

**EFFICIENT NANOSCALE VLSI STANDARD
CELL LIBRARY CHARACTERIZATION
USING A NOVEL DELAY MODEL**

A DISSERTATION

*Submitted in partial fulfillment of the
requirements for the award of the degree*

of

MASTER OF TECHNOLOGY

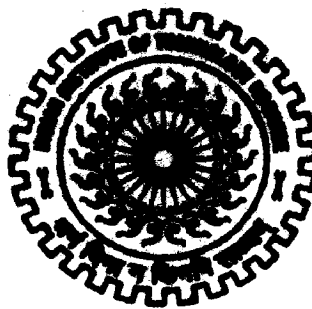
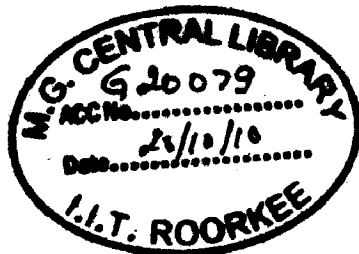
in

ELECTRONICS AND COMMUNICATION ENGINEERING

(With Specialization in Semiconductor Devices & VLSI Technology)

By

SANDEEP MIRYALA



**DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE -247 667 (INDIA)
JUNE, 2010**

CANDIDATE'S DECLARATION

I hereby declare that the work, which is being presented in the dissertation entitled, "***Efficient Nanoscale VLSI Standard Cell Library Characterization Using a Novel Delay Model***", which is submitted in the partial fulfillment of the requirements for the award of degree of ***Master of Technology in Semiconductor Devices & VLSI Technology***, submitted in the Department of Electronics and Computer Engineering, Indian Institute of Technology Roorkee (India), is an authentic record of my own work carried out under the guidance of Dr. Anand Bulusu, Assistant Professor, Department of Electronics & Computer Engineering, Indian Institute of Technology Roorkee.

The matter embodied in the dissertation report to the best of our knowledge has not been submitted for the award of any other degree elsewhere.

Dated : 7th JUNE 2010

Place : Roorkee

Sandeep
(Sandeep Miryala)

CERTIFICATE

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

B. Anand

Dr. Anand Bulusu
Assistant Professor

ACKNOWLEDGEMENTS

It gives me great pleasure to take this opportunity to thank and express my deep sense of gratitude to my guide Dr. Anand Bulusu, Assistant Professor, Department of Electronic and Computer Engineering, Indian Institute of Technology Roorkee, for sharing his industry experience and valuable guidance. It is under my professors valuable tutorship that i have learnt different aspects of research study.

I express my sincere thanks to Dr.A.K. Saxena, Professor, Dr. Sudeb Dasgupta, Assistant Professor and Dr. Sanjeev Manhas, Assistant Professor, Department of Electronic and Computer Engineering for their kind help and moral support throughout my dissertation work. Their suggestions helped me analyze my research work thoroughly.

I would like to thank research scholars of Semiconductor Devices and VLSI Technology group and the VLSI Lab staff for their help in learning HSPICE tool and installation of various tools for carrying out my dissertation work.

My special sincere heartfelt gratitude to my parents and brothers, Ram Prasad, Shyam Kumar and Varun Kumar whose best wishes, support and encouragement has been a constant source of strength to me during my work.

A special thanks to my friends Sandeep Talla, Deepthi, Satish, Abhilash, Kapila and Ashok Nandam. With out mentioning them this work would be incomplete. They have been with me in my good and bad times. I would also like to thank Uday Shanker Mudigonda, for his constant encouragement, motivating me to take up this area of work and sharing his industry experience.

Finally, I would like to thank all my batch-mates, my seniors, friends at Govind Bhawan of IIT Roorkee, whose constant presence made my work enjoyable.

(Sandeep Miryala)

ABSTRACT

Accurate estimation of delays in Static Timing Analysis (STA) using Non Linear Delay Model (NLDM) based Look Up Table (LUT) is a major challenge in nanometer range VLSI circuits. There are serious issues with NLDM based LUT due to the present method of arbitrary choice of input signal transition time t_{rin} and load capacitance (C_l) and resulting large number of HSPICE simulations with ad-hoc method adopted for achieving tolerable accuracy. In this dissertation, we present a systematic method to reduce standard cell library characterization time significantly achieving accuracy in a more systematic way. For this purpose we propose and use a simple and physics based logic gate delay model in which delay varies linearly with C_l and t_{rin} , where C_l is the load capacitance and t_{rin} is the input signal transition time of a standard cell. We also determine its region of validity in the (C_l, t_{rin}) space. We express the delay model coefficients and model's region of validity as a function of inverter (or logic gate) size. We validate our model for all basic gates such as inverter and NAND using HSPICE. We extend our work to multi-stage standard cells too. We do not use device current/capacitance models in our work and hence the method is general enough to be valid with scaling. With the help of this new model, we were able to save approximately of 60% SPICE simulations during the standard cell library characterization. We observe that the delay obtained using our LUTs is as accurate as that of the delay obtained through traditional LUTs with the said saving in simulation time.

Contents

Candidate's Declaration	ii
Certificate	ii
Acknowledgements	iii
Abstract	v
Table of Contents	vi
List of Figures	viii
List of Tables	x
List of Symbols	xi
1 Introduction	1
1.1 Introduction	1
1.2 Static Timing Analysis	2
1.3 Clocking Disciplines: Edge-Triggered Circuits	3
1.4 Delay Models	5
1.4.1 Alpha power law Delay Model	5
1.4.2 Polynomial DMs	8
1.4.3 Look Up Table Approach	8
1.5 Problem Statement	9
2 Novel Linear Delay Model and Its Region of Validity	11
2.1 Overview	11
2.2 Novel Linear Delay Model	11
2.3 Region of validity of linear delay model	14
2.4 Verification of linear delay model	15
3 Model for Output Signal Transition Time (T_{rout})	18
3.1 Overview	18
3.2 T_{rout} Model	18
3.3 Verification of T_{rout} model using HSPICE	21

4	Extension of the Models to Complex Gates	25
4.1	Overview	25
4.2	Linear Delay Model for NAND gate	25
4.3	T_{out} for NAND gate	28
4.4	Analysis for Multistage cells (Buffer)	29
5	Impact of technology scaling on the delay model	34
5.1	Overview	34
5.2	Delay Model Verification at 32nm Technology node	34
5.3	T_{out} model independent of technology node	36
6	Results and Conclusion	38
6.1	Overview	38
6.2	LUT using delay model	38
6.3	Reduction in number of simulations	39
6.4	Accuracy of LUT generated using the model	40
6.5	Delay comparison for a cascaded chain of inverters	40
6.6	Conclusion and Future work	41
	Bibliography	43

List of Figures

1.1	An example illustrating the application of CPM on a circuit with inverting gates.	2
1.2	Illustration of the clocking parameters of the flip-flop	4
1.3	Discharging waveform of an inverter	6
2.1	(a) CMOS Inverter, W_p and W_n are the widths of PMOS and NMOS respectively, (b) O/P of an inverter for ramp input.	12
2.2	(a) Delay variation with input signal transition time, (b) Delay variation with load capacitance. In both figures Points are simulated data and dotted lines are fitting of Delay Model.	15
2.3	(a) K_2 in Equation 2.11 with variation of W_n , (b) $K_1C_l + K_3$ in Equation 2.11 with W_n , (c) t_{rb} variation with load capacitance for various inverter sizes, (d) S_{trb} variation with W_n	16
2.4	Delay with realistic input versus delay with ramp input	16
3.1	The output transition of an inverter in Case 1(Not drawn to scale) . .	19
3.2	Inverter output transition for estimating T_{rout} in Case 2 (Not drawn to scale)	20
3.3	(a) Variation of T_{rout} with C_l before T_{rb} for different NMOS widths, (b) Variation of slope of T_{rout}	22
3.4	(a)Variation of T'_{rb} with Load Capacitance(C_l), (b)Variation of T_{rout} with t_{rin} for $t_{rin} \leq t_{rb}$	22
3.5	Variation of T_{rout} with load capacitance (C_l).	23
3.6	Variation of T_{rout} coefficients with device width (W_n).	24
4.1	A CMOS NAND gate circuit diagram.	26
4.2	(a) Input's B= V_{DD} and A=0 to V_{DD} , (b) Input's A= V_{DD} and B=0 to V_{DD}	26

4.3	(a) Comparison of slope in delay model of NAND to Inverter, (b) Comparison of intercept in delay model of NAND to Inverter.	27
4.4	Variation of t_{rb} with load capacitance for several device widths.	27
4.5	Slope of t_{rb} with respect to that of an inverter.	28
4.6	Variation of model coefficients with the load capacitance for NAND gate with input 'A' variable	29
4.7	Variation of model coefficients with the gate size for NAND gate with input 'A' variable	30
4.8	Variation model coefficients with the load capacitance for NAND gate with input 'B' variable	31
4.9	Variation model coefficients with the gate size for NAND gate with input 'B' variable	32
4.10	(a) Delay of 1 st stage, (b) Output transition of 1 st stage, (c) Delay of 2 nd stage, (d) Total delay of two stage buffer	33
5.1	(a) Delay variation with t_{rin} , (b) Delay variation with C_l . Points are simulated data and lines are fitting of our delay model.	35
5.2	Variation of coefficient K_2 in our delay model with device width.	35
5.3	Variation of slope of t_{rb} with device width.	36
5.4	Variation of T_{rout} model coefficients with load capacitance.	37
5.5	Variation of T_{rout} model coefficients with device width.	37
6.1	Comparison of delay for inverters of various size, load and t_{rin}	41

List of Tables

1.1	Delay LUT of a logic gate	9
1.2	T_{rout} LUT of a logic gate	9
6.1	Number of savings in HSPICE simulation while characterizing in- verter library using our DM.	39
6.2	Simulation savings in NAND gate with input 'A' variable.	39
6.3	Simulation savings in NAND gate with input 'B' variable.	40
6.4	Simulation savings in buffers.	40
6.5	Delay optimized LUT, Delay traditional and Delay HSPICE for 3,4, 5 and 11 stage inverter.	42

List of Symbols

W_n	Width of NMOS device
C_l	Load Capacitance
C_p	Parasitic Capacitance
μ	Mobility of electrons
C_{ox}	Gate Oxide Capacitance
I_{ds}	Drain current of MOSFET
V_{GS}	Gate to Source voltage of MOSFET
V_{DS}	Drain to Source voltage of MOSFET
t_{rin}	20% to 80% transition of input signal
T_R	Time taken by input ramp (linear) from 0 to V_{DD}
T_{rout}	Output signal transition time
L	Channel Length of MOSFET
V_{TH}	Threshold voltage of MOSFET
V_{DD}	Power Supply voltage
t_{rb}	For a given load, the maximum value of T_R for which our delay model is valid
C_{lb}	For a given T_R , the minimum value of C_l from which our delay model is valid
V_{out}	Output voltage
V_{in}	Input voltage

Chapter 1

Introduction

1.1 Introduction

The speed of an integrated circuit is characterized by its clock frequency. The setup time, hold time and the period of the clock impose delay constraints on a combinational data path. The setup time determines the maximum delay of the combinational data path called as setup time constraint, similarly the hold time of flip-flop imposes minimum delay constraint on the combinational data path called as hold time constraint. The delay of a combinational data must be estimated accurately to confirm that the constraints imposed by clock period, setup time and hold time are satisfied. The delay of the data path can be measured by two ways [1]:

1. Circuit simulations using SPICE can be used to estimate the delay of a circuit accurately. However, SPICE simulations need large cpu times to process an entire circuit having large number of transistors. SPICE takes few seconds to process individual transistors in a circuit, so the processing of an entire circuit takes large time.
2. An alternative method to using SPICE which estimates circuit delays reasonably accurately is *static timing analysis method (STA)*. STA makes use of the simple gate delay models to find the delay of the entire data path, hence takes lesser time.

1.2 Static Timing Analysis

To determine if a circuit meets timing constraints, it is necessary to find its critical path. Critical path is the path having maximum delay in traversing from primary inputs (PIs) to primary outputs (POs). Consider the combinational circuit shown in the Fig. 1.1. In STA the critical path is mostly found by a method which we elucidate here. Each block in the figure could be a simple logic gate or combinational block,

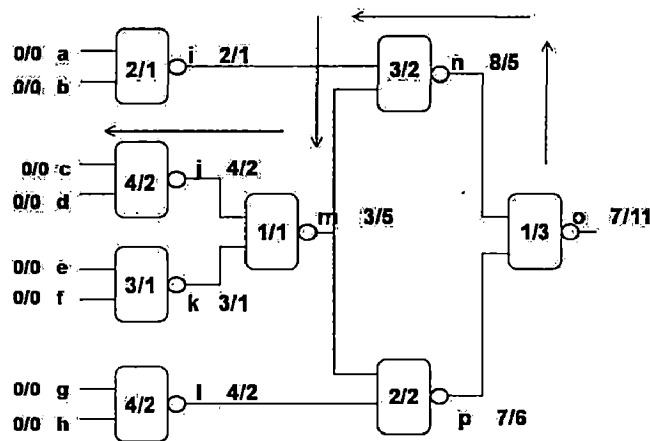


Figure 1.1: An example illustrating the application of CPM on a circuit with inverting gates.

and is characterized by delay from each input pin to output pin. Each block is an inverting type logic gate such as NAND or NOR. The numbers d_r/d_f inside each block represents the delays for output rise transition and output fall transition cases, respectively. These delay's are obtained through delay models. We also assume that all the primary inputs arrive at the time zero, so that the numbers "0/0" at each primary input represent the worst case rise and fall arrival times, respectively at each of these nodes.

