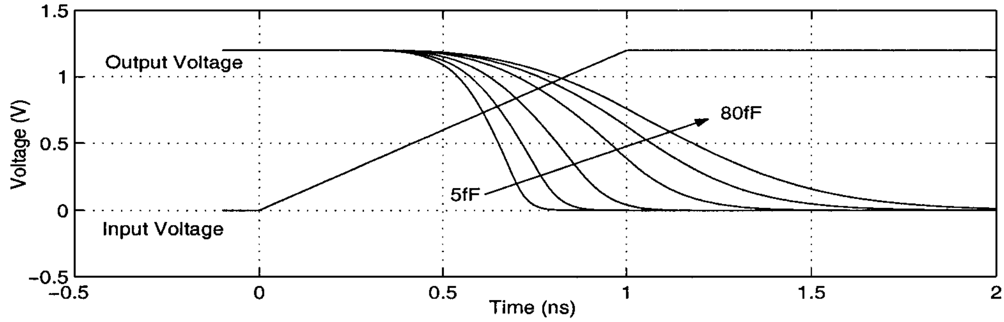
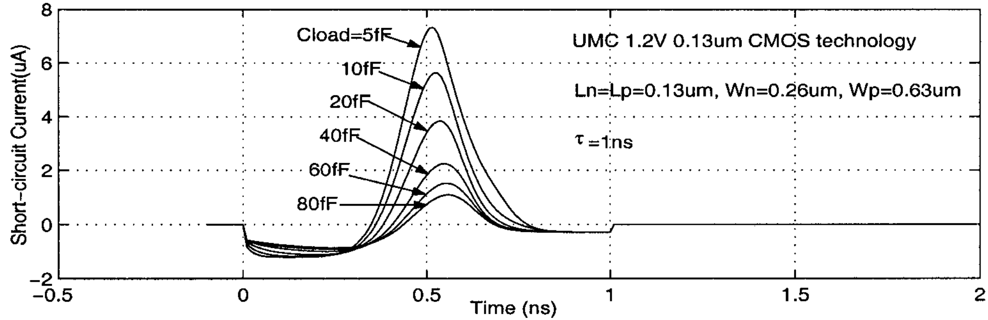


Figure 5.3 Inverter short-circuit energy dissipation in one switching cycle as a function of output loading capacitance



(a) Inverter transient input/output waveforms (rising input) for different loading capacitance



(b) Inverter transient short-circuit current waveforms (rising input) for different loading capacitance

Figure 5.4 Transient waveforms of the inverter in Figure 5.3 in the case of rising input for different loading capacitances. (a) Inverter transient input/output voltage waveforms. (b) Inverter transient short-circuit current waveforms.

5.2.2 Accuracy Validation Using Large Size Inverters

In Figure 5.5 and 5.6, two large size inverters are tested to validate the accuracy of the proposed model.

In Figure 5.5, the inverter with Wn=0.52μm and Wp=1.04μm can drive loading capacitance as large as 120fF, while the accuracy keeps below15%. In Figure 5.6, the inverter with Wn=1.04μm and Wp=2.08μm can drive loading capacitance as large as 240fF, while the maximum relative error is 20.3%. Results show the proposed model is accurate in the case of large size inverter driving big loading capacitance.

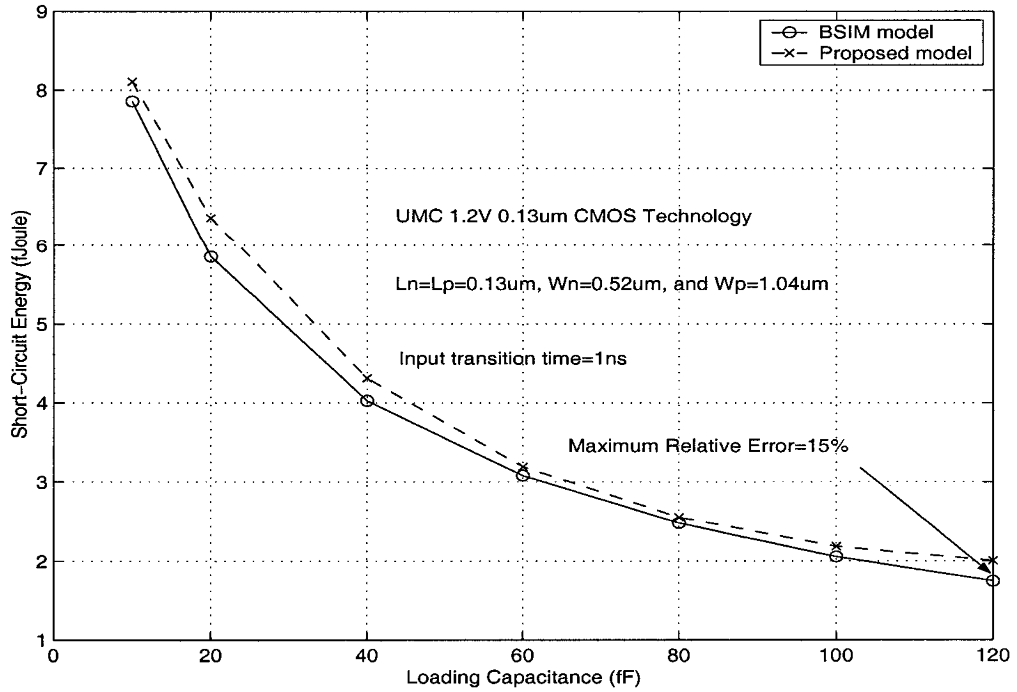


Figure 5.5 Large size inverter (Wn=0.52μm, and Wp=1.04μm) short-circuit energy dissipation in one switching cycle as a function of output loading capacitance