A block is said to be ready for processing when the signal arrival time information is available for all of its inputs. Therefore initially, only those blocks that are fed solely by primary inputs are ready for processing. In the example these correspond to i,j,k and l. Then out of all the blocks that are ready for processing, choose any of the block. We compute the worst case arrival time at output by adding the delay of the block to latest arriving input time. In this way we process the remaining blocks and through out the entire circuit. In our example the processing of blocks

that are chosen are i,j,k,l,m,n,p,o and the worst case delay for the entire block is $\max(7,11)=11$ units.

To find the critical path in the above example, we begin with the final gate output 'o', whose falling transition corresponds to the maximum delay. This transition is caused by the *rising transition at the output of gate n*, which must therefore precede 'o' on the critical path. Similarly, the transition at 'n' is effected by the *falling transition at the output of 'm'*, and so on. By continuing this process, the critical path from the input to the output is identified as being caused by a *falling transition at either c or d*, and then progressing as follows: rising j \rightarrow falling m \rightarrow rising n \rightarrow falling o. The arrows in the Fig. 1.1 indicates the critical path.

Thus in STA, critical path is determined by CPM method, which in turn makes use of the delay's of each logic gate. These delay's of logic gates are obtained by making use of Delay Models (DM). The fast calculation of delay of a data path in STA is because of the delay models that it uses. However, for the circuit design to be flawless, these DMs should also be as accurate as possible.

In this chapter we will discuss several DMs proposed in literature. Before doing this, we discuss the setup time and hold time constraints.

1.3 Clocking Disciplines: Edge-Triggered Circuits

The basic parameters associated with a flip-flop can be summarized as follows:

- Setup Time: The Data input of the register, commonly referred to as the D input, must receive incoming data at a time that is at least T_{setup} units before the onset of the latching edge of the clock. The data will then be available at the output node Q, after the latching edge. The quantity, T_{setup} is referred to as the setup time of the FlipFlop.
- Hold Time: The input 'D' must be kept stable for a time of T_{hold} units, where T_{hold} is called the hold time, so that the data are allowed to be stored correctly in the FlipFlop.
- Each latch has a delay between the time the data and clock are both available

at the input, and the time when it is latched; this is referred to as the clock-to-Q delay, T_q .

Consider two flip-flops i and j , connected only by pure combinational paths. Over all such paths $i \rightarrow j$, let the largest delay from FF i to FF j be $\bar{d}(i,j)$, and the smallest delay be $\underline{d}(i,j)$.

Let us denote the setup time, hold time, and the maximum and minimum clock-to-Q delay of any arbitrary flip-flop be T_{Sk} , T_{hk} , Δ_k and δ_k respectively. Data is available at the launching flip-flop, i after the clock-to-Q delay and will arrive at the latching flip-flop J , at a time no later than $\Delta_i + \bar{d}(i,j)$. For correct clocking, the data are required to arrive one setup time before the latching edge of the clock at FF j as shown in Fig 1.2 i.e at a time no later than $P - T_{Sj}$. Where P is the period of the clock. This will give us a relation as

$$\Delta_i + \bar{d}(i,j) \leq P - T_{Sj} \quad (1.1)$$

$$\bar{d}(i,j) \leq P - T_{Sj} - \Delta_i \quad (1.2)$$

This constraint is often referred to as the setup time constraint. Since this requirement places an upper bound on the delay of a combinational path, it also called as long path constraint. The data must be stable for an interval that is at least as long

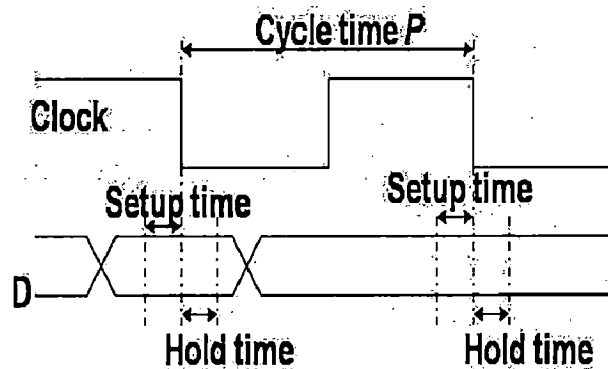


Figure 1.2: Illustration of the clocking parameters of the flip-flop

as the hold time after the clock edge, if it is to be correctly captured by the FF. Hence it is essential that the new data do not arrive at FF j before time T_{hj} . Since the earliest time that the incoming data can arrive is $\delta_i + \underline{d}(i,j)$, this gives us the

following hold time constraint:

$$\delta_i + \underline{d}(i, j) \geq T_{hj} \quad (1.3)$$

$$\underline{d}(i, j) \geq T_{hj} - \delta_i \quad (1.4)$$

Since this constraint puts a lower bound on the combinational delay on a path, it is referred to as a short path constraint. If this constraint is violated then the data in the current clock cycle are corrupted by the data from the next clock cycle. Thus the delay of the combinational data has to be measured by the designer to check whether the set-up time and hold time constraints are satisfied. The delay measurement is done using STA, which makes use of delay models. In the next section we discuss several DMs, their merits and demerits.

1.4 Delay Models

In order to find the delay of an entire combinational circuit using STA, we must determine the delay of its logic gates. These are obtained by using the models relating the delay, load capacitance and input transition time of a logic gate. These gate delay models are classified as:

1. Analytical Delay Models: The delay of a logic gate is found from the output voltage transition of logic gate across the load capacitor. They make use of the current equation of the MOSFET. The accuracy depends on accuracy of the current equation. Alpha power law delay model is a typical example.
2. Empirical Delay Models: Empirical models are based on the curve fitting on the simulation data obtained using SPICE. Polynomial delay model is a typical example.
3. Look Up Table (LUT) Method: Here we tabulate the gate delays for several values of input transition time (t_{rin}) and load capacitance (C_l). Having known t_{rin} and C_l we can pick the delay of that particular logic gate from the table.

We now discuss several important delay models proposed in the literature.

1.4.1 Alpha power law Delay Model

Alpha power Delay Model is a representative example of analytical DMs. This model is an extension of the Schokley's square-law MOS model in device saturation region.

The main advantage of the α -power delay model is that it takes into account of the velocity saturation effects which are dominant in the short-channel MOSFET's [2]. Using this model we can derive an equation for logic gate delay by taking into account of the input signal slope.

A full description of the model is given below:

$$I_D = 0 \quad (V_{GS} \leq V_{TH} : \text{cutoff region}) \quad (1.5)$$

$$= (I'_{D0}/V'_{D0})V_{DS} \quad (V_{DS} < V'_{D0} : \text{linear region}) \quad (1.6)$$

$$= I'_{D0} \quad (V_{DS} \geq V'_{D0} : \text{saturation region}) \quad (1.7)$$

where

$$I'_{D0} = I_{D0} \left(\frac{V_{gs} - V_{TH}}{V_{D0} - V_{TH}} \right)^\alpha \quad (1.8)$$

$$V'_{D0} = V_{D0} \left(\frac{V_{gs} - V_{TH}}{V_{D0} - V_{TH}} \right)^{\alpha/2} \quad (1.9)$$

The 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 $V_{GS} = V_{DS} = V_{DD}$). The time from a half- V_{DD} point of the input to a half- V_{DD} point of the output is defined as delay, t_{pHL} , in this discharging case. In the charging case the delay t_{pLH} is defined in the same way. Consider the Fig. 1.3 during the evaluation of the delay expression.

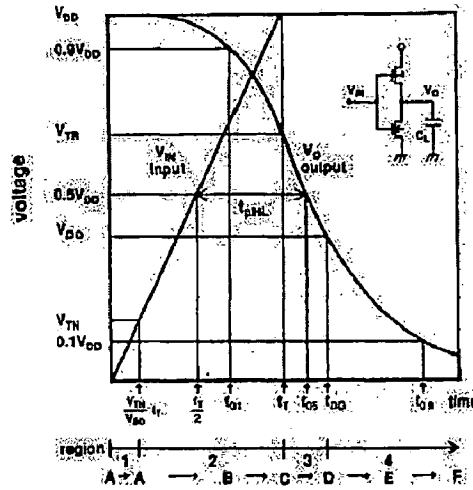


Figure 1.3: Discharging waveform of an inverter

Before the input reaches V_{TN} , NMOS is off and the output voltage V_o remains V_{DD} . (region1 in the above figure number). Then in the region 2, the input ramps

up linearly and the NMOS is operated in the saturation region. The output voltage V_o satisfies the following differential equation:

$$C_L \frac{dV_o}{dt} = -I'_{D0} = -I_{D0} \left(\frac{t/t_T - \nu_T}{1 - \nu_T} \right)^\alpha, \quad \nu_T = \frac{V_{TH}}{V_{DD}} \quad (1.10)$$

The solution is

$$V_o = V_{DD} - \frac{I_{D0} t_T}{C_L} \cdot \frac{1}{1 + \alpha} \cdot \frac{1}{(1 - \nu_T)^\alpha} \left(\frac{t}{t_T} - \nu_T \right)^{\alpha+1} \quad (\text{region2} : \nu_T t_T \leq t \leq t_T) \quad (1.11)$$

In region 3, the input is fixed at V_{DD} and the n-channel MOSFET is operated in the saturation region. Consequently, the output capacitance C_L is discharged by a constant current I_{D0} and the output voltage V_o changes linearly. The output equation in this region is given by

$$V_o = V_{DD} - \frac{I_{D0}}{C_L} \left(t - \frac{\nu_T + \alpha}{1 + \alpha} t_T \right) \quad (\text{region3} : t_T \leq t \leq t_T) \quad (1.12)$$

In the final region 4, the input is still fixed at V_{DD} but the operation mode of the n-MOSFET goes in to the linear region. As a result, the differential equation governing the discharging process can be written as

$$C_L \frac{dV_o}{dt} = -\frac{I_{D0}}{V_{D0}} V_o \equiv -\frac{1}{R3} V_o \quad (1.13)$$

The solution in this region has an exponential form and goes through the point (t_{D0}, V_{D0}) :

$$V_o = V_{D0} e^{\frac{-1}{C_L R3(t-t_{D0})}} \quad (\text{region4} : t_{D0} \leq t) \quad (1.14)$$

Denoting t_{05} as the time when the output reaches a half V_{DD} point, the delay t_{PHL} is calculated by using the previous derived expressions:

$$t_{PHL} = t_{05} - \frac{t_T}{2} = \left(\frac{\nu_T + \alpha}{1 + \alpha} - \frac{1}{2} \right) t_T + \frac{C_L V_{DD}}{2I_{D0}} \quad (1.15)$$

For t_{pLH} , the expression is exactly the same but the values of V_{TH} , α , and I_{D0} for the p-channel MOSFET should be used. Thus from the delay expression it can be seen that the delay is a linear combination of two terms. The first term is the input waveform dependent term, which is proportional to the input waveform transition time t_T , and the second term is the output capacitance term, which is proportional to the out capacitance C_L .

As we observe in α -power DM, the accuracy of analytical delay models depends on the accuracy of the MOSFET's current equation. MOSFET current equations for sub-90nm technologies are very complex in nature. In addition, their parameters are dependent on the terminal voltages of the device.

1.4.2 Polynomial DMs

Traditional methods for characterizing a cell driving a load use an equation of the form $K_1 C_l + k_2$, where k_1 is the characterized slope and k_2 is the intrinsic delay. However, such an equation neglects the effect of input transition on the delay.

The Non Linear Delay Model (NLDLM) from Synopsys uses the characterizing equations of the form $\alpha t_{rin} + \beta C_l + \gamma t_{rin} C_l + \delta$.

The Scalable Polynomial Delay Model (SPDM) developed by Synopsys uses a product of polynomials to fit the delay data. For example, for two parameters C_l and t_{rin} , a product of m^{th} order polynomial in C_l with an n^{th} order polynomial in t_{rin} , of the form $(a_0 + a_1 C_l + \dots + a_m C_l^m)(b_0 + b_1 t_{rin} + \dots + b_n t_{rin}^n)$ may be used [3].

To overcome this disadvantage industry is making use of LUTs to obtain delay of logic gate. In the next section we discuss details of look up table approach of finding the delay of any logic gate.