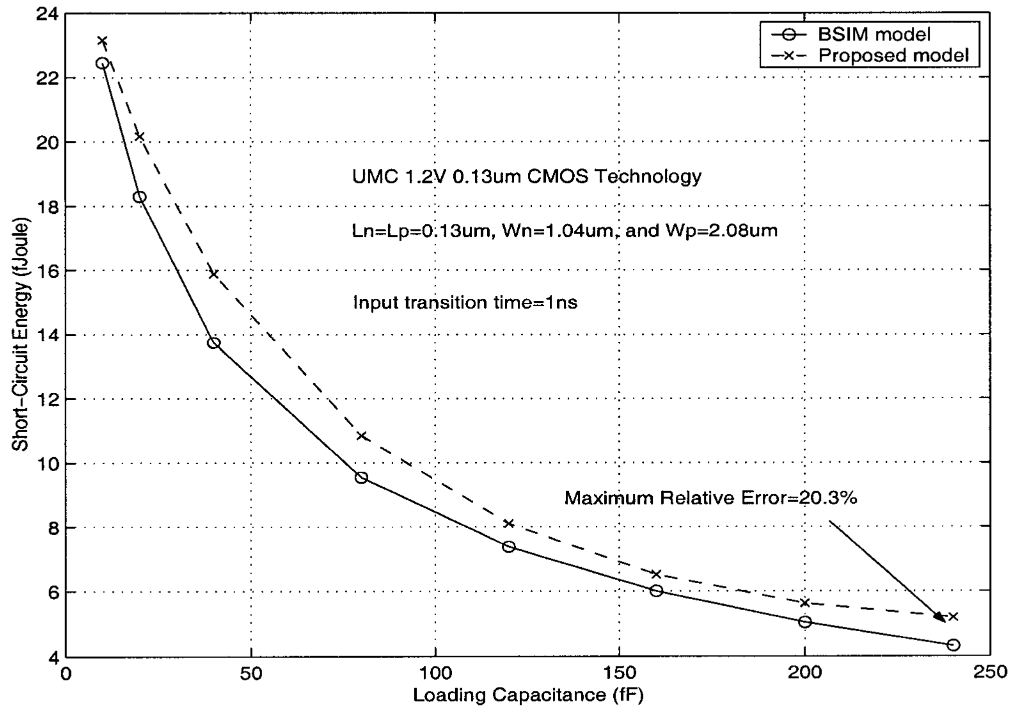


Figure 5.6 Large size inverter (Wn=1.04μm, and Wp=2.08μm) short-circuit energy dissipation in one switching cycle as a function of output loading capacitance

5.2.3 Accuracy Validation Using Inverters with Different Aspect Ratios

In Figure 5.7, the values of inverter short-circuit energy dissipation in one switching cycle as a function of input transition time for different inverter ratio (Wn=0.26μm, Wp/Wn = 0.5, 1, 2, and 4, i.e. Wp=0.15μm to 1.04μm) are compared. Note that in UMC 1.2V 0.13μm CMOS technology, the minimum width is 0.15μm, other than 0.13μm. Hence, when testing Wp/Wn=0.5, Wp virtually used is 0.15μm, not 0.13μm.

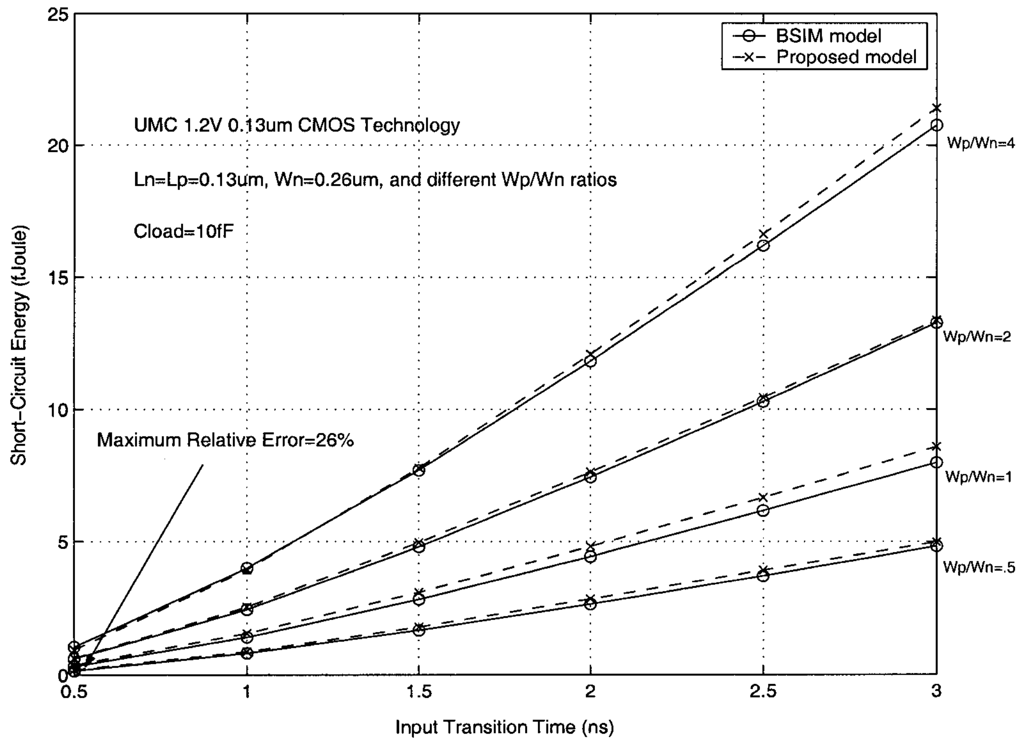


Figure 5.7 Inverter short-circuit energy in one switching cycle as a function of input transition time for different aspect ratio (Wp/Wn=.5, 1, 2, and 4).

The maximum relative error is 26% at the point τ=0.5ns, Wp/Wn=0.5. This occurs because the absolute value of short-circuit energy at this point is too small. The value computed using the proposed model is 0.185 fJoule, while the simulation result using BSIM3v3.2 is 0.147 fJoule. The absolute error between them is only 0.038 fJoule. Except for two points (τ=0.5ns, Wp/Wn=0.5 and τ=0.5ns, Wp/Wn=1), the errors at most points

are below 15%, showing good agreement between the model results and the "exact" values.

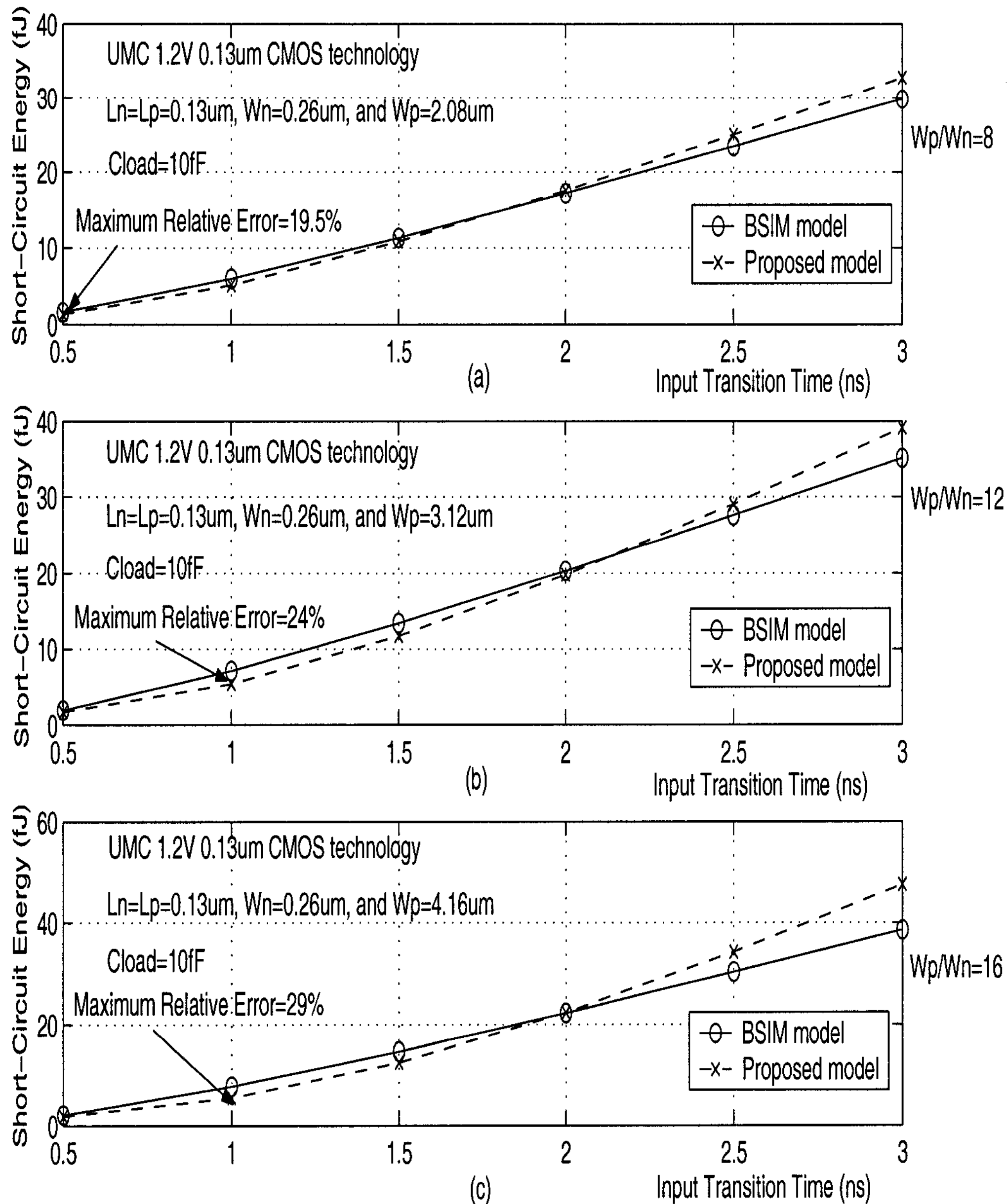


Figure 5.8 Inverter short-circuit energy in one switching cycle as a function of input transition time for different aspect ratio (Wp/Wn=8, 12, and 16).

As mentioned before, digital IC designers are mainly interested in designing circuits with transistors having widths ranging from minimum size up to 8 times the minimum width [33], i.e. 1.2μm for the UMC 0.13μm CMOS technology. To illustrate the robust

accuracy of the proposed model, inverters with a wider range of aspect ratios (Wp/Wn= 8, 12, and 16, i.e. Wp=2.08μm to 4.16μm) are tested, as shown in Figure 5.8. To avoid overlapping, the results are drawn in three subplots. The maximum relative errors responding to Wp/Wn=8, 12, and 16 (i.e. Wn=0.26μm, Wp=2.08μm, 3.12μm, and 4.16μm) are 19.5%, 24%, and 29%, respectively. Although the error increases with the aspect ratio as a whole, the accuracy of the proposed model is less than 20% for most cases in the interested range of device geometry for digital IC design.