1.4.3 Look Up Table Approach

LUTs with delays tabulated for several values of input transition times and load capacitances are today used in STA. These LUTs are popular due to the limitations of analytical and polynomial DMs discussed in Sections 1.4.1 and 1.4.2.

Look up table used for STA is a two dimensional table, where a gate's delay is characterized with respect to its load capacitance (C_l) and the input signal transition time (t_{r-in}) [4].

A general look up table looks as shown in the Table 1.4.3, tabulates delay with respect to the load capacitance and input signal transition time. Since delay's are tabulated as a function of t_{rin} and C_l , we need to estimate signal transition times of all circuit nodes. A look table of gate output transition time (T_{rout}) must also be estimated as a function of load capacitance and input transition time.

	C_{load1}	C_{load2}	C_{load3}	C_{load4}
t_{r-in1}	D11	D12	D13	D14
t_{r-in2}	D21	D22	D23	D24
t_{r-in3}	D31	D32	D33	D34
t_{r-in4}	D41	D42	D43	D44
t_{r-in5}	D51	D52	D53	D54

Table 1.1: Delay LUT of a logic gate

	C_{load1}	C_{load2}	C_{load3}	C_{load4}
t_{r-in1}	$t_{r-out11}$	$t_{r-out12}$	$t_{r-out13}$	$t_{r-out14}$
t_{r-in2}	$t_{r-out21}$	$t_{r-out22}$	$t_{r-out23}$	$t_{r-out24}$
t_{r-in3}	$t_{r-out31}$	$t_{r-out32}$	$t_{r-out33}$	$t_{r-out34}$
t_{r-in4}	$t_{r-out41}$	$t_{r-out42}$	$t_{r-out43}$	$t_{r-out44}$
t_{r-in5}	$t_{r-out51}$	$t_{r-out52}$	$t_{r-out53}$	$t_{r-out54}$

Table 1.2: T_{rout} LUT of a logic gate

However, there are several problems associated with the LUTs. We enumerate these problems in the next section.

1.5 Problem Statement

1. Presently t_{rin} and C_l values are selected in an ad-hoc manner, so there needs a systematic way of choosing these values for improving accuracy.
2. These characterization tables have to be regenerated due to variation in process, voltage and temperature. The LUTs are characterized for several sizes of logic gates. This requires huge characterizing effort.
3. The accuracy of the delay obtained by LUT approach depends on the size of the LUT. Increasing the size increases accuracy but increases the system's memory consumption and simulation time.

In this thesis, we propose a solution to the above problems with existing LUTs which focuses on optimum choice of t_{rin} and C_l values. We propose a model in which delay varies linear with t_{rin} and C_l and also determine its region of validity as a function

of its logic gate size. We show that our approach is valid for any single stage logic cell. We also extend our work to multistage standard cells. Using this model and region of validity we choose appropriate values of t_{rin} and C_l for generating LUTs. Our approach can also be used with the technology scaling.

Chapter 2

Novel Linear Delay Model and Its Region of Validity

2.1 Overview

In this chapter we derive a linear delay model and its region of validity. Our model is physically based and its coefficients are obtained using physical arguments and very few SPICE simulations. While developing the model we make various assumptions and later justify the use of all our assumptions. In the next section we derive our linear delay model.

2.2 Novel Linear Delay Model

In this section, using physical reasoning we show that delay varies linearly with load capacitance (C_l) and input transition time (T_R) when these parameters are within a certain range which we determine later. We use the symbol T_R to denote the time required for the input voltage to increase from 0 to V_{DD} . In this derivation we use an inverter circuit with its output being discharged as shown in Fig. 2.1(b). In this chapter, the word “delay” stands for 50% delay. In our work, we assume that $V_{GS} = \frac{V_{DD}}{T_R}t$, where V_{GS} is the gate-source voltage of the inverter’s NMOS device, V_{DD} is the power supply voltage and t is time. We relax this assumption later. The output discharge comprises of two regions: First, when the input transitions from 0 to V_{DD} and second, when the input voltage $V_{in}=V_{DD}$. We make an assumption that the load capacitance is chosen such that the NMOS device is in saturation region

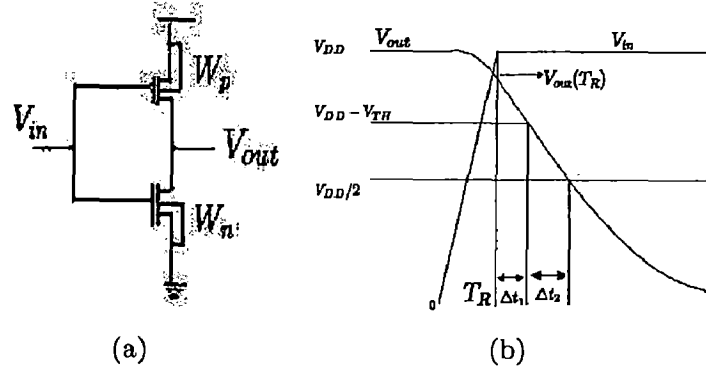


Figure 2.1: (a) CMOS Inverter, W_p and W_n are the widths of PMOS and NMOS respectively, (b) O/P of an inverter for ramp input.

from 0 to V_{DD} . The output discharge $\Delta Q(T_R)$ from 0 to T_R is,

$$\Delta Q(T_R) = \int_0^{T_R} I_{ds} dt \quad (2.1)$$

$$= T_R \int_0^1 f\left(\frac{V_{GS}}{V_{DD}}, \frac{V_{DS}}{V_{DD}} \cong 1\right) d\left(\frac{V_{GS}}{V_{DD}}\right) \quad (2.2)$$

$$= T_R \int_0^1 f(x, y = 1) dx \quad (2.3)$$

$$= S_T T_R \quad (2.4)$$

$$V_{out}(T_R) = V_{DD} - \frac{\text{Output discharge from } t = 0 \text{ to } t = T_R}{C_l + C_p} \quad (2.5)$$

$$V_{out}(T_R) = V_{DD} - \frac{S_T T_R}{C_l + C_p} \quad (2.6)$$

Here, $I_{ds} = f\left(\frac{V_{GS}}{V_{DD}}, \frac{V_{DS}}{V_{DD}}\right)$ is the NMOS drive current, S_T is a constant proportional to W_n since the NMOS current is proportional to the width of device ($I_{ds} \propto W_n$), $x = V_{GS}/V_{DD}$ and $y = V_{DS}/V_{DD}$. The generalized expression of current as a function of V_{GS} and V_{DS} enables us to include the second order effects into the expression. We assume that $y = V_{DS}/V_{DD} \cong 1$ for the NMOS device since it is operating in saturation regime. $V_{out}(T_R)$ is the output voltage at time $t = T_R$. We assume that PMOS device is very weak when compared to NMOS device due to rising transition at the input node. The output transition time can be further divided into two regions:

- When the NMOS device is in saturation, Δt_1 .

- When NMOS device operates in linear region, Δt_2 .

We need to find out the expressions for Δt_1 and Δt_2 to derive the complete delay expression.

We first assume that $V_{out}(T_R) \geq V_{DD} - V_{TH}$, where $V_{out}(T_R)$ is the output node voltage at time $t=T_R$. In other words, we assume that the device is in saturation from $t=0$ to $t=T_R$. From time $t=T_R$ to $T_R + \Delta t_1$, the device is in saturation. The output node is discharged from V_{DD} to $V_{DD}/2$ from time $t=0$ to $t=T_R + \Delta t_1 + \Delta t_2$. We now derive the values of Δt_1 and Δt_2 .

$$\Delta t_1 = \frac{\text{output discharge from } t = T_R \text{ to } t = T_R + \Delta t_1}{I_{ON}} \quad (2.7)$$

$$= \frac{(C_l + C_p)(V_{DD} - V_{TH}) - (C_l + C_p)V_{out}(T_R)}{I_{ON}} \quad (2.8)$$

$$\Delta t_1 = \frac{(C_l + C_p)(V_{TH}) - S_T T_R}{I_{ON}} \quad (2.9)$$

Here, C_p is the parasitic capacitance of the inverter seen at its output node and I_{ON} is the NMOS device ON current. We observe in Equation 2.9 coefficient of C_l is inversely proportional to W_n . Since C_p , S_T and I_{ON} are proportional to W_n , the remaining terms are independent of W_n ¹. We now consider the discharge from $V_{DD} - V_{TH}$ to $V_{DD}/2$ in Δt_2 . In this region the NMOS device operates in linear region and the output node can be assumed as an RC network. Hence, $\Delta t_2 \propto (C_l + C_p)$. Therefore, total delay is equal to

$$\text{Delay} = \frac{T_R}{2} + \Delta t_1 + \Delta t_2 \quad (2.10)$$

$$\text{Delay} = K_1 C_l + K_2 T_R + K_3 \quad (2.11)$$

where K_1 , K_2 and K_3 are constants which are extracted by fitting the model in the HSPICE simulation data. Further we make the following observations:

- Observation 1: K_1 and K_3 are linear functions of $1/W_n$.

¹Please note that $W_p=2W_n$ in inverter standard cells. Therefore, any change in W_n results in a proportional change in W_p

- Observation 2: K_2 is independent of W_n .

We validate these observations in Section 2.3 and use them in optimizing the delay LUT, as we have explained in Chapter 1. Here, we have made the assumption that the NMOS device is in saturation from 0 to T_R . This assumption imposes constraints on the region of validity of the model. We determine the region of validity of the model in the next section.

2.3 Region of validity of linear delay model

In this section, we determine the region of validity of our delay model. For this, we find the range of values of T_R and C_l in which the device operates in saturation regime. From Equation 2.5, we observe that,

$$V_{out}(T_R) = V_{DD} - \frac{S_T T_R}{C_l + C_p} \geq V_{DD} - V_{TH} \quad (2.12)$$

$$\Delta Q(T_R) = S_T T_R \leq (C_l + C_p) V_{TH} \quad (2.13)$$

For a given value of C_l , linear delay model of Equation 2.11 is valid for all the values of T_R which satisfy Equation 2.13. We denote the maximum value of T_R which satisfies Equation 2.13 as t_{rb} . From Equation 2.13, t_{rb} is a linear function of C_l .

$$t_{rb} = \frac{C_l V_{TH}}{S_T} + \frac{C_p V_{TH}}{S_T} \quad (2.14)$$

We extract the slope and intercept of this linear function by fitting in SPICE simulation data. We observe from Equation 2.14 that

- Observation 3: The slope of t_{rb} versus C_l plot is proportional to $1/W_n$.
- Observation 4: The intercept is a constant with W_n .

This is because $S_T \propto W_n$ and $C_p \propto W_n$. Using Equation 2.13 and a similar analysis one can derive the corresponding maximum value of C_l which Equation 2.13 holds. We denote this value of C_l as C_{lb} . In the next section, we discuss the verification of

linear delay model of Equation 2.11 and the region of validity expression given by Equation 2.13.

2.4 Verification of linear delay model

In this section we validate the results of Sections 2.2 and 2.3 using HSPICE simulations. We also extract coefficients of Equation 2.11 using the simulation data. We justify all the assumptions that we make in this chapter. We use 45nm PTM CMOS technology model files² in these simulations. We simulate inverters W_p and W_n adjusted such that the rise and fall transition times are equal.

Fig. 2.2(a) is a plot of simulated delay versus t_{rin} for several values of C_l . We use the symbol t_{rin} for 20% to 80% transition at input of logic gate. We show that Equation 4.1 fits well on data with an upper bound on t_{rin} , we also verify the validity of observations made Sections 2.2 and 2.3. In Fig. 2.3(a)- 2.3(b), our SPICE simulations confirm Observation 1 and Observation 2. Now we discuss this upper bound on t_{rin} i.e t_{rb} . The variation of $t_{rb}(C_l)$ with C_l is linear, as can be seen from the Figure 2.3(c).

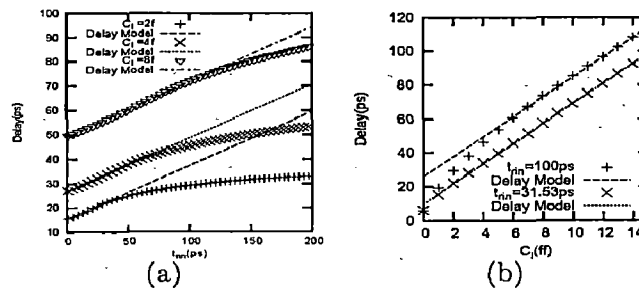


Figure 2.2: (a) Delay variation with input signal transition time, (b) Delay variation with load capacitance. In both figures Points are simulated data and dotted lines are fitting of Delay Model.

We show in Fig. 2.3 that the slope (S_{trb}) and intercept (C_{trb}) of this linear variation of delay with t_{rin} are independent and proportional to $1/W_n$, confirm with Equation 2.14 and Observations 3 and 4.

²Obtained from <http://www.eas.asu.edu/~ptm/>

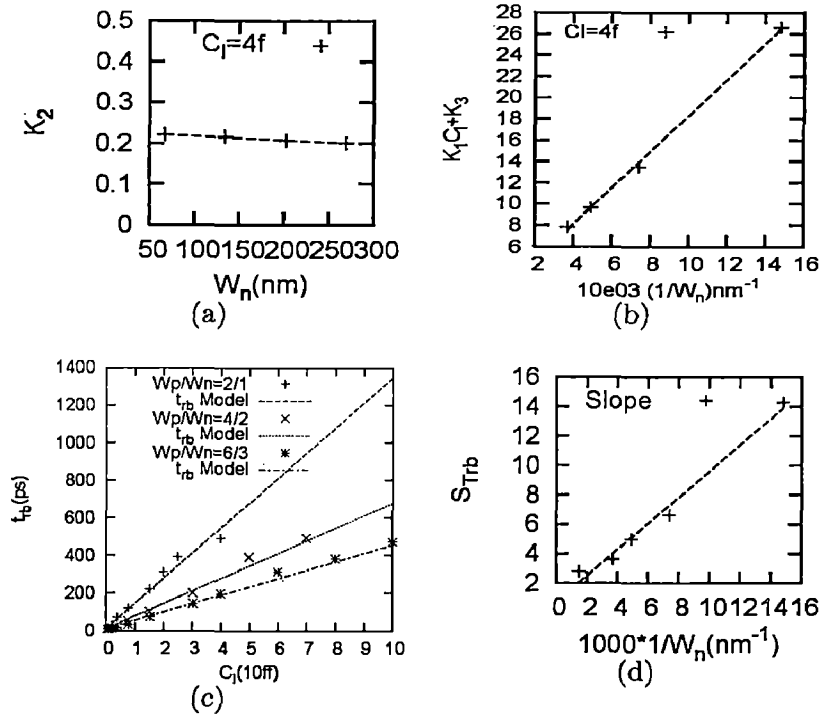


Figure 2.3: (a) K_2 in Equation 2.11 with variation of W_n , (b) $K_1 C_l + K_3$ in Equation 2.11 with W_n , (c) t_{rb} variation with load capacitance for various inverter sizes, (d) S_{trb} variation with W_n

Throughout this work we assume a linear (ramp) increase in input voltage from 0 to V_{DD} . We observe in Fig. 2.4 that delay with a ramp input and a more realistic input are linearly related. Therefore our delay model is valid for a realistic input signal.

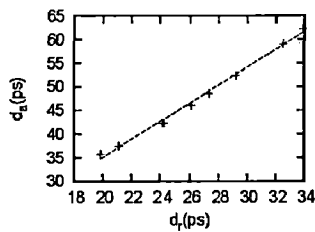


Figure 2.4: Delay with realistic input versus delay with ramp input

Say, for an inverter with a given size we know for load capacitances C_{l1} and C_{l2} the simulated values of delay for two values of t_{rin} each. Say that we also know

$t_{rb}(C_{l1})$ and $t_{rb}(c_{l2})$ from simulations. From these values we can deduce K_1, K_2, K_3 and $t_{rb}(c_l)$ for any values of C_l for the inverter using observations of Section 2.3.

The output signal of the logic gate acts as input signal for the logic gate that follows. Hence we need to find the relation between output signal transition time (T_{rout}) of the logic gate (T_{rout}) and the values of t_{rin} and C_l . In the next chapter we discuss these relations and verify them using HSPICE.

Chapter 3

Model for Output Signal Transition Time (T_{rout})

3.1 Overview

In the previous chapter we discussed our model for delay and its region of validity. We have also obtained the values of the model coefficients using simulation data. We use our delay model to optimize standard cell characterization in its region of validity. For this, we also need a simple and accurate model of gate's output transition time (T_{rout}) in the region of validity of our delay model. This is again a semi-empirical model which is physics based and the model coefficients are obtained by fitting the model on the simulated data. In this work we use HSPICE simulations of 45nm PTM CMOS technology. In the next section, we will explain our approach in deriving the T_{rout} model.

3.2 T_{rout} Model

The input transition time of a logic gate in a data-path is the output transition time of its driver stage. Therefore, an LUT of output transition time T_{rout} of logic gates expressed as a function of t_{rin} and C_l is also required in standard cell characterization data for STA. In this subsection, we express T_{rout} of an inverter as a simple function of t_{rin} and C_l for $t_{rin} \leq t_{rb}(C_l)$. In this work, we denote output's 80%-20% transition by T_{rout} .

There are two cases for output transition: First, where the entire 80-20% output transition occurs after $t = T_R$ and second, where a part of the output transition

occurs for time $t \leq T_R$. We analyze the two cases as follows:

Case 1: $V_{out}(T_R) \geq 0.8V_{DD}$: The 80-20% output transition occurs after the inverter's input voltage V_{in} has reached V_{DD} . This can be clearly seen in Fig. 3.1. Therefore, the load ($C_l + C_p$) is discharged from a voltage $0.8V_{DD}$ to $V_{DD} - V_{TH}$

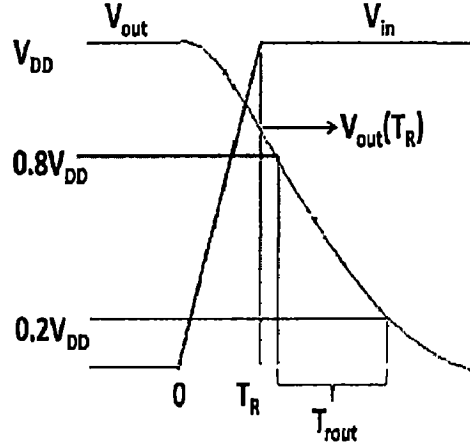


Figure 3.1: The output transition of an inverter in Case 1 (Not drawn to scale)

through the NMOS drive current I_{ON} . From $V_{DD} - V_{TH}$ to $0.2V_{DD}$, the device operates in linear regime. Therefore, T_{rout} varies linearly with C_l . All the parameters C_p , I_{ON} and the transistor's equivalent resistance in linear regime are proportional to W_n . Therefore, We make the following observations:

- Observation 5: The slope of variation of T_{rout} with C_l is proportional to $1/W_n$.
- Observation 6: Intercept is independent of $1/W_n$.

We validate these observations 5 and 6 using SPICE simulation data in Figures 3.3(a)- 3.3(b).

We denote the value of T_R for which $V_{out}(T_R)=0.8V_{DD}$ by T'_{rb} . From Equation of $V_{out}(T_R)$ we observe that

$$0.8V_{DD} = V_{out}(T'_{rb}) \quad (3.1)$$

$$= V_{DD} - S_T T'_{rb} / (C_l + C_p) \quad (3.2)$$

Therefore, T'_{rb} varies linearly with C_l . We also observe that the slope of variation of T'_{rb} varies linearly and intercept stays constant with W_n . We verify the linear

variation in Figure 3.4.

Case 2: $V_{out}(T_R) \leq 0.8V_{DD}$: This happens for values of $T_R > T'_{rb}$, as can be observed from the discussion on Case 1. Figure 3.2 gives more insight in to the Case 2. From

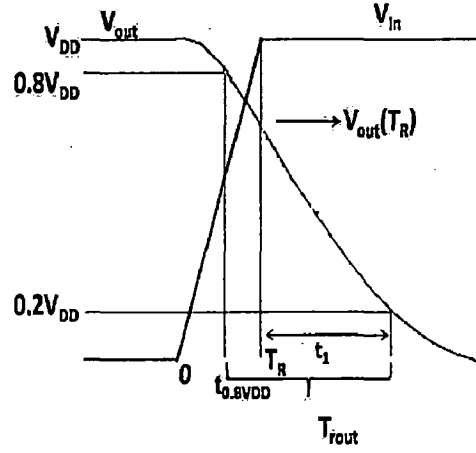


Figure 3.2: Inverter output transition for estimating T_{rout} in Case 2 (Not drawn to scale)

the figure 3.2 we can write

$$T_{rout} = T_R - t_{0.8V_{DD}} + t_1 \quad (3.3)$$

We need to find the expressions for $t_{0.8V_{DD}}$ and t_1 for the T_{rout} model to be complete. t_1 is the time during which the output transition happens from $V_{out}(T_R)$ to $V_{DD} - V_{TH}$. For a given value of $V_{out}(T_R)$, t_1 is proportional to C_l , as we explained in Case 1. Assuming that the output transition from $V_{out}(T_R)$ to $0.2V_{DD}$ is linear with time,

$$t_1 = \alpha (C) \frac{V_{out}(T_R) - 0.2V_{DD}}{0.8V_{DD} - 0.2V_{DD}} \quad (3.4)$$

$$= \alpha (C) \frac{V_{DD} - \frac{S_T T_R}{C} - 0.2V_{DD}}{0.6V_{DD}} \quad (3.5)$$

$$(3.6)$$

where α is inversely proportional to the device width (W_n). From time $t=0$ to $t=t_{0.8V_{DD}}$, the output voltage change across the load capacitor is $0.2V_{DD}$. Charge lost within this interval is $0.2V_{DD}C$.

$$0.2V_{DD}C = \int_0^{t_{0.8V_{DD}}} I_{ds} dt \quad (3.7)$$

$$= \mu C_{ox} \frac{W_n}{L} \int_0^{t_{0.8V_{DD}}} V_{GS} - V_{TH} dt \quad (3.8)$$

$$0.2V_{DD}C = \mu C_{ox} \frac{W_n}{L} \int_0^{t_{0.8V_{DD}}} \frac{V_{DD}}{T_R} t dt \quad (3.9)$$

$$0.2V_{DD}C = \mu C_{ox} \frac{W_n}{L} \frac{V_{DD}}{2T_R} t_{0.8V_{DD}}^2 \quad (3.10)$$

$$t_{0.8V_{DD}} = \sqrt{\frac{0.4CLT_R}{\mu C_{ox} W_n}} \quad (3.11)$$

We assume velocity saturation in this expression and also $V_{GS} \gg V_{TH}$. Making use of Equation 3.11 and Equation 3.5 in Eqn. 3.3, we obtain the final expression for T_{rout} given by

$$T_{rout} = aT_R + b\sqrt{T_R} + c \quad (3.12)$$

From Eqn. 3.12 the following observations can be seen,

1. The model coefficient 'a' is independent of load capacitance and width of NMOS device (W_n).
2. The model coefficient 'b' is directly proportional to $\sqrt{C_l}$ and inversely related $\sqrt{W_n}$.
3. The model coefficient 'c' is directly proportional to load capacitance and inversely proportional to width of NMOS device.

The coefficient values can be obtained fitting the model Eqn. 3.12 on the simulated data. After obtained the coefficients the observations enumerated above can be checked. In the next section we discuss the validity of the the T_{rout} model on the simulated data obtained from HSPICE.

3.3 Verification of T_{rout} model using HSPICE

Initially we verify the model presented in Case 1. In Case 1 we observed that T_{rout} before t'_{rb} is a linear function of load capacitance and inversely related to the width

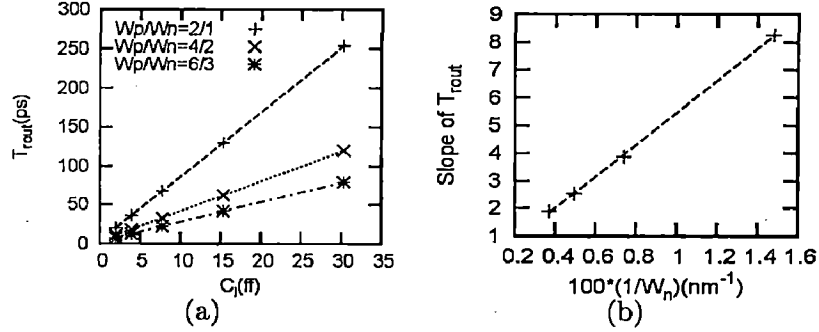


Figure 3.3: (a) Variation of T_{rout} with C_l before $T_{rb'}$ for different NMOS widths, (b) Variation of slope of T_{rout} .

of NMOS device. This is verified in Fig. 3.3.

For a given load capacitance, as t_{rin} is increased the output voltage $V_{out}(T_R)$ drops, Case 1 does not hold when t_{rin} becomes equal to t'_{rb} . Confirming our derivation of t'_{rb} model, it varies linearly with the load capacitance and inversely with the NMOS device width. This is verified in Fig. 3.4. We now verify our derivation of

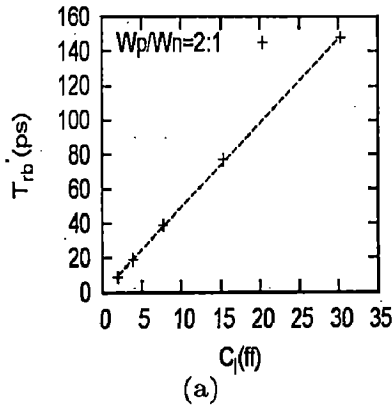


Figure 3.4: (a) Variation of T'_{rb} with Load Capacitance(C_l), (b) Variation of T_{rout} with t_{rin} for $t_{rin} \leq t_{rb}$.