5.2.4 Accuracy Comparisons with Latest Methods

Table 5.2 Short-circuit energy dissipation and total dynamic energy dissipation per switching event estimated using the method in [7]. Results computed using the proposed model in [7] are compared with ELDO simulation results for different CMOS inverters (with transistor widths in 0.8μm CMOS technology) under various conditions of input transition time τ and loading capacitance C_L. (Copied from [7]). Note that in this table, no error in short-circuit energy estimation is listed.

<table>
<tr><th rowspan="2">τ (ns)</th><th rowspan="2">Wn (μm)</th><th rowspan="2">Wp (μm)</th><th rowspan="2">C_L (fF)</th><th colspan="2">Short-circuit Energy E_{SC} (pJ)</th><th colspan="2">Total Dynamic Energy E (pJ)</th><th rowspan="2">E_{SC}/E (%)</th><th rowspan="2">Error in E (%)</th></tr>
<tr><th>Analytical Model in [7]</th><th>ELDO simulator</th><th>Analytical Model in [7]</th><th>ELDO simulator</th></tr>
<tr><td rowspan="8">2</td><td rowspan="4">6</td><td rowspan="2">6</td><td>100</td><td>0.951</td><td>1.29</td><td>4.45</td><td>4.51</td><td>28.6</td><td>1.3</td></tr>
<tr><td>250</td><td>0.837</td><td>0.900</td><td>8.07</td><td>7.88</td><td>11.4</td><td>2.4</td></tr>
<tr><td rowspan="2">12</td><td>100</td><td>1.65</td><td>2.10</td><td>5.63</td><td>5.67</td><td>37.0</td><td>0.70</td></tr>
<tr><td>250</td><td>1.50</td><td>1.55</td><td>9.22</td><td>8.87</td><td>17.4</td><td>3.9</td></tr>
<tr><td rowspan="4">12</td><td rowspan="2">12</td><td>200</td><td>1.94</td><td>2.65</td><td>8.878</td><td>9.079</td><td>29.1</td><td>2.2</td></tr>
<tr><td>350</td><td>1.88</td><td>2.17</td><td>12.56</td><td>12.35</td><td>17.5</td><td>1.7</td></tr>
<tr><td rowspan="2">36</td><td>200</td><td>3.32</td><td>4.29</td><td>11.24</td><td>11.40</td><td>37.6</td><td>1.4</td></tr>
<tr><td>350</td><td>3.26</td><td>3.63</td><td>14.92</td><td>14.49</td><td>25.0</td><td>2.9</td></tr>
<tr><td rowspan="2">4</td><td rowspan="2">6</td><td>6</td><td rowspan="2">250</td><td>1.99</td><td>2.74</td><td>9.23</td><td>9.72</td><td>28.1</td><td>5.0</td></tr>
<tr><td>12</td><td>3.42</td><td>4.52</td><td>11.2</td><td>11.80</td><td>38.3</td><td>5.0</td></tr>
</table>

In [7], an analytical model for estimation of inverter short-circuit power dissipation is proposed. The results of short-circuit energy dissipation and total dynamic energy dissipation per switching event are compared, as shown in Table 5.2, which is copied from [7]. In Table 5.2, only ten cases are compared, and only the errors in dynamic energy dissipation are listed. To illustrate the accuracy of the inverter short-circuit power model in [7], the relative errors of the short-circuit energy dissipation in Table 5.2 are calculated and listed in Table 5.3, where the first six columns are same as those in Table 5.2. In the ten cases compared in [7], the relative errors of six cases are between 20% and 30%.

Table 5.3 The relative errors of the inverter short-circuit energy estimation of the cases in Table 5.2.

τ (ns)	Wn (μm)	Wp (μm)	C_L (fF)	Short-circuit Energy E_{SC} (pJ)		Relative Error in E_{SC} =(simulator-model)/simulation (%)
				Analytical Model in [7]	ELDO simulator	
2	6	6	100	0.951	1.29	26.2
			250	0.837	0.900	7.0
		12	100	1.65	2.10	21.4
			250	1.50	1.55	3.2
	12	12	200	1.94	2.65	26.8
			350	1.88	2.17	13.4
		36	200	3.32	4.29	22.6
			350	3.26	3.63	10.2
4	6	6	250	1.99	2.74	27.4
		12		3.42	4.52	24.3

To compare their accuracy more directly, ten cases corresponding to Table 5.3 but using 0.13μm technology are calculated in Table 5.4, where inverters are in same W/L ratio as Table 5.3. For example, in Table 5.3, Wn/Ln=6/0.8=7.5, while in Table 5.4 Wn/

Ln=0.975/0.13=7.5. From 0.8μm to 0.13μm, the minimum feature size is reduced about 6 times. For easy calculation, the load capacitance value shrinks 5 times. The input transition time is 1ns and 2ns respectively.

Table 5.4 Inverter short-circuit energy dissipation results obtained by using the proposed model, in the cases corresponding to Table5.3.

τ (ns)	Wn (μm)	Wp (μm)	C_L (fF)	Short-circuit Energy E_{SC} (pJ)		Relative Error in E_{SC} =(simulator-model)/simulation (%)
				The Proposed Model	Spectre simulator	
1	0.975	0.975	20	10.197	9.329	9.3
			50	6.742	5.956	10.5
		1.95	20	17.323	16.559	4.6
			50	11.904	11.035	7.9
	1.95	1.95	40	23.239	20.923	11.1
			70	17.70	16.289	8.7
		5.85	40	61.716	48.672	27.0
			70	48.673	39.456	23.4
2	0.975	0.975	50	20.414	18.768	8.8
		1.95		35.825	33.444	7.1

Since BSIM3v3.1 model is used in [7], and my thesis use BSIM3v3.2 model, in fact, there is no big difference between these two versions of BSIM3v3 model [16]. By comparing Table 5.3 and Table 5.4, it can be found that the accuracy of the proposed model is better than that of the method [7], which requires complex parameter extraction and optimization.