T_{rout} model in Case 2. In Fig. 3.5 and Fig. 3.6 the variation of model coefficients with respect to the load capacitance and NMOS device width W_n , respectively confirm with our predictions.

In Chapter 2 we have derived a linear delay model and validated it, where as in

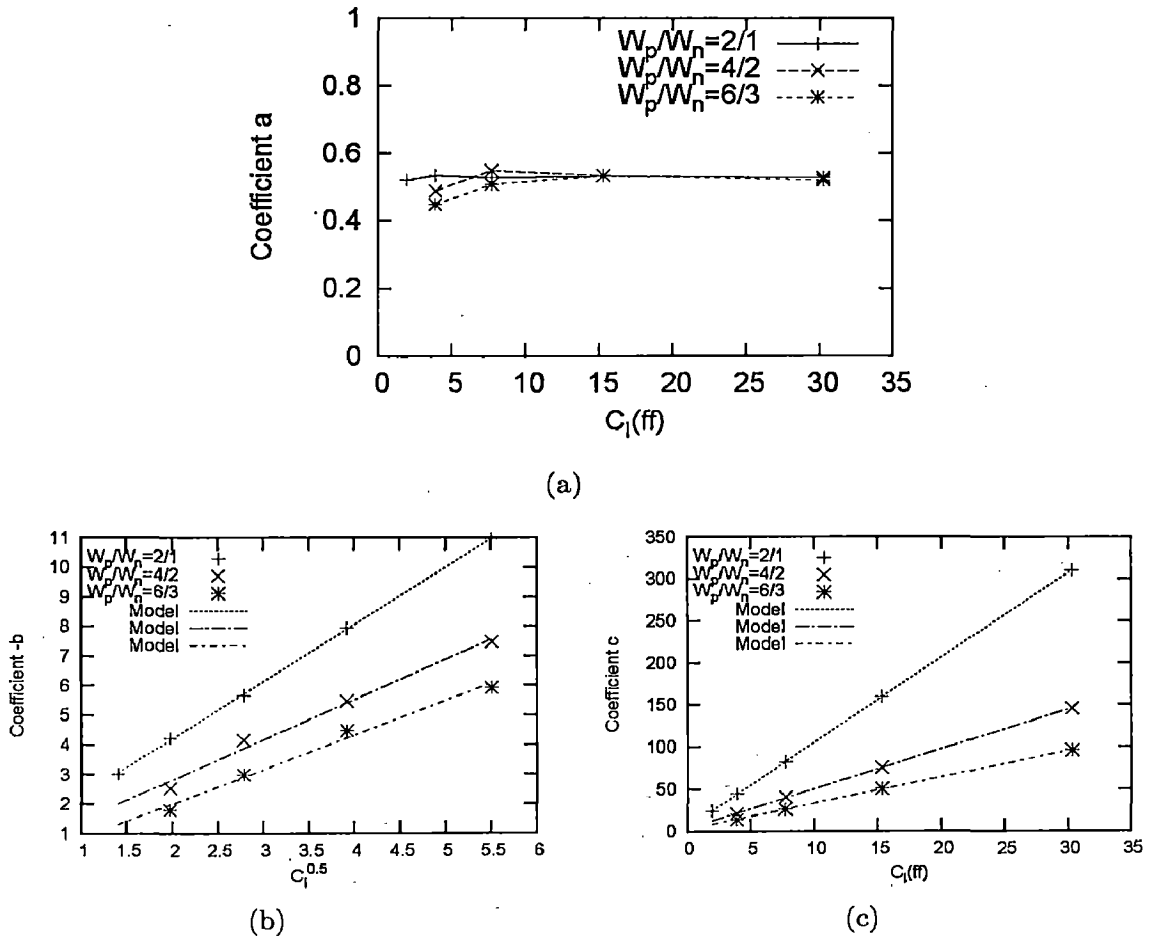


Figure 3.5: Variation of T_{rout} with load capacitance (C_l).

this Chapter we have derived a simple model for output signal transition time in (C_l, T_R) range in which our delay model is valid. The coefficients of the model are verified using HSPICE simulated data. In the next chapter we show a similar model holds for other logic gates such as NAND and NOR.

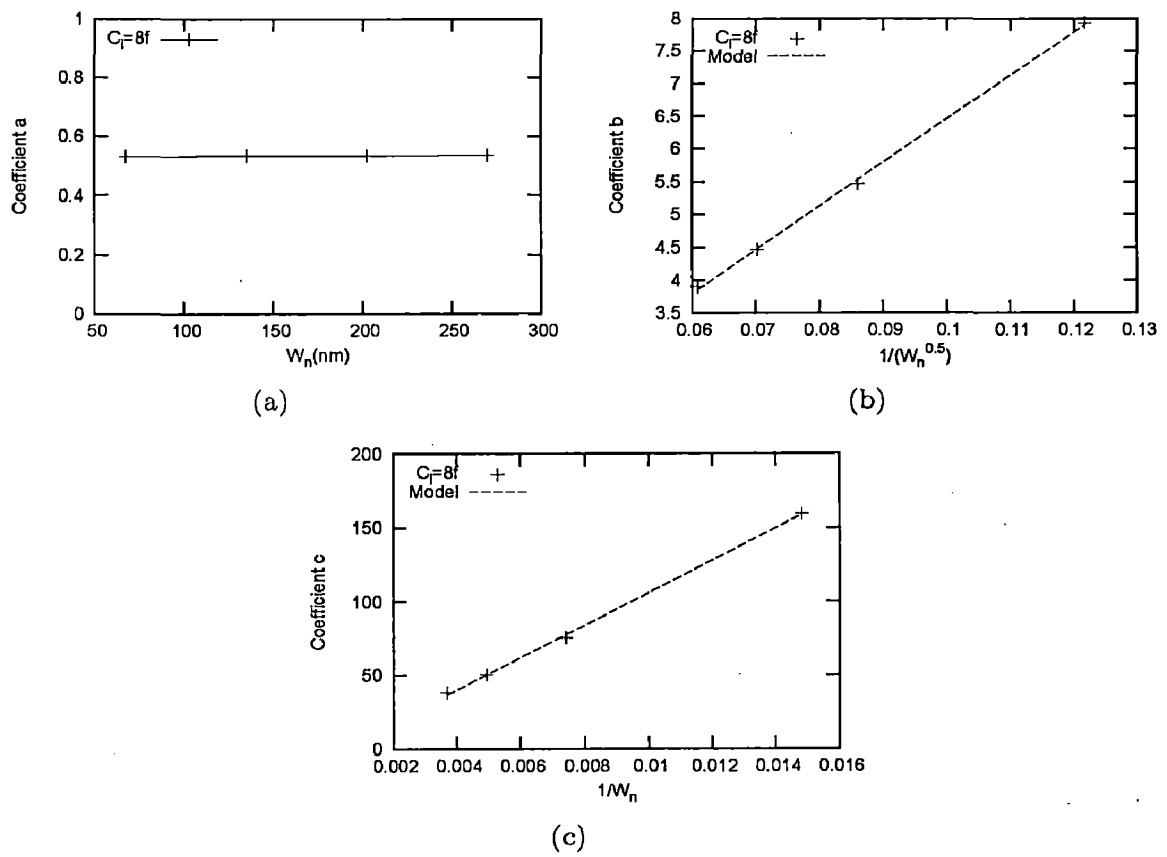


Figure 3.6: Variation of T_{rout} coefficients with device width (W_n).

Chapter 4

Extension of the Models to Complex Gates

4.1 Overview

In Chapters 2 and 3, we have derived and validated semi-empirical models for CMOS inverter delay and output signal transition time. Till this point we have only considered inverters of various sizes for analysis. However, in this chapter we would like to show that the similar models are valid for gates such as NAND and NOR.

4.2 Linear Delay Model for NAND gate

We first design a NAND gate with sizing of transistors such that its rise/fall transitions are matching each other. In line with [logical effort], we call an inverter with equal rise/fall delays as being its equivalent. The output rise and fall transitions due to switching of each input is similar to an inverter. Therefore, we propose that a delay model, similar to our inverter DM in its form, would be valid for NAND gate too. Then we compare results of all model coefficients of an equal inverter with that of the NAND gate. The CMOS NAND gate schematic is shown in Fig. 4.1. The linear delay model which was derived in Chapter 2 is given by

$$Delay = K_1 C_l + K_2 T_R + K_3 \quad (4.1)$$

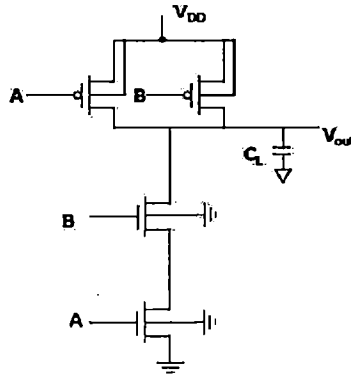


Figure 4.1: A CMOS NAND gate circuit diagram.

We obtain coefficients K_1 , K_2 and K_3 for NAND gates using HSPICE simulation data. This is shown in Fig 4.2.

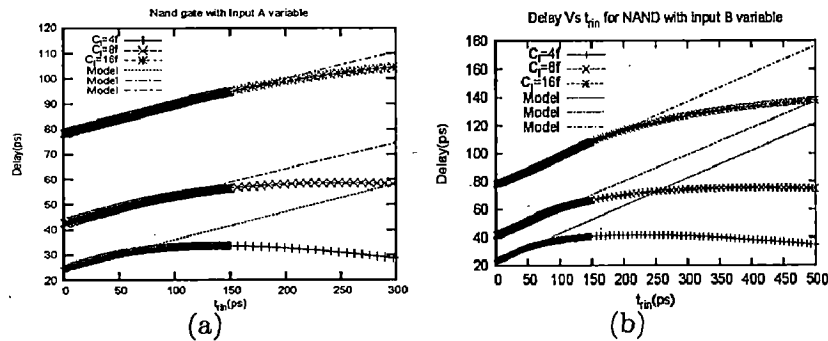


Figure 4.2: (a) Input's $B=V_{DD}$ and $A=0$ to V_{DD} , (b) Input's $A=V_{DD}$ and $B=0$ to V_{DD} .

At a given load capacitance, the coefficient K_2 will act as slope of equation and $K_1C_l + K_3$ will be its intercept. In Fig 4.3 we compare the slope and intercept of an inverter with that of NAND for various sizes. As the ratio of slope of NAND to that of an inverter is constant, showing that we can obtain the slope of NAND from an inverter.

The delay model is valid only in a particular range of (T_R, C_l) space. We have already discussed that for a given load the delay model is valid with an upper bound of T_R denoted by t_{rb} which we call breakpoint. The form of t_{rb} is given below

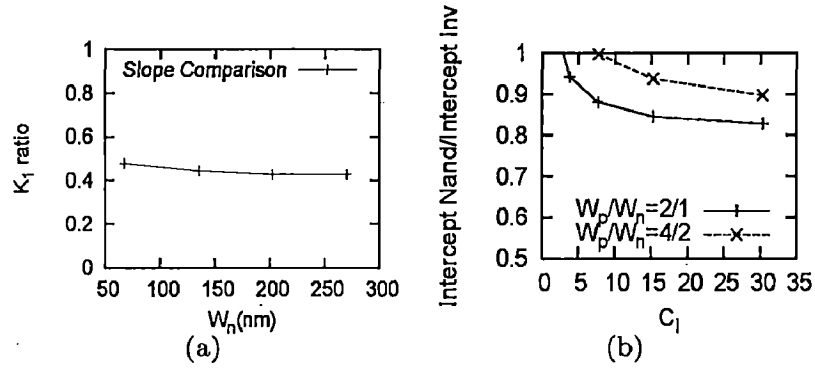


Figure 4.3: (a) Comparison of slope in delay model of NAND to Inverter, (b) Comparison of intercept in delay model of NAND to Inverter.

$$t_{rb} = \frac{V_{TH}C_l}{S_T} + \frac{C_pV_{TH}}{S_T} \quad (4.2)$$

We observe that in Equation 4.2, t_{rb} varies linearly with the load capacitance. In Fig. 4.4 we show the variation of breakpoint with load capacitance for the NAND gates of several sizes.

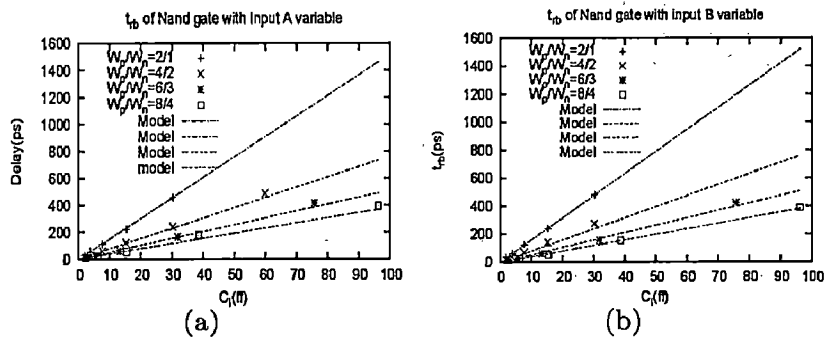


Figure 4.4: Variation of t_{rb} with load capacitance for several device widths.

The slope in Equation 4.2 is obtained for both NAND gates and their equivalent inverters. Fig 4.5(a) shows the comparison of slope values for several device widths. At this point we have compared all the coefficients in delay model of an inverter to that of a NAND gate. In the next section observe that the T_{rout} model derived for inverter is valid for NAND gate too.

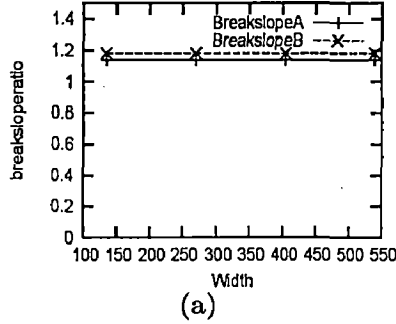


Figure 4.5: Slope of t_{rb} with respect to that of an inverter.

4.3 T_{rout} for NAND gate

In Chapter 3, we have derived a semi-empirical model for T_{rout} . The model equation is given by

$$T_{rout} = aT_R + b\sqrt{T_R} + C \quad (4.3)$$

As there are two input variables in NAND we denote these inputs by A and B. Here we have two cases

Case 1: A is changing from 0 to V_{DD} and B is kept constant at V_{DD} .

Case 2: B is changing from 0 to V_{DD} and A is kept constant at V_{DD} .

In both these cases we obtain the model coefficients a, b, c in the T_{rout} model. We observe that our prediction regarding model coefficients' variation with the load capacitance and the width of the device hold for NAND gate too. We then obtain model coefficients of Equation 4.3 using simulated data. The variation of these model coefficients with C_l is shown in Fig. 4.6.

In Fig. 4.7 we show that all our Observations in Case 1 regarding the variation model coefficients with the device width are valid for NAND gates too.

In Fig. 4.8 and Fig. 4.9, we show that the corresponding variations of the model coefficients with the load capacitance and the device width confirm with our Observations for Case 2 as predicted. Thus the observations are verified for Case 2 described in this section.

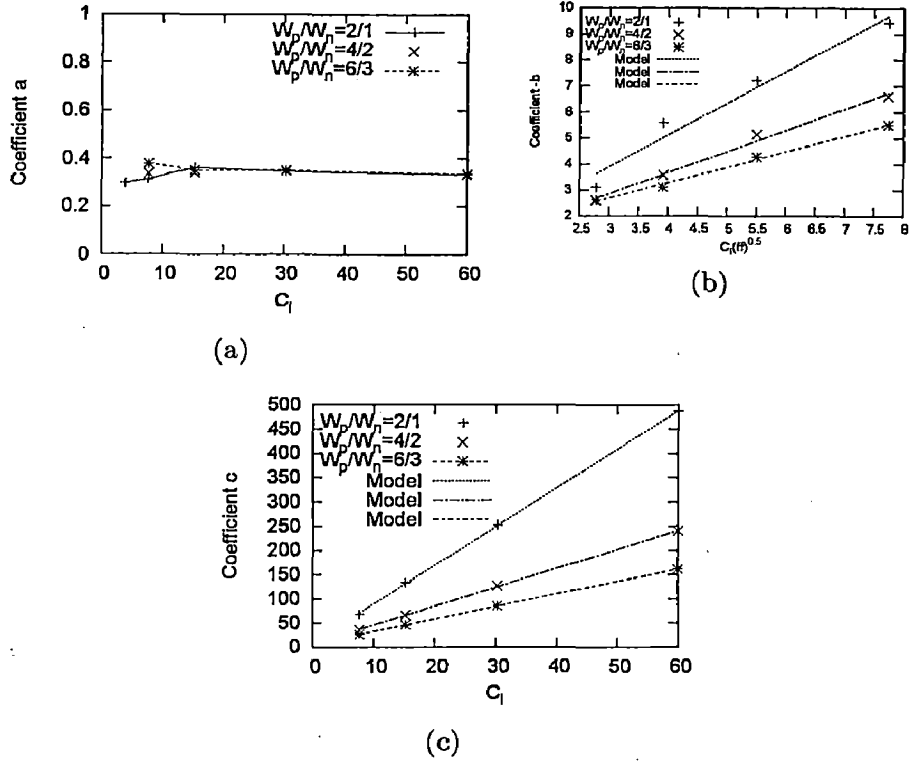


Figure 4.6: Variation of model coefficients with the load capacitance for NAND gate with input 'A' variable

4.4 Analysis for Multistage cells (Buffer)

In the previous section we have shown that our delay model is valid for NAND gates. In this section we extend our delay model to multistage standard cells such as buffers.

We consider a two stage buffer, input signal is applied to the first stage and the load capacitance is at the output of second stage. The load capacitance of the 1st stage of the buffer is the input gate capacitance of 2nd stage, which is almost constant (if we assume that the load capacitance of the 2nd stage has little impact). If this load is smaller than a certain value, $t_{rb}(C_l)$ of 1st stage denoted by t_{rb1} is a constant as described in Chapter 1. Therefore for $T_R < t_{rb}(C_l)$, delay of 1st stage, denoted by Delay1 is a linear function of T_R . Now we are interested in T_{rout} of 1st stage which acts as T_R for 2nd stage. We use our result from Chapter 1.

$$V_{out}(T_R) = V_{DD} - \frac{S_T T_R}{C} \quad (4.4)$$

For typical buffer 2nd stage loads Case 1 of T_{rout} model is valid. In this case T_{rout} is

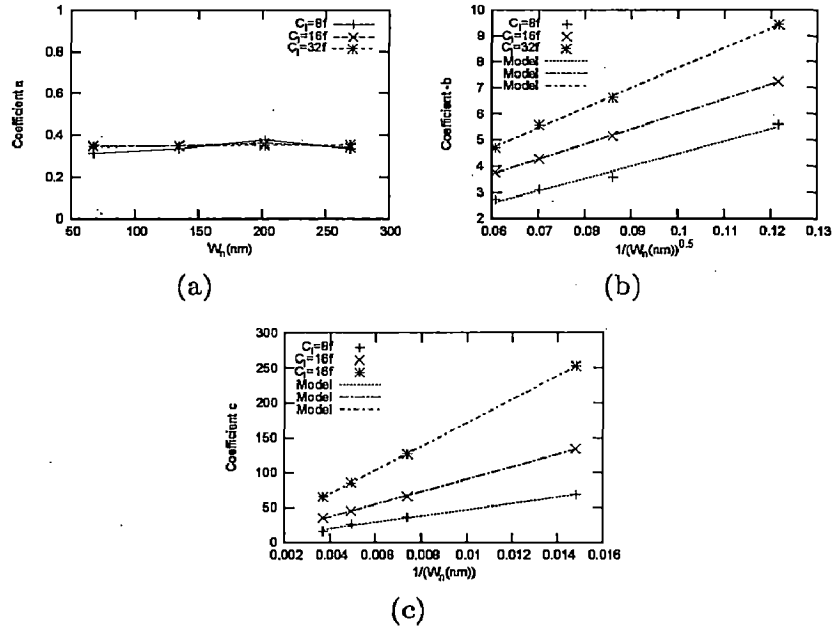


Figure 4.7: Variation of model coefficients with the gate size for NAND gate with input 'A' variable

a linear function of C_i . As C_i is fixed, T_{rout} of 1st stage must be constant.

For 2nd stage, its input T_R is fixed and hence for $C_i \geq C_{ib}$, delay of 2nd stage denoted by Delay2 is a linear function of C_i .

The total delay of the buffer for $T_R \leq t_{rb1}$ and $C_i \geq C_{ib}$ is linear function of T_R and C_i .

$$TotalDelay = Delay1 + Delay2 \tag{4.5}$$

$$= aT_R + c1 + bC_i + c2 \tag{4.6}$$

$$= aT_R + bC_i + c \tag{4.7}$$

All the observations are verified using HSPICE simulation data as shown in Fig. 4.10.

Finally all the models that are developed in Chapter 2 and 3 and all the observations that are made in the previous chapters for the inverter are valid for the NAND gate and Buffer too. In the next chapter, we see the reduction in the characterization effort during the Look Up Table (LUT) generation by making use of the semi-empirical models in the Chapters 2-4.

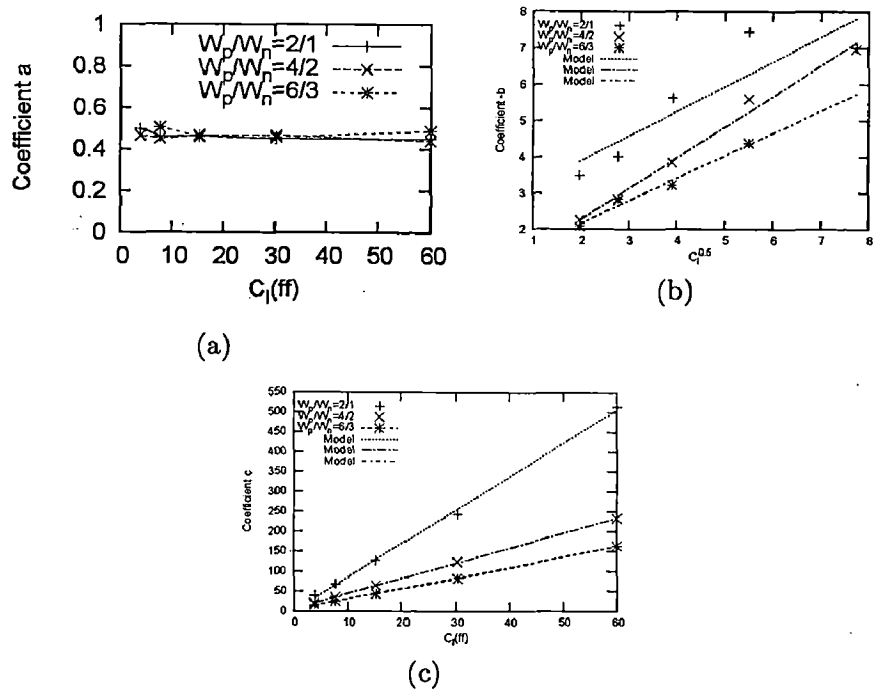


Figure 4.8: Variation model coefficients with the load capacitance for NAND gate with input 'B' variable

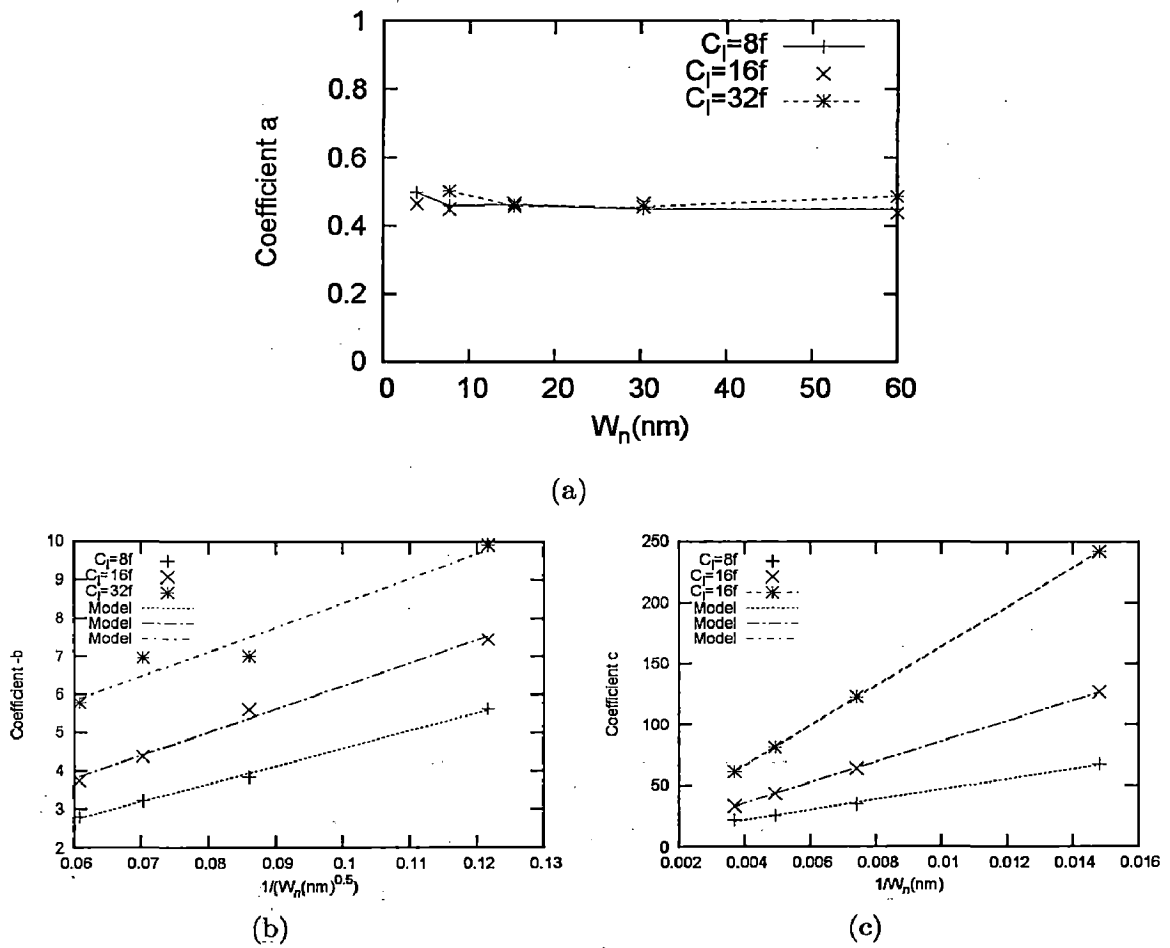


Figure 4.9: Variation model coefficients with the gate size for NAND gate with input 'B' variable