In [8], another method for the evaluation of the inverter short-circuit energy dissipation is presented and extended to static CMOS gates. The accuracy of the inverter model is compared with previous methods by using only one inverter for different input transition

time. The comparison result is copied in Figure 5.9. Based on this comparison, a conclusion that the error in most cases is less than 15% is then given out. However, the accuracy is not validated using inverters with different sizes and for different loading capacitances. Even further, the comparison is based on SPICE simulation using level-3 MOSFET model, which is no longer accurate enough for modern DSM technologies [16][17].

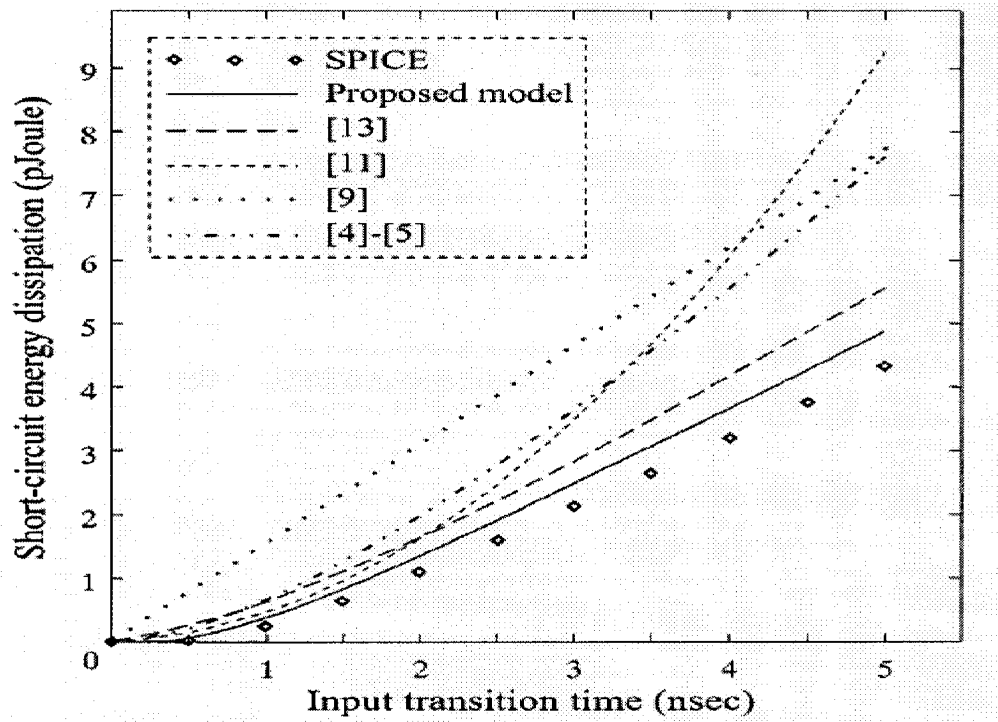


Figure 5.9 Comparison of inverter short-circuit energy in one switching cycle as a function of input transition time (Inverter size: Ln=Lp=0.8μm, Wn=4μm, Wp=6.55μm), using different method (Copied from [8]). Note that, in the legend of this figure, the reference number [13], [11], [9], [4] and [5] are corresponding to this thesis' reference [48], [24], [20], [22], and [23]

5.3 Model Portability Validation

To validate the portability, the TSMC 1.8V 0.18μm CMOS technology is used. (The TSMC 0.25μm CMOS technology was also considered. However, after I met a problem in

the course of simulation and sent a Bug Submission Form to CMC, I was noticed that CMC did not support the 0.25 micron CMOS design kit any longer. So I had to give up the use of TSMC 0.25μm CMOS technology.)

An inverter with Wn=0.4μm and Wp=0.8μm is randomly selected to verify the portability of the proposed model. With this technology, the inverter short-circuit energy in one switching cycle as a function of input transition time is compared in Figure 5.10, while that as a function of loading output capacitance compared in Figure 5.11. The maximum relative error in Figure 5.11 is 7%. In Figure 5.10, at most points, the relative errors are less than 15%. The maximum relative error is 19%. This occurs because of the same reason in Figure 5.7

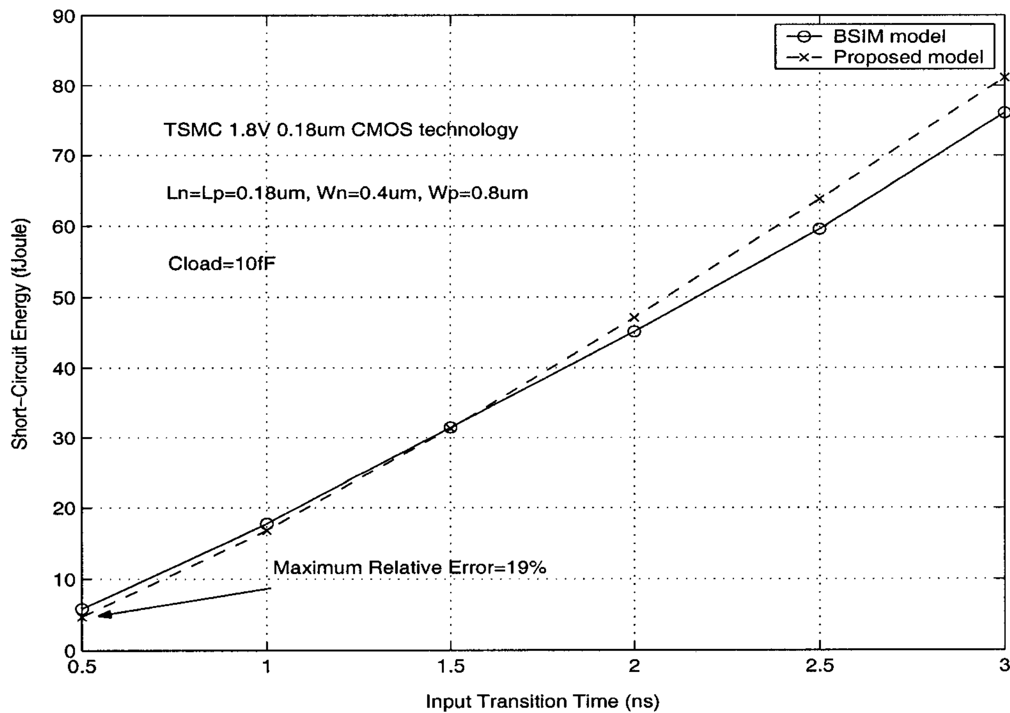


Figure 5.10 Inverter short-circuit energy in one switching cycle as a function of input transition time using 0.18μm CMOS technology.

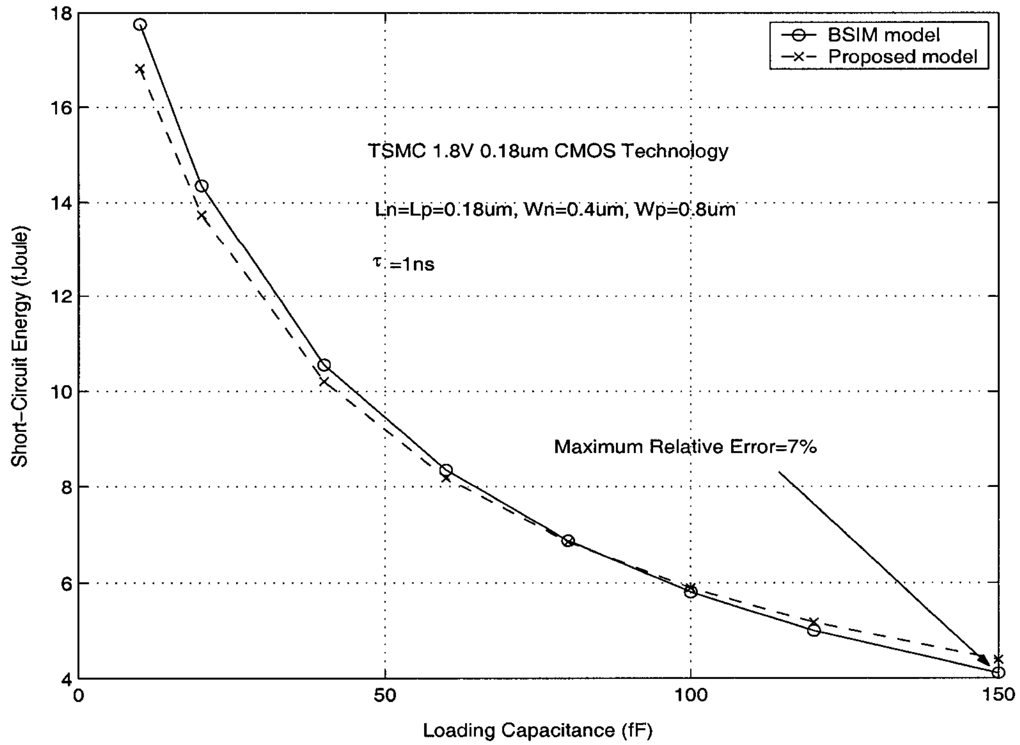


Figure 5.11 Inverter short-circuit energy in one switching cycle as a function of loading capacitance using 0.18μm CMOS technology.

5.4 CPU Time

The main advantage in using the proposed model rather than a circuit simulator, such as Spectre and HSPICE, is speed. For a given inverter configuration and input transition time, the CPU time required to simulate that circuit using Spectre or HSPICE is primarily dependent on two simulator settings -- the time step and the start/stop time (or the time duration of the transient analysis). To determine a reasonable estimate of the speedup achieved by the proposed model over Spectre, these two settings are selected as follows:

- the time step is set to 1ps, in order to ensure the smoothness of short-circuit current waveforms even when the input transition time is small, such as τ=0.5ns.

- the time duration of the transient analysis is firstly selected to be slightly larger than the input transition time in an initial simulation, in order to determine the values of t_1 and t_3, as shown in Figure 2.6. Then rerun the simulator with the start time and stop time settings equal to the values of t_1 and t_3, respectively.

Hence, for each case, to calculate the short-circuit energy dissipation in one switching cycle, the Spectre simulator must be run four times (two for rising input, another two for falling input transition), and the calculator tool in Cadence design kit must be used as well.

For simple comparison, the CPU time required in the initial simulation is chosen and is averaged. Many inverters of different sizes, load capacitances, and input transition times in two CMOS technologies are simulated. Results show that the proposed model, implemented in Maple, always offers about ten times improvements in average CPU time over Spectre simulation.

In [8], the speed of the method is not compared. In [7], the inverter model, run in MATLAB, offers about two orders of magnitude improvements in CPU time over HSPICE or ELDO.

Note that as a modern circuit simulator, the basic capabilities of the Spectre are similar in function and application to SPICE, but the Spectre is not descended from SPICE. The Spectre circuit simulator has many improvements over SPICE. As for speed, the Spectre is generally two to five times (sometimes over 10 times) faster than SPICE [47].

So, the speed of the proposed model can approach that of [7]. Both of them are much faster than other simulators, such as Spectre, HSPICE or ELDO, when calculating the inverter short-circuit power dissipation.