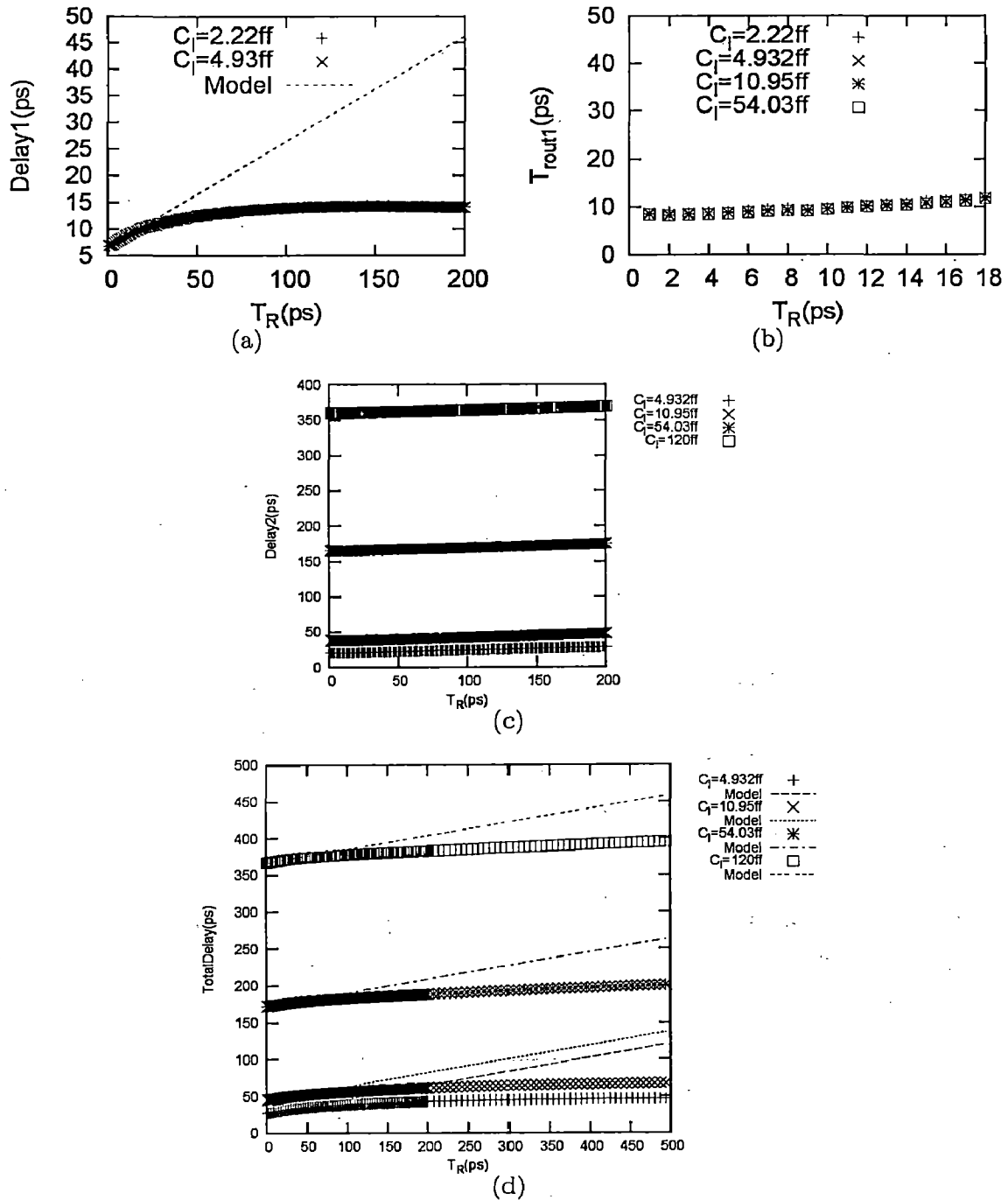


Figure 4.10: (a) Delay of 1st stage, (b) Output transition of 1st stage, (c) Delay of 2nd stage, (d) Total delay of two stage buffer

Chapter 5

Impact of technology scaling on the delay model

5.1 Overview

For a delay model to be useful in STA, it should maintain its accuracy with technology scaling. In this chapter we show using HSPICE simulations that our novel delay model is valid with technology scaling. We perform our analysis at 32nm technology node.

5.2 Delay Model Verification at 32nm Technology node

¹In this section, we verify the validity of our delay model and our observations regarding its coefficients using HSPICE simulations at 32nm technology node². Our linear delay is given by

$$Delay = K_1 C_l + K_2 T_R + K_3 \quad (5.1)$$

We extract the model coefficients using 32nm inverter HSPICE simulations. In Fig. 5.1(a) we plot the delay variation with the input signal transition time at a given load capacitance (C_l). We observe here that the delay model perfectly fits up to a value of $T_R = t_{rb}$. In Fig. 5.1(b) we plot the delay variation with the load

¹This work was done jointly with Baljit Kaur, Ph.D. student in EC Dept. of IIT Roorkee

²Device models are obtained from <http://www.eas.asu.edu/~ptm/>

capacitance at a given input signal transition time. We observe here that the delay model perfectly fits after a value of $C_l=C_{lb}$.

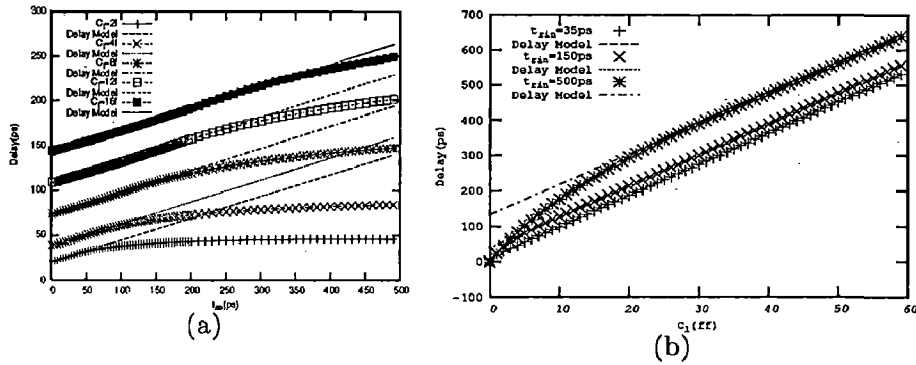


Figure 5.1: (a) Delay variation with t_{rin} , (b) Delay variation with C_l . Points are simulated data and lines are fitting of our delay model.

For a given load capacitance, the variation of delay with t_{rin} has a slope of K_2 . It was observed that K_2 is independent of inverter size represented by W_n . Fig 5.2 verifies this statement at 32nm technology node.

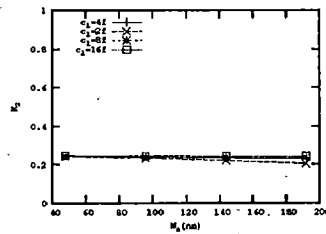


Figure 5.2: Variation of coefficient K_2 in our delay model with device width.

The equation for t_{rb} is derived in Chapter 2 is given by

$$t_{rb} = \frac{C_l V_{TH}}{S_T} + \frac{C_p V_{TH}}{S_T} \quad (5.2)$$

In Fig. 5.3 we show that t_{rb} is varying linearly with the load capacitance, and the slope of the equation denoted by S_{trb} is inversely related to the device width (W_n). We now verify whether the model for T_{rout} developed at 45nm technology is independent of technology node. In the next section we look at the variation of coefficients of the T_{rout} model.

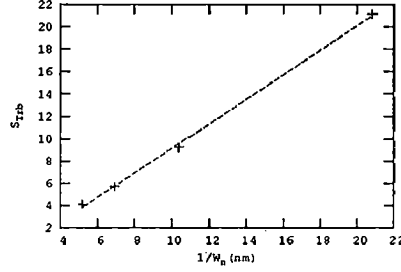


Figure 5.3: Variation of slope of t_{rb} with device width.

5.3 T_{rout} model independent of technology node

In chapter 3, we have identified two cases for T_{rout} modeling at 45nm technology node. The model for T_{rout} for the two cases at 32nm node are:

Case1: Here the entire transition of the output response of an inverter happens after $t = T_R$. We are interested on 80% to 20% transition time. During this regime NMOS operates in linear regime and the transition time is given by

$$T_{rout} = \alpha(C_l + C_p) \quad (5.3)$$

T_{rout} varies linearly with load capacitance and inversely with the device width W_n .

Case2: In this case part of the output transition happens prior to input reaching T_R . The model for T_{rout} as already developed in Chapter 3 is given by

$$T_{rout} = aT_R + b\sqrt{T_R} + c \quad (5.4)$$

The relationship of model coefficients a, b and c with the C_l and W_n are:

- a is independent upon the device width (W_n)
- a is independent upon the load capacitance
- b is proportional to $\sqrt{C_l}$
- b is inversely proportional to $\sqrt{W_n}$
- c is directly proportional to C_l
- c is inversely proportional to W_n

All the above observations are verified in Figures 5.4 and 5.5. This indicates that the T_{rout} model is independent of technology node.

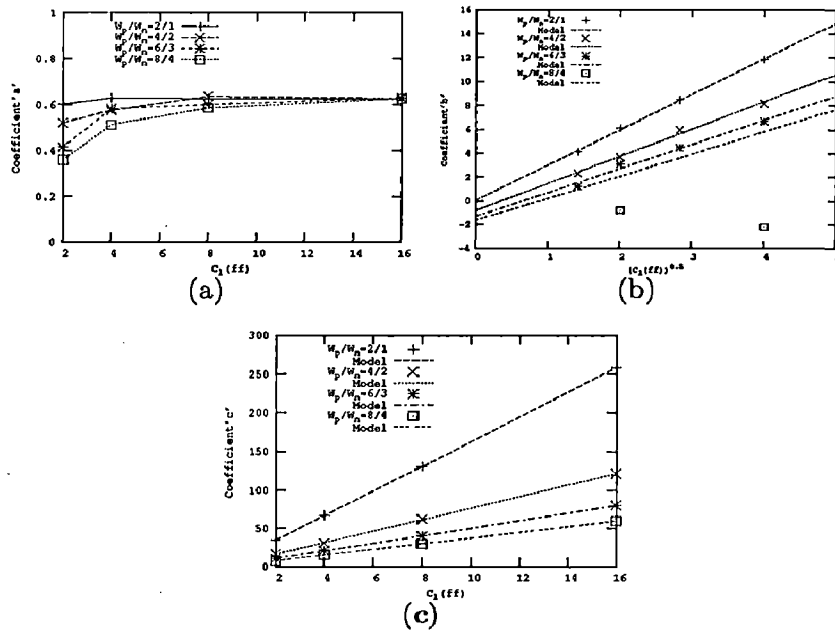


Figure 5.4: Variation of T_{rout} model coefficients with load capacitance.