Chapter 6
Conclusion

In an attempt to satisfy the growing need for rapidly and accurately estimating the short-circuit power dissipation in modern digital ICs, a technology portable analytical model for the DSM CMOS inverter has been developed.

Compared with previous works, which always require extracted or fitting parameters, the proposed model depends only on the inverter dimension and SPICE parameters, which are usually provided with the given technology, resulting in the technology portability of this model. To achieve better accuracy, the effect of the gate-to-drain coupling capacitance and some main DSM effects such as velocity saturation and mobility degradation are taken into account.

The accuracy and portability validations of the proposed model have been performed for two DSM CMOS technologies and for a wide range of input transition times, output loading capacitances and aspect ratios. The results produced by the proposed model show good agreement with Spectre simulation using BSIM3v3.2 model in different DSM CMOS technologies, indicating its very good portability. Its accuracy is better than that of latest extraction required methods. However, extra extraction or fitting is no longer required, indicating the superiority of the model's technology portability. Based on a Maple implementation, the proposed model always offers much less average CPU time than Spectre simulator.

References

[1] Tadahiro Kuroda and Takayasu Sakurai, "Overview of low-Power ULSI circuit techniques", IEICE Trans. on Electronics, vol.E78-C, no.4, April 1995.

[2] G. K. Konstadinidis, K. Normoyle, Samson Wong, S. Bhutani, H. Stuimer, T. Johnson, A. Smith, D. Y. Cheung, F. Romano, Shifeng Yu, Sung-Hun Oh, V. Melamed, S. Narayanan, D. Bunsey, Cong Khieu, K. J. Wu, R. Schmitt, A. Dumlao, M. Sutera, Jade Chau, K. J. Lin, and W. S. Coates, "Implementation of a third-generation 1.1GHz 64-bit microprocessor," IEEE Journal of Solid-State Circuits, vol. 37, Nov. 2002, Pages:1461 - 1469.

[3] Neil H. E. Weste and Kamran Eshraghian, Principles of CMOS VLSI Design: A System Perspective, 2nd Edition, Addison-Wesley Publishing Company, 1993.

[4] R. F. Pierret, Semiconductor Device Fundamentals, Reading, MA: Addison-Wesley, 1996.

[5] H. J. M. Veendrick, "Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits," IEEE J. Solid-State Circuits, vol. SC-19, pp. 468-473, Aug. 1984.

[6] A. Kabbani, D. Al-Khalili, and A.J. Al-Khalili, "Technology-portable analytical model for DSM CMOS inverter transition-time estimation," IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems, vol. 22, pp. 1177-1187, Sept. 2003

[7] Anas A. Hamoui, and Nicholas C. Rumin, "An analytical model for current, delay, and power analysis of submicron CMOS logic circuits", IEEE Trans. on

Circuits and Systems --II: Analog and Digital Signal Processing, vol.47, no.10, October 2000.

[8] L. Bisdounis, and O. Koufopavlou, "Short-circuit energy dissipation modeling for submicronmeter CMOS gates," IEEE Trans. on Circuits and Systems -I: Fundamental, theory and Applications, vol. 47, no. 9, Sept. 2000, pp. 1350-1361.

[9] A. Nabavi-lishi and N. C. Rumin, "Inverter models of CMOS gates for supply current and delay evaluation," IEEE Trans. Computer-Aided Design, vol.13, pp.1271-1279, Oct. 1994.

[10] A. Chatzigeorgiou, S. Nikolaidis, and I. Tsoukalas, "A modeling technique for CMOS gates," IEEE Trans. Computer-Aided Design, vol.18, May 1999.

[11] J. T. Kong and D. Overhauser, Digital Timing Macromodeling for VLSI Design Verification, Kluwer Academic, 1995, pp. 125-169.

[12] L. Bisdounis and O. Koufopavlou, "Modeling the dynamic behavior of series-connected MOSFET's for delay analysis of multiple-input CMOS gates," Proc. IEEE Int. Symp. Circuits Syst., vol. 6, June 1998, pp. 342-345.

[13] P. E. Gronowski, W. J. Bowhill, R. P. Preston, M. K. Gowan, and R. L. Allmon, "High-performance microprocessor design" IEEE J. Solid-Static Circuits, vol.33, no.5, pp.676-686, May 1998.

[14] D. Liu and C. Svensson, "Power consumption estimation in CMOS VLSI chips." IEEE J. Solid-State Circuits, vol. 29, pp.663-670, June 1994.

[15] K. Toh, P. Ko and R. G. Meyer, "An engineering Model for short-channel MOS device," IEEE J. Solid-State Circuits, vol. 23, no. 4, Aug. 1988, pp. 950-958.

[16] Yuhua Cheng, and Chenming Hu, "MOSFET Modeling & BSIM3 User's Guide" Kluwer Academic Publishers, 1999.

[17] Star-Hspice Manual, Avant! Corporation, Release 1998.2, July 1998.

[18] D. A. Hodges and H. G. Jackson, Analysis and Design of Digital Integrated Circuits, 2nd Edition, New York: McGraw Hill, 1988.

[19] H. Shichman and D. A. Hodges, "Modeling and simulation of insulated-gate field-effect transistor switching circuits," IEEE J. Solid-State Circuits, vol. SC-3, pp. 285-289, Sept. 1968.

[20] T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," IEEE J. Solid-State Circuits, vol. 25, no. 2, pp. 584-594, Apr. 1990.

[21] T. Sakurai and A. R. Newton, "A simple MOSFET model for circuit analysis," IEEE Trans. Electron Devices, vol. 38, no. 4, pp. 887-894, Apr. 1991.