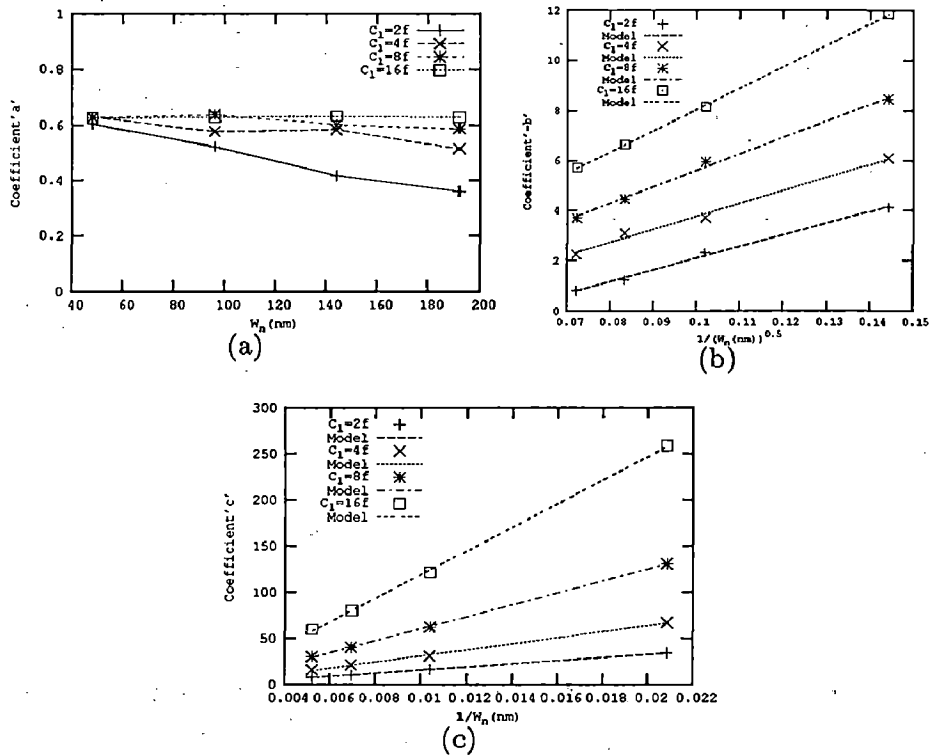


Figure 5.5: Variation of T_{rout} model coefficients with device width.

Chapter 6

Results and Conclusion

6.1 Overview

In previous chapters, we derived and validated simple models for the delay and output signal transition time of a logic gate. As the aim of the work is to reduce the characterization effort of LUT, we make use of the models in optimizing the process of generation of LUT for library characterization. Sample library consists of important logic gates such as an inverter, NAND, buffers of various sizes. In this chapter we look at the saving in simulation time using our method to generate the LUTs. Finally, we compare the delay values obtained from traditional LUTs, our LUTs and HSPICE for cascaded chain of logic gates.

6.2 LUT using delay model

We have explained in Chapter 1 the Look Up Table (LUT) approach for delay estimation in STA. For example consider a LUT of 9x9 matrix. The number of points in this LUT is 81. These points are obtained by using SPICE. Currently, industry obtains delays of all the points making use of SPICE and we call this LUT as traditional LUT. In optimized LUT or LUT using delay model, we make use of the simple models derived in chapter 2-4. However, for the points where the delay model is not valid we use HSPICE simulations. In the next section we enumerate the savings in the number of points of LUT obtained using simulations.

6.3 Reduction in number of simulations

In this section we use models derived in Chapters 2 and 3 during LUT characterization. Hence, the number of SPICE simulations required to characterize a standard cell library decreases. In our sample library we have built the LUTs for inverters, NAND gates and Buffers with ¹FO2, FO3 and FO4. We perform our experiments for LUTs with sizes 6x6, 7x7, 8x8 and 9x9. In Tables 6.1-6.4 we show savings in the LUT points requiring the simulation.

Size	6x6	7x7	8x8	9x9
Reference	26	35	44	58
2	24	32	43	55
3	23	31	40	52
4	22	39	39	48
5	22	30	39	48
12	22	29	37	46
16	21	29	37	44
24	21	27	34	43
48	19	25	33	41
64	19	25	33	41
128	18	25	33	39
256	15	20	28	38

Table 6.1: Number of savings in HSPICE simulation while characterizing inverter library using our DM.

NAND Size	6x6	7x7	8x8	9x9
1	25	28	40	54
2	24	28	40	45
3	24	28	40	45
4	24	28	40	45

Table 6.2: Simulation savings in NAND gate with input 'A' variable.

¹Where FO denotes fanout

NAND Size	6x6	7x7	8x8	9x9
1	26	35	45	57
2	12	16	21	28
3	14	19	23	28
4	15	21	28	35

Table 6.3: Simulation savings in NAND gate with input 'B' variable.

Second stage	6x6	7x7	8x8	9x9
Twice	12	21	24	27
Thrice	12	21	24	36
Four	12	21	24	36

Table 6.4: Simulation savings in buffers.

6.4 Accuracy of LUT generated using the model

In this section we consider inverter of several sizes, load capacitance and input signal transition times. Using the model and model coefficients derived in previous chapters we obtain the delay of logic gate. Then, we also obtain delays using HSPICE. Fig 6.1 shows the comparison of both the delays.

In Fig. 6.1, we have a straight line passing through the origin indicating that the model delay is very close to the HSPICE delay. In the next section we compare the delay calculation from optimized LUT and Traditional LUT for a cascaded chain of inverters.

6.5 Delay comparison for a cascaded chain of inverters

In this section we compare delays obtained using the optimized LUT and traditional LUT. For this, we use simulations of a cascaded chain of inverters. The inverter size

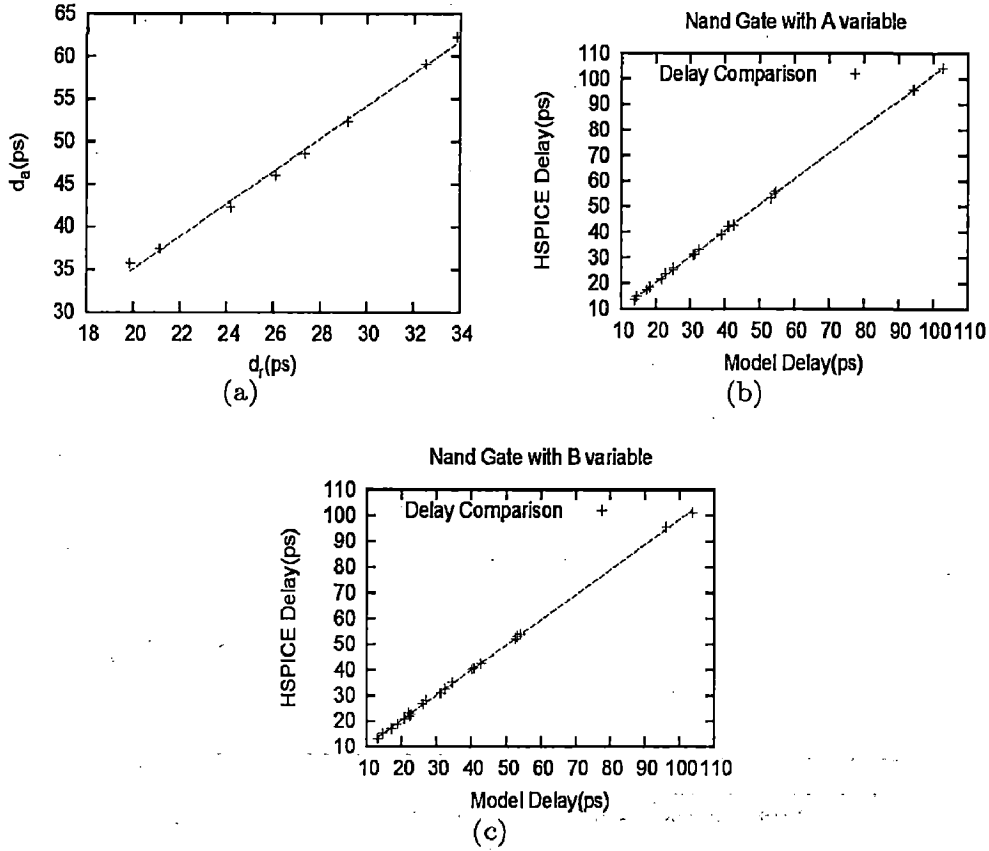


Figure 6.1: Comparison of delay for inverters of various size, load and t_{rin} .

in the chain is chosen randomly. The load, input signal transition time and inverter are also chosen randomly. Output transition T_{rout} of each stage is also obtained from the T_{rout} LUTs. Table 6.5 shows the comparison of the delay obtained using optimized LUTs, Traditional LUTs and HSPICE.

6.6 Conclusion and Future work

We show that if an upper bound on input transition time t_{rin} is followed, a simple delay model is valid for inverters, NAND and NOR gates which relates delay linearly to t_{rin} and load capacitance C_l . We also derive the relation of the delay model coefficients with inverter size W_n (assuming that the ratio of its NMOS and PMOS devices remains constant) and simple relations which express t_{rb} as a function of C_l and W_n . We derive similar relations which relate output transition time T_{rout}

No. of stages	Delay optimized LUT	Delay traditional LUT	Delay HSPICE
3	67.89ps	68.01ps	71.44ps
4	116.62ps	115.4099ps	122.7ps
5	83.67ps	85.4ps	91.34ps
11	261.64ps	255.69ps	277.3ps

Table 6.5: Delay optimized LUT, Delay traditional and Delay HSPICE for 3,4, 5 and 11 stage inverter.

to t_{rin} , C_l and W_n . We also extend this work to multi stage standard cells such as buffers. To derive these relations we did not use device currents/capacitances models. We use the topology of the gate and the charging/discharging phenomenon of the load stage. Therefore, these relations are general in nature and would not change with technology scaling. Using these relations, we show that standard cell library characterization can be done with a significantly lesser number of simulations (60% reduction) while maintaining accuracy. This is useful since numerous cycles of standard cell characterization would be needed at several Process, Voltage and Temperature (PVT) corners in deep sub-micron technologies. Another potential application of this work is in increasing the accuracy of standard cell characterization data in the form of LUT. This is because the LUT points in the region of validity of the linear delay models do not need simulations. Therefore, to increase accuracy of the LUT, simulations can be performed to obtain delay for additional points where delay is a highly non-linear function of (t_{rin}, C_l) . As a future work, we will extend the relations we obtained for the inverter to sequential circuit elements.

Bibliography

- [1] Louis Scheffer, Luciano Lavagno and Grant Martin, *EDA for IC implementation, circuit design and process technology*, CRC Press, 2006.
- [2] T. Sakurai and R. Newton, "Alpha-power law MOSFET model and its implications to CMOS inverter delay and other formulas," IEEE JSSC, pp. 584-594, April, 1990.
- [3] Feng Wang and Shie-Shen Chiang, "Scalable Polynomial Delay Model for Logic and Physical Synthesis," Proceedings of IEEE ICCD, August, 2000.
- [4] <http://www.opensourceliberty.com>
- [5] Ivan Sutherlands, Bob Sproull and David Harris, *Logical Effort: Designing Fast CMOS Circuits*, Morgan Kaufmann Publications, 1999.
- [6] JianChang, Louis G Johnson and Cheng Liu, "Piecewise Linear Delay Modeling of CMOS VLSI Circuits," IEEE IMSCAS, August, 2009.
- [7] Yangang Wang and Mark Zwolensky, "Analytical Transient Response and Propagation Delay Model for Nanoscale CMOS Inverter," IEEE ISCAS, November, 2009.
- [8] So Young Kim, Wong S.S., "Closed-Form RC and RLC Delay Models Considering Input Rise Time," IEEE Tran. on Circuits and Systems, Vol. 54, pp. 2001-2010, Sep, 2007
- [9] Kahng A.B., Masuko K., Muddu S., "Analytical delay models for VLSI interconnects under ramp input," IEEE ICCAD, pp. 30-36, November, 1996
- [10] B. Hoppe, G. Neuendorf, D. Schmitt-Landsiedel and W. Specks, "Optimization of High Speed CMOS Logic Circuits with Analytical Models for

- Signal Delay, Chip Area, and Dynamic Power Dissipation," IEEE Trans. on Computer Aided Design, Vol. 9, pp. 236-247, March, 1990
- [11] N. Hedenstierna and K.O. Jeppson, "CMOS Circuit Speed and Buffer Optimization," IEEE Tran. on Computer-Aided Design, Vol. 6, pp. 270-281, March, 1987.
- [12] John K.Ousterhout, "Switch-Level Delay Models for Digital MOS VLSI," IEEE Trans. on Computer Aided Design of VLSI Circuits and Systems, Vol. 4, pp. 336-349, 1985.
- [13] Vivek Raghavan, "Submicron Delay Calculation for Accurate Timing Analysis," IEEE WESCON Conference record, pp. 124-127, November, 1995.
- [14] Jorge Rubinstein, Paul Penfield Jr. and Mark Hortwitz, "Signal Delay in RC Tree Networks," IEEE Transaction on Computer-Aided Design, CAD-2(3), pp. 202-211, July, 1983.
- [15] Jessica Qian, Satyamurthy Pullela, and Lawrence Pillage, "Modeling the effective capacitance for the RC interconnect of CMOS gates," IEEE Tran. Computer-Aided Design of VLSI Circuits and Systems, Vol. 13, pp. 1526-1535, 1994.
- [16] Jan M. Rabey, Anantha Chandrakasan and Borivoje Nikolic, *Digital Integrated Circuits A design perspective*, 2nded., Prentice Hall of India Pvt Ltd, New Delhi, 2006.

G 20079.