[22] N. Hedenstierna and K. O. Jeppson, "CMOS circuit speed and buffer optimization," IEEE Trans. Computer-Aided Design, vol. CAD-6, pp. 270-281, Mar. 1987.

[23] N. Hedenstierna and K. O. Jeppson, "Comments on 'A module generator for optimized CMOS buffers'," IEEE Trans. Computer-Aided Design, vol. 12, pp. 180-181, Jan. 1993.

[24] S. R. Vemuru and N. Scheinberg, "Short-circuit power dissipation estimation for CMOS logic gates," IEEE Trans. Circuits Syst. I, vol. 41, no. 11, pp.762-765, Nov. 1994.

[25] S. R. Vemuru and N. Scheinberg, "Correction to 'Short-circuit power dissipation estimation for CMOS logic gates'," IEEE Trans. Circuits Syst. I, vol. 42, no. 4, pp.234, Apr. 1995.

[26] S. Turgis, N. Azemard, and D. Auvergne, "Explicit evaluation of short-circuit power dissipation for CMOS logic structures," in Proc. Int. Symp. on Low-Power Design, Apr. 1995, pp. 129-134.

[27] A. Hirata, H. Onodera, and K. Tamaru, "Estimation of short-circuit power dissipation and its influence on propagation delay for static CMOS gates," in Proc. IEEE Int. Symp. on Circuits and Systems, May 1996, vol. 4, pp.751-754.

[28] Q. Wang and S. B. K. Vrudhula, "On short circuit power estimation of CMOS inverters," in Proc. IEEE Int. Conf. Computer Design, Oct. 1998, pp. 62-67.

[29] L. Bisdounis, O. Koufopavlou, and S. Nikolaidis, "Accurate evaluation of the CMOS short-circuit power dissipation for short-channel devices," in Proc. IEEE Int. Symp. Low-Power Electron. Design, Aug. 1996, pp. 189-192.

[30] L. Bisdounis, S. Nikolaidis, and O. Koufopavlou, "Propagation delay and short-circuit power dissipation modeling of the CMOS inverter," IEEE Trans. on Circuits and Systems -I: Fundamental, theory and Applications, vol. 45, no. 3, pp. 259-270, Mar. 1998.

[31] Y. P. Tsividis, Operation and Modeling of the MOS Transistor, New York: McGraw-Hill, 1988, pp.123-130.

[32] Makram M. Mansour, and N.R.Shanbhag, "Simplified current and delay models for deep submicron CMOS digital circuits," ISCAS '02, IEEE International

Symposium on Circuits and Systems, vol.5, pp.109-112, 2002.

[33] Makram M. Mansour, Mohammad M. Mansour, and Amit Mehrotra, "Modified Sakurai-Newton current model and its applications to CMOS digital circuit design," Proceedings. IEEE Computer Society Annual Symposium on VLSI, pp.62-69, Feb. 2003.

[34] D. Liu and C. Svensson, "Trading speed for low power by choice of supply and threshold voltages," IEEE J. Solid-State Circuits, vol. 28, no. 1, pp. 10-17, Jan. 1993.

[35] A. Chandrakasan, I. Yang, C. Vieri, and D. Antoniadis, "Design considerations and tools for low-voltage digital system design," Design Automation Conference Proceedings 1996, 33rd, 3-7 June 1996, Pages:113 - 118.

[36] K.O. Jeppson, "Modeling the influence of the transistor gain ratio and the input-to-output coupling capacitance on the CMOS inverter delay," IEEE Journal of Solid-State Circuits, Volume: 29, Issue: 6, June 1994, Pages:646 - 654.

[37] BSIM3v3.2.2 MOSFET Model Users' Manual, University of California, Berkeley. http://www-device.eecs.berkeley.edu/~bsim3

[38] Jan M. Rabaey, Digital Integrated Circuits: A Design Perspective, Prentice Hall, 1996.

[39] G. Massobrio and P. Antognetti, Semiconductor Device Modeling with SPICE, 2nd Edition, McGraw-Hill Inc., 1993.

[40] D. A. Hodges and H. G. Jackson, Analysis and Design of Digital Integrated Circuits, 2nd Edition, New York: McGraw Hill, 1988.

[41] M. Shur, Physics of Semiconductor Devices, Englewood Cliffs, Prentice-Hall, 1990.

[42] M. S. Liang, J. Y. Choi, P. K. Ko, and C. Hu, "Inversion-Layer Capacitance and Mobility of Very Thin Gate-Oxide MOSFETs," IEEE Trans. Electron Devices, ED-33,409, 1986.

[43] R. S. Muller and T. I. Kamins, Device Electronics for Integrated Circuits, 2nd edition, New York: John Wiley, 1986.

[44] C. C. Lo, G. P. Li and J. H. Mulligan, "An approximation to the factor K in the Toh-Ko-Meyer MOS engineering model," IEEE J. Solid-State Circuits, vol. 29, no. 1, Jan. 1994, pp. 77-78.

[45] E.S. Cheb-Terrab, Computer Handbook of ODEs, http://lie.uwaterloo.ca/help/abel2a1.html.

[46] Yuhua Cheng, Min-Chie Jeng, Zhihong Liu, Kai Chen, Mansun Chan, Chenming Hu, and Ping Keung Ko, "An investigation on the robustness, accuracy and simulation performance of a physics-based deep-submicronmeter BSIM model for analog/digital circuit simulation," IEEE 1996 Custom Integrated Circuits Conference, 1996.

[47] Affirma Spectre Circuit Simulator User Guide, Cadence Design Systems Inc., June 2000.

[48] S. Turgis and D. Auvergne, "A novel macromodel for power estimation in CMOS structures," IEEE Trnas. Comput.-Aided Design, vol. 17., pp1090-1098